

Genomdiversität der dominierenden
Pseudomonas aeruginosa Klone C und PA14

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

genehmigte Dissertation

von

M. Sc. **Sebastian Fischer**

geboren am 05.11.1986 in Frankenberg

2015

Referent:

Korreferent:

Korreferent:

Datum der Promotion:

Prof. Dr. Dr. Burkhard Tümmler

Prof. Dr. Sebastian Suerbaum

Prof. Dr. Alexander Goesmann

13.11.2015

Danksagung

Mein Dank gilt an erster Stelle Herrn Prof. Dr. Dr. Burkhard Tümmler für die Bereitstellung dieses höchst interessanten Themas, sein fortwährendes Interesse und seine engagierte Unterstützung bei der Durchführung dieser Arbeit.

Den Herren Prof. Dr. Sebastian Suerbaum und Prof. Dr. Alexander Goesmann danke ich für die Übernahme des Korreferats und Prof. Dr. Thomas Scheper für die Übernahme des Prüfungsvorsitzes.

Weiterhin danke ich allen derzeitigen und ehemaligen Mitgliedern der Klinischen Forschergruppe für die erfolgreiche Zusammenarbeit, Hilfsbereitschaft und vielen schönen Stunden während und nach der Arbeit.

Dr. Jens Klockgether, Dr. Nina Cramer und Dr. Sarah Dethlefsen danke ich für die tolle Betreuung und ständige Hilfsbereitschaft bei sämtlichen auftretenden Problemen.

Sabrina Woltemate und Prof. Sebastian Suerbaum danke ich für die Durchführung der 454-Sequenzierung.

Rolf Hilker danke ich für die initiale Assemblierung des NN2 Genoms.

Klaas Hagemann danke ich für die ständige Erreichbarkeit bei allen wichtigen und unwichtigen Kleinigkeiten.

Ein besonderer Dank gilt meiner Familie und meinen Freunden für ihre Unterstützung und Kameradschaft in den letzten Monaten und Jahren.

Und vor allem möchte ich mich bei meiner Freundin Christiane für ihre Geduld bedanken und dafür, dass sie mir immer zur Seite steht.

Vorbemerkung

Die vorliegende Arbeit wurde im Zeitraum September 2011 bis Juni 2015 in der Naturwissenschaftlichen Fakultät der Leibniz Universität Hannover unter der Leitung von Prof. Dr. Dr. Burkhard Tümmler angefertigt.

Teile dieser Dissertation wurden als Posterbeiträge auf der Pseudomonas-Konferenz 2013, der Summer School on Infection Research 2014, der VAAM 2015 und in den folgenden Publikationen vorgestellt.

1.

Hilker R, Munder A, Klockgether J, Losada PM, Chouvarine P, Cramer N, Davenport CF, Dethlefsen S, **Fischer S**, Peng H, Schönfelder T, Türk O, Wiehlmann L, Wölbeling F, Gulbins E, Goesmann A, Tümmler B.

Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome from disease and environment.

Environ Microbiol. 2015 Jan;17(1):29-46. doi: 10.1111/1462-2920.12606. Epub 2014 Sep 29. PubMed PMID: 25156090.

2.

Sebastian Fischer*, Nina Cramer*, Jens Klockgether, Patricia Moran Losada, Sarah Dethlefsen, Colin Davenport, Marie Dorda, Alexander Goesmann, Rolf Hilker, Samira Mielke, Torben Schönfelder, Oliver Türk, Lutz Wiehlmann, Philippe Chouvarine and Burkhard Tümmler

Genome diversity of the major *Pseudomonas aeruginosa* clones C and PA14.

Eingereicht.

Zusammenfassung

Pseudomonas aeruginosa ist ein gramnegatives γ -Proteobakterium, welches sowohl in der Umwelt als auch als Humanpathogen nachgewiesen werden konnte. Vertreter des häufigsten klonalen Komplexes C wurden in früheren Arbeiten bereits charakterisiert und hinsichtlich genomischer Inseln untersucht, jedoch kein Vertreter vollständig sequenziert und annotiert. Ferner wurde die Genomdiversität von 20 unabhängigen Klonen untersucht, jedoch wurden bislang nicht Stämme innerhalb eines klonalen Komplexes verglichen. Da von den beiden häufigsten *P. aeruginosa* Klonen, Klon C und Klon PA14, einzig die Sequenz des Referenzstamms PA14 bekannt war, wurde innerhalb dieser Arbeit die Sequenz eines Isolates des häufigsten Klons C, NN2, entschlüsselt. Das mit 6,9 Mbp vergleichsweise große Genom (PA14: 6,5 Mbp, PAO1 6,2 Mbp) kodiert für mehr als 6.600 Gene und hat einen GC-Gehalt von 66,12%. Anhand einer RNA-Sequenzierung war es möglich, für 1.800 Gene die Transkriptionsstartpunkte zu ermitteln und zusätzlich 327 bislang unbekannte nicht-kodierende RNAs (ncRNA) zu identifizieren. Erstmals in *P. aeruginosa* konnten dabei neben intergenisch lokalisierten ncRNAs auch intragenische Vertreter beschrieben werden.

Ein Schwerpunkt der vorliegenden Arbeit war die Analyse von unabhängigen Isolaten der Klone C (57 Isolate) und PA14 (42), welche sowohl aus der Umwelt als auch aus Patientenproben isoliert wurden. Anhand von SOLiD-Sequenzierungen der einzelnen Stämme und anschließender Sequenzauswertung wurde die Genomdiversität der beiden klonalen Komplexe vergleichend untersucht. Vertreter beider Klone zeichneten sich durch eine hohe Konservierung des Kerngenoms mit einer durchschnittlichen Diversität von $8 \cdot 10^{-6}$ für Klon C und $2 \cdot 10^{-5}$ für Klon PA14 und einer variablen Zusammensetzung des akzessorischen Genoms aus. Das Verhältnis von synonymen zu nicht-synonymen Austauschen korrelierte mit dem Logarithmus der SNPs und näherte sich dem Wert vier an, welcher für den Genomvergleich von unverwandten *P. aeruginosa* Klonen beobachtet wird. Die Diversität des akzessorischen Genoms war zwischen Vertretern eines Klons genauso variabel wie zwischen unverwandten Stämmen.

Intraklonale Fitnessunterschiede wurden in kompetitiven Wachstumsexperimenten *in vitro* untersucht. Die Stämme mit der vergleichsweise höchsten Fitness waren entweder stammspezifisch mit Phagen, Plasmiden und zusätzlichen Genen ausgestattet oder hatten die zentralen Schaltergene inaktiviert, die alternativ ein Leben im Biofilm oder einen planktonischen Lebensstil mit der Freisetzung von Virulenzfaktoren festlegen.

Schlüsselwörter: *Pseudomonas aeruginosa*, Genomdiversität, Sequenzierung, intraklonale Fitness

Abstract

Pseudomonas aeruginosa is a Gram-negative γ -proteobacterium that thrives in aquatic inanimate habitats and on mucosal surfaces of animate hosts. The cosmopolitan clones C and PA14 are the dominant complexes in the *P. aeruginosa* population. An isolate from cystic fibrosis human lungs, termed NN2, was completely sequenced within this study to have in addition to the existing strain PA14 genome reference sequences at hand for a comprehensive investigation of the intraclonal genome diversity of the two major *P. aeruginosa* clonal complexes. The NN2 genome sequence was resolved to 6.9 Mbp with a GC-content of 66.12% (PA14: 6.5 Mbp; PAO1: 6.2 Mbp), and more than 6,600 genes were annotated within the genome. Using RNA-seq 1,800 transcriptional start sites and 327 yet undescribed ncRNAs were identified the majority of which being located within coding regions.

Next, 57 and 42 independent clone C or clone PA14 isolates from environmental and clinical habitats were sequenced with the accurate SOLiD technology and subsequently analyzed in their genetic repertoire. The core genome of each clone turned out to be highly conserved with an average sequence diversity of $8 \cdot 10^{-6}$ for clone C and $2 \cdot 10^{-5}$ for clone PA14. The ratio of synonymous to non-synonymous nucleotide substitutions compared to reference genomes NN2 and PA14 correlated with the logarithm of SNPs and converged to a maximal value of four which is characteristic for the comparison of independent clones. In contrast to the highly conserved core genome the constituents of the accessory genomes were nearly as variable within clones C or PA14 than between unrelated clones.

Intraclonal fitness diversity was assessed in comparative growth experiments *in vitro*. The strains with the comparably highest fitness were endowed with a strain-specific repertoire of weapons like phages or plasmids or extra genes for transport or metabolism and/or had inactivated key genes that govern the lifestyle of *P. aeruginosa* as either a sessile microbe in biofilms or a planctonic virulent bacterium.

key words: *Pseudomonas aeruginosa*, genome diversity, sequencing, intraclonal fitness

Inhaltsverzeichnis

| | |
|---|------------|
| Inhaltsverzeichnis | I |
| Abbildungsverzeichnis | IV |
| Tabellenverzeichnis | VI |
| Abkürzungsverzeichnis | VII |
| 1 Einleitung | 1 |
| 1.1 Bakterielle Populationsbiologie | 1 |
| 1.2 <i>Pseudomonas aeruginosa</i> | 3 |
| 1.3 Populationsstruktur von <i>P. aeruginosa</i> | 6 |
| 1.3.1 Das Kern- und das akzessorische Genom | 7 |
| 1.4 <i>Pseudomonas aeruginosa</i> Klon C | 7 |
| 1.5 Ziele der Arbeit | 10 |
| 2 Material und Methoden | 11 |
| 2.1 Stämme und Wachstumsmedien | 11 |
| 2.2 DNA-Präparation | 12 |
| 2.3 Genotypisierung | 12 |
| 2.4 DNA-Sequenzierungen | 12 |
| 2.4.1 Sequenzierungen des NN2 Genoms | 12 |
| 2.4.2 Sequenzierung der 100 Isolate | 15 |
| 2.5 <i>De novo</i> Assemblierung des NN2 Genoms | 17 |
| 2.6 Lokalisation der NN2 Contigs innerhalb der PAO1 Referenzsequenz | 18 |
| 2.7 Schließen des NN2 Genoms | 18 |
| 2.8 Annotation des NN2 Genoms | 19 |
| 2.9 RNA-Sequenzierung | 20 |
| 2.9.1 Kultivierung der Bakterien | 20 |
| 2.9.2 RNA-Isolierung und Depletion ribosomaler RNA | 21 |
| 2.9.3 Durchführung der RNA-Sequenzierung | 22 |

| | | |
|----------|--|-----------|
| 2.10 | Identifikation von Transkriptionsstartpunkten und kurzen nicht-kodierenden RNA-Molekülen | 23 |
| 2.11 | Differentielle Genexpressionsanalyse | 24 |
| 2.12 | Erstellen des Genomatlas für NN2 | 24 |
| 2.13 | Struktur nicht-kodierender RNAs (ncRNAs) | 25 |
| 2.14 | Verteilung von Nukleotidtetrameren im NN2 Genom | 25 |
| 2.15 | Vorhersage von Einzelnukleotidaustauschen | 25 |
| 2.16 | Gruppierung von Aminosäureaustauschen | 26 |
| 2.17 | Bestimmung von Haplotypen | 26 |
| 2.18 | Phylogenetische Bäume | 27 |
| 2.19 | Visuelle Darstellung der Zusammensetzung des akzessorischen Genoms (Heatmaps) | 27 |
| 2.20 | Konservierung des akzessorischen Genoms | 28 |
| 2.21 | Bestimmung von Deletionen größer als 1 kb | 28 |
| 2.22 | Analyse kleiner Insertionen und Deletionen (Indels) | 29 |
| 2.23 | Identifizierung stammspezifischer Gene | 30 |
| 2.24 | Kompetitive Fitnessexperimente | 31 |
| 2.24.1 | Messung der Pyocinproduktion | 31 |
| 2.24.2 | Kompetitive Wachstumsexperimente | 32 |
| 2.24.3 | Bestimmung der Anteile der einzelnen Isolate an der jeweiligen Kompetition | 32 |
| 3 | Ergebnisse und Diskussion | 33 |
| 3.1 | Die Entschlüsselung des Klon C NN2 Genoms | 33 |
| 3.1.1 | Sequenzierung und <i>de novo</i> Assemblierung des NN2 Genoms | 34 |
| 3.1.2 | Sortieren der Contigs zu einer konsekutiven Reihenfolge | 37 |
| 3.1.3 | Schließen des NN2 Genoms | 38 |
| 3.1.4 | Annotation des NN2 Genoms | 41 |
| 3.1.5 | Bestimmung von Transkriptionsstartpunkten und nicht- kodierenden RNAs (ncRNAs) | 43 |
| 3.1.6 | Genomkarte des Isolats NN2 | 46 |
| 3.1.7 | Differentielle Genexpressionsanalyse des Isolats NN2 | 48 |
| 3.1.8 | Tetranukleotidzusammensetzung des NN2 Genoms | 49 |
| 3.1.9 | Diskussion der Ergebnisse zur Entschlüsselung des NN2 Genoms | 50 |
| 3.2 | Intraklonaler Genomvergleich der dominanten Klone C und PA14 | 53 |
| 3.2.1 | Erhalt des chromosomalen Grundgerüsts | 55 |
| 3.2.2 | Intraklonale Genomdiversität des Kern- und des akzessorischen Ge- nomms | 58 |

| | | |
|----------|---|------------|
| 3.2.3 | Natürliche Selektion von Einzelnukleotidaustauschen innerhalb klonaler Komplexe | 69 |
| 3.2.4 | Intraklonale Bereiche hoher Mutationsfrequenz | 74 |
| 3.2.5 | Sequenzdiversität der nicht-kodierenden RNAs | 75 |
| 3.2.6 | Insertionen, Deletionen und Leserasterverschiebungen | 76 |
| 3.2.7 | Identifikation von stammspezifischen Genen | 78 |
| 3.2.8 | Diskussion der Ergebnisse des intraklonalen Genomvergleiches | 83 |
| 3.3 | Kompetitive Wachstumsexperimente | 89 |
| 3.3.1 | Diskussion der kompetitiven Wachstumsexperimente | 95 |
| 4 | Schlussfolgerungen und Ausblick | 98 |
| | Literaturverzeichnis | 101 |
| | Lebenslauf | 113 |
| | Anhang | 114 |

