Conservation genomics: Speciation of the Neotropical damselfly species *Megaloprepus caerulatus* – as a model for insect speciation in tropical rainforests

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von
Wiebke Feindt, M.Sc.

2019

To my grandma and my mama.

ZUSAMMENFASSUNG

Die Neotropen sind die vielfältigste Ökoregion der Erde. Diese bemerkenswerte biologische Vielfalt hat ihren Ursprung in einer Vielzahl von Artenbildungsereignissen, die durch eine komplexe geologische Geschichte und diverse Habitatstrukturen entstanden sind. Jedoch werden durch die derzeitigen Lebensraum- und Klimaveränderungen Arten in schockierenden Zahlen ausgelöscht. Daher muss moderne evolutionäre und ökologische Forschung, Grundlagenforschung mit modernster Naturschutzgenetik kombinieren. Von zentralem Interesse sind dabei in der Evolutionsbiologie Untersuchungen von Artbildung und die Entstehung phänotypischer Neuerungen. Um diese Zusammenhänge bei fliegenden Insekten zu untersuchen, ist die weltweit größte lebende Libellenart *Megaloprepus caerulatus* (Odonata: Zygoptera, Pseudostigmatidae) ein hervorragender Modellorganismus. Als einziger Vertreter der Gattung, hat *M. caerulatus* ein großes Verbreitungsgebiet von Mexiko bis Peru, jedoch aber eine kleine und konservierte ökologische Nische. Zudem ist *Megaloprepus* als Waldspezialist auf intakte, alte Regenwälder und wassergefüllte Baumlöcher angewiesen, um stabile, aber kleine Populationsgrößen zu erhalten. Seit über 150 Jahren wurde Artbildung in dieser Gattung wiederholt diskutiert. Infolge geringer aber regional begrenzter morphologischer Unterschiede, der engen ökologischen Nische und dem starken Wandel der neotropischen Ökoregion über die Zeit, erscheint dies als wahrscheinlich.

Im Mittelpunkt dieser Arbeit steht die Untersuchung der Auswirkungen von Paläogeographie und ökologischen Veränderungen auf Artbildungsmechanismen und phänotypische Veränderungen in der Gattung *Megaloprepus*. Dazu werden die Populationsstrukturen und Artgrenzen untersucht sowie neue Erkenntnisse in der RNA-Sequenzierung als Grundlage für vergleichende Studien im großen Maßstab präsentiert.

In einer ersten Studie wurden zwei mitochondriale Sequenzmarker und ein Mikrosatelliten-System verwendet, um die Populationsstruktur und die genetischen Verwandtschaftsverhältnisse von vier Populationen von Mexiko bis nach Panama zu untersuchen. Dabei zeigten die Ergebnisse eine relativ geringe genetische Diversität innerhalb der Populationen, jedoch eine starke Differenzierung zwischen Populationen, die die Hypothese der Artbildung unterstützen. Es folgte eine umfassende biogeographische Studie, um diese Hypothese zu verneinen oder zu verifizieren. Dabei wurden Proben aus 11 Museumssammlungen und neu gesammeltes Material aus 14 Populationen von Mexiko bis Peru gemeinsam in populations- und phylogenetischen Untersuchungen, Verbreitungs-modellierungen, bei Vergleichen der ökologischen Nischen und morphologisch analysiert. Die Ergebnisse bestätigten eindeutig die Artbildungshypothese und enthüllten vier Arten innerhalb der Gattung *Megaloprepus*. Zeitlich konnte die Auftrennung der Arten zum einen auf die Anhebung der Anden (10-8 Mya) und zum anderen mit der nördlichen Ausbreitung nach der endgültigen Schließung des Isthmus von Panama (3-2 Mya) in Verbindung gebracht werden. Die heutige Verbreitung der vier *Megaloprepus*-Arten ist stark begrenzt und kann durch die klimatischen Bedingungen im Pleistozän sowie die aktuelle

Habitatstruktur erklärt werden. Noch interessanter ist der zugrunde liegende Mechanismus der Artbildung. Starke Ähnlichkeiten der ökologischen Nischen deuten auf einen phylogenetischen Nischenkonservatismus hin und folglich müsste der Mechanismus der Artbildung nicht auf Anpassung beruhen (*non-adaptive speciation*). Aber dennoch sind die beobachteten Unterschiede der Flügelmusterung eine evolutionäre Neuerung, die höchstwahrscheinlich auf die Umwelt zurückzuführen ist, sich aber jetzt unter sexueller Selektion befindet. Die abschließende Artbeschreibung definiert alle vier Arten und illustriert unter anderem die Flügelfarbmuster und Formveränderungen der sekundären männlichen Geschlechtsorgane sowie des Prothorax. In einem ersten Versuch, die Variabilität von Artbildungsmechanismen innerhalb der Odonaten und um die signifikant unterschiedlichen Radiationsmuster der zwei Schwestergattungen *Megaloprepus* (4 Arten) und *Mecistogaster* (8 Arten) zu beleuchten, wurden drei mitochondriale Genome als wertvolle Ressource für zukünftige Forschung in der Artbildung und für phylogenetische Studien erzeugt. Der Fokus liegt dabei auf den Genen, die an der oxidativen Phosphorylierung beteiligt sind.

Diese Ergebnisse sind eine solide Voraussetzung um Artbildung, die Evolution von taxonomischen sowie genomischen Charakteren und die Radiationsmuster innerhalb der Familie der Pseudostigmatidae im Detail zu untersuchen. RNA-Seq ist dabei die Methode der Wahl für ökologisch bedeutende Organismen ohne genomische Grundlagen. Demzufolge wurde das Transkriptom eines Thorax einer einzelnen Larve erstellt. Eine hohe Vollständigkeit (93%) und die ersten Flügelgene für Odonaten wurden dabei entdeckt und dienen als wertvolle Basis für zukünftige Studien zur Flügelfärbung und Flügelentwicklung. Unter dem Aspekt, dass RNA-Seq eine häufig verwendete Methode ist und es außerdem für Publikationen obligatorisch ist die Rohdaten auf einer öffentlich zugänglichen Datenbank zu hinterlegen, wurde für diesen Prozess ein Leitfaden entwickelt. Diese ‚Guideline' enthält zwei umfassende Protokolle, in denen alle erforderlichen Schritte zum Hochladen von Daten bei dem National Center for Biotechnology Information (NCBI) erläutert werden.

RNA-Seq hat bereits das Verständnis von Anpassung, Artbildung, phänotypischer Variabilität und Populationsstrukturen revolutioniert und wird auch weiterhin zum Verständnis der Evolution beitragen. Die vielversprechendsten Ansätze sind dabei die Identifikation von neuen Transkripten und differenzielle Expressionsanalysen auch in Abhängigkeit von unterschiedlichen Umweltbedingungen. In dem Wettlauf gegen das Artensterben spielen fliegende Insekten eine bedeutende Rolle um Artbildung sowie die genetischen und die genomischen Strukturen von Diversität zu untersuchen.

Schlüsselwörter: Artbildung, Biogeography, RNA-Sequenzierungen, Odonata, *Megaloprepus*

ABSTRACT

The Neotropics are the most diverse ecoregion on earth. This remarkable biological diversity is associated with a great variety of speciation events through a complex geological history and habitat structure. Unfortunately, current changes to climate and habitat are erasing species at shocking rates. Consequently, modern evolutionary and ecological research must combine basic scientific research with state-of-the-art conservation genetics. In evolutionary biology, the study of speciation processes and how phenotypic novelties arise is of central interest. To approach this task in flying insects, the world's largest living odonate species, *Megaloprepus caerulatus* (Odonata: Zygoptera, Pseudostigmatidae) is an excellent model organism. *Megaloprepus caerulatus*, which is the only representative of its genus, has a wide distributional range, from Mexico to Peru, but a narrow and conserved ecological niche. As a forest specialist *Megaloprepus* is dependent on intact old growth rain forests and water filled tree holes to maintain stable but small population sizes. In the last 150 years' speciation processes in this genus were often under discussion. Because of small but regionally restricted morphological differences, the narrow ecological niche and the continuous conversion of the Neotropical ecoregion over time, speciation seems probable.

The central focus of this thesis is to study the effects of paleogeography and ecological changes over time on speciation and phenotypic changes in *Megaloprepus*. Therefore, the population genetic structures and species boundaries are studied, and new insights into RNA-Sequencing (RNA-Seq) are presented as a foundation for large-scale comparative studies.

A first study uses two mitochondrial sequence markers and a panel of microsatellites to investigate the population genetic structure of four *Megaloprepus* populations ranging from Mexico to Panama. The results showed relatively low genetic diversity within populations, but a strong differentiation among populations, supporting a speciation hypothesis. A comprehensive biogeographic study followed to falsify or verify this hypothesis. Samples from 11 museum collections and newly collected material across 14 populations from Mexico to Peru were analyzed simultaneously by applying phylogenetics, population genetics, species distribution models, niche comparisons and morphometrics. The results unambiguously proved the speciation hypothesis and revealed that the genus *Megaloprepus* consists of four species. Hereby the estimated diversification times suggest that the species splits were associated with the Andean uplift (10-8 Mya) and migration events following the closure of the Isthmus of Panama (3-2 Mya). The current distribution ranges of the four *Megaloprepus* species are restricted and can be explained by Pleistocene climatic variations as well as by today habitat structure. Even more interesting is the underlying mode of speciation. A strong niche similarity indicates phylogenetic niche conservatism and consequently sets the speciation mode to 'non-adaptive'. However, currently observable divergence in wing patterns is an evolutionary novelty, which are most likely related to the environment but now under sexual selection. The final species description covers all four species, including wing coloration patterns, and variation of shape in the male secondary

sexual organs and the prothorax. In a first attempt to approach the great variety of speciation patterns among odonates and to reveal the significantly different radiation patterns of the two sister genera *Megaloprepus* (4 species) and *Mecistogaster* (8 species), three mitochondrial genomes were generated as a valuable resource for future speciation and phylogenetic studies. Hereby, the focus will be on the genes involved oxidative phosphorylation.

These results are a solid prerequisite to study speciation, the evolution of taxonomic/genomic key characters and radiation patterns within the family of the Pseudostigmatidae in detail. RNA-Seq is the method of choice for studying ecologically important organisms that lack genomic resources. Consequently, the transcriptome of a single larval thorax is presented. This transcriptome has a high level of completeness (93%) and provides the first reported wing gene sequences for odonates and supplies a valuable resource for future studies on wing coloration and wing development. Furthermore, because RNA-Seq is a frequently used method and for publication it is obligatory to upload raw reads to a public database, a submission guideline for this process was developed. This guideline includes two all-inclusive protocols explaining all necessary steps to upload data to the National Center for Biotechnology Information (NCBI).

RNA-Seq has already revolutionized the understanding of adaptation, speciation, phenotypic variability and population structures, and will continue to contribute to the understanding of evolution. Novel transcript identification and differential expression analyses (also in dependence of environmental conditions) are the most promising approaches. In a race against extinction, flying insects own a significant role for studying speciation and the genomic patterns of diversity.

CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

%	Percent
°C	Degree celsius
16S rRNA	16S ribosomal RNA
app.	Approximately
AFLP	Amplified Fragment Length Polymorphism
ANTP class	Antennapedia, *Hox* genes
BLAST	Basic Local Alignment Search Tool
CAOS	Characteristic Attribute Organization System (CAOS-barcoding)
CBoL	Consortium for the Barcode of Life
cf.	*confer* (compare)
cm	Centimetre
CO1	Cytochrome c oxidase subunit I (*cox1*)
ddRAD-Seq	Double digested Restriction Site Associated DNA sequencing
DNA	Deoxyribonucleic acid
Dpp	Decapentaplegic
e.g.	*exempli gratia* (for example)
et al.	*et alia* (and others)
GABI	Great American Biotic Interchange
GIS	Geographic Information System
GMM	Geometric morphometrics
Hh	Hedgehog
i.e.	*id est* (in other words)
IUCN	International Union for Conservation of Nature
LCA	Lower Central America
mtDNA	Mitochondrial DNA
Mya	Million years ago
N	Notch
ND1	NADH Dehydrogenase Subunit 1 (*nad1*)
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
OXPHOS	Oxidative phosphorylation
PCR	Polymerase Chain Reaction
pH	Scale used to specify how acidic or basic a water-based solution is.
PNC	Phylogenetic Niche Conservatism
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
rRNA	Ribosomal Ribonucleic Acid

SDM	Species Distribution Models
SNP	Single-Nucleotide Polymorphism
spp.	*species pluralis* (multiple species)
SRA	Sequence Read Achieve
TSA	Transcriptome Shotgun Assembly
wg	Wingless

1. INTRODUCTION

> *"Throw up a handful of feathers, and all must fall to the ground according to definite laws; but how simple is this problem compared to the action and reaction of the innumerable plants and animals which have determined, in the course of centuries, the proportional numbers and kinds of trees now growing on the old Indian ruins!"*
>
> —*Charles Darwin*

The work presented in this thesis is located at the interface between ecology, evolution and developmental biology. It addresses theories and questions in population biology, phylogeography and speciation as well as methodological approaches for applying Next Generation Sequencing (NGS) data. In the center of this thesis stands the world's largest extant damselfly, *Megaloprepus caerulatus*, as a model system for primary rainforests. Biodiversity, speciation and the evolution of novel characters are highly interwoven fields in biological research, and great methodological efforts are needed to elucidate their exact mechanisms.

1.1 Biological Diversity

The modern definition of biodiversity[1] includes three components: diversity of species, genetic diversity (i.e. genetic variety within species) and diversity of ecosystem types. By far the most diverse animal group are the flying insects (Pterygota). The evolution of wings

---

[1] *Biodiversity after E. O. Wilson - a pioneer in biodiversity research and conservation: "Biodiversity is the variety of life at every hierarchical level and spatial scale of biological organizations: genes within populations, populations within species, species within communities, communities within landscapes, landscapes within biomes, and biomes within the biosphere." (Wilson EOe. Biodiversity: National Academy Press, Washington D.C., USA; 1988)*

has allowed them to adapt to nearly every habitat [1-3]. Yet biodiversity is not distributed equally on the planet, and the regions with the highest diversity are the tropics [4-6]. Here about 80% of the overall diversity is concentrated on 20% of the earth's surface [6, 7], which is directly linked to high local endemism and generally lower local abundances [4].

Although intact biodiversity is of particular importance to human survival (e.g. [8]), it is destroyed at alarming rates [5, 6, 9-12]. Anthropogenic-induced habitat loss and climate change are unequivocally the main threats to extant and future biodiversity (cf. [4, 5, 13-15]). Species are becoming extinct 1,000 times faster than the calculated background rates [4]. Since present biodiversity has evolved over millions of years and the current habitat change appears from an evolutionary perspective very fast, most species may not be able to adapt or migrate quickly enough to guarantee their survival.

On this basis, conservation biology aims to understand and protect biological diversity by combining research in taxonomy, ecology, phylogeography, genetics and evolution (e.g. [16]). Prerequisite for monitoring and evaluating the effects of habitat disturbance and subsequently prioritizing research and conservation efforts, is information on species abundances, species ranges, community structures, and species states [11, 17]. Here in particular, conservation genetics and genomics can reveal objective and otherwise 'hidden' information about species, population- and ecosystem-status.

1.1.1 *Speciation*

Speciation is the *sine qua non* of biological diversity on earth. With "*The Origin of Species*" Charles Darwin [18] set the cornerstone of modern evolutionary synthesis proposing natural selection as the main driver of diversification. Mayr [19] and Simpson [20] added that within-species processes (i.e. mutation or natural selection) are important features in speciation.

Speciation research aims to understand how barriers to gene flow and morphological novelties arise, and finally why a lineage evolves into two distinct and reproductively isolated lineages (e.g. [21, 22]). Scientific efforts over the last 20 years have increased knowledge on the causal mechanisms but many issues remain unresolved (e.g. [22-25]). For this reason, the Marie Curie SPECIATION Network proposed a set of future research efforts, which included studying (i) the circumstances leading to reproductive isolation, (ii) the underlying genomic changes, and (iii) the relationship between speciation and biodiversity [22].

Traditionally, modes of speciation are explained by either the geographic origin of reproductive barriers or by the genetic sources (cf. [21], Figure 1). More recently however, ecological and non-ecological speciation are two emerging but opposing principles [25-30]. Both are built upon the concept of fundamental ecological niches (cf. [29, 31-33]). Ecological speciation describes the adaptation through niche exploration [25], while in non-ecological speciation natural selection retains niche constraints and species do not adapt [34]. As a result, non-adaptive speciation is linked to phylogenetic niche conservatism

(PNC), to a slow trait divergence and cryptic species [26, 34, 35]. Equally important to evolutionary and conservation research are the consequences of non-adaptive speciation and PNC (e.g. [26, 35]). Its primary causes such as natural selection for niche stasis, lack of gene flow between separated lineages, pleiotropy, and lack of variability within populations [34, 35] imply that these species are likely not able to adapt to global or local changes in environmental conditions and may face rapid extinction.

Discussions about the need of defining categories, which describe speciation modes, are very active (e.g. [35-37]), though it has been acknowledged that "each speciation event is unique" [22]. Consequently, comprehensive studies of different species groups are necessary to identify repetitive patterns of speciation.
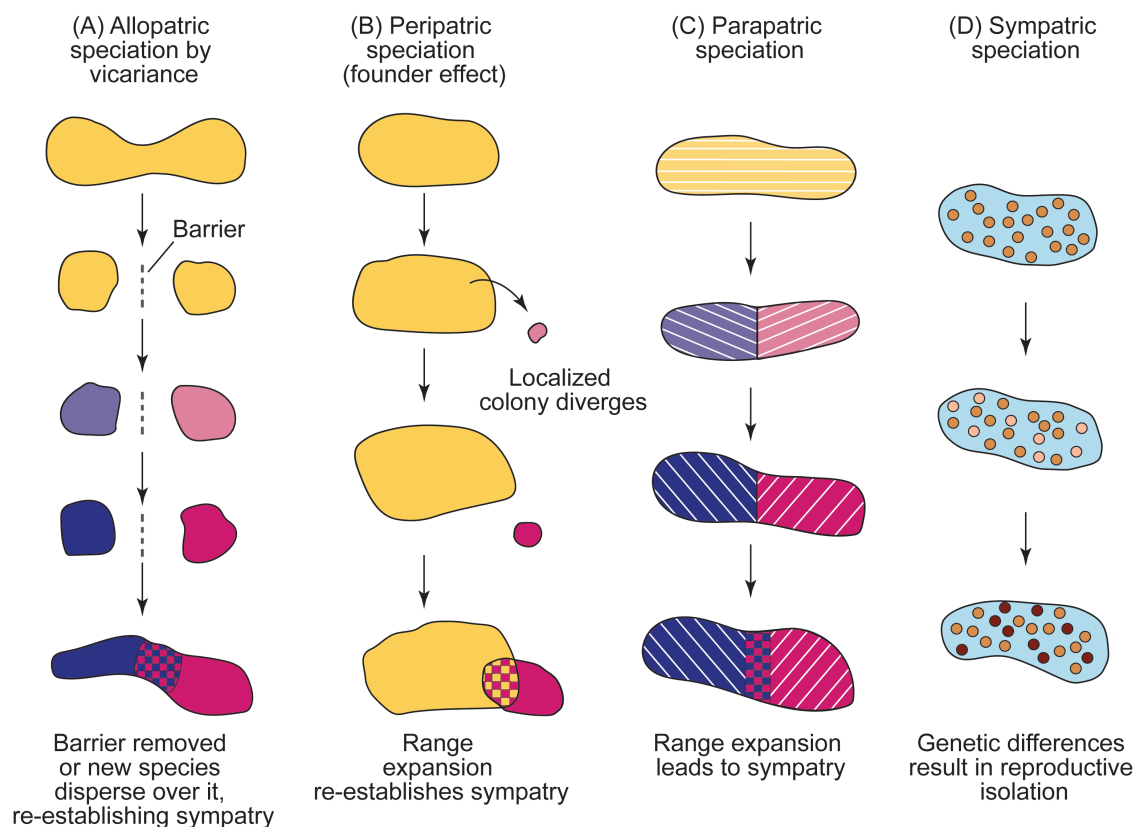
* Figure adopted from Futuyma 2009

Figure 1: Traditional classification of speciation modes by the geographic origin of reproductive barriers: **A – B)** allopatric, **C)** parapatric and **D)** sympatric. Please note that allopatric speciation has two modes and can either occur through vicariance or as a result of founder effects (peripatric). Hereby, however, different genetic mechanisms act at different time points and/or different geographic modes. The main genetic and causal effects involved in diversification are for example genetic divergence (i.e. genetic drift, peak shift, and natural selection), cytoplasmic incompatibility, and cytological divergence. Non-adaptive speciation might be most common among strict allopatric speciation (cf. [21]).

1.1.2 Unraveling species boundaries

The definition of species boundaries as fundamental units in biodiversity is the basic prerequisite for progress on research in conservation and systematics [38-40]. Although a great fraction of the unknown biodiversity remains to be discovered [2, 4, 9, 41], there are continuing difficulties in delimitating even the species that we are aware of. First, species that may be new to science are mostly hidden and may not belong to well-known groups. Second, due to the existence of a variety of species concepts, boundaries are difficult to set (e.g. [42-44]). Third, a challenge for taxonomists is still the phenomenon of cryptic species, as they constitute a significant percentage of the unknown biodiversity [38, 45]. The discovery of many cryptic species complexes within the last years (e.g. [46-51]) is a direct result of efforts to increase the distributional ranges investigated and the application of new molecular methods. Finally, there is a large discrepancy between taxonomic and molecular work. While classic comparative morphology alone often fails to discriminate cryptic species [38], taxonomists sometimes refuse to recognize species described solely with molecular methods [52].

Consequently, integrative approaches promise a higher accuracy, because they use information from different disciplines to serve more than one species concept [47, 53, 54]. In such a comprehensive analysis, critical criteria for species delimitation are investigated mutually and concordant divergence pattern in several characters allow for species determination (e.g. [43, 54, 55]). The taxonomic circle, for example, is one improved approach illustrating a strategy to resolve the conflict between different species concepts. It unites DNA, ecology, morphology, geography and reproduction in a circular workflow [43]. However, the choice of appropriate marker systems and methods of species discovery highly depend upon the system studied. Some of the best-known methods are discussed below.

The most important character for species delimitation is morphology, but morphology-based species descriptions can be subjective and overestimate species numbers. Consequently, objective and quantitative analyses are more often applied to describe phenotypic variation. For example, landmark-based geometric morphometrics (GMM) and linear morphometrics are central tools [56, 57]. Directly connected to taxonomy are also behavioral patterns. For example, coloration differences in insect wings modify mating success through sexual selection (e.g. [58, 59]). Secondly, the DNA barcoding is used for both, to identify and delimitate species. Traditionally a fragment of the *cytochrome c oxidase 1* - CO1 or *cox1* is used [60], because its high variability allows for the identification of species, populations, and sometimes even individuals [61]. The Consortium for the Barcode of Life (CBoL) has accepted CO1 as the standard DNA barcode region for vertebrates and insects (cf. [61, 62], but see [63]). However, another method, the character-based DNA barcoding (via the Characteristic Attribute Organization System '*CAOS-barcoding*'), is in comparison to distance-based methods more accurate through the identification of diagnostic characters at all taxonomic levels [61, 62, 64]. Such single-locus studies are recently complemented by multiple-locus comparisons [53, 64, 65]. Lastly, GIS-based methods and niche comparisons use environmental data to compare ecological niches, distributional ranges and ecological separation.

1.1.3 'Evolution' in molecular biology methods

In the last 10 years, molecular techniques 'evolved' rapidly, and scientists experienced intense accelerating changes. In the early 1990's, RFLPs (Restriction Fragment Length Polymorphisms) and AFLP's® (Amplified Fragment Length Polymorphisms) were used for population biology (e.g. [66, 67]). Subsequently they were largely replaced by microsatellites (e.g. [68-70]) and multiple sequence markers (e.g. [71-74]), which both became standard in conservation genetics (cf. [75-80]).

However, Next Generation Sequencing (NGS) methods are significantly outperforming the traditional methods because the obtained results have higher resolution and allow deeper insights into population structures, speciation, and developmental processes (e.g. [78, 81-85]). The continuous improvement of sequencing methods, analysis programs, and storage capabilities is revolutionizing molecular biology and has caused the shift from conservation genetics to conservation genomics. Although NGS methods continue to become more accessible, analyzing entire genomes is still expensive and time consuming. As a result, complete genome comparisons, are still mostly restricted to a few well-studied model organisms and well-equipped laboratories (e.g. ik5 project: atlasofthefuture.org/project/i5k-initiative/).

One alternative is transcriptomics or RNA-Sequencing (RNA-Seq). The relatively smaller size and lower repetitive content of transcriptomes in comparison to genomes makes transcriptomes easier to work with. Additionally, no prior genomic data is necessary for analyzing and assembling transcriptomes or for large-scale comparative studies (cf. [86]). Today transcriptomes are a valuable tool for ecologically important non-model organisms to study adaptation, speciation, phenotypic variation, organismal development, and the origin and maintenance of biodiversity (e.g. [87, 88]). The broad range of applications from functional annotation and novel gene identification (e.g. [89-93]) to single nucleotide polymorphism (SNP) [94, 95] and expression level comparisons [96-99] make RNA-Seq a method of choice to study the response of indicator species to environmental disturbances, because it goes beyond the single gene level towards the systems level and can link differential gene expression in nature to evolutionary processes.

1.2 The Neotropics and the Neotropical odonate diversity

1.2.1 The Neotropics

The New World tropics or Neotropics (from the Greek *neos* = new) define the tropical ecoregion of the American continent (Figure 2). The Neotropical ecozone is the richest biogeographical region on earth [100]. It includes the world's largest continuous rainforest block [101] and the most complex ecological communities with a high economic importance on both regional and global scales [102, 103]. The primary causes for this high level of

biodiversity and endemism are complex and associated with the multifaceted geological and geographic history i.e. Neogene tectonic events and climatic changes during the Pleistocene [104-108].

1.2.1.1 Changes in geography over time

The biogeographic history of the Neotropics represents an important background for approaching species-specific research questions. It started with the final break-up of the Paleocontinent Gondwana about 135 to 100 Mya [105, 109]. With the westward drift of the South American Plate and the eastward drift of the Nazca Plate an oceanic–continental subduction process occurred along the Pacific margin of South America (cf. Andean cycle, [110]) causing changes in the Amazon Carton and later the Andean formation [105, 110, 111]. During the Paleogene (~65 to 23 Mya), the Andean orogeny occurred slowly in discrete periods and different regions and in Late Miocene the Andes experienced a phase of fast mountain uplift at ~11-7 Mya [111]. This caused major climatic changes in the eastern Andean slopes and on the rising mountains tips [112], dramatically modifying the landscape evolution of northern South America [105]. The Andes reached its present elevation in the Pliocene (~3.5-3 Mya) [105, 107, 111]; but also see [113]. In summary, the Andes not only separate the Amazon basin from the Pacific coast, but also the complex orogeny resulted in a variable surface structure with high mountain peaks and deep river valleys summarized in at least 15 biogeographical regions (e.g. [106]).

The second large geological event shaping the Neotropical biota was the formation of the Isthmus of Panama[2]. During Paleocene Lower Central America (LCA) was an island archipelago. Continuous surface volcanism and landmass uplifts beginning in late Eocene-Miocene caused a gradual closure of the Isthmus [107, 113, 114]. There is an ongoing debate about the exact closure time, with time frames between 3.5-3 Mya (e.g. [107, 114]) and 8 Mya [115] being discussed. However, after the Isthmus was fully closed, the great biotic interchange (GABI) caused a confluence of flora and fauna [107, 108, 116-118].

Pleistocene climactic variation, with its accompanying sea level changes, further influenced flora and fauna in the Neotropics (e.g. [108, 119, 120]). Major glaciations caused a cooling between 5-8 °C, which in turn resulted in downwards shifts of montane fauna to lower elevations as the high mountain tips were covered with glaciers [107, 120]. In the lowland areas, however, forests were unaffected, fragmented or replaced by savannah depending on local climates [120, 121]. These habitat changes represented major isolation events and caused many recent animal radiations [74, 122-124].

---

[2] *"The Isthmus of Panama is a more effectual line of union, since it is hilly, well-watered, and covered with luxuriant vegetation; and we accordingly find that the main features of South American zoology are continued into Central America and Mexico." (Wallace, 1876, p. 38).*

1.2.1.2 Current Status quo

The Neotropics include seven biodiversity hotspots (Figure 2) [100]. The assignment of biodiversity hotspots for conservation priorities implies that those regions are under high threat, because by definition more than 70% of the original natural vegetation is lost [100]. The current picture of Neotropical forests is heavily characterized by destruction causing the tropical biodiversity crisis [11, 12, 125].

This environmental degradation has been related to rising standards of living and increasing human population sizes in developing countries, and economic globalization (e.g. [126-128]). The serious fragmentation is reflected by a loss of 30% wilderness in the Amazon [129, 130] and 70% of the total loss is due to large-scale commercial agriculture, mainly for supplying the food demands of the first world [126, 127, 131]. Consequently, most remaining Neotropical landscapes consist of a mosaic of forest patches embedded in pastures, agricultural and urban areas.

The destruction of Neotropical natural habitats represents a serious threat to biodiversity on local and global scales [128, 130, 132]. Local endemics are mostly eliminated [133]. The loss of continuous habitats creates geographic barriers, limits migration and gene flow. The resulting isolation of populations modifies species abundances as well as species richness, genetic diversity and enlarges extinction risks [133]. Inside the remaining forest patches, physical and hydrological characteristics change due to selective logging, edge effects, climate change, and wind disturbances [134, 135]. Consequently, tree mortality increases and ecosystem functioning decreases [133-138].

In summary, Neotropical biodiversity is under serious threat, but remains understudied. In a meta-analysis of 2,434 phylogeography publications, only few focused on the Neotropics (~3% in Central America and 6.3% in South America [139]), which demonstrates "*that the top two areas of vertebrate species richness, endemism and threat— the Tropical Andes and Mesoamerica [100]—are largely underrepresented*" [107].
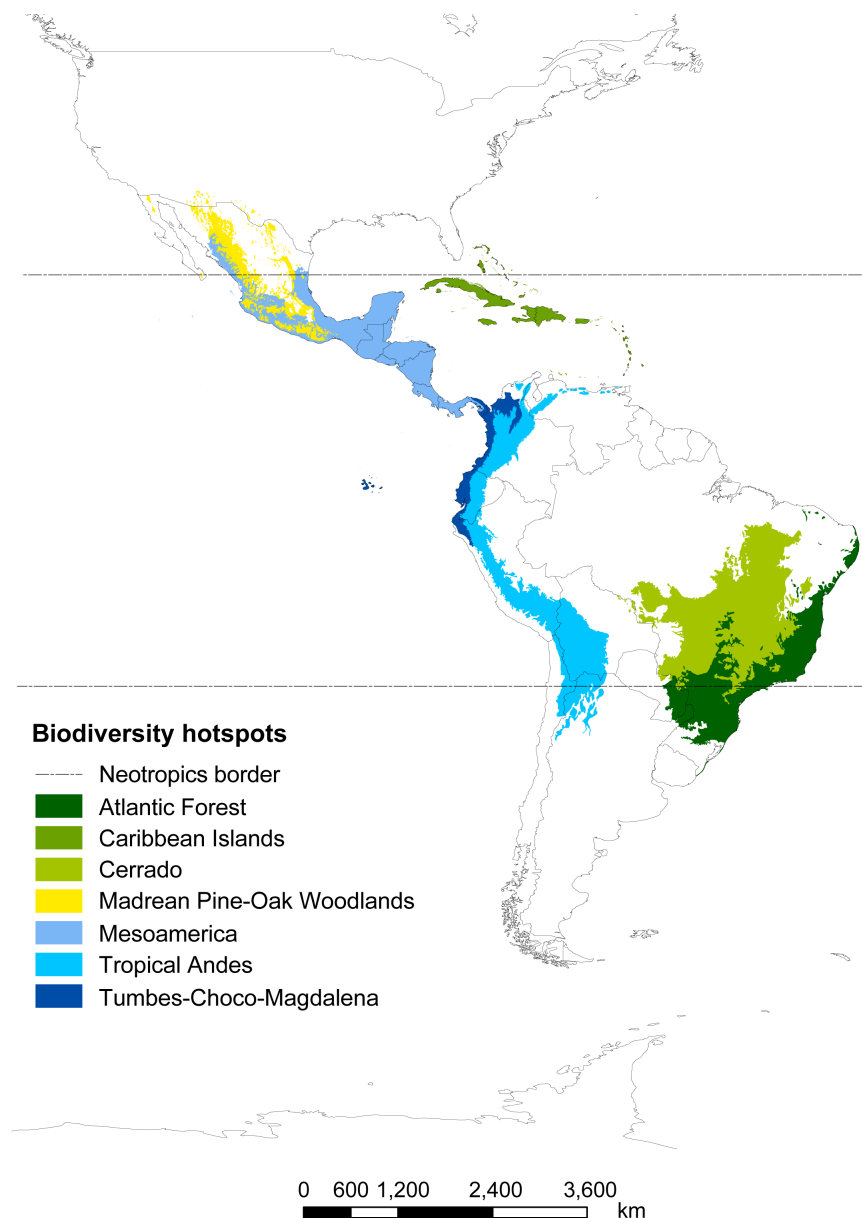
Figure 2: The Neotropical ecoregion with its seven Biodiversity Hotspots (e.g. [100]). The Neotropics extend from central Mexico to southern Brazil containing all Central America, the Caribbean Islands and most of Southern America including northern Argentina and Peru [140, 141], which is indicated by the gray dotted lines. From a strict perspective the transition zone in Mexico, a Nearctic region, and the Andes are not included, which is defined as the Neotropical region *sensu stricto* contrasting the Neotropical region *sensu lato* with both included [140].

1.2.2 Odonata as model systems

As conservation and molecular ecology has entered the '*omics*' era, odonates are a promising animal group, in which future genomic studies will certainly reveal new fundamental knowledge (cf. [142]). The order Odonata (Insecta, Pterygota, Palaeoptera: Odonata), which contains app. 6,000 described species, is a relatively small insect order

[143]. It consists of two main suborders, the Anisoptera (true dragonflies) and Zygoptera (damselflies) (e.g. [143] but see [144]).

Odonates are ecological and evolutionary supermodels. They combine specific characteristics that make them unique among flying insects (Pterygota). Odonate evolution dates back to the late Carboniferous period between 400-350 Mya and together with the Ephemeroptera (mayflies), odonates are among the earliest winged insects [1, 145]. Odonates as an order encompass a high ecological diversity. They are adapted to a wide variety of freshwater ecosystems ranging from arctic areas in Sweden and Canada to the subantarctic in Argentina. Their complex, hemimetabolous life cycle includes both an aquatic immature larval stage and a terrestrial imago. Differential habitat requirements of adults and larvae (i.e. stenoecious vs. euryoecious) [143, 146, 147] are linked to dispersal capabilities, which in turn allow for studies of vicariance and dispersal on different evolutionary timescales [148]. High ecological sensitivity allows odonates to serve as bioindicators for environmental health [147, 149]. Furthermore, reproduction is unique and complex in odonates; its discovery leads to a pioneering principle in evolution – sperm competition (cf. [150, 151]). Consequently, odonate mating systems, which are characterized by sexual conflicts, have grave effects on the evolution of genital and wing morphology (cf. [146, 148, 152-157]).

These exclusive features (among others) paved the way for odonates to emerge as model systems in recent ecological and evolutionary research (e.g. [142]). Recent work includes the origin of wings and wing coloration [90, 99, 154], body coloration and polymorphism [158-161], the evolution of *Hox* genes and body bauplan [99], vision [162, 163], niche conservatism and niche evolution [164, 165], and ecological adaptation and speciation [164-166].

1.2.2.1 Speciation research in odonates

Despite the above-mentioned studies, most odonatological research focuses on taxonomy, systematics, ecology and behavior (e.g. [147]) resulting for example in a recently high number of newly described species (e.g. [143, 167-170]). Even through of the great importance of taxonomic work for delimitating new species and species abundances, it is equally important to understand the mechanisms that lead to species diversity on local and regional scales.

Considering the high ecological and evolutionary diversity of odonates, it is surprising that Odonates have been under-represented in speciation research. With continuously high levels of scientific interest in this field (e.g. [24, 25, 30, 35, 82, 171]) and the fact that the first odonate genome [172] and several odonate transcriptomes [89, 90, 99, 163, 173] have recently been published, scientific attention will hopefully increase. So far radiations in odonates have been mostly related to ecological variability and sexual selection (i.e. adaptive speciation) [174-176]. To date only two reviews are introducing potential non-adaptive speciation processes in Odonata [164, 165]. First, Svensson [164] suggested non-ecological speciation as common in Odonata, while he considers non-adaptive speciation as

a process-based phenomenon potentially followed by adaptation and natural or sexual selection. Wellenreuther and Sánchez-Guillén [165] provided further elaboration and compared whether the sympatric living damselfly genera *Ischnura*, *Enallagma* and *Calopteryx* [177] speciated through non-adaptive speciation or not (see also [178]). They concluded that ecological niche diversification is low among species, but that reproductive isolation may have arisen due to sexual selection [165]. Despite various speciation hypotheses, 'true' allopatric, non-adaptive speciation has not been demonstrated in dragon- or damselflies.

1.2.2.2 Neotropical Odonate diversity

Odonate research in the Neotropics currently receives much attention (cf. [170, 179-186]). In 2012 there were 1,746 known species from that area (derived from the global species database Odonata [144]) and between 2006 and 2010 the majority of newly described odonate species (43%) were discovered in tropical America [143]. With still 400-500 undescribed species [144] 'hidden' in museum collections worldwide and unexplored areas such as southern Guyana or the Amazon there will be more to come.

The majority of dragon- and damselflies in the Neotropics are the forest species [149, 187]. Little is known about forest dragonflies; through it appears that approximately 80% of all genera include forest species, which represents the ancestral state from an evolutionary perspective [187]. Famous examples for forest odonates are the Polythoridae [159, 188] and the Pseudostigmatidae [149, 189].

1.2.2.3 The Pseudostigmatidae

The Pseudostigmatidae (Odonata: Zygoptera) is a relatively small damselfly family with only 25 described species (Table 1) arranged in six Neotropical genera (*Anonisma*, *Mecistogaster*, *Platystigma*, *Megaloprepus*, *Microstigma* and *Pseudostigma*) and one African genus (*Coryphagrion*) [167, 190-192].

The natural history of the Pseudostigmatidae makes this family an excellent model system for studying evolutionary patterns inside tropical rainforests. Members of this family are commonly known as helicopter damselflies. With wingspans up to 190 mm and abdominal lengths up to 12 cm they are the largest recent odonates worldwide. While the two genera *Platystigma* and *Mecistogaster* - each radiated into eight species inhabiting different forest environments (from moist forests to dry forests at different succession states), *Megaloprepus* remained monotypic and is restricted to old growth rain forests (e.g. [191, 193, 194]). However, in all forest types, the Pseudostigmatidae oviposit exclusively in phytotelmata [149, 187, 189]. They are small water accumulations inside terrestrial plants (*Bromeliaceae* or tree holes) or in dead plant material (fallen trees, leaves or fruit husks), and represent an additional aquatic habitat with a unique flora and fauna [195-198].

Today the phylogenetic position of the Pseudostigmatidae within the damselfly tree of life is unresolved. Earlier work by Groeneveld *et al.* [192] placed the Pseudostigmatidae close to the Eastern African species *Coryphagrion grandis* and identified this family as an

old Gondwana relict. Recent phylogenetic research [199] questioned the family status and included the Pseudostigmatidae into the large and diverse family of the Coenagrionidae – the narrow-winged damselflies. However, the use of a few partial marker genes, low node supports, and current understanding of integrative taxonomy contradicts this. Consequently, in the present work the Pseudostigmatidae are treated as a family until more comprehensive phylogenies will be published (please also compare [167]). Furthermore, the precise number of species within the Pseudostigmatidae is also unresolved. Consequently, taxonomic studies are highly in demand in the Pseudostigmatidae and other Neotropical odonates, not only because the Neotropics still contain many unexplored regions but also because many unclassified specimens exist in museum collections worldwide.

Table 1: The family of the Pseudostigmatidae consists of seven genera and 25 species. Here only true species are listed, but it is assumed that many undescribed species occur in tropical America.

Genus	Species	
Coryphagrion	*grandis**	MORTON, 1924
Anonisma	*abnorme*	SELYS, 1860
Mecistogaster	*amalia*	BURMEISTER, 1839
	linearis	FABRICIUS, 1776
	amazonica	SJÖSTEDT, 1918
	garleppi	FÖRSTER, 1903
	astica	SELYS, 1860
	lucretia	DRURY, 1773
	modesta	SELYS, 1860
	ornata	RAMBUR, 1842
Platystigma	*astictum**	SELYS, 1860
	buckleyi	MCLACHLAN, 1881
	martinezi	MACHADO, 1985
	pronoti**	SJÖSTEDT, 1918
	jocaste	HAGEN, 1869
	humaita	MACHADO & LACERDA, 2017
	minimum	MACHADO & LACERDA, 2017
	quadratum	MACHADO & LACERDA, 2017
Megaloprepus	*caerulatus*	DRURY, 1782
Microstigma	*anomalum*	RAMBUR, 1842
	calcipennis	FRASER, 1946
	maculatum	HAGEN, 1869
	rotundatum	SELYS, 1860
Pseudostigma	*aberrans*	SELYS, 1860
	accendens	SELYS, 1860

According to the IUCN Red List of Threatened Species: * vulnerable, ** critical endangered.
For further information please compare specific references [149, 167, 190, 191, 193, 194, 200-215]

1.2.3 *The genus* Megaloprepus

Within the Pseudostigmatidae, *Megaloprepus caerulatus*, DRURY, 1782 (Odonata: Zygoptera, Pseudostigmatidae) is one of the most impressive species (Figure 3). Its history is characterized by taxonomic disagreements [202-205, 216]. In recent times, the single species classification was called into question [149] based on character-based barcodes from the Hadrys' lab. These doubts were supported by sexually dimorphic wing coloration that occurred in specimens from Costa Rica and Panama, but not from Mexico [149]. Males in the dimorphic populations have an additional white wing band proximal to the dark metallic blue band (Figure 4), while females have two bright white dots on the wing tips (cf. [217]). It has been shown that the shape, size and position of the colored regions of *Megaloprepus'* wings and their UV reflectance contribute to complex territorial and sexual behaviors [157, 217], suggesting positive selection for larger wings in males. Consequently, sexual selection could have triggered large inter-population effects.
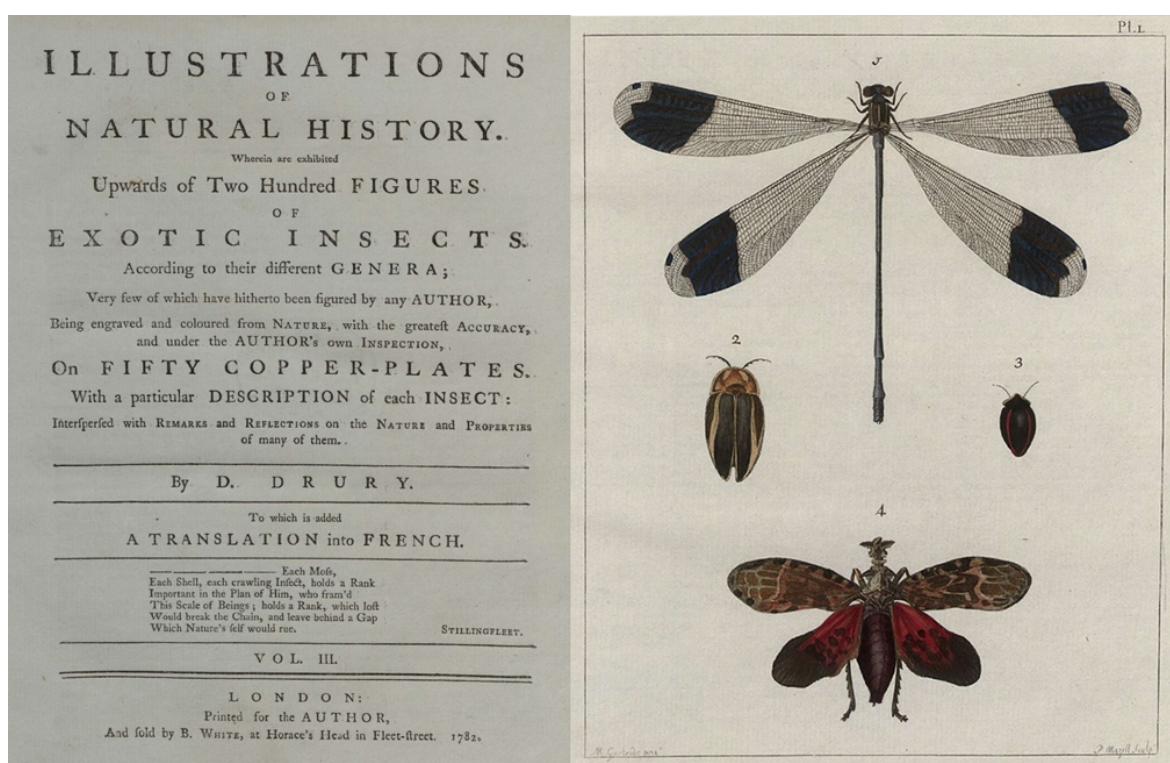
Figure 3: Original publication of Drury from 1782 describing *Megaloprepus caerulatus* as *Libellua caerulata* for the first time. Since Drury *Megaloprepus* was included in many ecological studies and it is probably the most studied damselfly genus in the Neotropics. Despite this, the presented research here is the first that includes also genetic and genomic research.

A broad ecological background (e.g. [68, 157, 189, 195, 218-220]) reveals that *M. caerulatus* is a niche specialist with a wide distributional range. Two niche-related traits make *M. caerulatus* an excellent model for speciation research. First, its dependence on water filled tree holes as a crucial larval habitat [218], whereat stable population sizes require a certain amount of tree species that produce large tree holes [189, 218, 219, 221]. Second, its low capacity to colonize secondary forests or to migrate from one forest patch to another apparently resulting from an inability to tolerate higher temperatures. Thus, *Megaloprepus* is locally restricted as soon as forest patches are disconnected (cf. [149, 189, 222]).

These niche constraints imply that past and current environmental changes may have negative effects on *Megaloprepus*. Consequently, population-level comparisons would allow real-time monitoring of the effects of habitat destruction and elucidate how geological changes over time have shaped evolution within sensitive forest insects.

Figure 4: *Megaloprepus* in its natural habitat. **A)** A male *M. caerulatus* from the Biological Research Station La Selva in Costa Rica. The wingspan of *Megaloprepus* is with 190 mm the largest worldwide. Obvious is here the milky white wing band proximal of the blue band. This white band is missing in all females and the males from Mexico, Honduras, Guatemala, the pacific Coast of Costa Rica as well as on the east side of the Andes. **B)** Light gap on Barro Colorado Island. Although *Megaloprepus* does not persist in open areas among forest sites, it occurs in light gaps, but within 'healthy' old-growth rainforests with a closed canopy providing stable microclimatic conditions. Here males are territorial and usually defend a tree hole. **C - E)** Water filled tree holes. Plant held waters (eg. Fruit husks, bromeliads, tree holes) are an important aquatic habitat in tropical rainforests where rivers and lakes are rare. **C)** A water filled tree hole inside a dead fallen tree (probably *Fabaceae*). **D)** Tree hole in a buttress root containing a *Megaloprepus* larva **E)**.

1.3 Open questions in Neotropical odonate research

Research demands in Neotropical odonatology are threefold. First, how many species exist in this area? For future research there is a great need to overcome taxonomic incompleteness. Many cryptic species are likely to be identified, and there is a high probability of identifying new genera or even families. Second, what are their distributional ranges? Biogeography and systematics would benefit from large sampling efforts. Lastly, what are the effects of past and current habitat changes? Here, traditional methods combined with large-scale NGS research would allow for precise explanations for evolutionary mechanism such as morphological novelties, adaptation, biodiversity patterns and speciation.

2. AIMS OF THE THESIS

The focus of this thesis is on *M. caerulatus* as a model species to study ecology, evolution and development in Neotropical rainforests, implementing the transition phase from conservation genetics towards genomic work. By combining traditional population genetics with modern taxonomy, ecology and geography, the population structure and speciation modes in *M. caerulatus* were studied and will be described. The research will contribute to the knowledge about Neotropical odonates and establishes a model system for modern research in tropical odonates. Furthermore, this work wishes to provide novel insights into the use of RNA-Seq, new genomic resources and candidate genes for wing development in odonates.

2.1 Phylogeography of the genus *Megaloprepus*

In the first three chapters the population genetic structures, the evolutionary history and the *status quo* of the genus *Megaloprepus* will be illuminated. Specifically, the following questions are addressed:

(i) What are the population genetic structures between populations from isolated geographic regions within the genus *Megaloprepus,* and is gene flow detectable?

(ii) If individual populations are genetically isolated, can this be correlated to past climatic change, recent fragmentation of rainforest habitats or large geographic barriers?

(iii) If populations have distinct geographical distributions, does the genus *Megaloprepus* still consists of a single species?

(i) Population genetic structure of the Neotropical damselfly *M. caerulatus*

To determine if the current habitat structure has impacted populations of *Megaloprepus*, population genetic structure and genetic diversity within and between four populations in Mesoamerica were compared (chapter 6.1). This included four sample sites from Mexico to Panama. Two standard methods of analyzing diversity indices on different taxonomic levels were used: mitochondrial sequence markers and microsatellite loci. The genetic distances were evaluated in relation to genetic distances among sister species within the Pseudostigmatidae (i.e. *P. jocaste*, *P. asticta* and *P. martinezi*, formerly belonging to *Mecistogaster*). In addition, ecological data for water filled tree holes in different sampling areas were compared, specifically tree hole size, occupancy, temperature, conductivity, pH and number of tree holes per hectare. The results showed three distinct genetic clusters indicating diversification at the species level, which led to a deeper look into the underlying (evolutionary) mechanisms.

(ii) *Megaloprepus'* phylogeography unravels cryptic speciation

In the consecutive study (chapter 6.2) the diversification patters of the genus and its potential causal mechanisms were examined in detail in order to illuminate the specific radiation pattern in *Megaloprepus*. Under this aspect, an integrative phylogeographic study scheme was designed combining population genetics, phylogenetics, morphometrics and species distribution modeling in order to prove multiple species in the genus and to reveal how speciation may have been affected by *Megaloprepus'* ecology and the geography in the Neotropics over time.

Newly collected samples from 14 populations were combined with museum material from four different collections. Together, the included specimens cover almost the entire distributional range of the genus. For a classical taxonomic overview, historical species descriptions [193, 202-206, 216] were compared to the initial genetic results (6.1) and obvious phenotypic characters of the specimens from the different regions. Based on these findings' specimens were divided into four clades (putative species): *M. caerulatus*, *M. brevistigma*, *M. latipennis* and *Megaloprepus* sp. nov.. Population- and phylogenetics should provide conclusions about genetic diversities and phylogenetic relationships, and the genetic patterns should support the geographic distributions. An accompanying time-calibrated phylogeny was used to determine the time of divergence, correlating past geological events with the observed present-day diversification. To quantify morphological diversification with respect to wing shape and size, landmark-based geometric morphometrics (GMM) and linear morphometrics were applied. Species distribution modeling was implemented to estimate potential distributions of *Megaloprepus* at different time scales (i.e. current and during the Pleistocene) as well as to compare the ecological niches among groups. Finally, in its complexity this work presents an example on how phylogeographic studies could be designed studying Neotropical insects in general.

(iii) Four in one – revalidation of the genus *Megaloprepus*

Based on the results of the two previous studies (6.1 & 6.2), a cryptic species complex must be assumed. By taking integrative taxonomy into account, the last manuscript of this section (6.3) aims to define species boundaries and establish new species states in the genus *Megaloprepus*.

The revision of the genus is based on extensive morphological analyses, and character-based DNA barcoding accompanied by a phylogeny using the Folmer barcoding region CO1 [60, 61]. Specimens from 11 different museum collections were combined with the newly collected material. Using those specimens, the precise aims were to (i) re-describe the nominal species *M. caerulatus*, (ii) identify the lectotypes for *M. latipennis* and *M. brevistigma* in Selys' collection and re-describe them while raising them to species level, and (iii) to define the male holotype for the fourth *Megaloprepus* – a new species. Finally, morphological variations are discussed and included into a taxonomic key to adult males.

2.2 Transcriptomics as a backbone for future Eco-Evo-Devo studies

The second section of this thesis aims to establish new approaches in RNA-Seq while answering the following questions:

- (iv) Is it possible to identify wing genes from the transcriptome of a larval thorax?
- (v) How can RNA-Seq data be easily submitted to the National Center for Biotechnology Information (NCBI) database?

(iv) Transcriptome profiling in *Megaloprepus*

Working towards future comparative transcriptomic studies (cf. Appendix), the manuscript in chapter 6.4 presents a comprehensive transcriptome profile of *M. caerulatus* in order to identify new candidate genes e.g. for wing development but also for creating a backbone to look deeper into temperature sensitivity and niche conservatism. To address these goals, the complete RNA of a single larval thorax (including the wing buds) was sequenced. Very stringent methods for read cleaning, assembly evaluation and annotation were applied to the resulting sequence data. To identify *Megaloprepus*-specific genes responsible for wing development and coloration, the transcriptome was screened for *Hox* genes of the Antennapedia (ANTP) class, four major wing developmental signalling pathways and the wing-patterning network.

(v) Submission of RNA-Seq data to NCBI

Although many questions can be addressed with RNA-Seq data, most research groups are interested in specific topics and thus leave this valuable resource underutilized. A responsible handling of this 'unused' data is to make it available to the scientific community. Although, most scientific journals oblige researchers to do so, when results are being published, RNA-Seq data submission to NCBI can be challenging and time consuming.

The purpose of the manuscript in chapter 6.5 was to design two protocols showing how researchers can submit RNA-Seq raw sequences and assemblies to the NCBI databases with the basic premise of encouraging researchers to submit their data to NCBI and facilitate this process.

2.3   Mitochondrial genomes–a deeper look into molecular diversity

Mitochondrial genomes (mtDNA) play an important role in modern population genetics, phylogeography, molecular systematics and evolutionary dynamics (e.g. [223-226]). They are the most studied genomic resource in insects and have revealed fundamental results in evolutionary research [225, 227]. Furthermore, because they are involved in oxidative phosphorylation (OXPHOS), genes of the mitochondria, together with nuclear genes, enable cellular respiration [228, 229]. The first complete mitochondrial genome of an odonate was published in 2010 from the damselfly *Euphaea formosa* [230]. By the summer of 2018, nearly 30 complete odonate mt genomes were deposited at NCBI.

Aiming to contribute to the mitogenomic dataset, three mitochondrial genomes were generated and described within the last three chapters of this thesis (6.6-6.8). Hereby three different odonate species with significant different ecological niche requirements were selected: (i) *Ischnura elegans* (VANDER LINDEN, 1820) a common and widespread European damselfly occurring in a wide range of aquatic habitats. It is an important model species for studies in the evolution of color polymorphism and wing development [97, 99, 160]. (ii) *Anax imperator* (LEACH, 1815) a species, which is changing its distributional range due to recent climate change [69]. (iii) *Megaloprepus caerulatus* an indicator of Neotropical forest health with a restricted and old phylogenetic niche [149]. For the assembly of all three mt genomes, a fraction of draft genomes was used and mapped onto a reference seed sequence. The final genomes were annotated and described.

3. SUMMARY OF RESULTS AND DISCUSSION

3.1 Phylogeography of the genus *Megaloprepus*

(Feindt *et al.* 2014)

(i) Population genetic structure of the Neotropical damselfly *M. caerulatus*

As the first genetic study of *M. caerulatus* this work sets the backbone for all following research. Results from two mitochondrial sequence markers (NADH dehydrogenase subunit I - ND1 and 16S ribosomal RNA -16S rRNA) and microsatellites revealed a similar pattern. First, low genetic variability within populations was found. Second, the four studied populations showed a split into three clusters with no gene flow and no shared haplotypes between clusters. Third, the microsatellites established for *M. caerulatus* from Panama (cf. [68]) failed to amplify in all samples from the other two clusters; the Biosphere Reserve Los Tuxtlas, Mexico and the Corcovado National Park, Costa Rica. These facts combined indicate genetic differentiation potentially at the species level.

The most plausible reason for this geographic isolation is *Megloprepus'* adaptation to a narrow ecological niche. Forest destruction [5, 231, 232] and lost continuous biological corridor of old growth forests have disrupted *Megaloprepus'* habitat. Adult *Megaloprepus* are unable to migrate from one forest patch to another if patches are more distant than 50 m (cf. [189, 222]). Consequently, migration between patches is not present.

A dramatic case was observed in the Los Tuxtlas region. Here the individuals show the lowest overall genetic diversity. This may reflect high rates of human disturbance over the last 50-100 years, which left many small and disconnected forest patches. In addition, comparison of ecological parameters of water filled tree holes showed differences between study sites. Although significant differences in pH, temperature and conductivity were discovered, their influence on *Megaloprepus* larvae is unclear. These variables were

collected at disjunct times of the year. Consequently, the observed differences could be natural variation in those forest sites, rather than results of climate change and forest patch size. But it points to the need for larger scale data collections.

(ii) *Megaloprepus'* phylogeography unravels cryptic speciation

The hypothesis of a past speciation event in *Megaloprepus* was proved through concordant results in genetic and morphological analyses; and further confirmed by *Megaloprepus'* current distribution and evolutionary theory.

Genetic differentiation measurements revealed that the four previously defined clades are now four isolated genetic clusters with genetic distances from 6 to 11% in the CO1 marker gene. The four groups have no shared haplotypes and no regions of co-occurrence. The relaxed molecular clock exhibits that the most recent common ancestor of *M. brevistigma* and the three Mesoamerican species probably diversified due to the uplift of the Andes 10-8 Mya, whereas the three Mesoamerican species differentiated 3-2 Mya, after the closure of the Isthmus of Panama (Figure 5). Both, linear morphometrics and GMM detected diverging wing patterns, which had high variation in the shape of the lower wing margin and the blue wing band. As a result, the genus now consists of four species: the nominal species *M. caerulatus*, the two previously described species with a long lasting unknown status *M. latipennis* and *M. brevistigma* and one true new species, *Megaloprepus* sp. nov..

The high estimated niche similarity among species is interesting, given that the three Mesoamerican species have been isolated for a minimum of two million years. This may reflect that speciation occurred through niche conservatism and that the three Mesoamerican species exhibit a phylogenetic niche conservatism. However, in the face of niche conservatism lineages tend to have a low potential to adapt and consequently, morphological inventions are expected to be rare [34, 35]. Exactly this was observed in *Megaloprepus*. But non-adaptive speciation can be followed by new random mutations, with a higher potential to become established when there is an ecological advantage (cf. [36, 134]). This could have happened to *M. caerulatus* as it is the only species with sexually dimorphic wing coloration and is also the most derived species. Wing patterns of *M. caerulatus* were related to sexual selection and territoriality [157, 217]. The appearance of sexual dimorphic wing traits could be related to small differences in climatic conditions (such as higher cloudiness in the Chocó-Darién) that were not detected by the SDM.
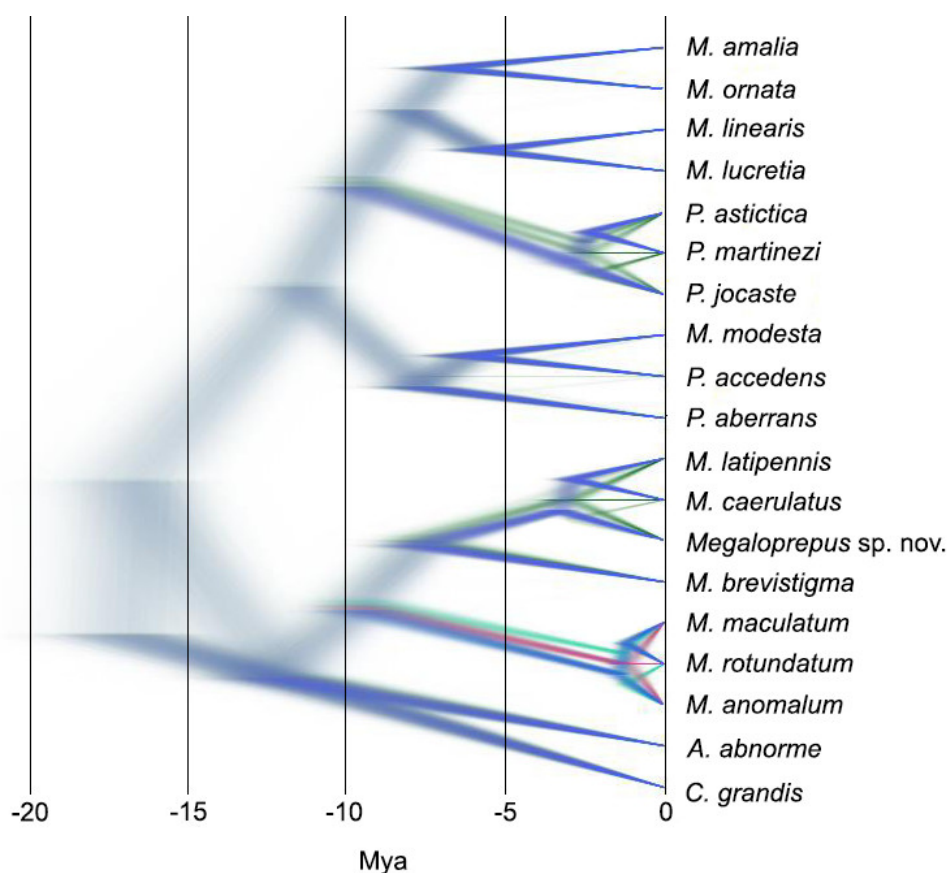
Figure 5: Time calibrated phylogeny for the Pseudostigmatidae using a log-normal relaxed molecular clock (please compare chapter 6.2).

Two factors could contradict the four species solution. First is the weak genomic differentiation, which however could have its probable origin in introgression due to hybridization during the Pleistocene. Second due to the reduced phenotypic evolution, because in other odonate species differences in wing shape were also observed among populations of one species (i.e. migratory vs. non-migratory [184]). But as said above without selection pressures or the need to adapt, morphology only diverges due to slow occurring genetic drift.

(iii) Four in one – revalidation of the genus *Megaloprepus*

The traditional species description presented in manuscript 6.3 defines distinct morphological characters for the four discovered *Megaloprepus* species. All previously defined species-specific characters by Selys [202, 203] and Ris [205] could be confirmed. This included the shape and size of the Pseudostigma [202, 203], the placement and the size of the blue wing stripe and the ratio of wing width to wing length [205] as the most prominent. However, new taxonomic characters were added. This includes the shape of the prothorax, the male appendages and the male ligula. In particular, shape variations in the male secondary sexual organs are correlated with mating success and sexual conflict (cf.

[146, 148]). Whereas the differences of the ligula among the four *Megaloprepus* seem not very distinctive on the first view, in comparison to the closest sister genera *Microstigma* and *Anomisma* the ligula appears even similar to those genera and variation among the well-defined *Microstigma* species is also low (cf. [191]).

Despite still cryptic, the taxonomic features alone could allow species delimitation, but in combination with the distinct character-based DNA barcodes, the current distribution and the evolutionary history, the (re-)discovery of 4 species is difficult to neglect. Consequently, *M. latipennis* and *M. brevistigma* now receive species status. In addition, *Megaloprepus diaboli* sp. nov. is named as the fourth species within the genus aside of the nominal species *M. caerulatus*. Its haplotype is from the Corcovado National Park in Costa Rica. These species may represent the first "cryptic" species complex described in a Neotropical odonate. Finally, by merging all information about the geographic origin of the four *Megaloprepus* species (i.e. old species descriptions, museum specimens and new collected material) the present work allows considerable insights into the distribution of the four species (Table 2, Figure 6).

Figure 6: Conservative estimation of the distributional ranges of the four *Megaloprepus* species based on the specimens i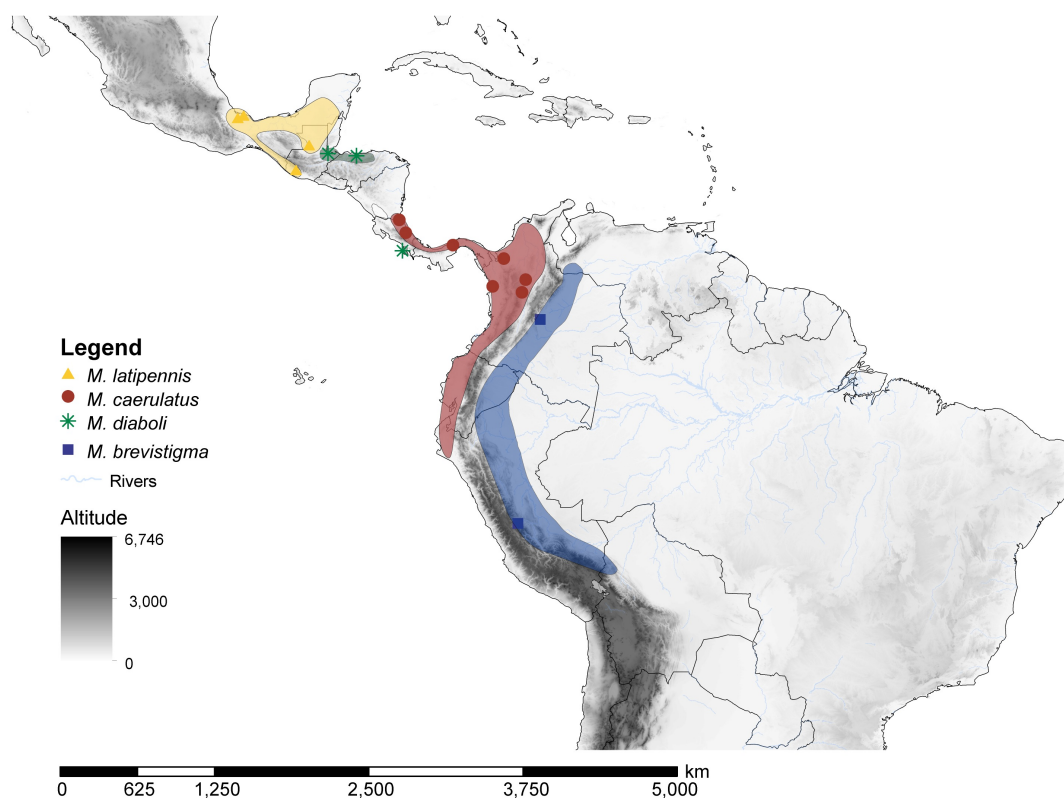ncluded into the species description. Today, the four species occur in distinct regions separated by mountain ranges and dry forests.

Table 2: Proved distributional ranges of the four *Megaloprepus* species based on old species descriptions and the recent morphological and genetic results. The nominal *M. caerulatus* was considered as single species genus distributed from southern Mexico to Bolivia. Now this range is occupied by four different species. Areas that should be confirmed by intensive fieldwork are: Mexico: Reserva de la Biosfera Calakmul, Chiapas; Guatemala: Reserva de Biosfera Maya; Belize; Honduras: Biological Reserve Río Plátano and Reserva Biologica Tawahka; northern Nicaragua: Biological Reserve Cayos Miskitos and Reserva Natural Bosawás; Colombia: Bogota region; West of the Andes: Amazon and Guyana (i.e. according to the SDM)

Species	Distributional range
M. diaboli sp. nov.	South Pacific Coast Costa Rica (i.e. Corcovado National Park and the Osa peninsula) and Caribbean Coast from Honduras (Pico Bonito National Park) to Guatemala (Cerro San Gil)
M. latipennis	Mexico (Veracruz), Northern Guatemala (Laguna Lachuá, Río Bravo)
M. brevistigma	East Cordillera South America (e.g. Colombia, Venezuela, Peru, Ecuador, Northern Brazil)
M. caerulatus	South Caribbean coast of Central America: Nicaragua (National Park Indio Maíz), Costa Rica (Tortuguero National Park, Braulio Carillo National Park), Panama (Barro Colorado Island), and the Pacific Coast of Colombia (Chocó-Darién) as well as Northern Colombia (Caribbean and West of the Cordillera Oriental)

3.2 Transcriptomics as a backbone for future Eco-Evo-Devo studies

(Feindt *et al.* 2018 a, Feindt *et al.* 2018 b)

(iv) Transcriptome profiling in *Megaloprepus*

Although RNA-Seq in non-model organisms promises relatively 'easy to achieve' results, several difficulties must be overcome. These include the rapid degradation and difficult isolation of RNA. Furthermore, transcriptomes usually have a bias towards highly transcribed genes such as the housekeeping genes, while rare genes may simply be not present. And, there is a high percentage of genes without a known function especially if there are no reference genomes from close relatives (cf. [92] and references therein).

In the present transcriptome profiling [90], which is based on RNA from a single larvae thorax that was collected from a natural water filled tree hole, the rigorous read cleaning, assembly evaluation and annotation resulted in a 93% complete transcriptome with the highest number of annotated genes so far (cf. [89, 97, 173] and Figure 7). Furthermore, genes involved in the four major wing developmental signaling pathways (Hedgehog: Hh, Decapentaplegic: Dpp, wingless: wg, and Notch: N) and the wing-patterning network are the first described for *Megaloprepus* and odonates in general.
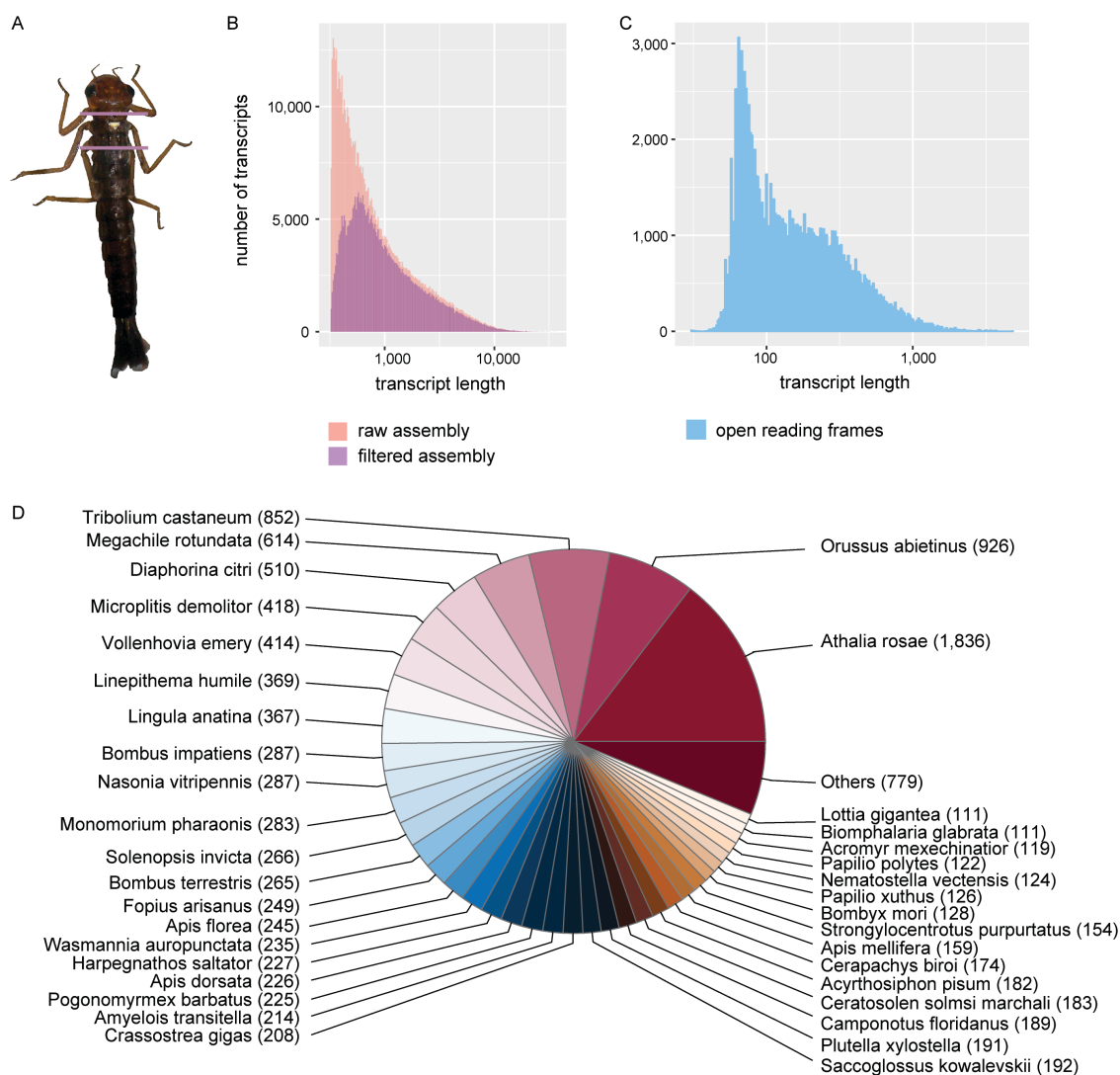
Figure 7: Thorax transcriptome of a *Megaloprepus caerulatus* larva. **A)** Exemplary illustration of a larva showing the thorax section, which was used for RNA extraction and sequencing. **B - C)** Depiction of number of transcripts over transcript length. The plots are displaying the filtering success from the raw reads to the open reading frames, whereby the strict and conserve filtering reduced redundancy and a high number of smaller transcripts. Please note: In B) both axes are logarithmic, whereas in C) only the x-axis. **D)** Functional annotation of the transcriptome showing species and the quantity to which *Megaloprepus* had at least 100 hits to in a graduated BLAST search. Graphs are taken from Feindt *et al.* [90].

Since this is the first genomic data generated for a Neotropical odonate species, it represents a solid backbone for future research on *Megaloprepus* and related odonates. It provides coding sequences, which allows other comparative studies to be undertaken especially in transcriptomics, phylogenomics, and phenotypic evolution. Hereby for example SNP's could identify selective sweeps in comparative analyses. Furthermore, the sequence data can be used for biomonitoring old growth rainforests via gene expression studies and gene evolution in response to climate change (cf. targeted RNA-sequencing).

Finally, because odonates are at the base of flying insects, the evolution of wings is of profound interest and future work could illuminate genes, pathways, and expression patterns related to morphological innovations.

(v) Submission of RNA-Seq data to NCBI

In this more technical project, a comprehensive step-by-step guide for the submission of RNA-Seq data to NCBI was successfully developed. The first protocol within the manuscript in chapter 6.5 displays the necessary steps to submit raw reads to the SRA (Sequence Read Achieve), which precisely includes first the generation of a user account, second the registration of the BioSample and the BioProject and at last the submission of the raw reads itself. The shorter second protocol shows the submission of assemblies to the TSA (Transcriptome Shotgun Assembly).

Most importantly, both protocols are easy to follow and useful for researchers at different level of knowledge in bioinformatics, i.e. because each protocol is accompanied by (i) a resource list for hard- and software, and information that is needed to complete all steps, (ii) internet links to NCBI, (iii) commands editing the data, and (iv) different submitting options. The second main element of this work is an accompanying webpage. Here checklists can be downloaded, and each protocol step is additionally explained visually by screen shots taken from the NCBI webpage. In conclusion, this work illustrates the advantages of an early submission as well as it tries to support scientific progress.

3.3   Mitochondrial genomes–a deeper look into molecular diversity

(Feindt *et al.* 2016 a; Feindt *et al.* 2016 b; Herzog *et al.* 2016)

In the third part of the present work, three mitochondrial genomes were successfully assembled, annotated and published. These genomes were from two zygopteran species *I. elegans* and *M. caerulatus,* and one anisopteran species *A. imperator* representing the two major orders of Odonata.

The gene structure and gene arrangements among odonate mitogenomes only show small differences (Figure 8). However, the length of the A+T rich (control) region is variable and difficult to assemble due to its repetitive structure. Therefore, the newly tested assembly approach, which was iterative mapping accompanied by a PCR based size verification, is considered as more reliable [233]. Using NGS approaches for reconstructing entire mitochondrial genomes are pointing to the future of mitogenomics (cf. mito-metagenomics [225, 226]). A second variable structure is the intergenic spacer regions. For example, *s5* was considered as a distinctive feature between Anisoptera and Zygoptera. However, this non-coding area within the mitochondrial genome was absent in *A. imperator* but present in other previous described dragonfly mitogenomes (cf. [234]). Although the numbering of

those spacer regions is not consistent among published mitogenomes, their presence or absence with the current knowledge cannot be used as a reliable phylogenetic informative character.
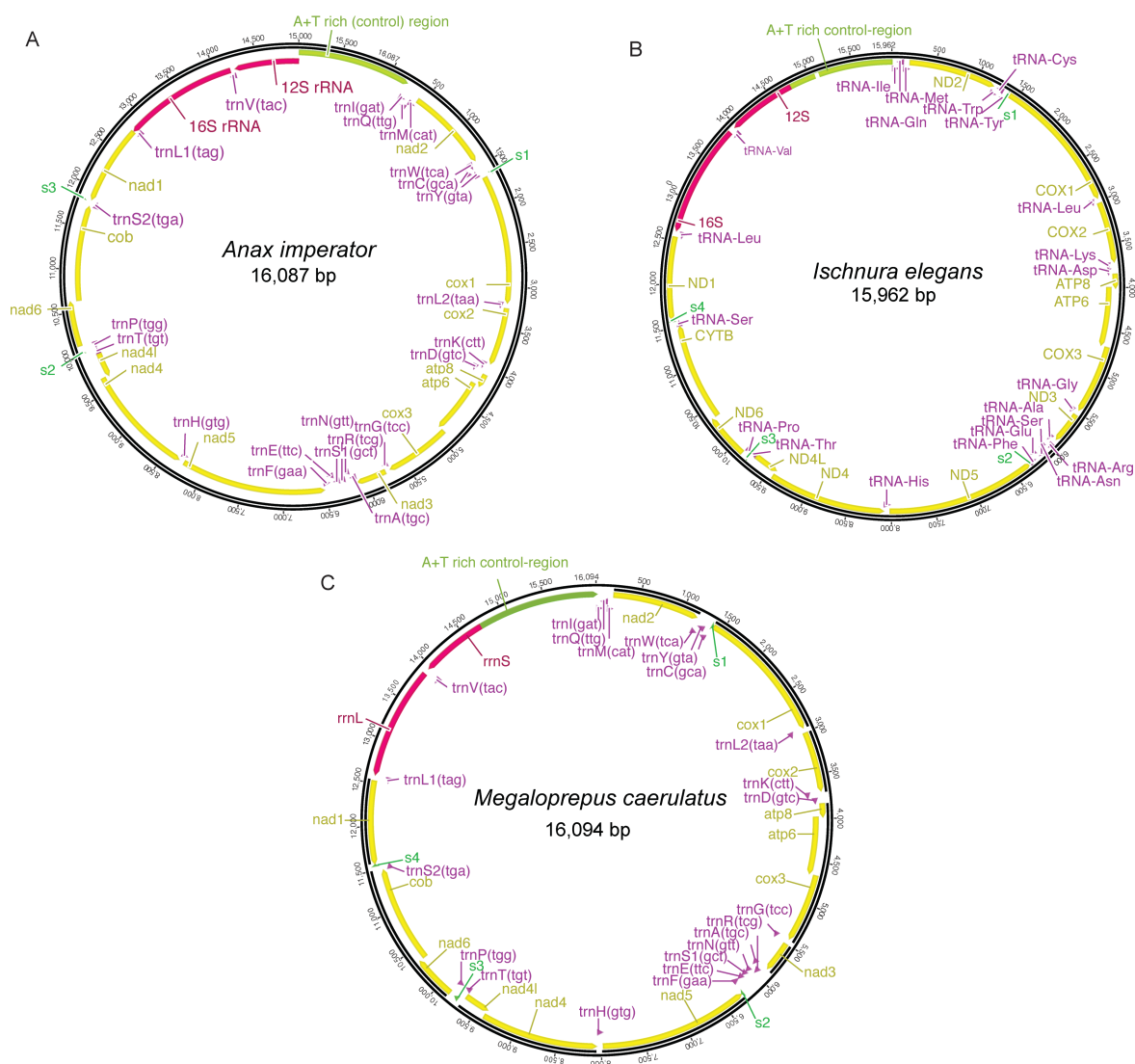
Figure 8: Gene maps for the mitochondrial genomes of **A)** *A. imperator*, **B)** *I. elegans* and **C)** *M. caerulatus*. All genomes are equal in the number of genes and gene arrangements. The differences are mostly in the length of the A+T rich control region and the intergenic spacer.

Finally, these mitogenomes represent a valuable resource for future research in phylogenetics and population genetics as well as in evolution. Mitochondrial and nuclear genes encode mutually for the oxidative phosphorylation (OXPHOS) pathway - the major energy generating system in animal cells [235]. This allows the conclusion, that for a well performance in odonate flight for example, high OXPHOS efficiency must be guaranteed. A general difference between dragon- and damselflies is flight performance, which may be related to the function of the OXPHOS complex and potential positive selection in the

nuclear genes as response to random mutations in the mitogenome [236]. A better flight performance (e.g. such as in *A. imperator*) and the capacity to expand distributional ranges may be related to strong positive selection in the nuclear OXPHOS genes and a general better co-performance. Consequently, future studies shall focus on both generating more mitogenomes of different species and the nuclear genes.

However, studies on the nuclear and mitochondrial OXPHOS genes specifically in *Megaloprepus* could give deeper insights into speciation. First comparisons of all mitochondrial protein coding genes showed genetic distances of 8-10% among *M. caerulatus* and *M. diaboli* sp. nov.. Because, a reduced fitness of interspecific hybrids in *Drosophila* and *Nasonia* wasps was observed [224, 226], similar effects could arise in *Megaloprepus* due to a potential mito-nuclear incomparability.

4. CONCLUSION AND OUTLOOK

The manuscripts presented in this thesis represent a solid frame and important backbones to lift research in Neotropical insects to the next level; particularly in the view of high extinction rates and declining abundances of tropical insect diversity [10, 237]. The methods applied ranged from classic taxonomy to modern NGS and bioinformatics.

Based on extensive fieldwork, modern state of the art taxonomy, and genetics, speciation in *Megaloprepus* could be proved and the originally monospecific genus is split into four independent and geographically distinct species. The diversification was driven by vicariant events in the face of strong ecological niche conservatism and the geological conditions in South and Central America during the Quaternary. Although the speciation mode must therefore be non-adaptive, strong sexual selection on wing coloration observed exclusively in *M. caerulatus* may be related to the prolonged rain season and a higher cloudiness in the forest areas in the Chocó-Darién.

Considering all these facts in the context of continuously ongoing habitat fragmentation as well as past and current environmental change, the present results have implications for both conservation and future research on the genomic base of speciation as well as the evolution of morphological traits. Describing speciation patterns and genes that can cause reproductive incomparability is of great scientific interest and could help identify biodiversity gradients in different animal groups. Because the process of speciation is complex, each study contributes to a specific aspect of the overall picture of speciation research. Little is known about speciation in odonates or about the evolution of phenotypic diversity in this important order at the base of flying insects. *Megaloprepus* could be a model for non-adaptive speciation research. Hereby, large-scale comparative *omic* studies could identify genomic regions under selection and consequently genes that could encode for speciation. Different speciation patterns could be investigated within the Pseudostigmatidae, because the sister genera *Platystigma* and *Mecistogaster* each radiated into at least eight species with different ecological niches. It would be interesting to identify the factors leading to adaptive radiation and which ones not. Furthermore, because it has

been assumed that different expression patterns could cause phenotypic novelties, it is promising to study in detail the effects of temperature changes on the phenotypic appearance. If due to climate change the temperature in the water filled tree holes increase or either decrease, it could be that during larval development pathways within the wing patterning networks (cf. [90, 238-240]) are modified and 'new' taxonomic characters may arise. However, if temperature changes are severe most likely larval development is negatively influenced and fitness could be reduced. This information would underline current conservation needs.

A backbone for such future studies is provided by the comprehensive transcriptome profiling for *Megaloprepus*, with its high annotation score and the first wing genes identified for odonates. Future evolutionary and biodiversity research will rely on genomic methods and because transcriptomics has broad applications, a great variety of research is possible. Expression comparisons across the larval stages could determine genes that are related to wing development or involved in important signaling cascades, which could contribute to the evolution of morphological diversity in basal flying insects. Parallel studies on odonate species with and without wing patches could detect specific coloration genes. Subsequently, knock down studies would show whether coloration could be changed. Furthermore, research on mitochondrial genomes and the potential role of the OXPHOS complexes in habitat disturbances and adaptability as well as in hybrid speciation and hybrid incomparability should receive more attention. Hereby comparisons between odonates with wide ecological ranges and restricted odonates are most promising. Most importantly, because odonates are receiving growing interest in the field of genomics [89, 97, 142, 173], large-scale cooperation's among odonatologists could lead to new concepts in biology or new evolutionary theories.

Odonates play an important role in conservation as bioindicators for environmental health because of their specific and species-dependent habitat requirements [185, 241-244]. Although cryptic species could distort results in a habitat quality assessment, so far only two genera encompassing cryptic dragonflies *Trithemis* [48, 245] and *Orthetrum* [246] have been discovered. But the proof of speciation in *Megaloprepus* as the first cryptic damselfly worldwide reveals species with highly restricted distributional ranges. Furthermore, species exhibiting strong niche conservatism have by definition a low potential to adapt to changing environmental conditions. These effects are multifaceted and depend on species-specific traits such as the genetic background and population size. In *Megaloprepus* particularly, clear cuttings modify the microclimate within forest sites, which in turn reduces tree survival. Extreme patchiness of forest sites further limits migration between populations. This is reflected by a reduced genetic diversity and high genetic differentiation between populations within each of the Mesoamerican species. In this, *M. latipennis* in the Los Tuxtlas region (Veracruz, Mexico) might be most affected as forests are highly fragmented and a decline in population size has already been observed (pers. comm. E. Gonzáles). Equally the population from La Selva in Costa Rica should receive urgent attention because of its low genetic diversity and current forest destruction by a hurricane. A study using, for example, genome-wide SNP's (via Double Digested Restriction Site Associated DNA

sequencing, ddRAD-Seq) could reveal population structures in more detail. As primary forests are highly endangered in the Neotropics (e.g. [5, 12]) the conservation status of all four species must be reconsidered by the IUCN red list.

Many natural ecosystems worldwide are currently in danger of destruction and with them a great number of species. Increasing extinction rates are alarming and socio-economic changes are urgently needed (e.g. [4, 10, 247, 248]). Consequently, two main approaches to conserving tropical forests are possible. First, simply protecting all Neotropical forests that are left today would retain both undiscovered and known biodiversity. Second, greater research efforts on taxonomy in less studied groups could resolve unknown biodiversity so that complete species lists, and distributional ranges could serve as the parameters for defining the conservation status of species and regions. However, there is such a great lack of biodiversity knowledge in Neotropical regions that taxonomists would need hundreds of years to describe the extant species. Consequently, the first approach would be the most efficient, but should be accompanied by taxonomic work.

In summary, because *Megaloprepus* as a genus unites several features it serves not only as a bioindicator for forest health but also as model organism for speciation research in the Neotropics and for studying how new phenotypic characters arise due to ecological variation. Both, conservation and future evolutionary studies are, to my perspective, the most important tasks for natural biologist of our time.

5. REFERENCES

1.   Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346(6210):763-767.
2.   Scheffers BR, Joppa LN, Pimm SL, Laurance WF. What we know and don`t know about Earth's missing biodiversity. Trends in ecology & evolution. 2012;27(9):501-510.
3.   Stork NE, McBroom J, Gely C, Hamilton AJ. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proceedings of the National Academy of Sciences. 2015;112(24):7519-7523.
4.   Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. The biodiversity of species and their rates of extinction, distribution, and protection. Science. 2014;344(6187):1246752.
5.   Laurance WF, Useche DC, Rendeiro J, Kalka M, Bradshaw CJA, Sloan SP, Laurance SG, Campbell M, Abernethy K, Alvarez P. Averting biodiversity collapse in tropical forest protected areas. Nature. 2012;489(7415):290-294.
6.   Butler RA, Laurance WF. New strategies for conserving tropical forests. Trends in ecology & evolution. 2008;23(9):469-472.
7.   Ghazoul J, Sheil D. Tropical rain forest ecology, diversity, and conservation2010.
8.   Council NR. Perspectives on biodiversity: valuing its role in an everchanging world: National Academies Press; 1999.
9.   Costello MJ, May RM, Stork NE. Can we name Earth's species before they go extinct? science. 2013;339(6118):413-416.
10. Lister BC, Garcia A. Climate-driven declines in arthropod abundance restructure a rainforest food web. Proceedings of the National Academy of Sciences. 2018. doi: 10.1073/pnas.1722477115.
11. Magurran AE, Dornelas M. Biological diversity in a changing world. The Royal Society; 2010.
12. Sloan S, Jenkins CN, Joppa LN, Gaveau DLA, Laurance WF. Remaining natural vegetation in the global biodiversity hotspots. Biological Conservation. 2014;177:12-24.
13. Newbold T, Hudson LN, Hill SL, Contu S, Lysenko I, Senior RA, Börger L, Bennett DJ, Choimes A, Collen B. Global effects of land use on local terrestrial biodiversity. Nature. 2015;520(7545):45.
14. Gibson L, Lee TM, Koh LP, Brook BW, Gardner TA, Barlow J, Peres CA, Bradshaw CJA, Laurance WF, Lovejoy TE. Primary forests are irreplaceable for sustaining tropical biodiversity. Nature. 2011;478(7369):378-381.

15. Deutsch CA, Tewksbury JJ, Huey RB, Sheldon KS, Ghalambor CK, Haak DC, Martin PR. Impacts of climate warming on terrestrial ectotherms across latitude. Proceedings of the National Academy of Sciences. 2008;105(18):6668-6672.

16. Frankham R, Briscoe DA, Ballou JD. Introduction to conservation genetics: Cambridge university press; 2002.

17. Gienapp P, Teplitsky C, Alho J, Mills J, Merilä J. Climate change and evolution: disentangling environmental and genetic responses. Molecular ecology. 2008;17(1):167-178.

18. Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London. 1859.

19. Mayr E. Systematics and the origin of species, from the viewpoint of a zoologist: Harvard University Press; 1942.

20. Simpson GG. The Major Features of Evolution: Columbia University Press: New York; 1955.

21. Futuyma DJ. Evolution: Andrew D. Sinauer; 2009.

22. Butlin R, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo CRF, Diao W, Maan ME, Paolucci S, Weissing FJ. What do we need to know about speciation? Trends in ecology & evolution. 2012;27(1):27-39.

23. Boughman JW. Ecological speciation: Selection and the origin of species. 2013.

24. Faria R, Renaut S, Galindo J, Pinho C, Melo‐Ferreira J, Melo M, Jones F, Salzburger W, Schluter D, Butlin R. Advances in ecological speciation: an integrative approach. Molecular ecology. 2014;23(3):513-521.

25. Nosil P. Ecological speciation: Oxford University Press; 2012.

26. Wiens JJ, Graham CH. Niche conservatism: integrating evolution, ecology, and conservation biology. Annu Rev Ecol Evol Syst. 2005;36:519-539.

27. Wiens JJ. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution. 2004;58(1):193-197.

28. Rundell RJ, Price TD. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution. 2009;24(7):394-399.

29. Peterson A, Soberón J, Sánchez-Cordero V. Conservatism of ecological niches in evolutionary time. Science. 1999;285(5431):1265-1267.

30. Nosil P, Gompert Z, Farkas TE, Comeault AA, Feder JL, Buerkle DA, Parchman TL. Genomic consequences of multible speciation processes in a stick insect. Proceedings of the Royal Society of London Series B. 2012;279:5058-5065.