Abbildungsverzeichnis

| | | |
|------|--|----|
| 1.1 | Populationsarten | 2 |
| 1.2 | <i>Pseudomonas aeruginosa</i> | 4 |
| 1.3 | Populationsstruktur von <i>P. aeruginosa</i> | 6 |
| 1.4 | Physikalische Karte Klon C | 8 |
| 2.1 | Arbeitsablauf der 454-Sequenzierung | 13 |
| 2.2 | Arbeitsablauf der Illumina-Sequenzierung. | 14 |
| 2.3 | Darstellung paired-end und mate pair Reads | 15 |
| 2.4 | Arbeitsablauf der SOLiD-Sequenzierung | 16 |
| 2.5 | Wachstumskurve im Fermenter | 21 |
| 2.6 | Anreicherung von primären Transkripten | 22 |
| 2.7 | Ablauf einer strangspezifischen Sequenzierung | 23 |
| 3.1 | Flussdiagramm NN2 Genomik | 34 |
| 3.2 | Contigs und Scaffolds | 35 |
| 3.3 | Position der 454 Contigs im Referenzgenom PAO1 | 36 |
| 3.4 | Beispiel für <i>in silico</i> Contigzusammenführung | 39 |
| 3.5 | Schema zum Schließen der letzten Lücke | 40 |
| 3.6 | Beispiele 5'-untranslatierter Regionen in NN2 | 43 |
| 3.7 | Längenvergleich 5'-untranslatierter Regionen zwischen NN2 und PA14 | 44 |
| 3.8 | Beispiele detektierter ncRNAs | 45 |
| 3.9 | Längenverteilung intragenischer ncRNAs | 46 |
| 3.10 | Genomatlas des NN2 Genoms | 47 |
| 3.11 | Tetranukleotidzusammensetzung von NN2 | 49 |
| 3.12 | Längenverteilung der Fragmentlängenpaare | 56 |
| 3.13 | Phylogenetische Bäume der klonalen Komplexe | 57 |
| 3.14 | Heatmaps genomischer Inseln | 66 |
| 3.15 | Heatmaps RGPs | 67 |
| 3.16 | Ähnlichkeit des akzessorischen Genoms | 68 |
| 3.17 | Verhältnis $\frac{s}{n}$ bei steigender SNP-Anzahl | 70 |
| 3.18 | Besetzung der Dayhoffklassen | 71 |
| 3.19 | SNP Verteilung im Referenzgenom | 74 |

| | |
|--|----|
| 3.20 SNPs in ncRNAs | 76 |
| 3.21 Beispiel einer großen Deletion | 76 |
| 3.22 Deletionen im Kerngenom | 77 |
| 3.23 Verteilung von Indels | 78 |
| 3.24 Flussdiagramm Fitnessexperimente | 89 |
| 3.25 Auswertung der Fitnessexperimente | 91 |

Tabellenverzeichnis

| | | |
|------|--|----|
| 3.1 | Sequenzierungsstatistiken | 35 |
| 3.2 | Vergleich Scaffolding der 3 kb und 300 bp Bibliotheken | 37 |
| 3.3 | Übersicht über die manuell vorhergesagte Länge der Lücken zwischen den einzelnen Contigs. | 38 |
| 3.4 | Übersicht über die geschlossenen Lücken | 40 |
| 3.5 | Leserasterverschiebungen in NN2 gegenüber dem Referenzgenom PAO1 . . | 42 |
| 3.6 | Eckdaten des NN2 Genoms. | 46 |
| 3.7 | Übersicht der verwendeten Stämme | 53 |
| 3.8 | SNPs pro Klon PA14 Isolat | 59 |
| 3.9 | SNPs pro Klon C Isolat | 60 |
| 3.10 | Allgemeine Übersicht zu SNPs und Indels | 63 |
| 3.11 | Übersicht zu SNPs und Indels des klonalen Komplexes C | 64 |
| 3.12 | Übersicht zu SNPs und Indels des klonalen Komplexes PA14 | 65 |
| 3.13 | Vergleich über die Besetzung der Dayhoffklassen | 72 |
| 3.14 | Häufigkeit der Aminosäureaustausche in funktionellen Kategorien | 73 |
| 3.15 | Stammspezifische Gene zugeordnet zu funktionellen Annotationsklassen . . | 79 |
| 3.16 | Stammspezifische Gene Klon PA14 | 81 |
| 3.17 | Stammspezifische Gene Klon C | 82 |
| 3.18 | Übersicht Pyocinloci | 90 |
| 3.19 | Fitnessergebnisse PA14 | 92 |
| 3.20 | Fitnessergebnisse Klon C | 94 |

Abkürzungsverzeichnis

| | | | |
|---------------|----------------------------|-----------------------------|--|
| 5'-UTR | 5'-untranslatierte Region | ml | Milliliter |
| bp | Basenpaare | mM | Millimolar |
| bzw. | beziehungsweise | min | Minuten |
| ca. | zirka | ncRNA | nicht-kodierende RNA |
| CF | zystische Fibrose | nm | Nanometer |
| CFU | koloniebildende Einheit | OD | optische Dichte |
| °C | Grad Celsius | ORF | offenes Leseraster |
| dATP | Desoxyadenosintriphosphat | <i>P. aeruginosa</i> | <i>Pseudomonas aeruginosa</i> |
| dCTP | Desoxycytidintriphosphat | PCR | Polymerase-Kettenreaktion |
| dGTP | Desoxyguanosintriphosphat | % | Prozent |
| dNTP | Nukleosidtriphosphate | RGP | Region mit Genomplastizität |
| dTTP | Desoxythymidintriphosphat | RNA | Ribonukleinsäure |
| DMSO | Dimethylsulfoxid | rRNA | ribosomale RNA |
| DNA | Desoxyribonukleinsäure | RIN | RNA-Integritätsnummer |
| ggf. | gegebenenfalls | rpm | Umdrehungen pro Minute |
| g/l | Gramm pro Liter | TE-Puffer | Tris-EDTA-Puffer |
| h | Stunde | TEX | Terminator 5'-Phosphat abhängige Exonuklease |
| Indel | Insertionen und Deletionen | tRNA | Transfer-RNA |
| kb | Kilobasenpaare | TSA | Casein-Soja-Pepton-Agar („Tryptic Soy Agar“) |
| LT | Life Technologies | TSB | Casein-Soja-Pepton-Broth („Tryptic Soy Broth“) |
| l/min | Liter pro Minute | v/v | Volumenprozent |
| LB | lysogeny broth | z.B. | zum Beispiel |
| Mbp | Megabasenpaare | | |
| µl | Mikroliter | | |
| µg | Mikrogramm | | |

Kapitel 1

Einleitung

1.1 Bakterielle Populationsbiologie

Die Populationsgenetik untersucht die genetische Variabilität biologischer Populationen. Dabei unterscheidet man zwischen diploiden Spezies, welche sich sexuell fortpflanzen, und haploiden Spezies, welche sich asexuell durch Zellteilung vermehren. Die genetische Diversität in diploiden Organismen ist eher klein, da die vorherrschende Rekombination zu einer Homogenisierung des Genpools führt. Hingegen wird in asexuellen Spezies die genetische Diversität sowohl durch Rekombination als auch durch *de novo* Mutationen vermittelt. Ein gebräuchliches Mittel zur Bestimmung der Rekombinationsfrequenz in Bakterien stellt die Rekombinationsrate dar, welche das Verhältnis von Rekombinationen zu *de novo* Mutationen innerhalb einer Spezies angibt (Milkman und Bridges, 1990). Diese Rekombinationsrate ist charakteristisch für einen Organismus und bestimmt die Populationsstruktur innerhalb einer bakteriellen Spezies. Sie stellt somit einen wichtigen evolutionären Faktor dar, kann jedoch in Unterarten und sogar zwischen separaten Subpopulationen schwanken. Bakterien weisen zumeist mittlere Rekombinationsraten auf, welche für die Evolution von Populationen zwar signifikant sind, jedoch das Auftreten klonaler Linien nicht ausschließen (Spratt und Maiden, 1999).

Man kann zwischen drei verschiedenen bakteriellen Populationsarten unterscheiden: klonal, epidemisch und panmiktisch (Abb. 1.1). Klonale Populationen weisen nur einen geringen genetischen Austausch zwischen den einzelnen Individuen aus. Aufgrund der geringen Rekombinationsrate zeigen typische Vertreter wie *Staphylococcus aureus* (Cohan, 2004) ein Kopplungsungleichgewicht zwischen den Allelen. Im Gegensatz dazu zeigen panmiktische Populationen eine hohe Rekombinationsrate mit einer randomisierten Assoziation von Allelen zwischen den einzelnen Loci. Hierbei herrscht ein Kopplungsgleichgewicht, welches jedoch klonale Gruppen nicht ausschließt. Panmiktische Populationen findet man unter anderem bei *Porphyromonas gingivalis* oder *Helicobacter pylori* (Köhler et al., 2003; Spratt und Maiden, 1999; Smith et al., 1993). Epidemische Populationen tragen Charak-

teristika der beiden Extreme. Sie haben einen panmiktischen Ursprung, weisen jedoch eine evolutionär junge Zunahme eines bzw. weniger Genotypen auf. Eine epidemische Populationsstruktur wurde beispielsweise bei *Pseudomonas aeruginosa* beobachtet (Pirnay et al., 2009).

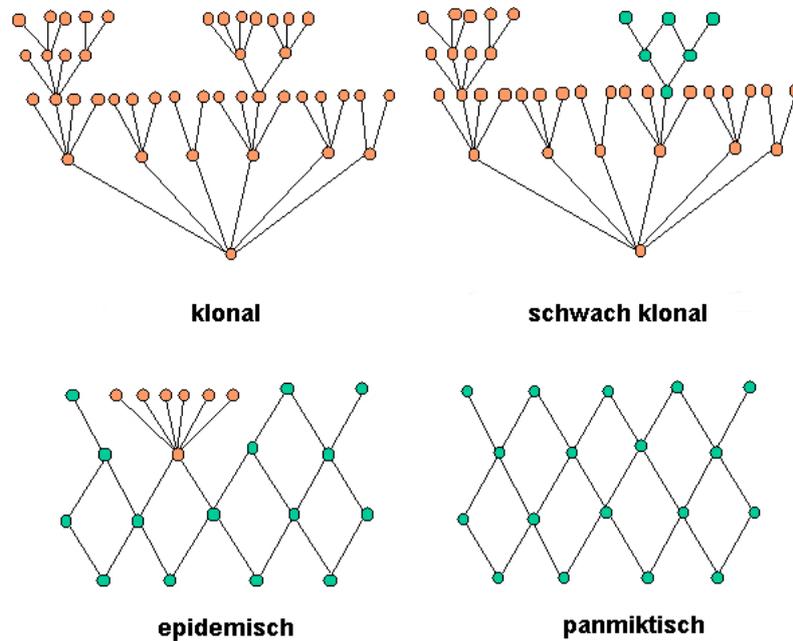


Abbildung 1.1: Schematische Darstellung der möglichen bakteriellen Populationsstrukturen (adaptiert von <http://www.ub.edu/microfar/webcatala/grups/angl/inici32.htm>).

Dabei stellt ein Klon eine Gruppe genetisch ähnlicher Zellen dar, welche rekombinationslos von einem gemeinsamen Vorfahren abstammen (Maynard Smith, 1995) und dabei in ausgewählten Genommarkern übereinstimmen. Die Zuordnung zu einem Klon kann zum Beispiel durch Multilokus-Sequenz-Typisierung (MLST) erfolgen. Dabei werden PCR-Fragmente von häufig sieben Housekeeping Genen amplifiziert, sequenziert und verglichen.

Im Forschungsfeld von *P. aeruginosa* wurde ein spezieller Mikroarray entwickelt, welcher die bekannten Variationen von 13 in der Population etablierten Einzelnukleotidaustauschen und zwei multiallelischen Loci (*fliC*, *fpvA*) des Kerngenoms sowie 38 Marker des akzessorischen Genoms repräsentiert (Wiehlmann et al., 2007). Zu einem klonalen Komplex in *P. aeruginosa* gehören nach diesem Test all jene Bakterien, welche in den Varianten des Kerngenoms übereinstimmen und somit einen gemeinsamen Genotyp des Kerngenoms aufweisen, sich jedoch in der Zusammensetzung des variablen akzessorischen Genoms unterscheiden.

Bakterien sind in der Lage, kleine und große DNA-Fragmente aus der Umgebung oder von anderen Mikroorganismen aufzunehmen und in ihr Genom zu integrieren. Als homologe Rekombination bezeichnet man dabei den Austausch zweier DNA-Abschnitte im Ganzen. Dabei werden bevorzugt Bereiche ähnlicher Sequenz ausgetauscht. Es wurden jedoch auch

schon Rekombinationen zwischen Stücken mit mehr als 20% Sequenzdiversität beobachtet (Feil und Spratt, 2001).

Innerhalb der Bakterien sind drei Arten des Genaustauschs beschrieben: Transduktion, Konjugation und Transformation. Bei der transduktionellen DNA-Übertragung werden Bakteriophagen als Vektor genutzt, deren DNA kann nach erfolgter Infektion der Wirtszelle durch homologe Rekombination in das Empfängergenom integriert werden kann. Bei der Konjugation wird die DNA über Plasmide oder Transposons direkt von einer zur anderen Zelle transferiert. Die Donor-DNA wird dabei in ein Plasmid integriert, welches in das Bakterienchromosom eingebaut werden kann. Nach erfolgter Wiederfreisetzung wird das Plasmid anschließend über einen Pilus in eine andere Zelle übertragen. Bei der Transformation wird freie DNA von den Bakterien aktiv aus der Umgebung aufgenommen und durch homologe Rekombination in das Genom integriert. Jedoch zeigt nicht jedes Bakterium die Kompetenz, DNA mittels Transformation aufzunehmen (Madigan und Martinko, 2006).

1.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa ist ein gramnegatives, Oxidase-positives, stäbchenförmiges Gammaproteobacterium (Abb. 1.2). Das Bakterium ist ein ubiquitärer Boden- und Wasserkeim und hat sich auch als bedeutendes nosokomiales Pathogen etabliert (Ramos, 2004). Der Name des Bakteriums wird auf die Bildung bläulich-grüner wasserlöslicher Pigmente wie Pyocyanin und Pyoverdin zurückgeführt. Die Stäbchen werden 2 – 4 µm lang und weisen polytrich-monopolare Flagellen auf. Zusätzlich ermöglichen Haftfimbrien die Anheftung an Oberflächen. Das Exopolysaccharid Alginate wirkt als Kapsel auf der äußeren Zellmembran und schützt das Bakterium somit vor Phagozyten und Antikörpern.