31. Harvey PH, Pagel MD. The comparative method in evolutionary biology: Oxford university press Oxford; 1991.

32. Hutchinson GE. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology. 1957;22:145–159.

33. Pearman PB, Guisan A, Broennimann O, Randin CF. Niche dynamics in space and time. Trends in Ecology & Evolution. 2008;23(3):149-158.

34. Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Jonathan Davies T, Grytnes JA, Harrison SP. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology letters. 2010;13(10):1310-1324.

35. Pyron RA, Costa GC, Patten MA, Burbrink FT. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. Biological Reviews. 2015;90(4):1248-1262.

36. Langerhans RB, Riesch R. Speciation by selection: a framework for understanding ecology's role in speciation. Current Zoology. 2013;59(1):31-52.

37. Sobel JM, Chen GF, Watt LR, Schemske DW. The biology of speciation. Evolution. 2009;64(2):295-315.

38. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution. 2007;22(3):148-155.

39. Mayr E. The growth of biological thought: Diversity, evolution, and inheritance: Harvard University Press; 1982.

40. Sáez AG, Lozano E. Body doubles. Nature. 2005;433(7022):111-111.

41. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How Many Species Are There on Earth and in the Ocean? PLoS Biol. 2011;9(8):e1001127.

42. Frankham R, Ballou JD, Dudash MR, Eldridge MD, Fenster CB, Lacy RC, Mendelson III JR, Porton IJ, Ralls K, Ryder OA. Implications of different species concepts for conserving biodiversity. Biological Conservation. 2012;153:25-31.

43. DeSalle R, Egan MG, Siddall M. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society of London Series B. 2005;360(1462):1905-1916.

44. De Queiroz K. Species concepts and species delimitation. Systematic biology. 2007;56(6):879-886.

45. Domingos FM, Bosque RJ, Cassimiro J, Colli GR, Rodrigues MT, Santos MG, Beheregaray LB. Out of the deep: cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. Molecular Phylogenetics and Evolution. 2014;80:113-124.

46. Ahmadzadeh F, Flecks M, Carretero MA, Mozaffari O, Böhme W, Harris DJ, Freitas S, Rödder D. Cryptic speciation patterns in Iranian rock lizards uncovered by integrative taxonomy. PloS one. 2013;8(12):e80563.

47. Arribas P, Andújar C, Sánchez-Fernández D, Abellán P, Millán A. Integrative taxonomy and conservation of cryptic beetles in the Mediterranean region (Hydrophilidae). Zoologica Scripta. 2013;42(2):182-200.

48. Damm S, Schierwater B, Hadrys H. An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. Molecular Ecology. 2010;19(18):3881-3893.

49. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(41):14812-14817.

50. Melville J, Smith K, Hobson R, Hunjan S, Shoo L. The Role of Integrative Taxonomy in the Conservation Management of Cryptic Species: The Taxonomic Status of Endangered Earless Dragons (Agamidae: *Tympanocryptis*) in the Grasslands of Queensland, Australia. PloS one. 2014;9(7):e101847.

51. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservación Guanacaste, Costa Rica. Proceedings of the National Academy of Sciences. 2008;105(17):6350-6355.

52. Löbl I. Overestimation of molecular and modelling methods and underestimation of traditional taxonomy leads to real problems in assessing and handling of the world's biodiversity. Zootaxa. 2014;3768(4):497-500.

53. Wiens JJ. Species delimitation: new approaches for discovering diversity. Systematic Biology. 2007;56(6):875-878.

54. Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. Integrative taxonomy: a multisource approach to exploring biodiversity. Annual review of entomology. 2010;55:421-438.

55. Padial JM, Miralles A, De la Riva I, Vences M. Review: The integrative future of taxonomy. Frontiers in Zoollogy. 2010;7(16):1-14.

56. Webster M, Sheets HD. A practical introduction to landmark-based geometric morphometrics. Quantitative Methods in Paleobiology. 2010;16:168-188.

57. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology. 2004;71(1):5-16.

58. Outomuro D, Adams DC, Johansson F. The evolution of wing shape in ornamented-winged damselflies (Calopterygidae, Odonata). Evolutionary biology. 2013;40(2):300-309.

59. Wellenreuther M, Svensson EI, Hansson B. Sexual selection and genetic colour polymorphisms in animals. Molecular ecology. 2014;23(22):5398-5414.

60. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology. 1994;3(5):294-299.

61. Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proceedings of the Royal Society B: Biological Sciences. 2008;275(1632):237-247.

62. Bergmann T, Rach J, Damm S, DeSalle R, Schierwater B, Hadrys H. The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. Molecular ecology resources. 2013;13(6):1069-1081.

63. Song H, Buhay JE, Whiting MF, Crandall KA. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the national academy of sciences. 2008;105(36):13486-13491.

64. Paknia O, Bergmann T, Hadrys H. Some 'ant'swers: Application of a layered barcode approach to problems in ant taxonomy. Molecular ecology resources. 2015.

65. Fontaneto D, Flot J-F, Tang CQ. Guidelines for DNA taxonomy, with a focus on the meiofauna. Marine Biodiversity. 2015;45(3):433-451.

66. Hadrys H, Balick M, Schierwater B. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular ecology. 1992;1(1):55-63.

67. Schulman AH. Molecular markers to assess genetic diversity. Euphytica. 2007;158(3):313-321.

68. Hadrys H, Schroth W, Schierwater B, Streit B, Fincke OM. Tree hole odonates as environmental monitors: Non-invasive isolation of polymorphic microsatellites from the neotropical damselfly *Megaloprepus caerulatus*. Conservation Genetics. 2005;6(3):481-483.

69. Hadrys H, Timm J, Streit B, Giere S. A panel of microsatellite markers to study sperm precedence patterns in the emperor dragonfly *Anax imperator* (Odonata: Anisoptera). Molecular ecology notes. 2007;7(2):296-298.

70. Schlötterer C. Evolutionary dynamics of microsatellite DNA. Chromosoma. 2000;109(6):365-371.

71. Damm S, Dijkstra K-DB, Hadrys H. Red drifters and dark residents: the phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa's changing environment (Odonata, Libellulidae, Trithemis). Molecular Phylogenetics and Evolution. 2010;54(3):870-882.

72. Damm S, Hadrys H. A dragonfly in the desert: genetic pathways of the widespread *Trithemis arteriosa* (Odonata: Libellulidae) suggest male-biased dispersal. Organisms Diversity & Evolution. 2012;12(3):267-279.

73. Feindt W, Fincke O, Hadrys H. Still a one species genus? Strong genetic diversification in the world`s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*. Conservation Genetics. 2014;15(2):469-481.

74. Villalobos F. Tree squirrels: A key to understand the historic biogeography of Mesoamerica? Mammalian Biology-Zeitschrift für Säugetierkunde. 2013;78(4):258-266.

75. Frankham R. Where are we in conservation genetics and where do we need to go? Conservation Genetics. 2010;11(2):661-663.

76. Hadrys H, Clausnitzer V, Groeneveld LF. The present role and future promise of conservation genetics for forest Odonates. In: Rivera AC, editor. Forest and dragonflies 4th WDA International Symposium of Odonatology. Sofia: Pensoft Publishers; 2006. p. 279-299.

77. Rowe G, Sweet M, Beebee T. An introduction to molecular ecology: Oxford University Press; 2017.

78. Schlötterer C. The evolution of molecular markers—just a matter of fashion? Nature reviews genetics. 2004;5(1):63.

79. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the entomological Society of America. 1994;87(6):651-701.

80. Sunnucks P. Efficient genetic markers for population biology. Trends in ecology & evolution. 2000;15(5):199-203.

81. Bassham S, Catchen J, Lescak E, von Hippel FA, Cresko WA. Repeated Selection of Alternatively Adapted Haplotypes Creates Sweeping Genomic Remodeling in Stickleback. Genetics. 2018:genetics. 300610.302017.

82. Butlin RK. Population genomics and speciation. Genetica. 2010;138(4):409-418.

83. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends in ecology & evolution. 2014;29(1):51-63.

84. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW, Blaxter ML, Mallet J, Jiggins CD. Genome‐wide patterns of divergence and gene flow across a butterfly radiation. Molecular ecology. 2013;22(3):814-826.

85. Wachi N, Matsubayashi KW, Maeto K. Application of next‐generation sequencing to the study of non‐model insects. Entomological Science. 2018;21(1):3-11.

86. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols. 2013;8(8). Epub 1512.

87. Oppenheim SJ, Baker RH, Simon S, DeSalle R. We can't all be supermodels: the value of comparative transcriptomics to the study of non‐model insects. Insect molecular biology. 2015;24(2):139-154.

88. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57-63.

89. Chauhan P, Hansson B, Kraaijeveld K, de Knijff P, Svensson EI, Wellenreuther M. De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. BMC genomics. 2014;15(1):808.

90. Feindt W, Oppenheim SJ, DeSalle R, Goldstein PZ, Hadrys H. Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world's largest wings. PloS one. 2018;13(1):e0189898.

91. Mehr SFP, DeSalle R, Kao H-T, Narechania A, Han Z, Tchernov D, Pieribone V, Gruber DF. Transcriptome deep-sequencing and clustering of expressed isoforms from *Favia* corals. BMC genomics. 2013;14(1):546.

92. Oppenheim SJ, Feindt W, DeSalle R, Goldstein PZ. De Novo characterization of transcriptomes from two North American Papaipema stem-borers (Lepidoptera: Noctuidae). PloS one. 2018;13(1):e0191061.

93. Simon S, Narechania A, DeSalle R, Hadrys H. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. Genome Biology and Evolution. 2012;4(12):1295-1309.

94. Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC genomics. 2008;9(1):1.

95. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC genomics. 2010;11(1):180.

96. Alves-Carvalho S, Aubert G, Carrere S, Cruaud C, Brochot AL, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J. Full‐length de novo assembly of RNA‐seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. The Plant Journal. 2015;84(1):1-19.

97. Chauhan P, Wellenreuther M, Hansson B. Transcriptome profiling in the damselfly *Ischnura elegans* identifies genes with sex-biased expression. BMC genomics. 2016;17(1):985.

98. Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, Monnin D, El Filali A, Carareto CM, Vieira C. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Research. 2016:gkw655.

99. Simon S, Sagasser S, Saccenti E, Brugler MR, Schranz ME, Hadrys H, Amato G, DeSalle R. Comparative transcriptomics reveals developmental turning points during embryogenesis of a hemimetabolous insect, the damselfly *Ischnura elegans*. Scientific Reports. 2017;7(1):2045-2322.

100. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;403(6772):853.
101. Merckx V, Smets E, Specht C. Biogeography and Conservation2013.
102. Laurance WF. Reflections on the tropical deforestation crisis. Biological conservation. 1999;91(2-3):109-117.
103. Norman M. The primary source: tropical forests and our future1984.
104. Mayhew PJ, Jenkins GB, Benton TG. A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. Proceedings of the Royal Society of London B: Biological Sciences. 2008;275(1630):47-53.
105. Hoorn C, Wesselingh F, Ter Steege H, Bermudez M, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson C, Figueiredo J. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. science. 2010;330(6006):927-931.
106. Hazzi NA, Moreno JS, Ortiz-Movliav C, Palacio RD. Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. Proceedings of the National Academy of Sciences. 2018:201803908.
107. Bagley JC, Johnson JB. Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. Biological Reviews. 2014;89(4):767-790.
108. Antonelli A, Zizka A, Carvalho FA, Scharn R, Bacon CD, Silvestro D, Condamine FL. Amazonia is the primary source of Neotropical biodiversity. Proceedings of the National Academy of Sciences. 2018:201713819.
109. Antonelli A, Sanmartín I. Why are there so many plant species in the Neotropics? Taxon. 2011;60(2):403-414.
110. Armijo R, Lacassin R, Coudurier-Curveur A, Carrizo D. Coupled tectonic evolution of Andean orogeny and global climate. Earth-Science Reviews. 2015;143:1-35.
111. Antonelli A, Nylander JA, Persson C, Sanmartín I. Tracing the impact of the Andean uplift on Neotropical plant evolution. Proceedings of the National Academy of Sciences. 2009;106(24):9749-9754.
112. Gregory-Wodzicki KM. Uplift history of the Central and Northern Andes: a review. Geological Society of America Bulletin. 2000;112(7):1091-1105.
113. Coates AG, Stallard RF. How old is the Isthmus of Panama? Bulletin of Marine Science. 2013;89(4):801-813.
114. Cione AL, Gasparini GM, Soibelzon E, Soibelzon LH, Tonni EP. The great American biotic interchange: a South American perspective: Springer; 2015.
115. Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences. 2015;112(19):6110-6115.
116. Marshall LG. Land mammals and the great American interchange. American Scientist. 1988;76(4):380-388.
117. Stehli FG, Webb SD. The great American biotic interchange: Springer Science & Business Media; 2013.
118. Woodburne MO. The Great American Biotic Interchange: dispersals, tectonics, climate, sea level and holding pens. Journal of Mammalian Evolution. 2010;17(4):245-264.
119. Rull V. Neotropical biodiversity: timing and potential drivers. Trends in ecology & evolution. 2011;26(10):508-513.
120. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405(6789):907.
121. Hooghiemstra H, van der Hammen T. Neogene and Quaternary development of the neotropical rain forest: the forest refugia hypothesis, and a literature overview. Earth-Science Reviews. 1998;44(3-4):147-183.
122. Camps GA, Martínez-Meyer E, Verga AR, Sérsic AN, Cosacov A. Genetic and climatic approaches reveal effects of Pleistocene refugia and climatic stability in an old giant of the Neotropical Dry Forest. Biological Journal of the Linnean Society. 2018;125(2):401-420.
123. Garzón‐Orduña IJ, Benetti‐Longhini JE, Brower AV. Timing the diversification of the Amazonian biota: butterfly divergences are consistent with Pleistocene refugia. Journal of biogeography. 2014;41(9):1631-1638.

124. Rull V. Pleistocene speciation is not refuge speciation. Journal of Biogeography. 2015;42(3):602-604.

125. Wright SJ, Muller‐Landau HC. The Future of Tropical Forest Species 1. Biotropica: The Journal of Biology and Conservation. 2006;38(3):287-301.

126. De Sy V, Herold M, Achard F, Beuchle R, Clevers J, Lindquist E, Verchot L. Land use patterns and related carbon losses following deforestation in South America. Environmental Research Letters. 2015;10(12):124004.

127. FAO. State of the World's Forests. Electronic Publishing Policy and Support Branch. Communication Division Roma; 2007.

128. Laurance WF. Have we overstated the tropical biodiversity crisis? Trends in Ecology & Evolution. 2007;22(2):65-70.

129. Aide TM, Clark ML, Grau HR, López‐Carr D, Levy MA, Redo D, Bonilla‐Moheno M, Riner G, Andrade‐Núñez MJ, Muñiz M. Deforestation and Reforestation of Latin America and the Caribbean (2001–2010). Biotropica. 2013;45(2):262-271.

130. Watson JE, Shanahan DF, Di Marco M, Allan J, Laurance WF, Sanderson EW, Mackey B, Venter O. Catastrophic declines in wilderness areas undermine global environment targets. Current Biology. 2016;26(21):2929-2934.

131. Gibbs HK, Ruesch AS, Achard F, Clayton MK, Holmgren P, Ramankutty N, Foley JA. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proceedings of the National Academy of Sciences. 2010;107(38):16732-16737.

132. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK. Global consequences of land use. science. 2005;309(5734):570-574.

133. Magrın G, Gay García C, Cruz Choque D, Giménez J, Moreno A, Nagy G, Nobre C, Villamizar A. Latin America. Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Forth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge. 2007:581-615.

134. Laurance WF. Edge effects in tropical forest fragments: application of a model for the design of nature reserves. Biological conservation. 1991;57(2):205-219.

135. Laurance WF. Forest-climate interactions in fragmented tropical landscapes. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2004;359(1443):345-352.

136. Bolt LM, Schreier AL, Voss KA, Sheehan EA, Barrickman NL, Pryor NP, Barton MC. The influence of anthropogenic edge effects on primate populations and their habitat in a fragmented rainforest in Costa Rica. Primates. 2018;59(3):301-311.

137. Mendoza E, Fay J, Dirzo R. A quantitative analysis of forest fragmentation in Los Tuxtlas, southeast Mexico: patterns and implications for conservation. Revista Chilena de Historia Natural. 2005;78(3).

138. Pulido F, Berthold P. Microevolutionary response to climatic change. Advances in Ecological Research. 2004;35:151-183.

139. Beheregaray LB. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Molecular Ecology. 2008;17(17):3754-3774.

140. Morrone JJ. Cladistic biogeography of the Neotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics. 2014;30(2):202-214.

141. Schultz J. The ecozones of the world: Springer; 2005.

142. Bybee S, Córdoba-Aguilar A, Duryea MC, Futahashi R, Hansson B, Lorenzo-Carballa MO, Schilder R, Stoks R, Suvorov A, Svensson EI. Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. Frontiers in zoology. 2016;13(1):46.

143. Kalkman VJ, Clausnitzer V, Dijkstra K-DB, Orr AG, Paulson DR, van Tol J. Global diversity of dragonflies (Odonata) in freshwater. Hydrobiologia. 2008;595(1):351-363.

144. Suhling F, Sahlén G, Gorb S, Kalkman VJ, Dijkstra K-DB, van Tol J. Order Odonata. Thorp and Covich's Freshwater Invertebrates (Fourth Edition): Elsevier; 2015. p. 893-932.

145. Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, Kück P, Ebersberger I, Walzl M, Pass G, Breuers S. A phylogenomic approach to resolve the arthropod tree of life. Molecular biology and Evolution. 2010;27(11):2451-2464.

146. Córdoba-Aguilar A. Dragonflies and Damselflies: Model Organisms for Ecological and Evolutionary Research. Oxford: Oxford University Press; 2008.

147. Bried JT, Samways MJ. A review of odonatology in freshwater applied ecology and conservation science. Freshwater Science. 2015;34(3):1023-1031.

148. Corbet PS. Dragonflies: Behaviour and Ecology of Odonata. Essex, England: Harley Books; 1999.

149. Fincke OM, editor Use of forest and tree species, and dispersal by giant damselflies (Pseudostigmatidae): their prospects in fragmented forests. Forest and dragonflies 4th WDA International Symposium of Odonatology Pensoft, Sofia; 2006.

150. Parker GA. Sperm competition and its evolutionary consequences in the insects. Biological Reviews. 1970;45(4):525-567.

151. Waage JK. Dual function of the damselfly penis: sperm removal and transfer. Science. 1979;203(4383):916-918.

152. Cordero-Rivera A, Córdoba-Aguilar A. Selective forces propelling genitalic evolution in Odonata. In: Leonard J, Cordoba-Aguilar A, editors. The evolution of primary sexual characters in animals: Oxford University Press, USA; 2010. p. 332-352.

153. Córdoba-Aguilar A, Uhía E, Rivera AC. Sperm competition in Odonata (Insecta): the evolution of female sperm storage and rivals' sperm displacement. Journal of Zoology. 2003;261(04):381-398.

154. Hosken DJ, Stockley P. Sexual selection and genital evolution. Trends in Ecology & Evolution. 2004;19(2):87-93.

155. Rivera AC, Andrés J, Córdoba-Aguilar A, Utzeri C, Noor M. Postmating sexual selection: allopatric evolution of sperm competition mechanisms and genital morphology in calopterygid damselflies (Insecta: Odonata). Evolution. 2004;58(2):349-359.

156. Simmons LW. Sexual selection and genital evolution. Austral Entomology. 2014;53(1):1-17.

157. Xu M, Fincke OM. Ultraviolet wing signal affects territorial contest outcome in a sexually dimorphic damselfly. Animal Behaviour. 2015;101:67-74.

158. Le Rouzic A, Hansen TF, Gosden TP, Svensson EI. Evolutionary time-series analysis reveals the signature of frequency-dependent selection on a female mating polymorphism. The American Naturalist. 2015;185(6):E182-E196.

159. Sanmartín-Villar I, Cordero-Rivera A. Female colour polymorphism and unique reproductive behaviour in *Polythore* damselflies (Zygoptera: Polythoridae). Neotropical entomology. 2016;45(6):658-664.

160. Svensson EI, Abbott J, Härdling R. Female polymorphism, frequency dependence, and rapid evolutionary dynamics in natural populations. The American Naturalist. 2005;165(5):567-576.

161. Willink B, Svensson E, editors. Intra-and intersexual differences in parasite resistance and female fitness tolerance in a polymorphic insect. Proc R Soc B; 2017: The Royal Society.

162. Futahashi R. Color vision and color formation in dragonflies. Current opinion in insect science. 2016;17:32-39.

163. Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T. Extraordinary diversity of visual opsin genes in dragonflies. Proceedings of the National Academy of Sciences. 2015;112(11):E1247-E1256.

164. Svensson EI. Non-ecological speciation, niche conservatism and thermal adaptation: how are they connected? Organisms Diversity & Evolution. 2012;12(3):229-240.

165. Wellenreuther M, Sánchez-Guillén RA. Nonadaptive radiation in damselflies. Evolutionary applications. 2016;9(1):103-118.

166. McPeek MA, Symes LB, Zong DM, McPeek CL. Species recognition and patterns of population variation in the reproductive structures of a damselfly genus. Evolution. 2011;65(2):419-428.

167. Machado AB, Lacerda DSS. Revalidation of *Platystigma* Kennedy, 1920, with a synopsis of the *quadratum* species group and the description of three new species (Odonata: Pseudostigmatidae). Zootaxa. 2017;4242(3):493-516.

168. Dijkstra K-DB, Kipping J, Mézière N. Sixty new dragonfly and damselfly species from Africa (Odonata). Odonatologica. 2015;44(4):447-678.

169. Bota-Sierra CA, Novelo-Gutierrez R, Amaya-Vallejo V. The rediscovery and redescription of Epigomphus pechumani Belle, 1970 (Odonata: Gomphidae), with a description of its female from the Western Colombian Andes. Zootaxa. 2017;4306(3):419-427.

170. Bota-Sierra CA. Two new species of the family Philogeniidae (Odonata: Zygoptera) from the Western Colombian Andes. International Journal of Odonatology. 2017;20(3-4):137-150.

171. Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P. Geographic mode of speciation and genomic divergence. Annual Review of Ecology, Evolution, and Systematics. 2013;44:73-97.

172. Ioannidis P, Simao FA, Waterhouse RM, Manni M, Seppey M, Robertson HM, Misof B, Niehuis O, Zdobnov EM. Genomic Features of the Damselfly *Calopteryx splendens:* Representing a Sister Clade to Most Insect Orders. Genome biology and evolution. 2017;9(2):415-430.

173. Shanku AG, McPeek MA, Kern AD. Functional Annotation and Comparative Analysis of a Zygopteran Transcriptome. G3: Genes| Genomes| Genetics. 2013;3(4):763-770.

174. Jordan S, Simon C, Polhemus D. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus Megalagrion (Odonata: Coenagrionidae). Systematic Biology. 2003;52(1):89-109.

175. Misof B. Diversity of Anisoptera (Odonata): infering speciation processes from patterns of morphological diversity. Zoology. 2002;105(4):355-365.

176. Svensson EI, Eroukhmanoff F, Friberg M. Effects of natural and sexual selection on adaptive population divergence and premating isolation in a damselfly. Evolution. 2006;60(6):1242-1253.

177. Wellenreuther M, Larson KW, Svensson EI. Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. Ecology. 2012;93(6):1353-1366.

178. McPeek MA. Determination of species composition in the *Enallagma* damselfly assemblages of permanent lakes. Ecology. 1990:83-98.

179. Callahan MS, McPeek MA. Multi-locus phylogeny and divergence time estimates of *Enallagma* damselflies (Odonata: Coenagrionidae). Molecular phylogenetics and evolution. 2016;94:182-195.

180. Guillermo-Ferreira R, Therézio EM, Gehlen MH, Bispo PC, Marletta A. The role of wing pigmentation, UV and fluorescence as signals in a Neotropical damselfly. Journal of insect behavior. 2014;27(1):67-80.

181. Herrera MS, Kuhn WR, Lorenzo-Carballa MO, Harding KM, Ankrom N, Sherratt TN, Hoffmann J, Van Gossum H, Ware JL, Cordero-Rivera A. Mixed signals? Morphological and molecular evidence suggest a color polymorphism in some neotropical *Polythore* damselflies. PloS one. 2015;10(4):e0125074.

182. Koroiva R, Pepinelli M, Rodrigues ME, de Oliveira Roque F, Lorenz-Lemke AP, Kvist S. DNA barcoding of odonates from the Upper Plata basin: Database creation and genetic diversity estimation. PloS one. 2017;12(8):e0182283.

183. Pires MM, Périco E, Renner S, Sahlén G. Predicting the effects of future climate change on the distribution of an endemic damselfly (Odonata, Coenagrionidae) in subtropical South American grasslands. Journal of Insect Conservation. 2018;22(2):303-319.

184. Suárez‑Tovar C, Sarmiento C. Beyond the wing planform: morphological differentiation between migratory and nonmigratory dragonfly species. Journal of evolutionary biology. 2016;29(4):690-703.

185. Valente-Neto F, de Oliveira Roque F, Rodrigues ME, Juen L, Swan CM. Toward a practical use of Neotropical odonates as bioindicators: Testing congruence across taxonomic resolution and life stages. Ecological indicators. 2016;61:952-959.

186. Vivas-Santeliz J, De Marmels J. Current knowledge of Odonata in Venezuela: Diversity and distribution of endemic taxa. Odonatologica. 2017;46(1/2):35-54.

187. Paulson D. The Importance of Forests to Neotropical Dragonflies. In: Rivera AC, editor. Forests and Dragonflies Fourth WDA International Symposium of Odonatology. Pontevedra (Spain): Pensoft Publishers; 2006. p. 79-101.

REFERENCES

188. Herrera MS, Realpe E, Salazar C. A neotropical polymorphic damselfly shows poor congruence between genetic and traditional morphological characters in Odonata. Molecular phylogenetics and evolution. 2010;57(2):912-917.

189. Fincke OM, Hedström I. Differences in forest use and colonization by Neotropical tree‐hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conservation. Studies on Neotropical Fauna and Environment. 2008;43(1):35-45.

190. Garrison RW, von Ellenrieder N. A synonymic list of the New World Odonata. Argia. 1991;3(2):30.

191. Garrison RW, von Ellenrieder N, Louton JA. Damselfly Genera of the New World, an Illustrated and Annotated Key to the Zygoptera. Baltimore, MD: The Johns Hopkins University Press; 2010. 490 p.

192. Groeneveld LF, Clausnitzer V, Hadrys H. Convergent evolution of gigantism in damselflies of Africa and South America? Evidence from nuclear and mitochondrial sequence data. Molecular Phylogenetics and Evolution. 2007;42(2):339-346.

193. Steinmann H. World Catalogue of Odonata, Vol. 1: Zygoptera. Wermuth H, Fischer M, editors. New York: Walter de Gruyter, Berlin; 1997.

194. Heckman CW. Encyclopedia of South American Aquatic Insects: Odonata - Zygoptera: Springer Netherlands; 2008. 693 p.

195. Yanoviak SP. Community structure in water-filled tree holes of Panama: effects of hole height and size. Selbyana. 1999:106-115.

196. Yanoviak S, Fincke O. Sampling methods for water-filled tree holes and their artificial analogues. Insect sampling in forest ecosystems Blackwell Publishing, London. 2005:168-185.

197. Kitching RL. Food webs and container habitats: the natural history and ecology of phytotelmata: Cambridge University Press; 2000.

198. Kitching R. An ecological study of water-filled tree-holes and their position in the woodland ecosystem. The Journal of Animal Ecology. 1971:281-302.

199. Dijkstra K-DB, Kalkman VJ, Dow RA, Stokvis FR, Van Tol J. Redefining the damselfly families: a comprehensive molecular phylogeny of Zygoptera (Odonata). Systematic Entomology. 2014;39(1):68-96.

200. Tsuda S. A distributional list of World Odonata, 1-362. Osaka Pref, Japan. 1991.

201. Sjöstedt Y. Wissenschaftliche Ergebnisse der schwedischen entomologischen Reise des Herrn Dr. A. Roman in Amazonas 1914-1915: I. Odonata1918.

202. Selys LEd. Révision du synopsis des Agrionines, premiére partie comprenant des légions Psuedostigma – Podagrion – Platycnemis et Protoneura. Mémoire Cour Académie Royale Belgique. 1886;38 (4):[1]+iv+233.

203. Selys LEd. Synopsis des Agrionines. Première Légion. – Pseudostigma –. Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux - Arts de Belgique. 1860;2(10):9-27.

204. Schmidt E. Odonata nebst Bemerkungen über die *Anomisma* und *Chalcopteryx* des Amazonas-Gebiets. In: Titschack E, editor. 1941-1942 Beiträge zur Fauna Perus nach der Ausbeute der Hamburger Südperu Expedition 1936. 21942. p. 225-276.

205. Ris F. Libellen (Odonata) aus der Region der amerikanischen Kordilleren von Costarica bis Catamarca. Archiv für Naturgeschichte. 1916;82A:1-197 + 192 pl.

206. Rambur M. Histoire Naturelle des Insectes. Névroptères. Paris1842. 534 p.

207. McLachlan R. II. Notes on Odonata, of the subfamilies Corduliina, Calopterygina, and Agrionina (Légion Pseudostigma), collected by Mr. Buckley, in the district of the Rio Bobonaza, in Ecuador. Transactions of the Royal Entomological Society of London. 1881;29(1):25-34.

208. McLachlan R. On some new and little-known forms of Agrionina (Légion Pseudostigma, de Selys). Entomologists Monthly Magazine. 1877;14:86-88.

209. Machado A. *Mecistogaster martinezi* n. sp. from the forests of Bolivia (Odonata–Pseudostigmatidae). Ciência e Cultura. 1985;37(7):854.

210. Hagen HA. Zur Odonaten-Fauna von Neu-Granada nach Lindig's Sammlungen. Stettin Entomologische Zeitung. 1869;30:256-263.

211. Fincke OM. Giant damselflies in a tropical forest: reproductive biologyof *Megaloprepus coerulatus* with notes on *Mecistogaster* (Zygoptera: Pseudostigmatidae). Advances in Odonatology. 1984;2(1):13-27.

212. Fabricius JC. Genera insectorum: eorumque characteres naturales secundam numerum, figuram, situm et proportionem, omnium partium oris adiecta mantissa specierum nuper detectarum1776. 310 p.

213. Drury D. Illustrations of natural history. London: White; 1782. 76 p.

214. Clausnitzer V, Lindeboom M. Natural history and description of the dendrolimnetic larva of *Coryphagrion grandis* (Odonata). International Journal of Odonatology. 2002;5(1):29-44.

215. Burmeister H. Handbuch der Entomologie. Zweiter Band. Besondere Entomologie.: Enslin, Berlin.; 1839.

216. Calvert P. Odonata. In: Goldman F, editor. Biologia Centrali-Americana Vol 50: Insecta, Neuroptera: Porter Dulau & Co., London; 1901-1908. p. 17-420.

217. Schultz TD, Fincke OM. Structural colours create a flashing cue for sexual recognition and male quality in a Neotropical giant damselfly. Functional Ecology. 2009;23(4):724-732.

218. Fincke OM. Interspecific competition for tree holes: consequences for mating systems and coexistence in neotropical damselflies. American Naturalist. 1992:80-101.

219. Fincke OM. Population regulation of a tropical damselfly in the larval stage by food limitation, cannibalism, intraguild predation and habitat drying. Oecologia. 1994;100(1-2):118-127.

220. Fincke OM, Hadrys H. Unpredictable offspring survivorship in the damselfly, *Megaloprepus coerulatus*, shapes parental behavior, constrains sexual selection, and challenges traditional fitness estimates. Evolution. 2001;55(4):762-772.

221. Fincke OM. Organization of predator assemblages in Neotropical tree holes : effects of abiotic factors and priority. Oxford, ROYAUME-UNI: Blackwell; 1999.

222. Khazan ES. Tests of biological corridor efficacy for conservation of a Neotropical giant damselfly. Biological Conservation. 2014;177:117-125.

223. Simon S, Hadrys H. A comparative analysis of complete mitochondrial genomes among Hexapoda. Molecular phylogenetics and evolution. 2013;69(2):393-403.

224. Gibson J, Niehuis O, Peirson B, Cash E, Gadau J. Genetic and developmental basis of F2 hybrid breakdown in Nasonia parasitoid wasps. Evolution. 2013;67(7):2124-2132.

225. DeSalle R, Schierwater B, Hadrys H. MtDNA: The small workhorse of evolutionary studies. Front Biosci (Landmark Ed). 2017;22:873-887.

226. Crampton-Platt A, Timmermans MJ, Gimmel ML, Kutty SN, Cockerill TD, Vun Khen C, Vogler AP. Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. Molecular biology and evolution. 2015;32(9):2302-2316.

227. Cameron SL. Insect mitochondrial genomics: implications for evolution and phylogeny. Annual review of entomology. 2014;59:95-117.

228. Morales HE, Pavlova A, Joseph L, Sunnucks P. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Molecular Ecology. 2015;24(11):2820-2837.

229. Sunnucks P, Morales HE, Lamb AM, Pavlova A, Greening C. Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. Frontiers in genetics. 2017;8:25.

230. Lin C-P, Chen M-Y, Huang J-P. The complete mitochondrial genome and phylogenomics of a damselfly, Euphaea formosa support a basal Odonata within the Pterygota. Gene. 2010;468(1):20-29.

231. DeClerck FAJ, Chazdon R, Holl KD, Milder JC, Finegan B, Martinez-Salinas A, Imbach P, Canet L, Ramos Z. Biodiversity conservation in human-modified landscapes of Mesoamerica: Past, present and future. Biological Conservation. 2010;143(10):2301-2313.

232. Harvey CA, Komar O, Chazdon R, Ferguson BG, Finegan B, Griffith DM, MartÍNez-Ramos M, Morales H, Nigh R, Soto-Pinto L, Van Breugel M, Wishnie M. Integrating Agricultural Landscapes with Biodiversity Conservation in the Mesoamerican Hotspot. Conservation Biology. 2008;22(1):8-15.

233. Feindt W, Herzog R, Osigus H-J, Schierwater B, Hadrys H. Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera). Mitochondrial DNA Part B. 2016;1(1):574-576.

234. Herzog R, Osigus HJ, Feindt W, Schierwater B, Hadrys H. The complete mitochondrial genome of the emperor dragonfly *Anax imperator* LEACH, 1815 (Odonata: Aeshnidae) via NGS sequencing. Mitochondrial DNA Part B. 2016;1(1):783-786.

235. Gibson JD, Niehuis O, Verrelli BC, Gadau J. Contrasting patterns of selective constraints in nuclear-encoded genes of the oxidative phosphorylation pathway in holometabolous insects and their possible role in hybrid breakdown in Nasonia. Heredity. 2010;104(3):310.

236. Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. Trends in ecology & evolution. 2004;19(12):645-653.

237. Seebacher F, White CR, Franklin CE. Physiological plasticity increases resilience of ectothermic animals to climate change. Nature Climate Change. 2015;5(1):61.

238. Yan S-J, Gu Y, Li WX, Fleming RJ. Multiple signaling pathways and a selector protein sequentially regulate *Drosophila* wing development. Development. 2004;131(2):285-298.

239. Shimmi O, Matsuda S, Hatakeyama M, editors. Insights into the molecular mechanisms underlying diversified wing venation among insects. Proc R Soc B; 2014: The Royal Society.

240. Abouheif E, Wray GA. Evolution of the gene network underlying wing polyphenism in ants. Science. 2002;297(5579):249-252.

241. Dijkstra K-DB, Monaghan MT, Pauls SU. Freshwater Biodiversity and Insect Diversification. Annual review of entomology. 2014;59:143.

242. Hassall C, Thompson DJ, French GC, Harvey IF. Historical changes in the phenology of British Odonata are related to climate. Global Change Biology. 2007;13(5):933-941.

243. Júnior CdSM, Juen L, Hamada N. Analysis of urban impacts on aquatic habitats in the central Amazon basin: adult odonates as bioindicators of environmental quality. Ecological indicators. 2015;48:303-311.

244. Sahlén G, Ekestubbe K. Identification of dragonflies (Odonata) as indicators of general species richness in boreal forest lakes. Biodiversity & Conservation. 2001;10(5):673-690.

245. Damm S, Hadrys H. *Trithemis morrisoni* sp. nov. and *T. palustris* sp. nov. from the Okavango and Upper Zambezi Floodplains previously hidden under *T. stictica* (Odonata: Libellulidae). International Journal of Odonatology. 2009;12(1):131-145.

246. Yong HS, Lim P-E, Tan J, Ng YF, Eamsobhana P, Suana IW. Molecular phylogeny of *Orthetrum* dragonflies reveals cryptic species of *Orthetrum pruinosum*. Scientific reports. 2014;4.

247. Crooks KR, Burdett CL, Theobald DM, King SR, Di Marco M, Rondinini C, Boitani L. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proceedings of the National Academy of Sciences. 2017;114(29):7635-7640.

248. Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PloS one. 2017;12(10):e0185809.

6. PUBLICATIONS AND MANUSCRIPTS UPON WHICH THIS THESIS IS BASED

Still a one species genus? Strong genetic diversification in the world´s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus.*

Wiebke Feindt[1,2 *], Ola Fincke[3] and Heike Hadrys[1,2,4]

[1]ITZ, Ecology & Evolution, University of Veterinary Medicine Hannover,
Bünteweg 17d, D-30559 Hannover, Germany
[2]American Museum of Natural History, The Sackler Institute for Comparative Genomics,
New York, NY 10024, U.S.A.
[3]Ecology and Evolutionary Biology Program, Department of Biology, University of
Oklahoma Norman, Oklahoma, OK 73019, U.S.A.
[4]Yale University, Department of Molecular, Cellular and Developmental Biology
165 Prospect Street, CT 06511, U.S.A.

[*] Author for correspondence
Email: wiebke.feindt@ecolevol.de
Phone: +49 511 953 8402
Fax: +49 511 953 8584

Still a one species genus? Strong genetic diversification in the world´s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*.

Abstract

Mesoamerican biodiversity is increasingly threatened by anthropogenic destruction of natural land cover. Habitat degradation and climate change are primary threats to specialized forest odonate species that are important model organisms for forest health and defining conservation units. The extreme niche specialization of *Megaloprepus caerulatus*, the world's largest extant odonate, makes it well suited as an indicator for changing environmental conditions. *Megaloprepus*, which is considered to be a monospecific genus, is highly dependent on old growth forests whose water filled tree holes are limiting reproductive resources for this species. Here, we focus on the question how historical and recent fragmentation events, strong niche conservatism and ecological conditions have affected population dynamics, viability and the species status in this evolutionarily old genus. Two mitochondrial sequence markers (ND1 and 16S rRNA) and a set of microsatellites were used to analyze population structure and genetic diversity of *M. caerulatus* in the northern part of its distributional range. Results suggested an absence of gene flow and no shared haplotypes among the study populations. Statistical parsimony indicated high sub-structuring among populations with sequence diversity similar to levels found at the species level compared to other odonates. In sum, the genetic data suggest that *Megaloprepus* may actually consist of more than one species. The taxonomic status of the group should be revised in light of the three distinct genetic clusters found in different forest regions. The results may also allow insights into the impact of recent and historical habitat fragmentation on a strong Neotropical forest restricted insect species.

Keywords: conservation genetics, speciation, Neotropical primary forests, Odonata

Introduction

Neotropical forests are among the most species rich biogeographical regions on earth (Myers et al. 2000). Mesoamerican tropical forests present a particularly high biodiversity, which results from the confluence of flora and fauna from the two major biogeographic regions of North and South America (Stehli & Webb 1985). South migration of North American forms (Nearctic species) dominated the Cenozoic and north expansion of Neotropical biota occurred in the late Pliocene (Rich & Rich 1983; Stehli & Webb 1985). Geological history, high climate variation over a small area, and a diverse geography is further contributing to the unique biodiversity and high levels of endemism found in Mesoamerica (Myers et al. 2000; Calderón et al. 2004; Mayhew et al. 2008). Not

surprisingly only a fraction of its forest biodiversity has been described as there are areas that have yet to be explored.

Today Neotropical rainforests face on-going destruction due to human activities that result in a contemporary assortment of forest patches surrounded by urban areas (e.g. Magurran & Dornelas 2010). High fragmentation rates along with climate change are the driving forces for species extinction, loss of biological dynamics, separation of populations, and declines in population size and viability (Pimm & Raven 2000; Wright & Muller-Landau 2006; Balint et al. 2011; Brodie et al. 2011). Highly specialized species that are sensitive to forest instability are valuable markers for conservation management, since forest disturbance can rapidly lead to impacts on population structure and ultimately to population decline (Pimm et al. 1995). Thus, research on the genetic diversity, population structure and dispersal of 'marker species' has become a promising approach to translate "conservation science into conservation practice" (Sutherland et al. 2009).

Insects, with their extraordinary species richness comprising an estimated 80% of total global biodiversity of macrofauna (Bisby et al. 2013), diverse habitat adaptations, and their role in essential ecosystem functions (Lewis & Basset 2007; Gullan & Cranston 2010) offer high potential for rapid monitoring of protected areas as well as indicators of the consequences of habitat loss or fragmentation on species assemblages in tropical forests (e.g. Brown 1997; Schulze et al. 2004; Hayes et al. 2009). Odonates (dragonflies and damselflies) in particular provide excellent model organisms for conservation ecology and evolutionary biology research due to their specific life history traits and habitat requirements (e.g. Corbet 1999; Cordero Rivera 2006; Córdoba-Aguilar 2008). Their presence or absence at a specific site is often directly linked to (micro) climate, geography, diversity of freshwater resources, and surrounding vegetation (Orr 2006; Hassall & Thompson 2008; Kalkman et al. 2008).

Among odonates, the most useful bioindicators for tropical forests are species with a wide distributional range but with a narrow ecological niche; the latter makes them sensitive to relatively small environmental changes. A forest specialist, which harbors this potential for such a model organism, is the Pseudostigmatid damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera). As the world´s largest extant odonate it is distributed throughout the Neotropics from Mexico to Bolivia (Davies & Tobin 1984). Across its geographic range its fundamental ecological niche is old growth forest with a closed canopy (Hedström & Sahlén 2001; 2003; Fincke & Hedström 2008). The literature describes only one species within the genus *Megaloprepus*: *M. caerulatus caerulatus*, with two potential subspecies from Mesoamerica and South America: *M. caerulatus brevistigma* and *M. caerulatus latipennis* (Steinmann 1997; Heckman 2008), whereat *M. caerulatus latipennis* is considered as a synonym for the nominal *M. caerulatus* (Garrison et al. 2010). The specific geographical distribution of the subspecies (i.e. *M. caerulatus brevistigma*) is in dispute (Steinmann 1997; Heckman 2008). Verification of its current taxonomic status, phylogeographic patterns, and contemporary distributional ranges with respect to forest history and destruction are needed.

Megaloprepus is one of only 19 described species within the family of giant damselflies, Pseudostigmatidae (Heckman 2008; Ingley et al. 2012); all are specialized to use water-

filled plant containers (i.e. phytotelmata) as larval habitats (Fincke 1998; reviewed by Fincke 2006). Within the Pseudostigmatidae only one genus (*Coryphagrion grandis*) inhabits eastern African coastal forests as primary ecological niche (Clausnitzer & Lindeboom 2002; Groeneveld et al. 2007). All other members of the family are registered to the Neotropics (Steinmann 1997; Fincke 2006; Heckman 2008). Whereas *Megaloprepus* seems to have remained a monospecific genus, inhabiting a highly conserved niche preserved over evolutionary time scales, *Mecistogaster* has radiated into 10 species (53% of the total species within the family Pseudostigmatidae), likely due to its ability to adapt to a variety phytotelmata and habitat conditions ranging from disturbed secondary to primary forest. On the other hand, species with strong niche conservatism may not be able to adapt sufficiently to new ecological conditions nor to rapid environmental changes caused by anthropogenic activity (Holt & Gomulkiewicz 2004; Wiens & Graham 2005). Rather, such changes lead to small, isolated populations, reduced genetic diversity with little potential for adaptation. Consequently, niche conservatism over large time scales should result in strict vicariant speciation, which occurs in geographic rather than ecological dimensions. To answer the question if independent evolutionary processes in populations of geographically isolated regions have led to diversification patterns and/or if the strong niche conservatism and recent forest fragmentation resulted in small, inbreed populations, patterns of genetic structure and genealogical relationships in and among four populations from Mexico to Central America were analyzed. Furthermore, a phylogenetic framework was constructed to gain deeper insights into the taxonomic status of the genus *Megaloprepus*. Results are discussed within the context of recent and past forest history.

Materials and Methods

Study areas and sampling methods

Tissue samples were collected at four localities in Central America and southern Mexico spanning *M. caerulatus´* northern geographical range during two breeding seasons in 2009/10 and 2011/12 (see Fig.1). Populations were in either primary (RBLT, CNP) or old growth secondary forests (BCI, LS). Anthropogenic impact and fragmentation rates as well as size and connectivity to other forest patches varied considerable between sites. On the Pacific slope, the Corcovado National Park (CNP, Área de Conservación Osa – ACOSA) including the sampling site around the Sirena Station, covered the largest forested area of about 41,800 ha surrounded by the ´Golfo Dulce Forest Reserve` (Carrillo et al. 2002). Selected sampling localities located on the Caribbean slope are smaller. The Biological Research Station La Selva, Costa Rica (LS, Organization for Tropical Studies), covered 1,536 ha with 55% primary forest, but is directly connected at its southern border to the ´Braulio Carrillo National Park` (McDade & Hartshorn 1994). Barro Colorado Island, Panama (BCI; Smithsonian Tropical Research Institute), with its 1,500 ha forest was formed between the Pacific and Atlantic slopes with the creation of Gatun Lake during the years 1910-1914. In the northeast, across the canal from BCI lies the National Park Soberania (12,000 ha) (Leigh 1999). On the Atlantic slope, within the Los Tuxtlas Biosphere Reserve,

Mexico (RBLT), the most northern locality, the collection site of Laguna Escondida, is contiguous with 640 ha of the Los Tuxtlas Tropical Biology Station and connected to the ´Sierra de San Martin` (Estrada 1982).
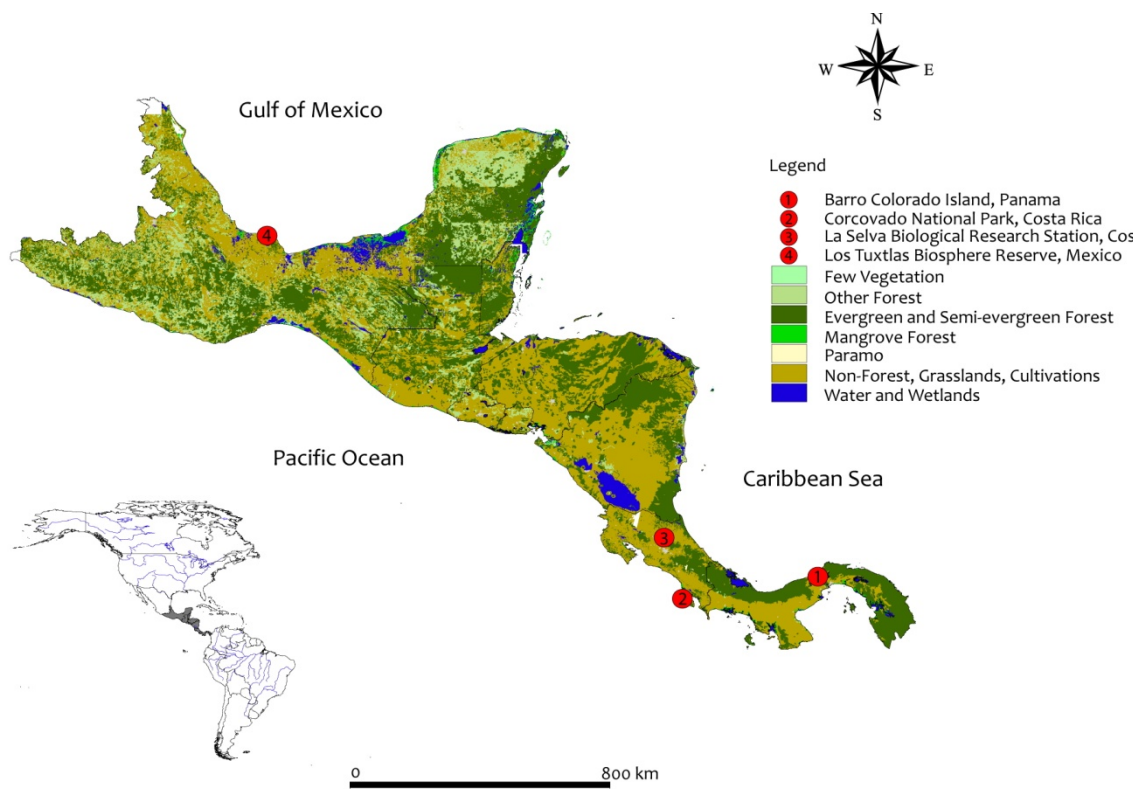
Fig.1 *Megaloprepus* sample sites in Central America and the South of Mexico. Map shows the frontiers of Mesoamerica, where all political borders displayed with black lines. Different sampling localities are visualized with red dots: Barro Colorado Island (BCI), Corcovado National Park (CNP), Biological Research Station La Selva (LS), and Los Tuxtlas Biosphere Reserve (RBLT). Furthermore, types of land-use are illustrated in different colors.
ArcView GIS 3.3 (ESRI 2002), source: CCAD (Comisión Centroamericana de Ambiente y Desarrollo), July 2010.

Within each locality, the size of monitored areas ranged from 64 ha at La Selva to 15 ha at Los Tuxtlas (CNP: 24 ha and BCI: 32 ha). During the field sessions the conditions of water filled tree holes were monitored from the ground to up to 2 m in height. The water volume was measured using a 3 L measuring cup; temperature, pH and conductivity were determined using a pH meter (PCE-228, PCE Inst.). Density was calculated as the number of tree holes over 0.1 L that were found within each sampling area. *Megaloprepus* larva occupancy was calculated as the percentage of the holes that contained at least one *Megaloprepus* larva and tree holes were classified as either large ($\geq$ 1 L of water) or small ($<$ 1 L).

Tissue samples were collected non-destructively (Fincke & Hadrys 2001) from adult individuals or larvae and stored in 98% ethanol. On BCI and La Selva material originate exclusively from the right middle tibia of adults. At Los Tuxtlas and Corcovado National

Park, most samples were from the caudal lamellae of larvae (terminal gills) (Table 1). To minimize the collection of siblings at the latter two sites, larvae were taken from 20 and 48 different tree holes, respectively. Within a given hole, multiple larvae sampled were always at different developmental stages (different instars). Tissue samples were collected from a total of 138 *Megaloprepus* individuals.

Table 1 Sampling properties for *Megaloprepus*. The number of tissue samples (*N*) are divided in sample types (larvae = *L*, adult individuals = *A*) per sampling locality. For each sampling position the geographical data and its used abbreviation in the text are presented.

Sampling locality, Abb.	Country	GPS Data Latitude	Longitude	N	Samples (N)
Barro Colorado Island, BCI	Panama	9°09′54.07′′N	79°50′12.27′′W	56	*A* = 56
Corcovado National Park, CNP	Costa Rica	8°32′32.42′′N	83°34′13.39′′W	31	*A* = 9, *L* = 22
Biological Research Station La Selva, LS	Costa Rica	10°28′58.65′′N	84°00′58.96′′W	32	*A* = 32
Los Tuxtlas Biosphere Reserve, RBLT	Mexico	13°35′06.35′′N	95°04′30.22′′W	19	*A* = 1, *L* = 18

DNA extraction, amplification, and sequencing

Isolation of total genomic DNA was conducted using a modified phenol-chloroform extraction protocol (Hadrys et al. 1992). Two mitochondrial genes, the 16S rRNA (16S) and the NADH-dehydrogenase subunit 1 including partial 16S rRNA and tRNA$_{Leu}$ (ND1) were used as molecular markers (e.g. Groeneveld et al. 2007; Damm et al. 2010a). Primer combinations for the 570 bp fragment of the conservative 16S rRNA fragment (P784 and P785) were described in Simon et al. (1994) and for the 610 bp long ND1 fragment in Abraham et al. (2001). Amplification of desired products were conducted as described in Damm et al. (2010a). All amplified products were purified by ethanol precipitation. Sequencing reactions were conducted bidirectionally using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). After purification with SephadexTM Gel Filtration (GE Healthcare), products were sequenced using the ABI PRISM 310 Genetic Analyzer following the manufacturer's protocol (Applied Biosystems). Three microsatellite loci originally developed for paternity analyses for the Barro Colorado population were checked for amplification products following Hadrys et al. (2005). Automated genotyping were carried out using 500 ROX™ Size Standard (Applied Biosystems, GeneScanTM) in ABI PRISM 310 Genetic Analyzer (Applied Biosystems); GeneScan Analysis Software (Applied Biosystems) was used to determine allele size relative to the standard.

Population genetic and phylogenetic analyses

Sequences were checked, assembled and edited manually using SeqMan II (vers. 5.03; DNAStar, Inc.). Multiple consensus sequences were aligned for each marker gene independently in Muscle vers. 3.6 (Edgar 2004) and edited in Quickalign (Müller & Müller 2003). Final alignments were analyzed using different programs for basic genetic statistics and general diversity information. The number of haplotypes (*H*), haplotype diversity (*h*), and nucleotide diversity (π) were calculated using DnaSP vers. 5.10 (Librado & Rozas 2009), whereas gaps were considered as the fifth state. To visualize genealogies at the population level, haplotype networks based on statistical parsimony algorithm were constructed using the default settings of TCS, vers. 1.13 (default parsimony connection limit of 95%) (Clement et al. 2000). A hierarchical analysis of molecular variance (AMOVA) (Excoffier et al. 1992) implemented in ARLEQUIN vers. 3.5 (Excoffier & Lischer 2010) was conducted to detect genetic sub-structuring within and among populations via a refined F_{ST} approach without any grouping of populations (1000 permutations). Finally, using the Kimura 2-parameter (K2P) model (Kimura 1980) pairwise sequence divergence between and within groups were calculated in MEGA 5 (Tamura et al. 2011).

Comparisons of population structure via microsatellites were conducted using ARLEQUIN (Excoffier et al. 1992). Allele frequencies were calculated in GENEPOP vers. 4.0.10 (Rousset 2008) and Bayesian multi-locus clustering was applied as implemented in STRUCTURE version 2.1 (Pritchard et al. 2000) placing all individuals into K populations. For correct estimation of K, the Δ K statistic was used (Evanno et al. 2005), runs with K values were repeated 20 times with a burn-in period of 10^5 steps followed by 10^5 Markov chain Monte Carlo replicates.

Different tree building algorithms were used to determinate the taxonomic status of *Megaloprepus*. The three most common haplotypes of each population were chosen and all existing sequences of *Coryphagrion*, *Pseudostigma*, *Mecistogaster*, *Anomisma*, and *Microstigma* species of the Pseudostigmatidae family were downloaded from GeneBank (Groeneveld et al. 2007; Ingley et al. 2012). The total data set included 24 ingroup taxa. *Teinobasis fortis* (Coenagrionidae, Odonata) was chosen as outgroup taxon (see Ingley et al. 2012). 14 species were included, and their corresponding sequence accession numbers itemized (see Table A1.1). Maximum Parsimony (MP) and Bayesian analyses (BA) were applied for each marker gene independently as well as combined. MP was performed using PAUP* vers. 4.0b8 (Swofford 2002). A full heuristic search was implemented under the 50% majority-rule, with 1000 bootstrap (BS) replicates as well as reconnection branch swapping option (TBR) (Felsenstein 1985). Bayesian analysis was performed in MrBayes vers. 3.1.2 (Ronquist & Huelsenbeck 2003). The most likely model of nucleotide substitution was searched separately for each locus using ModelTest vers. 3.7 (Posada & Crandall 1998) and the Akaike Information Criterion (AIC) (Akaike 1973). Finally, two independent runs were performed under the best fit-model (GTR+I+G) for 20×10^6 generations and each four Markov chains. Trees were sampled every 1,000 generations and the first 20,000 trees were discarded as `burn-in´ after reaching stationary. Remaining trees were used to calculate posterior probabilities as well as consensus topology.

Results

Site differences in tree hole characteristics and occupancy

The highest density of tree holes per ha was found in Corcovado National Park followed by Los Tuxtlas Biosphere Reserve (Table 2). The percentage of large tree holes (≥ 1 L) was highest in Corcovado and BCI (between 43% and 45%) and low in Los Tuxtlas (20%). At the latter site, 71% of tree holes were occupied with *Megaloprepus* larvae whereas in La Selva larvae were only found in 43% of water filled tree holes (Table 2). There were significant differences in pH ($p ≤ 0.01$), water temperature ($p ≤ 0.01$), and conductivity ($p ≤ 0.01$) of holes among the sampling sites, but no difference in water volume ($p = 0.11$) (one-way ANOVA).

Table 2 Water filled tree hole data. Shown are the numbers of tree holes per sampling locality (*N*), the percent of tree holes containing more than 1 L of water, its occupancy with *Megaloprepus* larvae, and the number of water filled tree holes per ha. In addition, the mean water temperature ($\bar{x}\ T°C$), the mean pH ($\bar{x}\ pH$), and the mean conductivity ($\bar{x}\ mS$). Sampling localities are: Barro Colorado Island (BCI), Corcovado National Park (CNP), Biological Research Station La Selva (LS), and Los Tuxtlas Biosphere Reserve (RBLT).

	N	tree holes ≥1L	occupancy	tree holes/ha	$\bar{x}\ T°C$ (+/-SD)	$\bar{x}\ pH$ (+/-SD)	$\bar{x}\ mS$ (+/-SD)	$\bar{x}\ ml$ (+/-SD)
BCI	11	45%	55%	0.34	26.30 (+/- 1.11)	7.07 (+/- 0.25)	0.44 (+/- 0.16)	2199.09 (+/- 2582.6)
CNP	48	43%	63%	2	25.73 (+/- 1.09)	6.08 (+/- 0.82)	1.98 (+/- 3.78)	1459.48 (+/- 1708.3)
LS	14	29%	43%	0.22	24.55 (+/- 1.45)	4.56 (+/- 0.46)	14.98 (+/- 26.98)	739.28 (+/-537.72)
RBLT	20	20%	71%	1.33	23.54 (+/- 0.59)	6.14 (+/- 0.36)	32.42 (+/- 26.88)	671.42 (+/-626.4)

Population genetic and phylogenetic analyses

The final ND1 and 16S rRNA alignments consisted of 109 and 111 sequences, respectively. For ND1, there was an alignment length of 572 bp including 68 parsimony-informative characters. For the 16S fragment the alignment was 488 bp long including 30 parsimony-informative sites.

Genetic diversity estimates indicated high variation among sampling sites. For the ND1 gene fragment the Corcovado population showed the highest haplotype diversity ($h = 0.84$) as well as the highest nucleotide diversity ($π = 0.29$). In contrast, the BCI population had the lowest number of haplotypes and nucleotide diversity ($h = 0.43$, $π = 0.09$). For 16S rRNA fragment high diversities were found on BCI ($h = 0.77$, $π = 0.26$), followed by Corcovado National Park ($h = 0.44$, $π = 0.26$) and La Selva ($h = 0.17$, $π = 0.04$). Lowest diversity ($h = 0.00$, $π = 0.00$) was received in the Los Tuxtlas populations (Table 3).

TCS statistical parsimony networks showing genealogical relationships between closely related haplotypes revealed strong sub-structuring (Fig. 2). Among the four populations, only those of BCI and La Selva were still connected, however with five mutational steps between the most common haplotypes, respectively. Los Tuxtlas and Corcovado split into separate networks indicating strong genetic isolation. Furthermore, all populations contained population specific haplotypes and no shared haplotypes were detected for either maker gene. For ND1, 30 haplotypes were found. Most haplotypes were in Corcovado (H = 10) and La Selva (H = 9) populations; fewer in Los Tuxtlas (H = 6) and BCI (H = 5). Only 13 haplotypes were defined for 16S rRNA (Table 3).

Table 3 Main characteristics of mitochondrial sequence marker ND1 and 16S rRNA for the four *Megaloprepus* populations: Barro Colorado Island (BCI), Corcovado National Park (CNP), Biological Research Station La Selva (LS), and Los Tuxtlas Biosphere Reserve (RBLT). Presented are number of individuals (N), number of (private) haplotypes (H), haplotype diversity (h), and nucleotide diversity (π) in percent per population.

| | **ND1** | | | | **16S** | | | |
	N	H	h (+/- SD)	$\pi\%$ (+/- SD)	N	H	h (+/- SD)	$\pi\%$ (+/- SD)
BCI	31	5	0.43 (+/- 0.10)	0.09 (+/- 0.03)	32	7	0.77 (+/- 0.07)	0.26 (+/- 0.19)
CNP	29	10	0.84 (+/- 0.05)	0.29 (+/- 0.06)	31	3	0.44 (+/- 0.08)	0.10 (+/- 0.01)
LS	30	9	0.76 (+/- 0.06)	0.28 (+/- 0.06)	30	2	0.19 (+/- 0.09)	0.04 (+/- 0.06)
RBLT	19	6	0.75 (+/- 0.09)	0.17 (+/- 0.03)	18	1	0.00 (+/- 0.00)	0.00 (+/- 0.00)

Pairwise nucleotide sequence divergences, measured using the K2P model, showed similar patterns for both marker genes (Table 4). Among populations, genetic distances for ND1 between BCI and La Selva were low (1.27%). In contrast, high distances were found between BCI (7.02%), La Selva (7.13%), as well as Los Tuxtlas (6.79%), and Corcovado. The highest genetic divergences in ND1 were between La Selva and Los Tuxtlas (ND1: 7.52%). In 16S rRNA distances among Los Tuxtlas and all other populations were high (BCI: 4.87%, La Selva: 4.74% and Corcovado: 4.50%) as well as between Corcovado and the other populations. Comparisons of F-statistics among all sampling localities showed significantly high ($p < 0.01$) genetic differentiation, indicating a lack of gene flow among populations. The lowest F_{ST} value found was between La Selva and BCI in the ND1 marker gene (F_{ST} = 0.85). All other values are higher; with F_{ST} = 0.99 between the Corcovado and BCI.
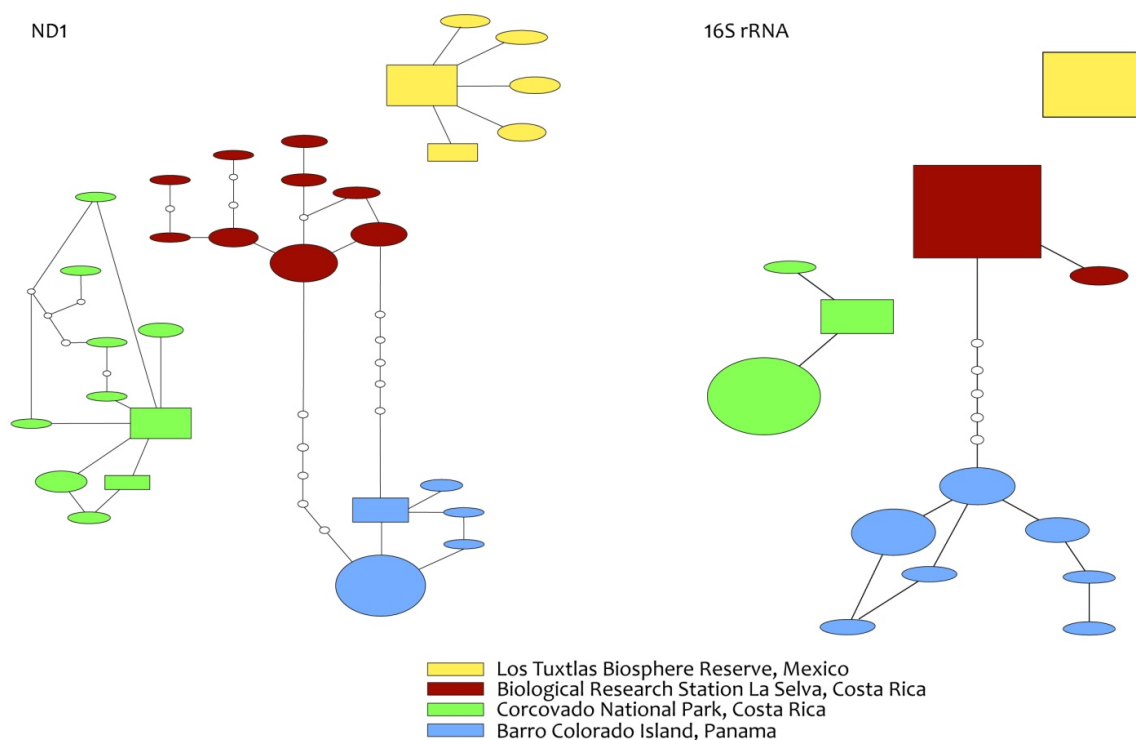
Fig. 2 Statistical parsimony haplotype networks based on two mitochondrial marker genes ND1 and 16S rRNA for all sampling localities. Each sampling point is illustrated in both networks with the same color. Haplotype diversity can be seen in the number of circles and rectangles. Strong phylogeographic structuring was obtained between populations. Barro Colorado Island and La Selva are still connected by mutational steps (blank circles). The considered ancestral haplotypes are depicted as rectangles, all other haplotypes as circles, whereas the sizes of rectangles and circles correlate with haplotype frequency within each network.

Additional microsatellite typing revealed 40 different alleles for three of the tested microsatellite loci (MeAB 3/11: 7 alleles, MeAB 12/15: 9 alleles, MeAB 5/19: 24 alleles) with 11 shared alleles for 56 individuals from BCI and 32 from La Selva. No products could be amplified for any locus of all individuals from Corcovado National Park ($N = 31$) as well as Los Tuxtlas ($N = 19$). Gene diversities obtained in BCI were higher (0.82) than in La Selva (0.65). Both populations showed significant ($p < 0.01$) deviation from the Hardy-Weinberg equilibrium where observed heterozygosity lied between 0.56 and 0.75. F_{ST} values showed high differentiation (11.89%) confirming the strong clusters obtained in STRUCTURE: La Selva and BCI (data not shown).

Phylogenetic relationships

The multiple sequence alignment contained in total 902 bp (16S rRNA: 432 bp and ND1: 470 bp) for 25 ingroup taxa and one outgroup taxon (Table A1.1). In total 398 variable sites including 273 parsimony-informative sites (16S: 123; ND1: 150) were found. Both tree reconstruction methods, MP and BA, exhibited identical topologies and well-supported nodes for both markers. Fig.3 shows the BA phylogram of the combined data set resolving three clades within the Pseudostigmatidae (see also Ingley et al. 2012). The three clades are composed of the genera: (i) *Coryphagion*, (ii) *Pseudostigma + Mecistogaster*, and (iii) *Anomisma*, *Microstigma + Megaloprepus*, with *Microstigma* as sister taxon to *Megaloprepus*. For *Megaloprepus* three distinct clusters with high support values (> 50% BS and > 0.73 PP) were obtained (Fig. 3) consistent with the haplotype networks and the results of the distance measurements.

Table 4 Genetic diversity estimates between populations for ND1 and 16S rRNA. Pairwise nucleotide sequence divergence in % between and within groups using Kimura's 2-parameter model (K2P) for all *Megaloprepus* populations (above) and population pairwise F_{ST} values (below) are shown. Significances are marked with an asterisk ($p \leq 0.001$). Population abbreviations: Barro Colorado Island (BCI), Corcovado National Park (CNP), Biological Research Station La Selva (LS), and Los Tuxtlas Biosphere Reserve (RBLT).

	ND1				16S rRNA			
	BCI	*CNP*	*LS*	*RBLT*	*BCI*	*CNP*	*LS*	*RBLT*
BCI	0.10				0.20			
CNP	7.02	0.30			3.98	0.01		
LS	1.27	7.13	0.30		1.43	4.08	0.04	
RBLT	7.50	6.79	7.52	0.20	4.87	4.50	4.74	0.00
BCI	-				-			
CNP	0.97*	-			0.99*	-		
LS	0.85*	0.96*	-		0.99*	0.98*	-	
RBLT	0.98*	0.96*	0.97*	-	0.97*	0.89*	0.96*	-

MP/BA
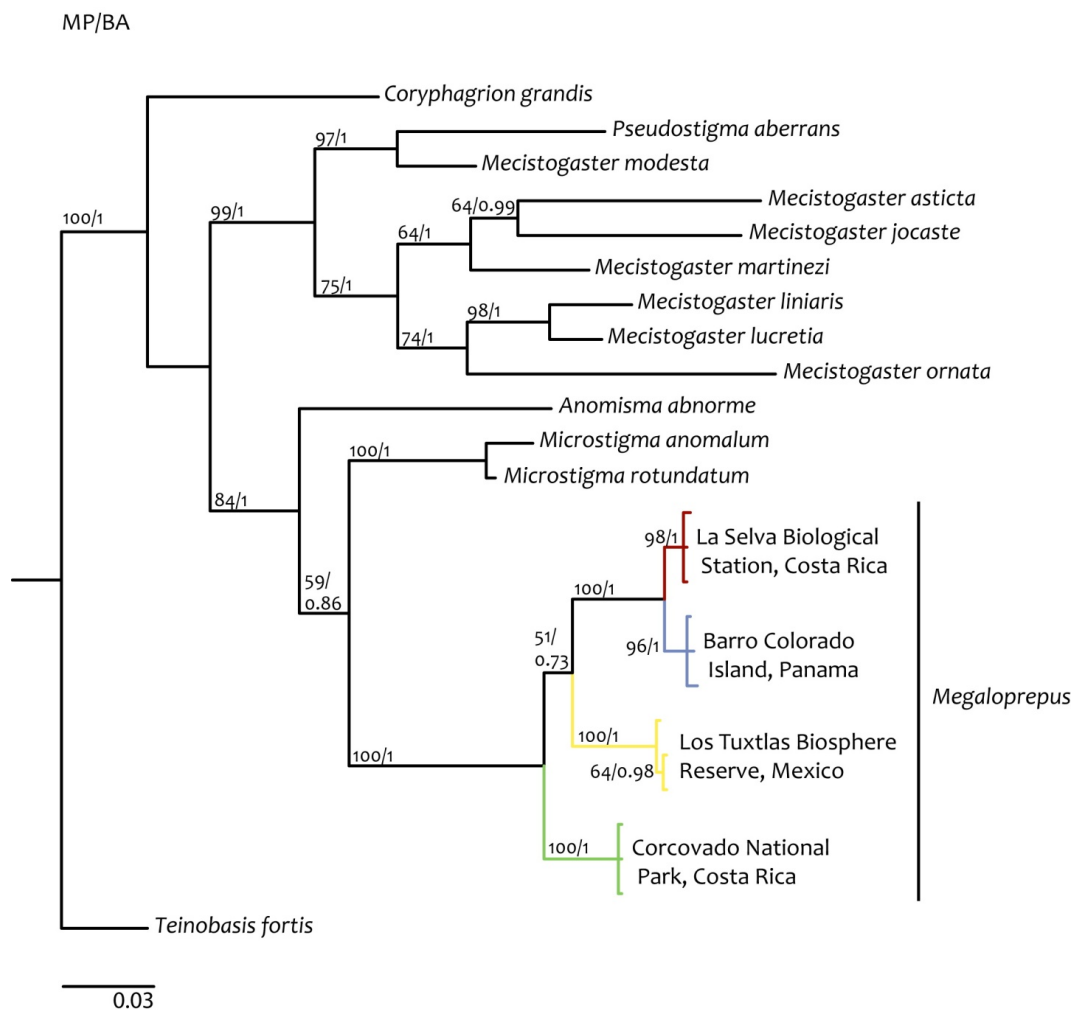
Fig. 3 Phylogenetic classification of the *Megaloprepus* "cluster". Bayesian consensus phylogram (16S rRNA and ND1) including 12 species of the family of the Pseudostigmatidae and the three most frequent haplotypes of each population of *Megaloprepus* data set. Additionally, *Teinobasis fortis* was used as outgroup to root the tree. Posterior probability values (PP) calculated on the basis of 20,002 ´post-burn-in` trees and maximum parsimony bootstrap supports (BS) above 50% (1,000 replicates) are presented for each node. Coloration for the different populations of *Megaloprepus* comport with the coloration in the haplotype network (see Fig. 2 above).

Discussion

In this study we evaluated, for the first time, the species status based on genetic data of a forest restricted odonate over its distributional range in Mesoamerican old growth forests. We found high evidence that, the narrow niche specialization of *Megaloprepus* has resulted in considerable population substructuring over a wide geographical range.

Genealogical relationships

Analysis of genealogical relationships among the study populations indicated high sub-structuring and complete genetic isolation among most populations. Statistical parsimony

networks separated three distinct clusters: (i) Los Tuxtlas Biosphere Reserve, (ii) Corcovado National Park and (iii) La Selva + Barro Colorado Island (still genetically connected). The F_{ST} values indicated nearly no gene flow and large genetic differentiation between all populations studied. The genetic distances among populations were high between the southern (BCI, LS) and the northern population (RBLT) as well as between the Corcovado National Park and all other populations. In contrast, lower distances were found between La Selva and Barro Colorado Island. Moreover, the phylogenetic tree topology supported three clades within *Megaloprepus*.

Further findings underline the discovered diversification pattern here. First, the genetic distances between the *Megaloprepus* populations were similar to the distance values found between described real sister species within the same family. In the conserved 16S rRNA fragment, we found high genetic distances among the suggested *Megaloprepus* clades of at least 3.98% (Table 4). Similarly, within the genus *Mecistogaster* the divergence values ranged from 3.08% to 4.03% between *M. jocaste*, *M. asticta*, *M. martinezi* (Table A1.2). Secondly, support is given by additional tests of the microsatellite system (Fincke & Hadrys 2001; Hadrys et al. 2005). The originally established species-specific microsatellite system for larvae and adults provided high resolution results in paternity analyses for the population on Barro Colorado Island (Fincke & Hadrys 2001). In the presented study microsatellite typing revealed 40 different alleles with 11 shared alleles for 56 individuals from BCI and 32 from La Selva and a high differentiation between the BCI and La Selva populations (F_{ST} value: 11.89%) confirming the structure detected by 16S rRNA and ND1 sequence markers. However, the repetitive failure to amplify products for any locus of all individuals (adults and larvae) from Corcovado National Park ($N = 31$) as well as Los Tuxtlas ($N = 19$) may further support the mitochondrial sequence marker applied. Microsatellites are well established molecular marker systems for a variety of odonate species and all are known to be highly species specific and work best at lowest taxonomic levels (e.g. for paternity and fine scale population genetic studies) (e.g. Hadrys et al. 2005; Giere & Hadrys 2006; Carballa et al. 2007; Hadrys et al. 2007a; Hadrys et al. 2007b; Damm & Hadrys 2012). In genome comparisons of wasps for example only 17 - 23% of possible microsatellites were shared by three sister species (Pannebakker et al. 2010). Furthermore, microsatellite abundance and distribution are variable between insect species (Pannebakker et al. 2010 and references therein). This indicates that a great percentage of randomly generated microsatellites may not appropriate for interspecific cross amplifications even across closely related sister species, mainly due to mutations in the flanking regions (Rutkowski et al. 2011). Consequently, the failure to detect amplification products for two geographic sites in *Megaloprepus* is consistent with ongoing large-scale speciation processes and the genetic distances detected by more conserved gene sequences. Nevertheless, the inability of microsatellite loci amplification may have several causes. While technical issues could be excluded, mutations in the flanking regions remain to be solved.

Finally, populations differed in phenotypic characters, e.g. wing coloration and size as discovered earlier (Fincke 2006; Schultz & Fincke 2009). Individuals from Barro Colorado Island are smaller than individuals from the Los Tuxtlas Biosphere Reserve (Fincke 1998;

2006) and sexually dimorphic wings are only present in populations from La Selva and Barro Colorado Island. In contrast, in Corcovado National Park and Los Tuxtlas populations, both females and males feature similar wing coloration and lack the characteristic matte-white band on all four wings in males, a taxon specific character. Consequently, we assume that *Megaloprepus* is in the process of speciating or has recently undergone speciation. Within its northern distributional range, the genus should not be considered a single species genus anymore. An integrative study of morphology and population ecology across different clusters could provide additional insights into the phenotypic differences underlying the detected genotypic pattern.

Speciation in the Neotropics

According to Groeneveld et al. (2007) the Pseudostigmatidae family is an old Gondwana relict. Consequently, it can be assumed that the historical distribution of the New World species (genera: *Pseudostigma*, *Mecistogaster*, *Anomisma*, *Microstigma*, and *Megaloprepus*) might have been in northern portion of South America. Due to the closing of the land bridge between south Nicaragua and Colombia (Bolivar Trench) in late Pliocene (about 3 million years ago) (Rich & Rich 1983) these genera may have dispersed northward (Kalkman et al. 2008). *Megaloprepus* probably inhabited the Caribbean as well as the Pacific Coast simultaneously. Support for this hypothesis is found in the close relatedness of *Megaloprepus* to *Microstigma* and *Anomisma*, genera which are distributed exclusively throughout South America. However, the narrowness of *Megaloprepus*` fundamental ecological niche has important implications on the distribution and connectivity between populations. When large primary habitats become interrupted by geographical events, gene flow between populations is inhibited and populations become geographically isolated. In the face of niche conservatism this separation can either led to the extinction of small populations or promotes genetic diversification leading to allopatric speciation (Kozak & Wiens 2010; Wiens et al. 2010). Beginning in the northwestern lowlands of Costa Rica up to southwest Mexico, dry forests with a prolonged dry season are common in the Pacific lowlands (Murphy & Lugo 1995). Subsequently, the moist forest in the southern Pacific coast of Costa Rica may have prevented *Megaloprepus* from dispersing northward, due to the species´ sensitivity to dry conditions. The extra tropical mountain ranges in the Center of Costa Rica (Talamancan Cordillera) could have constituted an additional ecological barrier between the Caribbean and Pacific regions (Coen 1983; Barrantes 2009). Historical geographic barriers combined with recent landscape structure might have led to the isolated range of *Megaloprepus* in the Corcovado National Park. In contrast, the split among populations from the north and south Caribbean Coast region seems to be more recent. The geographical distance from Los Tuxtlas in southern Mexico to La Selva and Barro Colorado Island, points to isolation by distance where past climatic conditions and restriction to small forest patches may have led to separation of populations. Studies of the genetic structure and ecology of other populations in Honduras, Guatemala or Belize could highlight this process.

Diversity within Populations

As a species sensitive to drying conditions and low dispersal ability over non-forest areas (Fincke 1994; 2006), *Megaloprepus* is susceptible to habitat disturbances (e.g. Fincke & Hedström 2008). Within the Los Tuxtlas Biosphere Reserve, high fragmentation rates have left mostly small remaining forests patches, surrounded by pastures and urban areas (Dirzo & García 1992; Urbina-Cardona et al. 2006; Solórzano García et al. 2012) leading to a rapid *M. caerulatus* population decline during the last decades (Fincke 2006; E. González Soriano personal communication). This concurs with the finding that this population had an overall low genetic diversity (Table 4). Although the number of tree holes found and sampled per ha was higher at Los Tuxtlas than in BCI and La Selva, but only 20% of these water filled tree holes exceeded one liter in volume, that which typically permits more than one larvae to survive to emergence at the same time (Fincke 1994). In contrast, the Corcovado population exhibited high genetic diversity for both maker genes, and similarly tree hole density appeared to be highest at that site. Within sampling localities water filled tree holes up to 2 m in height were monitored. Generally, *Megaloprepus* larvae have been found in tree holes as high as 7 m (Yanoviak 1999). Nevertheless, we assume that our samples of those under two meters is representative of those also found higher up in trees.

Population size of *Megaloprepus* depends in part on the number of tree holes inside a forest patch. The greatest recruitment of adults comes from large tree holes, which support multiple emerging adults during a given reproductive season (Fincke 1992; 1994; Fincke & Hadrys 2001). Additionally, intra-guild predation among co-occurring tree hole predators' limits *Megaloprepus* populations (Fincke 1992; 1994; 1998). Nevertheless, population size should be positively correlated with the number and size of tree species that harbor tree holes; larger trees of a given tree hole species typically have more water filled larval habitats (Fincke 2006). Hence, increasing forest fragmentation will result in greater isolation of sub-populations, along with concomitant changes in (micro) climate.

Implications for conservation

In tropical forest regions odonates are essential forest animals representing important environmental indicators (Paulson 2006). In the face of increasing patchiness of forests, and without the possibility for a range shift, forest odonates must adapt or go extinct (e.g. Gienapp et al. 2008). In Mesoamerica, roughly 80% of the original forest cover has already been lost (Harvey et al. 2008). Currently, the conservation status within this region includes 669 forests that represent 10.7% of the land area (Miller et al. 2001; DeClerck et al. 2010). Even protected areas face continuing threats as their borders become deforested, causing micro-climatic changes (Laurance et al. 2012). Important buffer zones as well as biological corridors that could provide long-term protection to forest biodiversity are insufficient in this region (DeClerck et al. 2010; Laurance et al. 2012). The Los Tuxtlas Tropical Biology Station is especially affected by human activities. As the most northern tropical moist forests with high rates of endemism, the lack of connectivity among forest patches, coupled with edge effects is threatening its endemic species (Mendoza et al. 2005). The rate of on-going fragmentation in this region is high, resulting in small forest patches (≤ 166.7 ha in mean)

(Solórzano García et al. 2012) and remaining patches are primarily reduced to mountain formations (i.e. precipices) unsuitable for pastures or agriculture. The *Megaloprepus* population found in this area is a possible site-endemic species and should receive priorities in conservation, especially given its continuing population decline. In contrast, in the relatively large Corcovado National Park, *Megaloprepus* exhibits greater genetic diversity. On Barro Colorado Island, which is protected from most human influences, the population seems to have remained stable over three decades. Here microsatellite monitoring over a 25-year period showed no significant changes in allelic diversities or allele frequencies (Feindt & Hadrys, unpublished data). Studies of European Odonates have shown a direct correlation of landscape structure and population isolation, as well as lack of genetic diversities in small populations and genetic isolation (Watts et al. 2004; Hadrys et al. 2006; Watts et al. 2007). In addition to habitat loss, many invertebrates are highly susceptible to climate change (Deutsch et al. 2008). In tropical regions the change of (micro) climates within forests, modifications of rain seasons, and increasing drying present serious challenges to their survival (e.g. Deutsch et al. 2008).

Effective conservation strategies or management decisions need a solid background regarding the status quo of a given species (DeSalle & Amato 2004; Hadrys et al. 2006; Lewis & Basset 2007; Pertoldi et al. 2007; Damm et al. 2010b). The discovery of 'new' evolutionary significant or conservation units in tropical forest sites may not only lead to reevaluation of the taxonomic status of the species but also of the conservation status of the forest itself. Furthermore, the data presented here revealed unexpected genealogical relationships and a potential speciation process within the highly specialized genus *Megaloprepus*. Recent forest status in Mesoamerica allows the assumption of numerous site-endemic species, calling for re-evaluations of some conservation priorities.

References

Abraham D, Ryrholm N, Wittzell H, Holloway JD, Scoble MJ, Löfstedt C (2001) Molecular Phylogeny of the Subfamilies in Geometridae (Geometroidea: Lepidoptera). *Molecular Phylogenetics and Evolution*, **20**, 65-77.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov and F. Csaki, eds., 2nd international symposium on information theory (Akademiai Kiado, Budapest).

Balint M, Domisch S, Engelhardt CHM, Haase P, Lehrian S, Sauer J, Theissinger K, Pauls SU, Nowak C (2011) Cryptic biodiversity loss linked to global climate change. *Nature Clim. Change*, **1**, 313-318.

Barrantes G (2009) The role of historical and local factors in determining species composition of the highland avifauna of Costa Rica and Western Panamá. *Revista de Biologia Tropical* **57**, 333-349.

Bisby F, Roskov Y, Culham A, Orrell T, Nicolson D, Paglinawan L, Bailly N, Kirk P, Bourgoin T, Baillargeon G, Hernandez F, De Wever A, Kunze T (2013) Species 2000 & ITIS Catalogue of Life. In: *Species 2000*, p. Digital resource at www.catalogueoflife.org/col/, Reading, UK.

Brodie J, Post E, Laurance WF (2011) Climate change and tropical biodiversity: a new focus. *Trends in ecology & evolution (Personal edition)*, **27**, 145-150.

Brown KS (1997) Diversity, disturbance, and sustainable use of Neotropical forests: insects as indicators for conservation monitoring. *Journal of Insect Conservation*, **1**, 25-42.

Calderón R, Boucher T, Bryer M, Sotomayor L, Kappelle M (2004) *Setting biodiversity conservation priorities in Central America: Action site selection for the development of a first portfolio*. The Nature Conservancy, Sna José, Costa Rica.

Carballa OL, Giere S, Cordero A, Hadrys H (2007) Isolation and characterization of microsatellite loci to study parthenogenesis in the citrine forktail, *Ischnura hastata* (Odonata: Coenagrionidae). *Molecular Ecology Notes*, **7**, 839-841.

Carrillo E, Saenz JC, Fuller TK (2002) Movements and activities of white-lipped peccaries in Corcovado National Park, Costa Rica. *Biological Conservation*, **108**, 317-324.

Clausnitzer V, Lindeboom M (2002) Natural history and description of the dendrolimnetic larva of *Coryphagrion grandis* (Odonata). *International Journal of Odonatology*, **5**, 29-44.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9 (10)**, 1657-1660.

Coen E (1983) Climate. In: *Costa Rican Natural History* (ed. Janzen DH), pp. 35-46. The University of Chicago Press.

Corbet PS (1999) *Dragonflies: Behaviour and Ecology of Odonata*. Harley Books, Essex, England.

Cordero Rivera A (2006) *Forests and Dragonflies. Fourth WDA International Symposium of Odonatology*. Pensoft Publishers, Sofia.

Córdoba-Aguilar A (2008) *Dragonflies and damselflies. Model organism for ecological and evolutionary research*. Oxford University Press, Oxford, UK.

Damm S, Dijkstra K-DB, Hadrys H (2010a) Red drifters and dark residents: The phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa´s changing environment (Odonata, Libellulidae, *Trithemis*). *Molecular Phylogenetics and Evolution*, **54**, 870-882.

Damm S, Hadrys H (2012) A dragonfly in the desert: genetic pathways of the widespread *Trithemis arteriosa* (Odonata: Libellulidae) suggest male-biased dispersal. *Organisms Diversity & Evolution*, **12**, 267-279.

Damm S, Schierwater B, Hadrys H (2010b) An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Molecular Ecology*, **19**, 3881-3893.

Davies DAL, Tobin P (1984) A synopsis of the dragonflies of the world: a systematic list of the extant species of Odonata. Volume 1, Zygoptera, Anisozygoptera. In: *Societas Internationalis Odonatologia Rapid Communications (Supplements)*. Utrecht.

DeClerck FAJ, Chazdon R, Holl KD, Milder JC, Finegan B, Martinez-Salinas A, Imbach P, Canet L, Ramos Z (2010) Biodiversity conservation in human-modified landscapes of Mesoamerica: Past, present and future. *Biological Conservation*, **143**, 2301-2313.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**, 702-712.

Deutsch CA, Tewksbury JJ, Huey RB, Sheldon KS, Ghalambor CK, Haak DC, Martin PR (2008) Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*, **105**, 6668-6672.

Dirzo R, García MC (1992) Rates of deforestation in Los Tuxtlas a neotropical area in south east Mexico. *Conservation Biology*, **6**, 84–90.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

ESRI (2002) ArcView GIS 3.3. (ed. Environmental Systems Research Institute I, Redlands), California, USA.

Estrada A (1982) Survey and Census of Howler Monkeys (*Alouatta palliata*) in the Rain Forest of "Los Tuxtlas", Veracruz, Mexico *American Journal of Primatology*, **2**, 363-372.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**, 783-791.

Fincke OM (1992) Interspecific competition for tree holes: consequences for mating systems and coexistence in neotropical damselflies. *American Naturalist*, **139**, 80-101.

Fincke OM (1994) Population regulation of a tropical damselfly in the larval stage by food limitation, cannibalism, intraguild predation and habitat drying. *Oecologia*, **100**, 118-127.

Fincke OM (1998) The population ecology of *Megaloprepus caerulatus* and its effect on species assemblages in water-filled tree holes. Dordrecht, Kluwer Academic Publ., NL.

Fincke OM (2006) Use of forest and tree species, and dispersal by giant damselflies (Pseudostigmatidae): their prospects in fragmented forests. In: *Forest and dragonflies. 4th WDA International Symposium of Odonatology*. (ed. Rivera AC), pp. 103-125. Pensoft Publishers, Sofia.

Fincke OM, Hadrys H (2001) Unpredictable offspring survivorship in the Damselfly, *Megaloprepus coerulatus*, shapes parental behavior, constrains sexual selection, and challenges traditional fitness estimates. *Evolution*, **55**, 762-772.

Fincke OM, Hedström I (2008) Differences in forest use and colonization by Neotropical tree-hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conversion. *Studies on Neotropical Fauna and Environment*, **43**, 35-45.

Garrison RW, von Ellenrieder N, Louton JA (2010) *Damselfly Genera of the New World: An Illustrated and Annotated Key to the Zygoptera* The Johns Hopkins University Press.

Gienapp P, Teplitsky C, Alho JS, Mills JA, MerilÄ J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167-178.

Giere S, Hadrys H (2006) Polymorphic microsatellite loci to study population dynamics in a dragonfly, the libellulid *Trithemis arteriosa* (Burmeister, 1839). *Molecular Ecology Notes*, **6**, 933-935.

Groeneveld LF, Clausnitzer V, Hadrys H (2007) Gigantism in damselflies of Africa and South America: convergent evolution or homologous structures? Evidence from nuclear and mitochondrial sequence data. *Molecular Phylogenetics and Evolution*, **42**, 339-346.

Gullan PJ, Cranston PS (2010) *The insects: an outline of entomology. 4th Edition*. Wiley-Blackwell.

Hadrys H, Balick M, Schierwater B (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology*, **1**, 55-63.

Hadrys H, Clausnitzer V, Groeneveld LF (2006) The present role and future promise of conservation genetics for forest Odonates. In: *Forest and dragonflies. 4th WDA International Symposium of Odonatology.* (ed. Rivera AC), pp. 279-299. Pensoft Publishers, Sofia.

Hadrys H, Schroth W, Schierwater B, Streit B, Fincke O (2005) Tree hole odonates as environmental monitors: Non-invasive isolation of polymorphic microsatellites from the neotropical damselfly *Megaloprepus caerulatus*. *Conservation Genetics*, **6**, 481-483.

Hadrys H, Timm J, Streit B, Giere S (2007a) A panel of microsatellite markers to study sperm precedence patterns in the emperor dragonfly *Anax imperator* (Odonata: Anisoptera). *Molecular Ecology Notes*, **7**, 296-298.

Hadrys H, Wargel A, Giere S, Kraus B, Streit B (2007b) A panel of microsatellite markers to detect and monitor demographic bottlenecks in the riverine dragonfly *Orthetrum coerulescens* F. *Molecular Ecology Notes*, **7**, 287-289.

Harvey CA, Komar O, Chazdon R, Ferguson BG, Finegan B, Griffith DM, MartÍNez-Ramos M, Morales H, Nigh R, Soto-Pinto L, Van Breugel M, Wishnie M (2008) Integrating Agricultural Landscapes with Biodiversity Conservation in the Mesoamerican Hotspot. *Conservation Biology*, **22**, 8-15.

Hassall C, Thompson DJ (2008) The impact of environmental warming on Odonata - a review. *International Journal of Odonatology*, **11**, 131-153.