P. aeruginosa bevorzugt aerobes Wachstum, kann jedoch in speziellen Lebensräumen wie zum Beispiel der Lunge von Patienten mit zystischer Fibrose (engl. cystic fibrosis, CF) auch mikroaerophil oder anaerob wachsen. In mikroaerophilen Habitaten liegt ein um wenige Prozent geringerer Sauerstoffanteil als in der Atemluft vor und die Zellatmung wird in diesen Fällen über terminale Oxidasen reguliert (Alvarez-Ortega und Harwood, 2007).

P. aeruginosa ist in der Lage, je nach Viskosität der Umgebung verschiedene Bewegungsformen anzuwenden. Dabei werden die Hauptbewegungsformen des Schwimmens in Flüssigmedium, des Schwärmens auf Geloberflächen und des Twitchens auf festen Strukturen durch Pili und/oder Flagellen vermittelt.

Virulenzfaktoren werden von *P. aeruginosa* hauptsächlich über die Sekretionssysteme des Typs III und VI sezerniert. In beiden Systemen sticht das Pathogen mit einer Art Nadelstruktur in die Wirtszelle und sezerniert direkt in das Zytosol Effektoren, welche zum Beispiel in den Zellzyklus eingreifen oder Apoptose induzieren (Büttner, 2012; Costa et al., 2015).



Abbildung 1.2: Elektronenmikroskopische Aufnahme von *P. aeruginosa* (Winsor et al., 2011).

Pseudomonas kann ein breites Spektrum an Substraten verwerten, ist allerdings aufgrund eines fehlenden Gens der Glykolyse nur über den Entner-Doudoroff-Weg in der Lage, auf Glukose zu wachsen (Berger et al., 2014). Stattdessen bevorzugt das Bakterium andere Kohlenstoffquellen wie Succinat oder Citrat. Die metabolische Versatilität von *P. aeruginosa* zeigt sich auch in den mit 6,0 - 7,2 Mbp großen Genomen (Stover et al., 2000; Hilker et al., 2015). *P. aeruginosa* ist aufgrund seiner sehr geringen Nährstoffansprüche ein klassischer Nasskeim und ist in der Lage, auch in entionisiertem Wasser mit sehr wenigen Kohlenstoffquellen zu wachsen (Favero et al., 1971). Ferner wurde *P. aeruginosa* sowohl aus verschiedensten Gewässern (Pellett et al., 1983) als auch von Pflanzenoberflächen isoliert (Ramos, 2004). Bemerkenswert ist, dass dieselben Virulenzmechanismen unabhängig vom Habitat genutzt werden (Costerton, 1980).

Immunkompetente Menschen sind für eine Pseudomonasinfektion in der Regel nicht anfällig, da das Bakterium die Immunabwehr nicht überwinden kann (Botzenhardt und Döring, 1993). Eine Ausnahme davon stellen Augeninfektionen dar. Eine solche Infektion kann auch bei gesunden Menschen eine starke Keratitis auslösen, welche unbehandelt zur Erblindung führen kann.

Im Gegensatz dazu befällt das opportunistische Pathogen hauptsächlich Personen mit einem geschwächten Immunsystem wie z.B. Träger des humanen Immundefizienz-Virus (HIV) oder Patienten, die nach Organspenden immunsuprimiert werden. Zusätzlich ist *P. aeruginosa* in der Lage Brandwunden zu infizieren. Einen Sonderfall stellen Patienten mit zystischer Fibrose dar. Deren Atemwege sind zwar immunkompetent, werden jedoch häufig von *P. aeruginosa* besiedelt, wobei sich eine chronische Infektion etabliert. In der CF-Lunge wird eine Infektion dadurch ermöglicht, dass der Abtransport des zähflüssigen Lungensekretes erschwert ist und ein geeignetes Nährmedium für Mikroorganismen

geschaffen wird. Dabei beginnt die Infektion stets mit der Initialbesiedlung der oberen Atemwege und breitet sich im weiteren Verlauf auch auf die unteren Atemwege aus (Döring et al., 2000). Die damit einhergehende Inflammation und die nachfolgende Destruktion des befallenen Lungengewebes stellen die Hauptgründe für die verminderte Lungenfunktion und eine verkürzte Lebenserwartung der betroffenen CF-Patienten dar (Govan und Deretic, 1996).

Infektionen mit *P. aeruginosa* sind oft nur schwer therapierbar, da das Bakterium viele Resistenzen gegen heutzutage gebräuchliche Antibiotika entwickelt hat. Da angenommen wird, dass eine Kolonisation mit *P. aeruginosa* von Umweltisolaten ausgeht (Gibson et al., 2003), sollte versucht werden, die Bakterien während der frühen Besiedlung zu eliminieren.

Die Antibiotikaresistenzen lassen sich durch Besonderheiten der Zellstruktur gramnegativer Bakterien erklären. So tragen diese über der Zellwand eine zusätzliche äußere Membran, welche eine Diffusionsbarriere für Antibiotika darstellt und nur durch spezielle Porine überwunden werden kann. Hier tritt eine Besonderheit von *P. aeruginosa* auf, dessen Porine besonders undurchlässig sind und einen zweiten Schutzmechanismus darstellen. Sollten Antibiotika doch ins Zelleninnere gelangen, verfügt *P. aeruginosa* über effektive Effluxsysteme, so genannte MDR-Transporter (engl. multidrug resistance) und kann auf diesem Weg Antibiotika aus den Zellen ausschleusen (De Kievit et al., 2001). Zusätzlich ist in einigen Fällen ein sogenanntes R-Plasmid beschrieben worden, welches Gene enthält, die für Komponenten des Antibiotikametabolismus oder -efflux kodieren (Madigan und Martinko, 2006).

Da sich *P. aeruginosa* im feuchten Milieu sehr gut vermehren kann, wozu auch Wasserleitungen, Luftbefeuchter, Beatmungsgeräte und Waschbecken zählen, stellt das Bakterium eine große Herausforderung für die Krankenhaushygiene dar. Prophylaktische Maßnahmen sollten somit im gesamten Klinikbereich einen hohen Stellenwert einnehmen und ständig überwacht werden. Erschwerend dazu ist *P. aeruginosa* in der Lage, Biofilme zu bilden (Botzenhardt und Döring, 1993). Als Biofilm bezeichnet man eine Schleimschicht auf Oberflächen, in der Mikroorganismen, aber auch Pilze und Algen leben. Ein Biofilm besteht hauptsächlich aus Wasser und extrazellulären polymeren Substanzen (EPS) sowie den darin lebenden Organismen. Im Gegensatz zu planktonischen Kulturen werden von den Bakterien im Biofilm die Flagellen abgeworfen und Gene des Oberflächenkontakts angeschaltet. Das genetische Repertoire der Mikroorganismen wird durch horizontalen Gentransfer mit Nachbarzellen moduliert, wodurch die Toleranz gegenüber extremen pH-Werten oder Antibiotika erhöht wird.

1.3 Populationsstruktur von *P. aeruginosa*

Die Populationsstruktur von *P. aeruginosa* ist als netzartig (Kiewitz und Tümmler, 2000) und epidemisch (Pirnay et al., 2009) beschrieben worden und wird sowohl durch horizontalen als auch vertikalen Gentransfer bestimmt. In der weltweiten Population finden sich mehrere dominante Klone wie der Klon PA14, benannt nach dem Referenzstamm PA14 (Rahme et al., 1995), einem epidemischen Klon der englischen Midlands LES (Cheng et al., 1996), der weit verbreitete Klon C und zusätzlich viele sporadische Klone (Römling et al., 2005; Wiehlmann et al., 2007). Die Mehrheit der klonalen Komplexe tritt jedoch nur selten auf (Wiehlmann et al., 2007; Selezska et al., 2012). Die 20 häufigsten Klone stellen nahezu 50% der gesamten Population dar und die sporadischen Klone die andere Hälfte. Die weite Verbreitung der einzelnen dominanten Klone sowohl in der geografischen Lokalisation auf verschiedenen Kontinenten, als auch in verschiedenen Isolationshabitaten (Umwelt, akute und chronische Infektion) verdeutlicht die hohe Versatilität der Spezies. Die bislang bekannten Klone segregieren zu drei nicht überlappenden Subgruppen (Abb. 1.3), wobei die beiden häufigsten Klone in unterschiedlichen Gruppen zu finden sind. Auf der anderen Seite wurden viele Klone noch nicht aus einem Patienten isoliert und zudem von vielen Klinikklonen noch kein entsprechendes Isolat in der Umwelt gefunden (Tümmler et al., 2014).

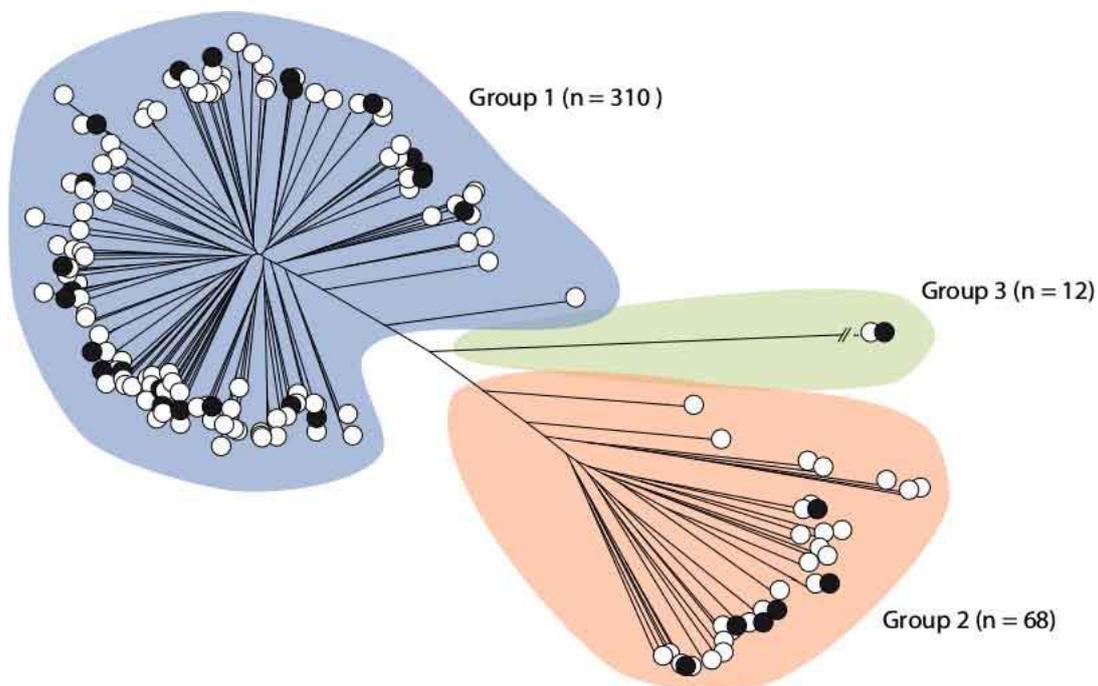


Abbildung 1.3: Phylogenetischer Baum von 390 *P. aeruginosa* Isolaten basierend auf SNPs innerhalb des Kerngenoms. Die Stämme wurden in drei Gruppen aufgespalten (blau, orange, grün) und die Anzahl der Stämme jeder Gruppe dargestellt. Schwarze Kreise symbolisieren bereits bekannte Genome, weiße Kreise symbolisieren in der zum Zeitpunkt der Niederschrift unpublizierte neu sequenzierte Genome. Mit freundlicher Genehmigung von R. Levesque (Universität Laval, Quebec, Kanada).

1.3.1 Das Kern- und das akzessorische Genom

Basierend auf Sequenzierungs- und Hybridisierungsexperimenten konnten zwei Bereiche des bakteriellen Genoms identifiziert werden: das konservierte Kerngenom (80% des jeweiligen Genoms) und das variable akzessorische Genom (20%). Das Kerngenom von *P. aeruginosa* beinhaltet mehr als 4.000 Gene (Hilker et al., 2015). Zusätzlich wiesen die in dieser Studie untersuchten Stämme in der Summe mehr als 10.000 stammspezifische Gene des akzessorischen Genoms auf. Das akzessorische Genom eines Stammes ist individuell aus Genominseln und mobilen Elementen wie Plasmiden, Phagen und Transposons zusammengesetzt. Diese mobilen Elementen integrieren gehäuft in bestimmte Genomregionen mit hoher Genomplastizität (engl. regions of genome plasticity, RGP) (Mathee et al., 2008; Kung et al., 2010). Dabei werden als Genominseln DNA-Fragmente bezeichnet, die nicht zum Kerngenom gehören, über horizontalen Gentransfer aufgenommen und in die RGP-Regionen integriert werden. Die Mehrzahl der genomischen Inseln integriert in spezifische Erkennungssequenzen (IS-Elemente) oder neben tRNA-Gene, welche an ihren Rändern häufig direkte DNA Wiederholungen (engl. direct repeats) tragen. Eine Subgruppe der genomischen Inseln wird als Pathogenitätsinseln bezeichnet, da diese für Virulenzfaktoren kodieren (Hacker und Kaper, 2002).

Das Kerngenom, in dem die essentiellen Informationen des Stoffwechsels und der Replikation kodiert sind, weist eine Sequenzdiversität von zirka 0,5% auf (Kiewitz und Tümmler, 2000; Spencer et al., 2003). Gene des Stoffwechsels tragen dabei deutlich weniger Punktmutationen als solche der Flagellin- und Siderophorbiosynthese (Spencer et al., 2003) oder der Komponenten des Typ-III Sekretionssystems (ExoS, ExoT, ExoU, ExoY). Die erhöhte Variabilität kann als starkes Indiz für eine diversifizierende Selektion während der Evolution dieser Genloci betrachtet werden (Ghysels et al., 2004; Smith et al., 2005). Zusammengefasst ergibt sich die genetische Variabilität demnach aus der Summe der Unterschiede des stammspezifischen akzessorischen Genoms und den, durch Punktmutationen und Genomrearrangements verursachten Veränderungen des Kerngenoms (Wiehlmann et al., 2007).

1.4 *Pseudomonas aeruginosa* Klon C

Klon C ist der weltweit am weitesten verbreitete Klon der *P. aeruginosa* Population. Entsprechende Isolate wurden sowohl in aquatischen als auch in klinischen Habitaten nachgewiesen (Römling et al., 1994). Anhand ausgewählter Klon C Isolate wurde schon vor mehr als 20 Jahren die Diversität von Bereichen des akzessorischen Genoms untersucht. Mit Pulsfeldgelelektrophorese und *SpeI-PacI-SwaI-I-CeuI*-Makrorestriktionskartierungen von 21 verschiedenen Klon C Isolaten konnten die Positionen von insertierten und deletierten DNA-Blöcken und zusätzlich auch die Endpunkte chromosomaler Inversionen bestimmt

werden (Römling et al., 1997). So konnte auch ohne die zu diesem Zeitpunkt kaum praktikablen Sequenzierungen der kompletten Genome der chromosomale Aufbau und die Plastizität des *P. aeruginosa* Klon C untersucht, sowie anhand der Makrorestriktionen eine erste physikalische Karte von Klon C erstellt werden (Abb. 1.4)(Schmidt et al., 1996). Die Verteilung der vier ribosomalen *rrn* Operons in den Klon C Isolaten war sehr ähnlich wie im Referenzstamm PAO1 (DSM1707) (Schmidt et al., 1996). Auch die Reihenfolge der Markergene des Kerngenoms zeigte zumeist zwischen Klon C und der Referenz PAO1 kaum Unterschiede.

Für einige Stämme des Klon C Komplexes konnten jedoch größere chromosomale Inversionen gezeigt werden (Schmidt et al., 1996). Allerdings wurden diese nur in Isolaten der CF-Lunge detektiert. Diese Inversionen könnten einen möglichen Wachstumsvorteil mit sich bringen, da einige Gene nun näher an der OriC-Region lokalisiert sind und somit aufgrund der Replikationsvorgänge in einer höheren Kopienzahl vorliegen, was eine höhere Transkriptionsrate erlaubt.

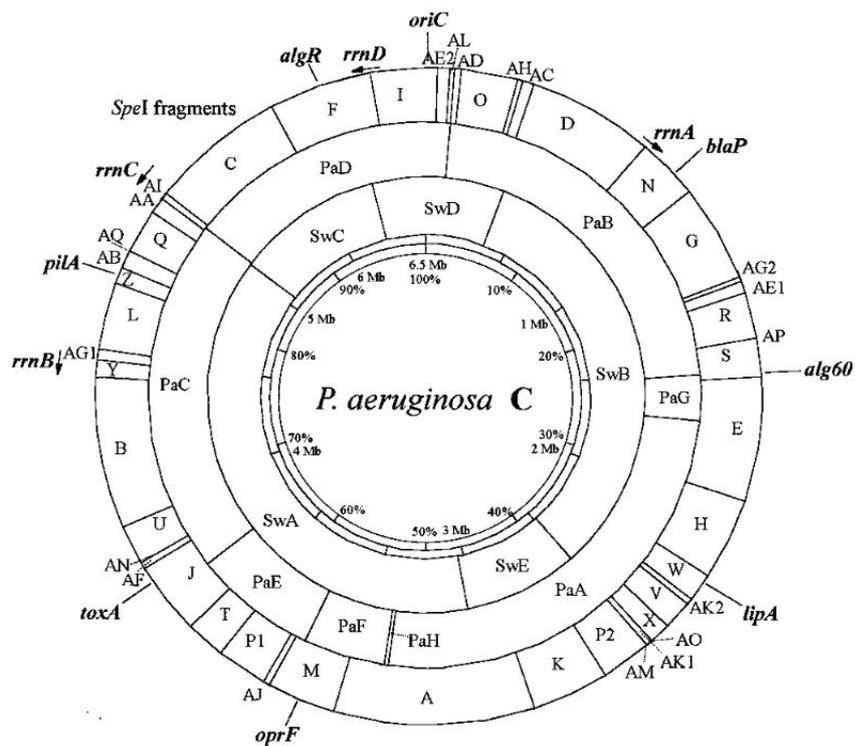


Abbildung 1.4: Physikalische Karte des *P. aeruginosa* Klon C Chromosoms (Schmidt et al., 1996). Die Schnittstellen der Restriktionsenzyme *SpeI*, *PacI* und *SwaI* sind vom äußeren zum inneren Ring dargestellt. Die zwei innersten Ringe zeigen die absolute und die relative Größenskala des Genoms. Zusätzlich sind die Positionen von acht Genen und den vier *rrn* Operons verzeichnet.

Die verschiedenen Genome der untersuchten Klon C Stämme sind im Vergleich zum PAO1-Referenzgenom um bis zu 700 kb größer und weisen intraklonal gegenüber dem Typenstamm C Insertionen mit Größen von 1 - 214 kb und in Einzelfällen auch chromosomale

Inversionen auf. Die Klon C Stämme konnten anhand des unterschiedlichen Spektrums an Insertionen und Deletionen vier Subgruppen zugeordnet werden (Römling et al., 1997). In der Region um den Replikationsursprung von Klon C konnten bislang keine großen Insertionen oder Deletionen detektiert werden. Im Gegensatz dazu war es möglich, drei hypervariable Regionen mit einer Akkumulation großer Genomveränderungen zu identifizieren. Alle drei Bereiche beinhalten im Vergleich zur Referenz PAO1 umfangreiche zusätzlich ins Genom integrierte DNA-Fragmente. So kodiert eine der variablen Regionen das integrierte Plasmid pKLC102 (Römling et al., 1997). Eine weitere Region zeigte in Isolaten aus der Lunge und der Umwelt (SG17M) jeweils ca. 100 kb große, stamm-spezifische Geninseln. Ergebnisse der Sequenzierung und Annotation einer dieser Inseln (PAGI-2) aus früheren Arbeiten (Larbig et al., 2002), wurden im Rahmen dieser Arbeit bei der Sequenzierung des kompletten Genoms des Isolats NN2 zum Schließen einiger Sequenzierungslücken und zur Annotation einiger Gene genutzt.

1.5 Ziele der Arbeit

Ziel dieser Arbeit ist es, die intraklonale Genomdiversität der beiden dominanten Klone in der *P. aeruginosa* Population, Klon C und Klon PA14, basierend auf Sequenzierungsergebnissen von 100 Isolaten zu untersuchen und miteinander zu vergleichen. Dabei sollen sowohl Unterschiede als auch Gemeinsamkeiten gefunden werden, welche einen Rückschluss auf die Diversität des Bakteriums ermöglichen. Zusätzlich sollen die gewonnenen Daten mit denen von 20 unverwandten Klonen abgeglichen werden.

Da vom häufigsten klonalen Komplex C bislang noch keine Genomsequenz bekannt war, soll das Isolat NN2 mit der 454- und der Illumina-Technologie sequenziert und basierend auf den Sequenzdaten das Genom von NN2 komplett entschlüsselt und Gene annotiert werden. Durch zusätzliche RNA-Sequenzierungen sollen Transkriptionsstartpunkte und nicht kodierende RNA in NN2 erkannt und mit dem Referenzstamm PA14 verglichen werden. Anhand dieser Gegenüberstellung sollte es möglich sein, klonspezifische Eigenschaften in der Transkription und Genregulierung abzuleiten.

Im Hauptteil dieser Arbeit sollen zusätzlich zu NN2 57 Stämme des Klons C und 42 Stämme des Klons PA14 mit der exakten SOLiD-Technologie sequenziert und Genomcharakteristika wie Einzelnukleotidaustausche (SNPs), kleine und große Deletionen sowie die Zusammensetzung des akzessorischen Genoms untersucht werden. Auf diesem Weg kann nach Hinweisen dafür gesucht werden, warum diese beiden klonalen Komplexe eine dominierende Stellung in der *P. aeruginosa* Population einnehmen. Die so gewonnen Erkenntnisse sollen ferner mit unverwandten Klonen verglichen werden, um nach einer Signatur innerhalb der *P. aeruginosa* Population zu suchen.

Abschließend sollen anhand von kompetitiven Wachstumsexperimenten Stämme identifiziert werden, welche einen Fitnessvorteil innerhalb der klonalen Komplexe C und PA14 haben und somit in der Lage sind, andere Stämme zu verdrängen. Für diese Gewinner-, aber auch die Verliererstämme soll anhand der zuvor untersuchten Genomcharakteristika gezeigt werden, welche erworbenen Eigenschaften für diese Wachstumsunterschiede verantwortlich sein könnten.

Kapitel 2

Material und Methoden

Anmerkung: Die Quellcodes aller in der Arbeitsgruppe geschriebenen Skripte befinden sich auf dem der Arbeit beigelegten Datenträger.

2.1 Stämme und Wachstumsmedien

P. aeruginosa Isolate von CF-Patienten und Patienten mit akuten Infektionen wurden durch Ausplattieren auf MacConkey-, Mueller-Hinton-, Casein-Soja-Pepton- (TSA) und Blutagar kultiviert (Cramer et al., 2011). Sekundäre Subkulturen der Isolate wurden als 0,3 ml Aliquots in Kryoröhrchen bei -80°C in LB-Medium mit einem 15%-igem (v/v) Glycerolanteil gelagert (Cramer et al., 2012).

Umweltisolate aus dem wässrigen Habitat wurden im Oktober 2007 und im Juli 2008 jeweils an zwei aufeinanderfolgenden Tagen isoliert (Selezska et al., 2012). Zuerst wurde das Wasser filtriert, die Filter anschließend auf Agarplatten mit selektivem Pseudomonas-medium (OXOID)(Pirnay et al., 2005) gelegt und für 36 h bei 37°C inkubiert. Anschließend wurden potentielle *P. aeruginosa* Stämme auf einem selektiven Medium aufgereinigt und durch eine spezifische PCR für *P. aeruginosa* (De Vos et al., 1997) verifiziert.

Als Standardwachstumsmedium für Agarplatten und Flüssigkulturen wurde LB-Medium (engl. lysogeny broth) verwendet. Agarplatten enthielten in der Regel 2% Agar. Für kompetitive Wachstumsexperimente wurde zusätzlich ein M9 Mineralmedium bestehend aus 0,681 g/l Na_2HPO_4 , 3 g/l KH_2PO_4 , 0,05 g/l NaCl , 0,1 g/l NH_4Cl , 0,1 g/l MgSO_4 , 0,01 mM FeSO_4 und 30 mM Natriumsuccinat als Kohlenstoffquelle verwendet.

2.2 DNA-Präparation

Genomische DNA wurde nach einem optimierten Protokoll für gramnegative Bakterien (Ausubel et al., 1994) aus einer bei 37°C in LB-Medium gewachsenen Übernachtskultur isoliert.

2.3 Genotypisierung

Um den Genotyp der einzelnen Isolate zu bestimmen, wurden vor dem Sequenzieren alle Isolate auf einem individuell gefertigten Mikroarray nach einem Standardprotokoll für *P. aeruginosa* typisiert (Wiehlmann et al., 2007).

2.4 DNA-Sequenzierungen

2.4.1 Sequenzierungen des NN2 Genoms

Um die vollständige Genomsequenz des Stamms NN2 zu erhalten, wurde das Isolat mit mehreren Techniken sequenziert. Dabei wurden Daten einer 454 Sequenzierung mit Daten aus Illumina Sequenzierungen von 300 bp paired-end und 3 kb mate pair Fragmentbibliotheken ergänzt.

454 Sequenzierung

Als Vorbereitung zur 454 Sequenzierung wurde die zu sequenzierende DNA in Stücke der Länge 300 - 800 bp fragmentiert und nach der Fragmentgröße sortiert. Anschließend wurden zwei verschiedene Oligonukleotid-Adaptoren an die Enden der Fragmente ligiert, welche als Primerbindestellen der Strangverlängerung dienten. Da einer der Adaptoren biotinyliert war, konnten einzelsträngige Fragmente vor der eigentlichen Sequenzierung eliminiert werden. Durch Binden der Fragmente an „DNA-Capture-Beads“ konnten diese in Tröpfchen einer Emulsion immobilisiert werden. Die mit der DNA überzogenen Beads wurden anschließend in einzelne Wells auf einem Lichtleitfaser-Träger geladen. Die eigentliche Sequenzierung basierte auf dem Prinzip der Pyrosequenzierung. Dabei wurde beim Einbau eines passenden Nukleotids durch die DNA-Polymerase und einer anschließenden Enzymkaskade Licht durch die Abspaltung von anorganischem Pyrophosphat (PPi) erzeugt und detektiert. Zur Identifizierung des passenden Nukleotids wurden diese im Wechsel über die Wells gespült und der Einbau detektiert. Die Anzahl und Intensität der erzeugten Lichtsignale war dabei proportional zur Anzahl der eingebauten Nukleotide (Abb. 2.1).

Die 454-Sequenzierung des NN2 Genoms wurde als Auftragsarbeit im Institut für medizinische Mikrobiologie der Medizinischen Hochschule Hannover durchgeführt.

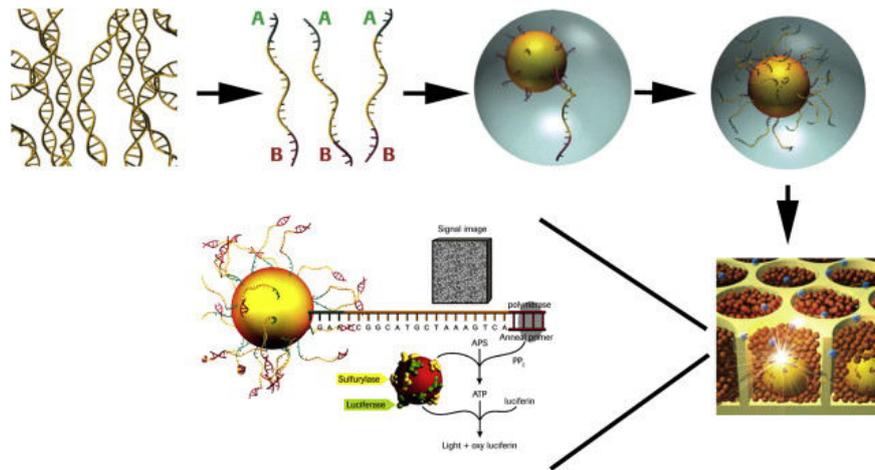


Abbildung 2.1: Schematischer Arbeitsablauf der 454-Sequenzierung. Nach der Ligat-ion der Adaptoren und dem Einfangen der Fragmente auf Beads werden diese in Wells auf einem Lichtleitträger platziert und nachfolgend die Sequenzierung gestartet (Quelle: <http://omicsspace.com>).