Hayes L, Mann DJ, Monastyrskii AL, Lewis OT (2009) Rapid assessments of tropical dung beetle and butterfly assemblages: contrasting trends along a forest disturbance gradient. *Insect Conservation and Diversity*, **2**, 194-203.

Heckman CW (2008) *Encyclopedia of South American Aquatic Insects: Odonata - Zygoptera*. Springer Netherlands.

Hedström I, Sahlén G (2001) A key to the adult Costa Rican helicopter damselflies Odonata: Pseudostigmatidae with notes on their phenology and life zone preferences. *Revista de Biología Tropical*, **49**, 1037-1056.

Hedström I, Sahlén G (2003) An extended description of the larva of *Megaloprepus caerulatus* from Costa Rica (Odonata: Pseudostigmatidae). *International Journal of Odonatology*, **6**, 23-31.

Holt RD, Gomulkiewicz R (2004) Conservation implications of niche conservatism and evolution in heterogeneous environments. In: *Evolutionary Conservation Biology* (eds. Ferriere R, Dieckmann U, Couvet D), pp. 244-264. Cambridge University Press, Cambridge, UK.

Ingley SJ, Bybee SM, Tennessen KJ, Whiting MF, Branham MA (2012) Life on the fly: phylogenetics and evolution of the helicopter damselflies (Odonata, Pseudostigmatidae). *Zoologica Scripta*, 00-00.

Kalkman VJ, Clausnitzer V, Dijkstra K-D, Orr A, Paulson D, Tol J (2008) Global diversity of dragonflies (Odonata) in freshwater. *In: Freshwater Animal Diversity Assessment. Springer Netherlands*, **595**, 351-363.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**.

Kozak KH, Wiens JJ (2010) Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters*, **13**, 1378-1389.

Laurance WF, Carolina Useche D, Rendeiro J, Kalka M, Bradshaw CJA, Sloan SP, Laurance SG, Campbell M, Abernethy K, Alvarez P, Arroyo-Rodriguez V, Ashton P, Benitez-Malvido J, Blom A, Bobo KS, Cannon CH, Cao M, Carroll R, Chapman C, Coates R, et a (2012) Averting biodiversity collapse in tropical forest protected areas. *Nature*, **489**, 290-294.

Leigh EG (1999) *Tropical Forest Ecology: A View from Barro Colorado Island*. Oxford University Press, New York.

Lewis OT, Basset Y (2007) Insect Conservation in Tropical Forests. In: *Insect Conservation Biology* (eds. Stewart A, New T, Lewis O), pp. 34-56. The Royal Entomological Society.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Magurran AE, Dornelas M (2010) Biological diversity in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3593-3597.

Mayhew PJ, Jenkins GB, Benton TG (2008) A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 47-53.

McDade LA, Hartshorn GS (1994) La Selva Biological Station. In: *La Selva: Ecology and Natural History of a Neotropical Rain Forest* (eds. McDade LA, Bawa KS, Hespenheide HA, Hartshorn GS), pp. 6-18. University of Chicago Press.

Mendoza E, Fay J, Dirzo R (2005) A quantitative analysis of forest fragmentation in Los Tuxtlas, southeast Mexico: patterns and implications for conservation. *Revista Chilena de Historia Natural*, **78**, 451-467.

Miller K, Chang E, Johnson N (2001) *Defining common ground for the Mesoamerican Biological Corridor*. World Resources Institute, Washington, D.C.

Müller J, Müller K (2003) QuickAlign: A New Alignment Editor. *Plant Molecular Biology Reporter*, **21**.

Murphy PG, Lugo AE (1995) Dry forests of Central America and the Caribbean. In: *Seasonally dry tropical forests* (eds. Bullock S, Mooney H, Medina E), pp. 9-34. Cambridge University Press, Cambridge, U.K.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853 - 858.

Orr AG (2006) Odonata in Bornean tropical rain forest formations: diversity, endemicity and implications for conservation management. In: *Forests and Dragonflies* (ed. Cordero-Rivera A), pp. 51–78. Pensoft, Sofia.

Pannebakker BA, Niehuis O, Hedley A, Gadau J, Shuker DM (2010) The distribution of microsatellites in the *Nasonia* parasitoid wasp genome. *Insect Molecular Biology*, **19**, 91-98.

Paulson D (2006) The Importance of Forests to Neotropical Dragonflies. In: *Forests and Dragonflies. Fourth WDA International Symposium of Odonatology* (ed. Rivera AC), pp. 79-101. Pensoft Publishers, Sofia.

Pertoldi C, Bijlsma R, Loeschcke V (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation*, **16**, 4147-4163.

Pimm SL, Raven P (2000) Biodiversity: Extinction by numbers. *Nature*, **403**, 843-845.

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. *Science*, **269**, 347-350.

Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945-959.

Rich PV, Rich TH (1983) The Central American Dispersal Route: Biotic History and Palaeogeography. In: *Costa Rican Natural History* (ed. Janzen DH), pp. 12-34. The University of Chicago Press.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources*, **8**, 103-106.

Rutkowski R, Szczuka A, Zalewsk M, Korczyńska J, Gryziak G (2011) Failure of microsatellite's cross-species amplification in common ground beetle *Pterostichus melanarius* (Illiger). *Baltic Journal of Coleopterology*, **11**, 17-24.

Schultz TD, Fincke OM (2009) Structural colours create a flashing cue for sexual recognition and male quality in a Neotropical giant damselfly. *Functional Ecology*, **23**, 724-732.

Schulze CH, Waltert M, Kessler PJA, Pitopang R, Veddeler D, Mühlenberg M, Gradstein SR, Leuschner C, Steffan-Dewenter I, Tscharntke T (2004) Biodiversity indicator groups of tropical land-use systems: Comparing plants, birds, and insects. *Ecological Applications*, **14**, 1321-1333.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Floors P (1994) Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Annals of the Entomological Society of America*, **87**, 651-701.

Solórzano García B, Ellis EA, Rodríguez-Luna E (2012) Deforestation and Primate Habitat Availability in Los Tuxtlas Biosphere Reserve, Mexico. *International Journal of Ecosystem*, **2**, 61-66.

Stehli FG, Webb SD (1985) *The great American biotic interchange*. Plenum Press, New York.

Steinmann H (1997) *World Catalogue of Odonata: Zygoptera*. Walter de Gruyter.

Sutherland WJ, Adams WM, Aronson RB, Aveling R, Blackburn TM, Broad S, Ceballos G, CÔTÉ IM, Cowling RM, Da Fonseca GAB, Dinerstein E, Ferraro PJ, Fleishman E, Gascon C, Hunter Jr M, Hutton J, Kareiva P, Kuria A, Macdonald DW, Mackinnon K, Madgwick FJ, Mascia MB, McNeely J, Milner-Gulland EJ, Moon S, Morley CG, Nelson S, Osborn D, Pai M, Parsons ECM, Peck LS, Possingham H, Prior SV, Pullin AS, Rands MRW, Ranganathan J, Redford KH, Rodriguez JP, Seymour F, Sobel J, Sodhi NS, Stott A, Vance-Borland K, Watkinson AR (2009) One Hundred Questions of Importance to the Conservation of Global Biological Diversity Cien Preguntas de Importancia para la Conservación de la Diversidad Biológica Global. *Conservation Biology*, **23**, 557-567.

Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). (ed. Sinauer Associates S, MA.).

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

Urbina-Cardona JN, Olivares-Pérez M, Reynoso VH (2006) Herpetofauna diversity and microenvironment correlates across a pasture–edge–interior ecotone in tropical rainforest fragments in the Los Tuxtlas Biosphere Reserve of Veracruz, Mexico. *Biological Conservation*, **132**, 61-75.

Watts PC, Rouquette JR, Saccheri IJ, Kemp SJ, Thompson DJ (2004) Molecular and ecological evidence for small-scale isolation by distance in an endangered damselfly, *Coenagrion mercuriale*. *Molecular Ecology*, **13**, 2931-2945.

Watts PC, Rousset F, Saccheri IJ, Leblois R, Kemp SJ, Thompson DJ (2007) Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale*: Odonata: Zygoptera) populations: analysis of 'neighbourhood size' using a more precise estimator. *Molecular Ecology*, **16**, 737-751.

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Jonathan Davies T, Grytnes J-A, Harrison SP, Hawkins BA, Holt RD, McCain CM, Stephens PR (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, **13**, 1310-1324.

Wiens JJ, Graham CH (2005) NICHE CONSERVATISM: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519-539.

Wright SJ, Muller-Landau HC (2006) The Future of Tropical Forest Species. *Biotropica*, **38**, 287-301.

Yanoviak SP (1999) Community structure in water-filled tree holes of Panama: effects of hole height and size. *Selbyana*, **20**, 106–115.

Cryptic non-adaptive speciation in a Neotropical rainforest odonate genus

Wiebke Feindt[1]*, Rob DeSalle[2], Heike Hadrys[1, 2, 3]

[1] University of Veterinary Medicine Hannover, ITZ, Division of Ecology and Evolution, Bünteweg 17d, Hannover, D-30559, Germany

[2] American Museum of Natural History, Sackler Institute for Comparative Genomics, Central Park West at 79th Street, New York, NY 10024, USA

[3] Yale University, Department of Ecology & Evolutionary Biology 165 Prospect Street, New Haven, CT 06511, USA

* Author for correspondence: Wiebke Feindt
Email: wiebke.feindt@ecolevol.de

This work is prepared for submission to Molecular Ecology.

Cryptic non-adaptive speciation in a Neotropical rainforest odonate genus

Abstract

Evolutionary research aims to understand speciation patterns for drawing conclusions about species distributions and species richness. One speciation mode where lineages diversify but maintain their ecological niches is referred as non-adaptive. The underlying causes are a lack of gene flow among lineages, missing adaptation capabilities to novel environmental conditions, limited genetic variation, and pleiotropy. In the present research we investigate non-adaptive speciation in old growth Neotropical rainforests using the worldwide largest odonate species *Megaloprepus caerulatus* as an example while intending to prove speciation in its genus and describe its demographic history. Our integrative survey based on both, field and museum collections across most of *Megaloprepus* distributional range from Southern Mexico to Peru support a cryptic species complex. The population genetics based on mitochondrial and nuclear genes show four distinct species. A relaxed molecular clock indicated that geological events in the Neotropics represented barriers to gene flow. The youngest split among the three Mesoamerican species was estimated at ~3-2 Mya. Relative comparisons of the ecological niches using species distribution modeling and similarity tests revealed highly similar ecological niches underlining phylogenetic niche conservatism among the four species. This ecological similarity could have reduced the development of morphological differences. However, the variation observed in wing coloration and wing shape via linear and geometric morphometrics, might be related to sexual selection. This is challenging the restriction to a single speciation mode towards a more variable pattern. Further research on insects that are restricted to narrow ecological niches over wide geographical ranges could reveal similar patterns, whereas current environmental changes have grave impacts on these species.

Keywords: Non-adaptive speciation, niche conservatism, Neotropics, Odonata, phylogeography, morphometrics

Introduction

'Why lineages differentiate?' and *'How species are distributed in ecological space?'* are central questions in evolutionary research (e.g. Butlin, et al. 2012; Nosil 2012; Pyron, et al. 2015). The ecological niche (cf. Hutohinson 1957) is thereby one key concept (e.g. Wiens and Graham 2005; Pearman, et al. 2008; Peterson, et al. 2011). Hence, revealing the precise pattern of niche evolution permits conclusions about species distribution and ecological conditions causing speciation.

When considering ecological niches over evolutionary time scales, two speciation modes are discussed. Whereas *adaptive or ecological speciation* arises through niche exploration from a standing genetic variation (Nosil 2012; Givnish 2015; Pyron, et al. 2015), the *non-ecological or non-adaptive mode* is characterized as another evolutionary pathway without niche evolution (Wiens 2004; Wiens and Graham 2005). In the latter case, niche conservatism (Ricklefs and Latham 1992; Peterson, et al. 1999) is sustaining the '*ecological status quo*' of lineages when they are isolated by either geographic or climatic barriers (Wiens 2004). The main causes for such niche conservatism are (i) a lack of adaptation capabilities to novel environmental conditions, (ii) limited genetic variation in niche related traits, (iii) high gene flow within but not among isolated lineages, and (iv) pleiotropy (e.g. Wiens 2004; Wiens, et al. 2010; Crisp and Cook 2012). Moreover, strong natural selection for niche constrains over evolutionary timescales results in a higher niche similarity among sister species than it would be expected based on their phylogenetic relationships (phylogenetic niche conservatism – PNC, (Harvey and Pagel 1991)). However, the consequences (Wiens, et al. 2010; Pyron, et al. 2015) and the precise conceptual frameworks (Sobel, et al. 2009; Wiens, et al. 2010; Crisp and Cook 2012; Langerhans and Riesch 2013; Pyron, et al. 2015) of non-adaptive speciation and PNC are still intensively discussed. Recently Pyron and colleagues (2015) linked PNC to a process-based phenomenon instead of a steady pattern, which continuously permits associating PNC with ecological speciation and sexual selection. At last, non-adaptive speciation has been related to slow trait divergence and the appearance of cryptic species (Rundell and Price 2009; Wiens, et al. 2010).

However, Wiens (2004) first mentioned PNC as playing a significant role in allopatric speciation and proved its evidence in North American salamanders (Kozak and Wiens 2006). From a simple ecological perspective, temperate zones hold lower niche diversities than the tropics (i.e. the Neotropics; cf. Bagley and Johnson 2014; Smith, et al. 2014; Antonelli, et al. 2018). Because of this high diversity in ecological and climatic niches on small areas, speciation could be expected to occur mostly through adaptation along different niche gradients while niche conservatism may appear more dominant in temperate zones. Specific examples are adaptive radiations along Neotropical mountain ranges (Hoorn, et al. 2010) or the vertical stratification of close related species inside tropical forests (e.g. Yanoviak 1999; Basset, et al. 2003). But gene flow, genetic potential and pleiotropy as well as constant temperatures over long periods in broad geographic ranges are neglected in this simplification. Today the high biodiversity in the Neotropics is related to both, ecological and non-ecological speciation, without a clear statement on the dominating mode (e.g. Antonelli, et al. 2018).

Phytotelmata or water-filled plant containers are particular microhabitats in forest areas because they represent a small aquatic refuge inside terrestrial environments. In old growth Neotropical rainforests water filled tree holes are common (Kitching 1971; Yanoviak 1999; Yanoviak and Fincke 2005), and their *inter alia* restricted size and relatively easy structure makes them a suitable model system for ecological and evolutionary research (e.g. Kitching 2000). Species compositions inside those tree holes are divers and include facultative and

obligate tree hole breeders, which both depend on the tree hole size and forest altitude (Yanoviak 1999; Kitching 2000). Combining the three factors: habitat specialism, the broad distribution of Neotropical rainforests with its complex geological history implies that allopatric non-adaptive speciation among obligate tree hole breeders is plausible.

Some species of the odonate family Pseudostigmatidae (Odonata: Zygoptera) depend exclusively on water filled tree holes in Neotropical rain forests. Although odonates are on the way for being model organisms in ecological and evolutionary research, speciation is rarely studied in this order (Svensson 2012). So far only three odonate genera are known of showing signs for PNC: *Enallagma*, *Ischnura* and *Calopteryx* (Svensson 2012; Wellenreuther and Sánchez-Guillén 2016). Interestingly in each case, species occur in sympatry and reproductive isolation is strongly related to sexual selection or learned mate preferences (Svensson 2012). Only the shape differences in male secondary reproductive organs of North American *Enallagma* species were strictly associated with drift (McPeek, et al. 2011).

Our objective in the present study was to test if niche conservatism over long evolutionary time scales and broad geographic ranges could have had an influence on odonates living in water filled tree holes. Consequently, we decided to investigate one specific tree hole breeding damselfly: *Megaloprepus caerulatus* (Drury 1782; Odonata: Zygoptera, Pseudostigmatidae) with the aim of identifying potential allopatric speciation and to describe its demographic history. The identification of phylogenetic niche conservatism in a Neotropical forest odonate species will represent a basis to pave the way for complex evolutionary speciation studies on the genomic base of non-adaptive speciation and may allow stronger statements in conservation of the tree hole fauna and other forest odonates.

However, the precise number of species within *Megaloprepus* is unresolved. Today *Megaloprepus* is known as a monotypic genus (Steinmann 1997; Garrison, et al. 2010). But about 150 years ago three species were described in the genus: *M. caerulatus* from the lower parts of Central America, *M. brevistigma* from South America and *M. latipennis* from Mexico (Selys Longchamps 1860), which definite states remained unclear over time (Calvert 1901-1908; Ris 1916; Fincke 2006). An earlier study revealed high genetic structuring between populations from Mexico, Costa Rica and Panama and assumed again speciation in this genus (Feindt, et al. 2014). The geographic origin of the investigated material from 150 years ago and of the 'newer' population study do only overlap in Mexico and the Caribbean coast of lower Central America. Subsequently, to obtain a closer picture of species states and their biogeographic history, we additionally aim to prove speciation in the genus *Megaloprepus* and to evaluate this process under the aspect of niche conservatism, time and morphological evolution.

Material and Methods

Taxon sampling

Our model species *M. caerulatus* is the world's largest damselfly. Although it occurs in a broad distributional range from Southern Mexico to Bolivia, it is a habitat specialist (e.g. Hedström and Sahlén 2001; Fincke 2006). It inhabits old growth rainforests with closed canopies that retain a constant microclimate and as a tree hole breeder *Megaloprepus* requires a high abundance of big trees that form such water filled tree holes (cf. Fincke 1984; Fincke 1992a, 1998). This niche restriction makes *M. caerulatus* an indicator species for forest health but at the same time vulnerable to habitat destruction (e.g. Fincke and Hedström 2008; Feindt, et al. 2014; Khazan 2014).

Tissue samples of 14 populations from southern Mexico to Peru were included, while we retained a greater focus on Mesoamerica (Table 1, Fig. 1). Sampling of entire individuals was reduced to a minimum for two reasons: (i) abundances of the *Megaloprepus* species strongly vary between sites over its occurrence and (ii) population sizes seem to decline especially in Los Tuxtlas, Mexico or the Pico Bonito National Park, Honduras. Therefore, our tissue samples originate from either an adult right leg or a caudal lamella (terminal gill) of larvae (Table A2.1.1) obtained from water filled tree holes (cf. Yanoviak and Fincke 2005).

To study the evolutionary history and the relationships among putative genetic groups, we included a phylogenetic classification of the Pseudostigmatidae. Additional sequences from the National Center for Biotechnology Information (NCBI) were downloaded if not collected by the authors or donated by collaborators (see Table A2.2.1).

The morphometric comparisons are based on adult specimens collected in the field and additional museum material. The latter originated from the Odonate Collection of the American Museum of Natural History (AMNH, New York), the University of Connecticut (UCONN), the National Insect Collection of the National Autonomous University of Mexico (UNAM, Mexico City), and the National Biodiversity Institute in Costa Rica (INBio, San José) (Table A2.3.1). Moreover, the basis for the species distribution modeling and niche comparisons is the geographic information of the field collected material and all museum samples (only if GPS coordinates were available).

Finally, for simplification we pre-sorted specimens by their morphological appearance based on the historical taxonomic descriptions (Selys Longchamps 1860; Ris 1916) and the previous population genetic distance measurements (Feindt, et al. 2014). This allowed us to separate the samples into four different putative *Megaloprepus* species: *M. caerulatus*, *M. latipennis*, *M. brevistigma* and *Megaloprepus* sp. nov.. The names for *M. latipennis* and *M. brevistigma* were adopted from the original description (Selys Longchamps 1860) because of their morphological similarity to our samples.

DNA extraction and Sequencing

DNA was extracted using phenol chloroform. For dry material the protocol was modified as follows: first the tissue underwent initially a 'water soaking step' preventing to raise dry

tissue dusk and therefore contamination during grinding and second the digestion step was extended up to three days, whereat each day additional 25 µl proteinase k was added.
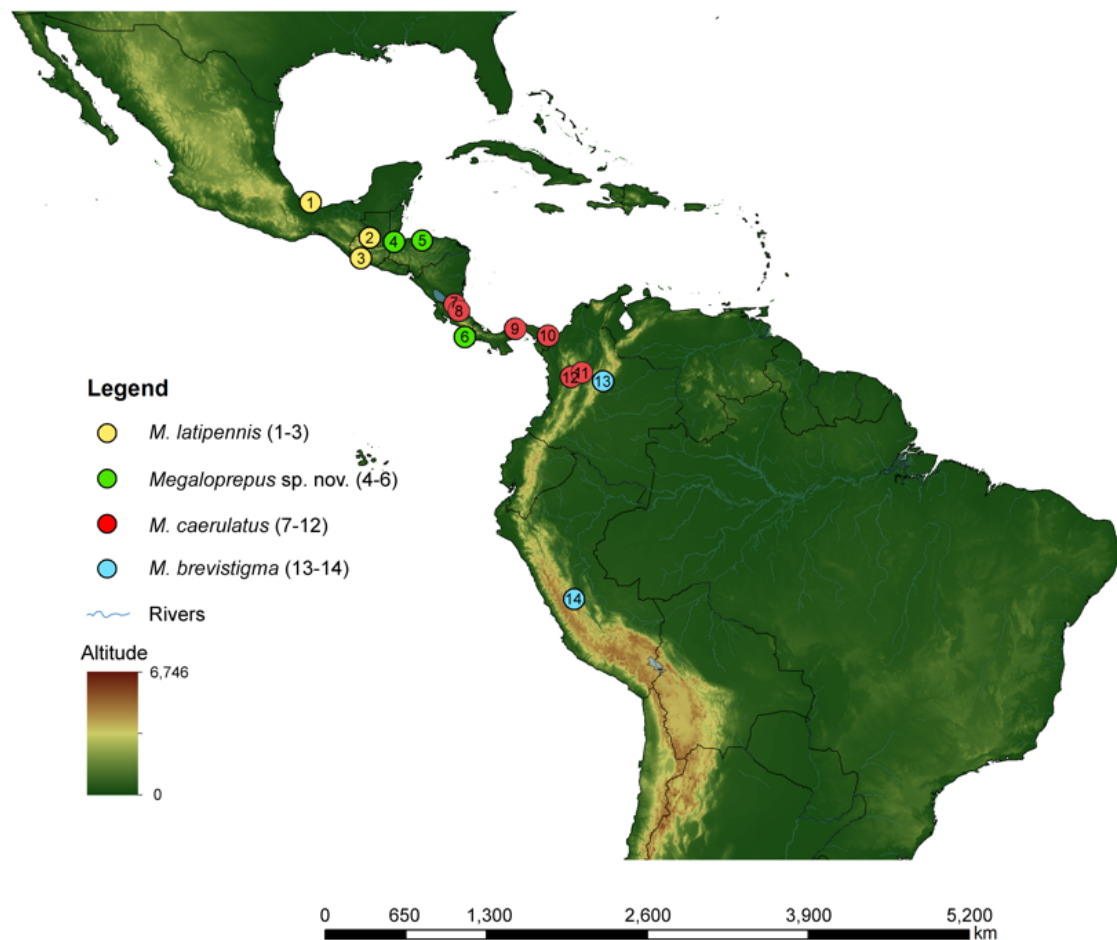
Figure 1: Sampling points for the genus *Megaloprepus*. The coloration of the points corresponds to the designated species and the numbers within the circles to the geographic localities*.
* 1: Biosphere Reserve Los Tuxtlas, Mexico; 2: National Park Laguna Lachuá, Guatemala; 3: Río Bravo, Guatemala; 4: Natural Reserve Cerro San Gil, Guatemala; 5: Pico Bonito National Park, Honduras; 6: Corcovado National Park, Costa Rica; 7: Biological Reserve Indio Maíz, Nicaragua; 8: Biological Research Station La Selva, Costa Rica; 9: Barro Colorado Island, Panama; 10: Chocó, Colombia; 11: Norcasia, Colombia; 12: Antioquia, Colombia; 13: Boyacá, Santa Maria, Colombia; 14: Pampa Hermosa Lodge, Peru.

The population comparisons are built upon three mitochondrial sequence markers: the barcoding region CO1 (Folmer, et al. 1994); the ND1 region (Abraham, et al. 2001); and a fragment of 16S ribosomal RNA (rRNA) (Simon, et al. 1994). Furthermore, two nuclear genes were amplified for a small subset of each putative species. We selected two overlapping regions of the elongation factor 1α (EF1α) using the primer pair ef7 and ef9 (Simon, et al. 2010) and EF1-F-2652 and EF1-R-3093 (Jordan, et al. 2003). As second nuclear marker we chose ITS I+II (internal transcribed spacer region; Damm, et al. 2010). For the phylogenetic reconstruction we used we used gene fragments for 12S rRNA

(Roehrdanz 1997) and 28S rRNA (Dijkstra, et al. 2014) in addition to the three mitochondrial sequence markers and EF1α described above.

All obtained PCR products were purified using either ethanol precipitation or polyethylene glycol (PEG) at a final concentration of 13%, before bidirectional Sanger sequencing at the DNA Analysis Facility on Science Hill, Yale University (dna-analysis.yale.edu/) or the Sackler Institute for Comparative Genomics, American Museum of Natural History.

Bioinformatics

Obtained raw sequences were assembled, reviewed, if necessary, edited in Geneious vers. 8.1.7 (Kearse, et al. 2012) and verified by local blast searches (Altschul, et al. 1997). We performed alignments for each locus using MUSCLE vers. 3.6 (Edgar 2004) implemented in Seaview vers. 4.6.1 (Gouy, et al. 2010), where we also trimmed overlapping ends. Local alignments were joined using MEGA7 (Kumar, et al. 2008). All sequences achieved in this study are deposited at NCBI (see Tables A2.1.1 for accession numbers).

Population genetics

We conducted the population genetic analyses for a concatenated data set including CO1, ND1 and 16S and for each mitochondrial and nuclear marker separately. For this purpose, additional sequences for ND1 and 16S were added from Feindt *et al.* (2014) (Table A2.1.1). Basic indices of genetic diversity such as nucleotide diversity (π) and haplotype diversity (Hd), number of haplotypes (*h*) and segregating sites (S) were estimated in DnaSP vers.5.10 (Librado and Rozas 2009) considering gaps as a fifth state.

To graphically display the genealogical relationships between populations we calculated a haplotype network based on statistical parsimony in TCS vers. 1.2.1 (Clement, et al. 2000) with the 95% default parsimony connection limit. The genetic relationships between populations were estimated using different methods: Genetic distance measures using the Kimura-2-Parameter substitution model (Kimura 1980) and uncorrected *p*-distances (cf. Srivathsan and Meier 2012) were performed in MEGA7 (Kumar, et al. 2008) applying a pairwise deletion of gaps. ARLEQUIN vers. 3.5.2.2 (Excoffier and Lischer 2010) was used to determine genetic differentiation parameters. We measured the pairwise differentiation among sample populations using the F_{ST} approach as indirect information of the level of migration among populations and the significance determined by 10,000 bootstrap replicates (Weir and Cockerham 1984; Cockerham and Weir 1993). An Analysis of Molecular Variance (AMOVA, Excoffier, et al. 1992) was performed to detect partitioning of variance among groups. We estimated the amount of genetic variation at different hierarchical levels attributable with the genetic differentiation among predefined groups (regions) (ΦCT), among populations within groups (ΦSC) and among all populations relative to the total sample (ΦST) and the corresponding fixation indices via Wright's *F* statistic (Wright 1949; Holsinger and Weir 2009).

Phylogenetic analyses

Into the phylogenetic estimations and divergence time estimations we included 15 of the so far 25 described Pseudostigmatidae (cf. Steinmann 1997; Fincke 2006; Garrison, et al. 2010; Ingley, et al. 2012; Machado and Lacerda 2017) and the four potential *Megaloprepus* species. The Pseudostigmatidae are a small and almost exclusive Neotropical damselfly family, but its position in the odonate tree of life is under discussion (Dijkstra, et al. 2014; Machado and Lacerda 2017). Its closest related species *Coryphagrion grandis* inhabiting eastern African coastal forests is linking this family as a Gondwana relict (Groeneveld, et al. 2007). However, here we present new genetic material for *Pseudostigma accedens*, *Microstigma maculatum* and *Mecistogaster amalia* and new sampling localities for eight other Pseudostigmatids. As outgroups we included the genera *Bromeliagrion* and *Telebasis*, because the recent work from Dijkstra and colleagues (2014) indicated them as close related and they had sufficient genetic information available on NCBI.

Phylogenies were estimated using both, maximum likelihood (ML) and Bayesian (BI) methods. For each gene we performed single ML gene tree to revise tree topologies before concatenating alignments. We selected the general time reversible (GTR) model of nucleotide substitution (Rodriguez, et al. 1990) including a gamma distributed rate heterogeneity (Γ) and an estimated proportion of invariable sites (GTRGAMMA) for the ML analysis realized in RaxML vers. 8.2.9 (Stamatakis 2014). The complete sequence matrix was partitioned by gene and support values were estimated via 1,000 bootstrap replicates. Prior Bayesian tree estimation in MrBayes vers. 3.2.6 (Ronquist, et al. 2012), we searched for the most likely substitution model for each sequence marker separately with jModelTest vers. 2.1.4 (Posada 2008; Darriba, et al. 2012). Under the corrected Akaike information criterion (AICc) the suggested models were either TPM ('3-parameter model' = K81, (Kimura 1981)) or TIM ('transitional model', (Posada 2003)) with rate variation among sites (+G) or invariable sites together with rate variation among sites (+I+G); (12S rRNA = TIM3+G, 16S rRNA = TIM3+I+G, 28S = TIM2+G, CO1 = TIM2+I+G, EF1α = TPM2+G, ND1 = TIM1+G). Finally, we run in the BI four independent chains for 20 million generations and sampled every 10,000 generations. The first 25% were discarded as burn-in for calculating Bayesian posterior probabilities and building the consensus trees. In addition, we calculated gene trees only for the mtDNA genes (12S rRNA, 16S rRNA, CO1 and ND1) and the nuclear genes (Ef1α and 28S). Hereby we used the same settings as described above. However, in the BI we reduced sampled generations to 2 million and sampled every 1,000 generations.

Time-calibrated phylogenies were computed to estimate the relative species splits within *Megaloprepus*. We reduced the data set to one representative per species and used *C. grandis* as outgroup. The concatenated alignment was partitioned 12S/16S, ND1/CO1 and 28S/EF1α for the substitutions and clock models. For those partitions we applied the following substitution rates: for 16S and 12S rRNA the standard insect mitochondrial DNA substitution rate of 0.0115 per site per My (= 2.3% My^{-1}) was used (Brower 1994), whereas for ND1 and CO1 we applied a revised divergence rate of 3.54% My^{-1} (Papadopoulou, et al. 2010). Unfortunately, no substitution rates are described for 28S rRNA and EF1α, which led

to a higher uncertainty. All estimations were calculated in BEAST vers. 1.8.4 (Drummond, et al. 2012) applying the GTR+G substitution model with six rate categories (Tavaré 1985).

A strict molecular clock model was realized as a first impression (cf. File A2.2). The obtained time estimations, however, appeared uncertain. Consequently, and because relaxed clocks promise more reliable results, we performed an uncorrelated, lognormal relaxed clock (Drummond, et al. 2006). Hereby we placed a lognormal prior on the mean (R (0.0115/0.25) for 12S/16S and (R (0.0177/0.25) for ND1/CO1), which centered the mean rate of relaxed clocks at 1.01% per million years (95% prior density: at 0.62-1.65% for 12S/16S and at 0.62-1.66% for ND1/CO1). For 28S rRNA and EF1α the relaxed clock had a broader lognormal prior for the mean (R (0.0115/1.0) setting the mean rate between 0.14% and 7.18%. For all, the standard derivation had an exponentially distributed prior with a mean of 0.1. As a tree prior we used a birth-death-model (Gernhard 2008) with an exponential prior on the growth *birth rate* (mean = 0.031) and a uniform prior on the relative death rate (U (0,1)). The remaining settings stayed as default. We run 6 independent runs with each 10×10^8 (MCMC) generations and sampled every 10,000. Tracer vers. 1.6.0 was used to determine stationery and convergence by monitoring the effective sample size (ESS) and the influence of our priors on the data was checked via an empty run using the priors only (Drummond, et al. 2006). Finally, the tree files were combined with LogCombiner vers. 1.8.4 using a conservative burn-in of 20% for each reach run and TreeAnnotator vers. 1.8.4 was used to summarize maximum clade credibility trees (MCC) with a posterior probability limit of 0.5. Trees were viewed in FigTree vers. 1.4.2 (Rambaut and Drummond 2015).

To additionally validate our results in a bigger context we calculated a relaxed molecular clock with one fossil calibration as described in Callahan & McPeek (2016). Hereby we used a dated fossil of the genus *Ischnura* that is between 16.4 – 20.5 My old (Mitchell 2007) and was found in Dominican amber (Bechly 2000). For a closer method description please see File A2.2.

Morphometrics

Linear and geometric morphometrics (GMM) were used to describe variation in wing morphology among the four putative species. To do so, we used standardized photographs (see Table A2.3.1 for a complete species list); but because most specimens were museum samples, wings had to be photographed still attached to the thorax. However, the pictures of single wings were taken in an angle of 90 degrees, with an appropriate background and a measurement scale. To guarantee a plane record of wing venation, the wings were weighted with glass object slides. Furthermore, although observed size differences were not significant, smaller females of one population showed overlaps with bigger males of an in general smaller population. Thus, only males were considered for the morphometrics to avoid bias of sexual variation in our analysis.

First 28 linear variables (Table A2.3.2) were analyzed for the four putative species: *M. caerulatus* ($N = 31$), *M. latipennis* ($N = 17$), *M. brevistigma* ($N = 10$) and *Megaloprepus* sp.

nov. (N = 8). Measurements were realized in ImageJ vers. 1.48 (Abramoff, et al. 2004) using the mean of three independent measurements for each variable. The analyses were carried out using R (RCoreTeam 2014) for hind and forewings separately since a covariance test indicated correlation. All values were log transformed prior statistical analyses. For an initial impression of groupings, both cluster analysis based on euclidean distances and principal component analysis (PCA) were performed. A multi-axis discriminant analysis (canonical variates analysis, CVA) was calculated and further accompanied by validation tests ("leave-one-out cross-validation") to obtain conservative group assessments and prediction accuracy by the percentages of individuals correctly assigned into their groups.

Secondly, GMM's were conducted to study the variation in wing shape among groups. On the *Megaloprepus* photographs two-dimensional landmarks were placed on 42 homologous wing venation points (Table A2.3.3) using tpsDig vers. 2.16 (Rohlf 2010). Prior group comparisons we tested for potential digitization and orientation errors (Adriaens 2007) and removed the size component from our data set (e.g. Klingenberg and McIntyre 1998; Mitteroecker, et al. 2013). Please compare File A2.3 for a more concrete description on the necessary data pre-processing. Inter- and intragroup differences were investigated first by a PCA. For the CVA pairwise distances between groups were calculated using both, Mahalanobis and Procrustes distances (applying 10,000 permutations for the *p*-value assessment). The following discriminant function analysis (DA) was accompanied by pairwise cross-validation tests. All three analyses (PCA, CVA and DA) were applied in MorphoJ vers.1.06b (Klingenberg 2011). To further verify significant group differences a nonparametric MANOVA (NPMANOVA) was conducted in PAST vers. 3.05 (Hammer, et al. 2001) using Euclidean distances. Hereby, the *p*-values were displayed using a Bonferroni adjustment (10,000 permutations), and the F-statistics were calculated (Goodall´s F-test) (Webster and Sheets 2010). Finally, thin-plate-splines (TPS) from the landmark configuration were used to visualize shape differences according to the PC and CV axes, respectively (Klingenberg 2011).

Species distribution modeling

The present species distribution model (SDM) is built on 156 revised, individual records for the complete genus whereof 67 belong to *Megaloprepus* sp. nov., 27 to *M. latipennis* and 62 to *M. caerulatus* with most of the localities represented in the genetic analyses (Fig. A2.4.1). Records are regionally restricted for the following reasons: First *Megaloprepus* occurs in low abundances and sightings are less common in rare forest patches, so that most records are derived from well-known collection areas. Furthermore, older collections (> 20 years old) mostly lack exact GPS occurrence data and couldn't be included. Finally, *M. brevistigma* was excluded because we had only three precise records (Proosdij, et al. 2015) and the SDMs showed unreliable results. Furthermore, we did not consider species interactions. But *Megaloprepus* is one of the top predators in the tree holes (Fincke 1992b, 2006) and being the top predator in water filled tree holes implies a higher probability that

Megaloprepus is preserving its ecological nice. In turn this underlines the suitability of SDMs for *Megaloprepus*.

We estimated *Megaloprepus'* potential geographic distribution under current climatic conditions and during Pleistocene ice ages using the following settings: for the current model we modeled the genus *Megaloprepus* and the three potential species separately. Whereas for the Last Interglacial (LIG, ~120,000-140,000 years BP), the Last Glacial Maximum (LGM, ~22,000 years BP) and Mid-Holocene (MH, ~6,000 years BP) we only used the complete 156 records. The according climatic variables were downloaded from the worldclim database with a spatial resolution of 30 arc-seconds for the current, MH and LIG (Otto-Bliesner, et al. 2006); and at 2.5 arc-minutes for the LGM (www.worldclim.org, Hijmans, et al. 2005).

Since our model species is a tree hole breeder, we are aware that tree hole specific variables such as water chemistry, water content and water temperature could have an influence on the distribution of *Megaloprepus*. But such data is rare and the tree species itself or animals living within the tree holes might additionally influence water chemistry. Nevertheless, we selected the variables because of their impact on the microhabitat and *Megaloprepus'* physiology (cf. Collins and McIntyre 2015). Furthermore, we excluded climatic variables that are highly correlated (Spearman's $|rho| < 0.8$). The final set of four bioclimatic variables included: annual precipitation (BIO 12), precipitation of driest quarter (BIO 17), annual mean temperature (BIO 1) and mean temperature of driest quarter (BIO 9). We additionally checked for an influence of the vegetation type on the potential distribution. The tested forest data (Global Land Cover (GLC-SHARE; Latham, et al. 2014) and MODIS Land Cover data (Friedl, et al. 2010; Channan, et al. 2014)) did not show an improvement of the model; rather we observed an overestimation of potential occurrences in the whole *Megaloprepus* model, but a significant reduction in the single models (see Fig. A2.4.2). This could be because the climatic variables define forest types (i.e. precipitation).

Biovariables and occurrence data were checked, prepared and cut into the sample area from 33°N to 55°S and from 122°W to 33°E using R (RCoreTeam 2014). The SDM itself was developed using MaxEnt vers. 3.3.3k (Phillips, et al. 2006). Hereby the MaxEnt default settings were retained except for the regularization parameter, which was changed from 1 to 1.5 to avoid model over-estimation (e.g. Norris 2014; Proosdij, et al. 2015) and the number of repetitions was increased to 20. The evaluation of the modeling was based on the cross-validation resampling method and the MaxEnt Area under the Receiver-Operating characteristic curve - AUC score (Fielding and Bell 1997).

Finally, to test for niche evolution and to infer potential adaptation we performed niche equivalency and background similarity tests implemented in the R package phyloclim (Heibl and Calenge 2013) based on the obtained densities of occurrences. Hereby Schoener's *D* (Schoener 1968) and the modified Hellinger's *I* as an index of niche space (van der Vaart 1998) are used to compare niche models with random models obtained from 100 pseudoreplicate data sets (Warren, et al. 2008).

Table 1: Overview of the *Megaloprepus* sampling localities and the associated genetic diversity indices *N* = number of individuals, *P* = nucleotide diversity in percent, h = number of haplotypes, Hd = haplotype diversity and S = number of polymorphic sites, based on the concatenated alignment for ND1, 16S and CO1 (1,593bp; CO1: 639 bp, ND1: 474 bp, 16S: 482 bp).

Species wide diversity indices for the entire mitochondrial data set are as follows:
M. latipennis: *N* = 31, *P* = 0.29 (± 0.0004), h = 16, Hd = 0.94 (± 0.021); *Megaloprepus* sp. nov.: *N* = 64, *P* = 0.56 (± 0.0003), h = 22, Hd = 0.82 (± 0.039); *M. caerulatus*: *N* = 91, *P* =0.72 (± 0.0004), h =37, Hd =0.81 (± 0.041), *M. brevistigma*: N = 2, P = 0.04 (± 0.0200), h = 2, Hd = 1.0 (± 0.500)
* Abbreviations of area codes are as used in the entire description of genetic results.

| putative species | Geographical location; abbreviation | Latitude / Longitude | *N* | mt DNA diversity | | | |
				h	*P*	Hd	S
M. latipennis	Biosphere Reserve Los Tuxtlas, Mexico; RBLT	N 13°35'06.35" / W 95°04'30.22"	27	14	0.28 ± 0.0005	0.93 ± 0.026	20
	National Park Laguna Lachuá, Guatemala; GuLL	N 15°56'49.7" / W 90°39'46.5"	3	1	0.00 ± 0.0000	0.00 ± 0.000	0
	Río Bravo, Guatemala; GuRB	N 14°24'52.1" / W 91°18'33.4"	1	1	/	/	/
Megaloprepus sp. nov.	Natural Reserve Cerro San Gil, Guatemala; GuCSG	N 15°38'22.9" / W 88°49'49.2"	15	7	0.37 ± 0.0005	0.79 ± 0.079	14
	Pico Bonito National Park, Honduras; HnPb	N 15°43'14.5" / W 86°44'35.5"	14	4	0.02 ± 0.0001	0.28 ± 0.148	3
	Corcovado National Park, Costa Rica; CrCNP	N 8°32'32.42" / W 83°34'13.39"	35	11	0.08 ± 0.0002	0.53 ± 0.103	19
M. caerulatus	Biological Reserve Indio Maíz, Nicaragua; NiBa	N 10°57'56.1" / W 84°20'13.3"	16	2	0.01 ± 0.0001	0.13 ± 0.106	0
	Biological Research Station La Selva, Costa Rica; CrLS	N 10°28'58.65" / W 84°00'58.96"	35	9	0.07 ± 0.0002	0.49 ± 0.103	14
	Barro Colorado Island, Panama; BCI	N 9°09'54.07" / W 79°50'12.27"	32	15	0.15 ± 0.0002	0.92 ± 0.026	15
	Choco, Colombia; CoCho	N 8°38'39" / W 77°21'35"	5	5	0.94 ± 0.0020	1. 00 ± 0.126	30
	Norcasia, Colombia; CoNor	N 5°34'1.2" / W 75°39'32.4"	1	1	/	/	/
	Antioquia, Colombia; CoA	N 5°53'7" / W 74°51'48"	5	5	0.31 ± 0.0007	1.00 ± 0.126	11
M. brevistigma	Boyacá, Santa Maria, Colombia; CoBo	N 5°14'16.8" / W 73°16'15.3"	1	1	/	/	/
	Pampa Hermosa Lodge, Peru; PePH	S 10° 59' / W 75° 25'	1	1	/	/	/
Σ			191	77	0.04 ± 0.0012	0.93 ± 0.012	322

Results

The present analyses give evidence for a past speciation event in *Megaloprepus* with at least four recent species. Furthermore, the morphological comparisons depict the appearance of diverging wing pattern in the light of ecological niche similarity.

Bioinformatics

Population genetics

The complete data set includes 191 tissue samples of *Megaloprepus* (Table 1). A total alignment length of 1,595 bp for CO1, ND1 and 16S rRNA was obtained, whereat 1,265 sites are invariable, and 322 characters are polymorphic but 251 are only parsimony informative. ITS and Ef1α, however, have with almost equal lengths of 659 bp and 668 bp, respectively, fewer informative sites (ITS: 22 polymorphic sites and 17 parsimony informative sites; EF1α: 22 polymorphic sites and 18 parsimony informative sites; $N = 48$ individuals from 8 populations). The genetic diversity measurements for the mitochondrial genes appear relatively high, however numbers vary dramatically if single sequence markers are considered (File A2.1). With respect to unequal sample sizes, the specimens collected in Nicaragua (NiBa) and Caribbean Costa Rica (CrLS) have the lowest nucleotide and haplotype diversities in all studied genes. In the two northern populations, from the Biosphere Reserve Los Tuxtlas (RBLT) and the Natural Reserve Cerro San Gil (GuCSG), nucleotide and haplotype diversity are higher.

The genealogical relationships between populations point towards a strict separation in the mtDNA sequence markers (Fig. 2). A total of 77 haplotypes were observed, whereas most interestingly no shared haplotypes among clusters either most populations were detected. Mexican and north Guatemalan samples clustered together into one network with 16 haplotypes, whereby 3 mutational steps separate Laguna Lachuá (GuLL) and 8 mutational steps the Pacific lowlands sample Río Bravo (GuRB) from Mexico. A second cluster contains samples from the Pacific Costa Rica (CrCNP, H = 11) that is connected via 13 mutational steps to Honduras (HnPb) and Guatemala (GuCSG). Samples from Nicaragua, Costa Rica, Panama and Colombia are in the third cluster. Although the Biological Research Station La Selva is geographically closer to Barro Colorado Island than to northern Colombia, the Colombian populations appear linking BCI and CrLS. Additionally, among all Colombian samples we observed a high separation from each other by up to 11 mutational steps (CoA, CoNor and CoCho). Most interestingly, the two samples from the east side of the Andes (Boyacá, Colombia and Pampa Hermosa Lodge, Peru) are situated separately within the network and take a special position. In contrast to the mtDNA, the nuclear sequence markers show a slightly different picture. Here differentiation is less obvious, and haplotypes are shared even between putative species (Fig. 2, Fig. A2.1.1). The ITS network has two separated clusters: all Mesoamerican samples oppose the eastern Andes specimens. Within the 'Mesoamerican cluster' 4 to 6 mutational steps separate the Corcovado National Park and Biological Reserve Los Tuxtlas. In EF1α, all samples cluster

into one single network together, whereat samples from CrLS, BCI, Co and RBLT share haplotypes, but not the CNP.
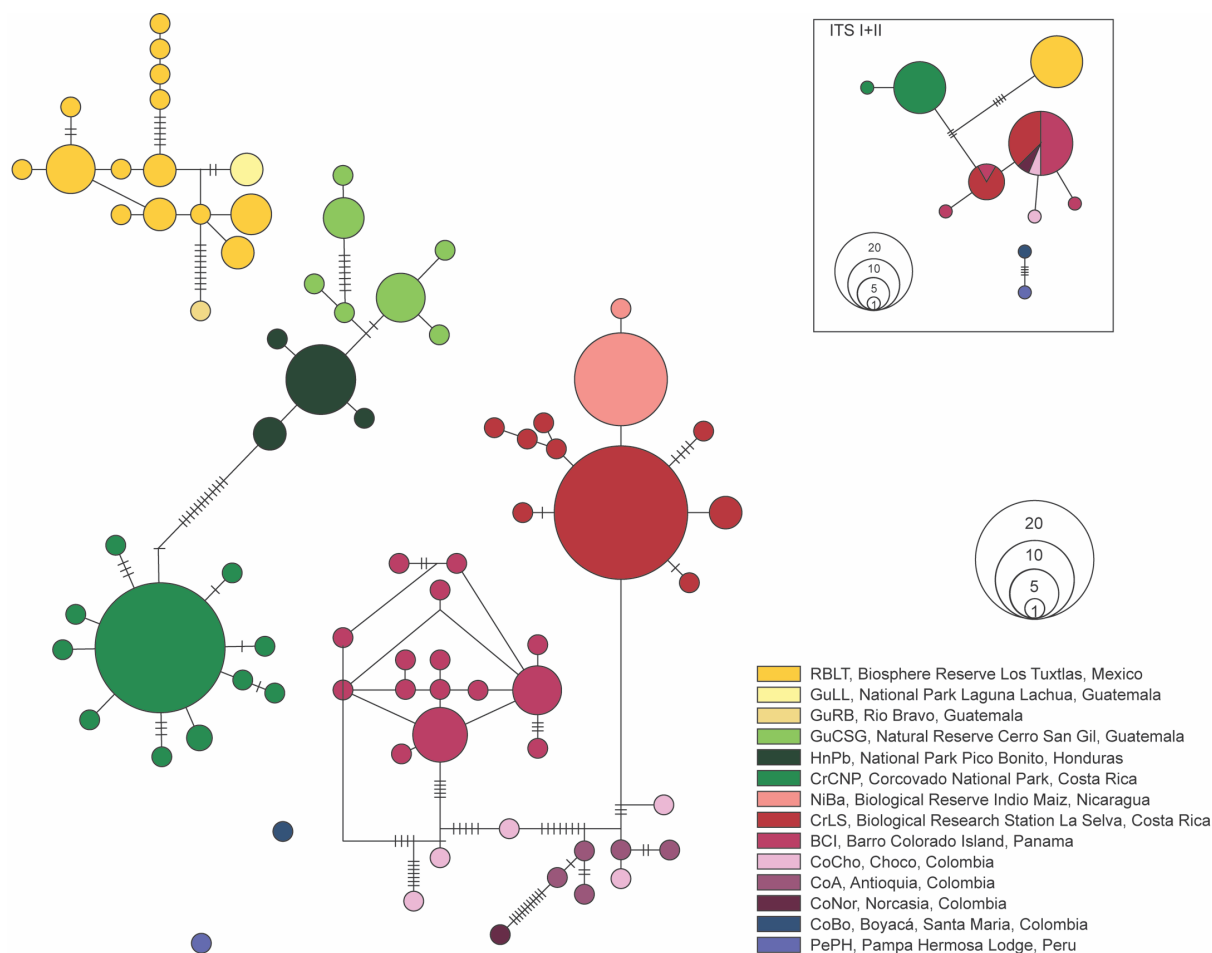
Figure 2: Haplotype networks showing the genealogical relationships within the genus *Megaloprepus* from southern Mexico to Peru based on statistical parsimony (95% connection limit) in TCS vers. 1.2.1 (Clement, et al. 2000). The genetic foundation in the main figure is the concatenated alignment for three mitochondrial sequence markers (1,593 bp: CO1 ND1, 16S) for in total 191 individuals. The smaller inserted haplotype network shows the relationships among 48 randomly selected individuals representing the four putative species nuclear gene marker ITS I+II (internal transcribed spacer region, 659 bp).

Each haplotype is represented by one circle, which size proportional to the number of individuals belonging to this haplotype. Black lines connect haplotypes; each representing a mutational step and each small cross line represents an additional mutational step. Finally, haplotypes are color coded in accordance with their origin.

Pairwise nucleotide divergences underline a genetic separation into distinct clusters mirroring genetic distances at species level (Table 2). While the genetic distances are high between clusters, they are low within clusters. The highest sequence divergences in the single marker comparisons were detected in CO1 12.84% between NiBa and PePH, in ND1 14.29% between GuLL and PePH and 10.40% in 16S between GuRB and PePH (cf. File

A2.1). However, the concatenated mtDNA sequence divergences between clusters range from 5.33 to 11.35%, whereas within clusters divergence ranged from 0.04 (CrLS vs. NiBa) to 1.96% (BCI vs. CoNor) (Table 2). The two eastern Andean samples (CoBo, PePH) show always a high genetic distance to all other samples and a moderate distance of 3.95% between each other. In the nuclear genes we observed the same tendency, although genetic distances were with maximal values between clusters of 3.08% in ITS and 2.03% in EF1α considerably lower than in the mtDNA (cf. File A2.1).

Our gene flow estimations via F_{ST}-values (Weir and Cockerham 1984) further reflect these relationships. In the present dataset, significant genetic isolations between clusters with F_{ST}-values ≥ 0.9 and a nearly complete isolation among populations within clusters was observed (Table 2). Exceptions are for example RBLT and GuRB, NiBa and CrLS, and the northern Colombian (CoCho, CoNor, CoA) populations.

Using the AMOVA we estimated how genetic differentiation is distributed within and among groups and populations (Table 3). At the hierarchical level of two groups (Mesoamerica vs. East Andes) the percentage of variation among regions was nearly as high as the variation among populations within regions. However, using four groups the highest amount of variation was explained among groups and the variation among populations within groups was substantially lower. Wright's *F* statistic (Wright 1949) revealed significant differentiation among regions and among populations within regions. Only in model 2 a lower differentiation among populations was observed with $\Phi ST = 0.51$ (*p* = 0.001).

Table 2: Estimates of evolutionary divergence between and within (bold) populations calculated for 14 populations from *Megaloprepus'* northern distributional range using a concatenated alignment for CO1, ND1 and 16S rDNA (1,595 bp). Analyses were conducted using uncorrected *p*-distances. Rate variation among sites was modelled with a gamma distribution in MEGA7 (Kumar, et al. 2016). The F_{ST}-values (italic) were determined in Arlequin vers. 3.5.2.2 (Excoffier, et al. 1992; Excoffier and Lischer 2010) indicate nearly no gene flow between populations. Significant values are indicated with an asterisk (*) and abbreviations of area codes follow Table 1.

	RBLT	GuLL	GuRB	GuCSG	HnPb	CrCNP	NiBa	CrLS	BCI	CoCho	CoNor	CoA	CoBo	PePH
RBLT	**0.28**	*0.45**	*0.65*	*0.95**	*0.97**	*0.97**	*0.98**	*0.98**	*0.97**	*0.95**	*0.96*	*0.96**	*0.97*	*0.97*
GuLL	0.27	**0.00**	*1.00*	*0.95**	*0.99**	*0.99**	*1.00**	*0.99**	*0.98**	*0.92**	*1.00*	*0.97**	*1.00*	*1.00*
GuRB	0.65	0.50	**n/c**	*0.93*	*0.99*	*0.99**	*1.00*	*0.99**	*0.98**	*0.87*	*1.00*	*0.96*	*1.00*	*1.00*
GuCSG	5.83	5.79	5.33	**0.37**	*0.42**	*0.83**	*0.98**	*0.98**	*0.97**	*0.94**	*0.96*	*0.96**	*0.96*	*0.96*
HnPb	5.91	5.87	5.41	0.35	**0.02**	*0.93**	*1.00**	*0.99**	*0.98**	*0.97**	*1.00*	*0.99**	*1.00*	*1.00*
CrCNP	6.05	6.01	5.56	0.96	0.93	**0.07**	*0.99**	*0.99**	*0.99**	*0.98**	*0.99*	*0.99**	*0.99*	*0.99*
NiBa	7.07	7.18	7.17	7.68	7.74	7.64	**0.01**	*0.55**	*0.92**	*0.68**	*0.99*	*0.84**	*1.00*	*1.00*
CrLS	7.09	7.20	7.19	7.70	7.76	7.65	0.04	**0.07**	*0.92**	*0.70**	*0.95*	*0.77**	*0.99*	*0.99*
BCI	7.08	7.18	7.10	7.40	7.42	7.38	1.30	1.32	**0.15**	*0.71**	*0.93*	*0.89**	*0.99*	*0.99*
CoCho	7.06	7.17	7.15	7.63	7.67	7.59	0.71	0.74	1.01	**0.94**	*0.27*	*0.26**	*0.91*	*0.92*
CoNor	7.38	7.49	7.48	8.13	8.18	8.13	1.20	1.23	1.96	1.29	**n/c**	*0.67*	*1.00*	*1.00*
CoA	7.15	7.26	7.24	7.87	7.94	7.83	0.39	0.43	1.48	0.85	0.95	**0.31**	*0.97*	*0.97*
CoBo	10.34	10.21	10.38	9.94	9.95	9.96	10.57	10.59	10.50	10.45	10.37	10.71	**n/c**	*1.00*
PePH	10.62	10.52	10.69	9.94	9.89	10.08	11.14	11.16	11.07	11.04	11.19	11.35	3.95	**n/c**

Phylogenetic analyses

The concatenated alignment with a total length of 4,117 bp (EF1α = 583 bp, 12S rRNA = 386 bp, 16S rRNA = 545, 28S = 1,398 bp, CO1 = 655 bp, ND1 = 550 bp) included 27 individuals from 19 Pseudostigmatids. Obtained tree topologies were highly similar for all analyses and similar to previous published research (e.g. Groeneveld, et al. 2007; Ingley, et al. 2012). Recent taxonomic classification within the Pseudostigmatidae is supported, but with one exception: *Mecistogaster modesta* clusters into the genus Pseudostigma. The remaining *Mecistogaster* and the newly redefined genus *Playstigma* (Machado and Lacerda 2017) appear as sister groups and form a monophyletic clade with *Pseudostigma*. On the opposite *Megaloprepus* forms a second clade together with the genera *Microstigma* and *Anonisma* (Fig. 3, Fig. A2.2.2 & A2.2.3).

More importantly, however, is the strict separation within the genus *Megaloprepus*. The two individuals from the eastern part of Andes belonging to *M. brevistigma* stand as a sister group to the three Mesoamerican samples. Within Mesoamerica we observed the same separation as seen in the genetic distance measurements and the haplotype network: Individuals from Mexico and north Guatemala are in one clade together with specimens from the Nicaragua and Costa Rica Caribbean, Panama, and Western Andes Colombia, defined as *M. latipennis* and *M. caerulatus*, respectively. Last, samples from CNP, HnPb and GuCSG associated with *Megaloprepus* sp. nov. constitute a sister clade to the previous. In the time estimations, all our analysis fulfilled the recommended threshold of an effective sample size (ESS > 200) in all parameters indicating stationary and enough MCMC runs. The estimated posterior mean divergence times of the relaxed molecular clock showed a split between *M. brevistigma* and *Megaloprepus* species from Central America at ~10-8 Mya during the Pliocene (Fig. 3). *Megaloprepus* sp. nov., *M. caerulatus*, and *M. latipennis* are more derived and split shortly before or either at the beginning of the Pleistocene (~3-2 Mya). However, in the Densi.Tree analysis, in which very thin lines represent individual posterior distributions, we observed uncertainties in the tree topologies within the genera *Microstigma*, *Platystigma, Pseudostigma* and *Megaloprepus*.

With the aid of the fossil calibration, we could support this result. The posterior mean divergence times for the different assumed *Megaloprepus* species were close to the ones estimated using the Pseudostigmatidae only data set (Fig. A2.2.5, A2.2.6). In addition, we were able to reproduce the results of Callahan and McPeek (2016).

Table 3: Analysis of molecular variance, AMOVA (Excoffier, et al. 1992) for the 14 sample populations. The four different models represent different partitions, where we separated the populations in 1, 2, 4 and 5 groups according to the genetic distance measures.

model	Component of differentiation	d.f.	Sum of squares	Variance components	Variation in %	FST
A	Among populations	13	7109.20	42.96	**97.19**	$\Phi ST = 0.97^{***}$
	Within populations	177	220.00	1.24	**2.81**	
	Total	190	7329.20	44.20		
B	Among regions	1	234.12	45.47	**51.33**	$\Phi CT = 0.51^{**}$
	Among populations within regions	12	6875.09	41.87	**47.26**	$\Phi SC = 0.97^{***}$
	Among all populations	177	220.00	1.24	**1.4**	$\Phi ST = 0.99^{***}$
	Total	190	7329.20	88.58		
C	Among regions	3	6370.88	51.04	**86.96**	$\Phi CT = 0.87^{***}$
	Among populations within regions	10	738.32	6.41	**10.92**	$\Phi SC = 0.83^{***}$
	Among all populations	177	220.00	1.24	**2.12**	$\Phi ST = 0.98^{***}$
	Total	190	7329.20	58.69		
D	Among regions	4	6402.38	51.33	**87.34**	$\Phi CT = 0.83^{***}$
	Among populations within regions	9	706.82	6.12	**10.54**	$\Phi SC = 0.98^{***}$
	Among all populations	177	220.00	1.23	**2.11**	$\Phi ST = 0.87^{***}$
	Total	190	7329.20	58.77		

[a] Using the AMOVA, we estimate the amount of genetic variation at different hierarchical levels attributable to genetic differentiation among predefined groups (regions) (ΦCT), among populations within groups (ΦSC) and among all populations relative to the total sample (ΦST).

[b] groups == regions

[c] significance values: ** $p < 0.001$, *** $p < 0.0001$

[d] Model explanation:

A = 1 group: all in one;

B = 2 groups: Mesoamerica vs. East Andes (RBLT, GuLL, GuRB, GuCSG, HnPb, CrCNP, NiBa, CrLS, BCI, CoCho, CoNor, CoA) vs. (CoBo, PePH);

C = 4 groups: (RBLT, GuLL, GuRB) vs. (GuCSG, HnPb, CrCNP) vs. (NiBa, CrLS, BCI, CoCho, CoNor, CoA) vs. (CoBo, PePH);

D = 5 groups (RBLT, GuLL, GuRB) vs. (GuCSG, HnPb, CrCNP) vs. (NiBa, CrLS, BCI, CoCho, CoNor, CoA) vs. (CoBo) vs. (PePH).

[*] Abbreviations of area codes follow Table 1.
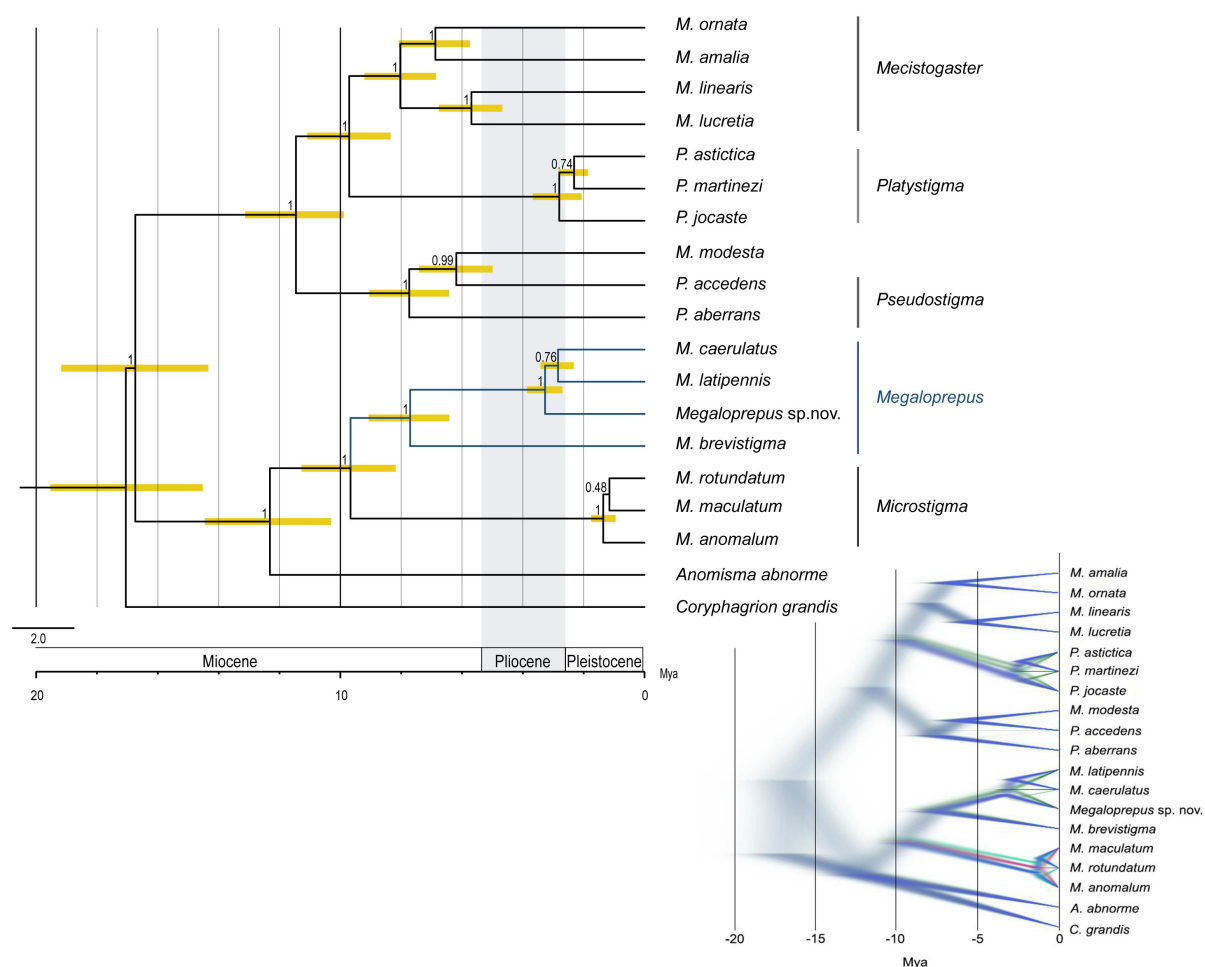
Figure 3: Time-calibrated MCC (Bayesian maximum clade credibility) tree with posterior mean node ages estimated by a lognormal relaxed molecular clock in BEAST vers. 1.8.4 (Drummond, et al. 2012) for the Pseudostigmatidae using *C. grandis* as an outgroup. Estimations are based on a concatenated matrix of four mitochondrial sequence marker (CO1/ND1 and 16S/12S) and two nuclear markers (28S/Ef1α). The branch lengths are scaled in time and the Bayesian posterior probabilities are depicted at each node. Node bars display the 95% HPD interval for that estimated node age.

The smaller tree on the right side shows the Densi.Tree illustration for the estimated node ages. Here the uncertainties among *Platystigma*, *Microstigma* and *Megaloprepus* are depicted.

Morphometrics

The linear morphometrics differentiated *M. brevistigma*, *M. caerulatus* and *M. latipennis* from each other. All three analyses (cluster, PCA, CVA) obtained congruent results (Fig. 4a, b). Only *Megaloprepus* sp. nov. shows an overlap with *M. latipennis* in the cluster analysis and the PCA (Fig. A2.3.1), but the CVA separated all four species. Thereby the highest variety among species was in wing width, width of the blue wing band in relation to the total wing length and in the pterostigma. Hereby, the prediction accuracy of correct grouping in the discriminant analyses was 100% (FW: 100%) indicating a very precise classification and the leave-one-out cross-validation also revealed 98.48% (FW: 95.45%) of correctly

identified specimens with only one *M. caerulatus* outlier that grouped into *M. latipennis* (Table A2.3.4). The accompanying MANOVA revealed significant ($p \leq 0.001$) separation between all four groups. To evaluate additional intragroup variation, comparisons were conducted between four *M. caerulatus* populations: Costa Rica (CrLS), Panama (BCI), Nicaragua (NiBa), and Colombia (Co). None of the analysis (cluster, PCA, CVA, MANOVA) obtained a strict separation between populations (results not shown).
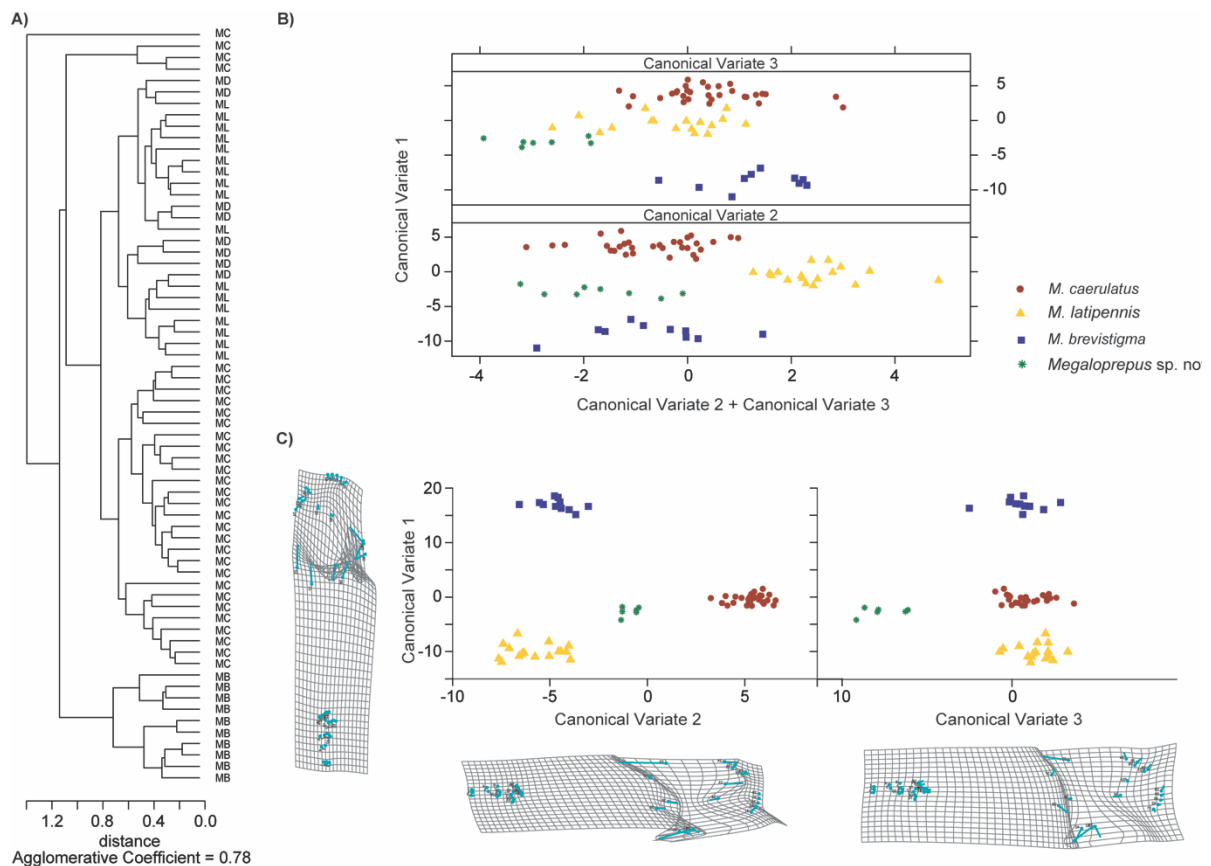
Figure 4: Analysis of morphological similarity between the putative *Megaloprepus* species based on measurements of 28 morphometric variables and the geometric morphometric landmarks analyzed in a cluster analysis and canonical variant analyses (CVA). Mapped are the results for the hind wings. (A) In the cluster dendrogram *M. brevistigma* is placed as a sister clade to the three remaining *Megaloprepus* species, while *M. latipennis* and *Megaloprepus* sp. nov. are clustered within one clade. (B) The three main axes of the CVA for the measurement comparisons show a complete separation of all four species. The CV1 (canonical variate) owns the highest proportion of variance (HW: CV1 = 84.97%, CV2 = 9.85%, CV3 = 5.18%; FW: CV1 = 89.27%, CV2 = 9.12%, CV3 = 1.60%). (C) A similar pattern was observed in the GGM analysis. The Scatter plots displaying the three canonical variates axes: CV1 horizontal versus CV2 and CV3 vertical (CV1 = 72.38%, CV2 = 22.95%, CV3 = 4.66%) illustrate the relative position of the investigated groups to each other (Eigenvalues: CV1 = 89.53, CV2 = 28.39, CV3 = 5.77). On each CV-axis the corresponding transformation grid illustrates the change in shape along the CV's. On each landmark point a vector indicates the orientation and the magnitude of variation with considerable differences in shape at the wing tips.
*Aberrations are: MB = *M. brevistigma*, MC = *M. caerulatus*, ML = *M. latipennis*, MD = *Megaloprepus* sp. nov..

Our GMM results are in concordance with the linear morphometrics, despite we had to reduce the data set for 6 specimens that showed damages at important congruent wing venation points due to the long storage in museums. Differences in wing shape were measured from 62 *Megaloprepus*: *M. brevistigma* (N = 11), *M. caerulatus* (N = 29), *M. latipennis* (N = 16) and *Megaloprepus* sp. nov. (N = 6). Superimposed landmarks overlap at the wing basis, but at the wing tips considerable LM clouds indicate higher shape differences. The PCA required seven principal components to cover over 90% of total variation whereat the first three PC's explain cumulative 74.82%. Nevertheless, a separation of groups was obtained and *M. brevistigma* was most distant to all other (Fig. A2.3.1). The same was observed in the CVA (Fig. 4c). Moreover, pairwise group comparisons for both Mahalanobis and Procrustes distances are highly significant (p < 0.001) (Table 4), which is further supported by the NPMANOVA (F = 20.42, p < 0.001). However, the DA revealed only partly explicit results between groups. While no misclassification of *M. brevistigma* was found, pairwise comparisons between *M. caerulatus*, *M. latipennis* and *Megeloprepus* sp. nov. could not allocate all individuals into its correct group (Table A2.3.5). Finally, three thin-plate-spline transformation grids are illustrating the variation in shape explained by the CV-axes (Fig. 4c).

We exemplary repeated the analysis with five selected forewings of each species, which confirmed the obtained results of the hind wings (results not shown). Despite of the wing morphology and general size, other taxonomic characters were found showing differences between species. The shape and coloration of the prothorax, the coloration of the mesothorax, and the male ligula and appendices are consistent for the distinct species (cf. species description Feindt and Hadrys submitted).

Table 4: Distances of shape (Procrustes and Mahalanobis distances and F-value) between species means based upon Euclidean distances.
Significance values are: * p < 0.001, ** p < 0.0001.

Pairwise taxons	Procrustes distances	Mahalanobis distances	F-statistics
M. brevistigma - *M. caerulatus*	0.08**	21.88**	40.77
M. brevistigma - *M. latipennis*	0.08**	27.77**	14.66
M. brevistigma – *Megaloprepus* sp. nov.	0.06**	22.08**	57.96
M. caerulatus - *M. latipennis*	0.04**	14.17**	4.62
M. caerulatus – *Megaloprepus* sp. nov.	0.03**	9.85**	12.32
M. latipennis - *Megaloprepus* sp. nov.	0.03*	11.54**	4.38

Species distribution modeling
Different model contributions the bioclimatic variables were observed: the highest impact was in most models the annual precipitation (BIO 12) followed by either precipitation of the driest quarter (BIO 17) or mean temperature of driest quarter (BIO 9). Although our modeling could be improved by including less locally restricted GPS occurrence points and remote sensing data on tree size and tree species to for example consider water filled tree

holes per grid, the MaxEnt predictions represent reliable results with high AUC values ≥ 0.95 (Table A2.4.1- A2.4.2).

The distribution modeled for the entire genus (Fig. 5, Fig. A2.4.2 & A2.4.3) shows a significant extrapolation into un-sampled environmental space in tropical South America. This includes South Venezuela, the Coastlines of French Guiana and northern Brazil, the northern regions of Colombia framing the three final slopes of the Andes (Cordillera Occidental, Cordillera Central and Cordillera Oriental) as well as the east and west sides of the Andes as far south as Bolivia. In Mesoamerica a high prevalence was detected in the Caribbean of South Mexico passing north of the Sierra Madre de Chiapas towards the Caribbean Lowlands of Guatemala at the Nature Reserve Cerro San Gil generating a possible contact zone between *M. latipennis* and *Megaloprepus* sp. nov. The southern slopes of the Sierra Madre de Chiapas mountain range in Guatemala and Honduras do not appear as very suitable, however most regions of Nicaragua, Costa Rica and Panama are. In lower Central America the Talamanca Cordillera (Chorotega volcanic front) displays a significant barrier between the Pacific and Atlantic. Interestingly, although *Megaloprepus* is known from the western Andes in Peru the model did not identify suitable habitats in those areas.

Potential distributions predicted to Mid Holocene, LGM and LIG showed high similarities for the LIG and Mid Holocene (Fig. 5), but in the Last Glacial Maximum a considerable expansion of the ecological niche was observed. *Megaloprepus* has a wide predicted distribution in the Amazon basin including today Guiana's and broad areas of the eastern slopes of the Andes. Furthermore, a high occurrence was observed in the Chocó region (including southern Panama Darién, northwestern coast of Colombia) and the Pacific slope of the Andes from Colombia to Peru. A continuous Caribbean corridor connecting all recent Mesoamerica populations is allowing broad overlaps between species (e.g. south of the Cordillera Volcánica Central in Costa Rica). In the last inter-glacial and during mid Holocene the potential distribution of *Megaloprepus* appear similar but more restricted and smaller (Fig. 5).

Our single distribution models showed for each species a similar occurrence map as for the genus, but with individual small differences (Fig. 6). Similarities are the exclusion of high mountain ranges. The slopes, however, appear suitable, as humidity is usually high in those regions. Our tests for niche similarity showed in all pairwise comparisons a considerable overlap of ecological niches. The background similarity test furthermore indicates niche conservatism (i.e. niches are more similar than expected based on their background). However, both indices *D* and *I* didn't obtain significant results. Only the *M. caerulatus* vs. *M. latipennis* comparison obtained a significant *I*–value indicating a decreased niche similarity (Table A2.4.3- A2.4.5, Fig. A2.4.3- A2.4.4).
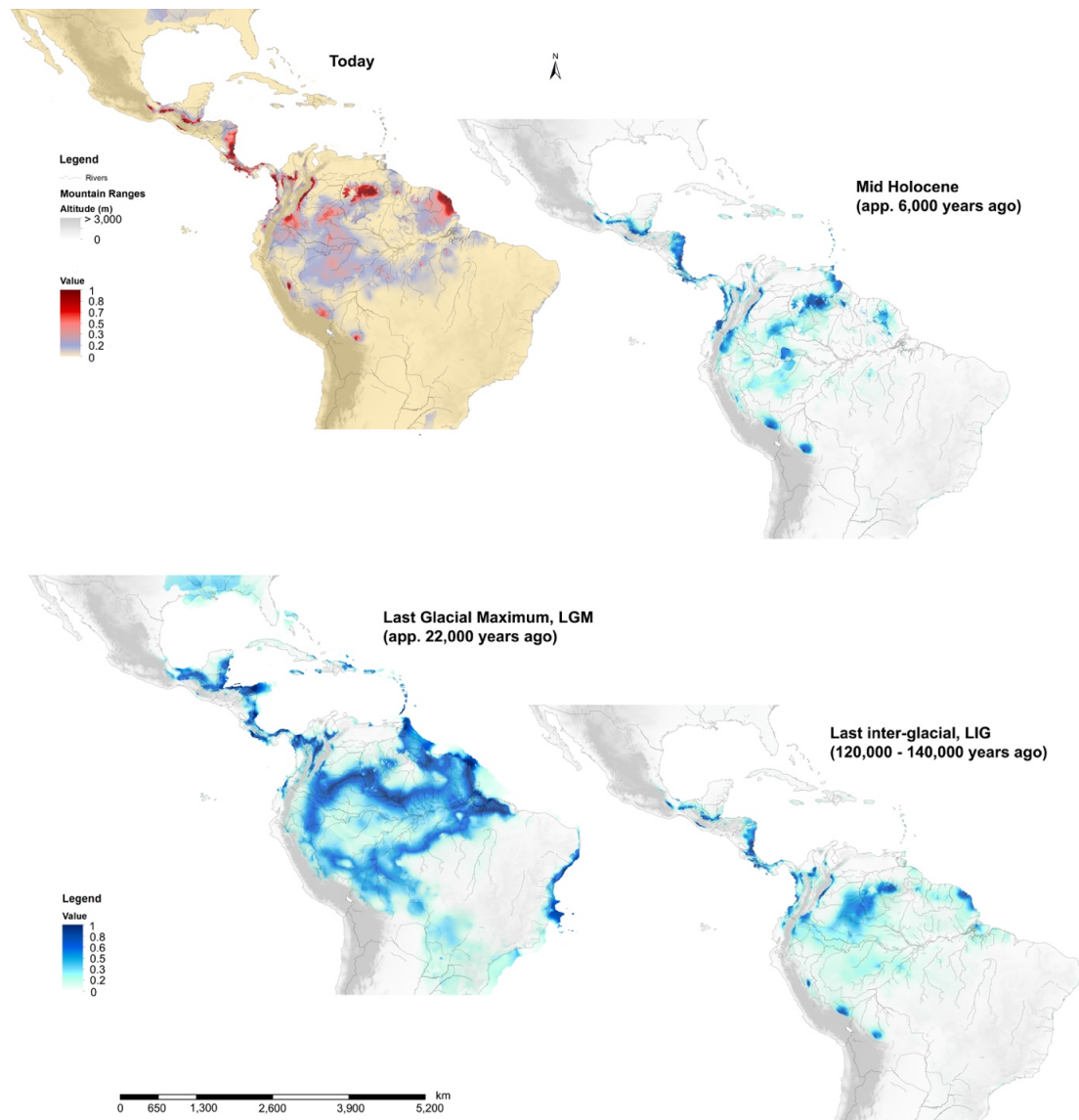
Figure 5: Predicted potential distribution of the genus *Megaloprepus* as generated in MaxEnt vers. 3.3.3k (Phillips, et al. 2006; Elith, et al. 2011) for today, Mid Holocene, Last Glacial Maximum and Last inter-glacial (Hijmans, et al. 2005; Otto-Bliesner, et al. 2006). All time frames were modelled using the entire coordinates from the genus (please compare File A2.4). In the hind casting the same most suitable variables as in the current model were used. The probability of occurrence is shown from 20% in light blue to 100% in dark red in the current distribution and in shades of blue for the past distribution.
*Maps were generated in ESRI ArcGIS 10.3.1. (http://www.esri.com/software/arcgis) and include mountain ranges, which were used from the GTOPO30, a global digital elevation model (DEM). The source is the U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, South Dakota. https://creativecommons.org/licenses/by-nc/3.0/.
**To project the potential past distribution of the genus during the LGM, we additionally compared three paleoclimatic data sets: the Community Climate System Model Version 4 (CCSM4, Gent, et al. 2011; Brady, et al. 2013), MPI-ESM-P (Max Planck Institute for Meteorology) and the Model for Interdisciplinary Research on Climate–Earth System Model (MIROC-ESM, Sueyoshi, et al. 2013), with all three data sets having almost identical results. Therefore, we only show the MPI-ESM-P.

Discussion

Megaloprepus species delimitation

The overall genetics point towards a past speciation event in the genus *Megaloprepus* and unambiguously indicate that the four previously defined groups shall be considered as true species. Consequently, the large distributional range of the nominal species *M. caerulatus* narrows to the Caribbean Coast of Nicaragua and Costa Rica, Panama and the Chocó-Darién. The remaining areas of the original distribution are now occupied by the three other species. *Megaloprepus brevistigma* was considered as subspecies of *M. caerulatus* in recent literature (Steinmann 1997) but as truly existing by Ris (1916). Now it occurs at the east side of the Andes from Colombia to Peru. From southern Mexico and northern Guatemala as well as Pacific Lowlands in Guatemala, we could verify *M. latipennis* – where the good species state appeared unlikely before (cf. Calvert 1901-1908). The fourth and strictly speaking new species *Megaloprepus* sp. nov. inhabits old growth forests at the Pacific coast of Costa Rica and the Caribbean of Guatemala and Honduras.

Although in concordance with the mitochondrial DNA, the nuclear DNA displays a less explicit picture. It shall be considered that we only investigated two nuclear genes representing approximately 0.00015% of an average odonate genome. Although ITS I+II and Ef1α are common markers in odonate research other genomic regions could be more diverged. However, the discrepancy concerning genetic differentiation between nuclear and mitochondrial gene data has got more attention in recent phylogeographic research (Toews and Brelsford 2012; Morales, et al. 2015; Figueiró, et al. 2017). Incomplete lineage sorting (ILS) and introgression have been associated as main causes (Toews and Brelsford 2012). Our results lead to the assumption that historical gene flow may be the reason in *Megaloprepus*. Temperature and precipitation as well as habitat changes during Pleistocene most likely caused potential overlapping distributional ranges among the different *Megaloprepus* species in Mesoamerica (c.f. Hooghiemstra and van der Hammen 1998; Bagley and Johnson 2014; Smith, et al. 2014), which is supported by the past distribution model (please compare demographic history). Furthermore, the topological uncertainties observed in the Densi.Tree favor a past interspecific gene flow (Fig. A2.2.4).

Another crucial factor supporting the four species explanation and playing a general role in speciation is the OXPHOS complex. Genes involved in oxidative phosphorylation (OXPHOS) enable cellular respiration, whereby the nuclear and mitochondrial genes encode mutually for electron transport chain proteins (Morales, et al. 2015; Sunnucks, et al. 2017). This infers that for organismal function the mito-nuclear compatibility is essential especially in energy consuming conditions (e.g. migration, mating, hunting and territorialism). In Odonata flight is the main energy demanding behavior and incomparability in the electron transport could lead to fitness disadvantages. For example, in *Nasonia* wasps a high mortality of hybrids was observed and related to incomparability between a nuclear-encoded OXPHOS gene (NADH dehydrogenase) and the mitochondria (Gibson, et al. 2013). Furthermore, the *Nasonia* genome working group showed an increased ratio of synonymous-to-nonsynonymous substitutions (*dN/dS*) in nuclear genes that cooperate with mitochondrial ribosomes and the OXPHOS complexes I and V (Werren, et al. 2010).

Between our *Megaloprepus* species we found in average 10% of genetic distances in the 13 mitochondrial-encoded OXPHOS genes (Feindt *et al.* in prep.). Aside of simple genetic drift, in allopatric occurring populations under slightly different micro-climatic conditions, a positive selection for the products of mitochondrial genes is still likely. Because co-adaptation of mitochondrial and nuclear OXPHOS genes is critical for survival for example in isolated populations, mito-nuclear incompatibilities among species or between populations of the same species would present a barrier to hybridization (cf. Hill 2016).

Demographic history

Megaloprepus' geographic patterns have similarities to other vicariant events in the Neotropics. The Isthmus of Tehuantepec dividing Sierra Madre Sur from Chiapas and Guatemala, the Nicaraguan depression isolating the Talamanca mountain range from lower Central America and the Andes are recognized as speciation zones on a large scale (e.g. Castoe, et al. 2009; Morrone 2014; Smith, et al. 2014; Hazzi, et al. 2018). However, only in lower Central America 31 phylogeographical breaks were identified in a recent review by Bagley and Johnson (2014). Phylogeographic studies of various Neotropical plant and animal species underlined repeatedly that diversification is triggered by the complex geological history and the change of climatic niches over time (e.g. Pennington, et al. 2004; Wang, et al. 2008; González, et al. 2011; Condamine, et al. 2012; Villalobos 2013). The accordance of the biogeographic barriers acting as diversification zones points towards the fact that similar evolutionary forces may have shaped biological and genetic diversity in the Neotropics (Bagley and Johnson 2014; Antonelli, et al. 2018). However, despite all these similarities, the complexity of this region also demands considering species specific ecological and evolutionary patterns as a key for precise conclusions and to reveal why reduced genetic connectivity in regions of speciation arose over time in a certain species or species group such as the tree hole breeders.

Time calibrated phylogenies can connect diversification with geological events, providing additional explanations for the species split and current distributions in *Megaloprepus*. But unfortunately, odonate research in the new world tropics is still underrepresented concerning the completeness of species lists and genetic studies, so that future molecular clocks will have to prove our results with a more complete sampling. However, the dated species splits in the relaxed clock and the fossil calibration reflect the geographic history of the Neotropics. Ancestors of the genus *Megaloprepus* were most likely distributed in South America. The diversification between today *M. brevistigma* and the most recent common ancestor of the three Mesoamerican *Megaloprepus* species of ~10-8 Mya is consistent with the uplift of the Andes (i.e. the Colombian Andes) during Late Miocene – Pliocene reaching half of its present elevation (~2,300 m) about 10 Mya (Gregory-Wodzicki 2000; Hoorn, et al. 2010). This mountain orogeny was accompanied by changes of local climates on the emerging mountain peaks (Gregory-Wodzicki 2000) and increasing rain at the eastern side of the Andes (Hoorn, et al. 2010). Both, dryer microclimatic conditions on the mountain tips and increasing geographic distances are

barriers to gene flow and could have caused the early split between eastern and western Andes *Megaloprepus*. Equally the stable and humid climate east of the Andes (i.e. in the Chocó region) during Quaternary (Hooghiemstra and van der Hammen 1998; Bagley and Johnson 2014) supported the northward migration of *Megaloprepus* after the closure of the Isthmus of Panama ~3 Mya (O'Dea, et al. 2016), but see Bacon *et al.* (2015). This migration was most likely unhindered, because moist forests in Central America apparently established before the final closure of the land bridge (Bagley and Johnson 2014). However, today's dry forests as potential barriers most likely speciated at a similar time period (Pennington, et al. 2004; Wang, et al. 2008; De-Nova, et al. 2012) as the Mesoamerican *Megaloprepus* diversified, which was 'shortly' after the closure of the Isthmus ~2 Mya. In summary, the diversification of the three Mesoamerican *Megaloprepus* species can be explained by northward migration pattern and geographic barriers such as dry forests, mountain ranges, Nicaraguan depression, and tectonic plate borders (e.g. Bagley and Johnson 2014) such as the Motagua-Polochic fault system – separating the Maya and Chortis Block in Central Guatemala (Castoe, et al. 2009; Rovito, et al. 2012).

However, still two struggling facts are (i) the 'unusual' disjunct distribution of *Megaloprepus* sp. nov. and (ii) the potential overlapping occurrences during Pleistocene among the Mesoamerican species. Its reason may be justified by the climatic variations during the Pleistocene. The stadial–interstadial temperature and rainfall changes are associated with north-south shifts of the Intertropical Convergence Zone, ITCZ (Hodell, et al. 2008), and have influenced animal migrations and reorganized tropical biota (e.g. Hewitt 2000). Consequently, an expansion of suitable forest habitats for *Megaloprepus* during glacial periods (cf. Hodell, et al. 2008; Mays, et al. 2017) could have caused overlapping occurrences as well as a second migration event allowing *Megaloprepus* sp. nov. to reach northern areas such as the Caribbean Coast in Honduras and Guatemala. In the hind casted species distribution models (Fig. 5) we observed similar distributions as today in the Interglacial period (LIG) and Mid Holocene, but an expansion of potential suitable habitats about 22,000 years BP. In theory glacial advances caused forest migration to lower elevations and most likely an expansion of open areas (Mays, et al. 2017). However, strict dry and wet conditions were not synchronous during glacial periods and probably also depended on the geographical location. More importantly the deglacial period after the LGM (18,000 – 11,000 years BP) appeared drier than the LGM itself (Bush, et al. 2009) and therefore rainforests were perhaps expanded 22,000 years BP (Hewitt 2000; Mayle, et al. 2004; Leite, et al. 2016).

Field trials with *M. caerulatus* revealed a low physiological tolerance to warm temperatures (Feindt *et al.* unpubl.) and impossible cultivation in green houses or insectaries, which had more sunlight, were warmer and dryer than the forest habitats. Today *Megaloprepus* occurs at elevations up to 2,500 m and rigorously depends on a closed canopy that retain relatively constant colder temperatures than for example selectively logged or young secondary forests (Fincke 1998). But also, a high precipitation is vital for larval survival, especially when considering that only a fraction of the rain reaches the tree holes and high temperatures can cause evaporation. In summary, wetter and relatively colder

temperatures positively affect *Megaloprepus* and this supports a wider distribution 22,000 years BP during colder but not extreme dry conditions.

Niche conservatism and morphological innovation

The SDM using current climatic variables revealed similar distributions for the three Mesoamerican species. Hereby one species predicted the distribution of its sister species to most of it extends. Additionally, the niche equivalency and background similarity tests validated the impression that the ecological niches are identical (Fig. 6, File A2.4). Consequently, at least the Mesoamerican species retained their ancestral niche over time and because *Megaloprepus* differentiated without adapting to new breeding habitats or changes in species interactions (pers. observation), the speciation mode is indeed non-adaptive.

It has to be considered that the water filled tree holes are unique and represent an exclusive habitat for several arthropods in forest sites. Rainfall and temperature apparently influence tree holes the most. For example, if in one forest site rainfall decreases and temperatures increase (also through edge effects) water chemistry would then change inside the tree holes or tree holes could fall dry. This would decrease survival rates and in an extreme scenario forests could lose important key stone species. The effects of variation in the water chemistry on *Megaloprepus* haven't been tested yet, but it could influence developmental rates, hatching success or adult body size.