Illumina-Sequenzierung

Die Technik der Illumina-Sequenzierung basiert auf der Anheftung randomisiert fragmentierter genomischer DNA-Stücke an die planare, optische Oberfläche einer Flusszelle (Flow Cell). An die DNA-Fragmente wurden chimäre Adaptoren ligiert, welche aus einer Primerbindungs- und einer zu den „Capture Oligos“ der Flusszellen komplementären Domäne aufgebaut waren. Die zu sequenzierenden DNA-Fragmente wurden mittels Hybridisierung durch die auf der Flow Cell lokalisierten Capture Oligos gebunden. Danach wurde ein antiparalleler Strang des DNA-Moleküls synthetisiert, wofür das Capture Oligo als Primer genutzt wurde. Der neu synthetisierte Strang diente als Matrize für die folgenden Amplifikationsschritte, indem er an ein weiteres komplementäres Oligo auf der Flow Cell ligiert wurde. Dieser Schritt wurde wiederholt, bis mehrere hundert Kopien der zu sequenzierenden DNA in so genannten Clustern vorlagen, wobei die Fragmentlängen die Anzahl der entstehenden Cluster bestimmten. Kleinere Fragmente (100 – 500 bp) wurden bevorzugt, weil dadurch mehr Cluster auf einer Flow Cell Platz fanden und somit die Ausbeute erhöht wurde. Waren ausreichend Kopien vorhanden, wurde der komplementäre Strang abgeschnitten, die freien Enden enzymatisch gekürzt und der eigentliche Sequenziervorgang gestartet. Dazu wurden farblich markierte Nukleotide von einer DNA-Polymerase in den zu synthetisierenden Strang eingebaut, wobei die Farbmoleküle als 3'-Terminator wirkten und demnach in jedem Zyklus nur eine Base verwendet werden konnte. Die Bestimmung des eingebauten Nukleotids erfolgte durch Fluoreszenz-Mikroskopie, wobei parallel zur Detektion auch das Farbmolekül abgespalten wurde und nachfolgend der nächste Syntheszyklus gestartet werden konnte (Abb. 2.2). Die maximale Leselänge wurde von der Templatequalität, sowie der Effizienz der Polymerase und der Farbstoffabtrennung reguliert. Da diese Fehler akkumulierten nahm die Verlässlichkeit, dass die

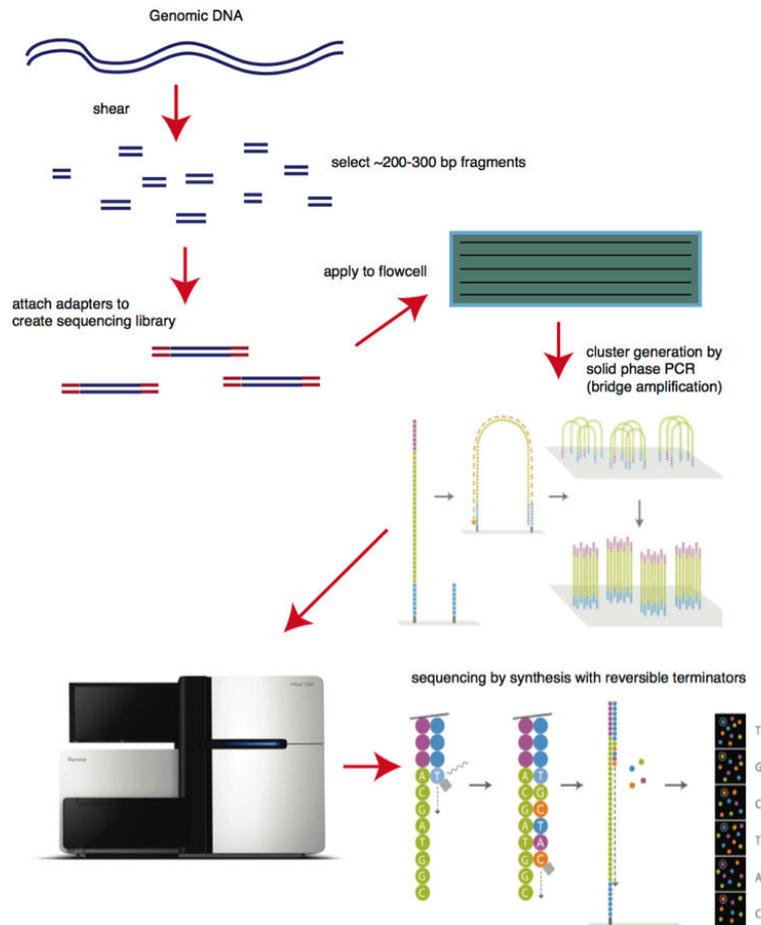


Abbildung 2.2: Schematischer Arbeitsablauf der Illumina-Sequenzierung (Quelle: www.bitesizebio.com).

detektierte Base auch der tatsächlichen Base an der entsprechenden Position in der Matrize entspricht, mit fortschreitender Readlänge ab. Daher war es speziell für die längeren Reads der 300 bp paired-end Bibliotheken nötig, die Sequenzreads auf eine Länge von 80 bp zu trimmen. Beim Stand der Technik zum Zeitpunkt der Sequenzierung wurden Reads mit Längen zwischen 50 und 100 Basenpaaren bevorzugt.

In dieser Arbeit wurde eine 300 bp paired-end und eine 3 kb mate pair Bibliothek genutzt. Paired-end bedeutet, dass Fragmente einer definierten Länge (in dieser Arbeit 300 Basenpaare) von beiden Enden her sequenziert wurden (je 101 bp), wobei die mittleren Basen unbekannt bleiben (Abb. 2.3). Die Kolokalisation zweier Reads auf einem Fragment erleichterte dabei die Assemblierung bzw. Zusammensetzung größerer Sequenzcontigs. Im Vergleich dazu wies die mate pair Bibliothek kürzere Reads (51 bp) bei einem deutlich größeren unbekanntem Bereich von 2898 bp auf. Dadurch ließen sich mit Hilfe dieser Bibliothek auch entferntere Contigs verbinden.

Alle Illumina-Sequenzierungen wurden als Auftragsarbeit durch die Firma GATC-Biotech (Konstanz, Deutschland) durchgeführt.

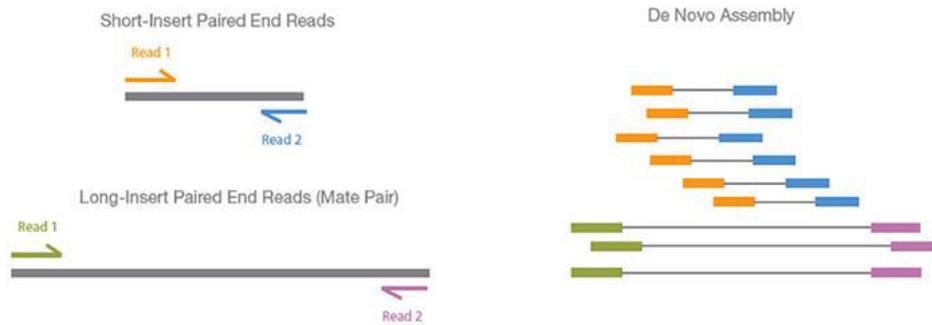


Abbildung 2.3: Darstellung von paired-end (links oben) und mate pair Bibliotheken (links unten). Die rechte Seite zeigt den Assemblierungsprozess (Quelle: illumina.com).

2.4.2 Sequenzierung der 100 Isolate

Die Sequenzierung der 57 Klon C und 42 Klon PA14 Isolate erfolgte mit einer SOLiD Maschine und wurde innerhalb der Arbeitsgruppe durchgeführt. Die SOLiD-Sequenzierung beruhte auf dem Prinzip der Sequenzierung durch Ligation. Dabei wurden mit Fluoreszenzfarbstoffen markierte achtbasige Oligonukleotide nacheinander mit dem Sequenzierprimer ligiert. An das 3'-Ende jeder Sonde, welche die Matrizensequenz komplementär zur vierten und fünften Base der Oligonukleotidsonde enthielt, war ein Fluoreszenzfarbstoff gekoppelt. Durch die Hybridisierung von Sonde und Matrize entstand ein Fluoreszenzsignal, welches von einem Laser detektiert wurde. Anschließend wurde die Sonde nach der fünften Base gespalten und dadurch die Fluoreszenzquelle entfernt. Dieser Vorgang wurde wiederholt, bis die gewünschte Readlänge von maximal 100 Basenpaaren erreicht wurde. Nacheinander wurden fünf dieser Durchläufe mit unterschiedlichen Startpositionen durchgeführt. Im zweiten Durchlauf wurde an der Position n-1 des Templates mit der Ligation gestartet. Durch die Verwendung der fünf Zyklen wurde sichergestellt, dass jede Base in zwei unabhängigen Ligationen durch zwei verschiedene Primer sequenziert wurde (Abb. 2.4). Dieser Umstand erhöhte die Genauigkeit der SOLiD-Sequenzierung drastisch. Eine weitere Steigerung der Genauigkeit wurde durch das Verwenden der „Exact Call Chemistry“ erreicht. Dabei wurde die Sequenzkodierung der Plattform abgeändert, was erlaubte, Messfehler zu korrigieren und mehr korrekte Reads zu produzieren (Massingham und Goldman, 2012).

Herstellung der DNA-Bibliotheken

1 µg bakterieller DNA wurde in einem Covaris S2 System (Covaris, USA) geschnitten. Endreparatur und Größenselektion der entstandenen Fragmente auf eine Durchschnittslänge von 200 bp wurden nach Standardprotokollen (Fragment library generation, Life technologies (LT)/Thermo, USA) durchgeführt. Die Standardprotokolle für die Generierung von Fragmentbibliotheken zur NGS-Anwendung verursachen jedoch eine Unterrepräsentation von GC-reichen Sequenzen, da die Ligation der Adaptoren an DNA-Fragmente

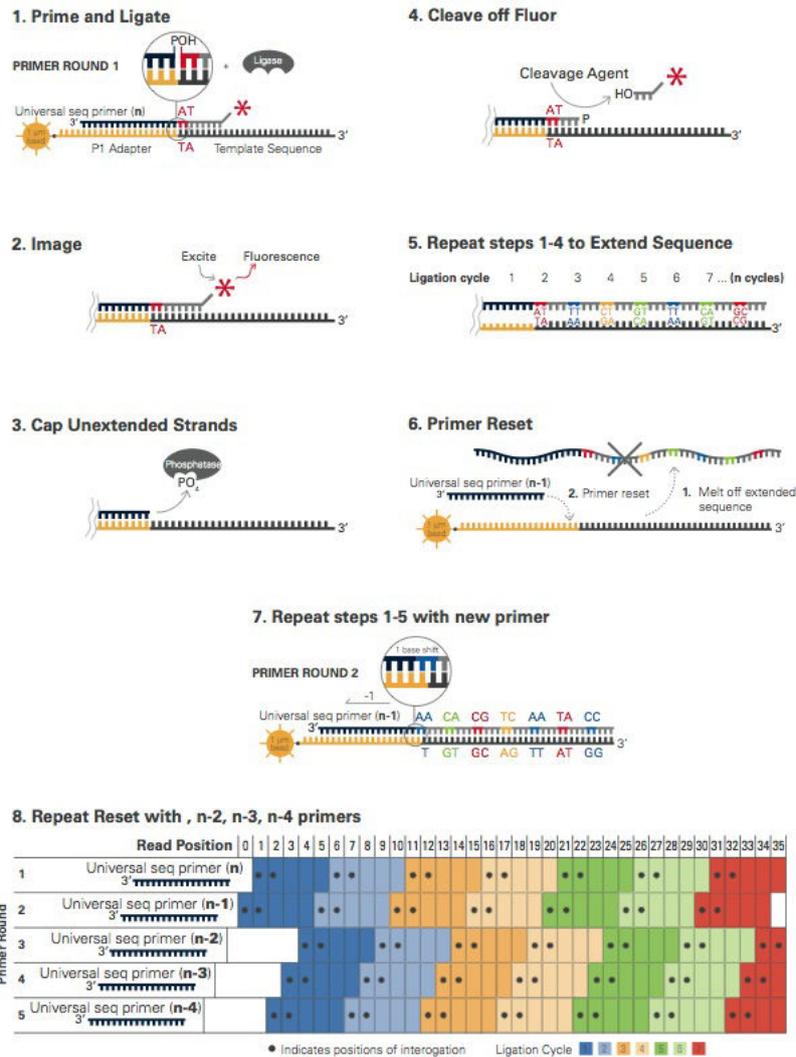


Abbildung 2.4: Schematischer Arbeitsablauf der SOLiD-Sequenzierung (Quelle: <http://seqanswers.com>).

mit mehr als 65% GC-Gehalt nur ineffizient abläuft. Zur Kompensation dieses Problems in *P. aeruginosa* (66% GC-Gehalt) wurde der Ligationsschritt des Standardprotokolls modifiziert. Die Reaktion welche ein Polyadenin an die Fragmente anhängt wurde in $\frac{1}{4}$ des Standardvolumens mit der Stratec Taq Polymerase anstelle der LT-dA-Tailing Enzyme durchgeführt (DNA 9 μ l; 5x Puffer (LT) 2,5 μ l; 10 mM dATP 0,25 μ l, Stratec Taq Polymerase 1,25 μ l; 30 min; 68°C). Die Inkubationsbedingungen der folgenden Ligation wurden abgeändert, um die Lebenszeit und Performance der T4 Ligase zu erhöhen (dA-tailed Reaktionsmix 13 μ l, 5x Puffer (LT) 0,75 μ l; jeder Adaptor (LT, 1:20-verdünnt) 0,5 μ l; 10 mM dNTP 0,3 μ l; T4 Ligase (NebNext, NEB) 0,8 μ l; bidest. Wasser 0,1 μ l; 12 h, 12°C gefolgt von Nick-Translation (20 min, 72°C)). Die anschließende Aufreinigung und Amplifikation (5 Zyklen) der generierten Fragmentbibliotheken wurde nach LT-Standardprotokollen durchgeführt.

Durchführung der SOLiD-Sequenzierung

Das Binden der Fragmentbibliotheken an Beads wurde nach Herstellerangaben durchgeführt (EZBead System (LT); E120 scale, P2 post enrichment 17%). Die Sequenzierung wurde auf einem SOLiD 5500 XL System (LT) mit einer Readlänge von 75 bp und implementierter ECC (Exact call chemistry (LT)) durchgeführt. Die produzierten Reads wurden mit dem SOLiD Accuracy Enhancer Tool (SAET) korrigiert, welches laut Hersteller die Anzahl der kartierbaren Reads für Genome der Größe 1 kb - 200 Mbp bei einer Abdeckung von 10 - 4000-fach und einer Readlänge von 25 - 75 bp durch Fehlerkorrekturen um 40 - 50% erhöhen soll.

Kartierung der SOLiD-Reads

Die Kartierung der ungetrimmten, mit SAET korrigierten Reads auf das jeweilige Referenzgenom wurde mit dem Programm NovoalignCS 1.04.02 (www.novocraft.com) und den Parametern

| | |
|---------------|--|
| -d | Pfad zur Referenzdatei |
| -f | Pfad zu Read- und Qualitätsdatei |
| -F CSFASTAnQV | spezifiziert Readdateien als .csfasta und QV.qual |
| -r Random | zufällige Readverteilung bei mehreren möglichen Positionen |
| -c 16 | Anzahl der Kerne, welche zur Berechnung genutzt werden |
| -o sam | Ausgabeformat: .sam |

durchgeführt. Aus den .sam-Dateien wurden mit der Softwaresammlung Picard Tools 1.6.8 (<http://broadinstitute.github.io/picard/>) binäre Alignment-Dateien (.bam) generiert. Duplikationen von Reads wurden mit dem Paket SAMtools rmdup (Li et al., 2009) eliminiert.

2.5 *De novo* Assemblierung des NN2 Genoms

Bei einer *de novo* Assemblierung werden überlappende Sequenzierungsreads zu größeren DNA-Stücken zusammengesetzt und sogenannte Contigs generiert. Für die 454 Reads erfolgte die Assemblierung mit Hilfe der Software Newbler 2.8 (Margulies et al., 2005). Die minimale erzeugte Contiglänge sollte dabei 150 bp sein, und es wurde die interne Read-Trimming-Funktion des Programms genutzt. Beim Trimming werden einzelne Basen am Readende bei niedrigen Qualitätswerten verworfen. Der Qualitätswert für jede Base eines Reads wurde als Phred-Score beim Sequenziervorgang angegeben, wobei ein geringer Wert für eine hohe Fehlerwahrscheinlichkeit stand. Die assemblierten Contigs wurden

mit Hilfe der Illumina paired-end Reads und des Programms SSPACE (Boetzer et al., 2011) zu größeren Scaffolds assembliert. SSPACE wurde dazu mit folgenden Parametern angewandt:

- x 1 wenn möglich Contigs verlängern
- p 1 Generierung einer .dot - Datei zur Visualisierung
- k 50 minimal 50 Verlinkungen um Scaffolds zu berechnen
- n 30 minimale Überlappung von 30 bp um Contigs zusammenzuführen

2.6 Lokalisation der NN2 Contigs innerhalb der PAO1 Referenzsequenz

Um die Positionierung der Contigs gegenüber dem Referenzstamm PAO1 (Stover et al., 2000) zu untersuchen und deren potentielle Reihenfolge zu ermitteln, wurde das Visualisierungsprogramm Circoletto (Darzentas, 2010) verwendet. Zur Darstellung wurden folgende Parameter genutzt:

- untangling_off um die Contigs unentwirrt darzustellen
- w_hits um nur die Contigs zu zeigen, welche auf das Referenzgenom kartieren

2.7 Schließen des NN2 Genoms

Mit Hilfe des Programms SSPACE und BLAST-Resultaten der Contigs gegen den Referenzstamm PAO1 konnten die Längen der unbekannt Bereiche innerhalb und zwischen den Scaffolds abgeschätzt werden. Für alle Lücken zwischen Contigs, deren Größe auf 0 bp vorhergesagt werden konnte, wurden die betroffenen Contigs *in silico* zusammengesetzt und eine Kartierung aller Illumina Reads mit Hilfe des speziell für kurze Reads konzipierten Kartierungsprogramms BWA (Li und Durbin, 2009) unter Standardeinstellungen gegen dieses vergrößerte Contig durchgeführt. Ob die *in silico* Zusammensetzung korrekt war, zeigte sich daran, ob ein Abbruch der Abdeckung an der Stelle des Zusammensetzens erkennbar war oder nicht (Abb. 3.4).

War keine Bruchkante der Abdeckung zu erkennen, wurde die Lücke als geschlossen betrachtet und zur weiteren Bestätigung der *in silico* Vorhersage exemplarisch drei solcher Bereiche durch PCR-Amplifikation und Sanger-Sequenzierung validiert. Alle so verifizierten „Lücken“ mit 0 bp sowie überlappende Contigs wurden manuell zusammengeführt. Die verbliebenen unbekannt Bereiche wurden mit Hilfe einer PCR überspannt, amplifiziert und per Sangersequenzierung durch die Firma GATC Biotech (Konstanz, Deutschland) sequenziert. Für die PCR wurde folgender Ansatz gewählt:

| | |
|---------------------------------------|-----------------------------------|
| 10 x Reaktionspuffer | 3 μ l |
| Primer [Konz. 5 μ M] | je 3 μ l |
| dNTP [Konz. 8 μ M] | 3 μ l |
| (je 2 mM dATP, dCTP, dGTP, dTTP) | |
| DMSO | 1 μ l |
| bakterielle DNA (10 μ g/ μ l) | 8 μ l |
| MgCl ₂ [50 mM] | 2 μ l |
| Goldstar-Polymerase | 0,2 μ l |
| bidest. Wasser | 6,8 μ l |
| <hr/> Endvolumen | <hr/> 30 μl |

Die PCR wurde nach dem folgenden Programm durchgeführt, wobei T_{An} die Hybridisierungstemperatur (58 - 62 °C) der jeweiligen Primer darstellt:

| | t [s] | T [°C] | |
|---------------|----------|----------|-----------|
| Denaturierung | 300 | 94 | 1 Zyklus |
| Annealing | 60 | T_{An} | |
| Elongation | 60 | 72 | 35 Zyklen |
| Denaturierung | 90 | 94 | |
| Annealing | 60 | T_{An} | |
| Elongation | 90 | 72 | 1 Zyklus |
| Kühlung | ∞ | 10 | |

Eine Lücke in einem Bereich der genomischen Insel PAGI-2 konnte anhand von Daten einer früheren Sequenzierung dieser Region geschlossen werden (Larbig et al., 2002). Eine weitere unbekannt Sequenz in einem hochrepetitiven Segment wurde aufgeklärt, indem für beide Contigenden passende Gegenstücke in den Illumina Reads gesucht und damit die Contigs verlängert wurden. Dieser Prozess wurde wiederholt, bis die Sequenz der Region bekannt war (Abb. 3.5). Bei gemischten Vorhersagen der Reads für einzelne Positionen, welche typischerweise durch punktuelle Abweichungen in den Repeats entstanden, wurde die häufigste Base verwendet. In nicht eindeutigen Fällen wurde die Base der Referenzsequenz PAO1 übernommen, falls diese in einem Teil der NN2 Reads zu finden war.

2.8 Annotation des NN2 Genoms

In der Genetik bezeichnet der Begriff Annotation eine funktionelle Zuordnung von offenen Leserastern zu kodierenden DNA-Sequenzen, sowie die Funktionszuweisung anhand von ähnlichen Genen und Proteinen aus Datenbanken. Basierend auf dem geschlossenen NN2 Genom wurde eine automatische Annotation durch den „Rast server“ durchgeführt (Aziz et al., 2008). Eine manuelle Korrektur erfolgte anhand der Pseudomonas-Datenbank (Stand März 2012) (Winsor et al., 2011). Dabei wurden alle Gene, welche ein bekanntes

Ortholog in den Referenzstämmen PAO1 bzw. PA14 (Lee et al., 2006) aufwiesen, bei einer abweichenden Genlänge an die Koordinaten der Referenz angepasst. Dafür musste sichergestellt werden, dass an den orthologen Genompositionen von NN2 auch ein Startkodon vorhanden war. Konnte kein Startkodon identifiziert werden, wurde die automatische Annotation beibehalten. Nachfolgend wurden Publikationen, welche nach März 2012 veröffentlicht wurden, auf neu aufgeklärte Genfunktionen hin durchgesehen und die zugewiesenen Proteinfunktionen gegebenenfalls aktualisiert. Die Annotation bekannter genomischer Inseln konnte aus der Literatur übernommen werden (Larbig et al., 2002). Bei allen Genen ohne Referenzhomolog wurde ebenfalls die automatische Annotation weiterverwendet.

Bislang nicht-annotierte Genomregionen, in denen offene Leseraster (engl. open reading frames, ORFs) von mehr als 150 Basenpaaren detektiert werden konnten, wurden abschließend genauer untersucht. Sequenzen dieser potenziellen ORFs wurden mit Hilfe eines Abgleichs gegen die NCBI-Datenbank (BLAST) auf Homologien in anderen Spezies überprüft. So identifizierte ORFs wurden zur Annotation hinzugefügt. Das annotierte Genom wurde als .gbk-Datei dieser Arbeit beigelegt (KlonC_NN2.gbk).

2.9 RNA-Sequenzierung

2.9.1 Kultivierung der Bakterien

Für die RNA-Sequenzierung wurden Bakterien in einem Fermenter (Biostat B Plus, Sartorius) unter standardisierten Bedingungen kultiviert. Hierfür wurden zunächst bakterielle Kulturen des Isolates NN2 auf TSA-Platten ausgestrichen und für 14 h bei 37°C inkubiert. Vorkulturen wurden mit einer optischen Dichte (OD) von 0,225 (550 nm) in 50 ml TSB-Medium angelegt und für eine Stunde bei 37°C kontinuierlich mit 125 rpm geschüttelt. Anschließend wurden die Vorkulturen mit etwa 10^7 CFU/ml zu 1,5 l TSB-Medium in den Fermenter gegeben und mit 750 µl des Antischaummittels „antifoam“ (Struktol) versetzt. Die Kultur wurde bei 37°C, konstantem pH-Wert (7,0), sowie gleichmäßiger Agitation (400 rpm) und Belüftung (komprimierte Luft 0,3 l/min) für 20 h inkubiert und das Wachstum separat durch die Bestimmung von OD und CFU-Zahlen, sowie im Fermenter durch Messung des Sauerstoffpartialdrucks (pO_2) und des Redoxstatus der Fermenterkultur überprüft. Für die RNA-Extraktion wurden jeweils drei Aliquots der Kultur während der mittleren exponentiellen Wachstumsphase und der frühen stationären Phase geerntet, wobei der Sauerstoffpartialdruck als Indikator für die jeweilige Wachstumsphase diente (Abb. 2.5).

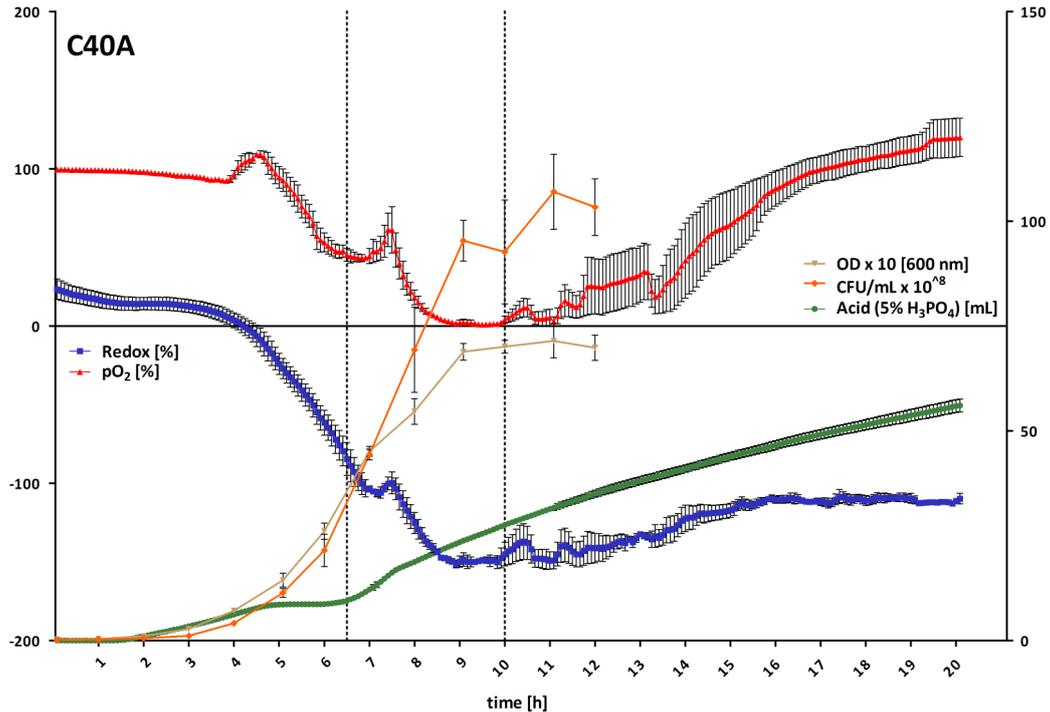


Abbildung 2.5: Wachstumskurve des Isolates NN2 bei konstanten Bedingungen in einem Fermenter. Die gepunkteten horizontalen Linien geben die beiden Probenahmezeitpunkte an.