High morphological similarities between *M. caerulatus*, *M. latipennis* and *Megaloprepus* sp. nov., which are more similar among each other than in comparison to *M. brevistigma*, mirror a slow trait divergence in the face of PNC. However, despite using a mixture of museum samples and newly collected specimens, the morphometric analyses differentiated the four species (Fig. 3, Table 4). Disparities in small and mostly obscure taxonomic characters may have been driven by sexual selection or learned sexual preferences (cf. Padial, et al. 2010; Svensson 2012; Wellenreuther, et al. 2012). Although on a broad scale the ecological niches are equal, there are little differences in local micro-conditions including light and seasonality, which is not observed by the SDM. As stated above, wing shape and wing coloration are the most striking variations among the *Megaloprepus* species. In Odonata as highly visible insects, different color pattern connected to wing size and their UV reflection are important secondary sexual traits (e.g. Schultz and Fincke 2009; Xu and Fincke 2015; Outomuro, et al. 2016). These traits influence mate choice, which consequently could lead to a directed selection (e.g. Svensson and Waller 2013; Guillermo-Ferreira, et al. 2014). But wings play also major roles in flight performance and endurance (e.g. Outomuro, et al. 2016; Suárez-Tovar and Sarmiento 2016). *Megaloprepus* usually occur in the understory of old growth forests and feed from spider webs wherefore they need good flight abilities. Consequently, broader wings could have increased flight performance and therefore success in territorial fights or feeding (e.g. Outomuro, et al. 2016; Suárez-Tovar and Sarmiento 2016).

In the Caribbean coast of Central America and the Chocó-Darién rainfall occurs throughout the whole year with drier months between February and April (McDade, et al.

1994; Matlock and Hartshorn 1999). In the southern Pacific Coast of Costa Rica and in the Los Tuxtlas region in Mexico two strong dry seasons occur from mid-December to mid-April (Wieder and Wright 1995; Sanchez-Azofeifa, et al. 2002) and from March to May (Bongers, et al. 1988). Because of these seasonal differences, a higher cloudiness in the Chocó-Darién and the Lower Central American Caribbean coast could have influenced light conditions within the forest and therefore the visibility of potential mates. An advantage under bad light conditions would be for example a broader wing with a bigger wing patch for a higher UV-reflectance. Indeed, *M. caerulatus* is the only species of its genus with a sexual dimorphism and the largest blue wing patch. In the genus *Megaloprepus* the original state was most likely no dimorphism, because *M. caerulatus* appears as the most derived species and all other do not show sexually dimorphic wing characters (cf. Figs. 3, A2.2.1-A2.2.5). This dimorphism could have developed as a natural variation which became a selective trait over time to better differentiate a male from a female and now counts as a barrier to reproduction (see the field experiments by Schultz and Fincke 2009; Xu and Fincke 2015). This is supported by Adams and colleagues (2009), which disproved a correlation between genetic divergence and morphological evolution. Consequently, phenotypic novelties are mostly related to ecological variation and sexual selection. But morphological trait evolution due to sexual selection is particularly successful when morphological signals are related to environmental conditions (Langerhans and Riesch 2013; Servedio and Boughman 2017). This is e.g. well studied in *Heliconius* butterflies, where wing coloration is linked to ecological adaptation and mating success (e.g. Jiggins 2008 and references herein).

In summary, although *Megaloprepus* speciated non-adaptively and the sister species exhibit a PNC, morphological novelties can occur without changing the ecological niche. Thus, a strict separation of speciation modes appears inconvenient and supports the process-based perspective of PNC (cf. Pyron, et al. 2015). However, the speed in which novelties evolve is far from concrete prediction. By theory it depends for example on population sizes, bottlenecks, genetic drift or selection (i.e. balancing selection or selection pressures) but not least at the study specimen itself. 'Fast' morphological novelties are associated with environmental conditions (Kashtan, et al. 2007). For example, the rapid divergence of phenotypes in birds and mammals is explained by the cold climate during Cenozoic (Clavel and Morlon 2017) and speciation rates in marine fish are highest in polar regions (Rabosky, et al. 2018). But still more precise statements demand large-scale comparisons of speciation rates within odonates, but also studying the genetic background of phenotypic evolution.
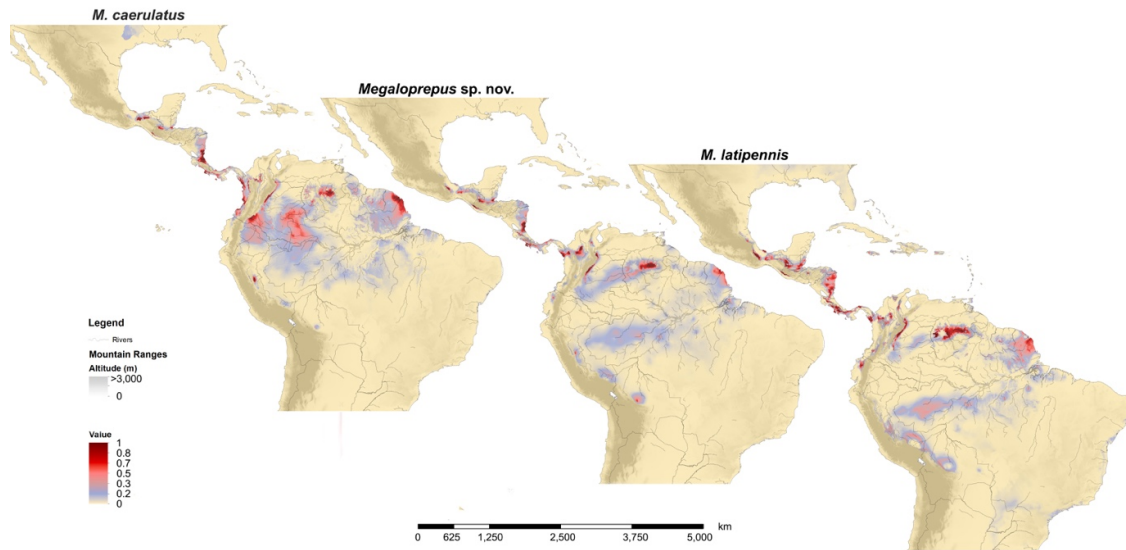
Figure 6: Potential distribution of *Megaloprepus* modelled in MaxEnt vers. 3.3.3k. (Phillips, et al. 2006; Elith, et al. 2011) according to the coordinates and climatic variables used (Hijmans, et al. 2005). The probability of occurrence is shown from 20% in light blue to 100% in dark red, whereat all three models sow a strong similarity in the distribution (please compare File A2.4).

*The map was generated in ESRI ArcGIS 10.3.1. (http://www.esri.com/software/arcgis) and include mountain ranges, which were used from the GTOPO30, a global digital elevation model (DEM). The source is the U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, South Dakota. https://creativecommons.org/licenses/by-nc/3.0/.

Future of (non-adaptive) speciation research in Odonata

Non-adaptive speciation is probably common in odonates (Svensson 2012), especially in damselflies when low flight endurance is combined with strong natural selection for niche constrains. But also, odonate species with wide ecological niches and higher capabilities for niche shifts such as *Crocothemis erythraea* (Sánchez-Guillén and Ott 2018) or *Coenagrion scitulum* (Swaegers, et al. 2015) are known. So far speciation research in odonates is limited, although genomics is advancing in this order. Future work could contribute to the general knowledge of speciation trough identifying trait loci or genes related to development or phenotypic pattern. However, non-adaptive and adaptive speciation appear not strictly separable rather as a continuum, where allopatric non-adaptive speciation could represent an initial step that might be followed by adaptation to e.g. changing habitats (e.g. Svensson 2012; Pyron, et al. 2015).

Megaloprepus could be one good model for such future research for forest odonates or highly specialized tropical insects, because of the combination of natural history traits like natural selection for niche constrains, the median age of the genus and the low capability to migrate between forest patches. Large-scale genome- or transcriptome-wide SNP data for all four species could reveal demographic patterns, micro-evolutionary changes and the genomic regions under positive selection in response to different environments.

M. brevistigma East of the Andes

Despite the clear species differentiation in Mesoamerica, *M. brevistigma* in South America shows an obscure picture. We only obtained two tissue samples from very distant populations of about 1,500 km between Colombia and southern Peru. Although equally high genetic distances to the Mesoamerican species were detected, a considerable genetic differentiation in ND1 (6.13%) and CO1 (5.16%) and a complete isolation (F_{ST} = 1) was found between those samples (Table 1, File A2.1). These facts point towards another complex evolutionary history at the east side of the Andes.

Recently, 15 biogeographical regions and 26 isolation events were described from the tropical Andes (Hazzi, et al. 2018). The most important is the Marañon River Valley separating the central from the northern Andes (Hazzi, et al. 2018). During Pleistocene Northern South American and Central Andes have experienced a disruption of continuous forest cover (e.g. Baker and Fritz 2015; Hazzi, et al. 2018; Winkler, et al. 2018), which may have caused a long-term separation of suitable habitats for *M. brevistigma.* However, due to our limited genetic sample size and the high homogeneity in the morphological comparisons, we consider *M. brevistigma* as one species in the most conservative way. But a future survey throughout the entire possible distribution of this species (cf. Fig. 6) could give additional insights into diversification pattern and forest movements on the East side of the Andes.

Conclusion

Speciation cannot be drawn in a simplistic way, and adaptive and non-adaptive speciation may occur successively or intertwined so that non-adaptive speciation may be completed by adaptation or *vice versa* (cf. Pyron, et al. 2015). Lacking either the need or the ability to adapt can consequently minimize morphological heterogeneity and cause cryptic speciation. Hidden species are often unnoticed by scientists as it has happened in *Megaloprepus*.

Megaloprepus fulfills some of the points that can cause non-adaptive speciation: conserved ecological niches and a lack of gene flow among lineages as well as time to speciate. Other causes such as pleiotropy and little genetic variability on niche important traits shall be proved in a comparative genomic study. With the current work, we were able to present a complex integrative study of the genus *Megaloprepus* throughout its whole distributional range and prove speciation in the genus. The new sister species show a phylogenetic niche conservatism, but also sexual selection has an influence on the morphology. These results demand conservation actions and higher support for taxonomic and genetic studies on the Neotropical arthropod fauna. Hereby the present study could serve as an example when cryptic speciation is assumed, and individuals are limited. But it also shows that Neotropical forest odonates may contribute significantly to the general understanding of speciation in the tropics.

Acknowledgements

References

Abraham D, Ryrholm N, Wittzell H, Holloway JD, Scoble MJ, Löfstedt C. 2001. Molecular phylogeny of the subfamilies in Geometridae (Geometroidea: Lepidoptera). Molecular Phylogenetics and Evolution 20:65-77.

Abramoff MD, Magalhaes PJ, Ram SJ. 2004. Image Processing with ImageJ. Biophotonics international 11:36-42.

Adams DC, Berns CM, Kozak KH, Wiens JJ. 2009. Are rates of species diversification correlated with rates of morphological evolution? Proceedings of the Royal Society of London B: Biological Sciences:rspb. 2009.0543.

Adriaens D. 2007. Protocol for error testing in landmark based geometric morphometrics (protocol link: http://www.fun-morph.ugent.be/Miscel/Methodology/Morphometrics.pdf). In.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25:3389-3402.

Antonelli A, Zizka A, Carvalho FA, Scharn R, Bacon CD, Silvestro D, Condamine FL. 2018. Amazonia is the primary source of Neotropical biodiversity. Proceedings of the National Academy of Sciences:201713819.

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences 112:6110-6115.

Bagley JC, Johnson JB. 2014. Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. Biological Reviews 89:767-790.

Baker PA, Fritz SC. 2015. Nature and causes of Quaternary climate variation of tropical South America. Quaternary Science Reviews 124:31-47.

Basset Y, Hammond PM, Barrios H, Holloway JD, Miller SE. 2003. Vertical stratification of arthropod assemblages. Arthropods of tropical forests:17-27.

Bechly G. 2000. A new fossil damselfly species (Insecta: Odonata: Zygoptera: Coenagrionidae: Ischnurinae) from Dominican amber. Stuttgarter Beitr. Naturk. Ser. B 299:1-9.

Bongers F, Popma J, Del Castillo JM, Carabias J. 1988. Structure and floristic composition of the lowland rain forest of Los Tuxtlas, Mexico. Vegetatio 74:55-80.

Brady EC, Otto-Bliesner BL, Kay JE, Rosenbloom N. 2013. Sensitivity to glacial forcing in the CCSM4. Journal of Climate 26:1901-1925.

Brower A. 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences 91:6491-6495.

Bush MB, Correa-Metrio AY, Hodell DA, Brenner M, Anselmetti FS, Ariztegui D, Mueller AD, Curtis JH, Grzesik DA, Burton C. 2009. Re-evaluation of climate change in lowland Central America during the Last Glacial Maximum using new sediment cores from Lake Petén Itzá, Guatemala. In. Past Climate Variability in South America and Surrounding Regions: Springer. p. 113-128.

Butlin R, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo CRF, Diao W, Maan ME, Paolucci S, Weissing FJ. 2012. What do we need to know about speciation? Trends in Ecology & Evolution 27:27-39.

Callahan MS, McPeek MA. 2016. Multi-locus phylogeny and divergence time estimates of *Enallagma* damselflies (Odonata: Coenagrionidae). Molecular Phylogenetics and Evolution 94:182-195.

Calvert P. 1901-1908. Odonata. In: Goldman F, editor. Biologia Centrali-Americana. Vol. 50. Insecta, Neuroptera. p. 51-57.

Castoe TA, Daza JM, Smith EN, Sasa MM, Kuch U, Campbell JA, Chippindale PT, Parkinson CL. 2009. Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. Journal of biogeography 36:88-103.

Channan S, Collins K, Emanuel W. 2014. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA 30.

Clavel J, Morlon H. 2017. Accelerated body size evolution during cold climatic periods in the Cenozoic. Proceedings of the National Academy of Sciences 114:4183-4188.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. Molecular ecology 9 (10):1657-1660.

Cockerham CC, Weir B. 1993. Estimation of gene flow from F-statistics. Evolution 47:855-863.

Collins SD, McIntyre NE. 2015. Modeling the distribution of odonates: a review. Freshwater Science 34:1144-1158.

Condamine FL, Silva-Brandão KL, Kergoat GJ, Sperling FA. 2012. Biogeographic and diversification patterns of Neotropical Troidini butterflies (Papilionidae) support a museum model of diversity dynamics for Amazonia. BMC evolutionary biology 12:82.

Crisp MD, Cook LG. 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? New Phytologist 196:681-694.

Damm S, Dijkstra K-DB, Hadrys H. 2010. Red drifters and dark residents: the phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa's changing environment (Odonata, Libellulidae, Trithemis). Molecular Phylogenetics and Evolution 54:870-882.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature methods 9:772-772.

De-Nova JA, Medina R, Montero JC, Weeks A, Rosell JA, Olson ME, Eguiarte LE, Magallón S. 2012. Insights into the historical construction of species-rich Mesoamerican seasonally dry tropical forests: the diversification of *Bursera* (Burseraceae, Sapindales). New Phytologist 193:276-287.

Dijkstra K-DB, Kalkman VJ, Dow RA, Stokvis FR, Van Tol J. 2014. Redefining the damselfly families: a comprehensive molecular phylogeny of Zygoptera (Odonata). Systematic Entomology 39:68-96.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS biology 4:e88.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969-1973.

Drury D. 1782. Illustrations of natural history. London: White.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792-1797.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources 10:564-567.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Feindt W, Fincke O, Hadrys H. 2014. Still a one species genus? Strong genetic diversification in the world`s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*. Conservation genetics 15:469-481.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation 24:38-49.

Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SH, Hughes GM, Komissarov A, Antunes A. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Science advances 3:e1700299.

Fincke OM. 1992a. Consequences of larval ecology for territoriality and reproductive success of a Neotropical damselfly. Ecology 73:449–462.

Fincke OM. 1984. Giant damselflies in a tropical forest: reproductive biology of *Megaloprepus coerulatus* with notes on *Mecistogaster* (Zygoptera: Pseudostigmatidae). Advances in Odonatology 2:13-27.

Fincke OM. 1992b. Interspecific competition for tree holes: consequences for mating systems and coexistence in neotropical damselflies. American Naturalist:80-101.

Fincke OM. 1998. The population ecology of *Megaloprepus coerulatus* and its effect on species assemblages in water-filled tree holes. In. Insect populations in theory and in practice: Springer. p. 391-416.

Fincke OM editor. Forest and dragonflies. 4th WDA International Symposium of Odonatology. Pensoft, Sofia. 2006.

Fincke OM, Hedström I. 2008. Differences in forest use and colonization by Neotropical tree-hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conversion. Studies on Neotropical Fauna and Environment 43:35-45.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology 3:294-299.

Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote sensing of Environment 114:168-182.

Garrison RW, von Ellenrieder N, Louton JA. 2010. Damselfly Genera of the New World, an Illustrated and Annotated Key to the Zygoptera. Baltimore, MD: The Johns Hopkins University Press.

Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M. 2011. The community climate system model version 4. Journal of Climate 24:4973-4991.

Gernhard T. 2008. The conditioned reconstructed process. Journal of theoretical biology 253:769-778.

Gibson J, Niehuis O, Peirson B, Cash E, Gadau J. 2013. Genetic and developmental basis of F2 hybrid breakdown in Nasonia parasitoid wasps. Evolution 67:2124-2132.

Givnish TJ. 2015. Adaptive radiation versus 'radiation'and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. New Phytologist 207:297-303.

González C, Ornelas JF, Gutiérrez-Rodríguez C. 2011. Selection and geographic isolation influence hummingbird speciation: genetic, acoustic and morphological divergence in the wedge-tailed sabrewing (*Campylopterus curvipennis*). BMC evolutionary biology 11:38.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27:221-224.

Gregory-Wodzicki KM. 2000. Uplift history of the Central and Northern Andes: a review. Geological Society of America Bulletin 112:1091-1105.

Groeneveld LF, Clausnitzer V, Hadrys H. 2007. Convergent evolution of gigantism in damselflies of Africa and South America? Evidence from nuclear and mitochondrial sequence data. Molecular Phylogenetics and Evolution 42:339-346.

Guillermo-Ferreira R, Therézio EM, Gehlen MH, Bispo PC, Marletta A. 2014. The role of wing pigmentation, UV and fluorescence as signals in a Neotropical damselfly. Journal of insect behavior 27:67-80.

Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4:9.

Harvey PH, Pagel MD. 1991. The comparative method in evolutionary biology: Oxford university press Oxford.

Hazzi NA, Moreno JS, Ortiz-Movliav C, Palacio RD. 2018. Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. Proceedings of the National Academy of Sciences:201803908.

Hedström I, Sahlén G. 2001. A key to the adult Costa Rican "helicopter" damselflies (Odonata: Pseudostigmatidae) with notes on their phenology and life zone preferences. International Journal of Tropical Biology and Conservation 49:1037-1056.

Heibl C, Calenge C. 2013. phyloclim: integrating phylogenetics and climatic niche modeling. R package ver. 0.9-4: http://cran.r-project.org/package=phyloclim.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. International journal of climatology 25:1965-1978.

Hill GE. 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. Ecology and evolution 6:5831-5842.

Hodell DA, Anselmetti FS, Ariztegui D, Brenner M, Curtis JH, Gilli A, Grzesik DA, Guilderson TJ, Müller AD, Bush MB. 2008. An 85-ka record of climate change in lowland Central America. Quaternary Science Reviews 27:1152-1165.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nature Reviews Genetics 10:639-650.

Hooghiemstra H, van der Hammen T. 1998. Neogene and Quaternary development of the neotropical rain forest: the forest refugia hypothesis, and a literature overview. Earth-Science Reviews 44:147-183.

Hoorn C, Wesselingh F, Ter Steege H, Bermudez M, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson C, Figueiredo J. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330:927-931.

Hutohinson G. 1957. A treatise on limnology: New York.

Ingley SJ, Bybee SM, Tennessen KJ, Whiting MF, Branham MA. 2012. Life on the fly: phylogenetics and evolution of the helicopter damselflies (Odonata, Pseudostigmatidae). Zoologica Scripta 41:637-650.

Jiggins CD. 2008. Ecological speciation in mimetic butterflies. AIBS Bulletin 58:541-548.

Jordan S, Simon C, Polhemus D. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus Megalagrion (Odonata: Coenagrionidae). Systematic Biology 52:89-109.

Kashtan N, Noor E, Alon U. 2007. Varying environments can speed up evolution. Proceedings of the National Academy of Sciences 104:13711-13716.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Khazan ES. 2014. Tests of biological corridor efficacy for conservation of a Neotropical giant damselfly. Biological Conservation 177:117-125.

Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proceedings of the National Academy of Sciences 78:454-458.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution 16:111-120.

Kitching R. 1971. An ecological study of water-filled tree-holes and their position in the woodland ecosystem. The Journal of Animal Ecology:281-302.

Kitching RL. 2000. Food webs and container habitats: the natural history and ecology of phytotelmata: Cambridge University Press.

Klingenberg CP. 2011. MorphoJ: an integrated software package for geometric morphometrics. Molecular ecology resources 11:353-357.

Klingenberg CP, McIntyre GS. 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution:1363-1375.

Kozak KH, Wiens JJ. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60:2604-2621.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in bioinformatics 9:299-306.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874.

Langerhans RB, Riesch R. 2013. Speciation by selection: a framework for understanding ecology's role in speciation. Current Zoology 59:31-52.

Latham J, Cumani R, Rosati I, Bloise M. 2014. Global land cover share (GLC-SHARE) database beta-release version 1.0-2014. FAO: Rome, Italy.

Leite YL, Costa LP, Loss AC, Rocha RG, Batalha-Filho H, Bastos AC, Quaresma VS, Fagundes V, Paresque R, Passamani M. 2016. Neotropical forest expansion during the last glacial period challenges refuge hypothesis. Proceedings of the National Academy of Sciences 113:1008-1013.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Machado AB, Lacerda DSS. 2017. Revalidation of *Platystigma* Kennedy, 1920, with a synopsis of the *quadratum* species group and the description of three new species (Odonata: Pseudostigmatidae). Zootaxa 4242:493-516.

Matlock RB, Hartshorn GS. 1999. La Selva Biological Station (OTS). Bulletin of the Ecological Society of America:188-193.

Mayle FE, Beerling DJ, Gosling WD, Bush MB. 2004. Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. Philosophical Transactions of the Royal Society of London B: Biological Sciences 359:499-514.

Mays JL, Brenner M, Curtis JH, Curtis KV, Hodell DA, Correa-Metrio A, Escobar J, Dutton AL, Zimmerman AR, Guilderson TP. 2017. Stable carbon isotopes (δ13C) of total organic carbon and long-chain n-alkanes as proxies for climate and environmental change in a sediment core from Lake Petén-Itzá, Guatemala. Journal of Paleolimnology 57:307-319.

McDade LA, Bawa KS, Hespenheide HA, Hartshorn GS. 1994. La Selva: ecology and natural history of a Neotropical rain forest: University of Chicago Press.

McPeek MA, Symes LB, Zong DM, McPeek CL. 2011. Species recognition and patterns of population variation in the reproductive structures of a damselfly genus. Evolution 65:419-428.

Mitchell T. 2007. The EDNA fossil insect database. See http://edna/. palass-hosting. org.

Mitteroecker P, Gunz P, Windhager S, Schaefer K. 2013. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. Hystrix, the Italian Journal of Mammalogy 24:59-66.

Morales HE, Pavlova A, Joseph L, Sunnucks P. 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Molecular ecology 24:2820-2837.

Morrone JJ. 2014. Cladistic biogeography of the Neotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics 30:202-214.

Norris D. 2014. Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil. Tropical Conservation Science 7:529-547.

Nosil P. 2012. Ecological speciation: Oxford University Press.

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD. 2016. Formation of the Isthmus of Panama. Science advances 2:e1600883.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A. 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. Science 311:1751-1753.

Outomuro D, Söderquist L, Nilsson-Örtman V, Cortázar-Chinarro M, Lundgren C, Johansson F. 2016. Antagonistic natural and sexual selection on wing shape in a scrambling damselfly. Evolution 70:1582-1595.

Padial JM, Miralles A, De la Riva I, Vences M. 2010. Review: The integrative future of taxonomy. Frontiers in Zoollogy 7:1-14.

Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Molecular Biology and Evolution 27:1659-1672.

Pearman PB, Guisan A, Broennimann O, Randin CF. 2008. Niche dynamics in space and time. Trends in Ecology & Evolution 23:149-158.

Pennington RT, Lavin M, Prado DE, Pendry CA, Pell SK, Butterworth CA. 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. Philosophical Transactions of the Royal Society of London B: Biological Sciences 359:515-538.

Peterson A, Soberón J, Sánchez-Cordero V. 1999. Conservatism of ecological niches in evolutionary time. Science 285:1265-1267.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. Ecological niches and geographic distributions (MPB-49): Princeton University Press.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190:231-259.

Posada D. 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25:1253-1256.

Posada D. 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. Current protocols in bioinformatics:6.5. 1-6.5. 14.

Proosdij AS, Sosef MS, Wieringa JJ, Raes N. 2015. Minimum required number of specimen records to develop accurate species distribution models. Ecography.

Pyron RA, Costa GC, Patten MA, Burbrink FT. 2015. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. Biological Reviews 90:1248-1262.

Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559:392.

Rambaut A, Drummond A. 2015. FigTree, ver. 1.4. 2. Available: http:/tree. bio. ed. ac. uk/software/figtree/Accessed on 28.

RCoreTeam. 2014. R: A language and environment for statistical computing. Vienna, Austria: URL http://www.r-project.org/.

Ricklefs RE, Latham RE. 1992. Intercontinental correlation of geographical ranges suggests stasis in ecological traits of relict genera of temperate perennial herbs. The American Naturalist 139:1305-1321.

Ris F. 1916. Libellen (Odonata) aus der Region der amerikanischen Kordilleren von Costarica bis Catamarca. Archiv für Naturgeschichte 82A:1-197 + 192 pl.

Rodriguez F, Oliver J, Marin A, Medina JR. 1990. The general stochastic model of nucleotide substitution. Journal of theoretical biology 142:485-501.

Roehrdanz RL. 1997. Identification of tobacco budworm and corn earworm (Lepidoptera: Noctuidae) during early developmental stages by polymerase chain reaction and restriction fragment length polymorphism. Annals of the entomological Society of America 90:329-332.

Rohlf FJ. 2010. tpsDig, Digitize Landmarks and Outlines, Version 2.16. In. Department of Ecology and Evolution, State University of New York, Stony Brook, NY.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61:539-542.

Rovito SM, Parra-Olea G, Vásquez-Almazán CR, Luna-Reyes R, Wake DB. 2012. Deep divergences and extensive phylogeographic structure in a clade of lowland tropical salamanders. BMC evolutionary biology 12:255.

Rundell RJ, Price TD. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution 24:394-399.

Sanchez-Azofeifa GA, Rivard B, Calvo J, Moorthy I. 2002. Dynamics of tropical deforestation around national parks: remote sensing of forest change on the Osa Peninsula of Costa Rica. Mountain Research and Development 22:352-358.

Sánchez-Guillén R, Ott J. 2018. Genetic consequences of range expansions along several fronts in Crocothemis erythraea. International Journal of Odonatology:1-11.

Schoener TW. 1968. The Anolis lizards of Bimini: resource partitioning in a complex fauna. Ecology 49:704-726.

Schultz TD, Fincke OM. 2009. Structural colours create a flashing cue for sexual recognition and male quality in a Neotropical giant damselfly. Functional Ecology 23:724-732.

Selys Longchamps Ed. 1860. Synopsis des Agrionines. Première Légion. – Pseudostigma –. Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux - Arts de Belgique 2:9-27.

Servedio MR, Boughman JW. 2017. The role of sexual selection in local adaptation and speciation. Annual Review of Ecology, Evolution, and Systematics 48:85-109.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the entomological Society of America 87:651-701.

Simon S, Schierwater B, Hadrys H. 2010. On the value of Elongation factor-1α for reconstructing pterygote insect phylogeny. Molecular Phylogenetics and Evolution 54:651-656.

Smith BT, McCormack JE, Cuervo AM, Hickerson MJ, Aleixo A, Cadena CD, Pérez-Emán J, Burney CW, Xie X, Harvey MG. 2014. The drivers of tropical speciation. Nature 515:406.

Sobel JM, Chen GF, Watt LR, Schemske DW. 2009. The biology of speciation. Evolution 64:295-315.

Srivathsan A, Meier R. 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28:190-194.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Steinmann H. 1997. World Catalogue of Odonata, Vol. 1: Zygoptera. New York: Walter de Gruyter, Berlin.

Suárez-Tovar C, Sarmiento C. 2016. Beyond the wing planform: morphological differentiation between migratory and nonmigratory dragonfly species. Journal of evolutionary biology 29:690-703.

Sueyoshi T, Ohgaito R, Yamamoto A, Chikamoto M, Hajima T, Okajima H, Yoshimori M, Abe M, O'ishi R, Saito F. 2013. Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM. Geoscientific Model Development 6:819-836.

Sunnucks P, Morales HE, Lamb AM, Pavlova A, Greening C. 2017. Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. Frontiers in genetics 8:25.

Svensson EI. 2012. Non-ecological speciation, niche conservatism and thermal adaptation: how are they connected? Organisms Diversity & Evolution 12:229-240.

Svensson EI, Waller JT. 2013. Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. The American Naturalist 182:E174-E195.

Swaegers J, Mergeay J, Van Geystelen A, Therry L, Larmuseau M, Stoks R. 2015. Neutral and adaptive genomic signatures of rapid poleward range expansion. Molecular ecology 24:6163-6176.

Tavaré S. 1985. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura RM, editor. Lectures on mathematics in the life sciences. p. 57-86.

Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular ecology 21:3907-3930.

van der Vaart AW. 1998. Asymptotic statistics: Cambridge university press.

Villalobos F. 2013. Tree squirrels: A key to understand the historic biogeography of Mesoamerica? Mammalian Biology-Zeitschrift für Säugetierkunde 78:258-266.

Wang IJ, Crawford AJ, Bermingham E. 2008. Phylogeography of the Pygmy Rain Frog (*Pristimantis ridens*) across the lowland wet forests of isthmian Central America. Molecular Phylogenetics and Evolution 47:992-1004.

Warren DL, Glor RE, Turelli M, Funk D. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62:2868-2883.

Webster M, Sheets HD. 2010. A practical introduction to landmark-based geometric morphometrics. Quantitative Methods in Paleobiology 16:168-188.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wellenreuther M, Larson KW, Svensson EI. 2012. Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. Ecology 93:1353-1366.

Wellenreuther M, Sánchez-Guillén RA. 2016. Nonadaptive radiation in damselflies. Evolutionary applications 9:103-118.

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Group NGW. 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 327:343-348.

Wieder RK, Wright SJ. 1995. Tropical forest litter dynamics and dry season irrigation on Barro Colorado Island, Panama. Ecology:1971-1979.

Wiens JJ. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution 58:193-197.

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Jonathan Davies T, Grytnes JA, Harrison SP. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology letters 13:1310-1324.

Wiens JJ, Graham CH. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annu. Rev. Ecol. Evol. Syst. 36:519-539.

Winkler I, Scheffer SJ, Lewis ML, Ottens KJ, Rasmussen AP, Gomes-Costa GA, Santillan LMH, Condon MA, Forbes AA. 2018. Anatomy of a Neotropical insect radiation. BMC evolutionary biology 18:30.

Wright S. 1949. The genetical structure of populations. Annals of Human Genetics 15:323-354.

Xu M, Fincke OM. 2015. Ultraviolet wing signal affects territorial contest outcome in a sexually dimorphic damselfly. Animal Behaviour 101:67-74.

Yanoviak S, Fincke O. 2005. Sampling methods for water-filled tree holes and their artificial analogues. Insect sampling in forest ecosystems. Blackwell Publishing, London:168-185.

Yanoviak SP. 1999. Community structure in water-filled tree holes of Panama: effects of hole height and size. Selbyana:106-115.

The damselfly genus *Megaloprepus* (Odonata: Zygoptera: Pseudostigmatidae): Revalidation and delimitation of species-level taxa including the description of one new species

WIEBKE FEINDT1, 2 * & HEIKE HADRYS1, 2, 3

[1] *University of Veterinary Medicine Hannover, ITZ, Division of Ecology and Evolution, Bünteweg 17d, 30559 Hannover, Germany*
E-mail: wiebke.feindt@ecolevol.de, heike.hadrys@ecolevol.de

[2] *American Museum of Natural History, The Sackler Institute for Comparative Genomics, Central Park West at 79th Street, New York, NY 10024, USA*

[3] *Yale University, Department of Ecology and Evolutionary Biology, 165 Prospect Street, New Haven, CT 06511, USA*

* Corresponding author: WiebkeFeindt

The damselfly genus *Megaloprepus* (Odonata: Zygoptera: Pseudostigmatidae): Revalidation and delimitation of species-level taxa including the description of one new species

Abstract

As the longest-winged odonate species of the extant world, *Megaloprepus caerulatus* Drury, 1782 has received attention by many entomologists. While the behavior and ecology of this species has been object of intense studies, biogeography and species status throughout its distributional range in old-growth Neotropical forests are less well known. For tropical forests this information is a sine qua non when estimating the impact of degradation and climate change. Recent population genetic analyses, quantitative morphometric – and traditional taxonomic studies rediscovered a "cryptic" species complex within the genus *Megaloprepus* Rambur, 1842 — up until now still regarded as a monotypic genus. Here we describe one new species *Megaloprepus diaboli* sp. nov. from the southern Pacific coast of Costa Rica and from the central Caribbean coast of Honduras and Guatemala. The holotype is from the Corcovado National Park, Costa Rica (N 8°28'55.62" W 83°35'13.92") and was deposited at the National Museum of Costa Rica in San José. Aside from the nominal *M. caerulatus*, two formerly described and later refused species within the genus were reevaluated and consequently raised to good species status: *Megaloprepus latipennis* Selys, 1860 is found in the northeastern regions of Mesoamerica and *M. brevistigma* Selys, 1860 in South America east of the Andes. The new species complex is illustrated, (re)described and compared.

Keywords: *Megaloprepus*, cryptic species complex, new species, new species states, Neotropics

Resumen

Debido a que se trata de la libélula con las alas más largas del mundo, *Megaloprepus caerulatus* Drury, 1782 ha recibido atención por parte de muchos entomólogos. Mientras el comportamiento y la ecología de esta especie han sido el objetivo de muchos estudios, el conocimiento sobre su biogeografía y el estado de la especie es menos conocido, a pesar de ser una información muy necesaria en tiempos de degradación del bosque tropical y el cambio climático. Amplios estudios recientes sobre la genética de poblaciones, morfometría y taxonomía tradicional han revivido el complejo críptico de especies del género *Megaloprepus* Rambur, 1892 — hasta ahora considerado como un género monotípico. En el presente trabajo hemos descrito una nueva especie, *Megaloprepus diaboli* sp. nov., la cual se distribuye en la costa Pacífico sur de Costa Rica y la costa central caribeña de Honduras y Guatemala. El holotipo es del Parque Nacional Corcovado, Costa Rica (N 8°28'55,62" W 83°35'13,92"), el cual fue depositado en el Museo Nacional de Costa Rica en San José. Aparte de la única

especie *M. caerulatus*, dos species anteriormente descritas y después descartadas dentro del género son ahora elevadas al estatus de especie: *Megaloprepus latipennis* Selys, 1860 se encuentra en la región noreste de Mesoamérica y *Megaloprepus brevistigma* Selys, 1860 en Suramérica al este de los Andes. El nuevo complejo de especies es ilustrado, redescrito y comparado.

Introduction

The New World Tropics are experiencing tremendous loss of forest cover due to accelerating rates of anthropogenic-induced landscape change (e.g., Gibson et al. 2011, Laurance et al. 2012). Here the primary goals of taxonomic studies are the description and documentation of species ranges, species communities or species states to understand evolutionary processes and to evaluate the effects of habitat loss and climate change. In the order Odonata (damselflies and dragonflies) species that inhabit forests are especially prone to extirpation due to the loss of forest cover (e.g., Paulson 2004, 2006). One prominent example is *Megaloprepus* Rambur, 1842, a Neotropical damselfly genus adapted to the endangered old-growth rain forests. Their exclusive breeding habitats are water filled tree holes (Ramírez 1997, Fincke 1998, Hedström & Sahlén 2003). Consequently, forest loss and climate change may have a significant affect on this genus (cf., Fincke 1998, Fincke 2006, Fincke & Hedström 2008, Khazan 2014). *Megaloprepus* is distributed throughout the Neotropics from southern Mexico to Bolivia including Venezuela, Guyana, Surinam and northern Brazil, but is absent in the Caribbean Islands (Davies & Tobin 1984, Hedström & Sahlén 2001, Fincke 2006). Within its broad geographical range, the preferred habitats are seasonal, semidry lowland forests, moist forests and non-seasonal wet forests from the coastline to mid-elevations (e.g., Hedström & Sahlén 2001).

Over the last 230 years, several major taxonomic studies focused on *Megaloprepus*. Drury (1782) described it first as Libellula caerulata. Sixty years later Rambur (1842) described the genus *Megaloprepus*, included Drury's *Libellula caerulata*, and placed it close to the genera *Mecistogaster* Rambur, 1842 and *Microstigma* Rambur, 1842. For his work he used a specimen from Colombia, which was apparently identical to the one used by Drury from Honduras. Subsequently, Selys included *Megaloprepus* in his 'Synopsis des Agrionines' (Selys 1860) and the 'Revision du Synopsis des Agrionines' (Selys 1886). In his first work he described for the first time three different species within the genus, *M. caerulatus* Drury, 1782, *M. brevistigma* Selys, 1860 and *M. latipennis* Selys, 1860, based on wing morphology and wing coloration as the most prominent taxonomic characters. However, Selys also remarked that those morphological differences might be intraspecific, phenotypic variations (Selys 1860). Consequently, his revision of the genus in 1886 resulted in the revocation of the genus' split into three species. Now he recognized *M. caerulatus* as the only species within the genus and *M. caerulatus brevistigma* as an additional race, which probably were undergoing speciation (Selys 1886). After 1900, taxonomists were less consistent in species or subspecies status. Whereas Calvert strictly rejected *M. latipennis* and *M. brevistigma* as good species (1901-1908), Ris (1916) did not have enough material from the northern distributional range of the genus for a clear statement concerning *M. latipennis*. However, Ris

also considered the *Megaloprepus* group as non-homogeneous and, based on a reanalysis of the five distinct characters described by Selys, he stated that *M. brevistigma* could be recognized (1916). In his contribution to the 'South American Fauna', Schmidt (1942) could not deny Ris` statement. Nowadays one subspecies (*M. caerulatus brevistigma*) is mentioned in the World Catalogue of Odonata (Steinmann 1997). *Megaloprepus caerulatus latipennis* applies as a synonym for *M. caerulatus brevistigma* according to the 2016 updated version of the list of New World Odonates by Garrison & von Ellenrieder (1991) and in the World Odonata List *M. caerulatus* race *brevistigma* and *M. caerulatus* race *latipennis* appear as synonyms for *M. caerulatus* (Schorr & Paulson 2016). However, most recent references still consider *Megaloprepus* as a monotypic genus (e.g., Heckman 2008, Garrison et al. 2010), whereas Fincke and colleagues support the subspecies status (Fincke et al. 2018).

Even though *Megaloprepus* is a well-known genus and a long-time research subject, only until recently its monotypic genus status was questioned again (cf., Fincke 2006). After a period of uncertainty, comparative molecular (Feindt *et al.* 2014), morphometric (Feindt *et al.* in prep.) and the present classical taxonomic analyses from multiple localities across tropical America as well as museum collections allowed the reevaluation of the genus. The integrative analyses of all data sets lead us back in history to accept the hypothesis of a multiple-species genus as suggested by Selys (1860). Based on morphological and genetic evidence, one new species, *Megaloprepus diaboli* **sp. nov.**, is delimited and *M. brevistigma* and *M. latipennis* are considered valid species along with the nominal species *M. caerulatus*. The 'historic species descriptions' were based on a few morphological characters, traditional measurements, and no detailed illustrations. In the present study we revise the genus with illustrations, additional morphological characters, morphometric measurements and genetic data.

Methods

Morphology. Specimens from 11 museum collections (Table 1) and field collections from Mexico, Guatemala, Nicaragua, Costa Rica and Panama were analyzed (Fig. 1). This combination of samples from a wide range of different localities covered most of the known geographical distribution of *Megaloprepus*. All original descriptions were revised, and the geographic origin of the studied museum material was verified as far as possible.

The type material of *M. diaboli* **sp. nov.** was deposited in the Odonata collection of the National Museum of Costa Rica. For *M. latipennis* and *M. brevistigma*, lectotypes were declared from the original collection of E. Selys Longchamps according to his notes from 1860. From this collection stored at the Royal Belgian Institute of Natural Sciences in Brussels, Belgium (RBINS) one specimen from Veracruz, Mexico, was designated as lectotype for *M. latipennis* and one male specimen from Bogota, Colombia, was selected for *M. brevistigma*.

Table 1: Odonate collections of different museums included in the present study and their abbreviation.

RBINS	Royal Belgian Institute of Natural Sciences (Brussels, Belgium)
AMNH	American Museum of Natural History (New York, New York, USA)
YPM	Yale Peabody Museum of Natural History, Yale University (New Haven, Connecticut, USA)
UCMS	University of Connecticut Biological Collections, Department of Ecology and Evolutionary Biology, University of Connecticut (Storrs, Connecticut, USA)
CNIN-UNAM	National Insect Collection (Colección Nacional de Insectos), National Autonomous University of Mexico (Mexico City, DF, Mexico)
RBLT	Odonate Collection of the Los Tuxtlas Biosphere Reserve (Los Tuxtlas, Veracruz, Mexico)
UNAH	Entomological Museum of the National Autonomous University of Honduras (Tegucigalpa, Central District, Honduras)
BIM	Butterfly & Insect Museum La Ceiba (La Ceiba, Atlántida, Honduras)
INBio	National Biodiversity Institute (Santo Domingo, Heredia, Costa Rica)
ANDES-E	Natural History Museum, University of the Andes (Museo de Historia Natural ANDES; Bogotá, Colombia)
FSCA	Florida State Collection of Arthropods, Florida Department of Agriculture & Consumer Services (Gainesville, FL, USA)

Basic morphological characters, such as the forewing and abdomen length (including appendages), were measured with a Vernier caliper. More detailed measurements for all museum specimens were taken digitally via ImageJ (Abràmoff *et al.* 2004) based on standardized photographs using a Canon EOS 600D. Here the dimensions are in millimeters (mm) and consist of the mean of three repeated measurements for each character and each specimen. The length of the blue stripe was measured at the costa as well as length and width of the pseudostigma. For all length measurements we show the mean and the standard derivation ($\pm$SD). Finally, male ligulae were removed from adult males ($N = 14$) and imaged using a Zeiss EVO 60 variable-pressure scanning electron microscope (SEM) at the Microscopy and Imaging Facility (MIF) of the American Museum of Natural History (AMNH). Differences in shape were compared among individuals as well as among species. For the present study, only a few field-collected females were available from Central America and none from Mexico, which did not allow conclusive studies on female morphology.

Terminology of wing venation follows Riek & Kukalova-Peck (1984) and of male genitalia Garrison *et al.* (2010). The following abbreviations are used: AL = abdominal length (including appendages), FW = forewing, FWL = forewing length, HW = hind wing, HWL = hind wing length, IR = intercalary vein, MP = media posterior, RA = radius anterior, RP = radius posterior, CuA = cubitus anterior, S1-10 = abdominal segments one to ten, s1-3 = male genital ligula segments one to three; and for the paraprocts: A = dorsal surface, B = ventral surface, D = base, L = length of the dorsal surface, L' = length of the ventral surface.
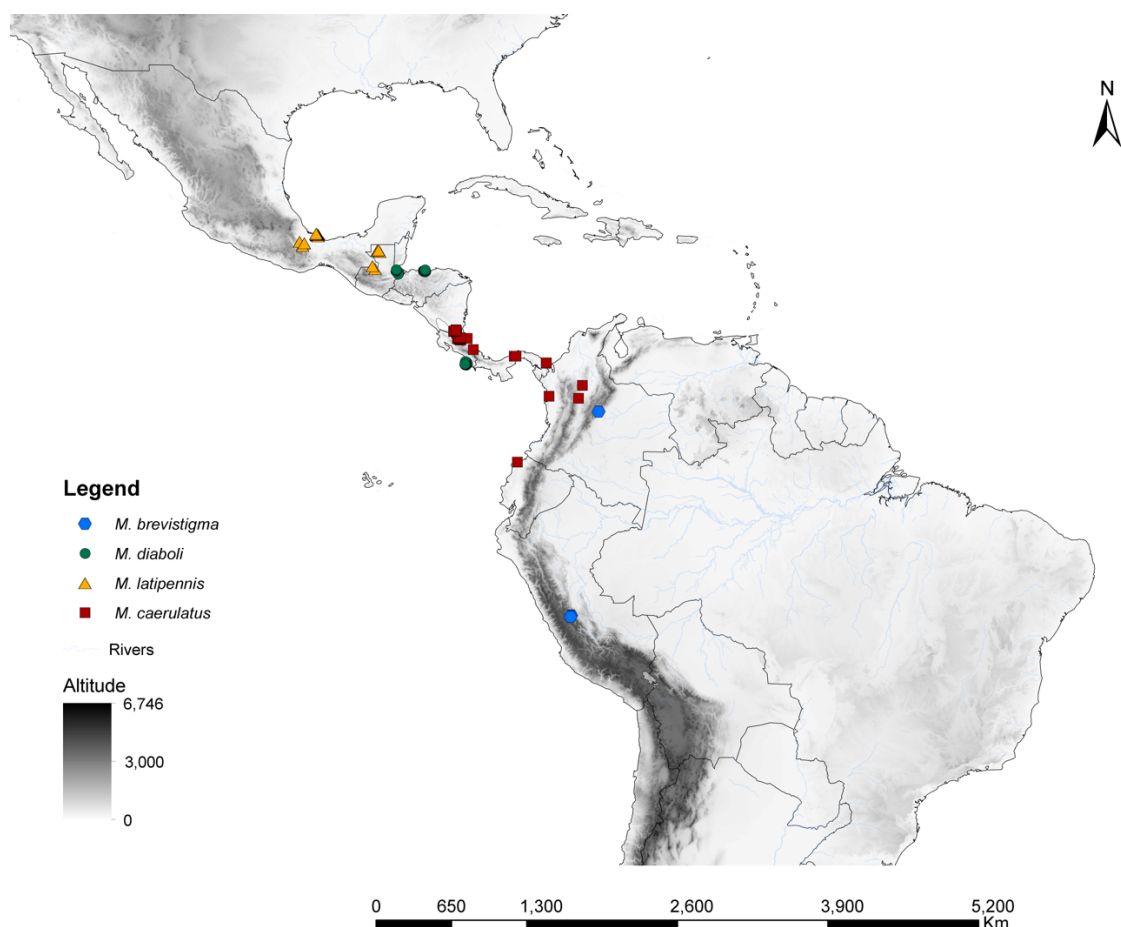
Figure 1: Geographic origin of the material included in this study. Each species is displayed using different colorations.

Phylogenetic reconstruction and character-based DNA barcoding.

In additional support of the proposed taxonomic (re)classification of the genus we reconstructed a phylogeny and generated species-specific character-based DNA barcodes. Therefore, sequences of the 'Folmer' barcoding region (Folmer *et al.* 1994) were used from Feindt *et al.* (in prep.), downloaded from NCBI (for *Mecistogaster linearis* Fabricius, 1776: KF369435 (Dijkstra *et al.* 2014), *Coryphagrion grandis* Morton, 1924: KC912402 (Groeneveld *et al.* 2007)) or donated from B. Willink *et al.*, (unpublished sequence for *M. brevistigma*). The mitochondrial cytochrome c oxidase subunit I (CO1) is a widely used marker gene in odonate population genetics and -barcoding, and allows specimen identification at genus, species and population level (e.g. Rach *et al.* 2008, Damm *et al.* 2010, Bergmann *et al.* 2013).

   Sequences were aligned in MEGA7 (Kumar *et al.* 2008) and a maximum likelihood tree was reconstructed with RAxML vers. 8 (Stamatakis 2006) using a GAMMA model of rate heterogeneity for 20 initial maximum likelihood searches and afterwards the best likelihood scored tree was used for bootstrapping (1,000 replicates) with *C. grandis* as outgroup. In a second step, a character-based barcode analysis was performed applying the CAOS (Character Attribute Organization System) algorithm following recommendations by Rach *et*

111

al. (2008). Therefore, the sequence alignment and the maximum likelihood tree were combined in a single nexus file using MacClade 4v. 4.07 (Maddison & Maddison 2000), which was processed using the CAOS-Analyzer and CAOS-Barcoder. For the final barcode, only pure characteristic attributes (CA) were selected.

Results

Characterization. *Megaloprepus* shares unique features with its five recent sister genera: *Mecistogaster, Microstigma, Platystigma* Kennedy, 1920*, Anomisma* McLachlan, 1877 and *Pseudostigma* Selys, 1860. Here one of the most prominent characters is the replacement of the pterostigma by a dense network of different-sized cells, the pseudostigma, which defined the original group name Pseudostigmatidae (Munz 1919, Tillyard & Fraser 1938-1940, Hedström & Sahlén 2001, Groeneveld *et al.* 2007, Ingley *et al.* 2012, Dijkstra *et al.* 2014). All species of this family oviposit in water-holding plant containers (phytotelmata), which most likely explains their especially long abdomen. Additional taxonomic characters defining the Pseudostigmatidae are an angled frons, convergent longitudinal veins towards the wing margin, a basal recession of IR_2 and RP_3, RA and RP_1 converging at the distal wing margin, and the terminal segment of the male ligula (s3) modified into a flagellum (please also compare Garrison *et al.* 2010 and references herein).

Megaloprepus is an easily recognizable genus in which males and females have almost identical coloration. Although very small individuals have been observed in exceptional cases on Barro Colorado Island, Panama (smallest male with FWL 43.0 mm and AL 60.5 mm; also compare Fincke 1984, Fincke 1992), *Megaloprepus* has the longest and broadest wings and at 190 mm the greatest wingspan of all recent Odonata. The cells of the pseudostigma are colored dark blue to black and are situated at the anterior wing apex in one single row or in a row of two cells. Proximal to the pseudostigma, a dark metallic blue gleaming band crosses all four hyaline wings, beginning at the most distal third of the wing length. The area between CuA and the lower wing margin is highly broadened, and CuA forks at least three times. The thorax and abdomen are dark, gleaming blue to black, and the abdomen varies in length from 60.5–109.6 mm.

Megaloprepus diaboli Feindt & Hadrys **sp. nov.**
(Figures 2, 6, 7, 8)

Specimens examined. (56 specimens, 1 male holotype, 55 paratypes).
Holotype: 1m#, Costa Rica, Peninsula de Osa, Corcovado National Park, N 8°28'55.62" W 83°35'13.92", 60 m, xii. 2011, leg. W. Feindt.
Paratypes: Costa Rica, Peninsula de Osa, Corcovado National Park: 5m#m# 5f#f#, N 8°28'53.52" W 83°35'15.72", 43–84 m, xii. 2011 and iv. 2012, leg. W. Feindt; 1m#, N 8°30'5.06" W 83°35'23.36", 50 m, iv. 1995, leg. L.D. Gomez (INBio); 1m# 1f#,

N 8°33'18.68" W 83°29'41.96", 50-345 m, iii. 1996, leg. L. Angulo (INBIOCRI002540295); Laguna Corcovado: 2m#m#, N 8°28'48.62" W 83°35'28.64", 1–100 m, xii. 1977, leg. D.H. Janzen (INBIOCRI001701902, INBIOCRI001701903); Sierpe: 1f#, N 8°40'44.75" W 83°34'00.17", 200–300 m, v. 1988, leg. A. Solis (INBIOCRI001008793); Rancho Quemado: 1m#, N 8°40'44.75" W 83°34'00.17", 200 m, x. 1993, leg. A. Gutiérrez (INB0004284807).—Honduras, Atlantida Province, Pico Bonito National Park, Rio Cangrejal: 1m#, N 15°42'0" W 86°47'0'', 490 m, ix. 2012, leg. R. Lehman (BIM); Atlantida Province, La Ceiba, C.U.R.L.A (Centro Universitario Regional del Litoral Atlántico) camp forest: 2m#m# 3f#f#, N 15°42'45.71" W 86°51'1.66", 248 m, 1994–1996, leg. R. Lehman (BIM).—Guatemala, Izabal, Morales: 1f#, N 15°30'57.80" W 88°52'15.87", 366 m, ix. 2010, leg. J. Monzón.—Panama: 9m#m# 2f#f# (RBINS: E. Selys Longchamps); Chiriqui: 9m#m# 2f#f# (RBINS: E. Selys Longchamps).—Central America: 7m#m# 2f#f#, leg. Weicht (RBINS: E. Selys Longchamps).

Note. The holotype was deposited in the Odonata collection of the National Museum of Costa Rica, and all paratypes will be placed at the AMNH in New York.

Etymology. This species was named after the common name for damselflies in Latin America: '*Caballitos del Diablo*'. It was given in the hope to raise attention for Odonates and their highly endangered habitats in the Neotropics to support their conservation.

Description of holotype.

Head. Labium yellow becoming dark brown at the most anterior margin. Mandible base yellow to greenish turning into black towards anterior margin. Gena greenish yellow. Labrum, anteclypeus, postclypeus and frons shining black. Antennae black with a yellow ring around their base. Vertex and rear of the head black, ocelli dark yellow to brown amber. Eyes in life bicolored, dorsally dark green turning almost black and ventrally green forming a line with the gena.

Thorax. Prothorax black. Anterior lobe contains additional small greenish yellow oval markings at both lateral ends, and a considerably larger oval central mark. Posterior lobe on both ends with yellow stripes and centrally a very small greenish yellow isosceles triangle with both sides of the triangle concave. Both anterior and posterior lobes smoothly rounded. Propleuron slightly convex with round to oval green coloration nearly covering it completely, distal margins broadly covered in light yellow spots (Fig. 2a, b). Pterothorax black. Yellow antehumeral stripe covering almost whole length of pterothorax, interrupted at distal-dorsal end of mesinfraepisternum; membranous area between posterior lobe of prothorax and mesostigmal plates with large greenish yellow mark appearing as extension of antehumeral stripe; lateral ends of mesostigmal plates somewhat pointed and with dark yellow spots; yellow stripe covering metepisternum also includes lower ventral proportion of mesepimeron and metastigma at its lowermost border, expands in width towards wing bases between FW and HW; metepimerum yellow with a smaller green stripe. Venter yellow with pruinescence continuing through coxae. Coxae yellow green with small dark brown markings. Remaining

internal and dorsal side of legs black but inner surface of femora green; external side of tibiae dark yellow. Tarsi, spines and claws entirely black. Two rows of spines on femora and tibiae, centrally directed.

Wings. Wings stalked, long and broad, with smoothly rounded wingtips (width/length ratio FW 0.23, HW 0.22); mostly hyaline except a gleaming metallic blue band (width at costa FW 14.7, HW 14.6; ratio width of blue stripe/wing length FW 0.18, HW 0.18) crossing last third of all four wings. Additional small, matte milky-white, rounded spots at wingtips, one between blue band and pseudostigma at anterior wing margin, other distal of blue band at posterior wing margin; venation black here. Pseudostigma dark blue and extends seven and six cells in FW and HW. Area basal of CuA broad, CuA forks five times, and secondary branching common. MP bifurcates 17 cells in FW and 19 cells in HW apart from wing margin. No sexual dimorphism in wing coloration.

Abdomen. Abdomen slender and elongate, black with glossy metallic blue as ground color. Abdominal segments S1–4 with additional light green coloration and a variety of small yellow and light bluish spots as follows: in S1 and S2 entire lateral terga covered; in S3 and S4 ventral parts of tergites colored greenish-yellow, expanding in width to distal parts of segments.

Genital ligula (Fig. 7). Second segment with very prominent inner fold, which contains sclerotized hair-like structures on both sides. Lateral lamina of distal s2 starts at lateral ends of inner fold, widely framing distal part of s2 and forming apical lobe. Filamentous whip-like distal segment thickened proximally and contains ventrally directed process close to base. Process of quadrate shape but with rounded edges.

Caudal appendages small (Fig. 8). Cercus haired, black, biramous, both ends acute in lateral view, dorsal branch slightly hairy directed dorso-distal, inner branches point parallel aligned to ventral and hidden between paraproct bases. Paraproct longer than cercus, slightly hairy, light brown or yellow-brown, triangular in lateral view with steeply, acuminate dark brown-black tip directed dorso-distal. Paraprocts in lateral view roughly triangular with A convex and B straight, D = 76% of L – L' length.

Measurements. AL 92.8; FWL 81.2; HWL: 80.6.

Variation in paratypes. Green coloration of mandible base can be more bluish green. Green coloration of propleuron always nearly rounded and size varies slightly. Greenish yellow central mark on posterior lobe of prothorax sometimes barely visible. Pseudostigma extends between four and seven cells in FW and HW. CuA forks between four and six times, MP bifurcates between 11–18 cells in FW and 11–20 cells in HW apart from wing margin (N = 10). Paraprocts always triangular, but A sometimes only smoothly convex (N = 8). Females as males except white dots at wingtips much more pronounced and brighter than in males. Also, secondary wing veins in this area white, and antehumeral stripe nearly complete.

Measurements. m# AL 73–99.5, mean: 87.3±7.7, FWL 59–94.4, mean: 76.2±9.2, N = 16. f# AL 66-89.5, mean: 80.1±7.0, FWL 56.5–79.3, mean: 69.1±6.8, N = 12. Wings: m# width/length ratio FW 0.23±0.02, HW 0.23±0.02, N = 10; blue band width at the costa FW

14.8±3.5, HW 13.5±3.2, *N* = 10; ratio width of blue band/wing length FW 0.19±0.04, HW 0.19±0.04, *N* = 10.

Range. Observed distributional range is old-growth rainforests of the Peninsula de Osa, the southern West Coast of Costa Rica, at the Atlantic Coast of Honduras, Atlántida Province (e.g. Pico Bonito National Park) and the Atlantic Coast of Guatemala between Morales and Puerto Barrios.

Figure 2: *Megaloprepus diaboli* sp. nov. color patterns: (a) lateral view of prothorax, pterothorax including mesostigmal plates, wing base and coxa; (b) dorsal view of prothorax.
al: prothorax anterior lobe. pl: prothorax posterior lobe. mp: mesostigmal plates. cx1–3: coxae.

Megaloprepus latipennis (Selys, 1860)

(Figures 3, 6, 7, 8)

Megaloprepus latipennis Selys, 1860: 14 (species description and diagnostic characters); Selys, 1886: 7 (revocation of *M. latipennis* and inclusion into *M. caerulatus*).
Megaloprepus caerulatus latipennis Ris 1916: 64, 69 (discussion on existence); Steinmann 1997: 459 (synonymic list); Heckman 2008: 201–203 (key to the genus); Schorr & Paulson 2016 (synonymic list).

Specimens examined. (14 specimens, 1 male lectotype, 13 paralectotypes).
Lectotype: 1m#, Mexico, Veracruz, M. Sallé (RBINS: E. Selys Longchamps).
Paralectotypes: Mexico: 5m#m# 4f#f# (RBINS: E. Selys Longchamps); 1f# Veracruz, (RBINS: E. Selys Longchamps).—Guatemala: 3m#m# (RBINS: E. Selys Longchamps).
Other material examined (61 specimens): Mexico: 2m#m# (RBINS); 6m#m# 6f#f#, leg. Don Genin (RBINS); Veracruz, Medellín: 1m#, N 18°34'46.36" W 95°6'23.70", 341 m, viii. 1967 (IBUNAM:CNIN:OD1784); Los Tuxtlas Biosphere Reserve, Tropical Biology Station Los Tuxtlas: 2m#m#, N 18°35'09.60" W 95°04'38.40", 199 m, iii. 2012, leg. W. Feindt; 8m#m# 3f#f#, N 18°34'–18°36' W 95°04'–95°09', 150–700 m, vii. 1985, leg. A. Figueroa (IBUNAM:CNIN:OD1781), ix. 1979, leg. J. Bueno (IBUNAM:CNIN:OD1780), vii. 1968, leg. H. González and A. Imada (IBUNAM:CNIN:OD1785), vii. 1985, leg. P. Sinaca (IBUNAM:CNIN:OD1792), vii. 1988, leg. R. García (IBUNAM:CNIN:OD1797), vii. 1975, leg. H.P. Ruiz (IBUNAM:CNIN:OD1799), vii.–viii. 1975, leg. E. González (IBUNAM:CNIN:OD1793, IBUNAM:CNIN:OD1795, IBUNAM:CNIN:OD1798), ix. 1979, leg. E. González (IBUNAM:CNIN:OD1796), v. 1981, leg. E. González (IBUNAM:CNIN:OD1789); 9m#m# 9f#f#, N 18°34'–18°36' W 95°04'–95°09', 150–700 m, 1985–1986, leg. E. Ramírez, R. Cedillo, C. Mayorga, V. Melendez, R. Mendoza, S. Sinaca, P. Sinaca, A. Ibarra (collection located at the Tropical Biology Station Los Tuxtlas); Municipio de San Andrés, Volcán San Martín Tuxtla: 1f#, N 18°33'5.02" W 95°11'30.22", 1249 m, viii. 1980, leg. P. Guzmán (IBUNAM:CNIN:OD1788); Arroyo Claro: 2m#m#, N 18°33'21.71 W 95°8'6.90", 920 m, viii. 1978, leg. E. González (IBUNAM:CNIN:OD1786) and N 18°35'26.12 W 95°5'25.68", 150 m, viii. 1979, leg. E. González (IBUNAM:CNIN:OD1787); Cerro Simaxtla Catemaco: 1f#, N 18°34'40.51" W 95°4'52.39", 455 m, vi. 1965 (IBUNAM:CNIN:OD1779); Dos Amantes: 1m#, N 18°28'54.06" W 95°4'33.70", 730 m, vii. 1966, leg. R.W. Cruden (IBUNAM:CNIN:OD1778); 2m#m#, N 18°29'11.06" W 95°4'42.04", 694 m, ix. 1969 and N 18°28'50.69" W 95°4'29.06", 770 m, xi. 1969, leg. R.G. Wind (PMNH, ENT 150638); Jalapa de Díaz Oax. Río Uluapan: 1m#, N 18°3'45.64" W 96°31'19.88", 130 m, x. 2010, leg. A. Ibarra (IBUNAM:CNIN:OD1794); Oaxaca, Chiltepec: 1f#, N 17°55'12.68" W 96°9'15.23", 264 m, v. 1980, leg. P. Guzmán (IBUNAM:CNIN:OD1782); 2f#f#, N 17°55'95.60" W 96°9'28.81, 290 m, v. and x. 1969, leg. R.G. Wind (PMNH: ENT 150639, ENT 150638); Oaxaca (a 5 km de San Mateo Yella, a 8 km de Valle Nacional): 1m#, N 17°44'12.16" W 96°17'26.69", 376 m, v. 1981, leg. H. Velazco (IBUNAM:CNIN:OD1783); Tezoyuca: 3m#m#, vi. and xi. 2005, leg. W.L. Tower (AMNH).––Guatemala: Suchitepéquez, Río Bravo: 1f#, N 14°24'52.1" W 91°18'33.4", vii 2012, leg. J.

Monzón; Cobán, National Park Laguna Lachuá: 1f#, N 15°56'49.7" W 90°39'46.5", ii 2012, leg. W. Feindt.

Note. The lectotype was chosen from the Selys collection (RBINS) according to his description from 1860. One male individual from Veracruz, Mexico, was identified as being collected by M. Sallé. Selys mentioned this individual as the voucher for *M. latipennis* in the first Synopsis des Agrionines, and therefore we designated it as lectotype.

Description of lectotype.

Head. Labium dark yellow, becoming dark brown at anterior margin. Mandible base yellow turning into black towards anterior margin. Gena yellow. Labrum, anteclypeus, postclypeus and frons black. Antennae black with a dark yellow ring around their base. Vertex and rear of head black.

Thorax. Prothorax black. Anterior lobe with yellow markings at both ends and a larger, semicircular central mark. Median lobe with round yellow spots in transition to propleuron. Posterior lobe with yellow spots at both ends, yellow triangle in center, the sides of which are strongly concave, forming almost a stripe. Both lobes rounded, posterior lobe additionally slightly arched upwards. Propleuron slightly convex with C-shaped dorsally directed green spot nearly covering half, distal margins yellow (as in Fig. 3a, b). Pterothorax black with a narrow yellow antehumeral stripe beginning at approximately 25% of pterothorax length, growing slightly wider towards wing base; lateral ends of mesostigmal plates somewhat pointed and tips yellow; membranous area between prothorax posterior lobe and mesostigmal plates with large yellow mark; second yellow stripe covering metepisternum also includes lower ventral proportion of mesepimeron and metastigma at its lowermost border, slender, expands in width towards wing bases between FW and HW; metepimerum yellow. Venter yellow with pruinescence continuing to coxae. Coxae yellow with small, light brown rounded marking. Remaining internal and dorsal side of legs black, inner surface of femora green, external side of tibiae yellow. Tarsi, spines and claws black. Two rows of spines on femora and tibiae centrally directed.

Wings. Wings stalked, long and broad, with wingtips smoothly rounded, slightly stretched posteriad (width/length ratio FW 0.24, HW 0.24); mostly hyaline except a gleaming metallic blue band (width at costa FW 12.2, HW 12.1; ratio width of blue stripe/wing length FW 0.14, HW 0.14) crossing last third of all four wings. Additional small, slight milky-white spots at wingtips: one between metallic blue band and pseudostigma at the anterior wing margin and one distal of metallic blue band at posterior wing margin; venation black here. Pseudostigma dark blue, in HW five cells long. Area basal of CuA broad, CuA forks five and six times in FW and HW, secondary branching is common. MP bifurcates 19 cells in FW and 19 cells in HW apart from wing margin. No sexual dimorphism in wing coloration.

Abdomen. Abdomen slender and elongate, black with a glossy metallic blue as ground color. S1–4 with additional light green coloration with a variety of small yellow and bluish influences as follows: in S1 and S2 entire lateral terga covered; in S3 and S4 ventral parts of tergites colored, 75–80% of total length in S3 and approximately 25% in S4.

Genital ligula (Fig. 7). Second segment with very prominent inner fold that contains sclerotized hair-like structures on both sides. Lateral lamina of distal s2 starts at lateral ends of inner fold, widely framing distal part of s2 and forming apical lobe. Filamentous whip-like distal segment with ventrally directed process close to base. Process with sharp quadrate shape.

Caudal appendages brown, small (Fig. 8). Cercus biramous, both ends acute in lateral view, dorsal branch slightly hairy directed dorso-distal, inner branches point parallel aligned to ventral and hidden between paraproct bases. Paraproct longer than cercus, slightly hairy, triangular in lateral view with a steep acuminate dark brown-black tip directed dorso-distal; smallest side of triangle slightly concave. Paraprocts in lateral view pear-like with A and B smoothly convex, D = 51.7% of L – L' length.

Measurements. AL 98.9; FWL: 89.1; HWL: 86.9.

Variation in paralectotypes. Antehumeral stripe sometimes very thin and straight. Black labrum of few males' green-bluish in the transition to the anteclypeus. Pseudostigma in FW five cells long (four cells in two cases and six cells in three cases), in HW predominantly four cells long but also five cells. CuA forks between three and six times, MP bifurcates between 11–20 cells in FW and 12–19 cells in HW apart from wing margin ($N = 17$). Paraprocts always pear-like, but in one case B nearly straight ($N = 10$). Female pseudostigma covers 7–8 cells in FW and 5–9 cells in HW, organized in two rows.

Measurements. m# AL 74.3–98.9, mean: 87.1±7.1, FWL 61.3–93.0, mean: 76.9±7.3, $N = 30$. f# AL 67–87.8, mean: 79.7±6.0, FWL 61.9–74.7, mean: 68.8±4.0, $N = 14$. Wings: #m width/length ratio FW 0.25±0.01, HW 0.25±0.01, $N = 17$; blue band width at the costa FW 12.6±1.6, HW 13.1±1.4, $N = 17$; ratio width of blue band/wing length FW 0.17±0.02, HW 0.18±0.01, $N = 17$.

Range. *Megaloprepus latipennis* occurs in the southern, tropical parts of Mexico, in particular in the Los Tuxtlas Biosphere Reserve (Veracruz), Chiapas and Oaxaca, the northern parts of Guatemala such as the National Park Laguna Lachua (Coban, Alta Verapaz, close to Chiapas in Mexico) and the western Guatemala such as the Río Bravo region.

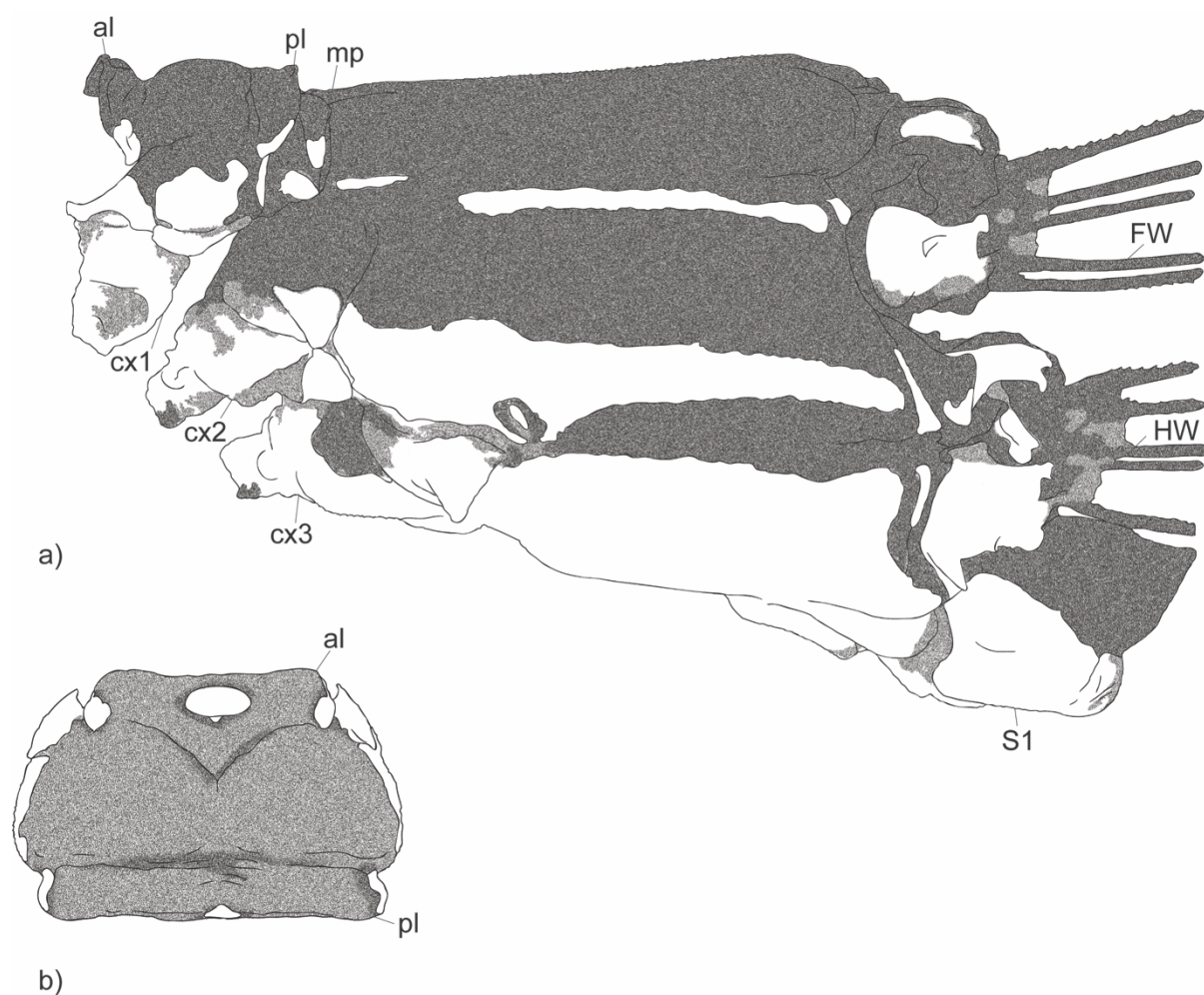**Figure 3:** *Megaloprepus latipennis* color patterns: (a) lateral view of prothorax, pterothorax including mesostigmal plates, wing base and coxa; (b) dorsal view of prothorax.
al: prothorax anterior lobe. pl: prothorax posterior lobe. mp: mesostigmal plates. cx1–3: coxae.

Megaloprepus brevistigma Selys, 1860

(Figures 4, 6, 7, 8)

Megaloprepus brevistigma Selys, 1860: 13–14 (species description and diagnostic characters).
Megaloprepus caerulatus race *brevistigma* Selys, 1886: 7–8 (refusal of species status); Ris 1916: 65–66, 189 (description of male and female, diagnostic characters); Schmidt 1942: 229 (reconsideration of origin of *M. brevistigma*); Steinmann 1997: 459 (synonymic list); Heckman 2008: 201–203, Fig. 3.1.210 (key to the genus); Schorr & Paulson 2016 (synonymic list).

Specimens examined. (16 specimens, 1 male lectotype, 15 paralectotypes).
Lectotype: 1m# (RBINS: E. Selys Longchamps), Bogota.
Paralectotypes: Colombia: 1m# (RBINS: E. Selys Longchamps); 4m#m# 9f#f#, Bogota (RBINS: E. Selys Longchamps).—Peru: 1m# (RBINS: E. Selys Longchamps).
Other material examined (29 specimens): Colombia: 1m#, leg. J. Müller (RBINS R.M.H.N.B. 16.364); 1m#, leg. Don Dr. P. Elsen (RBINS, IG: 28.483); 1m#, leg. Throt (RBINS, Reg. Mus. His. Nat. Belg.IG: 11.128); 10m#m# 10f#f#, 1934, leg. F. Ovalle (AMNH: Ac 33501);

Boyacá, Santa Maria: 1m#, N 5°14'16.8" W 73°16'15.348", 1050 m, iii. 2015, leg. D. García (ANDES-E 18803, Universidad de los Andes, Museo de Historia Natural ANDES).— Ecuador, Macas, Rio Upano, Morona-Santiago: 1m#, leg. F.M. Brown (AMNH); Puyo Oriente, Amazonía: 1m#, app. S 2°11'14.41" W 78°5'23.71", v. 1938, leg. F.M. Brown (AMNH).—Peru, La Merced, La Salud-Vía Chanchamayo: 1f#, (AMNH); Junín, Río Oxabamba, Pampa Hermosa Lodge: 2m#m#, S 10°59' W 75°25', 1900m, xii. 2008, leg. T.D. Donnelly (FSCA, Florida State Collection of Arthropods).

Notes. The lectotype was chosen from a specimen E. Selys assigned to the 'race *brevistigma*'. For the AMNH, Dr. Felipe Ovalle collected most specimens in a single expedition in 1934. The complete collection included about 30,000 specimens (Ac 33501), and no exact locality determination for most of these specimens was possible.

Description of lectotype.

Head. Labium yellow becoming dark brown at anterior margin. Mandible base yellow, turning black towards anterior margin. Gena yellow. Labrum, anteclypeus, postclypeus and frons shining black. Antennae black with a light ring around their base. Vertex and rear of head black.

Thorax. Prothorax black. Anterior lobe contains additional round to oval yellow markings at both ends and centrally an oval yellow mark. Posterior lobe with very small yellow spots at the most lateral surfaces (almost unapparent). Anterior and posterior lobes smoothly rounded. Middle lobe convex and shortened at the center. Propleuron slightly hairy, slightly convex with C-shaped, dorsally directed green area nearly covering half, distal margins with broad dark-yellow coloration (as in Fig. 4a, b). Pterothorax black with narrow, linear, yellow antehumeral stripe beginning at approximately 25% of length; ends of mesostigmal plates somewhat round pointed tips with small dark yellow spots; membranous area between posterior lobe of prothorax and mesostigmal plates black with a yellow mark; yellow stripe covering metepisternum also includes lower ventral part of mesepimeron and metastigma at its lowest border, expands in width towards wing bases between FW and HW; entire metepimerum yellow with a smaller green stripe. Venter yellow with pruinescence continuing to coxae. Coxae yellow with brown/black markings. Remaining internal and dorsal side of legs black, inner surface of femora yellow, external side of tibiae yellow. Tarsi, spines and claws entirely black. Two rows of spines on femora and tibiae, centrally directed.

Wings. Wings stalked, elongated and slender compared to the other *Megaloprepus* species (width/length ratio FW 0.20, HW 0.20), wingtips shallow rounded, without sexual dimorphism in coloration, mostly hyaline except for a metallic dark blue band crossing wing at most distal quarter of each wing, pseudostigma and two additional white spots at wingtips. Metallic dark blue band narrow (width at the costa FW 10.5, HW 10.3; ratio width of blue band/wing length: FW 0.16, HW 0.16), less gleaming. Two additional shining white, nearly round wing spots at wingtips: one at upper margin between pseudostigma and blue band, second at lower wing margin, distal from blue band but proximal of IR_1, here secondary venation also partly white. Pseudostigma dark blue, small, almost rectangular, covers 2 cells

(please compare Ris 1916). Area basal of CuA broad but smaller than in *M. caerulatus*; CuA forks five times, secondary branching less common. MP bifurcates 4 cells apart from wing margin.

Abdomen. Abdomen slender and elongate, black. S1–5 with a variety of green yellow markings as follows: in S1 and S2 entire lateral terga covered; in S3 and S4 ventral parts of tergites are colored, entire length of S3 and approximately 75–80% of both, S4 and S5.

Genital ligula Second segment with very prominent inner fold with sclerotized hair-like structures on both sides. Lateral lamina of distal s2 starts at lateral ends of inner fold, frames widely rounded around distal half of s2, and forms apical lobe. Filamentous whip-like distal segment contains ventrally directed process close to base. Process narrow, not much wider than distal segment, appears folded, with lateral emargination.

Caudal appendages brown, small (Fig. 8). Cercus biramous, both ends acute (lateral view), slightly hairy dorsal branch directed dorso-distal, inner parallel branches point aligned ventrally and hidden between paraproct bases. Paraproct longer than cercus, with few hairs, triangular in side view with a steep acuminate dark black-brown tip directed dorso-distal. Paraprocts in lateral view quadrangular with A angulated convex and B slightly concave (almost straight), D = 69% of L – L' length.

Measurements. AL 82.4, FWL 67.0, HWL 65.8.

Variation in paralectotypes. Membranous area between posterior lobe of prothorax and mesostigmal plates sometimes entirely black. In 95% of the Paralectotypes the metallic blue band proceeds proximal between C and RA for 4-11 cells, sometimes reaching the same length as the blue band itself. MP variable among individuals: in four out of 10 male hind wings MP not bifurcated. The number of cells between bifurcation and wing margin are 4–9 in FW and 0–11 in HW. Pseudostigma is never longer than 3.28 (FW) and 3.02 (HW) (please compare Ris 1916). Paraprocts always quadrangular, but tip sometimes sharper ($N = 10$).
Measurements. m# AL 79.0–98.5, mean: 87.1±7.7, FWL 63.0–94.2, mean: 77.3±9.0, $N = 10$. f# AL 71.2–94.7, mean: 86.4±7.4, FWL 66.9–79.8, mean: 74.4±5.3, $N = 11$. Wings: m# width/length ratio FW 0.21±0.01, HW 0.21±0.01, $N = 10$; blue band at the costa FW 9.5±1.0, HW 9.4±1.0, $N = 10$; ratio width of blue band/wing length: FW 0.12±0.01, HW 0.13±0.01, $N = 10$.

Range. All *M. brevistigma* specimens were located at the east side of the Andes leading towards the Amazon Basin. In Colombia museum samples from Bogota were assigned to *M. brevistigma* and *M. caerulatus*, which allows the assumption of possible overlapping regions.
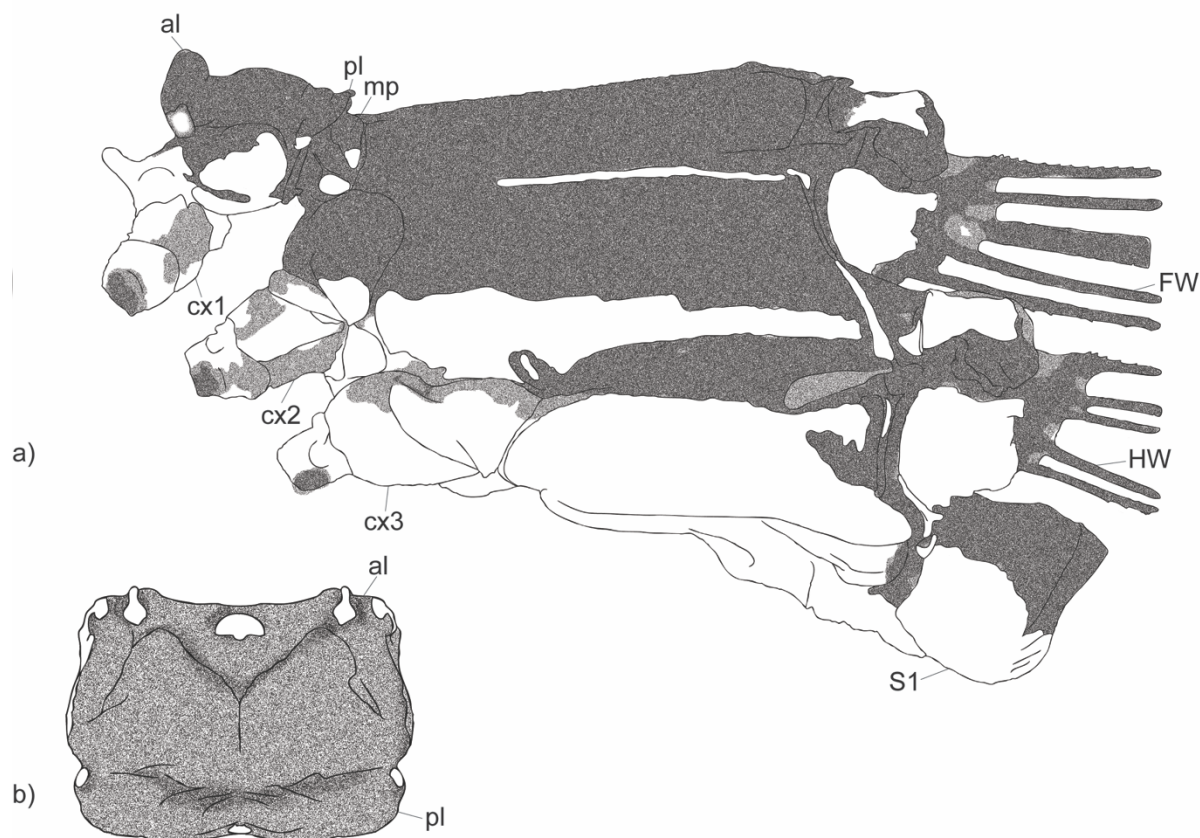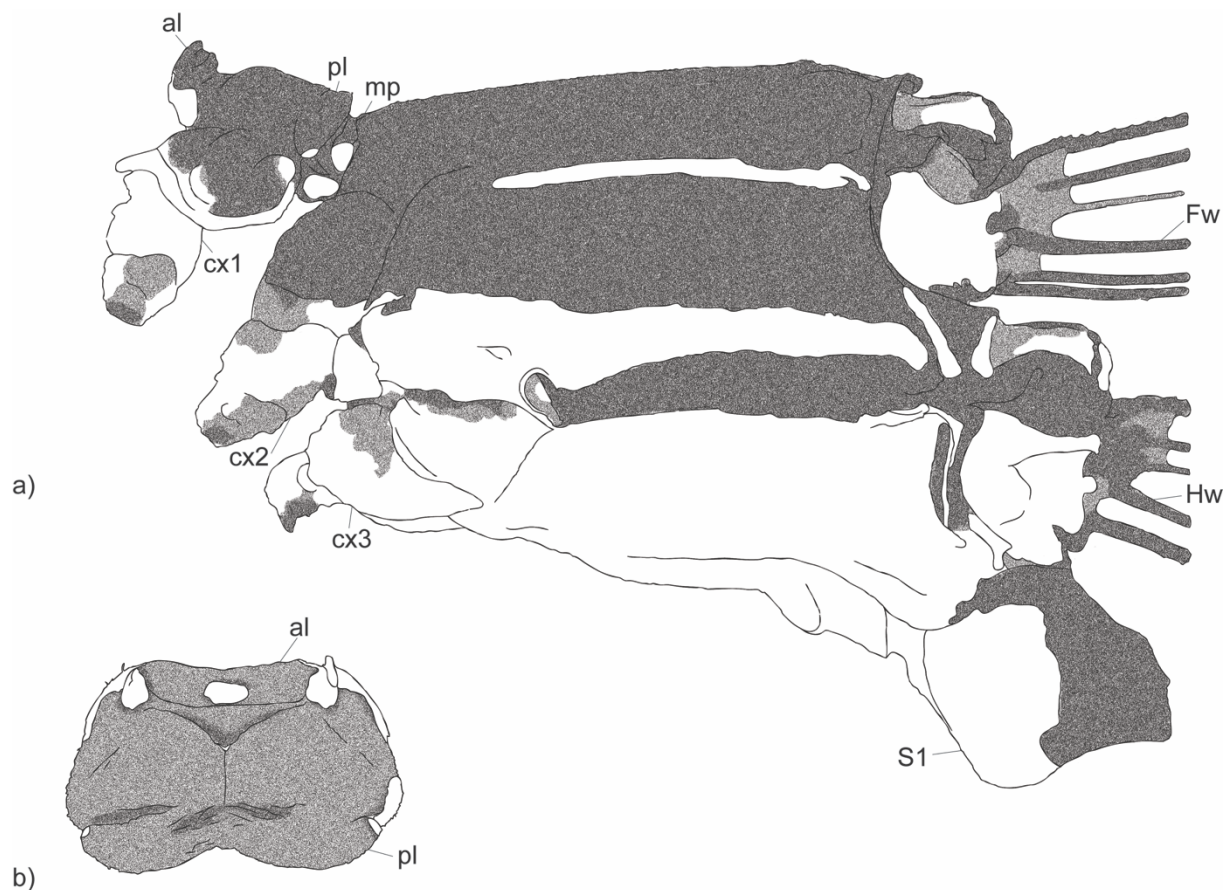
Figure 4: *Megaloprepus brevistigma* color patterns: (a) lateral view of prothorax, pterothorax including mesostigmal plates, wing base and coxa; (b) dorsal view of prothorax.
al: prothorax anterior lobe. pl: prothorax posterior lobe. mp: mesostigmal plates. cx1–3: coxae.

Megaloprepus caerulatus Drury, 1782

(Figures 5, 6, 7, 8, 9)

Libellula caerulata Drury 1782: 75, Plate 50, Fig. 1 (first species description and illustration of a male specimen).
Megaloprepus caerulatus Rambur, 1842: 290–291 (transfer to *Megaloprepus* and redescription); Selys 1860: 12–13 (species description and diagnostic characters); Selys 1886: 6–7 (remarks to the species description 1860); Calvert 1901–1908: 51–53, 417 (description); Ris 1916: 64–69, 189 (description of male and female, diagnostic characters); Ramírez 1997: 6 (description and illustration of the larvae); Steinmann 1997: 459 (synonymic list); Hedström & Sahlén 2001: 1037–1056, Figs. 2, 6a–b, 12a–b, 17a–b (description); Hedström & Sahlén 2003: 9 (extended description of the larvae); Heckman 2008: 201–203, Fig. 3.1.210 (key to the genus); Garrison *et al*. 2010: 394, Figs. 2544, 2546, 2572–2576, Map 106 (description); Schorr & Paulson 2016 (synonymic list).

Type. Museum national d'Histoire naturelle, Paris (not examined).
Specimens examined (81 specimens). 1m# 1f# (RBINS: E. Selys Longchamps); 1m#, San Carlos (RBINS, E. Selys Longchamps).—Central America: 2m#m# 1f#, leg. Weicht (RBINS: E. Selys Longchamps).—Nicaragua, Indio Maíz Biological Reserve, Refugio Bartola: 3m#m#

3f#f#, N 10°58'30.78" W 84°20'06.00", 77–110 m, xi. and xii. 2011, leg. W. Feindt.—Costa Rica, Conservation area Tortuguero, Tortuguero National Park: 1f#, vii. 1968 (AMNH); Estación Cuatro Esquinas: 1m# 1f#, N 10°32'22.38" W 83°30'23.33", 10 m, iii. 1989, leg. R. Aguilar (INBIOCRI000686309) and iv. 1989, leg. R. Delgado (INBIOCRI000575917); Conservation area Cordillera Volcánica Central, Puerto Viejo de Sarapiqui, Biological Research Station La Selva: 5m#m# 1f#, N 10°25'19.74" W 84°00'35.22", 61–97 m, xii. 2011 and iv. 2012, leg. W. Feindt; Selva Verde Lodge: 1f#, N 10°26'30.22" W 84°1'30.56", 62 m, 1991 (AMNH); Braulio Carrillo National Park, Estación Magsaysay: 2m#m#, N 10°24'04.52" W 84°02'57.53", 200 m, i. 1991, leg. A. Fernandez (INBIOCRI000218397, INBIOCRI000218396); 1m# 1f#, N 10°24'07.78" W 84°03'00.82", 160 m, iii. 1991, leg. M.A. Zumbado (INBIOCRI001112977, INBIOCRI001112978); Estación Cuarillo: 1f#, leg. A. Chacón (INBIOCRI001101165); Hitoy Cerere, Cuenca del Estrella: 1m#, N 9°40'15.73" W 83°01'34.16", 250 m, x. 2000, leg. L. Chavarría (INB0004284813). Cairo: 1m#, N 10°5'8.06" W 83°31'42.11", 304 m, i. 1931 (AMNH).—Panama: 4m#m# 4f#f# (RBINS: E. Selys Longchamps); Barro Colorado Nature Monument, Barro Colorado Island: 11m#m# 4f#f#, N 9°9'41.41"–9°9'15.48" W 79°50'30.97"–79°51'3.11", 140–180 m, xi. 1929 (AMNH), i. and xi. 1929, leg. C.H. Curran (AMNH), ii. and iii. 1936, leg. W.J. Gertsch (AMNH), ii. and iii. 1936, leg. F.E. Lutz (AMNH), xi. 1930, leg. E.I. Huntington (AMNH), xi. 1930, leg. Donato (AMNH: F30III2D), iii. 1933 (AMNH: F330313C).—Colombia: 9m#m# 1f#, 1934, leg. F. Ovalle (AMNH: Ac 33501); 1m# 1f#, 1934 (AMNH: Ac 4639); Bogota: 11m#m# 3f#f# (RBINS: E. Selys Longchamps); Caldas, Norcasia, Finca Germánica: 1f#, N 5°34'1.2" W 75°39'32.4", 800 m, v. 2015, leg. P. Cardozo (ANDES-E 21017).—Ecuador, Esmeraldas: 1m# 2f#f#, (RBINS: E. Selys Longchamps).

Other material examined (8 specimens). Costa Rica, Conservation area Cordillera Volcánica Central, Biological Research Station La Selva: 3f#f#, N 10°5'8.06" W 83°31'42.11", 304 m, vi. 2003, leg. R. Vargas (UCMS) and N 10°25'48.43" W 84°0'44.66", 63 m, i. 1998 and i. 1999, leg. R. Vargas and D. Wagner (UCMS); Braulio Carrillo National Park: 1m#, N 10°16'50.28" W 84°6'12.35", 1050 m, iii. 2001, leg. D. Brenes (UCMS); Rio Cantarana: 2f#f#, N 10°21'40.14" W 84°3'19.62", 324 m, iii. 2004, leg. E. Lopez (UCMS); El Ceibo Ranger Station: 1m#, N 10°18'1.42" W 84°5'38.37", 760 m, ii. 2003, leg. S. Gaimari (UCMS); Rio Bijagual: 1f# (UCMS), N 10°19'32.32" W 84°5'22.92", 507 m, iii. 2001, leg. E. Corrales (UCMS).