2.9.2 RNA-Isolierung und Depletion ribosomaler RNA

Die gesamte RNA (>18 Nukleotide) wurde mit RNA-Extraktionskits aus den Bakterienaliquots extrahiert (Qiagen, RNAProtect Bacteria Reagent und RNeasy Midi Kit). Zur Zellyse wurden 8 ml RNAProtect-Reagenz mit 5 ml Bakterienkultur vermischt. Nach fünfminütiger Inkubation bei Raumtemperatur wurde die Lösung zentrifugiert (5.000 x g, 10 min) und der Überstand dekantiert. Zellpellets wurden bis zur Prozessierung bei -80°C gelagert. Die Zellen wurden in 1 ml TE-Buffer, welcher 1,5 mg/ml Lysozym enthielt, und anschließend in 2 ml RLT-Puffer (Qiagen) mit 1% MeEtOH durch Zufügen von 1,5 vol-% absoluten Ethanol lysiert. Die RNA wurde nach Bindung an RNeasy-Säulen (Qiagen) durch einen zweimaligen zehnmütigen DNaseI Verdau aufgereinigt und schließlich mit einem finalen Volumen von 300 µl bidestilliertem Wasser eluiert. Die RNA-Integrität wurde auf einem Bioanalyser 2100 expert (Agilent, USA) überprüft, wobei alle RNA-Proben eine RNA Integritätsnummer (RIN) von mindestens 7,6 bei einem Durchschnitt von 9,3 hatten. Die RIN gab dabei an, in welchem Zustand sich die RNA befand. Stark degradierte RNA hatte einen Wert von 1, komplett intakte RNA hingegen den Maximalwert 10. Um verbliebene DNA-Kontaminationen zu eliminieren, wurden die Proben vor dem Entfernen der ribosomalen RNA (rRNA) noch einmal mit DNaseI verdaut. In Aliquots mit jeweils 5 µg totaler RNA wurde die ribosomale RNA mit Hilfe des RiboZero Kits für gramnegative Bakterien (Metabion) abgereichert. Mit RiboZero behandelte RNA wurde

durch Ethanolpräzipitation zurückgewonnen und in 12 µl bidestilliertem Wasser gelöst.

2.9.3 Durchführung der RNA-Sequenzierung

Die RNA aus den jeweils drei Proben jeder Kondition wurde gepoolt und durch die Firma GATC Biotech (Konstanz, Deutschland) auf einer Illumina Maschine nach dem Protokoll „TruSeq Stranded mRNA LT Sample Prep Kit“ (Illumina) sequenziert.

Die mit RiboZero behandelten RNA-Proben wurden benutzt, um zwei strangspezifische Bibliotheken herzustellen: eine unbehandelte und eine mit der Terminator 5'-Phosphat abhängigen Exonuklease (TEX, Epicentre) verdaut Probe. TEX ist eine prozessive 5'-3'-Exonuklease, die RNA mit einem 5'-Monophosphat verdaut (Abb. 2.6). Aus den Sequenzierungen der mit TEX behandelten Bibliotheken konnten die primären Transkripte identifiziert und somit die 5'-untranslatierten Regionen (5'-UTR) bestimmt werden. Strangspezifische Sequenzierungen wurden durchgeführt, um Aussagen treffen zu können,

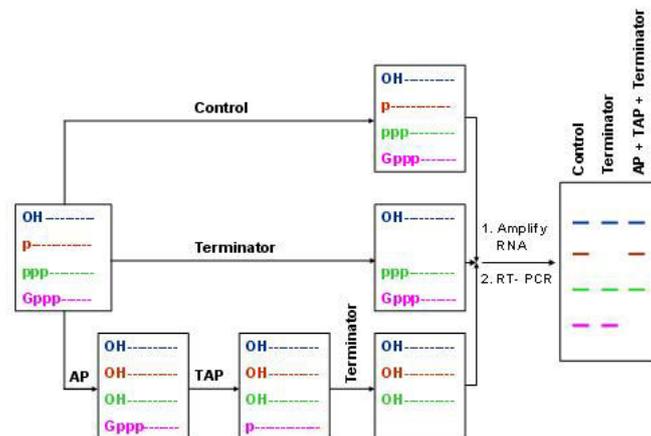


Abbildung 2.6: Schema, wie mit der Terminator 5'-Phosphat abhängigen Exonuklease (TEX) primäre Transkripte angereichert und solche mit Monophosphat verdaut werden (Quelle: <http://www.epibio.com>).

ob die transkribierte RNA auf den Leit- oder den Rückwärtsstrang des Bakteriengenoms kartiert wird (Parkhomchuk et al., 2009). Diese Stranginformation war unter anderem für die Analyse kurzer, nicht-kodierender RNA wichtig. Für eine strangspezifische Sequenzierung wurde aus der aufgereinigten RNA der erste Strang als cDNA synthetisiert. Im folgenden wurde die RNA abgespalten und bei der Synthese des zweiten Strangs Uracil anstelle von Thymin eingebaut. Nach einer Endreparatur und einer Strangverlängerung mit Polyadenin wurden Adaptoren ligiert und in den zweiten Strängen das Uracil verdaut. Da der Strang dadurch viele Lücken aufwies, wurde er degradiert und in der nachfolgenden PCR nicht amplifiziert. Auf diese Weise wurde sichergestellt, dass nur der erste Strang weiter prozessiert und sequenziert wird (Abb. 2.7).

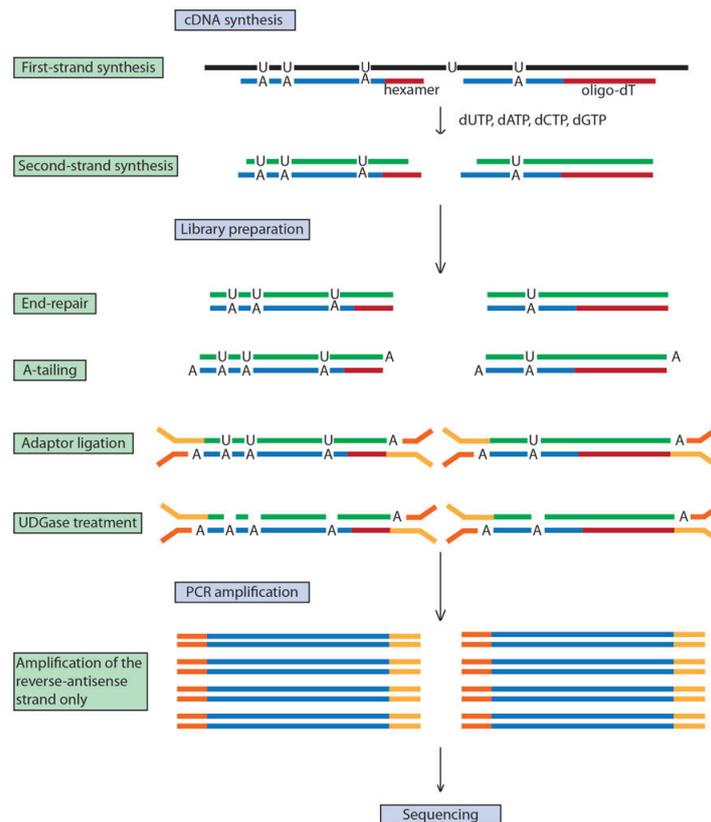


Abbildung 2.7: Schematischer Ablauf einer strangspezifischen Sequenzierung (Martin et al., 2013).

2.10 Identifikation von Transkriptionsstartpunkten und kurzen nicht-kodierenden RNA-Molekülen

Zur Auswertung der Transkriptomsequenzierung von NN2 wurden die strangspezifischen Reads mit dem Kartierungsprogramm BWA unter Standardeinstellungen gegen das NN2 Genom kartiert. Die Visualisierung, die Lokalisation der Reads im NN2 Genom und die manuelle Identifikation von Transkriptionsstartpunkten (engl. transcriptional start sites, TSS) und möglicher nicht-kodierender RNAs (engl. non-coding RNAs, ncRNAs) wurde in dem Genombrowser ReadXplorer (Hilker et al., 2014) durchgeführt. Dazu wurde vor dem eigentlichen Startcodon jedes Gens der, für eine 5'-UTR typische, sprunghafte Anstieg der Readabdeckung in den mit TEX-behandelten Proben gesucht.

Wenn diese charakteristischen Anstiege der Expression in großen intergenischen Bereichen auftraten, eine Expression in den unbehandelten Proben nachweisbar war und für diese Bereiche der Referenzsequenz kein ORF annotiert war, wurden transkribierte Abschnitte als intergenische ncRNAs klassifiziert. Intragenische ncRNAs wurden identifiziert, wenn die Sprünge der Readabdeckung innerhalb eines Gens lokalisiert waren und die unbehandelten Bibliotheken eine erhöhte Expression nur für einen Teilbereich des Gens zeigten.

Die Vorhersagen von intragenischen ncRNAs und deren Längen waren in diesem spezifischen Fall möglich, da vermutlich der dUTP-Degradationsschritt im Arbeitsablauf der Herstellung der strangspezifischen Bibliotheken unvollständig war. Somit lagen sowohl der gewünschte Zielstrang als auch der synthetisierte Folgestrang vor und wurden sequenziert. Die Zuordnung der ncRNAs zu einem der beiden DNA-Stränge des Genoms war nach Verlust der Strangspezifität der Sequenzreads allerdings nicht mehr möglich.

2.11 Differentielle Genexpressionsanalyse

Aus den RNA-Sequenzierungsdaten der nicht-spezifischen Fragmentbibliotheken (vgl. Kap. 2.9) konnte die Expression der einzelnen Gene errechnet und zwischen der frühen stationären und der mittleren exponentiellen Wachstumsphase verglichen werden. Dazu wurde die Expression als der RPKM-Wert mit dem Skript `rpkmforgenes.py` (Ramsköld et al., 2009) und den Parametern

| | |
|----------------|---|
| -i | Pfad zur Eingabedatei |
| -a | Pfad zur Hilfsdatei, welche die zu untersuchenden ORF-Koordinaten spezifiziert |
| -o | Pfad und Name der Ausgabedatei |
| -readcount | Anzahl der Reads für die ORFs soll ausgegeben werden |
| -rmnameoverlap | Überlappungen von ORFs werden verworfen, Isoformen werden weiter verwendet |

bestimmt. Der RPKM-Wert gibt die Anzahl der Reads pro ORF normiert auf 1 kb ORF-Länge pro eine Million kartierter Reads an (Mortazavi et al., 2008).

Die Normierung auf den GC-Gehalt und die Länge jedes ORFs wurde mit einem in der Arbeitsgruppe entwickelten Skript durchgeführt (`ExonicGC_fromTranscriptome.pl`). Zusätzlich wurden mit einem weiteren Skript die RPKM-Daten für die biologischen Replikate der beiden Konditionen zusammengefasst (`PrepareCountsForDESeq.pl`). Die differentielle, auf den GC-Gehalt normalisierte Genexpression wurde mit den R-Paketen `DESeq` und `EDASeq` ermittelt (`GC_DESeq.R`).

2.12 Erstellen des Genomatlas für NN2

Als Genomatlas bezeichnet man eine Übersicht über ein Genom, welche die Informationen über die Länge des Genoms, die Verteilung des GC-Gehalts sowie die Lokalisation von ORFs und RNA darstellt. Der Genomatlas von NN2 wurde mit der Online-Software `CGView` anhand der NN2-Sequenz, der Annotation der ORFs und der identifizierten RNAs erstellt (Stothard und Wishart, 2005).

2.13 Struktur nicht-kodierender RNAs (ncRNAs)

Die Struktur und die thermodynamische Stabilität von RNA-Molekülen sind ein Indikator dafür, ob eine RNA als real angesehen werden kann. Um die Plastizität der annotierten ncRNAs von NN2 zu prüfen, wurde deren potentielle Sekundärstruktur sowie deren thermodynamische Stabilität mit der Software „Vienna RNA websuite“ (Gruber et al., 2008) *in silico* ermittelt.

2.14 Verteilung von Nukleotidtetrameren im NN2 Genom

Die statistisch erhöhte Nutzung bestimmter kurzer Oligonukleotide ist typisch für verschiedene Organismen (Reva und Tümmler, 2004). Da die Genome verschiedener Organismen jedoch auch unterschiedliche GC-Gehalte aufweisen, ist es nötig, die Zahl der Oligonukleotide auf den GC-Gehalt der Genome zu normalisieren. Die Verteilung der Nukleotidtetramere des NN2 Genoms wurden mit einem Zählwerkzeug des Softwarepakets CLC Genomics Workbench 7.0 (<http://www.clcbio.com>) erfasst. Dieselbe Analyse wurde separat auch für die ORFs angewandt, für welche in dieser Arbeit eine 5'-UTR ermittelt werden konnte. Hierzu wurden alle Tetramere gezählt, welche in einem Bereich von 250 bp vor dem Startcodon eines ORFs sowie in den ersten 50 bp danach lokalisiert waren. Anschließend wurde die Häufigkeit, mit der ein Tetramer auftritt, auf den GC-Gehalt des gesamten NN2 Genoms von 66,12% normalisiert.

2.15 Vorhersage von Einzelnukleotidaustauschen

Einzelnukleotidaustausche (engl. Single Nucleotide Polymorphisms, SNPs) stellen Varianten eines Basenpaares innerhalb eines DNA-Strangs dar. Für die sequenzierten Isolate der klonalen Komplexe C und PA14 wurde aus den in Kapitel 2.4.2 beschriebenen .bam-Dateien die SNPs gegenüber der jeweiligen Referenzsequenz NN2 (Klon C) bzw. PA14 (Klon PA14) in einem dreistufigen Prozess mit dem Softwarepaket SAMtools bestimmt. Im ersten Schritt wurde mit der Funktion mpileup und den Optionen

- g Berechnung von Genotypwahrscheinlichkeiten und Ausgabe im .bcf Format
- B Deaktivierung der probabilistischen Rekartierung; empfohlen, um die Zahl falsch-positiver SNPs zu reduzieren
- f Pfad zur Referenzdatei

eine Liste von Genompositionen erstellt, die alle Informationen wie Abdeckung, Qualität und ggf. zu welcher Base mutiert wurde, enthält. Nachfolgend konnte mit der Funktion bcftools view und den Parametern

- c Variantenbestimmung durch Bayessche Statistik
- v nur Varianten ausgeben
- g Genotyp jeder Variantenposition bestimmen

die .bcf in eine .vcf Datei umgewandelt und dabei die SNP-Kandidaten und das Verhältnis zwischen Reads mit der Referenzsequenz und Reads mit einem Einzelnukleotidaustausch bestimmt werden. Im abschließenden Schritt mit der Funktion `vcfutils.pl varFilter -d 2` wurden die vollständigen SNP-Listen, welche die Informationen über alle Varianten der jeweiligen Stämme mit einer Mindestabdeckung von zwei Reads zeigten, generiert. Aus den vollständigen Listen wurden alle Einträge mit einer Abdeckung kleiner als 4, einem Qualitätswert unter 50 und einem Verhältnis zwischen Referenz zu Nukleotidaustausch von weniger als 0,7 verworfen. In den Qualitätswert fließen in diesem Fall die Qualität der einzelnen Basen der Sequenzierung und das Verhältnis zwischen Reads, welche die Sequenz der Referenz bzw. des SNPs zeigen, ein und er kann für jede Position den Maximalwert 222 erreichen. SNP-Statistiken wie die Lokalisation in ORFs, die Vorhersage, ob es sich um einen neutralen (synonymen) oder einen Aminosäuresequenz-verändernden (nicht-synonymen) Austausch, sowie um eine Transition oder Transversion handelte, wurden mit dem Programm `SNPeff 3.0` (Cingolani et al., 2012) erstellt.