Note. Analyzed specimens that do not already belong to a collection will stay at the ITZ-Division of Ecology and Evolution, University of Veterinary Medicine Hanover, Germany. The section 'other material' includes adult samples from the Biological Collection at the University of Connecticut, Storrs, USA.

Mature male.

Head. Labium yellow becoming dark brown at anterior margin. Mandible base greenish turning into black towards anterior margin. Gena greenish yellow. Labrum, anteclypeus,

postclypeus and frons shining black. Antennae black with a dark yellow ring around base. Vertex and rear of head black, ocelli dark yellow to light brown. Eyes bicolored, dorsal dark green turning almost black and ventral green forming a line with gena.

Thorax. Prothorax black. Anterior lobe at both sides with additional small oval yellow spots, disposed transversally, and a bigger rhombic central mark. Posterior lobe on both sides with yellow to green comma-shaped spots and a small central yellow isosceles triangle. Both anterior and posterior lobes smoothly rounded, distal margin of posterior lobe slightly turned upward. Propleuron slightly convex and mostly green coloration, distal margins broadly light yellow (Fig. 5a, b). Pterothorax black with yellow antehumeral stripe covering 80% of its total length, appearing interrupted at distal-dorsal end of mesinfraepisternum; outer ends of mesostigmal plates pointed, with dark yellow spots; short yellow stripe at proximal end of mesepisternum close to humeral suture; yellow stripe covering metepisternum also includes lower ventral portion of mesepimeron and metastigma at its lower border, expands in width towards wing bases between FW and HW; metepimerum yellow with a smaller green stripe. Venter yellow with pruinescence continuing to entire coxae. Coxae yellow green with small, light brown-black markings. Remaining internal and dorsal side of legs black, inner surface of femora green, external side of tibiae dark yellow. Tarsi, spines, and claws entirely black. Two rows of spines on femora and tibiae, centrally directed.

Wings. Wings stalked, long and very broad, wingtips smoothly rounded (greatest surface); mostly hyaline except a broad, dark metallic blue band crossing last third of all four wings. Milky-white coloration at wingtips minimized to two cell rows between blue band and pseudostigma at anterior wing margin and one distal of blue band at posterior wing margin, barely visible. Pseudostigma dark blue, extends between five and six cells in FW, four and five cells in HW. Area basal of CuA very broad, CuA forks 3–5 (rarely six) times, and secondary branching very common. MP forks distant from wing margin: number of cells between bifurcation and wing margin 9–19 in HW, 8–15 in FW ($N = 30$). Wings dimorphic, males with an additional matte milky-white band proximal to blue band, more than half width of blue band.

Abdomen. Abdomen slender and elongate, black with slightly metallic brown to bluish sheen as ground color. S1–4 with light green to yellow as follows: in S1 and S2 entire lateral terga; S3 with 80% and S4 with 25% of ventral parts of tergites dark yellow, expanding in width to posterior parts of segments.

Genital ligula (Fig. 7). Second segment with a very prominent inner fold, containing sclerotized hair-like structures on both sides. Lateral lamina of distal s2 starts at sides of inner fold, widely framing distal half of s2 and forming apical lobe. Filamentous whip-like distal segment slightly thickened proximally, containing ventrally directed flattened oval process close to base.

Caudal appendages small (Fig. 8). Cercus nearly hairless, black, biramous, both ends acute in lateral view, dorsal branch slightly hairy, directed dorso-distal, inner branches point parallel aligned to ventral, hidden between paraprocts base. Paraproct longer than cercus, slightly hairy, light brown to yellow, laterally viewed shape of a triangle with a steeply,

acuminate dark brown-black tip directing dorso-distal. Paraprocts in lateral view roughly triangular with A and B slightly convex and sub-equal, D = 66% of L – L' length.

Measurements. m# AL 60.5–106.1, mean: 84.4±9.0, FWL 43.0–87.1, mean: 69.0±9.0, *N* = 102. f# AL 61.0–97.1, mean: 76.9± 7.8, FWL 50.8–81.2, mean: 65.1± 6.9, *N* = 32. Wings: width/length ratio FW 0.25±0.01, HW 0.25±0.01, *N* = 30; blue band width at the costa FW 18.3±2.7, HW 17.8±2.7 mm, *N* = 30; ratio width of blue stripe/wing length FW 0.25±0.02, HW 0.25±0.03, *N* = 30.

Remarks. *Megaloprepus caerulatus* shows the largest size variation in abdominal length and wing length, mostly studied and observed on Barro Colorado Island, Panama. The yellow antehumeral stripe sometimes covers almost the whole pterothorax. The pruinescence on the ventral side of the thorax and the entire coxae can also include the femora. A few males have additional bluish green coloration on the black labrum in transition to the anteclypeus. Paraprocts sometimes very hairy and in one specimen A appeared straight (*N* = 10). In females the proximal white band is not present, but they have very bright, shining white dots at the wingtips in the emarginations of the blue band (at the costa between the blue band and the pseudostigma as well as on the posterior side of the wing distal to the blue band).

Range. The nominal species *M. caerulatus* has a high number of records. Its distribution covers in Central America the Southern Caribbean coast of Nicaragua (Indio Maíz Biological Reserve), the Caribbean slope of Costa Rica (National Park Tortuguero, Biological Research Station La Selva, Hitoy-Cerere Biological Reserve) and Barro Colorado Island. In South America it occurs in the Northern part of Colombia at the Caribbean side, and on the Pacific side of the Andes ranging from Colombia to Peru (West of the Andes).
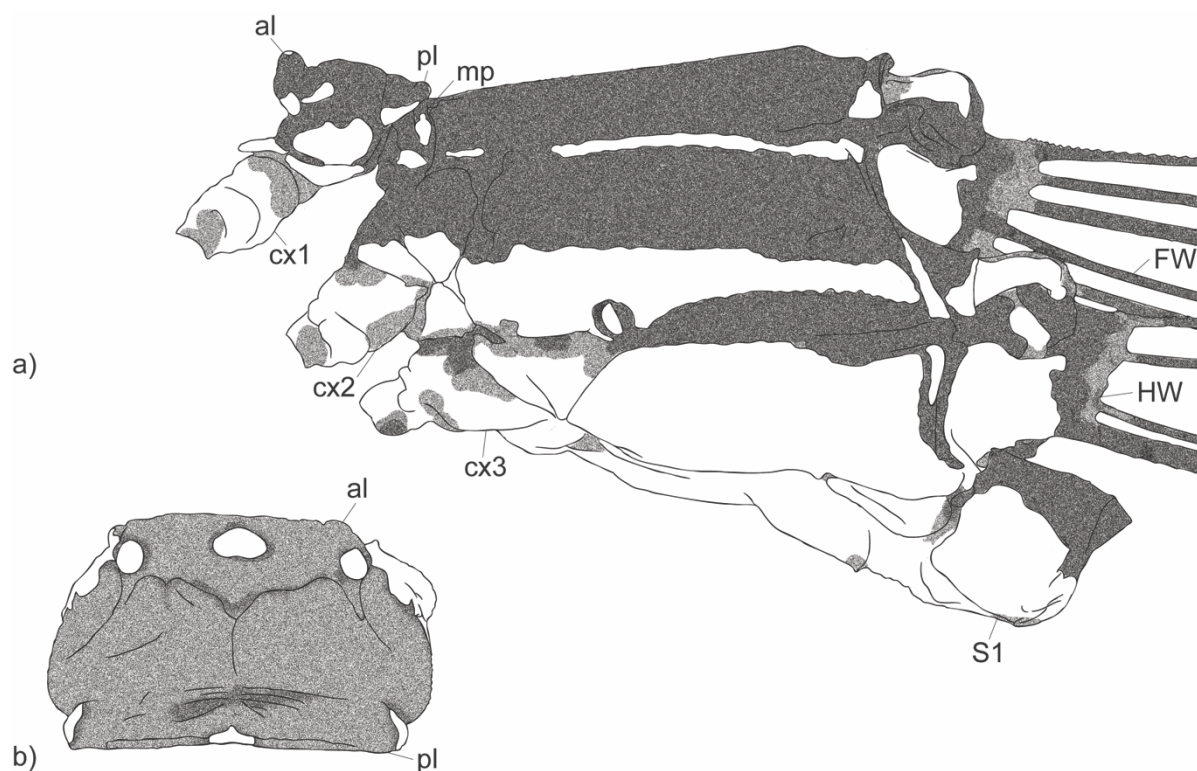
Figure 5: *Megaloprepus caerulatus* color patterns: (a) lateral view of prothorax, pterothorax including mesostigmal plates, wing base and coxa; (b) dorsal view of prothorax.
al: prothorax anterior lobe. pl: prothorax posterior lobe. mp: mesostigmal plates. cx1–3: coxae.
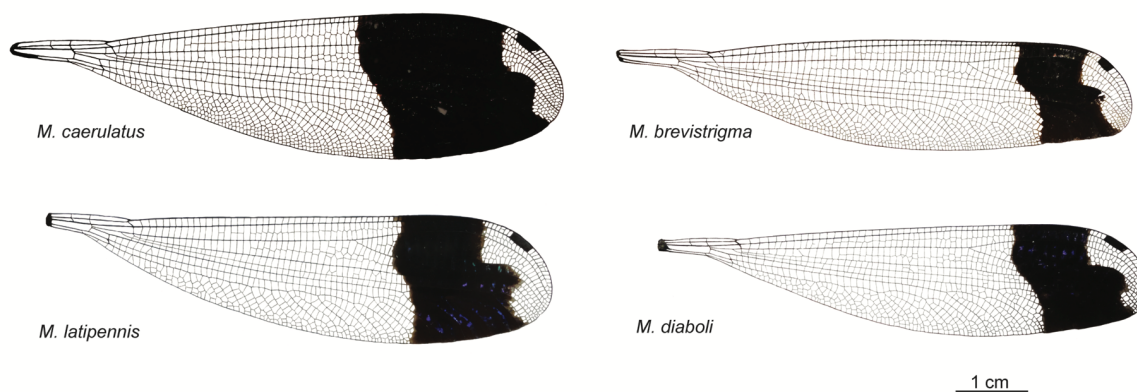
Figure 6: Left hindwings of the four *Megaloprepus* species. The metallic blue wing band proximal of the pseudostigma was integrated to visualize its variation, whereas the additional matte white band of *M. caerulatus* could not be pictured (but see Fig. 6).
The scale bar indicates 1 cm.

Diagnostic characters of the four taxa

The four species discussed here are not obviously distinguishable by their phenotypes in the field. *Megaloprepus brevistigma* is the most phenotypically distinct species compared to *M. caerulatus*, *M. latipennis* and *M. diaboli*. It's narrow metallic blue band with distal extensions at the costa and the small, quadrate pseudostigma are the most distinctive characters (cf. Fig. 6). In contrast, *M. caerulatus* has the broadest wings with the widest blue band, and it is the only species showing sexual dimorphism in wing color (please compare Selys 1860). Only males have an extra matte white band proximal to the blue band, which allows a quick discrimination of the two sexes from a distance (Fig. 9, also compare Table 2 for a character overview). Delimitation between *M. latipennis* from *M. diaboli* is – besides the geographic isolation – based on the shape of the antehumeral stripe, the coloration of the anterior and posterior lobes of the pronotum, the shape of the paraprocts, and the shape of the inner process of the male ligula.

Quantitative analyses found significant size differences between the species in FWL (*M. diaboli* vs. *M. caerulatus*: ANOVA, $p = 0.025$, 3 d.f.; *M. latipennis* vs. *M. caerulatus*: ANOVA, $p = 0.002$, 3 d.f.) and in AL (*M. caerulatus* vs. *M. brevistigma*: ANOVA, $p = 0.008$, 3 d.f.). In addition, Panamanian individuals from Barro Colorado Island show the greatest variation in size than any of other populations (AL 60.5–104 mm; FWL 43.0–78.4 mm).

Key to adult male *Megaloprepus*

1   Hind lobe of prothorax with recessed middle area 2
1´   Hind lobe of prothorax straight 3

2   Hind lobe of prothorax with a small recessed area in the middle; paraprocts in lateral view pear-like with A and B smoothly convex, D = 51.7% of L-L' length . . . . ***M. latipennis***
2'   Hind lobe of prothorax with a large recessed area in the middle, forming two rounded lateral lobes; paraprocts in lateral view quadrangular with A angulated convex and B slightly concave (almost straight), D = 69% of L-L' length; blue wing band narrow; pseudostigma nearly quadrate . . . . ***M. brevistigma***

3   Paraprocts in lateral view roughly triangular with A and B slightly convex and sub-equal, D = 66% of L – L' length wings broad; with additional milky-white band proximal of the metallic blue band . . . . ***M. caerulatus***
3'   Paraprocts in lateral view roughly triangular with A convex and B straight, D = 76% of L – L' length . . . . ***M. diaboli* sp. nov.**

Table 2: Summary of the most distinguishable characters among the four *Megaloprepus* species.

Wings	Ratio of wing width to wing length
	Width of metallic blue band and relative to wing length
	Size of pseudostigma: length and length/width ratio
	Presence of sexual dimorphism and coloration of white wing markings
	Distance of MP bifurcation from posterior wing margin
Prothorax	Middle area
Thorax	Coloration of pronotum
	Antehumeral stripe
Male ligula	Shape of process of distal segment
Paraproctus	General shape
	Curvature of A and B
	Length of D relative to L – L' length
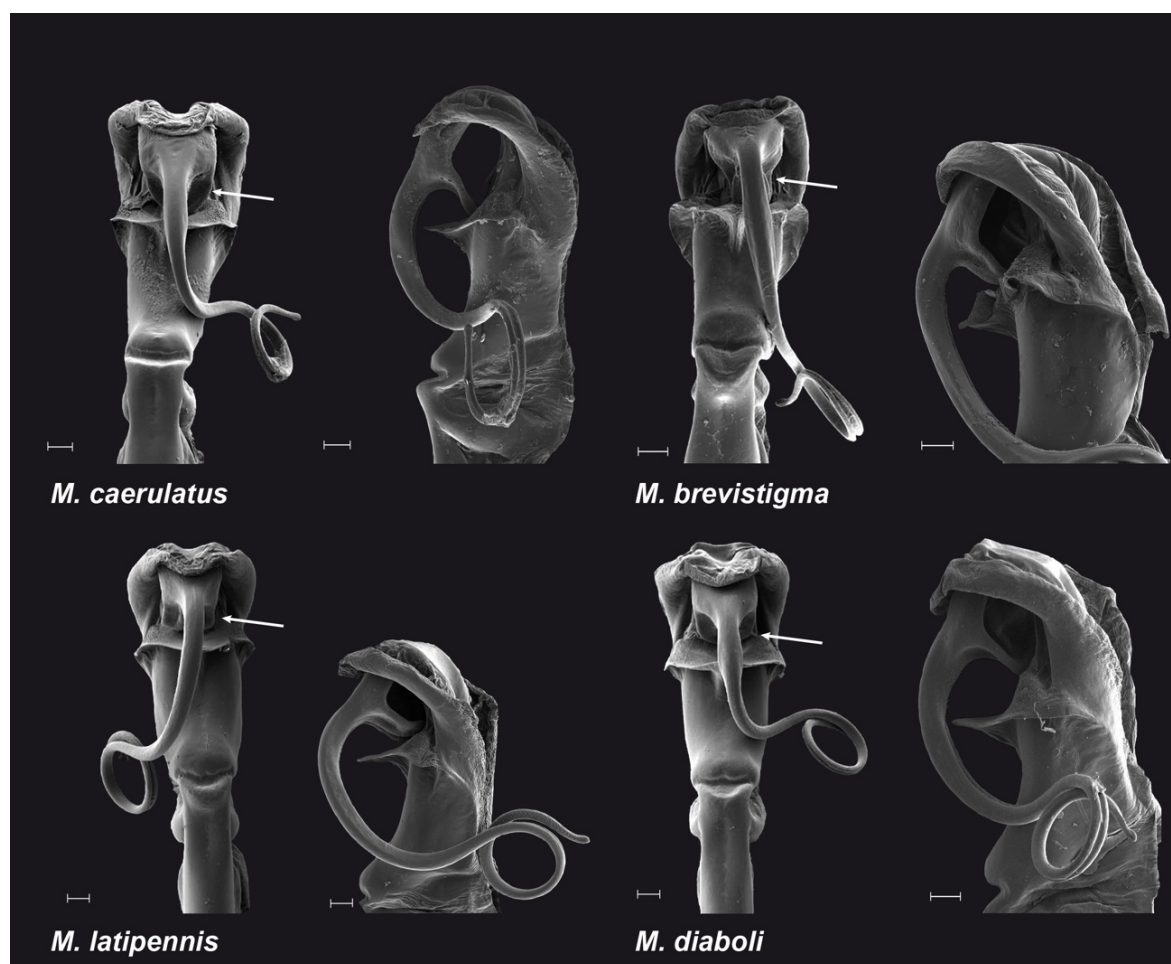
Figure 7: Ventral and lateral views of the genital ligula of the four *Megaloprepus* species. *M. caerulatus*, *M. latipennis* and *M. brevistigma* show significant differences in the inner process of the third segment (s3), *M. diaboli* **sp. nov.** constitutes an intermediate position.
The scale bar indicates 100 µm.
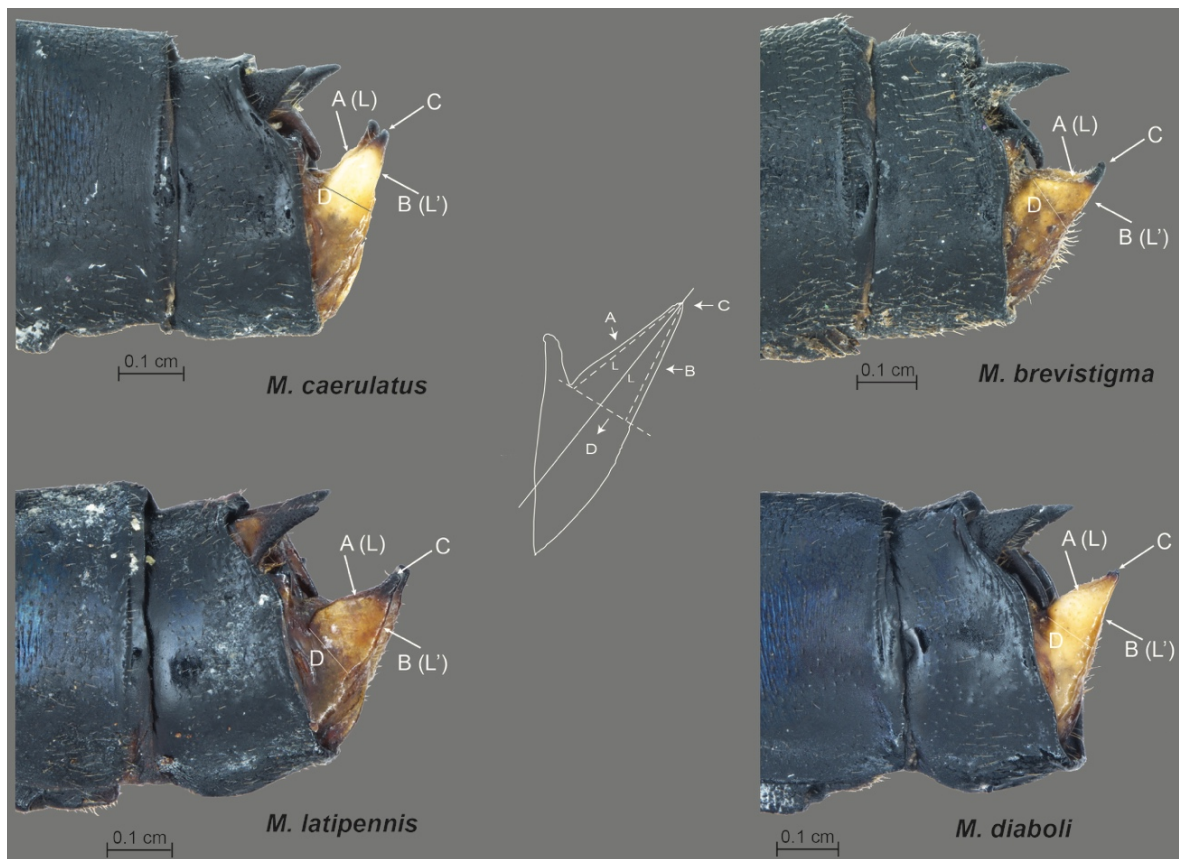
Figure 8: Lateral views of the male caudal appendices.

Figure 9: *Megaloprepus caerulatus* in copulation wheel on Barro Colorado Island (Panama) highlighting the sexual dimorphism.

Phylogenetic reconstruction and character-based barcoding.

The final alignment had a length of 428 bp and included 287 conserved sites and 76 parsimony informative characters. In the phylogenetic reconstruction a distinct separation between the four *Megaloprepus* species was detected (cf. Feindt *et al.* 2014, Feindt et al. in prep.). In accordance with the taxonomy *M. brevistigma* represents a sister clade to *M. caerulatus*, *M. diaboli* and *M. latipennis* (Fig. 10), while *M. caerulatus* and *M. latipennis* appear closely related to each other (based on CO1).

In addition, the character-based barcoding revealed a total of 141 variable nucleotide positions (VNPs) as diagnostic characters among the species. A selection of unique and most informative nucleotide positions is displayed in Fig. 10 (cf. S1 File). Varying numbers of species-specific simple pure characteristic attributes (CA) were identified to separate species (characters unique to one clade and that are absent from the other clade). To distinguish *Megaloprepus* from *C. grandis* 40 simple pure characters were documented. Within the genus 25 simple pure characters were found at the node separating *M. brevistigma* from *M. caerulatus, M. latipennis* and *M. diaboli*; 19 simple pure characters at the node between *M. diaboli* and *M. caerulatus* together with *M. latipennis*, and 32 characters between *M. caerulatus* and *M. latipennis*. Those numbers are comparable to other closely related odonate species (e.g. Damm *et al.* 2010, Bergmann *et al.* 2013, Rach *et al.* 2017). At the population level we observed in *M. diaboli* for example, eight simple pure characters between Guatemala (Protected Reserve Cerro San Gil) and Costa Rica (Corcovado National Park), but only one in *M. caerulatus* between Colombia (Pacific Coast) and Panama (Barro Colorado Island). The latter species shows no variable positions between Nicaragua (Biological Reserve Indio Maíz) and Costa Rica (Biological Station La Selva).
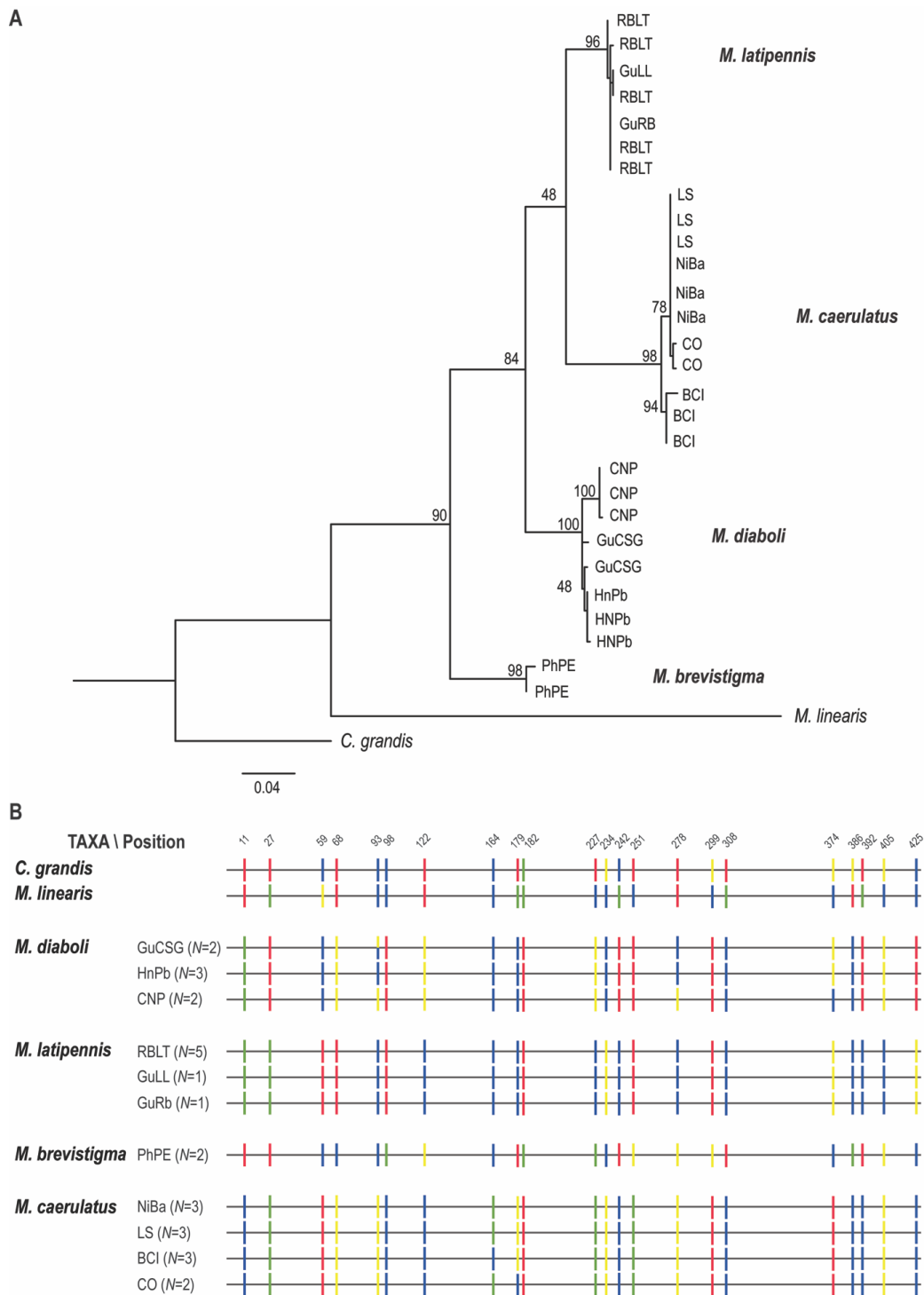
Figure 10: Phylogenetic relationships and the character-based barcodes for the mitochondrial CO1 sequence marker gene for the four species of the genus *Megaloprepus* (*M. diaboli*, *M. latipennis*, *M. brevistigma* and *M. caerulatus*) from several populations*, the close sister species *Mecistogaster linearis* and *Coryphagrion grandis* as outgroup.

A) The phylogeny reconstructed using maximum likelihood in RaxML (1,000 bootstrap replicates), places *Megaloprepus brevistigma* as a sister clade to the three mostly Central American species: *M. diaboli* **spec. nov**, *M. latipennis* and *M. caerulatus*. The present tree reconstruction further identifies *M. caerulatus* and *M. latipennis* as the closest relatives. B) Selection of single pure character attributes for the investigated species. Hereby a red line stands for A (adenine), blue for T (thymine), green for G (guanine) and yellow for a C (cytosine). Please see File S1 for the complete barcode.

*Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; GuLL = National Park Laguna Lachua, Guatemala; GuCSG = Natural Reserve Cerro San Gil, Guatemala; GuRB = Rio Bravo, Guatemala; HnPb = Pico Bonito National Park, Honduras; NiBa = Biological Reserve Indio Maíz, Nicaragua; LS = Biological Research Station La Selva, Costa Rica; CNP = Corcovado National Park, Costa Rica; BCI = Barro Colorado Island, Panama; CO = Pacific Region and West Cariban, Colombia; PhPE = Pampa Hermosa Lodge, Peru.

Discussion

Within the genus *Megaloprepus* Selys (1860) described in his Synopsis des Agrionines for the first time *M. latipennis* and *M. brevistigma* aside of Drury's *M. caerulatus*. Although his second monograph was contradictory and refused the species states of *M. latipennis* and *M. brevistigma* (1886), it was assumed repeatedly that the genus undergoes speciation (e.g. Ris 1916, Fincke 2006).

The present taxonomic study of museum material along with field-collected specimens supports this historical hypothesis. There is combined evidence for the existence of one new species and the re-evaluation of the two species originally described by Selys (1860) besides the nominal *M. caerulatus*. The inclusion of specimens covering nearly the complete geographic range of *Megaloprepus* for taxonomic and molecular analyses was crucial to reveal this "cryptic" species complex. In most previous publications, the material was restricted. However, sample material used in the monographs of Selys (1860) and Ris (1916) originated from regions like those in the present study, both obtaining similar results. The wing morphology described in these two publications was re-evaluated and augmented by differences in coloration of the thorax as well as shape of the paraprocts and the male ligula, resulting in the rise of *M. latipennis* and *M. brevistigma* to a good species status. Consequently, specimens from the Caribbean coast in Nicaragua and Costa Rica, as well as from Barro Colorado Island, Panama, and the Pacific coast of Colombia are assigned to the nominal species *M. caerulatus*. Individuals from the Los Tuxtlas region in Veracruz as well as Oaxaca in Mexico and northern Guatemala are considered *M. latipennis*, whereas those on the east side of the Andes from Colombia, Peru and Ecuador belong to *M. brevistigma*. A fourth group, from the southern Pacific lowlands of Costa Rica (Corcovado National Park) and the Caribbean lowlands of Honduras (Pico Bonito National Park) and Guatemala (Protected Reserve Cerro San Gil), had not been previously recognized. In this new species, *M. diaboli*, morphological differences are less distinctive than quantitative, morphometric data (Feindt *et al.* in prep.) and strong genetic differentiation (Feindt *et al.* 2014). Here the species status was assigned in the most conservative and integrative way based on the taxonomic circle (cf. DeSalle *et al.* 2005).

A significant variation between species was discovered in the male appendages and the prothoraxes. The prothorax has either a recessed middle area or a straight hind lobe, and the male paraprocts are more different in shape between species than within. Another important character is the male ligula, which shows variation in the inner process of the third segment among the four species. Variation in secondary genital structures in Odonata develops either through male-male competition or female-male interactions, whereat the latter directly

influences co-evolution (Cordero-Rivera & Córdoba-Aguilar 2010). Differences in shape could lead to reproductive incompatibility between the new species preventing hybridization (e.g., Cordero-Rivera & Córdoba-Aguilar 2010, Simmons 2014, Wellenreuther & Sánchez-Guillén 2016). As for the remaining genera of the Pseudostigmatidae (cf. Garrison *et al.* 2010, Machado & Lacerda 2017), the male ligula does not show much interspecific variation in *Megaloprepus*. This indicates that the differences found in *Megaloprepus* could indeed point towards a prezygotic barrier. On the other hand, variation among female secondary structures could assumed to be higher than in males as it has been shown, for example in *Calopteryx* Leach in Brewster, 1815 damselflies. Recently, co-evolution in genital traits by sexual antagonism has been found in the co-occurring *C. haemorrhoidalis* Vander Linden, 1825 and *C. splendens* Harris, 1780 (Cordero-Rivera 2017). The author found that during occasional hybridization the sperm removal rate from the female spermathecae was at least two times higher by an interspecific male compared to an intraspecific (Cordero-Rivera 2017). Unfortunately, for our study we did not obtain field samples of females from most sample sites to get conclusive results regarding female genitalia.

Considerable variation among species was also found in the wings: the metallic blue band, the ratio of wing width to wing length, and the size of the pseudostigma. Ris (1916) considered wing morphology and coloration as the most significant characters to separate the taxa, especially the width of the blue band and the ratio of wing length to wing width. The present results confirm his findings. Wing surface is the largest in *M. caerulatus*, with a relatively broad wing (mean width/length ratio = 0.25), whereas the wings of *M. brevistigma* are narrower, with the lower wing margin less curved (mean width/length ratio = 0.21). The highest variation detected was in the width of the blue band. Here *M. caerulatus* and *M. brevistigma* were most distinct from each other with *M. brevistigma* having a significantly smaller wing band (ANOVA, $p < 0.002$, 3 d.f., see Table S2a and b). Between *M. latipennis* and *M. diaboli* the difference was smaller. For delimitation of *M. brevistigma*, Selys (1860) also used the size of the pseudostigma as a taxonomic character (number of cells in the pseudostigma), but this character does not separate the northern / Mesoamerican species from each other. Finally, the relative position of the bifurcation of media posterior (MP) was described as informative (Ris 1916). Here we found a significant support for the separation of *M. brevistigma* from all other species as nearly no bifurcation of MP was observed in *M. brevistigma*.

One important character in interspecific differentiation and recognition based on behavioral patterns is wing coloration. A sexual dimorphism is exclusively present in the nominal *M. caerulatus*, whereas the other species are monomorphic. Previous studies by Fincke and colleagues have linked wing coloration and the reflectance of the wing spots to male-male or male-female interactions such as territoriality and mate recognition (Schultz & Fincke 2009, Xu & Fincke 2015, Fincke *et al.* 2018). A recent publication by Fincke *et al.* (2018) has shown that in both, monomorphic and dimorphic 'populations', males react to the extra white wing band with agonistic behavior and females do not select for a male phenotype. The authors subsequently discussed that wing coloration would therefore not be a barrier for mating (Fincke *et al.* 2018), if for example contact among distant populations of

different morphotypes would be reestablished. However, knowledge about female mate choice based on field studies is still limited. Consequently, the present species evaluation shall be corroborated in future by an intensive study of females including both behavioral and taxonomic characters.

In terms of the genetic distances, all four species are separated by almost equally high distance levels based on the mitochondrial CO1 marker for more than 190 individuals (Feindt *et al.* in prep.). Between *M. diaboli* and *M. latipennis* the uncorrected p-distance is nearly as high as between *M. caerulatus* and *M. latipennis* (7% vs. 9%, respectively), whereas among *M. caerulatus* and *M. diaboli* the p-distance is even slightly higher (10%). These species level distance values are confirmed by a calibration against established species within the Pseudostigmatidae (cf. Feindt *et al.* 2014). The additional barcode analyses strongly support the results furthermore. In contrast, the recent publication by Fincke *et al.* (2018) described three distinct haplotypes based on minimum spanning haplotype networks using data from two mitochondrial genes (CO1 and 16S ribosomal DNA) and one nuclear gene (Histon H3) for approximately 68 individuals. The authors concluded based on alignments that contained for example non-identifiable bases, subspecies states for the three Mesoamerican species. After a careful reanalysis of the available sequence data using conservative alignment building methods, parsimony based haplotype networks and a strict evaluation of population localities based on our here presented barcodes, the resulting genetic data rather mirrors the population genetic analyses by Feindt *et al.* (2014) and further supports the existence of four good species.

Conclusion

The present study of the Neotropical damselfly genus *Megaloprepus* supports the historical hypothesis that *Megaloprepus* is not a monotypic genus. Although most research on *Megaloprepus* has been carried out in Costa Rica and Panama, some distinct behavioral characters may be applicable to all four species within the entire range of the genus. In addition to the same ecological niche, one important feature shared is the sensitivity to heat and consequently low dispersal abilities over open areas. This behavioral pattern in combination with increasing forest fragmentation of old-growth forests in the Neotropics results in restricted distributional ranges and geographic isolation. Consequently, the conservation status of the four species should be re-evaluated, also by considering regional/local endemism

Acknowledgements

References

Abràmoff, M.D., Magalhães, P.J. & Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics international,* 11, 36-42.

Bergmann, T., Rach, J., Damm, S., DeSalle, R., Schierwater, B. & Hadrys, H. (2013) The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. *Molecular ecology resources,* 13, 1069-1081.

Calvert, P. (1901-1908) Odonata. *In:* F. Goldman (Ed), *Biologia Centrali-Americana. Vol. 50: Insecta, Neuroptera*. Porter Dulau & Co., London, pp. 17-420.

Cordero-Rivera, A. (2017) Sexual conflict and the evolution of genitalia: male damselflies remove more sperm when mating with a heterospecific female. *Scientific reports,* 7, 7844.

Cordero-Rivera, A. & Córdoba-Aguilar, A. (2010) Selective forces propelling genitalic evolution in Odonata. *In:* J. Leonard & A. Cordoba-Aguilar (Eds), *The evolution of primary sexual characters in animals*. Oxford University Press, USA, pp. 332-352.

Damm, S., Schierwater, B. & Hadrys, H. (2010) An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Molecular ecology,* 19, 3881-3893.

Davies, D.A.L. & Tobin, P. (1984) *The dragonflies of the world: A systematic list of the extant species of Odonata Vol. 1 Zygoptera, Anisozygoptera* (Vol. Rapid communications, 3): Societas Internationalis Odonatologica.

DeSalle, R., Egan, M.G. & Siddall, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London Series B,* 360, 1905-1916.

Dijkstra, K.-D.B., Kalkman, V.J., Dow, R.A., Stokvis, F.R. & Van Tol, J. (2014) Redefining the damselfly families: a comprehensive molecular phylogeny of Zygoptera (Odonata). *Systematic Entomology,* 39, 68-96.

Drury, D. (1782) *Illustrations of natural history.* (Vol. 3). London: White.

Fabricius, J.C. (1776) *Genera insectorum: eorumque characteres naturales secundam numerum, figuram, situm et proportionem, omnium partium oris adiecta mantissa specierum nuper detectarum.*

Feindt, W., Fincke, O. & Hadrys, H. (2014) Still a one species genus? Strong genetic diversification in the world`s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus.* *Conservation genetics,* 15, 469-481.

Fincke, O.M. (1984) Giant damselflies in a tropical forest: reproductive biologyof *Megaloprepus coerulatus* with notes on *Mecistogaster* (Zygoptera: Pseudostigmatidae). *Advances in Odonatology,* 2, 13-27.

Fincke, O.M. (1992) Consequences of larval ecology for territoriality and reproductive success of a Neotropical damselfly. *Ecology,* 73, 449–462.

Fincke, O.M. (1998) The population ecology of *Megaloprepus coerulatus* and its effect on species assemblages in water-filled tree holes. *In: Insect populations in theory and in practice.* Springer, pp. 391-416.

Fincke, O.M. (2006) Use of forest and tree species, and dispersal by giant damselflies (Pseudostigmatidae): their prospects in fragmented forests. *In: Forest and dragonflies. 4th WDA International Symposium of Odonatology. Pensoft, Sofia*, pp. 103-125.

Fincke, O.M. & Hedström, I. (2008) Differences in forest use and colonization by Neotropical tree‑hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conservation. *Studies on Neotropical Fauna and Environment,* 43, 35-45.

Fincke, O.M., Xu, M., Khazan, E.S., Wilson, M. & Ware, J.L. (2018) Tests of hypotheses for morphological and genetic divergence in *Megaloprepus* damselflies across Neotropical forests. *Biological Journal of the Linnean Society,* 125, 844-861.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology,* 3, 294-299.

Garrison, R.W. & von Ellenrieder, N. (1991) A synonymic list of the New World Odonata. *Argia,* 3, 30.

Garrison, R.W., von Ellenrieder, N. & Louton, J.A. (2010) *Damselfly Genera of the New World, an Illustrated and Annotated Key to the Zygoptera.* Baltimore, MD: The Johns Hopkins University Press.

Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F. & Lovejoy, T.E. (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature,* 478, 378-381.

Groeneveld, L.F., Clausnitzer, V. & Hadrys, H. (2007) Convergent evolution of gigantism in damselflies of Africa and South America? Evidence from nuclear and mitochondrial sequence data. *Molecular Phylogenetics and Evolution,* 42, 339-346.

Harris, M. (1780) An exposition of English insects with curious observations and remarks wherein each insect is particularly described, distinguished, and the natural history faithfully related. *Decad IV. London*, 100-138.

Heckman, C.W. (2008) *Encyclopedia of South American Aquatic Insects: Odonata - Zygoptera* (Vol. 1): Springer Netherlands.

Hedström, I. & Sahlén, G. (2001) A key to the adult Costa Rican "helicopter" damselflies (Odonata: Pseudostigmatidae) with notes on their phenology and life zone preferences. *International Journal of Tropical Biology and Conservation,* 49, 1037-1056.

Hedström, I. & Sahlén, G. (2003) An extended description of the larva of *Megaloprepus caerulatus* from Costa Rica (Odonata: Pseudostigmatidae). *International Journal of Odonatology,* 6, 23-31.

Ingley, S.J., Bybee, S.M., Tennessen, K.J., Whiting, M.F. & Branham, M.A. (2012) Life on the fly: phylogenetics and evolution of the helicopter damselflies (Odonata, Pseudostigmatidae). *Zoologica Scripta,* 41, 637-650.

Kennedy, C.H. (1920) Forty-two hitherto unrecognized genera and subgenera of Zygoptera. *Ohio Journal of Science,* 21 (2), 83-88.

Khazan, E.S. (2014) Tests of biological corridor efficacy for conservation of a Neotropical giant damselfly. *Biological Conservation,* 177, 117-125.

Kumar, S., Nei, M., Dudley, J. & Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics,* 9, 299-306.

Laurance, W.F., Useche, D.C., Rendeiro, J., Kalka, M., Bradshaw, C.J.A., Sloan, S.P., Laurance, S.G., Campbell, M., Abernethy, K. & Alvarez, P. (2012) Averting biodiversity collapse in tropical forest protected areas. *Nature,* 489, 290-294.

Leach, W.E. (1815) Entomology. *In: Edinburgh encyclopaedia, conducted by David Brewster*, pp. 57-172.

Machado, A.B. & Lacerda, D.S.S. (2017) Revalidation of *Platystigma* Kennedy, 1920, with a synopsis of the *quadratum* species group and the description of three new species (Odonata: Pseudostigmatidae). *Zootaxa,* 4242, 493-516.

Maddison, D.R. & Maddison, W.P. (2000) MacClade 4: Analysis of Phylogeny and Character Evolution (V. 4.06). *In: Sinauer, Sunderland, Massachusetts, USA*.

McLachlan, R. (1877) On some new and little-known forms of Agrionina (Légion Pseudostigma, de Selys). *Entomologists Monthly Magazine,* 14, 86-88.

Morton, K. (1924) A new genus and new species of dragonflies from East Africa belonging to the legion Podagrion (Odonata). *The Entomologist,* 57, 217-220.

Munz, P.A. (1919) *A venational study of the suborder zygoptera (Odonata): with keys for the identification of genera*: American Entomological Society.

Paulson, D. (2004) Critical species of Odonata in the Neotropics. *International Journal of Odonatology,* 7, 163-188.

Paulson, D. (2006) The Importance of Forests to Neotropical Dragonflies. *In:* A.C. Rivera (Ed), *Forests and Dragonflies. Fourth WDA International Symposium of Odonatology*. Pensoft Publishers, Pontevedra (Spain), pp. 79-101.

Rach, J., Bergmann, T., Paknia, O., DeSalle, R., Schierwater, B. & Hadrys, H. (2017) The marker choice: Unexpected resolving power of an unexplored CO1 region for layered DNA barcoding approaches. *PloS one,* 12, e0174842.

Rach, J., DeSalle, R., Sarkar, I.N., Schierwater, B. & Hadrys, H. (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B: Biological Sciences,* 275, 237-247.

Rambur, M. (1842) *Histoire Naturelle des Insectes. Névroptères.* (Vol. xvii). Paris.

Ramírez, A. (1997) Description and natural history of the Costa Rican Odonata larvae 5: *Megaloprepus caerulatus* (Drury, 1782)(Zygoptera, Pseudostigmatidae). *Odonatologica,* 26, 75-81.

Riek, E.F. & Kukalová-Peck, J. (1984) A new interpretation of dragonfly wing venation based upon Early Upper Carboniferous fossils from Argentina (Insecta: Odonatoidea) and basic character states in pterygote wings. *Canadian Journal of Zoology,* 62, 1150-1166.

Ris, F. (1916) Libellen (Odonata) aus der Region der amerikanischen Kordilleren von Costarica bis Catamarca. *Archiv für Naturgeschichte,* 82A, 1-197 + 192 pl.

Schmidt, E. (1942) Odonata nebst Bemerkungen über die *Anomisma* und *Chalcopteryx* des Amazonas-Gebiets. *In:* E. Titschack (Ed), *1941-1942 Beiträge zur Fauna Perus nach der Ausbeute der Hamburger Südperu Expedition 1936*, pp. 225-276.

Schorr, M. & Paulson, D. (2016) World Odonata List. *In: Revision 3. University of Puget Sound. Electronic Database accessible at* http://www.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resouces/dragonflies/world-odonata-list2/. *(accessed 17 November 2019)*.

Schultz, T.D. & Fincke, O.M. (2009) Structural colours create a flashing cue for sexual recognition and male quality in a Neotropical giant damselfly. *Functional Ecology,* 23, 724-732.

Selys, L.E.d. (1860) Synopsis des Agrionines. Première Légion. – Pseudostigma –. *Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux - Arts de Belgique,* 2, 9-27.

Selys, L.E.d. (1886) Révision du synopsis des Agrionines, premiére partie comprenant des légions Psuedostigma – Podagrion – Platycnemis et Protoneura. *Mémoire Cour. Académie Royale Belgique,* 38 (4), [1]+iv+233.

Simmons, L.W. (2014) Sexual selection and genital evolution. *Austral Entomology,* 53, 1-17.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22, 2688-2690.

Steinmann, H. (1997) *World Catalogue of Odonata, Vol. 1: Zygoptera.* (Vol. Das Tierreich / The Animal Kingdom 110: i-xxi, 1-500). New York: Walter de Gruyter, Berlin.

Tillyard, R.J. & Fraser, F.C. (1938-1940) A reclassification of the order Odonata. *Australian Zoologist,* 9, 125-169.

Vander Linden, P.L. (1825) *Monographiae Libellulinarum europaearum specimen*. Bruxelles: apud J. Frank.

Wellenreuther, M. & Sánchez-Guillén, R.A. (2016) Nonadaptive radiation in damselflies. *Evolutionary applications,* 9, 103-118.

Xu, M. & Fincke, O.M. (2015) Ultraviolet wing signal affects territorial contest outcome in a sexually dimorphic damselfly. *Animal Behaviour,* 101, 67-74.

Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world´s largest wings

Wiebke Feindt[1*], Sara J. Oppenheim[2], Robert DeSalle[2],
Paul Z. Goldstein[3], and Heike Hadrys[1,2,4]

[1] University of Veterinary Medicine Hannover, ITZ - Division of Ecology and Evolution, Hannover, Germany
[2] American Museum of Natural History, Sackler Institute for Comparative Genomics, New York, NY, United States of America
[3] Systematic Entomology Laboratory (USDA-ARS), National Museum of Natural History, Washington, DC, United States of America
4 Yale University, Department of Ecology & Evolutionary Biology, New Haven, Connecticut, United States of America

* Author for correspondence
Email: wiebke.feindt@ecolevol.de

Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world´s largest wings

Abstract

The evolution, development and coloration of insect wings remains a puzzling subject in evolutionary research. In basal flying insects such as Odonata, genomic research regarding bauplan evolution is still rare. Here we focus on the world's largest odonate species - the "forest giant" *Megaloprepus caerulatus,* to explore its potential for looking deeper into the development and evolution of wings. A recently discovered cryptic species complex in this genus previously considered monotypic is characterized by morphological differences in wing shape and color patterns. As a first step toward understanding wing pattern divergence and pathways involved in adaptation and speciation at the genomic level, we present a transcriptome profiling of *M. caerulatus* using RNA-Seq and compare these data with two other odonate species. The *de novo* transcriptome assembly consists of 61,560 high quality transcripts and is approximately 93% complete. For almost 75% of the identified transcripts a possible function could be assigned: 48,104 transcripts had a hit to an InterPro protein family or domain, and 28,653 were mapped to a Gene Ontology term. In particular, we focused on genes related to wing development and coloration. The comparison with two other species revealed larva-specific genes and a conserved 'core' set of over 8,000 genes forming orthologous clusters with *Ischnura elegans* and *Ladona fulva*. This transcriptome may provide a first point of reference for future research in odonates addressing questions surrounding the evolution of wing development, wing coloration and their role in speciation.

Introduction

The bauplan evolution of the Pterygota (flying insects) is one of the major challenging subjects of evolutionary research. Although the unique appearance of wings in Hexapods has led to the greatest adaptive radiations in the animal kingdom, the precise developmental mechanisms and their evolution are yet not fully understood [1]. A wide range of research is focusing on wing development, shape and coloration and their role in speciation, but so far most research has been limited to more derived model systems such as *Drosophila* sp.*, Tribolium* sp. and some Lepidoptera [1-7].

Today progress in high throughout sequencing, advancing analytical methods and an easy access to next generation sequence data, makes integrative approaches achievable for non-model organisms [7-10]. Specifically, transcriptomics are suitable because they enable

simultaneously the analysis of expression patterns of known developmental genes and the identification of new candidate genes [8, 11]. Moreover, interspecific transcriptome comparisons enhance our ability to infer the mechanisms underlying homologous structural and functional changes as well as allow to detect fundamental principles and conserved features [12, 13].

Among the oldest flying insects [14-16], Odonata (dragonflies and damselflies) with their exclusive set of bioindicator traits hold a key role as "non-model" organisms in ecological and evolutionary research [17, 18]. However, in the evolutionary-developmental context, the molecular basis of wing development and the evolution of morphological variation in odonate wings, so far received little attention. One species promising more insights into wing evolution is the Neotropical damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera, Pseudostigmatidae), because a recent study of *Megaloprepus* revealed a radiation into at least three geographically separated cryptic species ([19], Fig 1A). These species show differences in wing shape, i.e. in wing width, the curvature of the lower wing margin and width of the blue wing band (Feindt et al. in prep). In addition, only the nominal species *M. caerulatus* shows sexual dimorphism in wing coloration. It has been described that modified expression patterns or signaling cascades are responsible for the variation in wing morphology, since such changes are associated with downstream responses to supposedly conserved wing-pattern genes [1, 2, 7, 20].

Integrative research on the origin of morphological variation associated with diversification in odonates is rare and hampered by a shortage of primary data. '*Omic*' studies are still at their beginning and have focused so far on three species: *Enallagma hageni* [21], *Ischnura elegans* [22-25] and *Calopteryx splendens* [26]. Only one study addressed the importance of transcriptional information across embryogenesis to highlight gene sets involved in morphogenesis [24]. Undoubtedly there is a need to integrate developmental data into evolutionary research to - for example - obtain a broader knowledge of species and tissue specific expression patterns.

Thus, we here present a *de novo* transcriptome assembly from the larval thorax of *M. caerulatus* with the overall goal of detecting expressed genes related to wing patterning that might be relevant to the interplay between genomics, development and morphological variation. Specifically, we first focus on a high completeness of the transcriptome and catalogue the candidate wing genes in odonates found in *M. caerulatus*. Secondly, to portray larva-specific genes, we compared the transcriptome of *M. caerulatus* with that of two adult odonate species: *I. elegans* (Odonata: Coenagrionidae) and *Ladona fulva* (Odonata: Libellulidae).

Material and Methods
Sample collection, RNA isolation and sequencing
One individual *M. caerulatus* larva (Fig 1A, 1B) was collected from a natural tree hole [27] in a lowland rain forest at the La Selva Biological Station (OTS, Organization for tropical Studies) in Costa Rica (10° 26' N, 83° 59 W). The larva (total length = 1.96 cm) was

immediately euthanized and stored in RNA*later* (Thermo Fisher Scientific Inc., USA). Prior to RNA isolation, the larva was dissected on ice to isolate the thorax (including the dorso-lateral wing buds) from the head and abdomen (Fig 1B). The tissue was frozen in liquid nitrogen and ground with a pestle. Total RNA was extracted from the thorax using TRIzol reagent (Invitrogen, USA) in combination with RNeasy Micro kit (Qiagen Inc., USA) for subsequent RNA purification. Overall quality and quantity of the isolated RNA were assessed with the BioAnalyzer 2100 (Agilent Inc., USA). Although the larva was collected into RNA*later*, some RNA degradation had occurred. Therefore, we used a TruSeq Stranded Total Library Preparation kit (Illumina, Inc., USA) for library preparation, with Ribo-Zero treatment to select preferentially for mRNA transcripts. The cDNA libraries were paired-end sequenced (2x125bp) on an Illumina HiSeq 2500 (Illumina, Inc.).

Read cleaning and de novo assembly

Raw sequence reads were first checked for overall quality using a Phred-like score in FastQC [28] and, based on these results, adapters and low-quality reads were removed with Trimmomatic 0.33 [29] at the Q20 level. Reads containing ribosomal RNA (rRNA) sequences were erased from the dataset to avoid mis-annotation of rRNAs as putative proteins [30] using SortMeRNA version 2.0 [31]. The Kraken taxonomic sequence classification system version 0.10.5 [32] was applied to filter out prokaryotic sequences. Those reads belong potentially to microorganisms co-inhabiting tree holes or to the microbiome of the larva. Singleton reads (where only 1 member of a read pair remained after the previous cleanup steps) were further removed before assembly.

The *de novo* assembly was conducted using Trinity version 2.0.6 [33, 34] with default parameters except for setting the strand specific flag (RF), a read normalization, and a lower limit of 300 bp on contig size. Assembly quality and completeness were evaluated in several steps. General assembly summary statistics were calculated via *TrinityStats.pl* [34]. As a more reliable estimator of assembly completeness we also calculated additionally the ExN50 statistic. Reads were mapped back to the assembly [35] and following Haas et al. [34] the Ex90N50 was determined. This represents the N50-value at 90% of the total normalized contigs, which is excluding contigs with a low read coverage. For an evaluation of completeness BUSCO-Benchmarking Universal Single-Copy Orthologs [36] version 1.1 was used and the RSEM-EVAL package distributed with DETONATE [37] represented our reference-free evaluation method to calculate assembly scores. Because CD-Hit [38] reduced our BUSCO scores, we finally filtered the raw assembly by applying RSEM-EVAL's contig impact score [37]. Contigs with impact scores less or equal than zero were removed from the assembly using an in-house R script in RStudio [39] and the Bioconductor R package [40].

The cleaned raw reads are available under BioProject: PRJNA336267, BioSample: SAMN05507136 and the sequence read archive (SRA) SRR3997526. Our Trinity assembly used for all subsequent analyses is available in NCBI's Transcriptome Shotgun Assembly database under the TSA GEXY00000000.

Gene prediction and functional annotation

Open reading frames (ORFs) from start to stop codon on a six-frame translation were identified using TransDecoder (http://transdecoder.github.io) [34]. To further improve the ORF identification, the filtered assembly was first blasted against the arthropod data base (e-value cutoff: $1e^{-5}$) downloaded from UniProtKB [41]. This was followed by HMM (hidden Markov models) searches against the Pfam-A protein domain database [42] via Hmmer version 3.1 [43]. To maximize sensitivity, these results were retained as a basis for informing protein prediction in a second TransDecoder step (2-step prediction). The final predicted protein completeness was evaluated using BUSCO [36].

For functional annotation, initial sequence homology searches were performed with BLASTp (e-value cutoff: $1e^{-7}$) against an individually designed "insect reference data base". This customized data base contained the arthropod protein database from UniProtKB (including SWISS-PROT and TrEMBL, [41]) and protein databases for 4 Hemiptera, 21 Hymenoptera, 3 Lepidoptera and Coleoptera, and 26 Diptera species as the closest relatives to Odonata available from NCBI (data downloaded August 3rd, 2015). Sequences without a hit were additionally blasted against the non-redundant database nr - RefSeq: NCBI Reference Sequence Database (downloaded June 14th, 2016) using BLASTp and an e-value threshold of $1e^{-7}$. Putative protein sequences and BLAST results were uploaded to Blast2GO [44, 45], where InterProScan [46] searches were carried out. The InterProScan and BLAST results were used for Gene Ontology (GO) term mapping (http://geneontology.org/) [47].

Identification of key genes

The annotated transcriptome was screened for genes related to stress response, housekeeping genes, developmental genes, and genes responsible for wing development and coloration. Stress response and housekeeping genes were extracted from the annotated *M. caerulatus* transcriptome searching for keywords via Blast2GO [44].
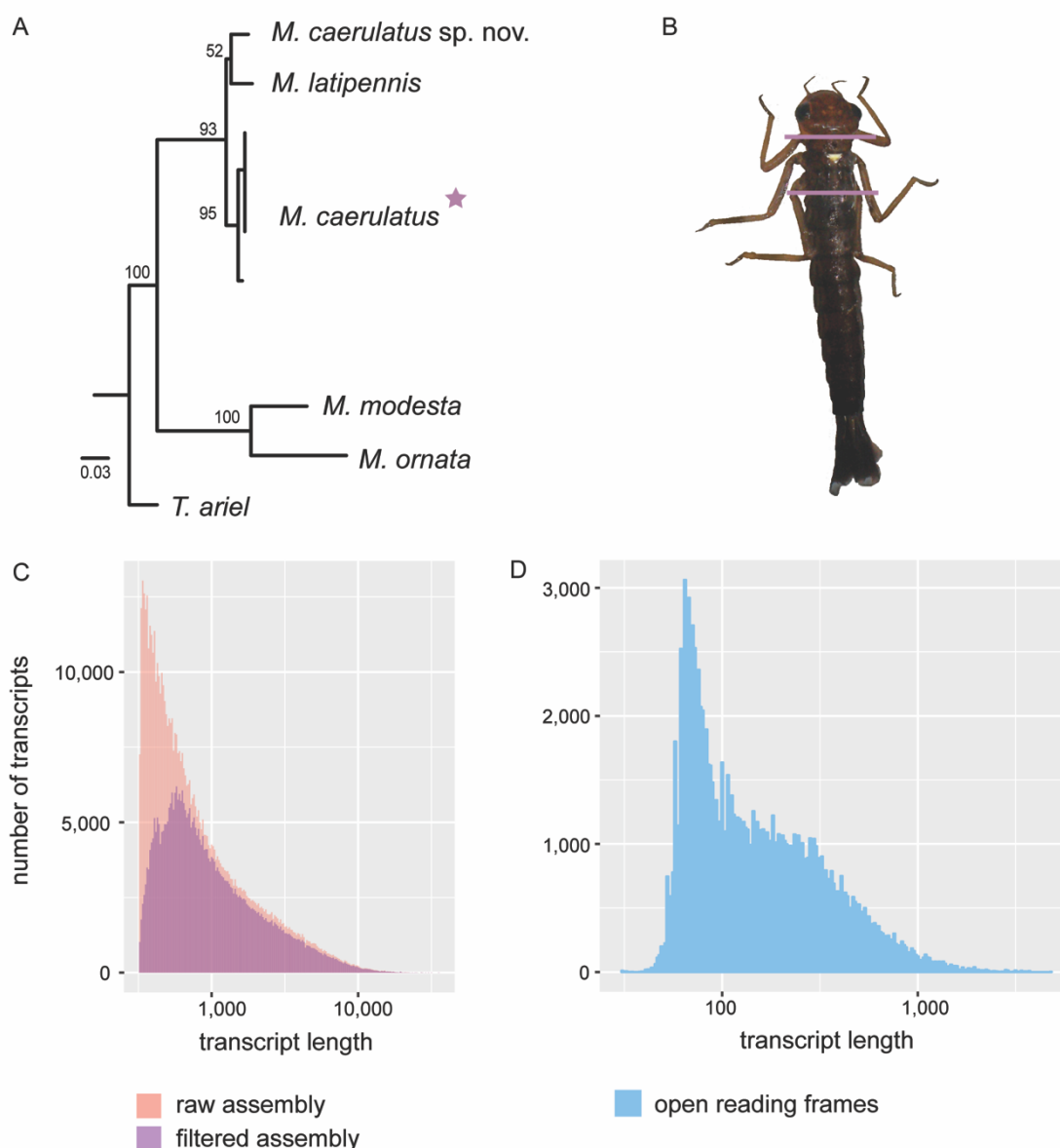
Fig 1. Thorax transcriptome of *Megaloprepus caerulatus*. A) Phylogeny based on the 16S rRNA gene showing the position of *Megaloprepus caerulatus* within the Pseudostigmatidae using *Teinobasis ariel* as outgroup (cf. [19]). The NCBI accession numbers are: KF895223, DQ642987, KF895193, JQ966660, KF895130, KF895162, JQ966657, DQ642983, JQ966662. B) Exemplary illustration of a *M. caerulatus* larva – here of about 2.5 – 3 cm in length. The section between the two lines indicates the tissue used for RNA extraction. C) Difference in the number of transcripts over transcript length between the raw assembly and the filtered assembly. The filtering reduced redundancy and the amount of shorter transcripts. D) Length distribution of the final predicted open reading frames. Note whereas in plot B both axes are logarithmic; in C, only the x-axis is logarithmic.

To detect genes involved in insect development, reference sequences were downloaded from the Homeobox database (HomeoDB; http://homeodb.zoo.ox.ac.uk/, [48, 49]). Hereby we focused on the HOXL subclass (*Hox* genes and *Hox*-derived genes) and NKL subclass (ParaHox gene cluster), both are fractions of the largest gene class Antennapedia (ANTP

class) within the homeobox genes. The HOXL subclass and NKL subclass reference sequences were blasted against the *M. caerulatus* transcriptome and hits were verified via local BLAST searches. In order to identify additional differences of gene expression between adults and larvae, the *Hox* gene and ParaHox gene cluster reference sequences were also blasted against the *I. elegans* (SRR1265958) and *L. fulva* (SRR1850403) transcriptomes (see section comparison with other Odonata).

Genes responsible for wing pigmentation and wing development including wing shape such as the wing gene regulatory network (wing-patterning network) and the four major wing developmental signaling pathways (Hedgehog: *Hh*, Decapentaplegic: *Dpp*, wingless: *wg* and Notch: *N*) were identified within the *M. caerulatus* transcriptome via reciprocal BLASTp searches. Thus, reference sequences were downloaded from Swiss-Prot [41] or NCBI and blasted against the transcriptome. All potential positive hits were verified via a local BLAST search or inside B2GO [44].

Comparison with other Odonata

The predicted open reading frames of *M. caerulatus* were compared to the damselfly *I. elegans* (SRR1265958) and to the dragonfly *L. fulva* (SRR1850403). Raw reads for both species were downloaded from the NCBI's sequence read archive (http://www.ncbi.nlm.nih.gov/Traces/sra), assembled *de novo* [33, 34] and open reading frames predicted (http://transdecoder.github.io) following the steps described above, again under the strict completeness control. The Trinity assemblies for *I. elegans* and *L. fulva* are available upon request.

Overlaps were determined via comparative sequence similarity applying a reciprocal BLAST search using an in-house Perl script that reverts to BLASTp with a significant e-value of $1e^{-7}$. OrthoVenn [50] was further applied to categorize the transcripts into orthologous clusters. It simultaneously annotates the clusters, which were extracted to compare among the three transcriptomes.

Results

Sequencing and *de novo* transcriptome assembly

Sequencing generated more than 14.4 Gbp of raw data consisting of ~115 million 125 bp paired-end reads. The cleanup steps used to filter the raw reads reduced their number by ~2%, for a final set of 112 million high-quality reads (see Table 1 for a detailed trimming report).

Using Trinity [33, 34] raw reads were assembled *de novo* into a transcriptome containing 567,572 contigs longer than 300 bp, with an N50 value of 1,956 bp (Table 2). Using Bowtie 2 [35] read support was assessed by mapping the reads back to the assembly and found that 73% of the reads mapped back in proper pairs. The Ex90N50 statistic was 2,478 bp and therefore higher than the traditional N50 measure. To evaluate the quality of the individual contigs, we used RSEM-EVAL [37], which is displaying impact scores as an estimate of read support for each contig and its contribution to the assembly. Some 84,000 low scoring

contigs were removed from the assembly, reducing the assembly size to 382,606 contigs (Fig 1C). These initial assembly evaluation steps are critical in *de novo* transcriptome studies, because false positives (the inclusion of misassembled contigs) will lead to errors in gene prediction, annotation, and further downstream analyses such as expression profiling. However, false negatives (the elimination of legitimate contigs) can reduce the completeness of the transcriptome; thus, evaluations should be repeated after each filter step (Table 2). The final assembly was ~93% complete based on BUSCO's [36] arthropod reference database of 2,675 single-copy orthologs present in >90% of the species (Table 2), which is consistent with results from other recently published insect transcriptomes (e.g. [22]).

Table 1. Trimming statistics using three different filtering steps.

Number PE raw reads	114,824,092
Read length in bp	125
Trimmomatic	
Number low quality reads	39,096
Percent low quality reads	0.03
Reads remaining after Trimmomatic	114,784,996
SortMeRNA	
Number rRNA reads	1,244,902
Percent rRNA reads	1.10
Reads remaining after rRNA removal	113,540,094
Kraken taxonomic sequence classification system	
Reads classified as contaminants	804,313
Percent classified as contaminants	0.70
PE reads remaining after cleanup	112,534,902

Gene prediction and functional annotation

TransDecoder.LongORFs [34] identified about 93,000 potential open reading frames (ORFs) in the final *M. caerulatus* assembly. The homology-based second step retaining BLAST [51] and Pfam [52] search results in *TransDecoder.Predict* [34] resulted in a final set of 61,560 predicted proteins longer than 100 amino acids (Fig 1D, Table 2).

The continuous BLASTp search against our custom 'RefSeq' database allowed the determination of gene functions of about 73.04% of our sequences. However, the 27% that had no hit to this database and were additionally blasted against the entire non-redundant database, which produced hits for another 1%. The top hit species distribution shows the highest number of hits against the basal Hymenopterans (Symphyta) *Athalia rosae* and *Orussus abietinus* (see Fig 2A). Accuracy of our assembly and the predicted protein coding genes were supported by consistently high e-values (42% of the blast hits had e-value >1e^{-100}; see e-value distribution in A4.1 Fig).

Our gene ontology (GO) term assignment [47] via Blast2GO [44] revealed 78% of the putative genes had an InterPro hit, and 46% had a GO annotation (Fig 2B). Longer

sequences were more likely to be annotated than shorter ones (see A4): approximately 50% of the sequences >200 amino acids were annotated, and almost all of those >500 AA.

Table 2. Assembly statistics during final assembly evaluation steps

	raw assembly	filtered assembly	predicted ORFs
Assembly assessment parameters			
Transcripts > 300 bp	567,572	382,606	61,560
Total contig length	674,031,026	539,335,401	66,236,823
Mean contig size (bp)	1,187.57	1,409.64	1,075.97
Number of contigs > 1000 nt	175,803	154,692	21,023
N50 contig length	1,956	2,162	1,605
Longest contig	35,790	35,790	24,318
Percent GC	38.56	38.59	45.73
BUSCO - annotation completeness via universal single-copy orthologous genes			
Complete Single-Copy BUSCOs	2,236	2,244	2,173
Complete Duplicated BUSCOs	1,497	1,406	1,175
Fragmented BUSCOs	266	270	304
Missing BUSCOs	173	161	198
Complete BUSCOs in %	83.59	83.89	81.23
Total BUSCOs in %	**93.53**	**93.98**	**92.60**
DETONATE – RSEM-EVAL's contig impact scores			
Score	-8,182,390,224.62	-7,955,408,062.90	
Prior score on contig sequences	-934,405,410.56	-747,677,625.16	
Expected aligned reads	39,576,297.27	39,708,827.69	
Contigs with no read aligned	84,119	78	
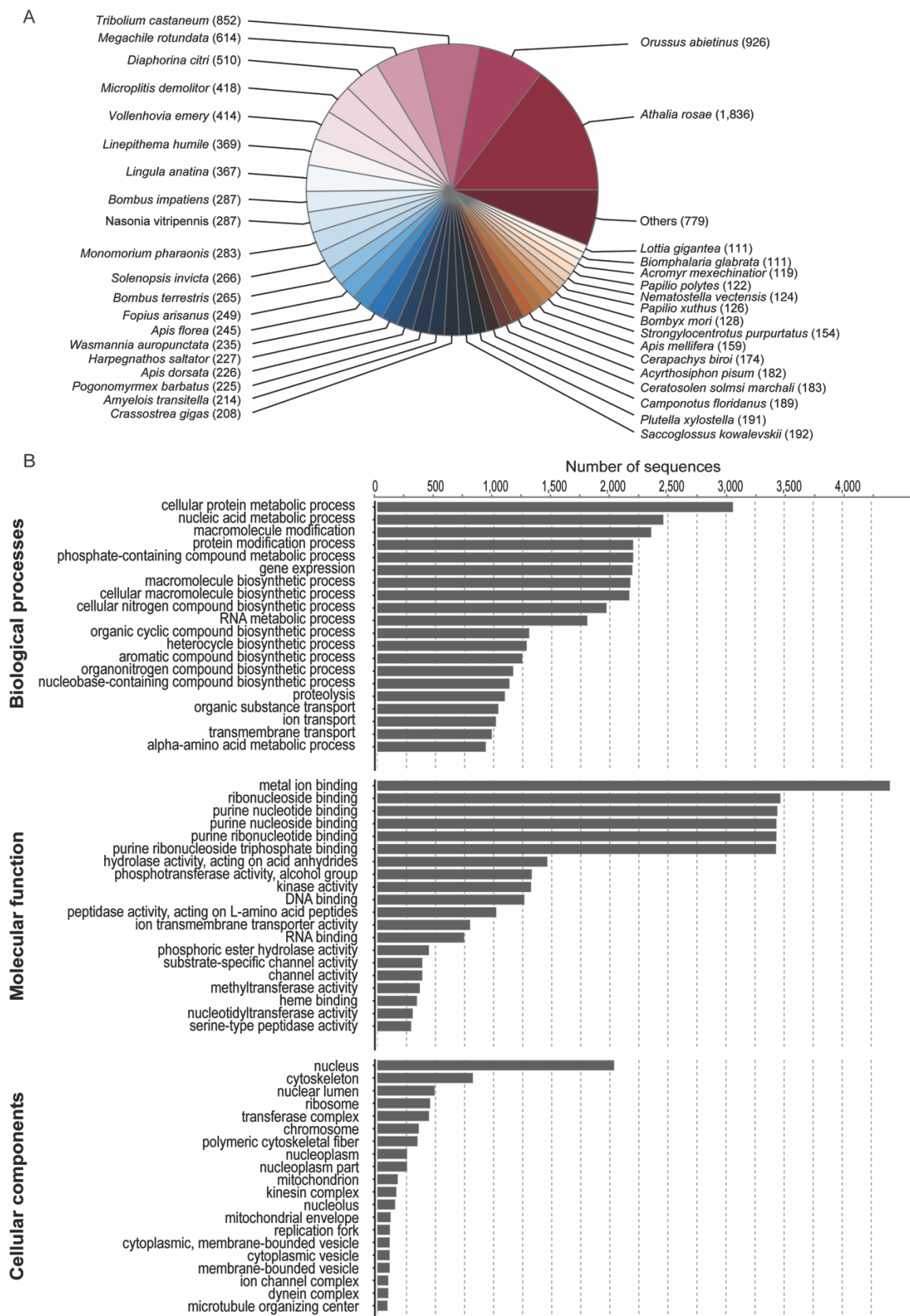
Fig 2. Functional annotation of the *M. caerulatus* transcriptome. A) Distribution of top hits shows all species to which *M. caerulatus* had at least 100 hits to. B) Classification of the functional annotation into the three gene ontology (GO) categories: molecular function (MF), cellular component (CC), and biological process (BP) at GO level 5. Displayed are the distribution of the top 20 GO terms and the number of sequences with the corresponding assignment.

Identification of candidate genes

Stress response genes. Environmental studies on insects frequently focus on heat shock proteins (HSPs), a large and highly conserved gene family involved in protein metabolism and insect survival through their roles in protein folding and repair (e.g. [53, 54]). Under cellular stress, HSP expression levels increase and their assessment in natural environments can help identify stress adaptation under climate change or habitat fragmentation [25]. We identified 23 HSP genes and 3 general stress response genes (see A4 File for the AA sequences and gene names).

Housekeeping genes. Basic cell functions are controlled by housekeeping genes, expressed in every tissue under most experimental conditions (e.g. [55]) and these serve as a baseline for normalizing quantitative real-time PCR or RNA-Seq gene expression experiments. We identified the majority of housekeeping genes commonly found in insects, including the ribosomal proteins S18 and L13a as well as ATP and actin genes (A4 File).

Developmental genes. Hox genes are of particular interest as they encode transcription factors that modulate bauplan development during early embryogenesis and determiners of cell fate (e.g. [56]). With the focus on the *Antp*-class genes, we identified *M. caerulatus* orthologs for only three *Hox* genes, including: Antennapedia (*Antp*, Hox6-8), Ultrabithorax (*Ubx*, Hox6-8) and Sex combs reduced (*Scr*, Hox5); and one *ParaHox* gene (*Nedx*). *Hox* genes were also identified in the *I. elegans* and *L. fulva* transcriptomes. An alignment of these first full-length homeodomain amino acid sequences for our 16 detected *Hox* and *ParaHox* genes in Odonata is shown in A4.4 File.

Wing genes. Insect wing development is controlled by the wing-patterning network (wing gene regulatory network) in which the *Hox* genes *Scr* and *Ubx* act jointly with cell signaling molecules, selector genes and transcription factors to modulate wing morphogenesis, differentiation and growth [2, 57]. These signaling molecules are further grouped into four main signaling pathways: Hedgehog (*Hh*), Decapentaplegic (*Dpp*), wingless (*wg*) and Notch (*N*) constituting overarching structures [5, 58, 59]. In addition, the wing-patterning network influences wing coloration, as developmental gene expression determines the activity of subsequent pigment genes (e.g. [2]).

We were able to identify most representatives of the pigmentation genes, 14 of the 21 genes described so far for the wing-patterning network, and the four main developmental signaling pathways. For the latter, we could discover 6 genes related to accurate cell differentiation and growth in the Hedgehog pathway, 8 genes that are included in the cell fate determination by the Notch pathway, 9 genes associated with the development of wings in the wingless pathway and 3 genes could be connected to dorsal/ventral patterning and development of the wing epithelia in the Decapentaplegic pathway (see Table 3 for a complete overview of wing genes and their related pathways and A4 File for the corresponding amino acid sequences). The reciprocal BLASTp searches against well-annotated patterning genes revealed an additional 19 genes described in wing coloration and

general pigmentation studies [2, 4, 60-64]. We detected pigmentation genes from the melanin pathway (*yellow, black, tan, pale*), the pteridine pathway (*henna, rosy, prat*), the ommochrome pathway (*vermillion, white, scarlet*) and pigment granule genes (*dor, garnet*). We also found phenol oxidases (*PO*), which contribute to melanization among other functions [6, 64] and the Ecdysone receptor (*EcR*), a hormone involved in wing growth [65, 66].

Comparison with other Odonata

Although direct comparative transcriptome analyses struggle with differences in sample preparation (e.g. different tissue collection, developmental stages, etc.) and by the difficulty of accurate ortholog detection, first comparisons amongst well-annotated sequences are appropriate for a selected set of questions (e.g. [13, 73]). Here, we compared our findings with the transcriptomes of *I. elegans* and *L. fulva* to search for unique gene expression. The damselfly *I. elegans* belongs to the Coenagrionidae - a sister family to the Pseudostigmatidae to which *M. caerulatus* is associated - while the more distantly related dragonfly *L. fulva* belongs to the suborder Anisoptera.

Our results reflect these relationships in that the highest overall sequence similarity is represented by the 12,569 reciprocal best hits between *M. caerulatus* and *I. elegans*. In the *M. caerulatus* / *L. fulva* search, 11,136 reciprocal best hits were obtained, similar to the results for *I. elegans* / *L. fulva* (Fig 3A). The comparison of both overlapping and unique orthologous clusters for each species and species pair showed a similar result (Fig 3B). Using OrthoVenn [50] we retrieved a total of 29,464 clusters, with 8,196 clusters containing genes from all three species. Of these, a functional annotation was available for 4,810 clusters.

To gain insights into larva-specific genes we focused on the unique orthologous clusters of *M. caerulatus*. In total 4,589 were detected, but only 1,168 clusters had a usable annotation. Those clusters were related primarily to general cell functions such as phosphorylation (serine/threonine-protein kinases, cytochrome oxidases) and signal transduction (receptor tyrosine kinases). Among potential larva-specific transcripts, we identified the following related to wings and general development: (i) *encore* regulates dorso-ventral polarity in embryos and larvae; (ii) *flightless-1* plays a structural role in indirect flight muscle; and (iii) *krueppel* is involved in gap class segmentation. Other interesting findings were the *O-mannosyl-transferase 2* that is responsible for somatic muscle development and the *Ryanodine receptor 44F* which is involved in proper muscle function, i.e., in larval body wall muscles, and is therefore essential for larval development [74, 75].
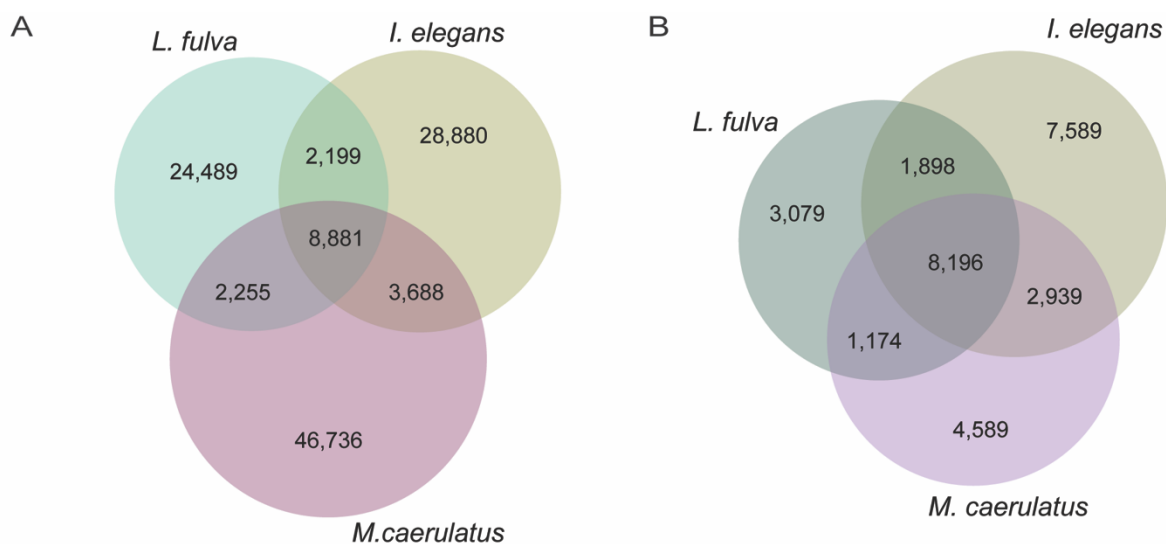
Fig 3. Comparisons among three odonate transcriptomes based on the open reading frames: ***Ladona fulva*, *Ischnura elegans*, and *M. caerulatus*.** A) The overall sequence similarity identified via reciprocal blast search among transcriptomes presented in a Venn diagram shows a greater number of overlapping genes between *M. caerulatus* and *I. elegans* than between the dragonfly *L. fulva* and the two damselflies. B) Overlapping orthologous gene clusters (OrthoVenn). Both analyses show a similar sized overlap among species.

Table 3. Selection of genes responsible for wing development in insects. Genes of the four major signaling pathways and functionally related genes ("Others") are arranged in the according columns. Members of the wing-patterning network [1, 2] known to be associated with a specific signaling pathway are shown in the grey row [3-12]. Genes identified in the *M. caerulatus* transcriptome are shown by bold; while the corresponding amino acid sequences and unigene IDs are given in A4.4 File and a description of primary gene functions are in the A4 Table. Genes that could not be identified in the *M. caerulatus* transcriptome are shown in black font.

	Hedgehog	Decapentaplegic	Notch	Wingless	Others	*Homeobox* genes
Wing-patterning network	**Hedgehog (*Hh*)**	**Decapentaplegic (*Dpp*)** **Optomotor-blind (*bi [omb]*)** **Spalt major (*salm*)** Spalt related (*salr*)	**Serrate (*Ser*)** **Cut (*cut*)** **Achaete (*ac*) / Scute (*sc*)**	**Wingless (*wg*)** **Scalloped (*sd*)** Vestigial (*vg*)	**Blistered (bs) syn: Serum Response Factor (*srf*)** **Spitz (*spi*)** Apterous (*ap*) Engrailed (*en*) Escargot (*esg*) Snail (*sna*)	**Ultabithorax (*Ubx*)** **Sex combs reduced (*Scr*)** **Extradenticle (*exd*)** Abdominal A (*Abd-A*) Distal-less (*Dll*)
	Patched (*Ptc*) **Smoothened (*Smo*)** **Cubitus interruptus (*ci*)** **Costa (*cos*)** **Fused (*Fu*)** **Suppressor of fused (*sufu*)**	**Mothers against Dpp (*mad*)** **Brinker (*brk*)** Medea (*Med*) Glass bottom boat (*Gbb*) Saxophone (*sax*) Punt (*put*) Thickveins (*tkv*)	**Notch (*N*)** **Delta (*Dl*)** **Hairless (*H*)** **Suppressor of hairless (*Su(H)*)** **Mastermind (*mam*)** **Notchless (*Nle*)** **Fringe (*fng*)** **Hairy (*h*)** Nipped-A (*Nipped-A*)	**Frizzled (*fz, fz2, fz3, fz7*)** **Disheveled (*dsh*)** **Armadillo (*arm*)** **Pangolin (*pan*)** **Wingful/Notum (*Wf*)** **Naked cuticle (*Nkd*)** **Nemo (*Nmo*)** **Axin (*Axn*)** **Adenomatous polyposis coli (*Apc*)** Arrow (*arr*) Dally (*Dlp*) Shaggy (*sgg*) GDI interacting protein 3 (*Gint3*)	**Homothorax (*hth*)** **Epidermal growth factor receptor (*Egfr*)** **Rhomboid (*rho*)** **Four-jointed (*fj*)** **Deadpan (*dpn*)** **Rap1 GTPase (*Rap1*)** **Small wing (*sl*)** **UV-resistance associated gene (*Uvrag*)** **Capricious (*caps*)** **Rickets (*rk*)** **Tolkin (*tok*)** Teashirt (*tsh*) Nubbin (*nub*) Pannier (*pnr*) Fat (*ft*) Gilgamesh (*gish*) Homeodomain interacting protein kinase (*Hipk*) Knirps (*kni*) Extra macrochaetae (*emc*) Net (*net*) Ventral veins lacking (*vvl*)	

Discussion

At the base of flying insects, Odonata have a long-standing record in ecological and evolutionary research. This head start should encourage their future role as non-model systems in integrative genomic research. The first transcriptome profiling of a larval tissue from *Megaloprepus caerulatus* represents a step towards this direction to study wing development and evolution, and speciation.

Transcriptome assembly and functional annotation

Odonate genomes are among the larger known genomes within winged insects [76], and dragonfly and damselfly transcriptomes likewise appear to be larger than those of prominent model species such as Diptera or Lepidoptera. Our assembly of raw sequences resulted in over 500,000 putative transcripts, which were reduced via strict evaluation to 61,560 high quality protein-coding genes. This number and the ExN50 expression value reflect some redundancy, but the transcriptome size is comparable to that of other odonates (e.g. [21, 22]). Furthermore, in comparison to previous odonate studies, function could be assigned to a greater number of genes which may reflect growing resources in genomics and the use of customized and frequently updated reference databases [77].

Beyond this, increasing tissue diversity should in turn increase the number of genes sequenced and annotated, but difficulties with the assembly of heterozygous sequences can limit the quality of the reconstructed transcripts and thereby impair the reliability of BLAST results. In spite of the comparatively successful assignment of function, some 25% of the putative proteins lack annotation. Some of these genes are probably misassembled transcripts that do not actually exist or, alternatively, represent odonate- or *Megaloprepus*-specific proteins that simply lack homologous sequences in current databases.

Candidate genes

Little is known about developmental genes such as *Hox* genes or those responsible for the pathways of wing development and coloration in odonates (but see [56]). In accordance with our expectations we found three *Hox* genes in the larval thorax. Interestingly, six *Hox* representatives could be detected in the adult *I. elegans* (A4 File), suggesting that *Hox* expression may be of functional importance in adults as well as larvae. Two of the *Hox* genes identified in *M. caerulatus* are involved in important wing traits: *Scr* suppresses wing development in the prothorax [1], while *Ubx* controls hind wing identity [72] and is an important modulator in the wing-patterning gene regulatory network [2]. It acts as a selector gene, influencing morphological characters such as wing venation and regulates wingless (*wg*), splat (*sal*) and vestigial (*vg*) in opposing mechanisms [71]. In *Drosophila* it facilitates the development of halteres and in *Tribolium* the sclerotized fore wings [72].

The development of wings and their shape is controlled by the wing-patterning network through the modulation of gene expression [6, 78]. It was originally described in *Drosphila melanogaster* and is supposedly largely conserved across holometabolous insects [2, 57]. However, in hemimetabolous insect orders information on wing differentiation across larval stages is limited [79]. In the *Megaloprepus* transcriptome, we identified 14 genes from the

wing-patterning network. So far *Dpp* has been described to inhibit *Dll* (Distal-less) in an early stage of the signaling cascade with the wing-patterning network, but later in the development of imaginal wing discs it activates *omb* (Optomotor-blind), *sal* (spalt) and *vg* (vestigial) to shape cell growth, vein positioning and intervein cell differentiation [2, 57]. In the pupal stage of holometabolous insects, a significant reorganization of tissues and organs takes place, while hemimetabolous insects undergo a more gradual developmental transition. Thus, some of the mechanisms of wing development in Odonata most likely differ from those of holometabolous insects, and further investigation of the timing and related genes may shed light onto the developmental changes that characterize the bauplan transition to holometaboly [1].

Wing coloration in odonates is highly variable across species. Some have only a colored pterostigma or different sized wing spots, while some other species show entirely colored wings. We identified 19 genes related to insect pigmentation. Furthermore, our data showed a higher relative expression of both, the *phenol oxidases* and *yellow* in comparison to the house keeping genes (S5 File). This could be an indication of polymerization of cuticular pigments following larval molt. However, since the larva was collected from its natural environment, this remains an assumption pending controlled experiments under laboratory settings to identify the genes and pathways responsible for coloration. Targeted RNA sequencing in parallel with in situ hybridization studies would thus provide deeper insights into gene expression during the course of odonate development.

Finally, some of the wing and developmental genes that were not identified in our analysis may simply lack expression at the time of collection, but most likely may indicate modified signaling pathways. However, we suggest that the genes identified here reflect an early stage of wing development, also because the *Megaloprepus* larva used bore visible wing buds on its thorax. In our comparison between the larvae and the two adult odonates, we found aside of the wing genes, proteins known as essential for larval development (*O-mannosyltransferase*, *Ryanodine receptor 44F*). However, many of those ~4,500 transcripts found only in *M. caerulatus* may be species-specific rather than larva-specific.

Conclusion

Megaloprepus caerulatus has a longstanding record in ecological and evolutionary research [19, 80-84]. The de novo transcriptome presented here is the first genomic resource for Neotropical odonates and may hopefully enhance future genomic research in odonates. For *M. caerulatus* comparative studies at different developmental stages involving the newly discovered species might reveal mechanisms of wing shape divergence, demographic patterns, micro-evolutionary changes and genomic regions under selection in changing environments.

Because of its close ecological association with Neotropical old growth rainforests, high vulnerability to climate shifts and forest disturbances, *Megaloprepus* is an effective bioindicator of the history (and future) of old growth rainforests. Genomic monitoring of

key genes combined with ecological data could provide early insights into the effects of environmental changes.

References

1. Engel MS, Davis SR, Prokop J. Insect wings: the evolutionary development of nature's first flyers. Arthropod Biology and Evolution: Springer; 2013. p. 269-98.
2. Connahs H, Rhen T, Simmons RB. Transcriptome analysis of the painted lady butterfly, *Vanessa cardui* during wing color pattern development. BMC genomics. 2016;17(1):1.
3. Ferguson LC, Maroja L, Jiggins CD. Convergent, modular expression of ebony and tan in the mimetic wing patterns of *Heliconius butterflies*. Development genes and evolution. 2011;221(5-6):297-308.
4. Kronforst MR, Papa R. The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. Genetics. 2015;200(1):1-19.
5. Liu S, Wei W, Chu Y, Zhang L, Shen J, An C. De novo transcriptome analysis of wing development-related signaling pathways in *Locusta migratoria manilensis* and *Ostrinia furnacalis* (Guenée). PloS one. 2014;9(9):e106770.
6. Massey J, Wittkopp PJ. The genetic basis of pigmentation differences within and between *Drosophila* species. Current topics in developmental biology. 2016;119:27-61. doi: 10.1016/bs.ctdb.2016.03.004. PubMed PMID: 27282023; PubMed Central PMCID: PMCPMC5002358.
7. Roux J, Rosikiewicz M, Robinson-Rechavi M. What to compare and how: Comparative transcriptomics for Evo-Devo. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2015;324(4):372-82.

8. Oppenheim SJ, Baker RH, Simon S, DeSalle R. We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. Insect molecular biology. 2015;24(2):139-54.

9. Abouheif E, Favé M-J, Ibarrarán-Viniegra AS, Lesoway MP, Rafiqi AM, Rajakumar R. Eco-evo-devo: The time has come. Ecological genomics: Springer; 2014. p. 107-25.

10. Gilbert SF, Bosch TC, Ledón-Rettig C. Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. Nature Reviews Genetics. 2015;16(10):611-22.

11. Alvarez M, Schrey AW, Richards CL. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Molecular ecology. 2015;24(4):710-25.

12. Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, et al. Conserved developmental transcriptomes in evolutionarily divergent species. Genome biology. 2010;11(3):R35.

13. Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, et al. Comparative analysis of the transcriptome across distant species. Nature. 2014;512(7515):445-8.

14. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346(6210):763-7.

15. Simon S. A Brief Guide to the evolution of flying insects. BioEssyas. 2013.

16. Simon S, Strauss S, von Haeseler A, Hadrys H. A phylogenomic approach to resolve the basal pterygote divergence. Molecular biology and evolution. 2009;26(12):2719-30.

17. Córdoba-Aguilar A. Dragonflies and Damselflies: Model Organisms for Ecological and Evolutionary Research. Oxford: Oxford University Press; 2008.

18. Bybee S, Córdoba-Aguilar A, Duryea MC, Futahashi R, Hansson B, Lorenzo-Carballa MO, et al. Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. Frontiers in zoology. 2016;13(1):46.

19. Feindt W, Fincke O, Hadrys H. Still a one species genus? Strong genetic diversification in the world`s largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*. Conservation Genetics. 2014;15(2):469-81.

20. van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-5.

21. Shanku AG, McPeek MA, Kern AD. Functional Annotation and Comparative Analysis of a Zygopteran Transcriptome. G3: Genes| Genomes| Genetics. 2013;3(4):763-70.

22. Chauhan P, Hansson B, Kraaijeveld K, de Knijff P, Svensson EI, Wellenreuther M. De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. BMC genomics. 2014;15(1):808.

23. Chauhan P, Wellenreuther M, Hansson B. Transcriptome profiling in the damselfly *Ischnura elegans* identifies genes with sex-biased expression. BMC genomics. 2016;17(1):985.

24. Simon S, Sagasser S, Saccenti E, Brugler MR, Schranz ME, Hadrys H, et al. Comparative transcriptomics reveals developmental turning points during embryogenesis of a hemimetabolous insect, the damselfly *Ischnura elegans*. Scientific Reports. 2017;7(1):2045-322.

25. Lancaster LT, Dudaniec RY, Chauhan P, Wellenreuther M, Svensson EI, Hansson B. Gene expression under thermal stress varies across a geographic range expansion front. Molecular ecology. 2016.

26. Ioannidis P, Simao FA, Waterhouse RM, Manni M, Seppey M, Robertson HM, et al. Genomic Features of the Damselfly *Calopteryx splendens:* Representing a Sister Clade to Most Insect Orders. Genome biology and evolution. 2017;9(2):415-30.

27. Yanoviak S, Fincke O. Sampling methods for water-filled tree holes and their artificial analogues. Insect sampling in forest ecosystems Blackwell Publishing, London. 2005:168-85.

28. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source. 2010.

29. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:btu170.

30. Tripp HJ, Hewson I, Boyarsky S, Stuart JM, Zehr JP. Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. Nucleic acids research. 2011;39(20):8792-802. doi: 10.1093/nar/gkr576. PubMed PMID: PMC3203614.

31. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.

32. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.

33. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology. 2013;29(7):644-52. doi: 10.1038/nbt.1883. PubMed PMID: 21572440; PubMed Central PMCID: PMCPmc3571712.

34. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols. 2013;8(8). Epub 1512.

35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

36. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:btv351.

37. Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, et al. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome biology. 2014;15(12):553.

38. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.

39. Racine JS. RStudio: A Platform-Independent IDE for R and Sweave. Journal of Applied Econometrics. 2012;27(1):167-72.

40. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80.

41. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. Plant Bioinformatics: Methods and Protocols. 2016:23-54.

42. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research. 2016;44(D1):D279-D85. doi: 10.1093/nar/gkv1344.

43. Wheeler T, Eddy S. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29(19):2487-9. doi: 10.1093/bioinformatics/btt403. PubMed PMID: 23842809; PubMed Central PMCID: PMCPmc3777106.

44. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.

45. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 2008;36(10):3420-35.

46. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17(9):847-8.

47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.

48. Zhong YF, Butts T, Holland PW. HomeoDB: a database of homeobox gene diversity. Evolution & development. 2008;10(5):516-8.

49. Zhong Yf, Holland PW. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evolution & development. 2011;13(6):567-8.

50. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic acids research. 2015;43(W1):W78-W84.

51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403-10.

52. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic acids research. 2011:gkr1065.

53. Hendrick JP, Hartl F-U. Molecular chaperone functions of heat-shock proteins. Annual review of biochemistry. 1993;62(1):349-84.

54. Zhao L, Jones W. Expression of heat shock protein genes in insect stress responses. Invertebrate Surviv J. 2012;90:93-101.

55. Yang C, Pan H, Liu Y, Zhou X. Stably expressed housekeeping genes across developmental stages in the two-spotted spider mite, *Tetranychus urticae*. PloS one. 2015;10(3):e0120833.

56. Hadrys H, Simon S, Kaune B, Schmitt O, Schöner A, Jakob W, et al. Isolation of hox cluster genes from insects reveals an accelerated sequence evolution rate. PloS one. 2012;7(6):e34682.

57. Abouheif E, Wray GA. Evolution of the gene network underlying wing polyphenism in ants. Science. 2002;297(5579):249-52.

58. Shimmi O, Matsuda S, Hatakeyama M, editors. Insights into the molecular mechanisms underlying diversified wing venation among insects. Proc R Soc B; 2014: The Royal Society.

59. Yan S-J, Gu Y, Li WX, Fleming RJ. Multiple signaling pathways and a selector protein sequentially regulate *Drosophila* wing development. Development. 2004;131(2):285-98.

60. Kronforst MR, Barsh GS, Kopp A, Mallet J, Monteiro A, Mullen SP, et al. Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. Pigment cell & melanoma research. 2012;25(4):411-33.

61. Nishikawa H, Iga M, Yamaguchi J, Saito K, Kataoka H, Suzuki Y, et al. Molecular basis of wing coloration in a Batesian mimic butterfly, *Papilio polytes*. Scientific reports. 2013;3:3184.

62. Reed RD, Nagy LM. Evolutionary redeployment of a biosynthetic module: expression of eye pigment genes *vermilion*, *cinnabar*, and *white* in butterfly wing development. Evolution & development. 2005;7(4):301-11.

63. True JR. Insect melanism: the molecules matter. Trends in ecology & evolution. 2003;18(12):640-7.

64. Wittkopp PJ, Carroll SB, Kopp A. Evolution in black and white: genetic control of pigment patterns in *Drosophila*. TRENDS in Genetics. 2003;19(9):495-504.

65. Herboso L, Oliveira MM, Talamillo A, Pérez C, González M, Martín D, et al. Ecdysone promotes growth of imaginal discs through the regulation of Thor in *D. melanogaster*. Scientific reports. 2015;5:12383.

66. Monteiro A, Tong X, Bear A, Liew SF, Bhardwaj S, Wasik BR, et al. Differential expression of ecdysone receptor leads to variation in phenotypic plasticity across serial homologs. PLoS Genet. 2015;11(9):e1005529.

67. Bray S. *Drosophila* development: Scalloped and Vestigial take wing. Current biology. 1999;9(7):R245-R7.

68. Kubota K, Goto S, Eto K, Hayashi S. EGF receptor attenuates Dpp signaling and helps to distinguish the wing and leg cell fates in *Drosophila*. Development. 2000;127(17):3769-76.

69. Ou J, Deng H-M, Zheng S-C, Huang L-H, Feng Q-L, Liu L. Transcriptomic analysis of developmental features of *Bombyx mori* wing disc during metamorphosis. BMC genomics. 2014;15(1):820.

70. Swarup S, Verheyen EM. Wnt/wingless signaling in *Drosophila*. Cold Spring Harbor perspectives in biology. 2012;4(6):a007930.

71. Tomoyasu Y. Ultrabithorax and the evolution of insect forewing/hindwing differentiation. Current Opinion in Insect Science. 2017;19:8-15.

72. Tomoyasu Y, Wheeler SR, Denell RE. *Ultrabithorax* is required for membranous wing identity in the beetle *Tribolium castaneum*. Nature. 2005;433(7026):643-7.

73. Berens AJ, Hunt JH, Toth AL. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. Molecular biology and evolution. 2014;32(3):690-703.

74. Lyalin D, Koles K, Roosendaal SD, Repnikova E, Van Wechel L, Panin VM. The twisted gene encodes *Drosophila* protein O-mannosyltransferase 2 and genetically interacts with the rotated abdomen gene encoding *Drosophila* protein O-mannosyltransferase 1. Genetics. 2006;172(1):343-53.

75. Sullivan KM, Scott K, Zuker CS, Rubin GM. The ryanodine receptor is essential for larval development in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 2000;97(11):5942-7.

76. Ardila-Garcia A, Gregory T. An exploration of genome size diversity in dragonflies and damselflies (Insecta: Odonata). Journal of Zoology. 2009;278(3):163-73.
77. Wadi L, Meyer M, Weiser J, Stein LD, Reimand J. Impact of knowledge accumulation on pathway enrichment analysis. bioRxiv. 2016:049288.
78. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. Nature reviews Genetics. 2009;10(2):141.
79. Medved V, Marden JH, Fescemyer HW, Der JP, Liu J, Mahfooz N, et al. Origin and diversification of wings: Insights from a neopteran insect. Proceedings of the National Academy of Sciences. 2015;112(52):15946-51.
80. Selys Longchamps Ed. Synopsis des Agrionines. Première Légion. – Pseudostigma –. Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux - Arts de Belgique. 1860;2(10):9-27.
81. Fincke OM, Hedström I. Differences in forest use and colonization by Neotropical tree-hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conversion. Studies on Neotropical Fauna and Environment. 2008;43(1):35-45.
82. Fincke OM. The population ecology of *Megaloprepus coerulatus* and its effect on species assemblages in water-filled tree holes. Insect populations in theory and in practice: Springer; 1998. p. 391-416.
83. Fincke O. Behavioral ecology of the giant damselflies of Barro Colorado Island, Panama (Odonata: Zygoptera: Pseudostigmatidae). In: Quintero D, Aiello A, editors. Insects of Panama and Mesoamerica: selected studies. Oxford, UK: Oxford University Press; 1992. p. 102-13.
84. Fincke OM, editor Use of forest and tree species, and dispersal by giant damselflies (Pseudostigmatidae): their prospects in fragmented forests. Forest and dragonflies 4th WDA International Symposium of Odonatology Pensoft, Sofia; 2006.

Good citizenship made easy: A step-by-step guide to submitting RNA-Seq data to NCBI

Wiebke Feindt[1*], Sara J. Oppenheim[2], Robert DeSalle[2] and Shaadi Mehr[2, 3*]

[1] ITZ, Division of Ecology and Evolution, University of Veterinary Medicine Hannover,
Bünteweg 17d, 30559 Hannover, Germany
[2] Sackler Institute for Comparative Genomics, American Museum of Natural History,
Central Park West and 79th St., New York, NY, 10024, USA
[3] Multiple Myeloma Research Foundation, 383 Main Ave # 5, Norwalk, CT 06851, USA

* Corresponding Authors:
wiebke.feindt@ecolevol.de, smehr@amnh.org

Good citizenship made easy: A step-by-step guide to submitting RNA-Seq data to NCBI

SIGNIFICANCE STATEMENT
The submission of sequencing data to the National Centre for Biotechnology Information (NCBI) is a pre-requisite for publication in most scientific journals and is, more fundamentally, an essential component of the transparency and data availability that drive scientific progress. Unfortunately, the process of submitting raw reads and the resulting assemblies to NCBI is complex, especially for first-time users. Here, we present step-by-step protocols to facilitate the timely submission of data by researchers. For each required step, we provide a clear list of requirements and provide easy-to-follow examples of all commands. We hope this contribution will allow scientists to more easily share the data they have worked so hard to generate and make NCBI an even more valuable resource than it is today.

ABSTRACT
The analysis of transcriptome data from non-model organisms contributes to our understanding of diverse aspects of evolutionary biology, including developmental processes, speciation, adaptation, and extinction. Underlying this diversity is one shared feature, the generation of enormous amounts of sequence data. Data availability requirements in most journals oblige researchers to make their raw transcriptome data publicly available, and the databases housed at the National Center for Biotechnology Information (NCBI) are a popular choice for data deposition. Unfortunately, the successful submission of raw sequences to the Sequence Read Archive (SRA) and transcriptome assemblies to the Transcriptome Shotgun Assembly (TSA) can be challenging for novice users, significantly delaying data availability and publication. Here we present two comprehensive protocols for submitting RNA-Seq data to NCBI databases, accompanied by an easy-to-use web site that facilitates the timely submission of data by researchers of any experience level.

INTRODUCTION
Next Generation Sequencing (NGS) has facilitated an unprecedented leap forward in all areas of molecular and evolutionary biology research, cancer research, and systems biology. The vast amount of data obtained from NGS studies has the potential to change the way we understand genomic and phenotypic variation, organismal development, speciation, and the

origin and maintenance of biodiversity. NGS data, particularly transcriptome sequences, can affordably be generated by even modestly funded laboratories, leading to an explosion in the number of RNA sequencing (RNA-Seq) studies published over the past several years (Oppenheim et al., 2015). The goals of such studies are diverse, and include the generation of EST databases (e.g. Jeukens et al., 2010; Kumar & Blaxter, 2010), functional annotation (e.g. Crawford et al., 2010), SNP discovery (e.g. Novaes et al., 2008; Parchman et al., 2010), novel gene identification (e.g. Feindt et al., 2018; Gruber et al., 2015; Mehr et al., 2013; Oppenheim et al., 2018), and gene expression profiling (Alves-Carvalho et al., 2015; Lopez-Maestre et al., 2016; Simon et al., 2017). Yet in spite of this diversity, all RNA-Seq studies share one feature: the generation of enormous amounts of sequence data.

While the number of research questions that can be addressed with transcriptome data is nearly infinite, individual researchers and research groups will typically be interested in a specific question that they will address with RNA-Seq data. As a result, most RNA-Seq data is vastly underutilized, because a data set that was generated to answer questions about, for example, the evolution of insect wings could also be used in the identification of insect-specific genes or for phylogenomic analyses of metazoans. One of the finest characteristics of the National Center for Biotechnology Information (NCBI Resource Coordinators, 2016) and other public databases is that large-scale comparative studies can be undertaken with little or no additional sequencing required. Given that the majority of evolutionary biology studies rely partly or wholly on public funding, openly shared, easily accessible raw data is the foundation on which all NGS studies should be built.

The timely submission of raw reads and resulting transcriptome assemblies into the NCBI databases is vital to ensuring the availability of high-quality data for downstream analyses and the transparency of published research. The NCBI databases most directly relevant to RNA-Seq data are the Sequence Read Archive (SRA), where raw sequencing reads are stored, and the Transcriptome Shotgun Assembly (TSA) Sequence Database, which allows researchers to share the assembly upon which their reported analyses are based. Although conceptually straightforward, the submission procedure can be quite cumbersome and may take novice users several weeks to complete. Because barriers to completing this vital process can significantly delay data availability and publication, there is an urgent need for clear guidance.

Although we recognize that other guides exist—ranging from informal "cheatsheets" on personal websites to the very thorough but dispersed information available at NCBI—we ourselves have felt the need for a detailed, all-in-one-place, presentation of the steps required to successfully submit RNA-Seq data to NCBI. Here we present two comprehensive protocols covering all the steps needed to upload raw RNA-Seq data and assembled transcriptomes to NCBI (Fig. 1) and provide detailed examples of all the underlying commands. It is provided as a complement to resources available on the NCBI webpage (*https://submit.ncbi.nlm.nih.gov/*). Finally, the accompanying website (see Internet Resources: *http://desalle.amnh.org/good_citizenship_rna*) provides detailed instructions, along with visual aids and examples of all necessary commands. The website also provides sample data files and a checklist.
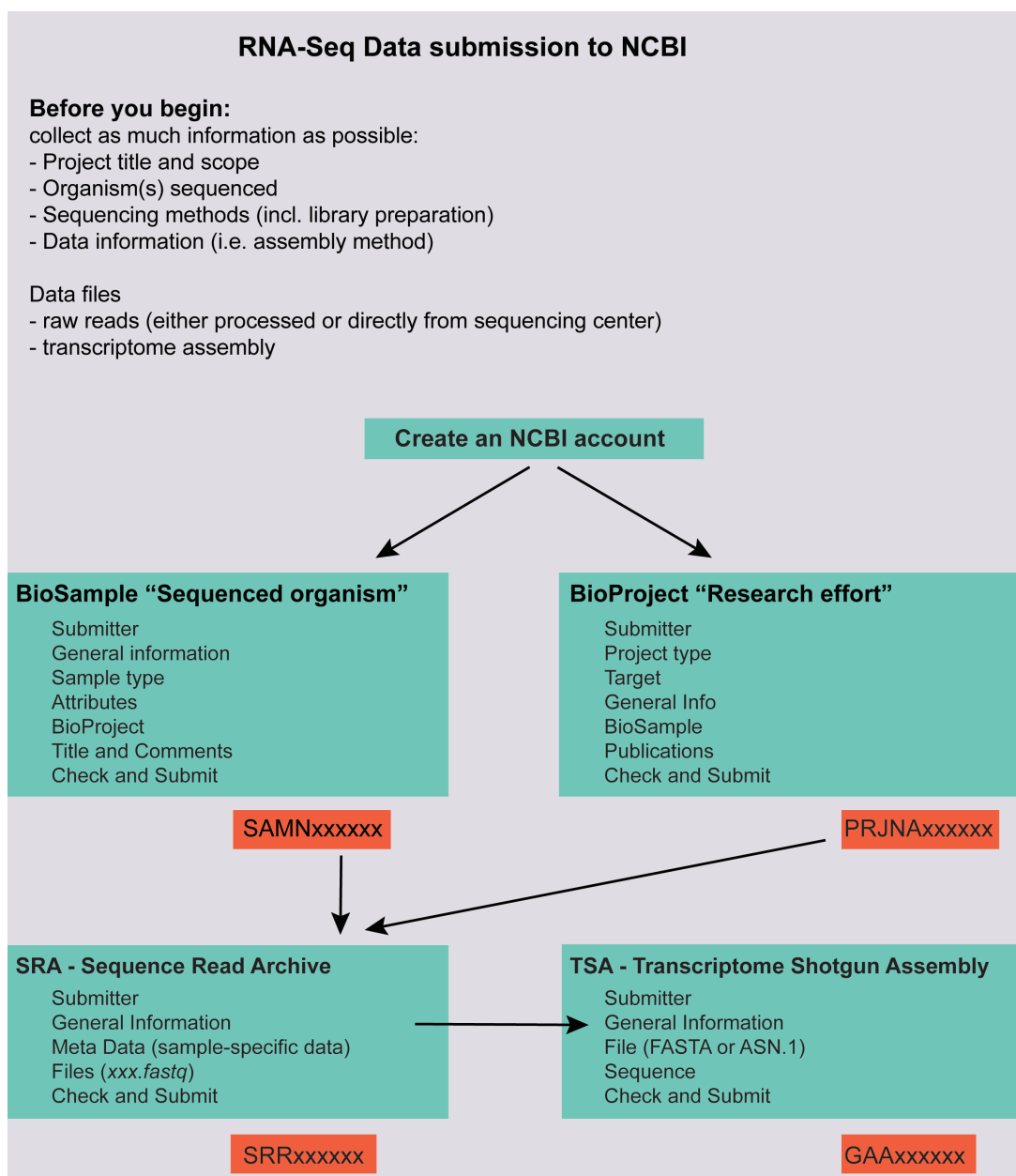
Figure 1: Overview of all necessary steps for submission of raw sequence reads and Transcriptome assemblies to NCBI.

BASIC PROTOCOL 1: STEP – BY – STEP GUIDE TO SEQUENCE READ ARCHIVE (SRA) SUBMISSION

This protocol is a step-by-step guide which will allow users to upload raw RNA-Seq data to NCBI's SRA. It details all the necessary steps, namely the creation of an NCBI user account, the registration of a BioProject and a BioSample, and the submission of sequence data. Four ways of uploading raw reads are described: direct upload with or without Aspera and command line upload via Aspera or FTP. For each step we describe the information the user has to provide and present useful commands as well as links to additional Internet resources.

Necessary Resources

Hardware:

- Unix, Windows, or Macintosh workstation with an Internet connection and web browser (Firefox, Chrome, Safari or Internet Explorer).

Software:

- Command line user interface (CLI) such as Terminal, iTerm, ConEmu, Console2, cmder, Cygwin, PuTTY, etc.
- Text editor (TextWrangler, BBEdit, Sublime Text, Notepad, Emacs, etc.)
- *Aspera* connect high-speed file transfer software: For command line version: go to *http://downloads.asperasoft.com/en/downloads/8?list* and chose the version for your operating system. For web-based version: download the plugin from *http://downloads.asperasoft.com/connect2//*.

Data:

- Sequence read data in standard FASTQ format
- Biological information about the sequenced organism
- Technical information about how the sequencing library was prepared

Create an NCBI account

1. Only registered users can submit data to NCBI. Create a user account at *https://www.ncbi.nlm.nih.gov/account/register/*. Registration requires choosing a user name and password and providing a security question (for password recovery) and an email address. Upon successful completion, the user will receive a confirmation e-mail from NCBI; clicking on the link provided will complete the registration process.

   The steps in the NCBI submission process have a modular structure, where the user is forwarded from tab to tab within the submission portal and can also navigate between tabs. Complete the mandatory fields that are marked with an asterisk, many fields are accompanied by a "?" icon that provides information about what should be entered (Fig. 2).

Create a BioSample

A BioSample ID is required for all SRA and TSA submissions (see Fig. 1). The BioSample information describes the biological source material used in the sequencing project. Also see the information provided by NCBI concerning BioSamples (https://www.ncbi.nlm.nih.gov /sra/docs/submitbio/).

2. Go to *https://submit.ncbi.nlm.nih.gov/subs/biosample/* and sign in to NCBI with your user account. Click on the *BioSample* link and choose *New submission.* During this step, the submitter should provide as much information as possible about the studied organism. It is difficult to edit this information after the process is complete, so users should carefully proofread all fields before submitting. The BioSample submission progresses through seven fillable forms (presented as tabs at the top of the page). In order, these are:

Figure 2: Tabular format of all NCBI submission steps. The example shown is for the creation of a BioSample.

a. *Submitter:* Provide information about the person submitting the data and the submitting organization (typically, this will be the submitter's organizational affiliation). An email address from the submitting organization's domain is required. If desired, a shared submission group can be created, allowing multiple authors to access and contribute to the submission.

b. *General info:* Here the user assigns a release date for the data, which can occur immediately upon submission or be delayed until publication (or until a specified future date). In addition, the user must choose between *Single BioSample* or

Batch/Multiple BioSamples submission. If a batch submission is selected, only samples that are part of the same project should be included.

c. *Sample type:* Here the user chooses among ten options giving a general description of the sample type. Researchers working with non-model invertebrates should choose *Invertebrate*, those working on model or non-model plants should choose *Plants*, those working on canonical model animals (e.g. *D. melanogaster* and *C. elegans*) should choose *Model organism or animal sample*, and those working on any non-model non-invertebrate animal should also choose *Model organism or animal sample*. There are a variety of more specialized descriptors available for metagenomes or pathogens that should be chosen if appropriate.

d. *Attributes:* If *Single BioSample* was chosen, this page is a fillable form. If *Batch/Multiple BioSamples* was selected, the user is prompted to download a fillable template file. In either case, the following fields are mandatory:

  i. *sample_name:* a short, unique descriptor of the sequenced sample; organism the scientific name of the organism to the most specific level available (standard "*Genus species*" if possible);

  ii. *collection_date:* the date when the sample was collected, from the field or lab as appropriate—basically the date the organism was sacrificed;

  iii. *geo_loc_name:* the site where the specimen was collected, in the general format Country:State:City;

  iv. *tissue:* the specific tissue from which RNA was extracted for sequencing.

  If any mandatory information is missing, user should enter "not collected", "not applicable" or "missing" as appropriate. Additional fields (e.g., sex, developmental stage, age, latitude/longitude of collection site) are available but not mandatory. The user is encouraged to provide as much data as possible. If Batch/Multiple BioSamples was selected, the filled template file must be saved as a TSV or TXT file before uploading—Excel format files will be rejected.

e. *BioProject:* The user may provide a BioProject ID at this point, but it is best to continue without entering a BioProject ID and link the BioSample and BioProject at a later stage.

f. *Description:* These fields will be visible to the public once your submission is complete and released for public viewing. Choose concise, descriptive titles.

g. *Overview:* Here the user can look over all the submission parts and decide if it is ready to submit.

Note: Before hitting "submit" check carefully for any errors—once the submission is complete, changes can only be made by emailing the BioSample help desk!

3. Once the user is satisfied with all entries, hit *Submit*. After the request is processed, the user will receive a confirmation email containing the BioSample ID(s) in the format SAMNxxxxxx.

Create a BioProject

A BioProject ID is required for all SRA and TSA submission (see Fig. 1). The BioProject can include diverse data types (e.g., the genome and transcriptome of a single species; transcriptomes from different life stages or strains of a single species). Thus, a single BioProject record can contain multiple BioSamples. Also see the information provided by NCBI concerning BioProjects (https://www.ncbi.nlm.nih.gov/sra/docs/submitbio/).

4. Go to *https://submit.ncbi.nlm.nih.gov/subs/bioproject/* and sign in to NCBI with your user account. Click on the *BioProject* link and choose *New submission.* During this step, the submitter will provide information about the sequencing project, including organism information and funding sources. The submission process progresses through seven fillable forms (presented as tabs at the top of the page). In order, these are:

   a. *Submitter:* Provide information about the person submitting the data and the submitting organization (typically, this will be the submitter's organizational affiliation). An email address from the submitting organization's domain is required. If desired, a shared submission group can be created.

   b. *Project type:* Many Project data types are available; at this stage, the user should select *Raw sequence reads.* Also on this tab, the user must define the "sample scope", which will depend upon the biological entities included in the study. *Monoisolate* studies involve a single individual OR projects that result in a single transcriptome assembly, even if multiple individuals were pooled for sequencing; *Multiisolate* studies involve multiple individuals or strains from a single species; *Multi-species* studies involve individuals from 2+ different species; *Environment* studies involve environmental samples (water, soil, etc.) whose species content is unknown; *Synthetic* studies involve synthetically created samples; and *Other* is a catchall that allows the user to manually define the sample scope.

   c. *Target:* Here information about the organism (or group of organisms) sequenced can be entered. The *Organism name* should be broad enough to cover all the included species (if multiple species within *Genus* are included, use *Genus* as the organism name).

   d. *General info:* Here the user assigns a release date for the data, which can occur immediately upon submission or be delayed until publication (or until a specified future date). In addition, the user must choose a *Project title* and a *Public description.* These fields will be visible to the public once your submission is complete and released for public viewing. The user here defines the *Relevance* of the BioProject (Agricultural, Evolution, etc.). Additional (optional) fields are available on this tab: *External Links*, where the user can provide relevant web pages (such as a project page for the sequencing project or a personal research page describing the research goals); *Select Your Grants*, where the user can provide funding information; *Consortium name*, if the study is part of a consortium project; and, *Data provider*, if the data were not generated by the submitter.

   e. *BioSample:* Here the user can provide a previously created BioSample ID. However, only single BioSamples can be registered this way. If the user's project involves

multiple samples (i.e. if it was a *Batch/Multiple BioSamples* submission), the user should complete the BioProject creation without including any BioSample information and then submit the BioSamples separately, including the BioProject ID in the raw read submission (see from step 6 onwards).

f. *Publications:* If publications related to the BioProject already exist, they can be entered here. Future publications can be added at a later date.

g. *Overview:* Here the user can review all the BioProject components and decide if it is ready to submit.

Note: Check carefully for any errors—once the submission is complete, changes can only be made by emailing the BioProject help desk!

5. Once the user is satisfied with all entries, hit *Submit*. Once the request is processed, the user will receive a confirmation email containing your Submission Title, Submission ID, and BioProject ID (in the form `PRJNAxxxxxx`).

SRA submission

At this point, the user is ready to submit the raw reads that are associated with the BioProject. Sequences can be submitted exactly as received from the sequencing center. Alternatively, processed reads (i.e. those that have been subjected to quality trimming, contaminant removal, etc.) may be submitted. If processed reads are submitted, is important to include information about all the processing procedures applied. Additional information on SRA submission can be found at NCBI (https://www.ncbi.nlm.nih.gov/ sra/docs/, https://www.ncbi.nlm.nih.gov/books/NBK51157/).

6. Go to *https://submit.ncbi.nlm.nih.gov/subs/sra/* and sign in to NCBI with your user account. Choose *New submission.* During this step, the submitter will provide information about the sequencing project, including organism information, funding sources, etc. The submission process progresses through five fillable forms (presented as tabs at the top of the page). In order, these are:

a. *Submitter:* Provide information about the person submitting the data and the submitting organization (typically, this will be the submitter's organizational affiliation). An email address from the submitting organization's domain is required.

b. *General info:* Here the user will enter the BioProject ID created above, indicate whether BioSample IDs have been created, and choose a release date. If *Release immediately following processing* is selected, the raw data will become publicly accessible right away. If the user wishes to delay release, a future release date must be selected.

c. *SRA metadata:* Here the submitter must provide a metadata file containing information about the sequencing procedures used. A template file can be

downloaded from the SRA site as a tab-delimited file or as an Excel file. The Excel file is easier to work with and provides helpful details. In either case, the user must save the edited template (the sheet called "SRA_Data" in the Excel file) as a tab-delimited text file. To save the SRA Data worksheet as a text file, use "Save As" "Tab Delimited Text (`.txt`)". The metadata template has 17 fields. The following 13 fields are mandatory for all data types (Fig. 3):

i. *bioproject_accession:* The BioProject ID associated with the raw reads.

ii. *biosample_accession:* The BioSample ID(s) associated with the raw reads.

iii. *library_ID:* A user-defined unique identifier. Each sequencing library must have its own unique ID.

iv. *title:* A short, publicly viewable description of the data. NCBI recommends the format "*<methodology>* of *<organism>*: *<sample info>*" (e.g. "RNA-Seq of *Drosophila melanogaster*: adult female antennae").

v. *library_strategy:* The user must choose from a provided set of options. For most transcriptome studies, the user should choose *RNA-Seq*.

vi. *library_source:* The nucleic acid type that was used to prepare the library. For most transcriptome studies, the user will choose "transcriptomic", but "metatranscriptomic" and "single cell transcriptomic" are also available choices.

vii. *library_selection:* The method of selection or enrichment used in preparing the sequencing library. For RNA-Seq studies using polyA selection for enrichment for messenger RNA (mRNA), the user should choose "PolyA". For other RNA-Seq methods, such as Total RNA, choose "cDNA". More specialized options are available as appropriate (e.g. "cDNA_oligo_dT").

viii. *library_layout:* Specify whether paired or single end sequencing was done.

ix. *platform:* The sequencing platform used (Illumina, PacBio, etc.).

x. *instrument_model:* The specific model of the sequencing instrument.

xi. *design_description:* A short methods section describing how the libraries were prepared. Users are encouraged to provide all relevant details, including, e.g., specific tissues extracted or whether sequencing represents pooled individuals. If processed reads are being submitted, describe the filtering or other steps carried out.

xii. *filetype:* The format of the raw reads file (typically FASTQ).

xiii. *filename:* The exact file name, including extension. This must match the name of the file you upload. In the case of paired read files (R1 and R2), the submitter will enter two filenames in the columns *filename* and *filename2*.

Upload the newly created `metadata.txt` file and click *Continue*.

Figure 3: SRA metadata file. The yellow columns contain drop-down lists of acceptable entry values.

d. *Files:* Here the user uploads the raw reads.

There are several ways to do this, and the choice of which procedure to use will depend on where the reads files are stored (local hard drive versus remote server), the size of the reads files, and the number of reads files being uploaded. Locally stored files smaller than 2GB can be uploaded directly with a web-based interface, while larger files will require the Aspera connect plugin. Files stored on a server can be uploaded using a command line interface, using either an FTP connection or the Aspera Connect command line transfer tool. The command line options are preferred, because they work with files of any size and do not require the files to be stored on the user's hard drive. Within the command line upload options, the Aspera version is preferred if multiple files are to be uploaded. We here describe each option more thoroughly:

   i.   Direct upload of raw reads

Direct upload: On the "Files" page, choose *I will upload all the files now via HTTP/Aspera*, and then use the *Browse* button to select the files stored locally. After all the files are transferred, click on *Continue* to move to the final tab.

Aspera plugin for files > 2GB: Note: The Aspera connect plugin must be installed. On the "Files" page choose *I will upload all the files now via HTTP/Aspera* and click *Browse* to automatically launch Aspera and let the user select a locally stored file. Multiple files can be selected at once. After selecting the files, the user must confirm the transfer by allowing Aspera to connect to the NCBI web page. After all the files are transferred, click on *Continue* to move to the final tab.

   ii.   Command line upload of raw reads

Command line FTP upload: The command `ftp` is standard in Unix and Linux environments.

On the "Files" page, choose *I have all files preloaded for this submission*, and then click on *FTP upload instructions*. This will create a temporary NCBI user directory, and display all the information required for logging in. The information is a numbered list (1 through 7) and provides variable names that

will be entered as described below (Fig. 4). Keep this webpage open (or copy all the information). For more detailed instructions, see the accompanying website (*http://desalle.amnh.org/good_citizenship_rna/*). Now the user should open a session on their local server and connect via ftp to the NCBI server. In the following the example the "$" symbol represents the Unix prompt and variables in **bold** must be replaced with appropriate values by the user.

`$ cd `**`reads_directory`**

> (Navigate to the local directory where the user's FASTQ files for SRA upload are stored.)

`$ ftp `**`Address`**

> (Note: For "**Address**" enter the address provided on the SRA submission page under "2. Establish an FTP connection using the credentials below: Address:".)

`Name (ftp-private.ncbi.nih.gov:userid): `**`Username`**

> (Note: For "**Username**" enter the username provided on the SRA submission page under "2. Establish an FTP connection using the credentials below: Username:")

`Password: `**`Password`**

> (Note: For "**Password**" enter the password provided on the SRA submission page under "2. Establish an FTP connection using the credentials below: Password:". If there are difficulties in having the password accepted, try copying it directly from the SRA page.)

`ftp> cd `**`Folder`**

> (Note: For "**Folder**" enter the folder name provided on the SRA submission page under "3. **Navigate to your account folder**:")

`ftp> mkdir `**`New_folder`**

> (Note: For "**New_folder**" enter a name for the folder that will contain your FASTQ files for SRA upload. Using the submission name from the SRA web page (SUBxxxxxx) is a good choice, but any meaningful name is acceptable.)

`ftp> cd `**`New_folder`**

`ftp> put `**`File`**

> (Note: For "**File**" enter the name of a FASTQ file located in the user's local directory. Only one file at a time can be entered, so the user must do a separate `put` statement for each file.)

`ftp> `**`quit`**

After all the files are transferred, exit the FTP session ("**`quit`**") and return to the SRA submission webpage. Click on *Select preload folder* to see and select the folder that was just created. Click on *Use selected folder* and *Continue* to move to the final tab.

Figure 4: Command line FTP upload of raw reads to the SRA. The user must type the username and password as described above (basic protocol step 6.4 – command line upload).

Command line Aspera upload (see also Fig.5): If the user has installed the Aspera connect software on the server where the FASTQ files are stored, a single command can be used to transfer all the FASTQ files at once, while the accompanying folders on the NCBI server are generated automatically. On the "Files" page, choose *I have all files preloaded for this submission*, then click on *Aspera command line upload instructions*. Click on the link *Get the key file* to download the obligatory key file called "aspera.openssh". This file must be transferred to the folder where the FASTQ reads are stored, which can be done with the command "scp" from the terminal or using free FTP software such as FileZilla (available at *https://filezilla-project.org/*).

Make sure that all the FASTQ files for upload are in a single folder that contains nothing else, and then execute the following command, which will upload all the FASTQ files to a preload folder on the SRA submission website:

```
$ ascp -i /path/to/key_file -QT -l100m -k1 \
-d /path/to/reads_directory
```

(Note: For "**/path/to/key_file**" enter the address provided on the SRA submission page under "Aspera command line upload instructions". For "**/path/to/reads_directory**" enter the absolute path to the local folder where the files to upload are stored.)

Figure 5: Command line Aspera upload of raw reads to the Sequence Read Achieve. The FASTQ files must be in a single folder.

After all the files are transferred, return to the SRA submission webpage. Click on *Select preload folder* to see and select the folder that was just created and afterwards on *Continue* to move to the final tab.

e. *Overview:* Here the user can review all the provided information and check for errors.

7. Once satisfied with all entries, the user should click *Submit*. When the process is complete, the user will receive an email containing the final SRA ID in the form SRAxxxxxx.

BASIC PROTOCOL 2: STEP – BY – STEP GUIDE TO TRANSCRIPTOME SEQUENCE ASSEMBLY (TSA) SUBMISSION

This protocol describes how to submit an assembled transcriptome to the Transcriptome Shotgun Assembly database on NCBI. It first details the steps necessary to prepare the assembly file for submission, and then gives step-by-step instructions for submitting the assembly. Assemblies can be submitted as either FASTA or ASN.1 (Abstract Syntax Notation One) files—instructions are provided for each format.

Necessary Resources
Hardware:

- Unix, Windows, or Macintosh workstation with an Internet connection and web browser (Firefox, Chrome, Safari or Internet Explorer).

Software:

- Command line user interface (CLI) such as Terminal, iTerm, ConEmu, Console2, cmder, Cygwin, PuTTY, etc.

- *Aspera* connect high-speed file transfer software: For command line version: go to *http://downloads.asperasoft.com/en/downloads/8?list* and chose the version for your operating system. For web-based version: download the plugin from http://downloads.asperasoft.com/connect2//.
- tbl2asn (*ftp://ftp.ncbi.nih.gov/toolbox/ncbi_tools/converters/by_program/tbl2asn/*)

Data:

- Identifiers generated in the Basic Protocol 1: BioProject (`PRJNAxxxxxx`), BioSample (`SAMNxxxxxx`), SRA accession(s) (`SRRxxxxxx`)
- A transcriptome assembly based on the raw reads
- Technical information about how the transcriptome assembly was generated

The submission of an assembled transcriptome to TSA has many advantages for both the submitter and the scientific community. Different methods of read processing and different assembly algorithms can produce very different assemblies. If the user is planning to publish any analyses that directly rely on the assembly (e.g., number of contigs, number of predicted proteins, number of single copy genes), it is a good idea to make the underlying assembly available so that others can easily reproduce the analyses. An important caveat is that for TSA submissions, the submitter must have generated the raw reads themselves or be a member of the group that did so—the submission of assemblies based on publicly available raw reads is not permitted.

When submitting an assembly to TSA, there are some important differences in comparison to the submission of raw reads to SRA. TSA submissions are evaluated with VecScreen (*https://www.ncbi.nlm.nih.gov/tools/vecscreen/*) *to identify* contaminants, while SRA submissions are not. This step represents an additional quality assurance for further downstream analyses. See *https://www.ncbi.nlm.nih.gov/genbank/tsaguide* for additional information about TSA submissions from NCBI.

To minimize the number of errors that will be encountered, users should only attempt to submit high quality, contaminant-free assemblies. Recommendations about how to filter and assemble raw reads are beyond the scope of this work, but some useful tools include FastQC for quality control checks on raw reads (Andrews, 2010), Trimmomatic for adapter removal and quality trimming (Bolger et al., 2014), and Kraken (Wood & Salzberg, 2014) or NCBI's VecScreen (*https://www.ncbi.nlm.nih.gov/tools/vecscreen/*) for the identification and removal of contamination; also see (Conesa et al., 2016) for a recent discussion of best practices for assembling RNA-Seq data.

Preparation Steps

1. For submission to TSA, certain formatting conventions must be adhered to. Failure to follow these conventions will result in a failed submission, so it is important to properly format the data before submitting.
   a. Contig criteria: Contigs must be longer than 199 bp; must not contain more than 10% N's; must not start or end with N; and, must not contain stretches of more than 14

N's in a row. Contigs that do not meet these requirements should be removed from the assembly before attempting TSA submission.

b. Sequence definition line: This is the header line of each contig, which starts with ">" and ends with a newline character. The definition line must not be longer than 50 characters, including spaces, and must begin with a unique identifier (e.g. ">contig_001"). Additional modifiers can be used (see *https://www.ncbi.nlm.nih.gov/Sequin/modifiers.html* for a complete list). These follow the format "[modifier=text]" and can include *organism*, *sex*, and other details. NCBI advises that all TSA submissions include "[moltype=mRNA]" and "[tech=TSA]" in the definition line. Many assembly programs will produce definition lines that contain information about contig length, assembly path, etc. All these values must be removed. See the accompanying webpage for more details.

c. File name: The assembly must have the extension .fsa, not .fasta.

d. File format: The user must decide between submitting the assembly as a FASTA file or as an ASN.1 file. The ASN.1 format is mandatory if the submitter also plans to provide annotation; otherwise, either method can be chosen. We recommend the ASN.1 format because it embeds data in the submission that otherwise (if FASTA is chosen) must be entered manually on the TSA submission page.

A TSA submission can contain only one assembly. Thus, if a BioProject contains, for example, three BioSamples (with corresponding SRA files for each), three distinct TSA submissions will be required.

Protocol Steps

2. Once the assembly file is properly prepared, go to *https://submit.ncbi.nlm.nih.gov/* to begin the TSA submission. Sign in to NCBI with your user account, choose *TSA* and click on *New Submission.*
During this step, the submitter will provide information about the assembly, including the BioSample, BioProject, and SRA identifiers generated in Basic Protocol 1. The submission process progresses through either five (ASN.1) or six (FASTA) fillable forms. We first present the steps for FASTA files, followed by the steps for ASN.1 files.

3. Submitting a FASTA assembly:
   a. *Submitter:* Provide information about the person submitting the data and the submitting organization (typically, this will be the submitter's organizational affiliation). An email address from the submitting organization's domain is required.
   b. *General info:* Here the user will provide the BioProject identifier (PRJNAxxxxxx), the BioSample identifier (SAMNxxxxxx), and the SRA accession identifier (SRRxxxxxx), and choose a release date. In addition, the user must provide "Assembly metadata", which includes:

     i.   *Information about the assembly method:* Provide the name of the assembly program used (e.g. Trinity, Abyss, etc.) and the version number (or date of assembly, if program version is not known);

    ii.   *Assembly name* (optional).

    iii.   *Assembly coverage* (optional).

    iv.   *Description of assembly method* (required): This should be as detailed as possible and include and read processing steps, whether default program settings were used, and any other information that would be required to exactly reproduce the assembly process; and,

    v.   *Sequencing Technology:* The platform used for sequencing (i.e. Illumina HiSeq, PacBio, etc.).

c. *File:* Here the user must choose between submitting the assembly as a FASTA file or as an ASN.1 file. Choose *File type* FASTA and click *Continue.*

d. *Sequence:* Click on *Browse* to select an `assembly.fsa` file stored on the user's local machine. An Aspera connect window will open to display the progress of the upload. Once the upload is complete, the message "Please wait! Processing the data" is displayed as an initial TSA validation check is conducted. If errors are displayed, the user will need to filter the assembly to remove any sequences that have been flagged by NCBI's screening process. Detailed instructions on filtering are provided in the accompanying webpage (*http://desalle.amnh.org/good_citizenship_rna/*). If no errors are displayed, click *Continue.*

e. *References:* Provide the name(s) of the *Sequence authors*, i.e. the people responsible for generating the raw reads upon which the assembly is based, and information about publications (if any) that include the assembly.

f. *Overview:* Here the user can look over all the provided information and decide if changes are needed. If the user is satisfied, click *Submit.*

4. Submitting an ASN.1 assembly:

Before submitting an ASN.1 file, additional preparations are required. The following files must be generated for each assembly that will be submitted:

    i.   Create a "GenBank Submission Template" (SBT) file: Go to *https://submit.ncbi.nlm.nih.gov/genbank/template/submission/* and complete the required fields. The same BioProject identifier can be used for as many assemblies as are part of the project, but each BioSample identifier should have its own SBT file. After filling all fields, click *Create Template* to download the SBT file. Save the file with a name that reflects which assembly it is for.

    ii.   Create a "Structured Comment - Non Genome Submissions" (CMT) file: Go to *https://submit.ncbi.nlm.nih.gov/structcomment/nongenomes/* and complete the required fields. The *Assembly name* field should be specific to a single assembly. Click *Download* to download the CMT file. Save the file with a name that reflects which assembly it is for.

iii. Place the following files in a single folder (either on the user's local machine or on a server): The correctly formatted *assembly.fsa* file, as described in Preparation Steps; the SBT and CMT files just created.

iv. From within this newly created folder, run the following command:

`$ ./path/to/version.tbl2asn` -t **assembly.sbt** \

-i **assembly.fsa** -a s -V v -w **assembly.cmt**

(Note: for "`./path/to/version.tbl2asn`" enter the path to the version of the tbl2asn executable downloaded from ftp://ftp.ncbi.nih.gov/toolbox/ncbi_tools/converters/by_program/tbl2asn/.)

This command will generate two new files:

`assembly.val`, a validation file that will report errors—if this file is empty, no errors were detected;

`assembly.sqn`, which is the ASN.1 file for TSA submission.

v. Once the `assembly.sqn` file is prepared, go to *https://submit.ncbi.nlm.nih.gov/* to begin the TSA submission.

a. *Submitter:* As described in 3a (Basic Protocol 2).

b. *General info:* As described in 3b (Basic Protocol 2).

c. *File:* As described above, but choose *File type* ASN.1 and click *Continue*.

d. *Sequence:* Click on *Browse* to select an *assembly.sqn* file stored on the user's local machine. An Aspera connect window will open to display the progress of the upload. Once the upload is complete, the message "Please wait! Processing the data" is displayed as an initial TSA validation check is conducted. If errors are displayed, the user will need to filter the assembly to remove any sequences that have been flagged by NCBI's screening process. Detailed instructions on filtering are provided at the accompanying website (*http://desalle.amnh.org/good_citizenship_rna/*). If no errors are displayed, click *Continue*.

e. *Overview:* Here the user can review all the provided information and decide if changes are needed. If the user is satisfied, click *Submit*.

5. The submission, (whether FASTA or ASN.1), will now undergo quality assessment at NCBI including a complete VecScreen analysis. This process can take 12 or more hours. If there are no problems, NCBI will send a confirmation email with a TSA accession number in the format `GAAxxxxxx`.

6. If the assembly fails to pass the post-submission checks, the submitter will receive an email stating that the submission has failed. An included link directs the user to a detailed error file describing the type of contaminant identified and listing the corresponding sequence identifiers. Detailed instructions on removing flagged sequences are provided at *http://desalle.amnh.org/good_citizenship_rna/*. Once the offending

sequences have been trimmed or removed, the user can return to step 4d (or step 3d for the FASTA format) to resubmit the corrected files.

COMMENTARY

The deposition of raw reads and assemblies on a public platform is a fundamental requirement for the publication of research involving NGS data. Currently there are more than 2.5 million BioSample IDs and over 33,500 SRA entries registered at NCBI. This reflects a concomitant increase in the amount of SRA data from 5 TB in 2009 to over 20,000 TB in 2018 (of which some 8,000 TB are publicly available). Depositing raw reads and assemblies has obvious advantages for researchers because it provides a free and reliable way to archive research data. For the research community as a whole, openly shared, easily accessible raw data can lay the foundation for large comparative studies that would otherwise be prohibitively expensive.

In spite of these manifest advantages to making one's raw reads and assemblies freely available on NCBI databases, there are some limitations to consider. First, because users can submit either "raw" reads (directly from the sequencing machine) or "processed" reads (that have been filtered for contamination and quality), the use of SRA data is very much a case of caveat emptor ("let the buyer beware"). A simple solution would be for the SRA submission process to require information about any filtering steps that were carried out or, instead, to explicitly accept only raw data as delivered from the sequencing machine. Second, since a major value of SRA data is that it can be used by researchers other than those who generated the raw data, it seems counterproductive that only the owner of the raw sequence data is permitted to submit assemblies based on the data. As assembly algorithms improve, the same raw reads can generate much better assemblies than were possible at the time of sequencing, and a mechanism should be devised that allows improved assemblies to be submitted by whoever generated them.

INTERNET RESOURCES

Web version of this guide: *http://desalle.amnh.org/good_citizenship_rna/*

NCBI links:

BioSamples: *https://www.ncbi.nlm.nih.gov/sra/docs/submitbio/*

BioProjects: *https://www.ncbi.nlm.nih.gov/sra/docs/submitbio/*

SRA: *https://www.ncbi.nlm.nih.gov/sra/docs/,*
https://www.ncbi.nlm.nih.gov/sra/docs/submit/,
https://www.ncbi.nlm.nih.gov/sra/docs/submitspfiles/,
https://www.ncbi.nlm.nih.gov/sra/docs/submitquestions/,
https://www.ncbi.nlm.nih.gov/books/NBK51157/

TSA: *https://www.ncbi.nlm.nih.gov/genbank/tsa/*,

https://www.ncbi.nlm.nih.gov/genbank/tsaguide

ACKNOWLEDGEMENTS

LITERATURE CITED

Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A.-L., Jacquin, F., Klein, A., Martin, C., Boucherot, K., Kreplak, J., Silva, C., Moreau, S., Gamas, P., Wincker, P., Gouzy, J., & Burstin, J. (2015). Full-length de novo assembly of RNA-seq data in pea (Pisum sativum L.) provides a gene expression atlas and gives insights into root nodulation in this species. *The Plant Journal, 84*(1), 1-19. doi:10.1111/tpj.12967

Andrews, S. (2010, 08/03/2016). FastQC. A quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol, 17*, 13. doi:10.1186/s13059-016-0881-8

Crawford, J. E., Guelbeogo, W. M., Sanou, A., Traoré, A., Vernick, K. D., Sagnon, N. F., & Lazzaro, B. P. (2010). De novo transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology. *PloS one, 5*(12), e14202.

Feindt, W., Oppenheim, S. J., DeSalle, R., Goldstein, P. Z., & Hadrys, H. (2018). Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world's largest wings. *PloS one, 13*(1), e0189898.

Gruber, D. F., Gaffney, J. P., Mehr, S., DeSalle, R., Sparks, J. S., Platisa, J., & Pieribone, V. A. (2015). Adaptive evolution of eel fluorescent proteins from fatty acid binding proteins produces bright fluorescence in the Marine Environment. *PloS one, 10*(11), e0140972.

Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W., & Bernatchez, L. (2010). The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular ecology, 19*(24), 5389-5403.

Kumar, S., & Blaxter, M. L. (2010). Comparing de novo assemblers for 454 transcriptome data. *BMC genomics, 11*(1), 1.

Lopez-Maestre, H., Brinza, L., Marchet, C., Kielbassa, J., Bastien, S., Boutigny, M., Monnin, D., Filali, A. E., Carareto, C. M., Vieira, C., Picard, F., Kremer, N., Vavre, F., Sagot, M.-F., & Lacroix, V. (2016). SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res, 44*(19), e148-e148.

Mehr, S. F. P., DeSalle, R., Kao, H.-T., Narechania, A., Han, Z., Tchernov, D., Pieribone, V., & Gruber, D. F. (2013). Transcriptome deep-sequencing and clustering of expressed isoforms from *Favia* corals. *BMC genomics, 14*(1), 546.

NCBI Resource Coordinators, N. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res, 44*(D1), D7-D19. doi:10.1093/nar/gkv1290

Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Sederoff, R. R., & Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC genomics, 9*(1), 1.

Oppenheim, S. J., Baker, R. H., Simon, S., & DeSalle, R. (2015). We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect molecular biology, 24*(2), 139-154.

Oppenheim, S. J., Feindt, W., DeSalle, R., & Goldstein, P. Z. (2018). De Novo characterization of transcriptomes from two North American Papaipema stem-borers (Lepidoptera: Noctuidae). *PloS one, 13*(1), e0191061.

Parchman, T. L., Geist, K. S., Grahnen, J. A., Benkman, C. W., & Buerkle, C. A. (2010). Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC genomics, 11*(1), 180.

Simon, S., Sagasser, S., Saccenti, E., Brugler, M. R., Schranz, M. E., Hadrys, H., Amato, G., & DeSalle, R. (2017). Comparative transcriptomics reveals developmental turning points during embryogenesis of a hemimetabolous insect, the damselfly *Ischnura elegans*. *Scientific reports, 7*(1), 2045-2322.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol, 15*(3), R46-R46. doi:10.1186/gb-2014-15-3-r46

The complete mitochondrial genome of the Neotropical helicopter damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera) assembled from next generation sequencing data

Wiebke Feindt[1], Hans-Jürgen Osigus[1], Rebecca Herzog[1],
Christopher E. Mason[2, 3] and Heike Hadrys[1,4]

[1] ITZ, Division of Ecology and Evolution, University of Veterinary Medicine Hannover, Bünteweg 17d, 30559 Hannover, Germany
[2] The Department of Physiology and Biophysics and the HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, 1305 York Avenue, Weill Cornell Medical College, New York, New York 10065, USA
[3] The Feil Family Brain and Mind Research Institute, 413 East 69th St, Weill Cornell Medical College, New York, New York 10065, USA
[4] The Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA

* Author for correspondence: Wiebke Feindt
Email: wiebke.feindt@ecolevol.de

The complete mitochondrial genome of the Neotropical helicopter damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera) assembled from next generation sequencing data

Abstract

Odonata (dragonflies and damselflies) is a small order at the base of flying insects Pterygota). Resolving family-level phylogenetic relationships within this order receives great attention. Hereby, genetic data already resulted in various changes, which are however still under discussion. Mitochondrial genomes may further enhance such phylogenies. This study presents the complete mitochondrial genome of the Neotropical damselfly *Megaloprepus caerulatus* based on next generation sequencing (NGS) data on total genomic DNA. The total length comprises 16,094bp and includes the standard metazoan set of 37 genes together with a 1376bp long A+T rich (control) region. Gene content, gene arrangement and base frequency are consistent with other odonate mitochondrial genomes. It further contains four intergenic spacer regions, indicating a possible family specific feature for the Coenagrionidae and its close relatives.

The relatively small insect order Odonata owns a key position in the evolution of winged insects and as sensitive indicator organisms for freshwater ecosystems. Robust phylogenies are one ultimate requirement for studies in evolution, ecology and developmental biology. Over the last 10 years many molecular data-based attempts have already resulted in various reorganizations of phylogenetic relationships (e.g. Dumont et al., 2010; Carle et al., 2015). However, today's phylogenies still harbour open questions concerning family-level as well as deeper taxonomic positions. Mitochondrial genome projects in odonates may help to unravel such unresolved phylogenetic relationships by constructing more robust phylogenies based on complete gene and genome comparison (e.g. Simon & Hadrys, 2013; 2014). One family that has received great attention in the past are the Pseudostigmatidae (giant damselflies) (e.g. Groeneveld et al., 2007; Ingley et al., 2012). They were recently placed into the Coenagrionidae based on three sequence markers (Dijkstra et al., 2014). Here, more robust genomic data are needed not only to clarify their taxonomic position but also to facilitate genomic-based studies on Neotropical forest health indicators. In this work, we present a complete mitochondrial genome of *Megaloprepus caerulatus* as the first member of this group. *Megaloprepus caerulatus* was already included in many ecological (e.g.

Fincke & Hedström, 2008) and evolutionary studies in tropical habitats, but yet little research has been done on its genetics (e.g. Fincke & Hadrys, 2001; Feindt et al., 2014).

Via a standard Phenol-Chloroform extraction (Hadrys et al., 1992) total genomic DNA from flight muscles from a single *M. caerulatus* individual collected at the Biological Research Station La Selva (OTS), Costa Rica (N 10°25'19.74" W 84°00'35.22") was extracted. Library preparation and DNA sequencing (100 bp mate pairs with different insert sizes, Illumina HiSeq2500, Illumina Inc., San Diego, CA) was performed at the Weill Cornell Medical College in New York. The mitochondrial genome was assembled using Geneious vers. 8.1 (http://www.geneious.com); while mapping a fraction of the cleaned reads onto a seed sequence (here *cox1*: KF895301.1 and *nad1*: KF895193.1) allowing strand extension using varying iterates. Hereby, the settings included a minimum overlap of 60bp, a minimal overlap identity of 90%, and a variable word size between 35 and 50. The mitochondrial genome was annotated using the MITOS WebServer (mitos.bioinf.uni-leipzig.de/index.py) and verified via BLAST (Altschul et al., 1990) against the NCBI database (http://www.ncbi.nih.gov) or additionally to published mitochondrial genomes of other odonate species. Transfer RNA genes were identified by a tRNA covariance model implied on the tRNAscan-SE vers. 1.21 Search Server (http://lowelab.ucsc.edu/tRNAscan-SE; Lowe & Eddy, 1997) and ARWEN vers. 1.2 (http://mbio-serv2.mbioekol.lu.se/ARWEN; Laslett & Canbäck, 2008). Phylogenetic relationships were reconstructed using a selection of odonate mitochondrial genomes and a mayfly *Parafronurus youi* as outgroup (Figure 1). All 13 protein coding genes and rRNA genes were aligned independently, then concatenated and a maximum parsimony tree was calculated (1,000 replicates) in Paup (Swofford, 2002).

The obtained complete mitochondrial genome of *M. caerulatus* (NCBI: KU958377) is the first of a tropical odonate species. It has a total length of 16,094bp, and is the second largest aside of *Vestalis melania* (16,685bp, NC_023233). In all observed parameters the presented mitochondrial genome shows a strong similarity to all other already published odonate mitochondrial genomes (e.g. Lorenzo-Carballa et al., 2014; Tang et al., 2014; Yu et al., 2014; Chen et al., 2015; Feindt et al., 2016; Herzog et al., 2016). It contains the common arrangement of 37 genes including 13 protein-coding genes, 2 rRNA (16S and 12S rRNA) genes, 22 tRNA genes, and an A+T - rich (control) region of 1,376bp in length. Comparing with the previously released single genes we observed 100% identity to *nad1* (DQ642992.1) and *16S* rRNA (DQ642987.1) (Groeneveld et al., 2007) in, and in *nad1* (JQ966612.1), *16S* (JQ966660.1) and *12S* rRNA (JQ966647.1) a similarity between 98 and 100% to genes illustrated in Ingley et al., (2012). The overall A+T - content of the mitochondrial genome is 79.9% (A: 43.1%, C: 14.2%, G: 9.8%, T: 24.1%) and therefore it is similar to the base frequencies of the protein coding genes (AT: 74.4%). Hereby, *atp8* encompasses the highest A+T - content with 81.1% and *cox1*, as it was described in other odonates (Lorenzo-Carballa et al., 2014), with 68.9% the lowest. All protein coding genes start with characteristic invertebrate specific mitochondrial start codons: *cox1*, *atp6*, *cox3*, *nad4*, *nad4L*, and *cob* use ATG; *nad2*, *cox2*, and *nad6* start with ATA; *nad3* and *nad1* start with TTG; *nad5* starts with ATT and *atp8* starts with ATC. The standard stop codon TAA was used eight times (*nad2*,

cox1, *atp8*, *atp6*, *nad4L*, *nad6*, *cob*, *nad1*), whereas TAG was used only once by *nad3*. An incomplete stop codon with a single T was found in four cases (*cox2*, *cox3*, *nad5*, and *nad4*). Length of the tRNA genes in *M. caerulatus* ranges from 64bp in to 74bp and except for *trnL1* (Leu), *trnS1* (Ser), *trnG* (Ser) all of which employ the typical clover-leaf secondary structures.
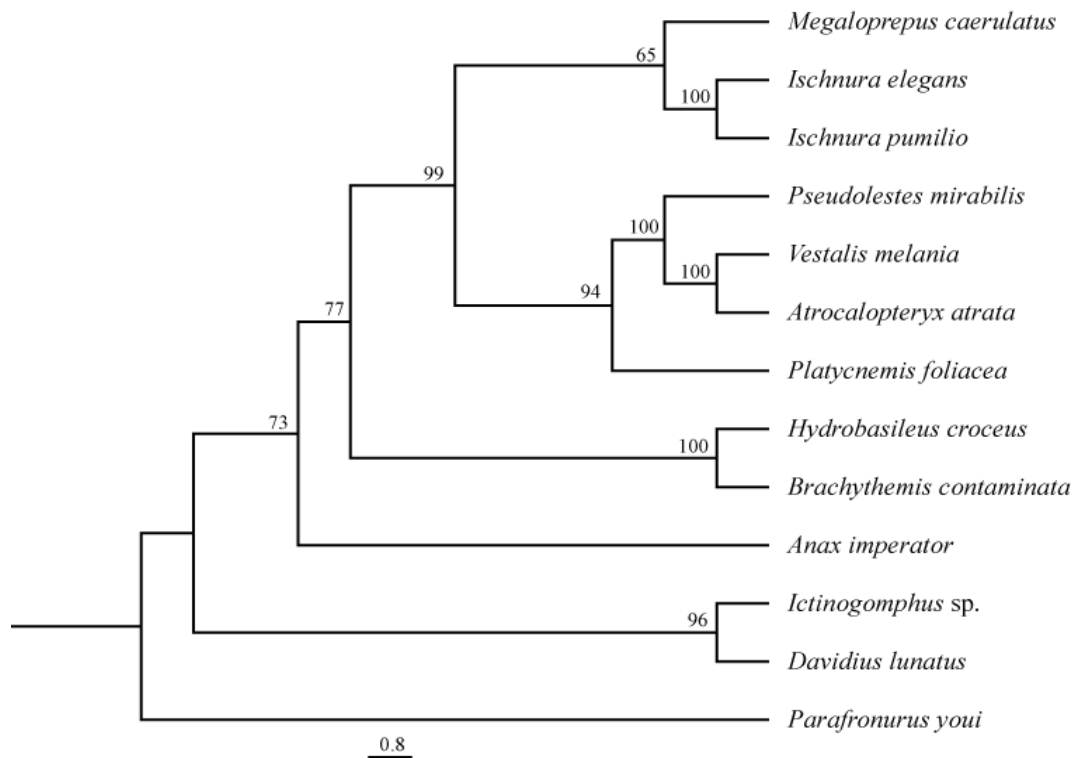
Figure 1: Phylogenetic relationships of odonate species based on maximum parsimony analysis of concatenated mitochondrial protein coding genes and rRNA sequences: *Brachythemis contaminata* (NC_026305), *Ictinogomphus* sp. (KM244673), *Hydrobasileus croceus* (NC_025758), *Davidius lunatus* (NC_012644), *Ischnura pumilio* (NC_021617), *Pseudolestes mirabilis* (NC_020636), *Atrocalopteryx atrata* (NC_027181), *Vestalis melania* (NC_023233), *Platycnemis foliacea* (NC_027180), *Anax imperator* (KX161841), *Ischnura elegans* (KU958378), and *Parafronurus youi* (EU349015.1) as outgroup. The heuristic search (under the 50 % majority-rule with 1,000 bootstrap replicates) placed *Megaloprepus* as a sister species to *Ischnura* spp, whereat relationships are displayed as in other phylogenies.

A difference in numbers of intergenic spacer regions is described in the two large odonate orders (Lorenzo-Carballa et al., 2014). They seem to provide a phylogenetic signal for the split of Anisoptera (dragonflies) from Zygoptera (damselflies). The lack of the intergenic spacer region *s5* seems to be a damselfly specific character. In *M. caerulatus* we detected four spacer regions at *trnY*/*cox1*, *trnF*/*nad5*, *trnT*/*trnP* and *trnS2*/*nad1* with a total length of 106bp. With this we proved that the unique spacer *s4* between *trnF* and *nad5*,

which is also present in *I. pumilio* (NC_021617), *Pseudolestes mirabilis* (NC_020636) and *I. elegans* (KU958378) might not only be a specific character for the Coenagrionidae. The phylogenetic tree (Figure 1) places *M. caerulatus* as a sister species to *Ischnura* spp. with low support. More mitochondrial genomic data is needed to resolve relationships within and between families. However, the mitochondrial genome presented in this study is a valuable resource for future population genomic studies and also owns potential for answering phylogenetic questions on higher taxonomic levels.

Disclosure statement

Collecting permit was processed from the Ministerio de Ambiente y Energía, Costa Rica (MINAE) and the permit for using the genetic material the individual was approved by Comisión Nacional para la Gestión de la Biodiversidad, Costa Rica (CONAGEBIO). Furthermore, the authors state no conflicts of interest.

Acknowledgements

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology, 215, 403-410.

Carle FL, Kjer KM, May ML (2015) A molecular phylogeny and classification of Anisoptera (Odonata). Arthropod systematics & phylogeny, 73, 281-301.

Chen M-Y, Chaw S-M, Wang J-F, Villanueva RJT, Nuneza OM, Lin C-P (2015) Mitochondrial genome of a flashwing demoiselle, *Vestalis melania* from the Philippine Archipelago. Mitochondrial DNA, 26, 720-721.

Dijkstra KDB, Kalkman VJ, Dow RA, Stokvis FR, Van Tol J (2014) Redefining the damselfly families: a comprehensive molecular phylogeny of Zygoptera (Odonata). Systematic Entomology, 39, 68-96.

Dumont HJ, Vierstraete A, Vanfleteren JR (2010) A molecular phylogeny of the Odonata (Insecta). Systematic Entomology, 35, 6-18.

Feindt W, Fincke O, Hadrys H (2014) Still a one species genus? Strong genetic diversification in the world's largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*. Conservation genetics, 15, 469-481.

Feindt W, Herzog R, Osigus HJ, Hadrys H (2016). Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* (Vander Linden, 1820)**.** Mitochondrial DNA. *In press*.

Fincke OM, Hadrys H (2001) Unpredictable offspring survivorship in the damselfly, *Megaloprepus coerulatus*, shapes parental behavior, constrains sexual selection, and challenges traditional fitness estimates. Evolution, 55, 762-772.

Fincke OM, Hedström I (2008) Differences in forest use and colonization by Neotropical tree-hole damselflies (Odonata: Pseudostigmatidae): Implications for forest conversion. Studies on Neotropical Fauna and Environment, 43, 35-45.

Groeneveld LF, Clausnitzer V, Hadrys H (2007) Convergent evolution of gigantism in damselflies of Africa and South America? Evidence from nuclear and mitochondrial sequence data. Molecular Phylogenetics and Evolution, 42, 339-346.

Hadrys H, Balick M, Schierwater B (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular ecology, 1, 55-63.

Herzog R, Osius HJ, Feindt W, Schierwater B, Hadrys H (2016) The complete mitochondrial genome of the emperor dragonfly *Anax imperator* LEACH, 1815 (Aeshnidae: Odonata) via NGS sequencing. Mitochondrial DNA, *in press*

Ingley SJ, Bybee SM, Tennessen KJ, Whiting MF, Branham MA (2012) Life on the fly: phylogenetics and evolution of the helicopter damselflies (Odonata, Pseudostigmatidae). Zoologica Scripta, 41, 637-650.

Laslett D, Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics, 24, 172-175.

Lorenzo-Carballa MO, Thompson DJ, Cordero-Rivera A, Watts PC (2014) Next generation sequencing yields the complete mitochondrial genome of the scarce blue-tailed damselfly, *Ischnura pumilio*. Mitochondrial DNA, 25, 247-248.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research, 25, 955-964.

Simon S, Hadrys H (2013) A comparative analysis of complete mitochondrial genomes among Hexapoda. Molecular phylogenetics and evolution, 69, 393-403.

Simon S, Hadrys H (2014). Phylogeny of the most species rich group on Earth, the Pterygota: ancient problems, living hypotheses and bridging gaps, Deep Metazoan Phylogeny: The Backbone of the Tree of Life: New insights from analyses of molecules, morphology, and theory of data analysis. De Gruyter

Swofford DL (2002) PAUP* phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland

Tang M, Tan M, Meng G, Yang S, Su X, Liu S, Song W, Li Y, Wu Q, Zhang A (2014) Multiplex sequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mito-metagenomics. Nucleic acids research, 42, e166-e166.

Yu P, Cheng X, Ma Y, Yu D, Zhang J (2014) The complete mitochondrial genome of *Brachythemis contaminata* (Odonata: Libellulidae). Mitochondrial DNA, 1-2.

Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera)

Wiebke Feindt[1]*, Rebecca Herzog[1]*, Hans-Jürgen Osigus[1],
Bernd Schierwater[1,2] and Heike Hadrys[1,2]

* The authors contributed equally to the manuscript.

[1] ITZ - Forschungsstätte 'Alter Bahnhof Schapen' Braunschweig, University of Veterinary Medicine Hannover, Hannover/Braunschweig, Germany
[2] Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06511, USA

* Author for correspondence: Heike Hadrys
E-mail: heike.hadrys@ecolevol.de

Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera)

Abstract

Damselflies of the genus *Ischnura* emerge as organisms with high potential in ecological, evolutionary and developmental research at the base of flying insects. *Ischnura elegans* and *Ischnura hastata* are for example one of the few odonate species where a complete life cycle over generations can be reared under laboratory conditions. We here report the complete mitochondrial genome of *Ischnura elegans* as a valuable genomic resource for future eco-evo-devo studies at the base of flying insects. The genome has a total length of 15,962 bp and displays all typical features of Odonata (dragonflies and damselflies) mitochondrial genomes in gene content and order as well as A+T content. Start and stop codons of all protein coding genes are consistent. Most interestingly, we found four intergenic spacer regions and a long A+T rich (control) region of 1,196bp, which is almost double the size of the close relative *Ischnura pumilio*. We assume that the adequate insert size and iterative mapping may be more efficient in assembling this duplicated and repetitive region.

Keywords: Mitochondrial genome, Odonata, iterative mapping, *Ischnura elegans*, A+T rich (control) region

The blue-tailed damselfly *Ischnura elegans* is a small, widely distributed European damselfly of the family Coenagrionidae. The females of this species exhibit a color polymorphism with three different color morphs (e.g. Andres et al. 2000), which put them in the center of research concerning the evolution of color polymorphism (e.g. Hammers & Van Gossum 2008; Cordero-Rivera & Sánchez-Guillén 2007). Furthermore, the complete life cycle of this species can be cultured in the lab bridging the gaps between developmental, environmental and evolutionary studies (Simon & Hadrys 2013; 2014). It was the first odonate species for which genomic data in terms of ESTs (Simon et al. 2009) and a transcriptome (Chauhan et al. 2014) were available. A complete genome would facilitate state of the art future studies in eco-evo-devo. In a first attempt, we here present the assembly of the mitochondrial genome of *I. elegans*.

Total genomic DNA was extracted from the flight muscles of a single individual using a standard Phenol-Chloroform extraction (Hadrys et al. 1992). The specimen was collected in Schapen, northern Germany (52°16′7.95"N, 10°31′36.37"E). The library preparation and whole genome sequencing was conducted at Yale University in the Center for Genome Analyses (YCGA, http://www.ycga.yale.edu) on an lllumina HiSeq2000 (Illumina Inc.)

platform generating 75bp paired-end reads with an insert size of approximately 450bp. For the assembly of the complete mitochondrial genome Geneious vers. 8.1.5 (http://www.geneious.com) was applied as follows: one published mitochondrial gene sequence served as seed (*cox2*, KC430130), and a fraction of the cleaned reads were mapped onto the seed using iterative mapping. Hereby the iterations were increased with length of the seed sequence (from 5 to 25) as well as the overlap identity, which was initially placed at 90 %. A maximum overlap of 45 – 50bp was chosen to prevent miss-mapping. In addition, the length of the A+T rich (control) region was confirmed via PCR. The continuous annotation was conducted using the MITOS WebServer (mitos.bioinf.uni-leipzig.de/index.py) and verified via BLAST (Altschul et al. 1990) against GenBank, already published mitochondrial genomes (e.g. Tang et al. 2014; Yu et al. 2014; Chen et al. 2015) and especially against the closest relative *Ischnura pumilio* (NC_021617; Lorenzo-Carballa et al. 2014). Transfer RNAs were predicted using the tRNAscan-SE vers.1.21 Search Server (http://lowelab.ucsc.edu/tRNAscan-SE; Lowe & Eddy 1997) and ARWEN vers. 1.2 (http://mbio-serv2.mbioekol.lu.se/ARWEN; Laslett & Canbäck 2008). Finally, a phylogeny was reconstructed using four other selected Odonata species and *I. elegans* (Figure 1). Based on a concatenated alignment of all 13 protein coding genes and the rRNA genes a maximum parsimony tree was calculated using PAUP vers. 4.0b10 (Swofford 2002) with a heuristic search under the 50 % majority-rule and 1,000 bootstrap replicates.

The present mitogenome has a total length of 15,962bp (GenBank accession number: KU958378) and is 712 bp longer than *I. pumilio*'s with a high overall similarity of 86.4%. It displays the typical metazoan gene content with 13 protein coding genes (PCGs), 2 rRNA genes (16S and 12S) and 22 tRNA genes with an identical gene order to all other odonate mitogenomes published to date (Tab. 1). Furthermore, four intergenic spacer regions were detected which are consistent in position with *I. pumilio* and *Megaloprepus caerulatus* (Lorenzo-Carballa et al. 2014, Feindt et al. 2016) but differ in size. Base composition of the *I. elegans* mitogenome is AT biased (A: 40.3%, T: 32.5%, G: 12.3%, C: 14.8%) so as all protein coding genes (71.4%), rRNAs (75.1%) and tRNAs (72.1%) on average. Except for *cox1* (TTA) all PCGs initiate with standard invertebrate mitochondrial start codons: *cox3*, *nad4*, *nad4L* and *cytb* with ATG; *nad2*, *cox2*, *nad5*, with ATT; *atp8*, *atp6*, *nad3*, *nad1* with TTG; and *nad6* with ATC. Complete stop codons terminate nine genes (TAA: *nad2*, *cox1*, *atp8*, *atp6*, *nad4L*, *nad6*, *cytb*, *nad1*; TAG: *nad3*) whereas four proteins use incomplete stop codons with post-transcriptional polyadenylation (*cox2*, *cox3*, *nad5*, *nad4*). Transfer RNAs vary in size from 65 – 72bp and all of them fold into the characteristic clover-leaf secondary structure. Overlapping gene junctions were observed for 13 genes, the longest overlap between *atp6* and *atp8* is 13bp.
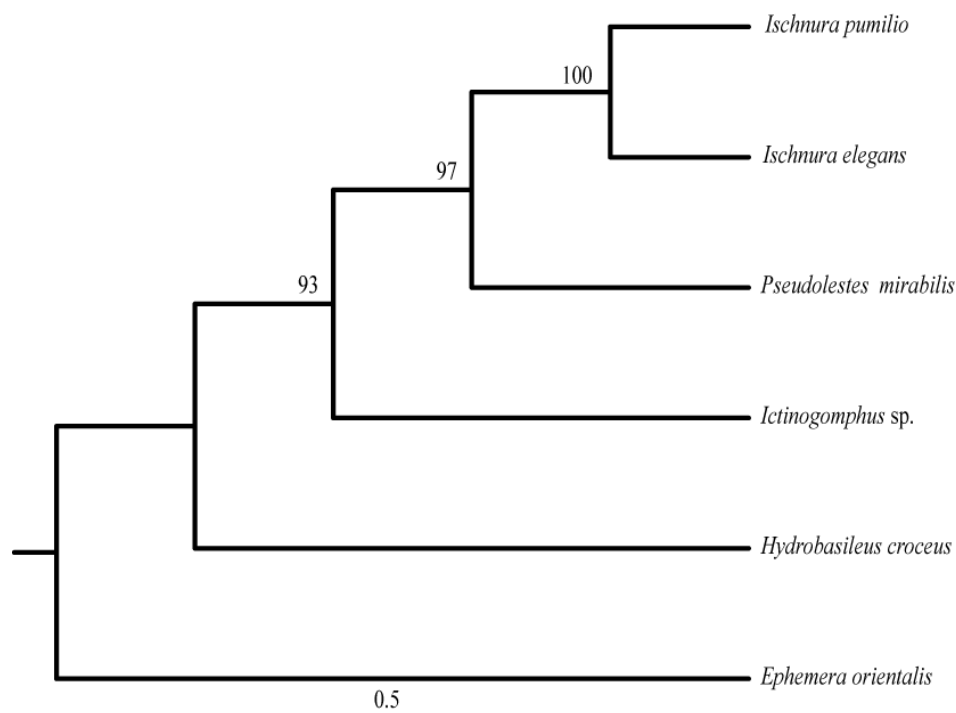
Figure 1: Phylogenetic position of *I. elegans* and *I. pumilio* (NC_021617), *Pseudolestes mirabilis* (NC_020636), *Hydrobasileus croceus* (NC_025758), *Ictinogomphus* sp. (KM244673) using the mayfly *Ephemera orientalis* (NC_012645) as outgroup. MP bootstrap supports are shown for each node. *I. elegans* and *I. pumilio* exhibit 86.4 % identity in the mt-genes used.

The different length of the presented mt genome compared to *I. pumilio* is mainly based on the almost two times longer A+T rich (control) region. Since we proved the length of the A+T rich (control) region via PCR, we assume that the combination of an appropriate insert size and iterative mapping may be more accurate for the assembly of long repetitive and duplicated regions. These usually tend to challenge genome assembly software. The control region comprises a triplicated motive of in total almost 600bp, which could only be resolved correctly with consideration to the insert size. This could be an important aspect for future library preparation on Illumina platforms.

Acknowledgements

Declaration of interest

The authors declare no conflict of interest to other working groups.

Table 1: Organization of the mitochondrial genome of *Ischnura elegans*. The table displays the gene order with information about the gene boundaries as well as start and stop codons, whereas incomplete stop codons are displayed as T(aa). All transfer RNAs are named according their corresponding amino acid. The intergenic spacer regions are named continuously from *s1* to *s4*.

Name	strand	Start position	Stop position	Length (bp)	Start codon	Stop codon
tRNA-Ile	+	1	67	67	/	/
tRNA-Gln	-	65	132	68	/	/
tRNA-Met	+	132	200	69	/	/
NAD2	+	201	1,196	996	ATT	TAA
tRNA-Trp	+	1,195	1,263	69	/	/
tRNA-Cys	-	1,256	1,321	66	/	/
tRNA-Tyr	-	1,322	1,389	68	/	/
s1	n.a.	1,390	1,406	17	/	/
COX1	+	1,407	2,967	1561	TTA	TAA
tRNA-Leu	+	2,963	3,028	66	/	/
COX2	+	3,029	3,716	688	ATT	T(aa)
tRNA-Lys	+	3,717	3,788	72	/	/
tRNA-Asp	+	3,789	3,854	66	/	/
ATP8	+	3,852	4,016	165	TTG	TAA
ATP6	+	4,004	4,687	684	TTG	TAA
COX3	+	4,687	5,473	787	ATG	TA(a)
tRNA-Gly	+	5,474	5,538	65	/	/
NAD3	+	5,539	5,892	354	TTG	TAG
tRNA-Ala	+	5,891	5,957	67	/	/
tRNA-Arg	+	5,956	6,021	66	/	/
tRNA-Asn	+	6,023	6,089	67	/	/
tRNA-Ser	+	6,089	6,159	71	/	/
tRNA-Glu	+	6,161	6,227	67	/	/
tRNA-Phe	-	6,226	6,291	66	/	/
s2	n.a.	6,292	6,303	12	/	/
NAD5	-	6,304	8,020	1717	ATT	T(aa)
tRNA-His	-	8,021	8,086	66	/	/
NAD4	-	8,087	9,429	1343	ATG	TA(a)
NAD4L	-	9,423	9,716	294	ATG	TAA
tRNA-Thr	+	9,719	9,784	66	/	/
s3	n.a.	9,785	9,794	10	/	/
tRNA-Pro	-	9,795	9,862	68	/	/
NAD6	+	9,865	10,380	516	ATC	TAA
CYTB	+	10,380	11,513	1134	ATG	TAA
tRNA-Ser	+	11,514	11,578	65	/	/
s4	n.a.	11,579	11,598	20	/	/
NAD1	-	11,599	12,549	951	TTG	TAA
tRNA-Leu	-	12,551	12,618	68	/	/
16S	-	12,619	13,902	1284	/	/
tRNA-Val	-	13,903	13,973	71	/	/
12S	-	13,974	14,766	793	/	/
A+T rich (control) region	n.a.	14,767	15,962	1196	/	/

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of molecular biology, 215, 403-410.

Andrés JA, Sánchez-Guillén RA, Cordero Rivera A. 2000. Molecular evidence for selection on female color polymorphism in the damselfly *Ischnura graellsii*. Evolution, 54(6), 2156-2161.

Chauhan P, Hansson B, Kraaijeveld K, de Knijff P, Svensson EI, Wellenreuther M. 2014. De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. BMC genomics, 15(1), 1.

Chen M-Y, Chaw S-M, Wang J-F, Villanueva RJT, Nuneza OM, Lin C-P. 2015. Mitochondrial genome of a flashwing demoiselle, *Vestalis melania* from the Philippine Archipelago. Mitochondrial DNA, 26, 720-721.

Hadrys H, Balick M, Schierwater B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular ecology, 1, 55-63.

Hammers M, Van Gossum H. 2008. Variation in female morph frequencies and mating frequencies: random, frequency-dependent harassment or male mimicry? Animal Behaviour, 76(4), 1403-1410

Laslett D, Canbäck B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics, 24, 172-175.

Lorenzo-Carballa MO, Thompson DJ, Cordero-Rivera A, Watts PC. 2014. Next generation sequencing yields the complete mitochondrial genome of the scarce blue-tailed damselfly, *Ischnura pumilio*. Mitochondrial DNA, 25, 247-248.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research, 25, 955-964.

Rivera AC, Sánchez-Guillén RA. 2007. Male-like females of a damselfly are not preferred by males even if they are the majority morph. Animal Behaviour, 74(2), 247-252.

Simon S, Strauss S, von Haeseler A, Hadrys H. 2009. A phylogenomic approach to resolve the basal pterygote divergence. Molecular biology and evolution, 26(12), 2719-2730.

Simon S, Hadrys H. 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda. Molecular phylogenetics and evolution, 69(2), 393-403.

Simon S, Hadrys H. 2014. Phylogeny of the most species rich group on Earth, the Pterygota: ancient problems, living hypotheses and bridging gaps, Deep Metazoan Phylogeny: The Backbone of the Tree of Life: New insights from analyses of molecules, morphology, and theory of data analysis. De Gruyter

Swofford DL. 2002. PAUP* phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland

Tang M, Tan M, Meng G, Yang S, Su X, Liu S, Song W, Li Y, Wu Q, Zhang A. 2014. Multiplex sequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mito-metagenomics. Nucleic acids research, 42, e166-e166.

Yu P, Cheng X, Ma Y, Yu D, Zhang J. 2014. The complete mitochondrial genome of *Brachythemis contaminata* (Odonata: Libellulidae). Mitochondrial DNA, 1-2.

The complete mitochondrial genome of the emperor dragonfly
Anax imperator **LEACH, 1815 (Aeshnidae: Odonata) via NGS sequencing**

Rebecca Herzog[1], Hans-Jürgen Osigus[1], Wiebke Feindt[1],
Bernd Schierwater[1,2], and Heike Hadrys[1, 2]

[1]ITZ, Ecology and Evolution, University of Veterinary Medicine Hannover, Bünteweg 17d,
30559 Hannover, Germany
[2]EEB, Yale University, 165 Prospect Street, New Haven, CT 06511, USA

* Author for correspondence: Rebecca Herzog
E-mail: rebecca.herzog@ecolevol.de

The complete mitochondrial genome of the emperor dragonfly

Anax imperator LEACH, 1815 (Aeshnidae: Odonata) via NGS sequencing

Abstract

Here we report the complete mitochondrial genome of the emperor dragonfly, *Anax imperator* (Odonata: Aeshnidae) as the first of its genus. Data were generated via next generation sequencing (NGS) and assembled using an iterative approach. The typical metazoan set of 37 genes (13 proteincoding genes, 22 tRNA genes, and 2 rRNA genes) was detected in the same gene order as in other odonate mitogenomes. However, only three intergenic spacer regions are present in *A. imperator* lacking the distinct s5 spacer, which was regarded as informative feature of the odonate suborder Anisoptera (dragonflies) but absent in Zygoptera (damselflies). With 16,087 bp, it is the longest anisopteran mitogenome to date, mainly due to the long A+T-rich control region of 1,291bp.

The emperor dragonfly, *Anax imperator,* is a widespread and common species in the old world inhabiting all types of standing and slow running freshwater ecosystems. It was one of the first odonate species for which a recent range shift northward (*e.g.* Parr 2010) and towards higher altitudes (Westermann 2003c; Hunger et al. 2006) was noticed due to global climate change. The first records of this species in Sweden were 2002 (Ott 2010). In only 11 years *A. imperator* crossed a distance of 970 km northwards through Scandinavia (Nielsen 1998; Lejfelt-Sahlén 2007). The larvae of this large dragonfly species are known to be very aggressive (*e.g.* Beutler 1985) and will invade and influence the native species composition of freshwater ecosystems. Genetic and comparative genomic studies on range shift, expansion and adaptive potential of this species are of great interest to further elucidate the impact of global change on flying insects. To date for *A. imperator* a panel of 10 nuclear microsatellite loci and partial mitochondrial genes (*cox1, nad1* and both *rRNAs*) were established so far to serve in various phylogenetic studies (Misof et al. 2001; Hadrys et al. 2007; Fleck et al. 2008; Rach et al. 2008; Bergmann et al. 2013). To consequently proceed towards a comparative genomic approach one first step is the unraveling and comparison of mitogenomes, e.g. their gene content, arrangements and genealogical relationships.

As for the *A. imperator* mitogenome a standard phenol-chloroform protocol by Hadrys et al. (1992) was used to extract total genomic DNA from flight muscles of a single individual collected in Southern France (43°36'17.7"N 4°48'34.4"E). DNA was submitted for library

preparation and whole genome sequencing on an llumina HiSeq2000 (75 bp paired-end reads) to the Yale Center for Genome Analyses (YCGA, http://www.ycga.yale.edu). Different mitochondrial gene sequences containing partial *nad1, cox1, 12S rRNA* and *16S rRNA* genes (accession numbers: KC912228.1, KF584974.1, EU477652.1 and EU183256.1) were used as reference seeds for a subsequent assembly employing Genious v.8.1.5 (http://www.geneious.com/). For mitochondrial genome annotation the MITOS WebServer (mitos.bioinf.uni-leipzig.de/index.py) was applied and results were checked manually using BLAST (Altschul et al. 1990) and available odonate mitochondrial genomes (e.g. Yu et al. 2014; Chen et al. 2015). Transfer RNA genes were predicted using both, the tRNAscan-SE v.1.21 Search Server (Lowe & Eddy 1997) and ARWEN v.1.2 (Laslett & Canbäck 2008).

The complete circular mitochondrial genome sequence of *A. imperator* (GenBank accession number #KX161814) with the length of 16,087 bp is the largest known among Anisoptera. It exhibits the standard metazoan gene content of 37 genes, comprising 13 protein-coding genes, 22 tRNA genes and two rRNA genes which are identically arranged as in the few other odonate mitochondrial genomes (e.g. Simon et al. 2013; Lorenzo-Carballa et al. 2014; Chen et al. 2015; Yu et al. 2014; Feindt et al. 2016). Overall base frequency is 76.0% AT-biased, for the 1,291 bp long control (A+T rich) region even 93.5%. All standard mitochondrial invertebrate start codons are found, in detail ATT (*nad5*), ATA (*nad2, nad3*), TTG (*cox1, nad1*), ATC (*atp8, nad6*) and ATG (*cox2, atp6, cox3, nad4, nad4l, cob*). Two proteins (*cox2, nad5*) possess a single T as an incomplete stop codon, requiring post-transcriptional polyadenylation whereas all others protein-coding genes use TAA as stop codon (Table 1). The gene length of tRNA genes ranges from 65 bp to 73 bp and all tRNAs can be folded in the typical cloverleaf structure, except the D-replacement tRNA *trnS1*. Further, two pseudo-tRNA genes were detected by the tRNA prediction software ARWEN v.1.2 (Laslett & Canbäck 2008) which were both D-Loop tRNAs and located inside the *cox2* sequence and in *trnA/trnR*, respectively. Therefore, their functionality remains questionable.

However, in contrast to the known other anisopteran mitogenomes, only three intergenic spacer regions were discovered (see Table 1). These are located between *trnY/cox1, trnT/trnP* and *trnS2/nad1*. They are also present in other odonates (Anisoptera and Zygoptera), e.g. *Ischnura elegans* (Feindt et al. 2016), *Ischnura pumilio* (Lorenzo-Carballa et al. 2014), *Megaloprepus caerulatus* (Feindt et al. 2016) or *Brachythemis contaminata* (Yu et al. 2014). The latter, an anisopteran species additionally shows a fourth spacer region between *nad1/trnL2* that is asserted to be typical for Anisopterans and lacking in Zygopterans (Lin et al. 2010). This spacer, commonly called *s5* (though counting and numbering spacer regions is not consistent between most mitogenome publications) is not present in *Anax*. Consequently, the absence of this spacer refutes the theory of being a putative distinctive feature between Anisoptera and Zygoptera and stresses the necessity to analyze more mitogenomes within Odonata to allow stronger, reliable assumptions about phylogenetically informative mtDNA characteristics. The phylogenetic position of A. imperator in the context of all available anisopteran mitogenomes to date (3 May 2016) is displayed in Figure 1 and so far consistent with other gene tree phylogenies.
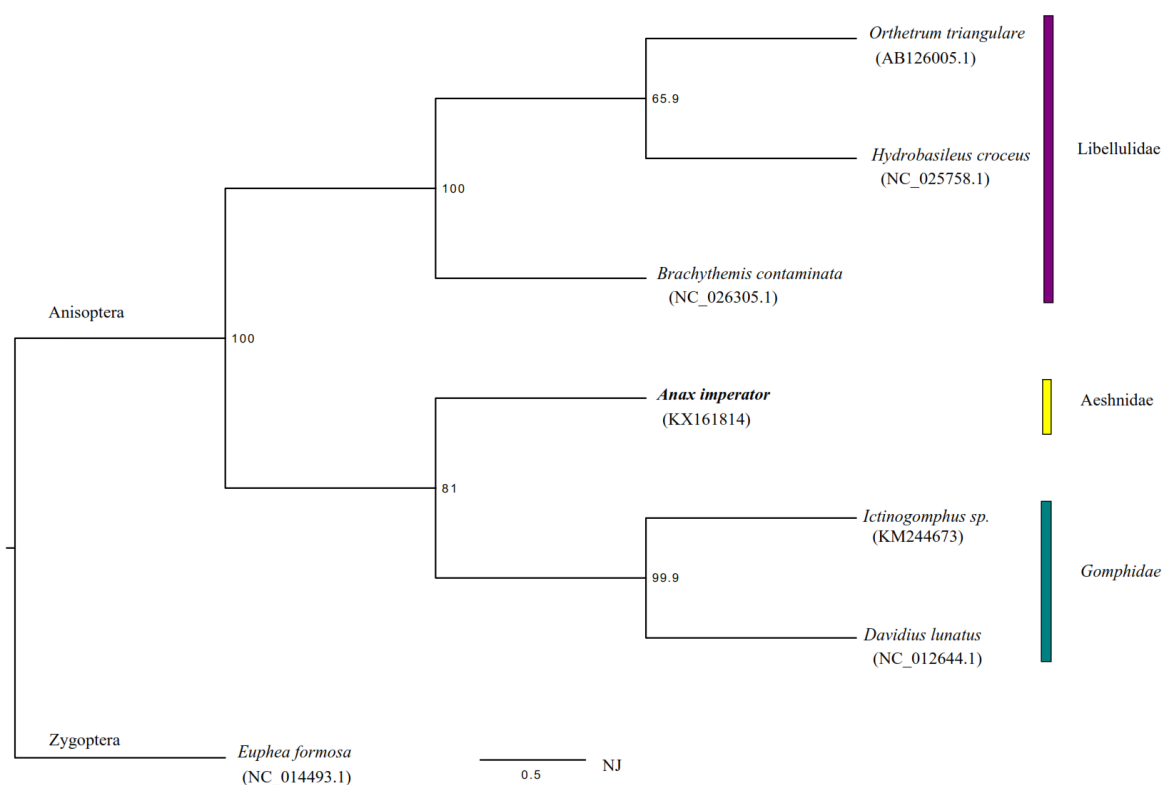
Figure 1. Neighbour-Joining Tree of *A. imperator* within all available anisopteran odonate species (03 May 2016): *Orthetrum triangulare* (AB126005.1), *Hydrobasileus croceus* (NC_025758.1), *B. contaminata* (NC_026305.1), *Ictinogomphus* sp. (KM244673) and *Davidius lunatus* (NC_012644.1). The phylogeny was reconstructed based on 13 mitochondrial protein-coding genes via Paup with 1,000 bootsrap replicates and *Euphea formosa* (NC_014493.1) as an outgroup.

Disclosure statement

The authors declare no conflict of interest to other working groups.

Funding

Table 1. Mitochondrial genome organization and gene content of *A. imperator* with detailed description of gene boundaries, strand, gene length (in bp) as well as start and stop codons for protein-coding genes and anticodons for tRNA genes*, respectively.

Gene/region	Strand	Start position	Stop position	Length (bp)	Anti/Start codon	Stop codon
trnI	+	214	281	68	GAT	/
trnQ	-	278	347	70	TTG	/
trnM	+	352	420	69	CAT	/
nad2	+	424	1,419	996	ATA	TAA
trnW	+	1,417	1,487	71	TCA	/
trnC	-	1,479	1,543	65	GCA	/
trnY	-	1,545	1,613	69	CTA	/
s1	n.a.	1,614	1,653	40	/	/
cox1	+	1,654	3,192	1,539	TTG	TAA
trnL2	+	3,187	3,256	70	TAA	/
cox2	+	3,256	3,943	688	ATG	T(aa)
trnK	+	3,944	4,016	73	CTT	/
trnD	+	4,016	4,084	69	GTC	/
atp8	+	4,084	4,245	162	ATC	TAA
atp6	+	4,239	4,916	678	ATG	TAA
cox3	+	4,916	5,704	789	ATG	TAA
trnG	+	5,704	5,768	65	TCC	/
nad3	+	5,766	6,122	357	ATA	TAA
trnA	+	6,122	6,190	69	TGC	/
trnR	+	6,189	6,258	70	TCG	/
trnN	+	6,258	6,324	67	GTT	/
trnS1	+	6,325	6,392	68	GCT	/
trnE	+	6,392	6,460	69	TTC	/
trnF	-	6,459	6,526	68	GAA	/
nad5	-	6,525	8,254	1,730	ATT	T(aa)
trnH	-	8,255	8,322	68	GTG	/
nad4	-	8,322	9,665	1,344	ATG	TAA
nad4l	-	9,659	9,952	294	ATG	TAA
trnT	+	9,954	10,022	69	TGT	/
s2	n.a.	10,023	10,045	23	/	/
trnP	-	10,046	10,111	66	TGG	/
nad6	+	10,113	10,634	522	ATC	TAA
cob	+	10,634	11,767	1,134	ATG	TAA
trnS2	+	11,766	11,832	67	TGA	/
s3	n.a.	11,833	11,849	17	/	/
nad1	-	11,850	12,800	951	TTG	TAA
trnL1	-	12,801	12,868	68	TAG	/
l-rRNA	-	12,810	14,180	1,371	/	/
trnV	-	14,167	14,236	70	TAC	/
s-rRNA	-	14,239	15,008	770	/	/
A+T-rich (control) region	n.a.	15,009	212	1,291	/	/

* Transfer RNAs are given in the one-letter amino acid code with the corresponding anticodons. Intergenic spacer regions are numbered (s1–s3).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Bergmann T, Rach J, Damm S, DeSalle R, Schierwater B, Hadrys H. 2013. The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. Mol Ecol Resources. 13:1069–1081.

Beutler H. 1985. [Freiland-Daten zur Koexistenz von Aeshnidenlarven]. Entomol Nachrichten Und Berichte. 29:73–75. [in German].

Chen M-Y, Chaw S-M, Wang J-F, Villanueva RJT, Nuneza OM, Lin C-P. 2015. Mitochondrial genome of a flashwing demoiselle, *Vestalis melania* from the Philippine Archipelago. Mitochondrial DNA. 26:720–721.

Feindt W, Herzog R, Osigus H-J, Hadrys H. 2016. Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* (Vander Linden, 1820). Mitochondrial DNA. 1:547–576.

Feindt W, Osigus H-J, Herzog R, Mason CE, Schierwater B, Hadrys H. 2016. The complete mitochondrial genome of the Neotropical helicopter damselfly *Megaloprepus caerulatus* Odonata: Zygoptera) assembled from next generation sequencing data. Mitochondrial DNA Part B. 1:497–499.

Fleck G, Ullrich B, Brenk M, Wallnisch C, Orland M, Bleidissel S, Misof B. 2008. A phylogeny of anisopterous dragonflies (Insecta, Odonata) using mtRNA genes and mixed nucleotide/doublet models. J Zool Syst Evol Res. 46:310–322.

Hadrys H, Balick M, Schierwater B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecol. 1:55–63.

Hadrys H, Timm J, Streit B, Giere S. 2007. A panel of microsatellite markers to study sperm precedence patterns in the emperor dragonfly *Anax imperator* (Odonata: Anisoptera). Mol Ecol Notes. 7:296–298.

Hunger H, Schiel F-J, Kunz B. 2006. [Verbreitung und Ph€anologie der Libellen BadenW€urttembergs (Odonata)]. Libellula Suppl. 7:15–188. [in German].

Laslett D, Canb€ack B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics. 24:172–175.

Lejfelt-Sahlén A. 2007. [Trollsländefauna i förvandling]. Fauna Och Flora. 102:44–44. [in Swedish].

Lin CP, Chen MY, Huang JP. 2010. The complete mitochondrial genome and phylogenomics of a damselfly, *Euphaea formosa* support a basal Odonata within the Pterygota. Gene. 468:20–29.

Lorenzo-Carballa MO, Thompson DJ, Cordero-Rivera A, Watts PC. 2014. Next generation sequencing yields the complete mitochondrial genome of the scarce blue-tailed damselfly, *Ischnura pumilio*. Mitochondrial DNA. 25:247–248.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Misof B, Rickert AM, Buckley TR, Fleck G, Sauer KP. 2001. Phylogenetic signal and its decay in mitochondrial SSU and LSU rRNA gene fragments of Anisoptera. Mol Biol Evol. 18:27–37.

Nielsen OF. 1998. [Guldsmede-nyt fra Danmark 1997]. Nordic Odonatol Soc Newslett. 4:4. [in Danish].

Ott J. 2010. Dragonflies and climatic change-recent trends in Germany and Europe. BioRisk. 5:253.

Parr AJ. 2010. Monitoring of Odonata in Britain and possible insights into climate change. In: Ott J (Ed) Monitoring Climatic Change with Dragonflies. BioRisk. 5:127–139.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proc Royal Soc London B: Biol Sci. 275:237–247.

Simon S, Hadrys H. 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda. Mol Phylogenet Evol. 69:393–403.

Westermann K. 2003. [Zum Status der Großen Königslibelle im höheren Schwarzwald]. Naturschutz Am Südlichen Oberrhein. 4:81–85. [in German].

Yu P, Cheng X, Ma Y, Yu D, Zhang J. 2014. The complete mitochondrial genome of *Brachythemis contaminata* (Odonata: Libellulidae). Mitochondrial DNA. 27:1–2.

APPENDIX

A1 Population genetic structure of the Neotropical damselfly *M. caerulatus*

Table A1.1: Pseudostigmatidae species and *Teinobasis fortis* (outgroup) included in the phylogenetic analyses

Table A1.2: Pairwise nucleotide sequence divergence in percent between *Mecistogaster* species

Table A1.1 Pseudostigmatidae species and *Teinobasis fortis* (outgroup) included in the phylogenetic analyses and its corresponding individual GeneBank accession numbers for the two marker genes (16S rRNA, ND1).

Genus	Species	16S	ND1
Coryphagrion	*grandis*	DQ642980.1	DQ642998.1
Pseudostigma	*aberrans*	DQ642984.1	DQ642995.1
Mecistogaster	*modesta*	JQ966657.1	/
	lucretia	JQ966655.1	JQ966607.1
	linearis	DQ642982.1	DQ642993.1
	ornata	DQ642983.1	DQ642994.1
	jocaste	JQ966659.1	JQ966611.1
	asticta	JQ966654.1	/
	martinezi	JQ966656.1	JQ966609.1
Anonisma	*abnorme*	JQ966651.1	/
Microstigma	*anomalum*	EU055119.1	JQ966613.1
	rotundatum	JQ966661.1	/
Megaloprepus	*"populations"*	Accession numbers will be added.	
Teinobasis	*fortis*	JQ966663.1	JQ966614.1

Table A1.2 Pairwise nucleotide sequence divergence in percent between *Mecistogaster* species of the Pseudostigmatidae family using Kimura's 2-parameter model (K2P) for the 16S rRNA sequence marker.

	M. modesta	*M. asticta*	*M.jocaste*	*M. martinezi*	*M. liniaris*	*M. lucretia*	*M. ornata*
M. modesta							
M. asticta	4.98						
M. jocaste	6.95	3.09					
M. martinezi	4.98	0.00	3.09				
M. liniaris	5.30	4.04	5.99	4.03			
M. lucretia	5.95	4.67	6.65	4.67	12.11		
M. ornata	6.95	7.64	8.32	7.64	4.66	4.65	

A2   *Megaloprepus'* phylogeography unravels cryptic speciation

File A2.1: Population Genetics
File A2.2: Phylogenetics
File A2.3: Morphometrics
File A2.4: Species Distribution Model

File A2.1. Supplementary material for the population genetics
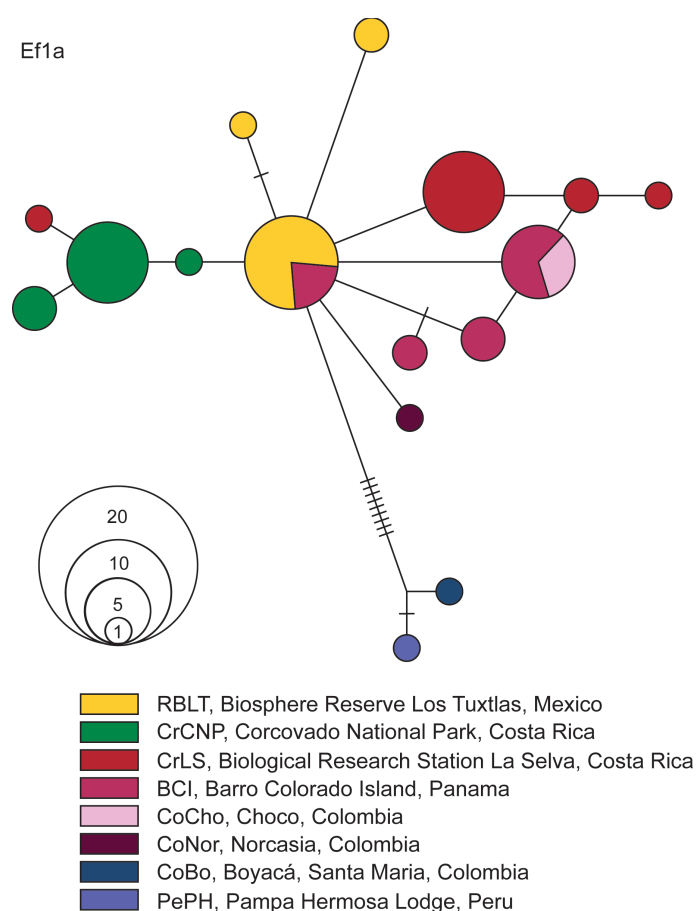
Figure A2.1.1: Haplotype network showing the genealogical relationships of *Megaloprepus* species from southern Mexico to Peru using the nuclear sequence marker Ef1α (668 bp) visualized by using statistical parsimony with a 95% connection limit in TCS vers. 1.2.1. Here a small section of 48 individuals was used and the individuals are displayed in only one single network, where the eastern Andean specimens are distant by either 10 or 11 mutation steps.
A circle represents each haplotype, which size correlates positively to the numbers of individuals owning the same haplotype. Their coloration corresponds to the geographical origin of the haplotype. If two or more populations share haplotypes, the coloration of the haplotype appears in a pie diagram representing the percentage of each population containing the haplotype.

Table A2.1.1: Sample origin for all individuals used in the population comparison with their corresponding NCBI accession numbers. Tissue samples taken from larvae were terminal grills and in adults always the right middle leg. The accession numbers in grey are downloaded from NCBI. For the GPS coordinates please compare the species description of *Megaloprepus* (Feindt and Hadrys submitted).

Sample location	Species ID	collection date	sex / larvae	collected by	Accession numbers				
					16S	ND1	CO1	Ef1α	ITS I+II
	LT 1	4/20/10	larvae	A. Schlötelburg	KF895223	KF895333	MH939548	MH939851	MH939851
	LT 2	4/20/10	larvae	A. Schlötelburg	KF895224	KF895334	MH939549	MH939852	MH939852
	LT 3	4/20/10	larvae	A. Schlötelburg	KF895225	KF895335	MH939550	/	/
	LT 4	4/20/10	larvae	A. Schlötelburg	KF895226	KF895336	MH939551	/	/
	LT 6	4/20/10	larvae	A. Schlötelburg	KF895227	KF895337	MH939552	MH939853	MH939853
	LT 7	4/20/10	larvae	A. Schlötelburg	KF895228	KF895338	MH939553	/	/
	LT 10	4/21/10	larvae	A. Schlötelburg	KF895229	KF895339	MH939554	/	/
	LT 13	4/22/10	larvae	A. Schlötelburg	KF895230	KF895340	MH939555	/	/
	LT 14	4/23/10	larvae	A. Schlötelburg	KF895231	KF895341	MH939556	/	/
	LT 15	4/23/10	larvae	A. Schlötelburg	KF895232	KF895343	MH939557	/	/
	LT 16	4/23/10	larvae	A. Schlötelburg	KF895233	KF895344	MH939558	MH939854	MH939854
Biosphere Reserve Los Tuxtlas (Veracruz), Mexico	LT 17	4/23/10	larvae	A. Schlötelburg	KF895234	KF895345	MH939559	MH939855	MH939855
	LT 18	4/23/10	larvae	A. Schlötelburg	KF895235	KF895346	MH939560	/	/
	LT 19	4/24/10	female	A. Schlötelburg	KF895236	KF895347	MH939561	MH939856	MH939856
	LT 20	4/24/10	female	A. Schlötelburg	KF895237	KF895348	MH939562	MH939857	MH939857
	LT 21	4/24/10	female	A. Schlötelburg	KF895238	KF895349	MH939563	MH939858	MH939858
	MeLT_MC42	3/26/12	male	W. Feindt	MH939462	MH939739	MH939564	/	/
	MeLT_MC44	3/26/12	male	W. Feindt	MH939463	MH939740	MH939565	MH939860	MH939859
	MeLT_L11	3/25/12	larvae	W. Feindt	KF895239	KF895331	MH939566	/	/
	MeLT_L13	3/25/12	larvae	W. Feindt	MH939464	MH939741	MH939567	/	/
	MeLT_L14	3/26/12	larvae	W. Feindt	MH939465	MH939742	MH939568	/	/
	MeLT_L15	3/26/12	larvae	W. Feindt	MH939466	MH939743	MH939569	/	/
	MeLT_L17	3/27/12	larvae	W. Feindt	KF895240	KF895332	MH939570	MH939859	MH939860
	MeLT_OF1	7/4/98	female	O. Fincke	MH939467	MH939744	MH939571	/	/
	MeLT_OF2	7/4/98	female	O. Fincke	MH939468	MH939745	MH939572	/	/
	MeLT_OF3	7/4/98	male	O. Fincke	MH939469	MH939746	MH939573	/	/
	MeLT_OF4	7/4/98	male	O. Fincke	MH939470	MH939747	MH939574	/	/
National Park Laguna Lachuá,	GuLL_L8	2/20/12	larvae	W. Feindt	MH939471	MH939748	MH939575	/	/
	GuLL_L10	2/20/12	larvae	W. Feindt	MH939472	MH939749	MH939576	/	/

Location	Sample	Date	Type	Collector					
Guatemala	**GuLL_L11**	2/20/12	larvae	W. Feindt	MH939473	MH939750	MH939577	/	/
Rio Bravo, Guatemala	**GuRB_MC45**	7/20/12	female	J. Monzón Sierra	MH939489	MH939766	MH939593	/	/
Natural Reserve Cerro San Gil, Guatemala	**GuCSG_L3**	2/3/12	larvae	W. Feindt	MH939474	MH939751	MH939578	/	/
	GuCSG_L5	2/3/12	larvae	W. Feindt	MH939475	MH939752	MH939579	/	/
	GuCSG_L6	2/3/12	larvae	W. Feindt	MH939476	MH939753	MH939580	/	/
	GuCSG_L10	2/3/12	larvae	W. Feindt	MH939477	MH939754	MH939581	/	/
	GuCSG_L12	2/3/12	larvae	W. Feindt	MH939478	MH939755	MH939582	/	/
	GuCSG_L13	3/3/12	larvae	W. Feindt	MH939479	MH939756	MH939583	/	/
	GuCSG_L13.1	3/3/12	larvae	W. Feindt	MH939480	MH939757	MH939584	/	/
	GuCSG_L16	11/15/12	larvae	W. Feindt	MH939481	MH939758	MH939585	/	/
	GuCSG_L17	11/15/12	larvae	W. Feindt	MH939482	MH939759	MH939586	/	/
	GuCSG_L18	11/15/12	larvae	W. Feindt	MH939483	MH939760	MH939587	/	/
	GuCSG_L19	11/15/12	larvae	W. Feindt	MH939484	MH939761	MH939588	/	/
	GuCSG_L20	11/15/12	larvae	W. Feindt	MH939485	MH939762	MH939589	/	/
	GuCSG_L22	11/16/12	larvae	W. Feindt	MH939486	MH939763	MH939590	/	/
	GuCSG_L23	11/16/12	larvae	W. Feindt	MH939487	MH939764	MH939591	/	/
	GuCSG_MC46	7/9/10	male	J. Monzón Sierra	MH939488	MH939765	MH939592	/	/
Pico Bonito National Park, Honduras	**HnPb_L1**	1/27/12	larvae	W. Feindt	MH939490	MH939767	MH939594	/	/
	HnPb_L2	12/4/12	larvae	W. Feindt	MH939491	MH939768	MH939595	/	/
	HnPb_L3	12/4/12	larvae	W. Feindt	MH939492	MH939769	MH939596	/	/
	HnPb_L6	12/9/12	larvae	W. Feindt	MH939493	MH939770	MH939597	/	/
	HnPb_L7	12/9/12	larvae	W. Feindt	MH939494	MH939771	MH939598	/	/
	HnPb_L8	12/9/12	larvae	W. Feindt	MH939495	MH939772	MH939599	/	/
	HnPb_L9	12/9/12	larvae	W. Feindt	MH939496	MH939773	MH939600	/	/
	HnPb_L10	12/9/12	larvae	W. Feindt	MH939497	MH939774	MH939601	/	/
	HnPb_L12	12/9/12	larvae	W. Feindt	MH939498	MH939775	MH939602	/	/
	HnPb_L13	12/9/12	larvae	W. Feindt	MH939499	MH939776	MH939603	/	/
	HnPb_L14	12/9/12	larvae	W. Feindt	MH939500	MH939777	MH939604	/	/
	HnPb_L15	12/10/12	larvae	W. Feindt	MH939501	MH939778	MH939605	/	/
	HnPb_L17	12/14/12	larvae	W. Feindt	MH939502	MH939779	MH939606	/	/
	HnPb_L18	12/14/12	larvae	W. Feindt	MH939503	MH939780	MH939607	/	/
	CNP 3	12/15/09	larvae	W. Feindt	KF895162	KF895283	MH939658	/	/
	CNP7	12/20/09	larvae	W. Feindt				MH939861	MH939861
	CNP8	12/20/09	larvae	W. Feindt				MH939862	MH939862
	CNP 13.1	12/16/09	larvae	W. Feindt	KF895164	MH939802	MH939659	MH939863	MH939863

Location	Sample ID	Date		Collector					
Corcovado National Park, Costa Rica	**CNP 13.2**	12/16/09	larvae	W. Feindt	KF895165	MH939803	MH939660	MH939864	MH939864
	CNP 14	12/16/09	female	W. Feindt	KF895166	KF895285	MH939661	/	/
	CNP 15	12/16/09	female	W. Feindt	MH939525	KF895286	MH939662	/	/
	CNP16.1	12/16/09	larvae	W. Feindt	MH939526	MH939804	MH939663	/	/
	CNP16.2	12/16/09	larvae	W. Feindt	MH939527	MH939805	MH939664	/	/
	CNP16.3	12/16/09	larvae	W. Feindt	MH939528	MH939806	MH939665	/	/
	CNP16.4	12/16/09	larvae	W. Feindt	MH939529	MH939807	MH939666	/	/
	CNP 18	12/16/09	larvae	W. Feindt	KF895167	KF895287	MH939667	/	/
	CNP 20	12/17/09	larvae	W. Feindt	KF895168	KF895288	MH939668	/	/
	CNP 25	12/17/09	larvae	W. Feindt	KF895169	KF895289	MH939669	MH939866	MH939866
	CNP 29	12/18/09	larvae	W. Feindt	KF895170	KF895290	MH939670	/	/
	CNP 30	12/18/09	larvae	W. Feindt	KF895171	KF895291	MH939671	/	/
	CNP 31	12/18/09	larvae	W. Feindt	KF895172	KF895292	MH939672	/	/
	CNP 32	12/19/09	male	W. Feindt	KF895173	KF895293	MH939673	/	/
	CNP 36	12/19/09	larvae	W. Feindt	KF895174	KF895294	MH939674	/	/
	CNP 45	12/20/09	larvae	W. Feindt	KF895175	KF895295	MH939675	/	/
	CNP 48.1	12/21/09	larvae	W. Feindt	KF895177	MH939808	MH939676	MH939867	MH939867
	CNP 48.2	12/21/09	larvae	W. Feindt	MH939530	KF895297	MH939677	/	/
	CNP 51.1	12/22/09	male	W. Feindt	KF895178	MH939809	MH939678	MH939868	MH939868
	CNP 51.2	12/22/09	male	W. Feindt	KF895179	KF895299	MH939679	/	/
	CNP 51.3	12/22/09	female	W. Feindt	KF895180	KF895300	MH939680	/	/
	CrCNP_MC5	12/18/11	female	W. Feindt	KF895186	KF895277	MH939681	MH939869	MH939869
	CrCNP_MC7	12/18/11	female	W. Feindt	KF895187	KF895278	MH939682	MH939870	MH939870
	CrCNP_MC9	12/20/11	male	W. Feindt	KF895188	KF895279	MH939683	MH939871	MH939871
	CrCNP_MC10	12/20/11	female	W. Feindt	KF895189	MH939810	MH939684	/	/
	CrCNP_MC11	12/22/11	male	W. Feindt	KF895190	KF895280	MH939685	/	/
	CrCNP_MC12	12/22/11	male	W. Feindt	MH939531	KF895281	MH939686	/	/
	CrCNP_MC13	12/22/11	female	W. Feindt	KF895191	KF895282	MH939687	/	/
	CrCNP_L3	12/15/11	larvae	W. Feindt	KF895181	KF895272	MH939688	/	/
	CrCNP_L14	12/18/11	larvae	W. Feindt	KF895182	KF895273	MH939689	/	/
	CrCNP_L15	12/18/11	larvae	W. Feindt	KF895183	KF895274	MH939690	/	/
	CrCNP_L19	4/13/12	larvae	W. Feindt	KF895184	KF895275	MH939691	MH939865	MH939865
	CrCNP_L20	4/13/12	Larvae	W. Feindt	KF895185	KF895276	MH939692	/	/
Biological Reserve Indio	**NiBa_MC14**	12/30/11	male	W. Feindt	MH939506	MH939783	MH939610	/	/
	NiBa_MC15	12/31/11	male	W. Feindt	MH939507	MH939784	MH939611	/	/

Maíz, Nicaragua	**NiBa_MC16**	1/1/12	male	W. Feindt	MH939508	MH939785	MH939612	/	/
	NiBa_MC18	1/8/12	female	W. Feindt	MH939509	MH939786	MH939613	/	/
	NiBa_MC19	1/9/12	male	W. Feindt	MH939510	MH939787	MH939614	/	/
	NiBa_MC20	1/9/12	male	W. Feindt	MH939511	MH939788	MH939615	/	/
	NiBa_MC21	1/9/12	male	W. Feindt	MH939512	MH939789	MH939616	/	/
	NiBa_MC22	1/10/12	male	W. Feindt	MH939513	MH939790	MH939617	/	/
	NiBa_MC23	1/10/12	male	W. Feindt	MH939514	MH939791	MH939618	/	/
	NiBa_MC24	1/10/12	male	W. Feindt	MH939515	MH939792	MH939619	/	/
	NiBa_MC26	1/11/12	male	W. Feindt	MH939516	MH939793	MH939620	/	/
	NiBa_MC27	1/11/12	female	W. Feindt	MH939517	MH939794	MH939621	/	/
	NiBa_MC28	1/17/12	male	W. Feindt	MH939518	MH939795	MH939622	/	/
	NiBa_MC29	1/17/12	female	W. Feindt	MH939519	MH939796	MH939623	/	/
	NiBa_L4	1/7/12	larvae	W. Feindt	MH939504	MH939781	MH939608	/	/
	NiBa_L5	1/7/12	larvae	W. Feindt	MH939505	MH939782	MH939609	/	/
	LS 1	2/17/10	male	B. Gericke	KF895193	KF895301	MH939624	MH939840	MH939840
	LS 2	2/22/10	male	B. Gericke	KF895194	KF895302	MH939625	MH939841	MH939841
	LS 3	2/22/10	male	B. Gericke	KF895195	KF895303	MH939626	/	/
	LS 4	2/23/10	male	B. Gericke	KF895196	KF895304	MH939627	MH939842	MH939842
	LS 5	2/23/10	male	B. Gericke	KF895197	KF895305	MH939628	/	/
	LS 6	3/6/10	male	B. Gericke	KF895198	KF895306	MH939629	MH939843	MH939843
	LS 7	3/9/10	female	B. Gericke	KF895199	KF895307	MH939630	MH939844	MH939844
	LS 8	3/10/10	male	B. Gericke	KF895200	KF895308	MH939631	MH939845	MH939845
	LS 9	3/11/10	female	B. Gericke	KF895201	KF895309	MH939738	MH939846	MH939846
Biological Research Station La Selva, Costa Rica	**LS 10**	3/17/10	male	B. Gericke	KF895202	KF895310	MH939632	/	/
	LS 11	3/17/10	male	B. Gericke	KF895203	KF895311	MH939633	MH939847	MH939847
	LS 12	3/17/10	male	B. Gericke	KF895204	KF895312	MH939634	MH939848	MH939848
	LS 13	3/24/10	male	B. Gericke	KF895205	KF895313	MH939635	/	/
	LS 14	3/24/10	male	B. Gericke	MH939520	KF895314	MH939636	/	/
	LS 15	3/25/10	male	B. Gericke	KF895206	KF895315	MH939637	/	/
	LS 16	3/25/10	male	B. Gericke	KF895207	KF895316	MH939638	/	/
	LS 17	3/29/10	male	B. Gericke	KF895208	KF895317	MH939639	MH939849	MH939849
	LS 18	3/30/10	male	B. Gericke	KF895209	KF895318	MH939640	/	/
	LS 19	4/8/10	male	B. Gericke	KF895210	KF895319	MH939641	/	/
	LS 20	4/8/10	male	B. Gericke	KF895211	KF895320	MH939642	/	/
	LS 21	4/10/10	male	B. Gericke	KF895212	KF895321	MH939643	/	/

	ID	Date	Sex	Collector					
	LS 22	4/13/10	male	B. Gericke	KF895213	KF895322	MH939644	/	/
	LS 23	4/13/10	male	B. Gericke	KF895214	KF895323	MH939645	/	/
	LS 24	4/13/10	male	B. Gericke	KF895215	KF895324	MH939646	/	/
	LS 25	4/15/10	male	B. Gericke	KF895216	KF895325	MH939647	/	/
	LS 26	4/15/10	male	B. Gericke	KF895217	KF895326	MH939648	/	/
	LS 27	4/15/10	male	B. Gericke	KF895218	KF895327	MH939649	/	/
	LS 28	4/18/10	male	B. Gericke	KF895219	KF895328	MH939650	/	/
	LS 29	4/18/10	male	B. Gericke	KF895220	MH939797	MH939651	/	/
	LS 30	4/18/10	female	B. Gericke	KF895221	MH939798	MH939652	/	/
	LS 31	4/19/10	male	B. Gericke	KF895222	KF895329	MH939653	MH939850	MH939850
	LS 32	4/22/10	male	B. Gericke	MH939521	KF895330	MH939654	/	/
	CrLS_MC34	4/1/12	male	W. Feindt	MH939522	MH939799	MH939655	/	/
	CrLS_MC37	4/3/12	male	W. Feindt	MH939523	MH939800	MH939656	/	/
	CrLS_MC39	4/5/12	male	W. Feindt	MH939524	MH939801	MH939657	/	/
Barro Colorado Island (BCI; Smithsonian Tropical Research Institute), Panama	**BCI 1**	12/28/09	male	W. Feindt	KF895130	KF895241	MH939693	MH939829	MH939877
	BCI 2	12/29/09	male	W. Feindt	KF895131	KF895242	MH939694	/	/
	BCI 3	12/29/09	male	W. Feindt	KF895132	KF895243	MH939695	MH939830	MH939878
	BCI 4	12/31/09	male	W. Feindt	KF895133	KF895244	MH939696	/	/
	BCI 5	12/31/09	male	W. Feindt	KF895134	KF895245	MH939697	MH939831	MH939879
	BCI 6	1/1/10	male	W. Feindt	KF895135	KF895246	MH939698	/	/
	BCI 7	1/1/10	male	W. Feindt	KF895136	KF895247	MH939699	MH939832	MH939880
	BCI 8	1/1/10	male	W. Feindt	KF895137	KF895248	MH939700	/	/
	BCI 9	1/1/10	female	W. Feindt	KF895138	KF895249	MH939701	MH939833	MH939881
	BCI 10	12/31/09	male	W. Feindt	KF895139	KF895250	MH939702	/	/
	BCI 11	12/31/09	male	W. Feindt	KF895140	KF895251	MH939703	MH939834	MH939882
	BCI 12	12/31/09	female	W. Feindt	KF895141	KF895252	MH939704	/	/
	BCI 12 II	12/31/09	male	W. Feindt	KF895142	KF895253	MH939705	/	/
	BCI 14	1/2/10	female	W. Feindt	KF895143	MH939811	MH939706	MH939835	MH939883
	BCI 15	1/2/10	male	W. Feindt	KF895144	KF895254	MH939707	/	/
	BCI 16	1/3/10	male	W. Feindt	KF895145	KF895255	MH939708	/	/
	BCI 17	1/4/10	male	W. Feindt	KF895146	KF895256	MH939709	/	/
	BCI 19	1/4/10	female	W. Feindt	KF895148	KF895257	MH939710	/	/
	BCI 20	1/1/10	female	W. Feindt	KF895149	MH939812	MH939711	/	/
	BCI 21	1/2/10	male	W. Feindt	MH939532	MH939813	MH939712	/	/
	BCI 22 w	1/3/10	male	W. Feindt	KF895150	KF895258	MH939713	/	/

	BCI 23	1/2/10	female	W. Feindt	KF895151	KF895260	MH939714	/	/
	BCI 24	1/4/10	male	W. Feindt	KF895152	KF895261	MH939715	MH939836	MH939884
	BCI 25	1/2/10	male	W. Feindt	KF895153	KF895262	MH939716	MH939837	MH939885
	BCI 26	1/5/10	female	W. Feindt	KF895154	KF895263	MH939717	/	/
	BCI 27	1/6/10	female	W. Feindt	KF895155	KF895264	MH939718	/	/
	BCI 28	1/6/10	female	W. Feindt	KF895156	KF895265	MH939719	/	/
	BCI 29	1/6/10	male	W. Feindt	KF895157	MH939814	MH939720	/	/
	BCI 30	1/9/10	male	W. Feindt	KF895158	KF895266	MH939721	/	/
	BCI 31	1/10/10	male	W. Feindt	KF895159	KF895267	MH939722	/	/
	BCI 33	1/11/10	male	W. Feindt	MH939533	KF895268	MH939723	/	/
	BCI 39	1/13/10	female	W. Feindt	MH939534	MH939815	MH939724	/	/
	BCI 49	1/23/10	male	W. Feindt	/	/	/	MH939838	MH939886
	BCI 101	1/28/10	female	W. Feindt	/	/	/	MH939839	MH939887
Chocó, Colombia	**CoCho_MC50**	12/6/05	adult	C. Botero	MH939543	MH939824	MH939733	MH939872	MH939872
	CoCho_MC51	5/28/05	adult	C. Botero	MH939544	MH939825	MH939734	MH939873	MH939873
	CoCho_MC53	10/13/09	adult	C. Bota	MH939539	MH939820	MH939729	/	/
	CoCho_MC54	10/12/09	adult	C. Bota	MH939541	MH939822	MH939731	/	/
	CoCho_MC55	8/1/11	adult	C. Bota	MH939542	MH939823	MH939732	/	/
Norcasia, Colombia	**CoNor_MC52**	4/16/15	female	C. Botero	MH939545	MH939826	MH939735	MH939874	MH939874
Antioquia, Colombia	**CoA_MC56**	12/27/09	adult	C. Bota	MH939535	MH939816	MH939725	/	/
	CoA_MC57	9/12/12	adult	C. Bota	MH939536	MH939817	MH939726	/	/
	CoA_MC58	12/29/09	adult	C. Bota	MH939537	MH939818	MH939727	/	/
	CoA_MC59	12/18/08	adult	C. Bota	MH939538	MH939819	MH939728	/	/
	CoA_MC60	12/3/11	adult	C. Bota	MH939540	MH939821	MH939730	/	/
Boyacá, Santa Maria, Colombia	**CoBo_Mcb2**	5/21/15	male	D. Garcia	MH939547	MH939828	MH939737	MH939876	MH939876
Pampa Hermosa Lodge, Peru	**PePH_Mcb1**	12/10/08	male	T.D. Donnelly	MH939546	MH939827	MH939736	MH939875	MH939875

Table A2.1.2: Genetic diversity indices for the four *Megaloprepus* species separated per sampling locality. Displayed are the indices for ND1 (474 bp), CO1 (641 bp) and 16S (484 bp): *N* = number of individuals, *P* = nucleotide diversity in percent, h = number of haplotypes and Hd = haplotype diversity and S = number of polymorphic sites.

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; GuLL = National Park Laguna Lachua, Guatemala; GuRb = Rio Bravo, Guatemala; GuCSG = Natural Reserve Cerro San Gil, Guatemala; HnPb = Pico Bonito National Park, Honduras; CrCNP = Corcovado National Park, Costa Rica; NiBa = Biological Reserve Indio Maíz, Nicaragua; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoA = Antioquia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru
** last row indicates diversity indices for the entire data set per sequence marker.

Locality	putative species	N	16S				CO1				ND1			
			h	P	Hd	S	h	P	Hd	S	h	P	Hd	S
RBLT	*M. latipennis*	27	5	0.55 ± 0.0018	0.49 ± 0.108	10	5	0.20 ± 0.0004	0.75 ± 0.050	6	5	0.10 ± 0.0003	0.44 ± 0.110	4
GuLL		3	1	0.00 ± 0.0000	0.00 ± 0.000	0	1	0.00 ± 0.0000	0.00 ± 0.00	0	1	0.00 ± 0.0000	0.00 ± 0.000	0
GuRB		1	1	/	/	/	1	/	/	/	1	/	/	/
GuCSG	*Megaloprepus* sp. nov.	15	2	0.40 ± 0.0008	0.48 ± 0.092	4	4	0.34 ± 0.0004	0.73 ± 0.067	5	5	0.40 ± 0.0005	0.73 ± 0.089	5
HnPb		14	1	0.00 ± 0.0000	0.00 ± 0.000	0	2	0.02 ± 0.0002	0.14 ± 0.119	1	2	0.03 ± 0.0003	0.14 ± 0.119	1
CrCNP		35	2	0.01 ± 0.0001	0.06 ± 0.053	1	6	0.10 ± 0.0004	0.32 ± 0.101	1	5	0.11 ± 0.0006	0.22 ± 0.092	8
NiBa	*M. caerulatus*	16	1	0.00 ± 0.0000	0.00 ± 0.000	0	1	0.00 ± 0.0000	0.00 ± 0.000	0	2	0.03 ± 0.0003	0.13 ± 0.106	1
CrLS		35	4	0.12 ± 0.0006	0.26 ± 0.094	8	1	0.00 ± 0.0000	0.00 ± 0.000	0	6	0.12 ± 0.0005	0.27 ± 0.098	6
BCI		32	6	0.23 ± 0.0003	0.73 ± 0.044	4	7	0.19 ± 0.0004	0.68 ± 0.076	9	3	0.03 ± 0.0002	0.12 ± 0.078	2
CoCho		5	3	0.10 ± 0.0045	0.70 ± 0.218	11	4	0.85 ± 0.0024	0.90 ± 0.161	10	5	0.98 ± 0.0018	1.00 ± 0.126	9
CoNor		1	1	/	/	/	1	/	/	/	1	/	/	/
CoA		5	2	0.17 ± 0.0010	0.40 ± 0.237	2	4	0.44 ± 0.0014	0.90 ± 0.161	6	3	0.30 ± 0.0010	0.80 ± 0.164	3
CoBo	*M. brevistigma*	1	1	/	/	/	1	/	/	/	1	/	/	/
PePH		1	1	/	/	/	1	/	/	/	1	/	/	/
Σ		191	31	0.03 ± 0.0010	0.86 ± 0.014	73	39	0.06 ± 0.0011	0.88 ± 0.015	136	41	0.05 ± 0.0010	0.88 ± 0.015	80

Table A2.1.3: Genetic diversity indices for the four *Megaloprepus* species separated per sampling locality. Displayed are the indices for two nuclear genes Ef1α (670 bp) and ITS I+II (661 bp): N = number of individuals, h = number of haplotypes, P = nucleotide diversity in percent, Hd = haplotype diversity and S = number of polymorphic sites.

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; CrCNP = Corcovado National Park, Costa Rica; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru
** last row indicates diversity indices for the entire data set per sequence marker.

Locality	putative species	N	Ef1a				ITS I+II			
			h	P	Hd	S	h	P	Hd	S
RBLT	*M. latipennis*	10	3	0.0011 ± 0.0005	0.511 ± 0.164	3	1	0.0000 ± 0.0000	0.000 ± 0.000	0
CrCNP	*Megaloprepus* sp. nov.	11	3	0.0009 ± 0.0003	0.564 ± 0.134	2	2	0.0003 ± 0.0002	0.182 ± 0.144	1
CrLS	*M. caerulatus*	11	4	0.0015 ± 0.0005	0.600 ± 0.154	4	2	0.0009 ± 0.0001	0.545 ± 0.072	1
BCI		11	4	0.0021 ± 0.0004	0.800 ± 0.075	3	4	0.0005 ± 0.0005	0.491 ± 0.175	3
CoCho		2	1	0.0000 ± 0.0000	0.000 ± 0.000	0	2	0.0016 ± 0.0008	1.000 ± 0.500	1
CoNor		1	1	/	/	/	1	/	/	/
CoBo	*M. brevistigma*	1	1	/	/	/	1	/	/	/
PePH		1	1	/	/	/	1	/	/	/
Σ		48	18	0.0040 ± 0.0009	0.840 ± 0.036	22	14	0.0039 ± 0.0014	0.710 ± 0.042	22

Table A2.1.4: Estimates of evolutionary divergence between and within populations calculated for 14 populations across *Megaloprepus*' distributional range using a concatenated alignment for CO1, ND1 and 16S rDNA (1,595 bp). Analyses of genetic diversification were conducted using the Kimura 2-parameter model (Kimura 1980). Rate variation among sites was modelled with a gamma distribution in MEGA7 (Kumar *et al.* 2016).

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; GuLL = National Park Laguna Lachua, Guatemala; GuRb = Rio Bravo, Guatemala; GuCSG = Natural Reserve Cerro San Gil, Guatemala; HnPb = Pico Bonito National Park, Honduras; CrCNP = Corcovado National Park, Costa Rica; NiBa = Biological Reserve Indio Maíz, Nicaragua; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoA = Antioquia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru

	RBLT	GuLL	GuRB	GuCSG	HnPb	CrCNP	NiBa	CrLS	BCI	CoCho	CoNor	CoA	CoBo	PePH
RBLT	0.28													
GuLL	0.27	0.00												
GuRB	0.66	0.51	n/c											
GuCSG	6.29	6.23	5.71	0.37										
HnPb	6.39	6.33	5.81	0.35	0.02									
CrCNP	6.54	6.50	5.97	0.98	0.94	0.08								
NiBa	7.74	7.88	7.87	8.45	8.53	8.40	0.01							
CrLS	7.77	7.91	7.88	8.47	8.56	8.42	0.04	0.07						
BCI	7.76	7.88	7.78	8.11	8.14	8.08	1.32	1.34	0.15					
CoCho	7.73	7.86	7.84	8.39	8.45	8.34	0.72	0.75	1.02	0.95				
CoNor	8.12	8.26	8.23	9.00	9.08	9.01	1.22	1.25	2.01	1.31	n/c			
CoA	7.84	7.97	7.95	8.69	8.78	8.64	0.40	0.43	1.50	0.86	0.97	0.32		
CoBo	11.76	11.58	11.80	11.23	11.25	11.26	12.03	12.07	11.92	11.87	11.78	12.23	n/c	
PePH	12.13	12.00	12.22	11.22	11.16	11.41	12.81	12.84	12.69	12.68	12.90	13.11	4.16	n/c

Table A2.1.5: Estimates of Evolutionary Divergence over sequence pairs for the four *Megaloprepus* species. Calculations are based on the concatenated alignment of three mitochondrial sequence marker for 191 nucleotide sequences (CO1, ND1 and 16S rDNA; 1,595 bp) in MEGA7 (Kumar *et al.* 2016). The number of base differences per site from averaging over all sequence pairs within each group are shown used the Kimura 2-parameter model (*K2P*, Kimura 1980) and uncorrected *p*-distances.

K2P	*M. latipennis*	*M. caerulatus*	*Megaloprepus* sp. nov.	*M. brevistigma*
M. latipennis	**0.30**			
M. caerulatus	6.42	**0.56**		
Megaloprepus sp. nov.	7.78	8.35	**0.73**	
M. brevistigma	11.93	11.28	12.40	**4.16**

p-distances				
M. latipennis	**0.30**			
M. caerulatus	5.95	**0.55**		
Megaloprepus sp. nov.	7.10	7.59	**0.72**	
M. brevistigma	10.47	9.98	10.84	**3.95**

Table A2.1.6: Estimates of evolutionary divergence between and within populations calculated for 8 populations of *Megaloprepus* using the mitochondrial sequence marker **Ef1α**. Analyses were conducted using the Kimura 2-parameter model (K2P, Kimura 1980) and uncorrected *p*-distances. Rate variation among sites was modelled with a gamma distribution in MEGA7 (Kumar *et al.* 2016).

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; CrCNP = Corcovado National Park, Costa Rica; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru

K2P	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT	**0.11**							
CrCNP	0.39	**0.09**						
BCI	0.25	0.52	**0.21**					
CrLS	0.14	0.36	0.22	**0.15**				
CoCho	0.21	0.48	0.15	0.15	/			
CoNor	0.21	0.48	0.34	0.23	0.30	/		
CoBo	1.90	1.89	2.03	1.89	1.99	1.99	/	
PePH	1.74	2.02	1.87	1.76	1.83	1.83	0.45	/

p-distance	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT	**0.11**							
CrCNP	0.39	**0.09**						
BCI	0.25	0.52	**0.21**					
CrLS	0.14	0.36	0.22	**0.15**				
CoCho	0.21	0.48	0.15	0.15	**0.00**			
CoNor	0.21	0.48	0.34	0.23	0.30	/		
CoBo	1.86	1.85	1.99	1.85	1.95	1.95	/	
PePH	1.71	1.97	1.84	1.73	1.80	1.80	0.45	/

Table A2.1.7: Population pairwise F_{ST}-values for the nuclear sequence marker **Ef1α** were modelled in Arlequin vers. 3.5.2.2 (Excoffier & Lischer 2010; Excoffier *et al.* 1992) with 1000 permutations indicating nearly no gene flow between populations. Significant values are indicated with an asterisk (*).

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; CrCNP = Corcovado National Park, Costa Rica; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru

	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT								
CrCNP	0.74*							
BCI	0.36*	0.71*						
CrLS	0.55*	0.76*	0.52*					
CoCho	0.56*	0.83*	0.00	0.59*				
CoNor	0.46	0.81	0.39	0.61	1.00			
CoBo	0.94	0.95	0.90	0.93	1.00	1.00		
PePH	0.93	0.95	0.89	0.92	1.00	1.00	1.00	

Table A2.1.8: Estimates of Evolutionary Divergence over sequence pairs for the four *Megaloprepus* species. Calculations are based on the nuclear sequence marker **Ef1α** for 48 nucleotide sequences in MEGA7 (Kumar *et al.* 2016). The number of base differences per site from averaging over all sequence pairs within each group are shown used the Kimura 2-parameter model (K2P, Kimura 1980) and uncorrected *p*-distances.

K2P	*M. latipennis*	*M. caerulatus*	*Megaloprepus* sp. nov.	*M. brevistigma*
M. latipennis	**0.11**			
M. caerulatus	0.20	**0.20**		
Megaloprepus sp. nov.	0.39	0.44	**0.09**	
M. brevistigma	1.82	1.89	1.95	**0.45**

p-distance	*M. latipennis*	*M. caerulatus*	*Megaloprepus* sp. nov.	*M. brevistigma*
M. latipennis	**0.11**			
M. caerulatus	0.20	**0.20**		
Megaloprepus sp. nov.	0.39	0.44	**0.09**	
M. brevistigma	1.78	1.85	1.91	**0.45**

Table A2.1.9: Estimates of evolutionary divergence between and within populations calculated for 8 populations of *Megaloprepus* using the nuclear sequence marker **ITS I+II**. Analyses were conducted using the Kimura 2-parameter model (K2P, Kimura 1980) and uncorrected *p*-distances. Rate variation among sites was modelled with a gamma distribution in MEGA7 (Kumar *et al.* 2016).

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; CrCNP = Corcovado National Park, Costa Rica; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru

K2P	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT	**0.00**							
CrCNP	0.01	**0.03**						
BCI	0.63	0.33	**0.11**					
CrLS	0.56	0.26	0.10	**0.09**				
CoCho	0.71	0.41	0.14	0.15	**0.16**			
CoNor	0.63	0.33	0.06	0.07	0.08	/		
CoBo	2.66	2.64	2.83	2.75	2.91	2.83	/	
PePH	2.83	2.84	3.00	2.92	3.08	3.00	0.47	/

p-distance	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT	**0.00**							
CrCNP	0.01	**0.03**						
BCI	0.63	0.33	**0.11**					
CrLS	0.56	0.26	0.10	**0.09**				
CoCho	0.71	0.41	0.14	0.15	**0.16**			
CoNor	0.63	0.33	0.06	0.07	0.08	/		
CoBo	2.58	2.57	2.75	2.67	2.83	2.75	/	
PePH	2.75	2.76	2.91	2.83	2.99	2.91	0.47	/

Table A2.1.10: Population pairwise F_{ST}-values for the nuclear sequence marker **ITS I+II** were modelled in Arlequin vers. 3.5.2.2 (Excoffier & Lischer 2010; Excoffier *et al.* 1992) with 1000 permutations indicating nearly no gene flow between populations. Significant values are indicated with an asterisk (*).

* Abbreviations are as follows: RBLT = Biosphere Reserve Los Tuxtlas, Mexico; CrCNP = Corcovado National Park, Costa Rica; CrLS = Biological Research Station La Selva, Costa Rica; BCI = Barro Colorado Island, Panama; CoCho = Choco, Colombia; CoNor = Norcasia, Colombia; CoBo = Boyacá, Santa Maria, Colombia; PePH = Pampa Hermosa Lodge, Peru

	RBLT	CrCNP	BCI	CrLS	CoCho	CoNor	CoBo	PePH
RBLT								
CrCNP	0.99*							
BCI	0.95*	0.91*						
CrLS	0.96*	0.92*	0.05					
CoCho	0.99*	0.95*	0.11	0.33				
CoNor	1.00	0.96	-0.90	-0.20	-1.00			
CoBo	1.00	1.00	0.99	0.99	0.98	1.00		
PePH	1.00	1.00	0.99	0.99	0.98	1.00	1.00	

Table A2.1.11: Estimates of Evolutionary Divergence over sequence pairs for the four *Megaloprepus* species. Calculations are based on the nuclear sequence marker **ITS I+II** for 48 nucleotide sequences in MEGA7 (Kumar *et al.* 2016). The number of base differences per site from averaging over all sequence pairs within each group are shown used the Kimura 2-parameter model (K2P, Kimura 1980) and uncorrected *p*-distances.

K2P	*M. latipennis*	*M. caerulatus*	*Megaloprepus* sp. nov.	*M. brevistigma*
M. latipennis	**0.00**			
M. caerulatus	0.61	**0.10**		
Megaloprepus sp. nov.	0.01	0.31	**0.03**	
M. brevistigma	2.74	2.88	2.74	**0.47**

p-distance	*M. latipennis*	*M. caerulatus*	*Megaloprepus* sp. nov.	*M. brevistigma*
M. latipennis	**0.00**			
M. caerulatus	0.60	**0.10**		
Megaloprepus sp. nov.	0.01	0.30	**0.03**	
M. brevistigma	2.67	2.80	2.67	**0.47**

File A2: Supplementary material for the phylogenetic analyses

Methods

Divergence Time Estimations

Although the approximation of divergence times has become one important method in phylogeography, they are sensitive to the selection of substitution models, genetic markers and sampling. To obtain a first impression on the data set we applied a strict molecular clock on the concatenated matrix for the Pseudostigmatidae (Table S2.1). In BEAST vers. 1.8.4 (Drummond *et al.* 2012) we linked within the matrix 12S with 16S, ND1 with CO1 and 28S rRNA with EF1α in the site and clock models. For each partition we fixed the clock rates for the mitochondrial DNA to 0.0115 (12S/16S) and 0.0177 (ND1/CO1) following the insect mitochondrial divergence rates of Brower (2.3% My^{-1}; (Brower 1994)) and Papadopoulou *et al.* (3.54% My^{-1}; (Papadopoulou *et al.* 2010)). Because no divergence rates are known for 28S rRNA/EF1α, we used an uncorrelated, lognormal relaxed clock (Drummond *et al.* 2006) with a broad lognormal prior for the mean (*R* (0.0115/1.0) for these two genes. This set the mean rate of the 95% prior density between 0.14% and 7.18%. The standard derivation had an exponentially distributed prior with a mean of 0.1.

We run 6 independent chains with each 10×10^8 (MCMC) generations and sampled every 10,000. Tracer vers. 1.6.0 was used to determine stationery and convergence by monitoring the effective sample size (ESS) and the influence of our priors on the data was checked via an empty run using the priors only (Drummond *et al.* 2006). Finally, the tree files were combined with LogCombiner vers. 1.8.4 using a conservative burn-in of 20% for each reach run and TreeAnnotator vers. 1.8.4 was used to summarize maximum clade credibility trees (MCC) with a posterior probability limit of 0.5. Trees were viewed in FigTree vers. 1.4.2 (Rambaut & Drummond 2015) and DensiTree vers. 2.2.6 (Bouckaert & Heled 2014).

To verify our relaxed molecular clock (see main document) we applied one fossil calibration in a second molecular relaxed clock following Callahan and McPeek (2016). We used one known fossil for the genus *Ischnura* from Dominican amber (Bechly 2000), which was dated to be between 16.4 – 20.5 My old (Mitchell 2007). Consequently, we downloaded 43 sequences for 16 species for four closes related odonate families (*Ischnura*, *Enallagma*, *Teinobasis* and *Telebasis*), whereat the genus *Neurobasis* served as outgroup. This however reduced our genetic sampling to only two mitochondrial (16S, 537 bp; CO1, 644 bp) and one nuclear (28S, 1406 bp) sequence marker (cf. Table S2.2, Fig. S2.5, S2.6).

For a first impression of family relationships phylogenetic trees using maximum likelihood (ML) and Bayesian methods (BI) were calculated (cf. Fig. 2.5). Subsequently, the time-calibrated phylogeny was calculated from the concatenated alignment in BEAST vers. 1.8.4 (Drummond *et al.* 2012). Hereby, for each gene independently the GTR+G substitution model with six rate categories and an uncorrelated, lognormal relaxed clock was applied. The priors for the clock were exponentially distributed for the mean and the

standard derivation of with means of 0.1. Most importantly, a stem-based calibration was set at the node ancestral to the most recent common ancestor for the *Ischnura* sequences with an exponentially distributed prior with a mean=2.0 and an offset=16.4 setting the 95% of the prior probability to 23.78 Mya. Equal to Callahan and McPeek (2016) we further tested a mean=10 (95% prior probability to 16.65 - 53.29 My). The birth-death model was applied as tree prior using an exponential prior on the birth rate (mean = 0.031) and uniform prior on the relative death rate (U (0, 1)). All remaining settings were left as default. Finally, we run 6 independent chains with each 10×10^8 (MCMC) generations, whereby the downstream analysis followed the description for the strict clock.

Results

In both, the strict and the fossil calibration we obtained a high ESS (effective sample sizes > 200) and all parameters reached stationary. The tree topologies are consistent to the ML or BI analysis (cf. Fig. S2.1 - S2.3).

The node ages in the strict molecular clock (Fig. S2.4) are significantly older than in the two relaxed clocks. This divergence time reflects the split between the American and the African species of the Pseudostigmatidae and therefore supports the Gondwana relict description (cf. Groeneveld *et al.* 2007). However, diversification times among the *Megaloprepus* species appear overestimated. In the fossil calibration the node among *Ischnura* and *Enallagma* correspond to observations from Callahan and McPeek (2016). However, the divergence times for the nodes appear older than in the 'Pseudostigmatidae only' relaxed molecular clock but are mainly within the 95% HPD intervals (Fig. S2.5). Using a greater mean in the steam calibration results in older nodes (Figure not shown).
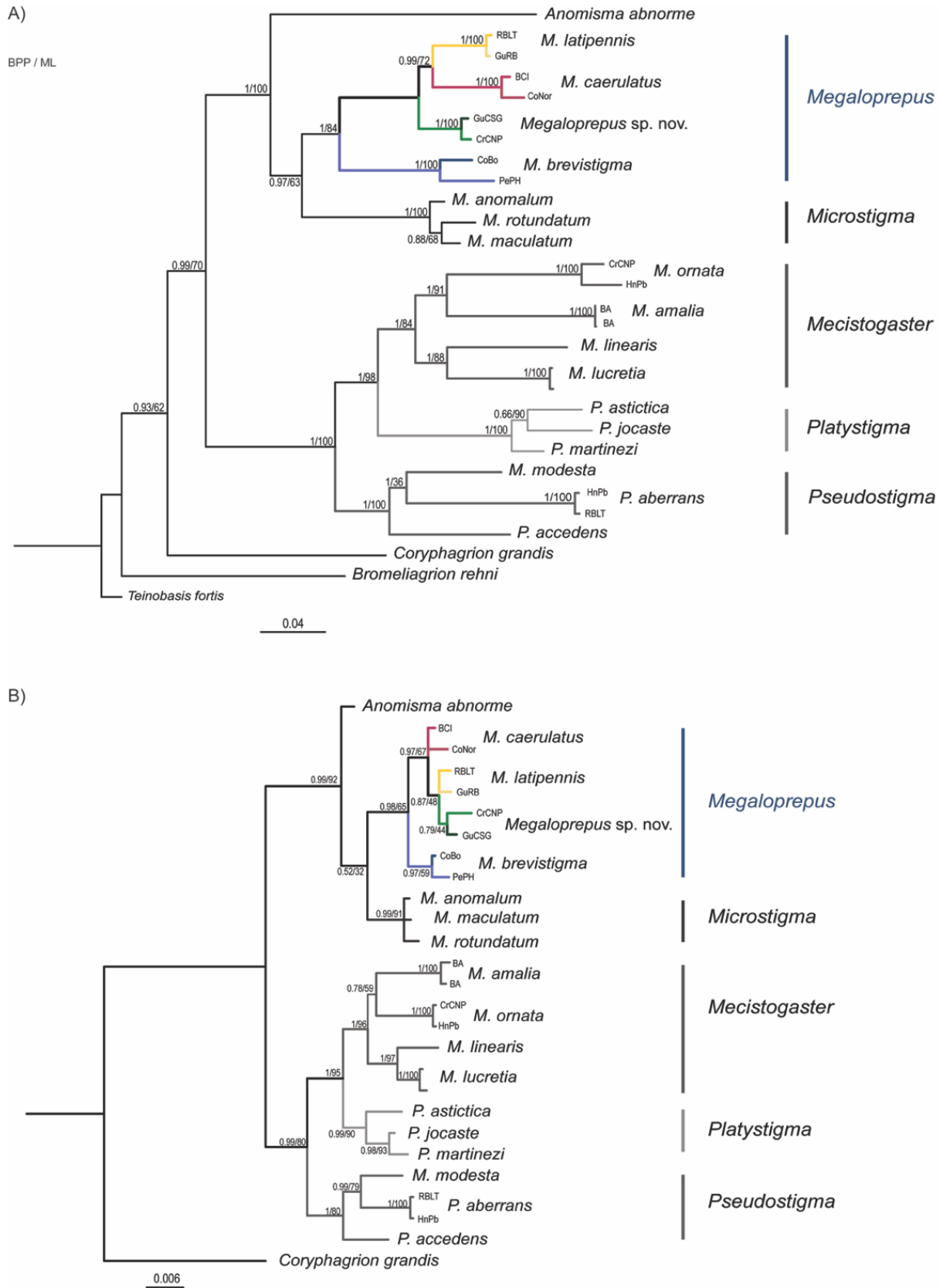
Figure A2.2.1: Phylogeny of the Pseudostigmatidae based on the mitochondrial genes CO1, ND1, 12S rRNA, 16S rRNA (A) and the nuclear genes Ef1α and ITS I+II (B) estimated via maximum likelihood (ML) in RAxML vers. 8.2.9 (Stamatakis 2014) and Bayesian interference using MrBayes vers. 3.2.6 (Ronquist *et al.* 2012). The node support values at the corresponding nodes were

estimated running 4 independent MCMC chains for 2 million generations, sampled every 1,000 generations and a burn-in of 25% (Bayesian posterior probabilities, BPP) and 1,000 ML bootstrap replications. Both gene sets were able to separate the six genera of the family, but topological uncertainties were detected inside *Platystigma* and *Microstigma*.

Figure A2.2.2: Phylogenetic relationships among the *Megaloprepus* species and their position within the Pseudostigmatidae estimated via maximum likelihood and Bayesian methods yielding the same topologies with their support values (Bayesian posterior probabilities, BPP and ML bootstrap values for 1,000 replications, ML) mentioned at the corresponding nodes. For tree reconstructing the concatenated alignment of four mitochondrial (CO1, ND1, 12S rRNA, 16S rRNA) and two nuclear genes (Ef1α and ITS) was used. The node support for the two big clades is high (ML bootstrap-values ≥ 70, PPBI ≥ 0.97).

All species cluster according to their taxonomic classification. The genera *Mecistogaster* and *Platystigma* represent the biggest radiation within the Pseudostigmatidae with 8-9 species each. Both are a sister clade to *Pseudostigma*. Inside *Megaloprepus*, the eastern Andean *M. brevistigma* stands as a sister clade to the Mesoamerican species, while *M. caerulatus* and *M. latipennis* appear as the closest related.

BPP / ML
16S rRNA, 28S rRNA, CO1

0.06

Figure A2.2.3: Phylogeny based on Maximum likelihood methods and Bayesian interference of the Pseudostigmatidae and close related damselflies used as starting information for the time calibration. As outgroup two *Neurobasis* species were used. Maximum likelihood and Bayesian interference were calculated as mentioned in the family tree. The tree topology within the helicopter damselflies and the other families is in concordance to previous estimations (cf. Callahan & McPeek 2016; Dijkstra *et al.* 2014).
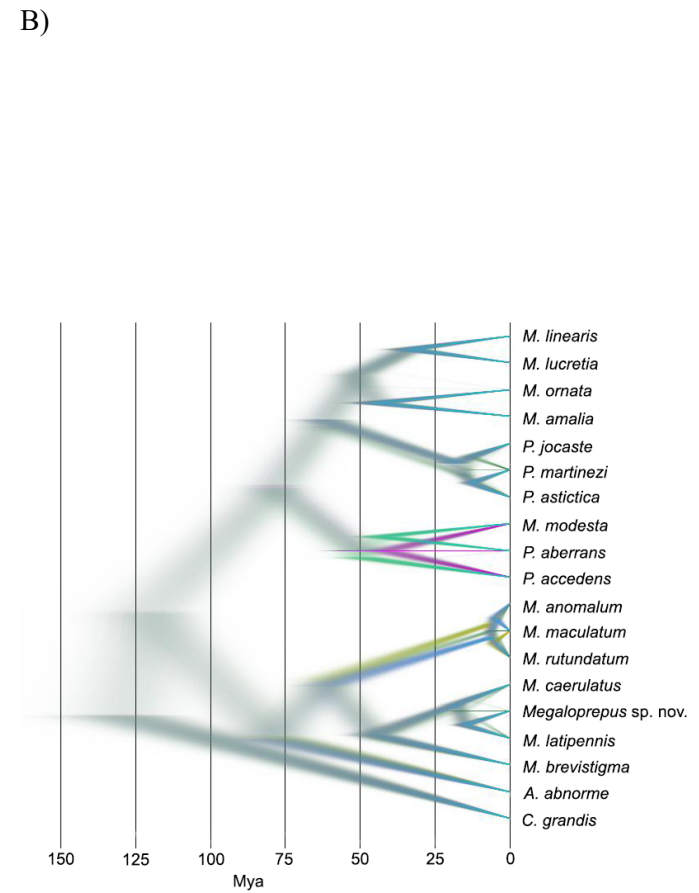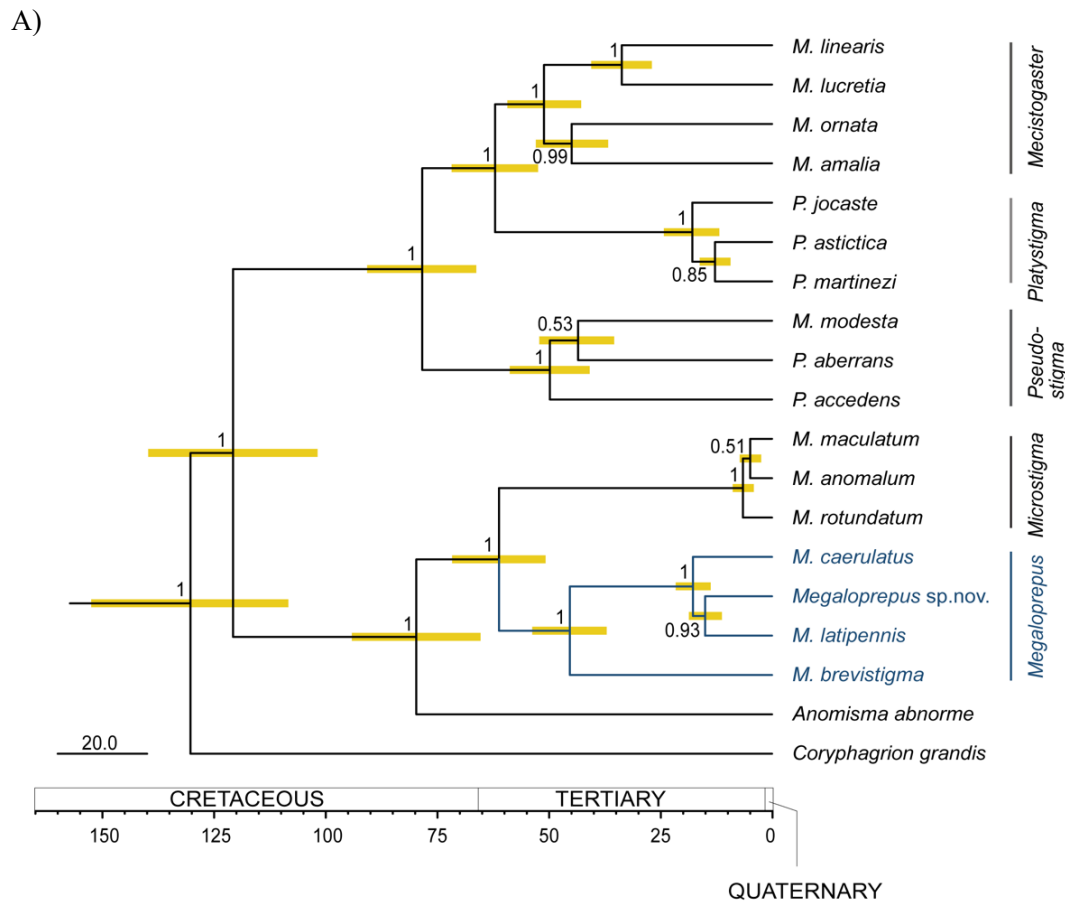
Figure A2.2.4: A) Time calibrated MCC (Bayesian maximum clade credibility) with posterior mean node ages estimated by a strict molecular clock for the Pseudostigmatidae using *C. grandis* as outgroup. The BEAST (Drummond *et al.* 2012) estimation is using a concatenated matrix of four mitochondrial sequence marker under the strict approach (CO1, ND1, 16S and 12S) and two nuclear markers under a relaxed approach (28S and Ef1α). The Bayesian posterior probabilities are depicted at each node and the node bars display the 95% HPD interval for that estimated node age. B) The strict time calibration displayed via DensiTree vers. 2.2.6 (Bouckaert & Heled 2014) showing each individual tree by a thin line highlighting tree density, the posterior distribution of trees and topological uncertainties.
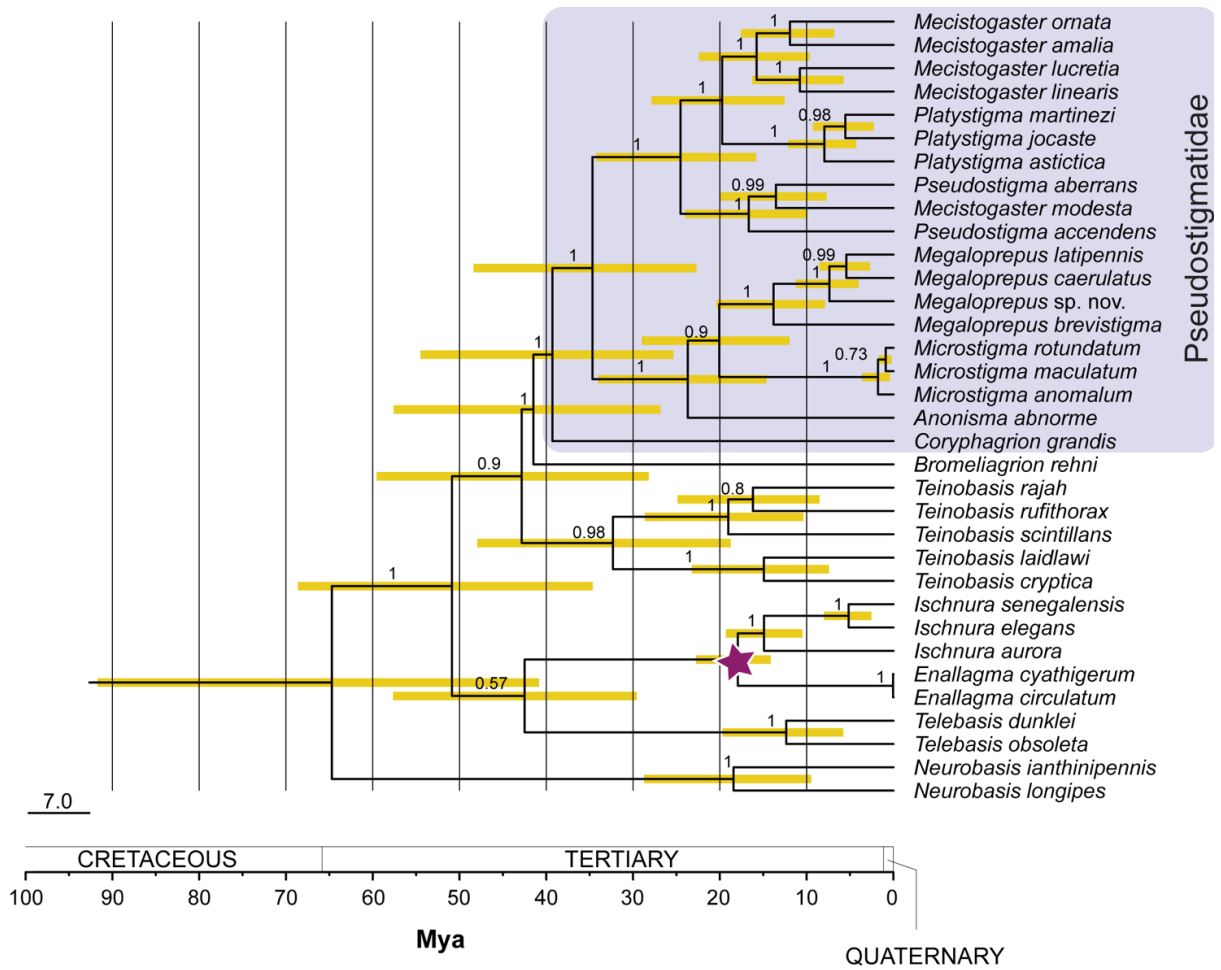
Figure A2.2.5: Time calibrated phylogeny based on a lognormal, relaxed molecular clock and three loci (16S, CO1 and 28S). An *Ischnura* fossil from 16.4-20.5 Mya was used to calibrate the most recent common ancestor of the *Ischnura* samples (purple star) with a mean of 2.0 for the offset exponential distribution on the age of the stem node. The MCC (Bayesian maximum clade credibility) tree shows the Bayesian posterior probabilities at each node and the node bars display the 95% HPD interval for that estimated node age.

File A3. Supplementary material for the morphometric analyses

Methods

Both, linear morphometrics and geometric morphometrics (GMM), are objective tools used to investigate a variety of evolutionary questions such as species boundaries and species determination (Adams *et al.* 2004; Mitteroecker & Gunz 2009), and diversification, hybridization, and phylogenetic relationships (Duda *et al.* 2009; Outomuro *et al.* 2013; Wappler *et al.* 2012). Beside of many studies in mammals and reptiles where the method is widely accepted (Tabatabaei Yazdi & Adriaens 2013), recently a larger amount of research focused on insect wings (e.g. Baracchi *et al.* 2011; Bots *et al.* 2012; Outomuro *et al.* 2013). In odonates, wings and their venation are not only discriminant morphological characters but play a role in flight performance and sexual and natural selection (Bots *et al.* 2012; Outomuro *et al.* 2013; Outomuro *et al.* 2016). Here we used the shape analysis of wings to describe phenotypic variation and looking for differentiation among species. These analyses are based on standardized photographs on museum samples and field collected samples.

Geometric Morphometrics (GMM) - Preparation of data files for the shape analysis
Initially potential landmark digitization and orientation errors were evaluated via replicate image taking and repeat digitizing (e.g. Bots *et al.* 2012; Tabatabaei Yazdi & Adriaens 2013). For this purpose, we followed the protocol suggested by Adriaens (2007). Our errors were significantly smaller than the natural shape variation explained by group differences and the means for digitization and orientation error were 4.32% and 6.2%, respectively. Additionally, significant differences obtained in a Procrustes ANOVA ($p < 0.05$) further indicate negligible errors.

    The following statistics included were performed to generate 'size free' shape variables for the statistical comparisons of groups. Therefore, first a Generalized Procrustes Analysis (GPA) was performed to remove location, size, and orientation differences between samples (e.g. Klingenberg & McIntyre 1998). Procrustes variables were validated through an outlier search in MorphoJ vers.1.06b (Klingenberg 2011) and a cluster analysis in PAST vers. 3.05 (Hammer *et al.* 2001), both showing no extremes. To identify allometry multiple regressions of procrustes coordinates on centroid size were performed (Drake & Klingenberg 2008; Mitteroecker *et al.* 2013). A linear correlation was obtained in the complete data set ($p = 0.019$) as well as in the *M. caerulatus* samples alone ($p = 0.002$). Consequently, the size component was removed by performing a multivariate linear regression which is maintaining the residuals (Outomuro *et al.* 2013; Viscosi & Cardini 2011). Based on those 'size free' shape variables all following statistical analyses were conducted.

Table A2.3.1: Listed individuals from different Museum collections used for the morphometric analyses and the geometric morphometrics. Number of individuals for each species is variable *M. caerulatus* (*N* = 32), *M. latipennis* (*N* = 17), *M. brevistigma* (*N* = 10) and *Megaloprepus* sp. nov. (*N* = 8). For the GMM number of individuals were reduced due to wing damages to 62: *M. brevistigma* (*N* = 11), *M. caerulatus* (*N* = 29), *M. latipennis* (*N* = 16) and *Megaloprepus* sp. nov. (*N* = 6).

* Museum abbreviations are as follows:

AMNH = American Museum of Natural History (New York, New York, USA); INBio = National Biodiversity Institute (San José, Costa Rica); UCMS = University of Connecticut Biological Collections, Department of Ecology and Evolutionary Biology, University of Connecticut (Storrs, Connecticut, USA); UNAM = National Insect Collection of the National Autonomous University of Mexico (Mexico City, DF, Mexico); YPM = Yale Peabody Museum of Natural History, Yale University (New Haven, Connecticut, USA); ITZ, TIHO = Division of Ecology and Evolution, University of Veterinary Medicine Hannover

(cf Feindt and Hadrys in prep.)

species	ID	Museum	accession number	collected by	collection date	country	Region / Conservation Area
M. caerulatus	AMNH_Mcc_001	AMNH	N.A.	CH Curran	Nov-29	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_002	AMNH	N.A.	Donato	Nov-30	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_003	AMNH	F30III2D	Donato	Nov-30	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_004	AMNH	N.A.	N.A.	Nov-29	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_005	AMNH	N.A.	CH Curran	Jan-29	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_006	AMNH	N.A.	WJ Gertsch	Mar-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_007	AMNH	N.A.	WJ Gertsch	Feb-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_009	AMNH	N.A.	FE Lutz	Mar-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_010	AMNH	N.A.	WJ Gertsch	Mar-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_011	AMNH	N.A.	FE Lutz	Mar-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_012	AMNH	N.A.	FE Lutz	Feb-36	Panama	Barro Colorado Nature Monument
	AMNH_Mcc_017	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_018	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_020	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_021	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_022	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_023	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_024	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcc_025	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	InBio_23	INBio	INBIOCRI001112977	MA Zumbado	Mar-91	Costa Rica	Braulio Carillo National Park

	InBio_31	INBio	INBIOCRI000218397	A Fernandez	Jan-91	Costa Rica	Braulio Carillo National Park
	InBio_32	INBio	INBIOCRI000218396	A Fernandez	Jan-91	Costa Rica	Braulio Carillo National Park
	InBio_43	INBio	INB0004284813	L Chavarría	Jan-00	Costa Rica	Reserva Biologica Hitoy Cerere
	InBio_49	INBio	INBIOCRI000686309	R Aguilar	Mar-89	Costa Rica	National Park Tortuguero
	UCMS_2	UCMS	N.A.	D Wagner, M Thomas, B Haber	Feb-08	Costa Rica	Biological Reserve Tirimbina
	UCMS_3	UCMS	N.A.	S Gaimari	Feb-03	Costa Rica	Braulio Carillo National Park
	AMNH_Mcc_031	AMNH	N.A.	N.A.	Nov-34	Costa Rica	Cordillera Volcánica Central
	Mcaer1_CrLS	ITZ, TiHo	N.A.	W Feindt	Dec-09	Costa Rica	Biological Research Station La Selva
	Mcaer3_CrLS	ITZ, TiHo	N.A.	W Feindt	Dec-11	Costa Rica	Biological Research Station La Selva
	Mcaer36_CrLS	ITZ, TiHo	N.A.	W Feindt	Apr-12	Costa Rica	Biological Research Station La Selva
	Mcaer14_NiBa	ITZ, TiHo	N.A.	W Feindt	Dec-11	Nicaragua	Biological Reserve Indio Maíz
	Mcaer15_NiBa	ITZ, TiHo	N.A.	W Feindt	Dec-11	Nicaragua	Biological Reserve Indio Maíz
M. brevistigma	AMNH_Mcb 011	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 013	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 015	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 016	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 017	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 018	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 019	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 020	AMNH	Ac33501	F Ovalle	Nov-34	Colombia	N.A.
	AMNH_Mcb 021	AMNH	N.A.	FM Brown	Jan-39	Ecuador	Sangay National Park
	AMNH_Mcb 022	AMNH	N.A.	FM Brown	Jul-38	Ecuador	Sangay National Park
M. latipennis	AMNH_Mcl 002	AMNH	Ac28144	N.A.	1927	Mexico	Tezoyuca
	AMNH_Mcl 003	AMNH	N.A.	WL Tower	Nov-05	Mexico	Tezoyuca
	AMNH_Mcl 004	AMNH	N.A.	WL Tower	Jun-05	Mexico	Tezoyuca
	YPM_1	Peabody Museum	N.A.	RG Wind	Nov-69	Mexico	Dos Amantes, Veracruz
	Mcaer30_MeLT	ITZ, TiHo	N.A.	W Feindt	Mar-12	Mexico	Los Tuxtlas Biosphere Reserve
	UNAM_001	UNAM	IBUNAM:CNIN:OD1778	RW Cruden	Jul-66	Mexico	Los Tuxtlas Biosphere Reserve
	UNAM_003	UNAM	IBUNAM:CNIN:OD1780	J Bueno	Sep-79	Mexico	Los Tuxtlas Biosphere Reserve
	UNAM_004	UNAM	IBUNAM:CNIN:OD1781	A Figuproa	Jul-85	Mexico	Los Tuxtlas Biosphere Reserve

UNAM_007	UNAM	IBUNAM:CNIN:OD1784	N.A.	Aug-67	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_008	UNAM	IBUNAM:CNIN:OD1785	H Gonzáles, A Imada	Jul-68	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_009	UNAM	IBUNAM:CNIN:OD1786	E Gonzáles	Aug-78	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_010	UNAM	IBUNAM:CNIN:OD1787	E Gonzáles	Aug-79	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_012	UNAM	IBUNAM:CNIN:OD1789	E Gonzáles	May-81	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_014	UNAM	IBUNAM:CNIN:OD1791	P Guzmán	Aug-88	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_018	UNAM	IBUNAM:CNIN:OD1795	E Gonzáles	Jul-75	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_019	UNAM	IBUNAM:CNIN:OD1796	E Gonzáles	Apr-79	Mexico	Los Tuxtlas Biosphere Reserve
UNAM_021	UNAM	IBUNAM:CNIN:OD1798	E Gonzáles	Jul-75	Mexico	Los Tuxtlas Biosphere Reserve
Megaloprepus **sp. nov.** InBio_12	INBio	INBIOCRI001701903	DH Janzen	Jul-77	Costa Rica	Corcovado National Park, ACOSA
InBio_13	INBio	INBIOCRI001701902	DH Janzen	Jul-77	Costa Rica	Corcovado National Park, ACOSA
InBio_22	INBio	N.A.	LD Gomez	Apr-95	Costa Rica	Corcovado National Park, ACOSA
InBio_37	INBio	INB0004284807	A Gutiérrez	Oct-93	Costa Rica	Peninsula de Osa, ACOSA
InBio_44	INBio	N.A.	LS Angulo	Feb-96	Costa Rica	Peninsula de Osa, ACOSA
Mcaer8_CrCNP	ITZ, TiHo	N.A.	W Feindt	Dec-11	Costa Rica	Corcovado National Park, ACOSA
Mcaer6_CrCNP	ITZ, TiHo	N.A.	W Feindt	Dec-11	Costa Rica	Corcovado National Park, ACOSA
Mcaer9_CrCNP	ITZ, TiHo	N.A.	W Feindt	Dec-11	Costa Rica	Corcovado National Park, ACOSA

Table A2.3.2: Summary of morphological variables measured in the four putative *Megaloprepus* species: *M. caerulatus*, *M. latipennis*, *M. brevistigma* and *Megaloprepus* sp. nov.. The number of included measurements varies between groups, whereat in *M. caerulatus* the most individuals and highest number of populations (N = 4) were included. The description, morphological terminology concerning wing venation follows Riek & Kukalova-Peck (cf. Rehn 2003).

Variables*	*M. caerulatus* (N = 31)	*M. brevistigma* (N = 10)	*M. latipennis* (N = 17)	*Megaloprepus* sp. nov. (N = 8)
AL	88.79 ± 8.64	93.09 ± 10.26	87.90 ± 5.65	88.54 ± 6.12
FWL	73.66 ±7.76	78.82 ± 9.76	75.78 ± 5.17	78.43 ± 8.52
FWW	18.54 ± 2.24	16.58 ± 2.01	18.91 ± 1.41	17.75 ± 1.72
FWL/FWW	3.98 ± 0.15	4.76 ± 0.15	4.01 ± 0.15	4.42 ± 0.20
FW-PtL	4.00 ± 0.69	2.68 ± 0.39	3.95 ± 0.62	3.87 ± 0.80
FW-PtW	1.17 ± 0.24	1.21 ± 0.17	1.29 ± 0.21	1.33 ± 0.18
FW-PtL/FW-PtW	3.49 ± 0.52	2.23 ± 0.32	3.10 ± 0.49	2.98 ± 0.85
FW-BSL	18.81 ± 2.59	9.77 ± 0.84	12.50 ± 1.56	13.54 ± 2.05
FWL/FW-BSL	3.94 ± 0.27	8.06 ± 0.52	6.12 ± 0.67	5.84 ± 0.43
FW-RP1a	33.68 ± 3.45	39.63 ± 5.57	32.11 ± 2.17	36.55 ± 4.06
FW-RP1b	31.77 ± 4.90	29.04 ± 3.70	34.59 ± 3.23	33.05 ± 5.31
FW-RP2	29.46 ± 4.54	27.38 ± 3.45	32.67 ± 3.11	30.90 ± 4.83
FWL/FW-RP1a	2.19 ± 0.12	1.99 ± 0.08	2.36 ± 0.11	2.15 ± 0.13
FWL/FW-RP2	2.52 ± 0.19	2.88 ± 0.11	2.33 ± 0.11	2.56 ± 0.20
HWL	71.67 ± 7.70	76.55 ± 9.57	72.88 ± 5.29	16.73 ± 7.05
HWW	17.97 ± 2.18	16.06 ± 2.01	18.29 ± 1.65	4.50 ± 1.55
HWL/HWW	4.00 ± 0.14	4.77 ± 0.16	4.00 ± 0.18	3.41 ± 0.06
HW-PtL	3.31 ± 0.68	2.34 ± 0.38	3.29 ± 0.52	1.29 ± 0.64
HW-PtW	1.09 ± 0.23	1.12 ± 0.13	1.26 ± 0.18	2.66 ± 0.15
HW-PtL/HW-PtW	3.03 ± 0.48	2.04 ± 0.20	2.62 ± 0.29	12.71 ± 0.47
HW-BSL	18.19 ± 2.59	9.59 ± 0.83	12.99 ± 1.39	5.95 ± 1.61
HWL/HW-BSL	3.97 ± 0.31	7.98 ± 0.72	5.64 ± 0.43	35.27 ± 0.37
HW-RP1a	32.51 ± 3.41	38.58 ± 5.45	30.62 ± 1.76	30.63 ± 3.71
HW-RP1b	29.93 ± 4.55	27.55 ± 3.69	33.25 ± 3.77	29.25 ± 4.70
HW-RP2	28.60 ± 4.96	26.44 ± 3.54	31.94 ± 3.63	2.14 ± 4.35
HWL/HW-RP1a	2.22 ± 0.32	1.99 ± 0.07	2.38 ± 0.14	2.59 ± 0.13
HWL/HW-RP2	2.54 ± 0.21	2.90 ± 0.14	2.29 ± 0.13	2.59 ± 0.04

Aberrations of variables: Abdomen length (AL), Forewing length (FWL), Forewing width (FWW), Ratio Forewing length/Forewing width (FWL/FWW), FW Pt length (FW-PtL), FW Pt width (FW-PtW), Ratio FW Pt length /FW Pt width (FW-PtL / FW-PtW), FW blue stripe length** (FW-BSL), Ratio FW length/FW blue stripe length (FWL/FW-BSL), FW length RP1*** (FW-RP1a), FW length of RP1 (FW-RP1b), FW length of RP2 (FW-RP2), Ratio FW length/ FW length RP1 (FWL/FW-RP1a), Ratio FWlength/ FW length of RP2 (FWL/FW-RP2), Hind wing length (HWL), Hind wing width (HWW), Ratio Hind wing length/Hind wing width (HWL/HWW), HW Pt length (HW-PtL), HW Pt width (HW-PtW), Ratio HW Pt length/HW Pt width (HW-PtL/HW-PtW), HW blue stripe length (HW-BSL), Ratio HW length/HW blue stripe length (HWL/HW-BSL), HW length RP1 (HW-RP1a), HW length of RP1 (HW-RP1b), HW length of RP2 (HW-RP2), Ratio HW length/ HW length RP1 (HWL/HW-RP1a), Ratio HW length/ HW length of RP2 (HWL/HW-RP2), Ratio FW length/HW length (FWL/HWL)

* Pt length was measured at the wing costa.
** Blue stripe length was measured at the wing costa.
*** FW-RP1a: length from the subnodus to the origin of RP2.
**** Rehn AC (2003) Phylogenetic analysis of higher-level relationships of Odonata. Systematic Entomology 28, 181-240.

Table A2.3.3: Localization of landmarks on the *Megaloprepus* wings. Morphological terminology concerning wing venation follows Riek and Kukalova-Peck*.

M. latipennis

1 cm
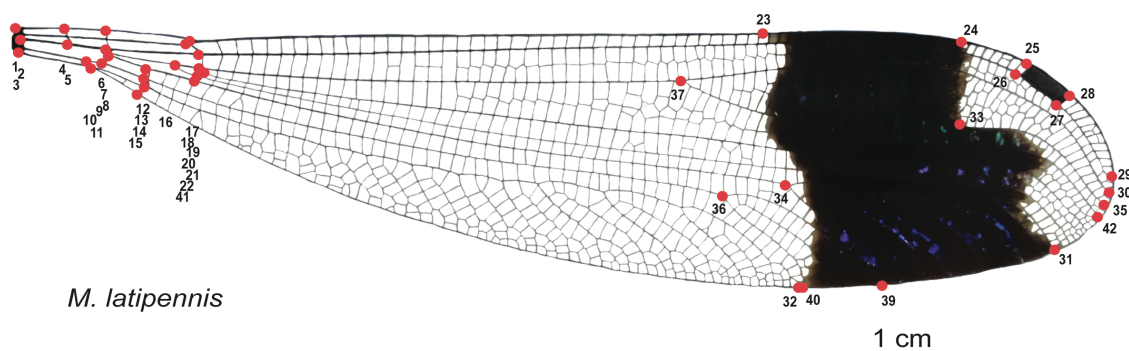
1	Proximal end of the Costa / C	22	Intersection nodal crossvein and media posterior
2	Proximal end of the Subcosta / ScP-	23	Proximal end of the blue stripe at the costal margin
3	Proximal end of the anal anterior and posterior / AA & AP+	24	Distal end of the blue stripe at the costal margin
4	1st antenodal at the costa / intersection CA & CP & ScA´ - ax1	25	Aantero-lateral and proximal end of the pterostigma
5	1st Antenodal at the radius anterior / intersection RA & RP+ - ax1	26	Postero-lateral and proximal end of the pterostigma
6	2nd antenodal at the costa / intersection CA & CP & ScA´ - ax2	27	Postero-lateral and distal end of the pterostigma
7	2nd antenodal at the radius anterior / intersection RA & RP + - ax2	28	Antero-lateral and distal end of the pterostigma
8	Dorsal directed end of the anterior margin of the quadrangle / intersection Arculus - MA+	29	Distal end of the radius anterior (RA)
9	Ventral directed end of the anterior margin of the quadrangle 7 intersection Arculus MP- (& CuA)	30	Distal end of the radius posterior (RP1)
10	Dorsal vestige of CuP-	31	Distal end of the blue stripe at the anal posterior
11	Ventral vestige of CuP-	32	Proximal end of the blue stripe at the anal vein
12	Dorsal directed end of the posterior margin of the quadrangle / intersection MA+ - ddv	33	Maximal notch of the distal end of the blue stripe at IR1
13	Ventral directed end of the posterior margin of the quadrangle / intersection MP-(& CuA) - ddv	34	Distal intersection of RP3
14	Proximal end of the cubitus anterior - CuA+	35	Distal end of RP2
15	Anal anterior - AA2+	36	Distal intersection of media anterior (MA)
16	Origin of the third branch (RP1) at the first branch (RP3) - RP1 - RP3	37	Distal intersection of RP1
17	Nodus (N)	38	Distal intersection of MP
18	Distal nodus (N)	39	Distal end of MP
19	Subnodus (SN)	40	Distal end of MP2
20	Intersection subnodus - RP1	41	Intersection nodal crossvein and media anterior
21	Origin of intercalar vein 2 (IR2)	42	Distal end of RP2

A)

B)

▲ *M. latipennis*   ● *M. caerulatus*   ■ *M. brevistigma*   ✳ *Megaloprepus* sp. nov.
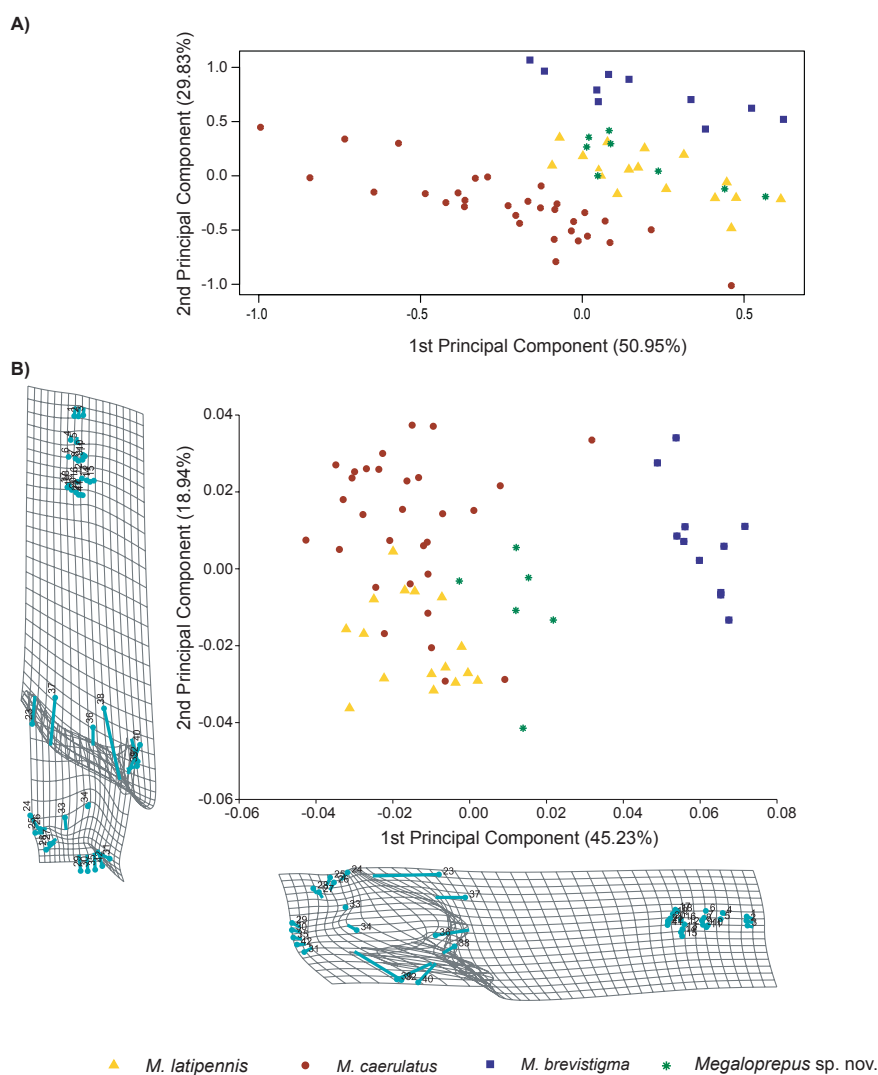
Figure A2.3.1: Graphic presentation of the principal component scores for the measurement comparisons (traditional morphometrics) and the GMM analyses for the four putative *Megaloprepus* species. (A) The PCA scatter plot displays the two main axes for the PCA whereat the first two PC components encountered for 80.78% of total variability in the hind wings (FW: PC1 = 51.17%, PC2 = 27.41%, PC3 = 11.97%).
(B) In the GGM the first two components explain 64.16% of variability (PC1 = 46.96%, PC2 = 16.61%, PC3 = 11.26). The PC shape changes are further displayed by transformation grids whereat the higher alternations are closer to the wing tips.

Table A2.3.4: CVA classification for hind wing measurement comparisons between the four putative *Megaloprepus* species indicates correctness the predefined groups. Predicted accuracy is 100%, and after cross validation the predicted accuracy is 98.48%.

	M. brevistigma	*M. caerulatus*	*M. latipennis*	*Megaloprepus* sp. nov.	*N*
M. brevistigma	10	0	0	0	10
M. caerulatus	0	31	0	0	31
M. latipennis	0	0	17	0	17
Megaloprepus sp. nov.	0	0	0	8	8

From cross-validation:

	M. brevistigma	*M. caerulatus*	*M. latipennis*	*Megaloprepus* sp. nov.	*N*
M. brevistigma	10	0	0	0	10
M. caerulatus	0	30	1	0	31
M. latipennis	0	0	17	0	17
Megaloprepus sp. nov.	0	0	0	8	8

Table A2.3.5: Results from the discriminant function analysis comparing pairwise group assignments in the GMM analysis. After cross validation the comparisons between *M. caerulatus*, *M. latipennis* and *Megaloprepus* sp. nov. are not 100 percent accurate, which could be justified at first by the close relationship of samples and second the number of individuals used and last age of the individuals.

	From discriminant function:			From cross-validation:		
True group	**Allocated to**		**True group**	**Allocated to**		**Total**
	M. brevistigma	*M. caerulatus*		*M. brevistigma*	*M. caerulatus*	
M. brevistigma	11	0	*M. brevistigma*	11	0	11
M. caerulatus	0	29	*M. caerulatus*	0	29	29
	M. brevistigma	*M. latipennis*		*M. brevistigma*	*M. latipennis*	
M. brevistigma	11	0	*M. brevistigma*	11	0	11
M. latipennis	0	16	*M. latipennis*	0	16	16
	M. brevistigma	*Megaloprepus* sp. nov.		*M. brevistigma*	*Megaloprepus* sp. nov.	
M. brevistigma	11	0	*M. brevistigma*	11	0	11
Megaloprepus sp. nov.	0	6	*Megaloprepus* sp. nov.	0	6	6
	M. caerulatus	*M. latipennis*		*M. caerulatus*	*M. latipennis*	
M. caerulatus	29	0	*M. caerulatus*	26	3	29
M. latipennis	0	16	*M. latipennis*	0	16	16
	M. caerulatus	*Megaloprepus* sp. nov.		*M. caerulatus*	*Megaloprepus* sp. nov.	
M. caerulatus	29	0	*M. caerulatus*	20	9	29
Megaloprepus sp. nov.	0	7	*Megaloprepus* sp. nov.	2	4	6
	M. latipennis	*Megaloprepus* sp. nov.		*M. latipennis*	*Megaloprepus* sp. nov.	
M. latipennis	16	0	*M. latipennis*	15	1	16
Megaloprepus sp. nov.	0	6	*Megaloprepus* sp. nov.	3	3	6

File A2.4: Supplementary material for the species distribution modeling (SDM)

Species distribution modeling (SDM) is a widely used method in modern phylogeography to estimate potential distributional ranges of species. It is based on the fundamental concept of stable ecological niches. However, the here presented maps were generated in ESRI ArcGIS 10.3.1. (http://www.esri.com/) and include mountain ranges, which were used from the GTOPO30, a global digital elevation model (DEM). The source is the U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, South Dakota. https://creativecommons.org/licenses/by-nc/3.0/.
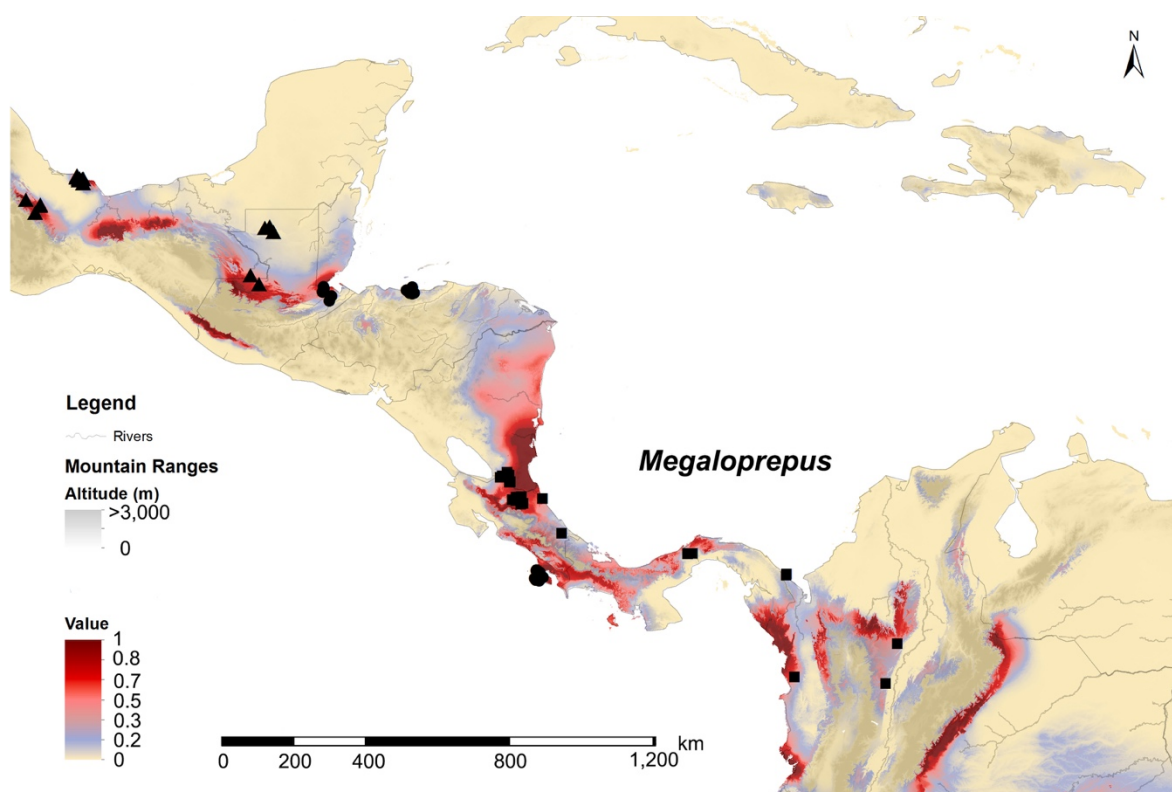
Figure A2.4.1: Modelled ecological suitability for the genus *Megaloprepus* using current climatic conditions generated in MaxEnt vers. 3.3.3k based on 190 GPS coordinates. The map presents a 'zoomed in' version of Mesoamerica, whereat the triangles stand for *M. latipennis*, dots for *Megaloprepus* sp. nov. and squares for *M. caerulatus*.
These coordinates were used together for the prediction of the current *Megaloprepus*' model and in the hind casting (cf. Figure 6 main text). In the species-specific models and for the niche comparison the coordinates were split according to the designated species.

Table A2.4.1: Estimations of relative contributions of the environmental variables to the MaxEnt model. Each variable contribution is shown in percent and permutation importance: BIO 1 = Annual mean temperature, BIO 9 = Mean temperature of driest quarter, BIO 12 = Annual precipitation, BIO 17 = Precipitation of driest quarter. Furthermore, the average AUC -value (area under the curve) for the replicate runs is shown for each model. The AUC is a measure of model performance, while AUC values lower than 0.5 indicate a random prediction, values higher than 0.75 picture well-fitted models (Elith *et al.* 2011). Biolimatic variables originate from http://worldclim.org/ (Hijmans *et al.* 2005) and the MODIS forest data from Global Land Cover Facility (Channan *et al.* 2014; Friedl *et al.* 2010).

current distribution				
Model	**Variable**	**contribution in %**	**Permutation importance**	**AUC**
Megaloprepus	BIO 12	78.9	44.9	
	BIO 17	11.4	5	
	BIO 1	6.1	16.6	
	BIO 9	3.6	33.5	0.96 +/- 0.03
M. caerulatus	BIO 12	79.3	21.8	
	BIO 9	9.4	65.4	
	BIO 1	9	12	
	BIO 17	2.3	0.7	0.98 +/- 0.02
M. latipennis	BIO 12	51.3	46.7	
	BIO 17	35.3	47.2	
	BIO 1	11	1.3	
	BIO 9	2.4	4.8	0.95 +/- 0.11
Megaloprepus sp. nov.	BIO 12	79.5	66.5	
	BIO 17	19.1	21.2	
	BIO 9	1.1	12	
	BIO 1	0.3	0.2	0.97 +/- 0.06
current distribution including forest coverage				
Model	**Variable**	**contribution in %**	**Permutation importance**	**AUC**
Megaloprepus	BIO 12	77.5	45.4	
	BIO 17	10.1	3.9	
	BIO 1	5.3	16.3	
	Modis - Forest coverage	4.1	0.3	
	BIO 9	2.9	34	0.97 +/- 0.03
M. caerulatus	BIO 12	78.1	26.4	
	BIO 1	9.5	12.4	
	BIO 9	9.1	59.9	
	BIO 17	1.6	1.3	
	Modis - Forest coverage	1.6	0.1	0.98 +/- 0.02
M. latipennis	BIO 12	52.3	58.5	
	BIO 17	32.7	32.2	
	BIO 1	11.8	1.8	
	BIO 9	1.6	6.4	
	Modis - Forest coverage	1.6	1.1	0.95 +/- 0.12
Megaloprepus sp. nov.	BIO 12	73.7	90.8	
	BIO 17	13.9	4.3	
	Modis - Forest coverage	11.4	1.9	
	BIO 9	0.9	2.9	
	BIO 1	0.1	0.1	0.96 +/- 0.06

Table A2.4.2: Estimations of relative contributions of the environmental variables to the MaxEnt model for the hind casting based on all GPS-records of all *Megaloprepus* samples. Contribution in percent and permutation importance is shown for BIO 1 = Annual mean temperature, BIO 9 = Mean temperature of driest quarter, BIO 12 = Annual precipitation and BIO 17 = Precipitation of driest quarter; the same variables as in the current model. In addition, the average test AUC for the replicate runs allows estimations about model performance, with AUC values higher than 0.75 indicating well-fitted models (Elith *et al.* 2011).

Biolimatic variables originate from http://worldclim.org/ (Hijmans *et al.* 2005), whereat variables for the Mid Holocene and LGM models are provided by MPI-ESM-P (Max Planck Institute for Meteorology) and for the LIG by Otto-Bliesner and colleagues (Otto-Bliesner *et al.* 2006).

Mid Holocene			
Variable	**Contribution in %**	**Permutation importance**	**AUC**
BIO 12	54.4	57.3	
BIO 17	23.7	5.9	
BIO 9	19.1	25.4	
BIO 1	2.9	11.3	0.97 +/- 0.02
Last glacial maximum (LGM)			
Variable	**Contribution in %**	**Permutation importance**	**AUC**
BIO 17	52.4	43.3	.
BIO 12	25.1	18.8	.
BIO 9	16.7	35.1	.
BIO 1	5.8	2.7	0.91 +/- 0.05
Last inter-glacial (LIG)			
Variable	**Contribution in %**	**Permutation importance**	**AUC**
BIO 12	90.8	92.3	
BIO 17	7.8	5.8	
BIO 1	0.8	0.9	
BIO 9	0.6	1.1	0.96 +/- 0.03
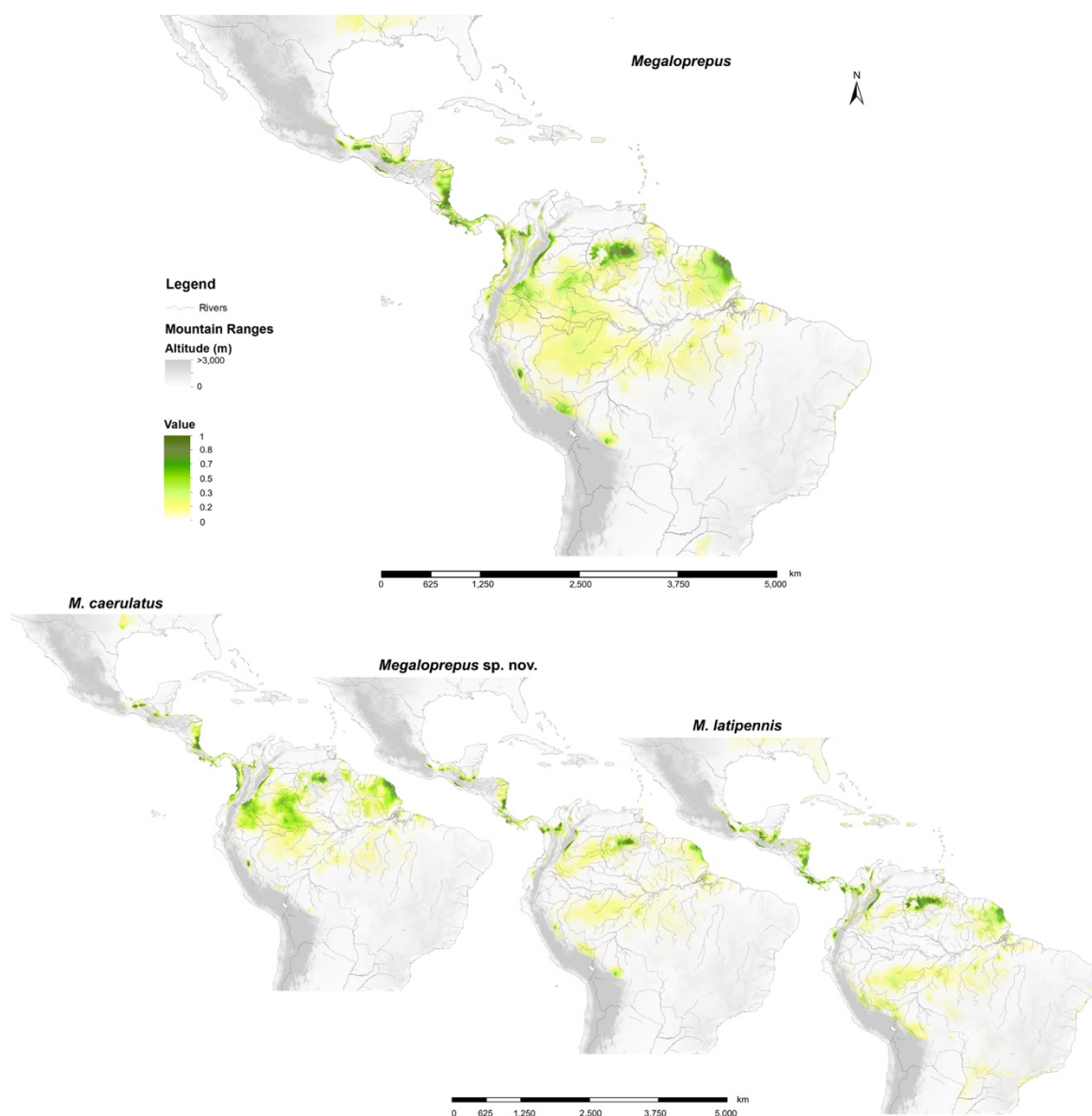
Figure A2.4.2: Modeled current distribution and ecological suitability of the three designated *Megaloprepus* species and the genus in MaxEnt vers. 3.3.3k based on four bioclimatic variables (BIO 1, BIO 9, BIO 12, and BIO 17) and the MODIS Land Cover data (Channan *et al.* 2014; Friedl *et al.* 2010). Using additionally the land cover data did not significantly change the model outcome or the potential environmental space of *Megaloprepus*. The potential distribution in each species model appear more restricted as using the climatic data alone. The probability of occurrence is shown from 20% in yellow to 100% in dark green.
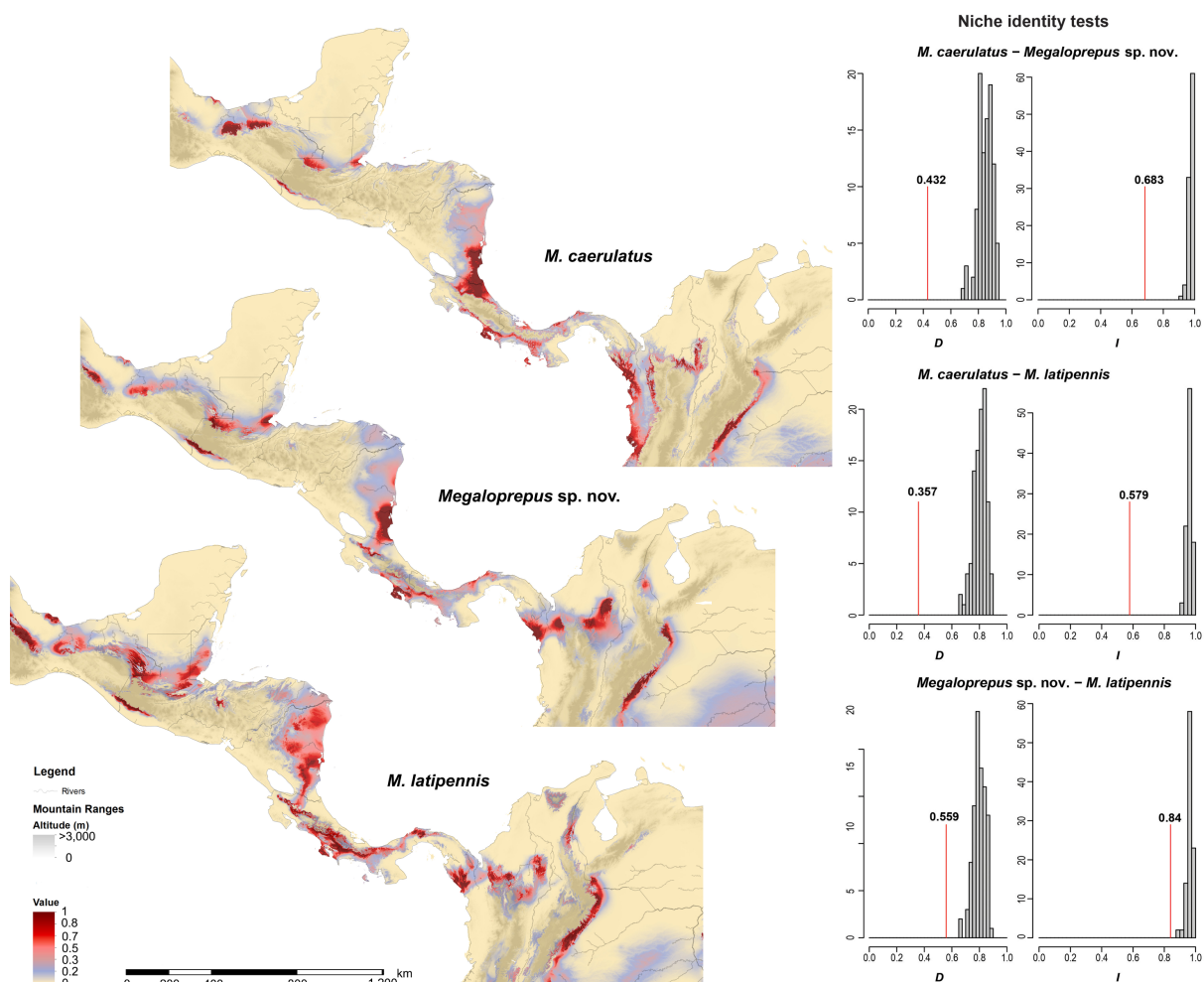
Figure A2.4.3: Niche analysis showing the modeled current distribution of the three *Megaloprepus* species for the entire Neotropics in a zoomed in version of Mesoamerica together with the results of the niche equivalency test.

The SDM was generated in MaxEnt vers. 3.3.3k using the four bioclimatic variables BIO 1, BIO 9, BIO 12, and BIO 17. The probability of occurrence is shown from 20% in light blue to 100% in dark red. The estimated ecological suitability for the three species is very similar and one species almost predicts the distribution of its sister species. However, larger differences are in central Guatemala and the Talamanca mountain range in central Costa Rica. On the right site are the results of the pairwise niche equivalency tests (see Table S4.3 for the statistics). The plots show the observed Schoener's *D* and the modified Hellinger distance's *I* with red lines and their corresponding values on the top. The histogram however indicates the random null distribution of ecological niche distance. All observed similarities are smaller than the null distribution, but not significant. Therefore, niches are considered as identical (Warren *et al.* 2008).

Table A2.4.3: Summary of the pairwise niche equivalency tests calculated using the package phyloclim (Heibl & Calenge 2013) in R Studio (R version 3.3.3, RCoreTeam 2014) based on the SDMs. Significant results in the Schoener's *D* and the modified Hellinger distance's *I* indicate identity of niches.

a	b	*D*	*I*	Niche *D/I*
M. caerulatus	*M. latipennis*	0.357***	0.579***	niche models are more similar
M. caerulatus	*Megaloprepus* sp. nov.	0.432***	0.682***	niche models are more similar
M. latipennis	*Megaloprepus* sp. nov.	0.559***	0.839***	niche models are more similar

Table A2.4.4: Pairwise comparisons testing if species ranges are more different than probable based on the ecological background differences. The background similarity test, a two tailed test, was calculated using the R package phyloclim (Heibl & Calenge 2013) in R Studio (R version 3.3.3, RCoreTeam 2014). 100 randomizations were performed to generate a null distribution of of overlap values.

Shown are the Schoener's *D* and the modified Hellinger distance's *I*, whereat H_0 of similarity is rejected if similarity falls outside of the 95% interval. All comparisons show that niches are similar. But among *M. caerulatus* and *M. latipennis* a divergence was observed in the reciprocal comparison (bold). Please see Fig. S4.5.

a	b	D	a → b		b → a		I	a → b		b → a		Niche
			2.5%	97.5%	2.5%	97.5%		2.5%	97.5%	2.5%	97.5%	
M. caerulatus	*M. latipennis*	0.35	0.14	0.23	0.27	0.31	**0.56***	0.31	0.47	0.54	0.60	similar / similar
M. caerulatus	*Megaloprepus* **sp. nov.**	0.46	0.15	0.21	0.22	0.27	0.71	0.35	0.44	0.48	0.54	similar / similar
Megaloprepus **sp. nov.**	*M. latipennis*	0.58	0.20	0.28	0.26	0.32	0.86	0.45	0.56	0.53	0.61	similar / similar
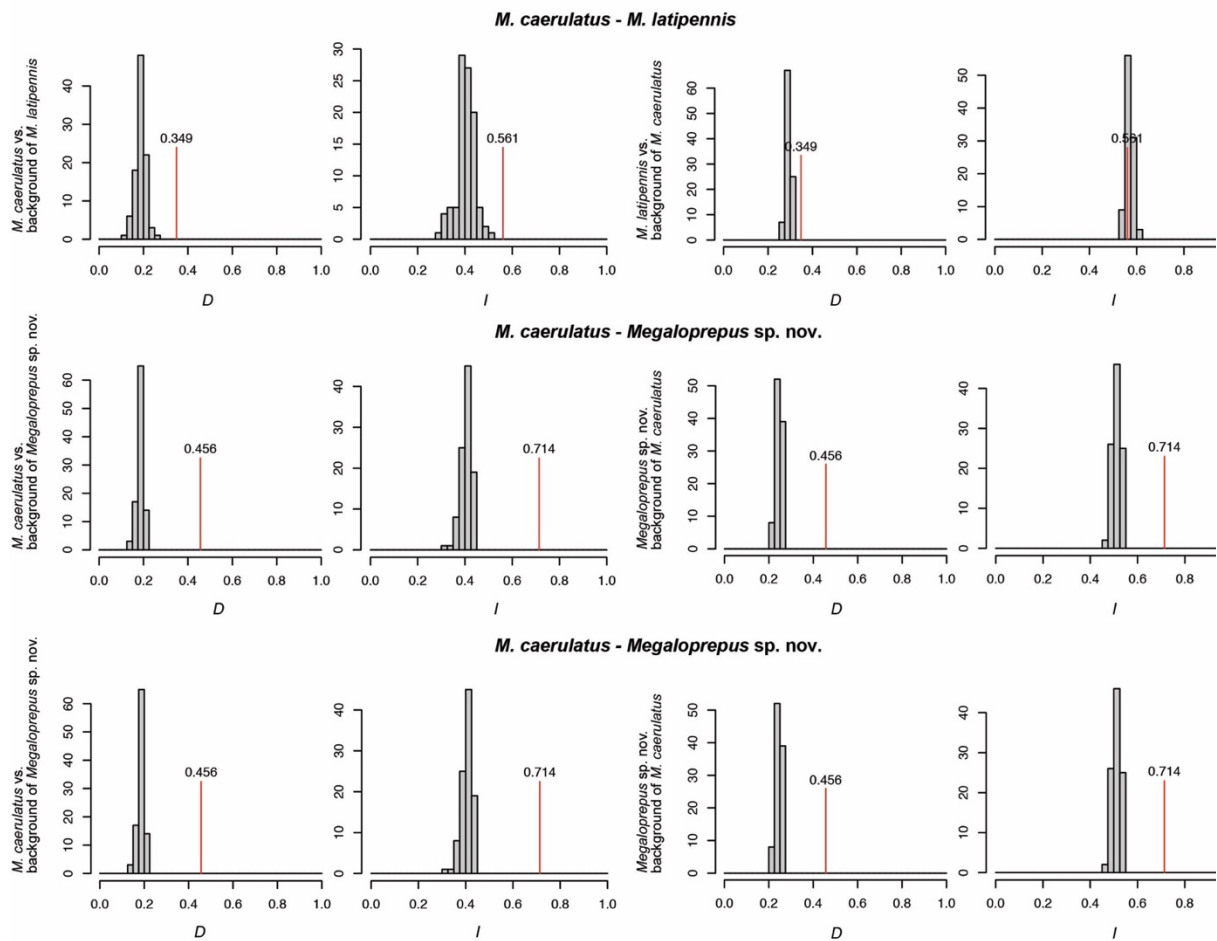
Figure A2.4.4: Results of the background similarity test (please see the statistics in Table S4.4) calculated using the package phyloclim (Heibl & Calenge 2013) in R Studio (R version 3.3.3, RCoreTeam 2014) based on the SDMs.

Hereby the Schoener's *D* and the modified Hellinger distance's *I* are estimated and plotted on the random niche distribution of the second species. The plots show the observed *D*'s and *I*'s with red lines and their corresponding values on the top. Most estimated values are larger than the background indicating niche similarity, but no value is significant. The only significant Hellinger *I* was observed in the comparison of *M. latipennis* against the background of *M. caerulatus*. Here they are significantly similar.

Table A2.4.5: Results of the Niche.Equivalency tests using all environmental variables used to develop the SDM (either without forest data and with), as background for the three *Megaloprepus* species: *M. caerulatus*, *M. latipennis* and *Megaloprepus* sp. nov. using the R package 'dismo' (Hijmans *et al.* 2017) in R studio with the R version 3.3.3 (RCoreTeam 2014). Both, Schoener's *D* and modified Hellinger distance's *I* (in bold) (Schoener 1968) are not significant ($p > 0.05$) between species (100 replications) implying that niche models are identical (Warren *et al.* 2008).

	M. caerulatus	M. latipennis	Megaloprepus sp. nov.
without forest			
M. caerulatus		**0.55**	**0.81**
M. latipennis	0.29		**0.88**
Megaloprepus sp. nov.	0.52	0.59	
with forest			
M. caerulatus		**0.57**	**0.78**
M. latipennis	0.30		**0.88**
Megaloprepus sp. nov.	0.51	0.60	

A3   Four in one – revalidation of the genus *Megaloprepus*

A3.1: Total barcode for the four *Megaloprepus* species and *Mecistogaster linearis* and *Coryphagrion grandis*

A3.2: Statistic comparisons of abdominal length and wing length.

A3.1: Total barcode for the four *Megaloprepus* species and *Mecistogaster linearis* and *Coryphagrion grandis* as out groups using the mitochondrial Folmer barcoding region of the cytochrome oxidase 1 (CO1). Shown are simple pure (sPu) and simple private (sPr) character within the alignment.

TAXA \ Position	C. grandis	M. linearis	M. diaboli GuCSG		M. diaboli HnPb			M. diaboli CNP			M. latipennis RBLT				GuLL	GuRB	M. brevistigma PhPe	M. caerulatus NiBa			LS			BCI			CO	
5	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
8	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
11	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	T	T	T	T	T	T	T	T	T	T	T
14	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
17	A	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A
20	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
23	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
24	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
26	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
27	A	G	A	A	A	A	A	A	A	A	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G
29	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
33	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T
35	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A
44	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
47	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
50	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
53	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
56	A	A	A	A	A	A	A	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
59	T	C	T	T	T	T	T	T	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A
62	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
65	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G
68	A	A	C	C	C	C	C	C	C	C	A	A	A	A	A	A	T	C	C	C	C	C	C	C	C	C	C	C
69	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
71	T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
77	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
80	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
86	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
93	T	T	C	C	T	T	T	C	C	C	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
98	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	T	T	T	T	T	T	T	T	T	T	T
101	C	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
105	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T
109	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
110	C	A	A	C	C	C	C	C	C	C	T	T	T	T	T	T	C	T	T	T	T	T	T	C	C	C	T	T
114	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
117	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
119	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
120	G	A	G	G	G	G	G	C	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G
121	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
122	A	A	C	C	C	C	C	C	C	C	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T
125	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
127	G	G	G	G	G	G	G	C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
128	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
129	T	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
131	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A
137	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G
140	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
144	G	G	G	G	G	G	G	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
146	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A
149	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T	T	T
152	C	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

Pos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
155	C	A	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
161	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
164	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	G	G	G	G	G	G	T	T	T	G	G
167	T	T	T	T	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
170	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T
173	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
174	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
176	A	T	G	A	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
179	A	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	C	C	C	C	C	C	C	C	C	T	T
182	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A
184	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
185	A	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
188	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	T	T	T	C	C
191	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
194	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
200	A	A	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
202	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
203	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
206	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
209	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
212	T	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
213	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
218	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
221	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
224	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
227	A	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	G	G	G	G	G	G	G	G	G	G	G
228	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
230	A	A	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
233	C	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
234	C	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C
236	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
239	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
242	T	G	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
248	C	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
251	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	G	G	G	G	G	G	G	G	G	G	G
254	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
255	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T
260	G	G	A	A	A	A	A	A	A	A	G	G	G	A	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A
263	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
266	T	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
272	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
278	A	A	T	T	T	T	T	C	C	C	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C
281	T	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
282	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
283	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
284	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
290	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T
293	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
297	T	A	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
299	C	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A
302	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
305	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
307	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
308	A	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T
312	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
314	A	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
321	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
325	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	A	C

Pos																													
326	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
327	T	T	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T
329	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
335	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
338	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
339	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
341	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
345	G	G	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A
347	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
353	T	G	C	C	C	C	C	C	C	C	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
357	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
359	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
363	C	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T
365	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
369	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	A	T	
374	C	T	C	C	C	C	C	T	T	T	C	C	C	C	C	C	C	T	A	A	A	A	A	A	A	A	A	A	A	A
375	C	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	
377	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
380	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
383	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	
384	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	
386	C	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	T	
389	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
392	A	G	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	
398	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	
401	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	
405	C	T	C	C	C	C	C	C	C	C	A	A	A	A	A	A	A	T	C	C	C	C	C	C	C	C	C	C	C	
407	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	
413	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	
419	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
422	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
425	C	T	A	A	A	A	A	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	
428	T	T	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	

A3.2 Statistic comparisons of abdominal length and wing length.

Comparisons of forewing length and abdomen length in males from different localities were compared by a multifactorial ANOVA.

The Table S2 displays measurements for all studied male individuals. Although sample sizes vary strongly among species, significant results were obtained. Male FW length comparisons showed significant differences between *M. caerulatus* and *M. diaboli* **sp. nov.** as well as *M. caerulatus* and *M. latipennis* ($p = 0.025$ and $p < 0.002$, respectively; TukeyHSD). Species comparisons of abdomen length showed significant differences only between *M. caerulatus* and *M. brevistigma* ($p = 0.008$; TukeyHSD).

Table A3.2: Size of male specimens measured in the field and in museum collections of forewing length (FWL) and abdomen length (AL). Extremes and means are shown, with standard derivation ($\pm$ sd). The sample size for all calculations varies from the total number of examined material because some specimens could not be measured accurately.

	M. diaboli $N = 16$		*M. latipennis* $N = 30$		*M. brevistigma* $N = 10$		*M. caerulatus* $N = 102$	
	FWL	AL	FWL	AL	FWL	AL	FWL	AL
min	59.00	73.00	61.25	74.25	63.00	79.00	43.00	60.50
max	94.40	99.50	92.95	98.40	94.20	98.50	87.10	106.10
mean	76.15	87.30	76.88	87.13	77.29	87.05	69.02	84.40
$\pm$ sd	9.22	7.68	7.34	7.12	9.04	7.68	8.99	8.99

A4 **Transcriptome profiling in *Megaloprepus***

A4.1 Table. Genes related to wing development and coloration in insects. Included are the descriptions of the main functions for genes involved in pigmentation (A4.1A) and genes of the wing-patterning gene network (GRN; A4.1B) found in *M. caerulatus* including their Unigene IDs and FPKMs.

A4.1 Fig. E-value distribution for BLAST hits against *Megaloprepus'* predicted proteins.
A4.2 Fig. Length distribution of annotated sequences within the *M. caerulatus* transcriptome.

A4.1 File. Amino acid sequences of *M. caerulatus* stress response genes.
  Is available at: https://doi.org/10.1371/journal.pone.0189898.s004
A4.2 File. Amino acid sequences of *M. caerulatus* house keeping genes.
  Is available at: https://doi.org/10.1371/journal.pone.0189898.s005
A4.3 File. Homeobox genes (HOXL subclass and NKL subclass) homeodomain amino acid alignment for *I. elegans*, *L. fulva* and *M. caerulatus*.
  Is available at: https://doi.org/10.1371/journal.pone.0189898.s006
A4.4 File. Amino acid sequences of wing genes found in *M. caerulatus*. Included are representatives of the wing-patterning gene network and the four signaling pathways: Notch (N), wingless (wg), Decapentaplegic (Dpp) and Hedgehog (Hh), and pigmentation genes.
  Is available at: https://doi.org/10.1371/journal.pone.0189898.s007
A4.5 File. In silico quantification of *Megaloprepus* expression levels.

A4.1A Table. Summary of genes involved in pigmentation of insect wings and other body parts identified within the *M. caerulatus* transcriptome. For each gene the corresponding Unigene ID, FPKM values and some of their main functions are shown.

gene	Unigene ID	read count	FPKM	Main function
Aristaless (*al*)	TR167515	8	0.77	*Homeobox* protein, precedes *wg* expression - wing pattern evolution
Black (*b*)	TR93729	12	0.65	Cysteine sulfinic acid decarboxylase, melanin pathway
Ebony (*e*)	TR1248	35	1.40	Inhibits melanization in non-black wing regions
Tan (*t*)	TR192472	44	1.81	Melanin pathway
Yellow (*yellow*)	TR182547	2,205	61.44	Melanin pathway
	TR183295	2,866	71.30	
Ecdysone Receptor (*EcR*)	TR215139	1,262	41.31	Ligand-dependent transcription factor, steroid receptor super family
Araucan (*ara*)	TR162214	32	2.24	Transcription factor, imaginal disc-derived and wing vein specification
Henna (*Hn*)	TR231264	874	24.74	Pteridine pathway - regulates phenylalanine
bric a brac 1 (*bab*)	TR165906	61	3.00	Axis patterning, involved in female specific pigmentation in *Drosophila* sp.
bric a brac 2 (*bab*)	TR206843	149	8.67	
Dor (*dor*)	TR165868	1,574	28.41	Pigment granule gene, vacuolar protein sorting-associated protein 18 homolog
Pale protein (*ple*)	TR106565	217	4.18	Tyrosine 3-monooxygenase, melanin pathway
Phenoloxidase 2 (*PO*)	TR158701	32,165	606.37	Melanization, final polymerization cuticular pigments, immune response
	TR200723	10,582	181.28	
	TR201540	441	8.99	
Rosy (*ry*)	TR177287	1,035	13.73	Xanthine dehydrogenase, pteridine pathway
	TR202403	136	3.21	
White (*w*)	TR203851	93	4.56	Ommochrome pathway - ommochrome precursor transporter gene
Scarlet (*st*)	TR217941	774	23.58	Ommochrome pathway, transport of pigment precursors
	TR182410	189	4.21	
Garnet (*g*)	TR16501	89	4.20	Pigment granule gene, organ pigmentation
Mal (*mal*)	TR151410	97	2.23	Molybdenum cofactor sulfurase, cofactor for synthetic processes: essential for xanthine dehydrogenase (XDH) and aldehyde oxidase (ADO)
	TR151410	35	1.51	
Prat (*Prat*)	TR15147	21	1.54	Amidophosphoribosyltransferase, purine metabolism, pteridine pathway
Vermillion (*v*)	TR155117	845	30.07	Ommochrome pathway, ommochrome enzyme gene, brown eye pigment, tryptophan 2,3-dioxygenase

A4.1B Table. Genes of the wing-patterning gene network found in *M. caerulatus* with their corresponding Unigene ID, FPKM values and some of their main functions.

gene	Unigene ID	read count	FPKM	Main function
Hedgehog (*Hh*)	TR219975	265	10.63	Short range signaling molecule, embryonic segmentation, segment polarity
Decapentaplegic (*Dpp*)	TR23365	19	1.98	Long range signaling molecule; embryonic dorsal/ventral patterning; boundaries between appendage compartments
	TR199155	67	2.56	
Optomor-blind (*bi* [*Omb*])	TR174327	226	20.03	Transcription factor, wing differentiation and wing formation
Spalt major (*salm*)	TR119349	79	2.93	Transcription factor, *Dpp* signaling, controls vein-specific expression of knirps and iroquois gene complex - controls position of the wing veins
	TR211792	132	6.63	
Serrate (*Ser*)	TR178981	1,344	28.57	Short range signaling molecule, notch ligand; initiation of wing margin (dorsal-ventral) by controlling cell proliferation
Cut (*cut*)	TR148092	98	3.88	Selector gene (homeodomain transcription factor), wing and follicle cell morphogenesis
Achaete / Scute (*ac/sc*)	TR159427	610	20.65	Transcription factor, involved in the determination of the neural fate; (scute alone responsible for sex determination)
Wingless (*wg*)	TR153309	159	9.53	Long range signaling molecule, segment polarity gene, wing morphogenesis, mediating dorso-ventral compartment boundary
Scalloped (*sd*)	TR191046	573	19.87	Selector gene (homeodomain transcription factor), downstream effector molecule (wingless pathway)
Blistered [Serum Response Factor] (*bs* [*srf*])	TR175967	390	15.56	Selector gene (homeodomain transcription factor), vein / intervein formation
	TR175378	67	2.22	
Spitz (*spi*)	TR160328	60	2.67	Ligand for the epidermal growth factor receptor, differentiation of ommatidial cell types
Ultabithorax (*Ubx*)	TR214266	21	1.88	Selector gene (*Hox* gene), 'correct' dorsal and ventral appendage patterning in the third thoracic segment
Sex combs reduced (*Scr*)	TR157439	65	4.37	Selector gene (*Hox* gene), first thoracic segment development
Extradenticle (*exd*)	TR187890	498	31.83	Transcription factor, interacts with homeotic proteins for specificity of homeotic protein binding to DNA - cofactor of *Ubx*, transformation of segmental identities
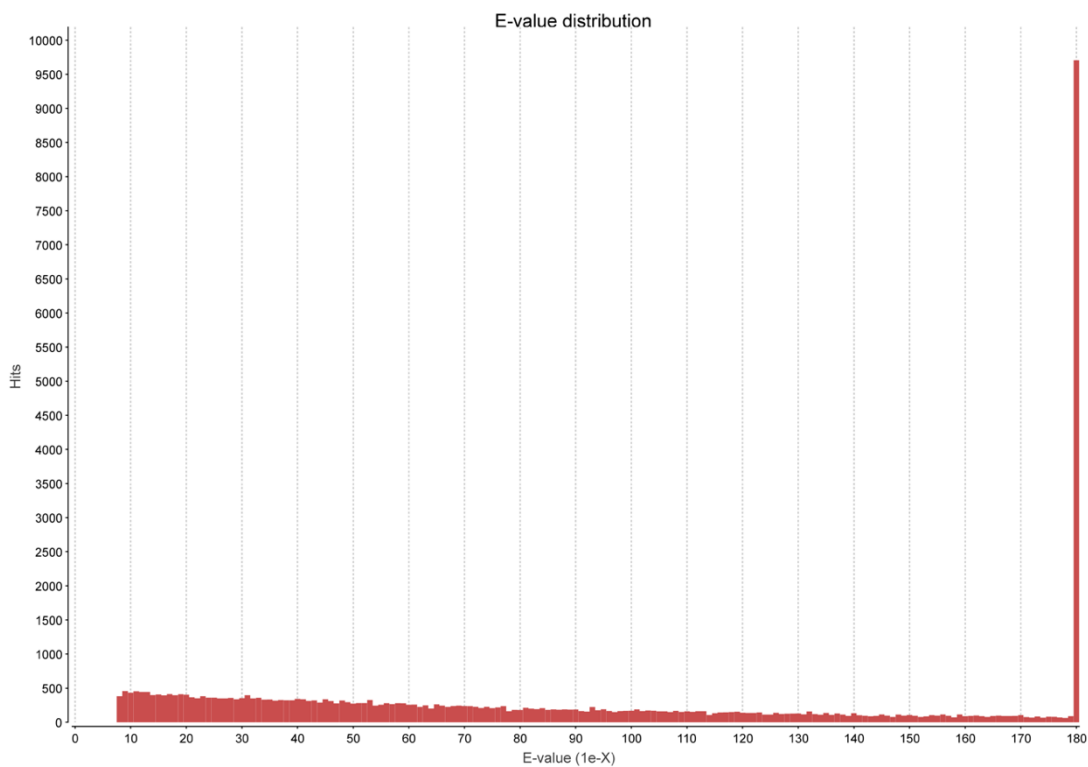
A4.1 Fig. E-value distribution for BLAST hits against *Megaloprepus'* predicted proteins.
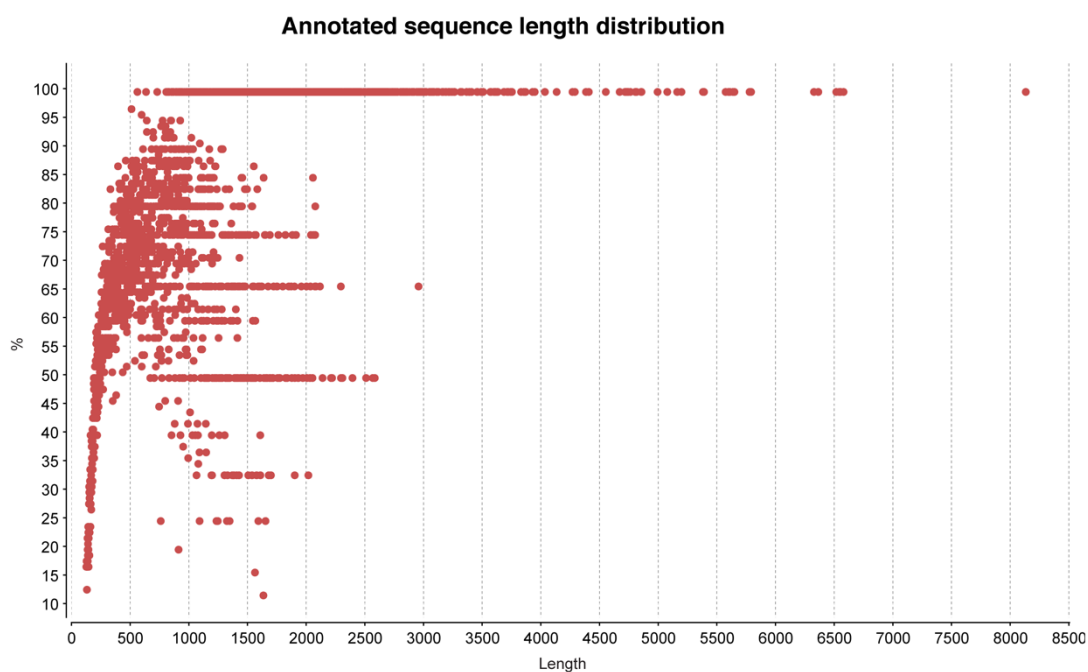
A4.2 Fig. Length distribution of annotated sequences within the *M. caerulatus* transcriptome.

A4.5 File. In silico quantification of *Megaloprepus* expression levels

Although we did not design this as an expression study, transcripts with very high or very low read coverage can give a gross indication of comparative expression levels. In order to detect these, we calculated FPKM (fragments per kilobase of transcript per million mapped reads) for the functionally annotated proteins following the suggestions of [85]. Expression levels ranged from 1 to over 1 million, suggesting a wide range of expression involving few highly expressed genes. The 50 transcripts with the highest read counts were mostly giant muscle proteins, such as titin isoforms, and twitchin, which belong to the titin/connectin superfamily. A deeper look to the wing genes is depicted in S5B. Future expression studies with appropriate biological replication are needed to potentially connect these patterns to developmental processes.
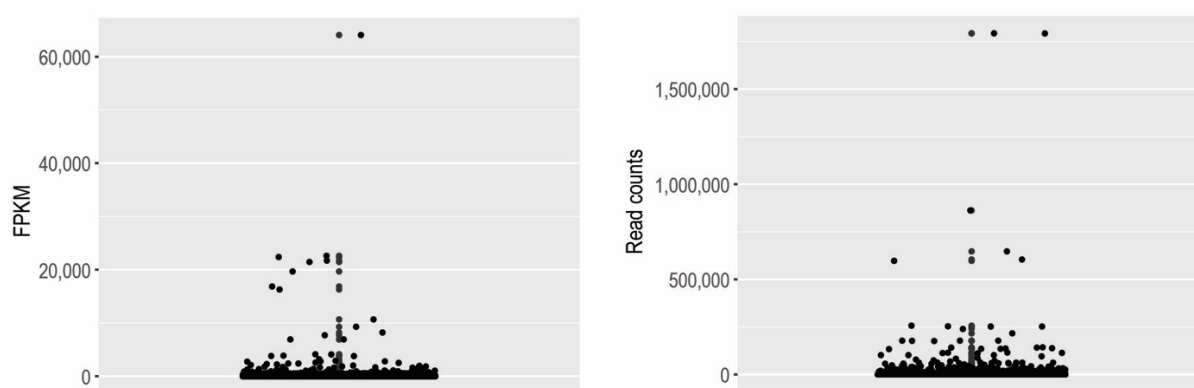
A4.5A. FPKM (fragments per kilobase of transcript per million mapped reads) and read count of annotated transcripts of the *M. caerulatus* transcriptome.
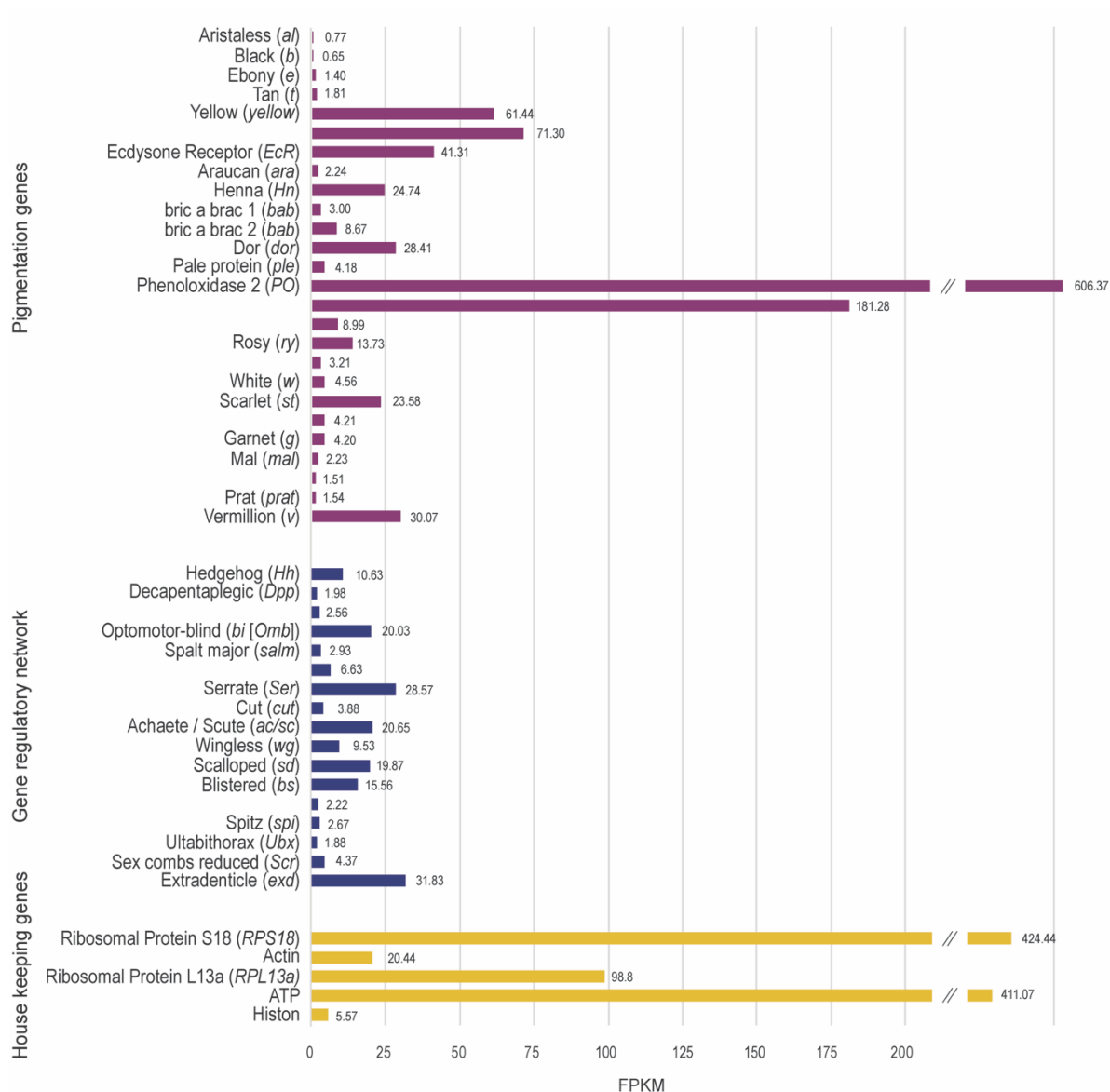
A4.5B. Genes related to wing pigmentation (A) and the gene regulatory network (B) identified within the *M. caerulatus* transcriptome. Shown are relative expression values per gene (fragments per kilobase of transcript per million mapped reads: FPKM) relative to a selection of house keeping genes.

A5 RNA-Seq goes into detail!

RNA-Seq goes into detail! – A short note on future speciation research in the Neotropical damselfly genus *Megaloprepus*

Background

Studies aiming to reveal the molecular mechanism behind new phenotypic traits are of high interest in contemporary research (e.g. [1]). One reason is the increasing improvement of next generation sequencing, which not only provides deeper insights but also allows more accurate answers. However, animal coloration and the appearance of new color pattern is directly linked to diversification and local adaptation.

Today the molecular mechanisms and pathways for coloration are still not fully understood, whereas coloration changes are most likely to evolve by a complex interplay of environment, development and selection. In arthropods, it has been assumed that a few genes with large effects are responsible for color variation [2]. Structural and regulatory changes seem to simultaneously influence novel phenotypic pattern, but it stays unanswered which genes are involved in which pattern and to which extend [3]. Although, in Lepidoptera research on wing coloration has a long history (e.g. [2, 4, 5]), only recently more precise results on the genomic basis of wing coloration were published. In the Swallowtail butterfly *Papilio polytes* (Lepidoptera: Papilionidae), changes in the gene networks (cf. [6]) were described to modify red and yellow color patterns in female wings [7]. Precisely an up-regulation of the biosynthetic genes for kynurenine and N-β-alanyldopamine (NBAD) and the *Toll* signaling genes was found in a differential expression analysis [7]. A second work has identified the *cortex* gene to play another important role in wing coloration [5, 8]. *Cortex* belongs to a fast-evolving subfamily of the conserved *fizzy* family that is a cell-cycle regulator. Although its function among flying insects may not be conserved, it has become a target gene to in color pattern variation (cf. *Heliconius* moths [5]). In British peppered moths *Biston betularia* (Lepidoptera: Geometridae) a large, tandemly repeated, transposable element in *cortex* was found causing its increased transcription and consequently adult melanization [8]. If such new traits become advantageous, coloration changes represent a prerequisite leading to diversification.

To my knowledge, there are no studies investigating coloration in odonate wings or the appearance of novel phenotypic traits on its genomic base, even though odonates are important ecological model organism [9]. One key indicator species could be the genus *Megaloprepus* (Odonata: Pseudostigmatidae). In Mesoamerica this genus has experienced a radiation into three species before the ice ages. This radiation was non-adaptively but is now most likely accompanied by selection for wing traits only in *M. caerulatus*. In order to understand the variation between different species, comparative *de novo* RNA-Seq studies could help to analyze and highlight the following questions:

(i)   Are there expressed genes, which might be responsible for wing development and/or coloration in *Megaloprepus* transcriptomes?

(ii)   If yes, are there any structural changes that could give a hind to phenotypic evolution?

(iii)   Are the transcriptomes of *Megaloprepus* useful to identify genes or pattern that are involved in speciation?

Material and methods

As one step towards this aim, three types of comparisons: (i) intra-species and intra-population, (ii) intra-species and inter-population and (iii) inter-species (cf. Figure 1) were established. In total four larval transcriptomes were compared in the hope to get insights to the evolution of new wing patterns and divergence in the genus *Megaloprepus*. To do so, *M. caerulatus* from the Biological Research Station La Selva (Costa Rica) and *M. diaboli* from the Corcovado National Park (Costa Rica) and the Pico Bonito National Park (Honduras) were selected.
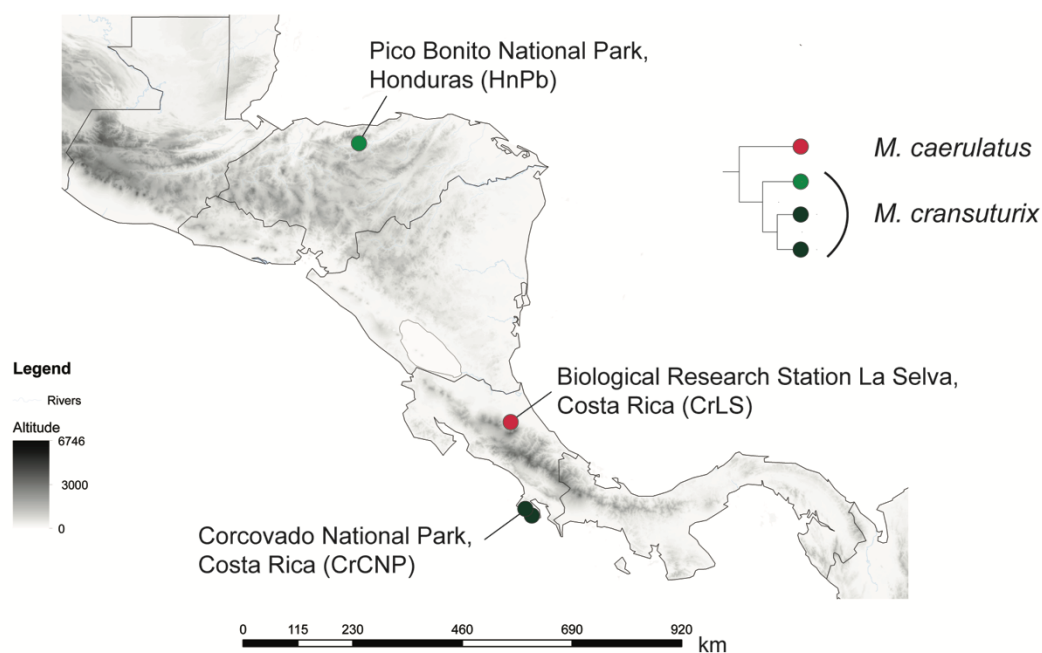
Figure 1: Geographical origin for the two *Megaloprepus* species used in the comparative study. *Megaloprepus diaboli* originates from the Corcovado National Park (Costa Rica) and the Pico Bonito National Park (Honduras), and *M. caerulatus* from the Biological Research Station La Selva (Costa Rica).

Sampling and RNA-Sequencing

Larvae were collected from natural tree holes during the field season 2011/12. After collection, individuals were immediately euthanized and stored in RNAlater (Thermo Fisher Scientific Inc., USA). Unfortunately, field conditions in remote areas limit the direct access

to appropriate sample storage. Consequently, samples had to rest at field temperatures for at maximum two weeks before they were stored at -80°C. This lead us the assumption of an unknown degradation (i.e. at the ployA region) and therefore all steps from extraction to library preparation were adapted to this condition. Individual RNA was extracted from the thorax including the wing buds of in total three larvae (two from CNP and one from HnPb) using a modified RNA isolation protocol (cf. [6]). Hereby, the larva was defrosted and dissected on ice. The thorax was frozen in a liquid nitrogen bath and grinded frozen. For the extraction of the total RNA we joined TRiZOL reagent (Invitrogen, USA) for the isolation with the RNeasy Micro kit (Qiagen Inc., USA) for purification. Quality and quantity were assessed via the BioAnalyzer 2100 (Agilent Inc., USA). The New York Genome Center (nygenome.org) conducted library preparation and sequencing. Hereby, in addition to the Illumina's TruSeq Total RNA kit (Illumina, Inc., USA) a Ribo Zero treatment (Ribo-Zero rRNA Removal Kit, Illumina, Inc., USA) was selected to remove superfluous rRNA. The final sequencing was performed as paired-end and strand specific (2x125 bp) on an Illumina HiSeq 2500 (Illumina, Inc., USA).

De novo *assemblies, assembly evaluation and functional annotation*

Before comparative analysis are possible RNA-Seq data must undergo a principle three-step protocol, which includes (i) a quality check, (ii) read cleaning and trimming, and (iii) the assembly and its evaluation. The quality of raw reads was checked with FastQC [10] using a Phred-like score and low-quality reads below the Q20 level were trimmed together with the adapter sequences using Trimmomatic [11] while retaining a minimum read length of 30 bp. Potential rRNA was removed from the dataset using Sort-MeRNA version 2.0 [12]. Kraken taxonomic sequence classification system version 0.10.5 [13] was applied to filter additional prokaryotic sequences via kmer congruence. Kraken aims to remove contamination from the sequence data as they disturb the assembly and continuous analyses. The strict quality control of the reads reduced approximately 2-7.5% of raw reads for each transcriptome and app. $100x10^6$ PE reads were used for individual assemblies (Table 1).

Individual *de novo* assemblies were conducted using the software Trinity vers. 2.0.6 [14] including a normalization step and retaining only transcripts with a minimum length of 300 bp. The assembly evaluation followed the steps suggested by Feindt *et al.* (cf. [6, 15]), which included the determination of the general eukaryotic core gene content via BUSCO (Benchmarking Universal Single-Copy Orthologs) vers. 1.1 [16]. The RSEM contig impact score was calculated in DETONATE to identify transcripts with no read information and their continuous removal was implemented using an in-house R-script [17]. Additional redundant transcripts were removed using CD-HitEST version 2.17.0 [18] on the 98% level. Finally, in the last step open reading frames (ORF) were determined on the six-frame translation via TransDecoder.LongORFs (http://transdecoder.github.io) [19]. Quality of the open reading frames were further improved by a second homology-based search. Here the results from a BLASTx (e-value cutoff: 1e-5) against an UniProtKB [20] arthropod data base and a homology search agaisnt Pfam protein domain database [21] using Hmmer version 3.1 [22] were retained in TransDecoder.Predict [19]. The transcriptomes were annotated using a BLAST search [23] against a specially designed insect genome database and Blast2GO [24, 25] following the descriptions in Feindt *et al.* [6].

Table 1: Trimming statistics for the three *M. diaboli*. Using three different trimming methods [10-13] low quality reads, rRNA reads and contaminants were removed.

ID	CrCNPL1	CrCNPL17	HnPbL17
Region	Corcovado National Park, Costa Rica		Pico Bonito National Park, Honduras
Read Quality			
Number of PE reads before trimming	104,950,058	108,103,742	116,757,288
Read length	125	125	125
Trimmomatic			
Number of cut low quality	35,452	41,922	43,084
Percent low quality reads	0.03	0.04	0.04
Reads remaining after Trimmomatic	104,914,606	108,061,820	116,714,204
SortMeRNA			
Number rRNA reads	1,973,106	762,177	787,526
Percent rRNA reads	1.88	0.71	0.67
Reads remaining after rRNA removal	102,941,500	107,299,643	115,926,678
Kraken taxonomic sequence classification system			
Reads classified as contaminants	2,975,940	490,680	284,127
Percent classified as contaminants	2.89	0.46	0.25
PE reads remaining after clean-up	99,371,930	106,643,058	115,514,400

Transcriptome Comparisons and SNP analysis

For the comparisons we additionally downloaded the transcriptome for *M. caerulatus* (CrLSL17, GEXY00000000) from the Transcriptome Shotgun Assembly database. First, sequence similarity among the four transcriptomes was compared by a reciprocal BLAST search using an e-value of $1e^{-7}$ [6]. Furthermore, the transcripts were classified into orthologous clusters via OrthoVenn [26], which were compared, and their overlap determined.

A variant analysis via Single-Nucleotide Polymorphism (SNP) was performed following the GATK (Genome Analyzer Toolkit) Best Practices Guideline for RNA-Seq data (https://www.broadinstitute.org/gatk/ guide/article?id=3891; [27, 28]). Hereby a very strict filtering was applied: only heterozygote SNP's were retained that had read a coverage > 20. This strict setting may have deleted some true positives, but also false positives were removed. In addition, two comparisons were run allowing different insights: in setting (A) a larva from the Corcovado National Park (CrCNPL17) was used as a reference to compare intra-specific differences and setting B included the *M. caerulatus* (CrLSL17) as reference allowing to detect differentiated genes (inter-species comparison).

Differential expression analysis

Although the present research is not suited for differential expression analyses (DE), the estimates between individuals may highlight genes of interest. To estimate similar and differential expressed genes between species as well as between individuals from one species, combined assemblies for the genus *Megaloprepus* and for *M. diaboli*, respectively,

were conducted. For this purpose, the trimmed and filtered reads of *M. diaboli* (presented here) and of *M. caerulatus* (SRR3997526) were used together for a *de novo* assembly in Trinity. The assembly evaluation followed the steps described above accompanied by an additional abundance estimation [19] and CD-Hit clustering at 98% level of similarity. For the final expression comparisons, the Trinotate pipeline [19] was implemented. Hereby the individual reads were mapped back to the joined assembly and the R package edgeR [17] was accessed for comparisons at the 4x, 100x and 1,000x change-level. Furthermore, important wing genes (wing development and coloration genes) were extracted and their expression patterns were compared (cf. [6]).

Results and Discussion

Transcriptome Comparisons and SNP analysis

The assembly resulted into three high quality transcriptomes with BUSCO core gene contents of at minimum 93% and between 60,000 and 70,000 predicted genes (CrCNPL17: 60,593; HnPbL17: 65,238; CrCNPL1: 71,855).

Table 2: Assembly statistics for the three Trinity *de novo* assemblies.

ID	CrCNPL1	CrCNPL17	HnPbL17
Assembly assessment parameters			
Transcripts > 300 bp	574,659	557,516	569,595
Total contig length	643,468,588	680,000,643	749,417,451
Mean contig size (bp)	1,119.7	1,219.7	1,315.7
Number of contigs > 1,000 nt	160,365	177,215	194,542
N50 contig length	1,914	2,035	2,328
Longest contig	42,344	44,296	38,368
BUSCO - annotation completeness via universal single-copy orthologous genes			
Complete Single-Copy BUSCOs	2,256	2,261	2,248
Complete Duplicated BUSCOs	1,507	1,483	1,528
Fragmented BUSCOs	246	251	272
Missing BUSCOs	173	163	155
Total BUSCOs in %	93.5	93.9	94.2

The sequence similarity obtained by the reciprocal BLAST searches on the predicted genes was high. The three *M. diaboli* species share 16,800 genes, and 15,523 genes are shared between the two *Megaloprepus* species (Figure 2 A, B). Contrary, the OrthoVenn revealed 7,535 and 5,684 shared gene clusters (Figure 2 C, D). Hereby it is obvious that the specimen from Honduras has the highest number of species-specific genes. This might be because this larva could have been in a different life stage such as before or after mold, whereas the other larvae not. Such differences are difficult to observe in the field. Consequently, observations concerning the larval stage and the special expression pattern shall take place under laboratory settings.
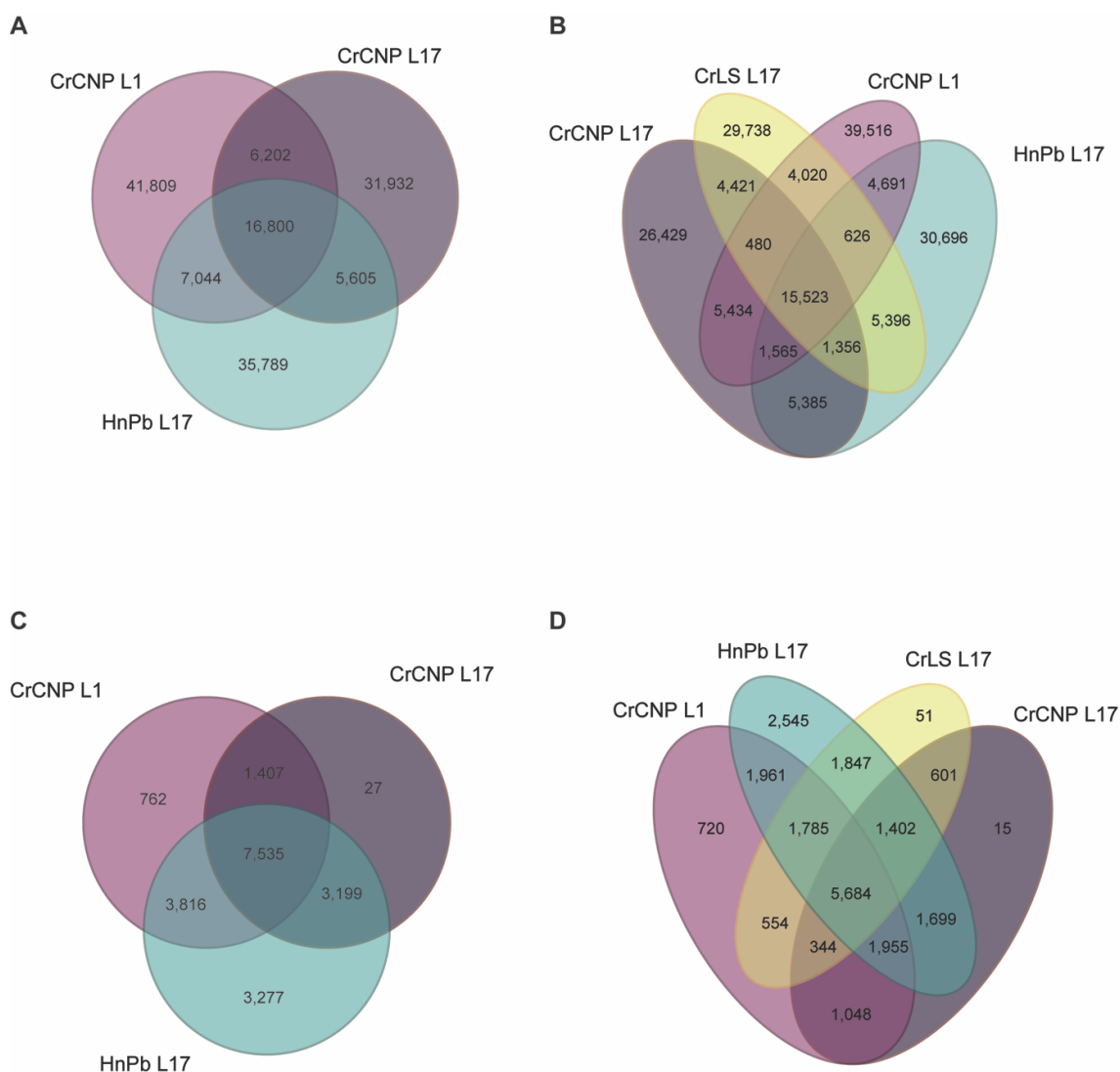
Figure 2: Sequence similarity estimated by a reciprocal BLASTp search (A and B) and via OrthoVenn (C and D). Most obvious is the high number of species-specific genes in HnPbL17.

The SNP analysis revealed a positive correlation between the number of SNP's and phylogenetic distance (intra-species: 2,178 SNP's, inter-population: 4,183 SNP's, inter-species: 9,003 SNP's). Furthermore, using the single *M. caerulatus* larva as reference 2,201 of putative genes that include SNP's were detected, which have an overlap among the three *M. diaboli* species. These genes could be under selection. Therefore, they should deserve a higher attention in future speciation and coloration studies, especially because the wing genes ebony and yellow are among them.

Differential expression analysis and wing genes

Using the Trinity software [19] 6,328 differential expressed transcripts were found at a 100-fold-change expression level (Figure 3), whereas at the 4-fold level 55,928 transcripts were estimated. Addtionally, the expression pattern of the wing genes (Figure 4) showed a great variation among individuals.

This study was not designed for a differential expression analysis. However, the genes detected here could give a hint for future targeted RNA-Seq studies or can be compared to the variant analysis. In general, differentially expressed genes are promising to compare the ecotypes (or populations) because genes under directional selection are assumed to change expression.
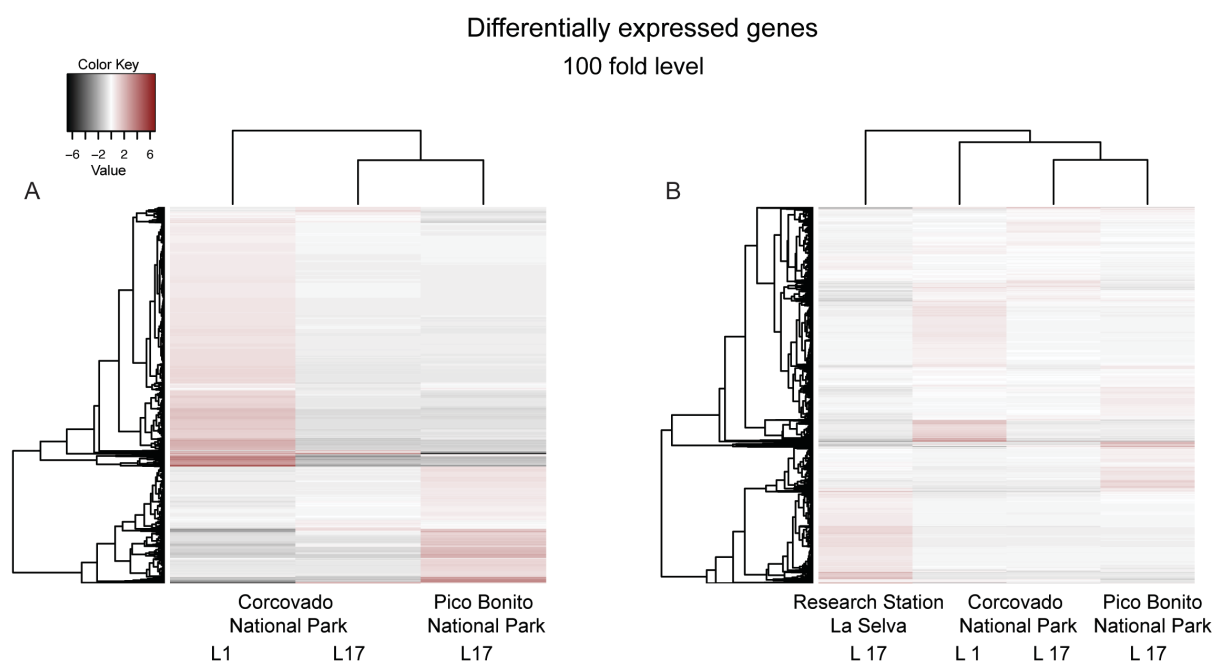
Figure 3: Differential expressed genes between 3 *M. diaboli* individuals (A) and between 4 individuals of two *Megaloprepus* species (B).
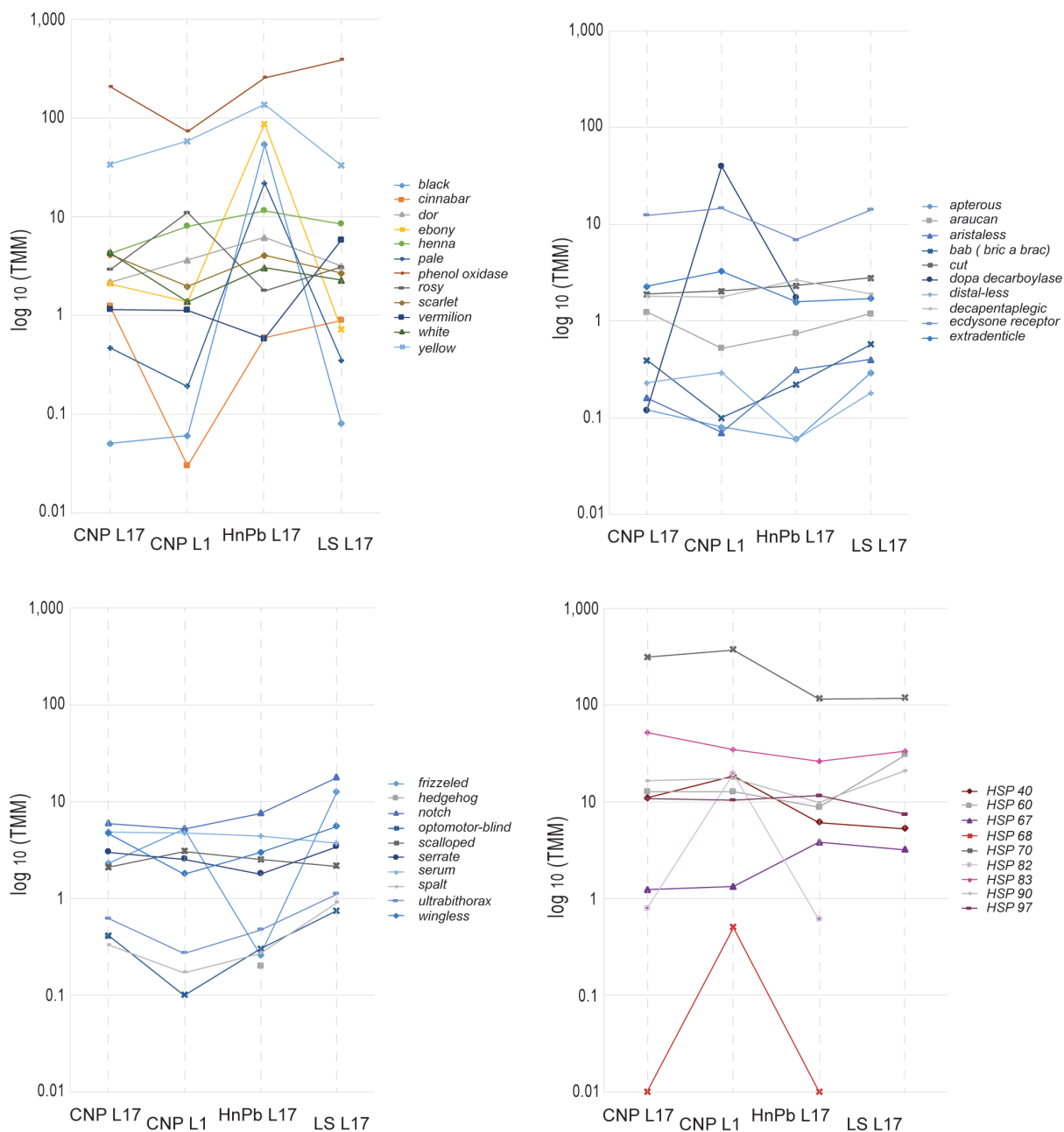
Figure 4: Differential expressed wing genes and heat shock proteins (HSPs) at the 100-fold-change levels estimated from the four *Megaloprepus* larvae.

Conclusion

To identify the genes that are responsible for interspecific differences, new phenotypic traits, color evolution and speciation in the post-genomic era is still a challenging field [29]. But a large amount of sequence data, that was sampled and sequenced under well-performed settings, promise interesting results. The transcriptomes shortly described above represent a groundwork for future evolutionary studies. Here we sequenced and annotated three transcriptomes of one Neotropical damselfly species from two different populations and compared them to its closest sister species. The overlaps were high, although we found more

than 3,200 species-specific genes in HnPbL17. Most interestingly, the SNP analyses revealed an increasing number of SNP's with increasing phylogenetic distance. Though transcriptomes only reflect a certain moment of expression, it is difficult to distinguish between genes that are simply not expressed or absent. Consequently, controlled comparisons across developmental stages would give more precise insights.

Another factor that could reveal interesting insights into coloration is to study how temperature influences color. Butterfly wings change their coloration with the season. In spring butterflies hatch tan and late summer/autumn butterflies hatch in dark red [30]. Daniels *et al.* showed in a continuous expression study that although this color changes occur from broad physiological responses, there could also be an effect to temperature [30]. This implies the importance of the environment, and its effects on permanent coloration changes could be indeed present. Understanding the genetic background of the evolution of different shapes and colorations of insect wings are of interest and allow a direct link between genetics, adaptation and speciation.

References

1. Oppenheim SJ, Baker RH, Simon S, DeSalle R. We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. Insect molecular biology. 2015;24(2):139-54.
2. Kronforst MR, Barsh GS, Kopp A, Mallet J, Monteiro A, Mullen SP, et al. Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. Pigment cell & melanoma research. 2012;25(4):411-33.
3. Wittkopp PJ. Evolution of gene expression. The Princeton guide to evolution, Princeton University Press, Princeton New Jersey, USA. 2013:413-9.
4. Jiggins CD. Ecological speciation in mimetic butterflies. AIBS Bulletin. 2008;58(6):541-8.
5. Nadeau NJ, Pardo-Diaz C, Whibley A, Supple MA, Saenko SV, Wallbank RW, et al. The gene *cortex* controls mimicry and crypsis in butterflies and moths. Nature. 2016;534(7605):106-10.
6. Feindt W, Oppenheim SJ, DeSalle R, Goldstein PZ, Hadrys H. Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world's largest wings. PloS one. 2018;13(1):e0189898.
7. Nishikawa H, Iga M, Yamaguchi J, Saito K, Kataoka H, Suzuki Y, et al. Molecular basis of wing coloration in a Batesian mimic butterfly, *Papilio polytes*. Scientific reports. 2013;3:3184.
8. van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-5.
9. Bybee S, Córdoba-Aguilar A, Duryea MC, Futahashi R, Hansson B, Lorenzo-Carballa MO, et al. Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. Frontiers in zoology. 2016;13(1):46.
10. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source. 2010.
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:btu170.
12. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.
13. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.
14. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature

biotechnology. 2013;29(7):644-52. doi: 10.1038/nbt.1883. PubMed PMID: 21572440; PubMed Central PMCID: PMCPmc3571712.

15. Oppenheim SJ, Feindt W, DeSalle R, Goldstein PZ. De Novo characterization of transcriptomes from two North American Papaipema stem-borers (Lepidoptera: Noctuidae). PloS one. 2018;13(1):e0191061.

16. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:btv351.

17. RCoreTeam. R: A language and environment for statistical computing. In: Computing RFfS, editor. Vienna, Austria: URL http://www.r-project.org/; 2014.

18. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.

19. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols. 2013;8(8). Epub 1512.

20. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. Plant Bioinformatics: Methods and Protocols. 2016:23-54.

21. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research. 2016;44(D1):D279-D85. doi: 10.1093/nar/gkv1344.

22. Wheeler T, Eddy S. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29(19):2487-9. doi: 10.1093/bioinformatics/btt403. PubMed PMID: 23842809; PubMed Central PMCID: PMCPmc3777106.

23. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-402.

24. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.

25. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 2008;36(10):3420-35.

26. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic acids research. 2015;43(W1):W78-W84.

27. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8.

28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010.

29. Wittkopp PJ, Carroll SB, Kopp A. Evolution in black and white: genetic control of pigment patterns in *Drosophila*. TRENDS in Genetics. 2003;19(9):495-504.

30. Daniels EV, Murad R, Mortazavi A, Reed RD. Extensive transcriptional response associated with seasonal plasticity of butterfly wing patterns. Molecular ecology. 2014;23(24):6123-34.

ACKNOWLEDGEMENTS

CURRICULUM VITAE

W I E B K E   F E I N D T

EDUCATION
since 11/11   **Leibniz University Hannover**

Ph.D. in Biological Science (Dr. rer. nat.)
Title: *"Conservation genomics: Speciation of the Neotropical damselfly species* Megaloprepus caerulatus *– as a model for insect speciation in tropical rainforests."* *at the ITZ, Division of Ecology and Evolution, University of Veterinary Medicine Hannover, Foundation, Germany*

Field research: *Collection of ecological parameters and tissue samples from larvae and adult* M. caerulatus *in different National parks and biological research stations in Panama, Costa Rica, Nicaragua, Honduras, Guatemala, Mexico (12/11 – 05/12 and 11/12 – 12/12)*
Research stays: *Sackler Institute for Comparative Genomics, American Museum of Natural History (AMNH), New York, USA (09/13 – 04/14 und 06 – 12/15); Royal Belgian Institute of Natural Sciences, Brussels, Belgium (09/16)*
Teaching assistant: *Supervision in Laboratory courses "Introduction to molecular genetics methods for ecology and evolutionary biology" (2011 – 2017), "Species conservation and environmental policy" (Crau Camargue: 2012, 2013)*
"Invertebrate zoology" (AMNH; 04/14), Introduction to phylogeny by the example of Heliothis moths (AMNH; 06 – 09/15); Co-Supervision of different individual projects, Bachelor and Master theses on European and tropical dragonflies and damselflies

10/07 – 12/10   University of Veterinary Medicine Hannover, Foundation

Master of Sciences (M.Sc.) in Animal Biological and Biomedical Sciences
Title: *"Gene expression studies and population genetics in the tree hole breeding damselfly* Megaloprepus caerulatus *(Odonata: Pseudostigmatidae) as an*

environmental monitor." at the ITZ, Division of Ecology and Evolution, University of Veterinary Medicine Hannover, Foundation, Germany

Exchange program: *Instituto Internacional en Conservación y Manejo de Vida Silvestre (ICOMVIS), Universidad Nacional, Heredia, Costa Rica (08/08 – 07/09) as DAAD Scholarship holder (Internationale Studien- und Ausbildungspartnerschaften)*

10/04 – 08/07 Leibniz University Hannover

Bachelor of Sciences (B.Sc.) in Biologie
Title: *"Eutrophierung von Waldgebieten im Nationalpark De Hoge Veluwe, Niederlande" at the Institute of Geobotany, Leibniz University Hannover, Germany*

EDUCATION / NON-SCIENTIFIC ACTIVITIES
10/99 – 09/02 Krankenhaus Region Hannover, GmbH, Germany

Professional training as Certified Nurse, intensive Medicine

Work experience: *Krankenhaus Region Hannover GmbH*, Hannover, Germany
06/14 – 03/15	part time work: intensive-care unit
07/05 – 12/11	part time work: intensive-care unit
10/02 – 06/05	full time work: intensive-care unit

ATTENDED WORKSHOPS
08/18	*QGIS* workshop, Graduate School of Natural Sciences, Leibniz University Hannover, Germany
10/16 – 07/17	*Promotion plus+ qualifiziert* (Ph.D. and beyond), Management Competencies for Non-University Careers, Graduate Academy, Leibniz University Hannover, Germany
11/16	*Team management*, Graduate Academy, Leibniz University Hannover, Germany
08/16	*Species distribution modelling under climate change*, Faculty of Science, Copenhagen University and Natural History Museum of Denmark, Center for Macroecology, Evolution, and Climate, Copenhagen
04/16 – 07/16	*Promotion Plus Coaching* Life/Work Planning, Graduate Academy, Leibniz University Hannover, Germany
09/13 – 01/14	Bioinformatics, American Museum of Natural History, New York
01/13	Workshop on Molecular Evolution, Evolution and Genomics, Intensive and comprehensive training workshops, Ceský Krumlov, Czech Republic

GRANTS AND FELLOWSHIPS
10/16 – 06/17	Participation scholarship for the course Promotion plus qualifiziert (PhD and beyond), given by the Dr. Heinz Lindemann Stiftung
08/16 – 09/16	Participation scholarship for the course "Modelling species distributions under climate change" given by the Faculty of Science, Copenhagen University and the Natural History Museum of Denmark, Center for Macroecology, Evolution, and Climate, Copenhagen
06/16	Travel grand for the European Congress of Odonatology in Tyringe, Sweden (ECOO16) given by the Graduate Academy, Leibniz University Hannover, Germany

11/16 – 07/16	Otto Bütschli Scholarship given by the ITZ, Division of Ecology and Evolution of the University of Veterinary Medicine Hannover, Foundation
06/15	Travel grand for the International Congress of Odonatology (ICO 2015), La Plata, Argentina given by the Graduate Academy, Leibniz University Hannover, Germany
06/14 – 10/15	Annette Kade Graduate Fellowship, Student Fellowship Program to support exchange of graduate students given by the Richard Gilder Graduate School of the American Museum of Natural History: "RNA-Seq goes into detail! Comparative larval transcriptome profiling to identify genes involved in speciation within the damselfly genus *Megaloprepus*."
12/13 – 04/14	Scholarship for a research stay abroad at the American Museum of Natural History given by the Graduate Academy, Leibniz University Hannover, Germany
09/13 – 11/13	Annette Kade Graduate Fellowship, Student Fellowship Program to support exchange of graduate students given by the Richard Gilder Graduate School of the American Museum of Natural History: "The giant damselfly family Pseudostigmatidae (Zygoptera): An integrative and comparative approach to study speciation dynamics in the Neotropical rainforests"
06/13	Travel grand for the International Congress of Odonatology (ICO 2013), Freising, Germany given by the Society of Friends of the University of Veterinary Medicine Hannover, Foundation
12/11 – 12/12	Scholarship for material resources for field work in Central America "Speciation and adaptation in fragmented Neotropical rainforests using the world's largest damselfly *Megaloprepus* caerulatus (Odonata: Zygoptera, Pseudostigmatidae) as an example"
03/12 – 06/12	Otto Bütschli Scholarship given by the ITZ, Division of Ecology and Evolution of the University of Veterinary Medicine Hannover, Foundation
12/11 – 02/12	DAAD Scholarship for field work in Central America "Speciation and adaptation in fragmented Neotropical rainforests using the world's largest damselfly *Megaloprepus* caerulatus (Odonata: Zygoptera, Pseudostigmatidae) as an example"
04/11 – 11/11	Otto Bütschli Scholarship given by the ITZ, Division of Ecology and Evolution of the University of Veterinary Medicine Hannover, Foundation
08/08 – 07/09	DAAD scholarship for an exchange year in Costa Rica *(ISAP, Internationale Studien- und Ausbildungspartnerschaften)*

LIST OF PUBLICATIONS & PRESENTATIONS

ORIGINAL PUBLICATIONS

SJ Oppenheim, PZ Goldstein, **W Feindt**, B Murray, and R DeSalle "Reducing the streetlight effect in comparative analyses of lepidopteran diet breadth." (*submitted to PloS Genetics*)

W Feindt and H Hadrys "A legend turns back to history: The legendary monotypic damselfly genus *Megaloprepus* (Odonata: Zygoptera; Pseudostigmatidae) no longer consists of a single species!" (*submitted to Zootaxa*)

W Feindt, SJ Oppenheim, R DeSalle and S Mehr "GoodCitizenship.com: A step-by-step guide to submitting RNA-Seq data to NCBI" (*accepted by Current Protocols in Bioinformatics*)

W Feindt, SJ Oppenheim, R DeSalle, PZ Goldstein and H Hadrys (2018) Transcriptome profiling with focus on potential key genes for wing development and evolution in *Megaloprepus caerulatus*, the damselfly species with the world's largest wings. *PloS one*. 13 (1), e0189898. doi: 10.1371/journal.pone.0189898.

SJ Oppenheim, **W Feindt**, R DeSalle and PZ Goldstein (2018) De Novo characterization of transcriptomes from two North American *Papaipema* stem-borers (Lepidoptera: Noctuidae). *PloS one*, 13 (1), e0191061. doi: 10.1371/journal.pone.0191061.

W Feindt, HJ Osigus, R Herzog, CE Mason, B Schierwater and H Hadrys (2016) The complete mitochondrial genome of the Neotropical helicopter damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera) assembled from next generation sequencing data. *Mitochondrial DNA Part B*. 1(1), 497-499. doi: 10.1080/23802359.2016.1192504.

W Feindt, R Herzog, HJ Osigus, B Schierwater and H Hadrys (2016) Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera). *Mitochondrial DNA Part B*. 1(1), 574-576. doi: 10.1080/23802359.2016.1192510.

R Herzog, HJ Osigus, **W Feindt**, B Schierwater and H Hadrys (2016) The complete mitochondrial genome of the emperor dragonfly *Anax imperator* LEACH, 1815 (Aeshnidae: Odonata) via NGS sequencing. *Mitochondrial DNA Part B*. 1 (1), 783-786. doi: 10.1080/23802359.2016.1186523.

Feindt W, Fincke O, Hadrys H (2015) Still a one species genus? Strong genetic diversification in the world's largest living odonate, the Neotropical damselfly *Megaloprepus caerulatus*. *Conservation Genetics*. 15(2), 469-481. doi: 10.1007/s10592-013-0554-z

PUBLICATIONS IN PREPARATION

W Feindt, R DeSalle and H Hadrys "Cryptic non-adaptive speciation with low rates of phenotypic evolution in a Neotropical rainforest odonate genus" (*prepared for Molecular Ecology*)

W Feindt, SJ Oppenheim, R DeSalle and H Hadrys "Comparative transcriptomics among two *Megaloprepus* species."

PRESENTATIONS

003/18	Annual meeting of Dutch Odonatologists, Utrecht, Netherlands: "The genetics behind the northward expansion of the Scarlet Darter *Crocothemis erythraea*, Odonata: Libellulidae (Brullé, 1832)."
07/16	4th European Congress on Odonatology (ECOO16) Tyringe, Sweden: "The genetics behind the northward expansion of the Scarlet Darter *Crocothemis erythraea*, Odonata: Libellulidae (Brullé, 1832)."
03/16	Department of Biology, University of Lund, Sweden: "Odonate speciation in the Neotropics: New insights into the genus *Megaloprepus*."
02/16	European Conference of Tropical Ecology (GTÖ Jahreskonferenz 2016), Göttingen, Germany: "Four in one: cryptic species in the genus *Megaloprepus* (Odonata: Zygoptera)."
11/15	International Congress of Odonatology (ICO 2015), La Plata, Argentina: "The first insight into comparative transcriptome analysis of the Neotropical damselfly genus *Megaloprepus*."; (poster presentation): "Four in one: cryptic species in the genus *Megaloprepus* (Odonata: Zygoptera)."
09/14	107. Annual meeting of the German Zoological Society (Jahrestagung der Deutschen Zoologischen Gesellschaft), Göttingen: "Integrative taxonomy favors divergent evolution in the largest living Odonate species; the Neotropical damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera)"
06/13	International Congress of Odonatology (ICO 2013), Freising, Germany: Radiation despite niche conservatism in the Neotropical damselfly genus *Megaloprepus*?
02/11	European Conference of Tropical Ecology (GTÖ Jahreskonferenz 2011), Frankfurt, Germany (poster presentation): "Speciation in the Neotropical giant damselfly *Megaloprepus caerulatus* reflects forest fragmentation (Pseudostigmatidae: Odonata)."
07/10	"Münster Meeting on Stress and Evolution", Münster, Germany (poster presentation): Physiological response to temperature stress – Gene-expression studies in the Neotropical tree-hole breeding damselfly *Megaloprepus caerulatus*.