2.16 Gruppierung von Aminosäureaustauschen

Gefundene Aminosäureaustausche wurden nach der Dayhoff-Ähnlichkeits-Index-Matrix, welche die Frequenz der Aminosäureaustausche zwischen nah verwandten Proteinen beschreibt (Dayhoff, 1978), eingeordnet. Dabei hatten drastische Aminosäureänderungen, welche mit einer höheren Wahrscheinlichkeit die Struktur und Funktion des betroffenen Proteins verändern, einen geringen Wert und eher neutrale Austausche einen hohen Wert.

2.17 Bestimmung von Haplotypen

Als Haplotypen bezeichnet man in diploiden Organismen Segmente mit absoluter Sequenzidentität. Um Bereiche 100%-iger Sequenzidentität im Kerngenom der beiden klonalen Komplexe zu ermitteln, wurden aus den SNP-Listen die Positionen im akzessorischen Genom entfernt. Ferner wurden die verbliebenen SNPs reduziert und dabei alle Einträge mit einer Abdeckung kleiner als 6, einer SNP-Qualität kleiner 150, einer Basenqualität kleiner als 30, sowie alle Insertions-, Deletions- (Indel) und Heterozygotenvorhersagen verworfen. Durch eine in der Arbeitsgruppe entwickelte Vorgehensweise konnten die Haplotypblöcke wie nachfolgend beschrieben berechnet werden (`haplotypes_identification.pl`).

Der erste Schritt war die Bildung einer Initialmatrix, deren Spalten sämtliche SNP-Positionen aus allen Isolaten und deren Reihen die Resultate für die einzelnen Stämme

enthielten. Dabei konnten in dieser Matrix nur die Zahlen 1 (Nukleotidaustausch zur Referenzsequenz) und 0 (Sequenz der Referenz) stehen. Anschließend wurden in paarweisen Vergleichen die Muster der Nukleotidaustausche zwischen den Stämmen verglichen. Die Haplotypen wurden ermittelt, indem innerhalb dieses binären Systems alle gemeinsamen Positionen bis zur ersten Abweichung gezählt wurden. Diese Zahlen der gemeinsamen Nukleotidaustausche N wurden als Elemente einer zweiten Matrix geschrieben, in der die SNP-Positionen in den jeweiligen Spalten und die paarweisen Vergleiche in den Reihen standen. Aus dieser Matrix wurden die Haplotypen der paarweisen Vergleiche extrahiert, um deren Länge und die Verteilung der Häufigkeit einer bestimmten Haplotyplänge zu ermitteln.

2.18 Phylogenetische Bäume

Die zur Haplotypanalyse verwendeten SNPs der einzelnen Isolate wurden jeweils durch ein Skript in die zugehörige Referenz PA14 bzw. NN2 inkorporiert (`SequenceReplacer.java`), um die virtuelle Sequenz des Kerngenoms jedes Stammes zu ermitteln. Anhand dieser stammspezifischen Sequenzen konnte der phylogenetische Baum der klonalen Komplexe C und PA14 mit Hilfe des Programms `Splitstree 4.13.1` berechnet werden (Huson und Bryant, 2006).

2.19 Visuelle Darstellung der Zusammensetzung des akzessorischen Genoms (Heatmaps)

Die SOLiD-Reads aus den Sequenzierungen der einzelnen Stämme wurden mit dem Kartierungsprogramm `NovoalignCS 1.04.02` gegen die Sequenzen der bekannten genomischen Inseln und Regionen hoher Genomplastizität (RGPs) von *P. aeruginosa* kartiert. Nicht abgedeckte Regionen wurden durch die Funktion `genomeCoverageBed` des Softwarepakets `BEDTools` (Quinlan und Hall, 2010) ermittelt und der prozentuale Anteil der abgedeckten Bereiche durch ein in der Arbeitsgruppe geschriebenes Skript berechnet (`LengthCoverage.pl`). Eine Normalisierung war notwendig, da die berechnete Sequenzabdeckung durch die unterschiedliche Readanzahl der Stammsequenzierungen limitiert wurde. Dazu wurden die prozentualen Anteile durch die Wahrscheinlichkeit, dass eine genomische Position durch mindestens einen Read abgedeckt ist, geteilt:

$$P = 1 - \left(\frac{1 - L}{G}\right)^N$$

In der Formel steht dabei L für die Readlänge, G für die Genomlänge des Referenzgenoms (NN2, PA14) und N für die Readanzahl. Da angestrebt wurde, für jede genomische Insel

oder RGP-Insertion eine Aussage über die An- bzw. Abwesenheit zu treffen, wurde die Readabdeckung in partiell abgedeckten Elementen des akzessorischen Genoms durch eine inverse logarithmische Transformation normiert. Durch die Transformation wurden hohe Abdeckungswerte verstärkt und die Werte schwach abgedeckter Regionen verringert. Dadurch ließ sich eine klarere Aussage über die An- und Abwesenheit der genomischen Inseln und RGPs treffen. Die Heatmaps der transformierten Abdeckungsdaten der genomischen Inseln und RGPs wurden mit dem `gplots` R-Paket erstellt (`heatmap.R`).

2.20 Konservierung des akzessorischen Genoms

Um zu vergleichen, welche genomischen Inseln oder RGPs die einzelnen Stämme aufweisen, wurden für alle in Kapitel 2.19 untersuchten genomischen Varianten deren An- bzw. Abwesenheit untersucht. Dabei wurden alle genomischen Inseln oder RGPs als in dem jeweiligen Stamm vorhanden angesehen, wenn mindestens eine Abdeckung von 50% vorlag. Anschließend wurde die An- bzw. Abwesenheit der genomischen Region in paarweisen Vergleichen zweier Isolate des selben Klons untersucht. Für jede Übereinstimmung wurde 1 addiert, für jeden Unterschied 1 subtrahiert. Anschließend wurde für jeden paarweisen Vergleich durch Summierung von Unterschieden und Gemeinsamkeiten ein globaler Score ermittelt.

2.21 Bestimmung von Deletionen größer als 1 kb

Zur Bestimmung großer Deletionen gegenüber den Referenzgenomen NN2 und PA14 wurde für jedes Isolat die Readanzahl in einem variablen Fenster, beginnend bei Genomposition 1, von 1.000 bp bestimmt, welches über das gesamte Genom in Schritten von 500 bp Länge verschoben wurde. Das Zählen der Reads wurde mittels des Skriptes `rpkmformgenes.py` (Ramsköld et al., 2009) und den Parametern

```
-i          Pfad zur Eingabedatei
-a          Pfad zur Hilfsdatei, welche die zu untersuchenden
            Genomintervalle spezifiziert
-o          Pfad und Name der Ausgabedatei
-readcount  Anzahl der Reads in den Intervallen soll ausgegeben werden
```

durchgeführt. Anschließend wurde eine manuelle Inspektion aller Fenster mit weniger als zehn Reads durchgeführt, um zu überprüfen, ob die geringe Anzahl an Reads aufgrund eines lokal erhöhten GC-Gehaltes auftrat oder eine tatsächliche Deletion vorlag.

2.22 Analyse kleiner Insertionen und Deletionen (Indels)

Indelvorhersagen basierten auf den Kartierungen der SOLiD-Reads auf die jeweilige Referenz (PA14, NN2). Positionen mit Indel-Vorhersage wurden aus den .sam - Dateien mit dem Softwarepaket SAMtools extrahiert. Aus dieser Indel-Kandidatenliste wurde all diejenigen Positionen verworfen, deren Readabdeckung kleiner als 6 war oder wenn mehr als 50% der Reads der Indel-Vorhersage widersprachen. Die verbliebenen Kandidaten wurden manuell mit dem Integrative Genome Viewer (IGV) (Thorvaldsdottir et al., 2013) untersucht, mit dem die Kartierungen der Reads auf die Referenzsequenz visualisiert werden konnten. Indelvorhersagen wurden als nicht verifizierbar oder falsch positiv angesehen, wenn eines der folgenden Ereignisse auftrat:

- a. Position des Indels zu nah (innerhalb einer Readlänge) an einer nicht-abgedeckten Region
- b. Locus in der Mehrzahl der Isolate schlecht (<6 Reads) oder gar nicht abgedeckt
- c. Position nur durch Readenden abgedeckt
- d. vermischte Resultate zwischen Wildtyp- und Indel-Vorhersage an der jeweiligen Position in allen Isolaten
- e. Vorhersage verschiedener Indels an derselben Position für ein Isolat
- f. Position trotz hoher Abdeckung nur durch ein paar Reads als Indel angezeigt, wobei andere Reads aufgrund von anderen Varianten nicht als Wildtyp gezählt wurden.

Wenn gleiche Indel-Vorhersagen in mehreren Isolaten einer klonalen Linie detektiert worden waren, wurden die entsprechenden Positionen auch in den übrigen Isolaten inspiziert, um zu überprüfen, ob diese tatsächlich die Wildtyp-Sequenz tragen oder keine Indel-Vorhersage aufgrund von schlechter Abdeckung oder schlechten Qualitätswerten getroffen wurde. Ferner wurden die Indels, welche in allen Stämmen präsent waren, entweder als Sequenzierungsfehler oder spezifische Sequenzvarianten des jeweiligen Referenzisolates betrachtet.

Die verifizierten Indels wurden mittels SNPeff 3.0 annotiert, um die betroffene Genomregion und potentielle Leserasterverschiebungen in kodierenden Regionen zu bestimmen. Indels, welche Protein-kodierende ORFs betrafen, wurden nach Annotationsklassen der betroffenen Gene in der jeweiligen Referenz sortiert. Für PA14 konnten die Annotationsklassen von der PA14 Genom Webseite übernommen werden

(<http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/annotation/start.cgi>).

Für Klon C wurde für alle PAO1-orthologen Gene die jeweilige Klasse aus der Annotation des PAO1-Sequenzierprojektes (Winsor et al., 2011) der Referenz übernommen. Alle nicht-PAO1 Gene, deren Funktion als hypothetisch vorhergesagt wurde, wurden in Gruppe 4 eingeordnet. Für die verbliebenen ORFs wurde in der Literatur nach experimentellen

Verifizierungen gesucht, um diese in die Klassen 1 oder 2 einzusortieren. Konnten keine Funktionszuordnungen gefunden werden, wurden die ORFs in Kategorie 3/4 einsortiert. Für Vertreter der Klasse 1 wurde in *P. aeruginosa* die Funktion des Gens nachgewiesen. Für ORFs der Klasse 2 konnte von sehr sequenzähnlichen Genen in anderen Organismen die Funktion aufgeklärt werden. Für ORFs der Klassen 3 und 4 wurde lediglich eine *in silico* Vorhersage über die Lokalisation und Funktion des Genproduktes getätigt.

2.23 Identifizierung stammspezifischer Gene

Als stammspezifische Gene werden alle ORFs bezeichnet, welche weder in den Referenzsequenzen NN2 bzw. PA14, noch in den untersuchten genomischen Inseln und RGPs kodiert sind. Um die stammspezifischen Gene zu identifizieren, wurden alle Reads, welche weder auf das jeweilige Referenzgenom C oder PA14 noch auf die bekannten genomischen Inseln und RGPs kartiert werden konnten und somit stammspezifische DNA repräsentieren, aus den .bam-Dateien mit dem Softwarepaket SAMtools view und den Optionen

```
-b .bam Ausgabeformat
-f 4 extrahiere nur nicht kartierte Reads
```

extrahiert. Für die *de novo* Assemblierung dieser Reads wurde das Programm Velvet (Zerbino und Birney, 2008) mit den K-mer Werten 27, 29, 31, 33, 35, 37, 39 und 41 verwendet. Für jeden Stamm wurde das beste Assemblierungsergebnis der verschiedenen K-mer Werte ausgewählt. Dieses war dadurch charakterisiert, dass es den höchsten N50-Wert bei maximaler Contiglänge aufwies. Der N50-Wert gibt den Median der Contiglänge an, welchen 50% der Contigs aufweisen. Um falsch-positive Contigs zu eliminieren, welche z.B. dadurch entstanden sein konnten, dass Reads initial aufgrund von schlechten Qualitätswerten nicht auf die Referenz kartiert werden konnten, wurden diese mit dem Lang-Read Kartierungsskript CUSHAW2 (Liu und Schmidt, 2012) unter Standardeinstellungen noch einmal gegen das Referenzgenom, sowie die genomischen Inseln und RGPs kartiert. Alle in diesem Durchlauf kartierten Reads wurden verworfen.

Für die verbliebenen Contigs wurde ein BLASTx-Vergleich mit dem Programm blastall gegen eine Datenbank, welche alle bis 2012 sequenzierten Bakteriengenome enthielt, mit den Parametern

```
-p blastx BLAST einer Nukleotidsequenz gegen eine Proteindatenbank
-d       Pfad zur Proteindatenbank
-i       Pfad zur Eingabedatei
-m 8    Ausgabe in tabellarischer Ansicht
-e 1E-10 Erwartungswert
-a 8    Anzahl der genutzten Prozessorkerne
-o       Pfad und Name der Ausgabedatei
```

durchgeführt. Aus diesem Datensatz wurden alle ORFs identifiziert, welche die gleiche Sequenz in unterschiedlichen Organismen kodierten. Solche Einträge wurden als falsch-positive ORFs erkannt und zu einem Treffer zusammengefasst. Anhand der BLASTx Resultate wurden alle ORFs mit einem Skript (`gather.py`) zu einem stammspezifischen Gen zusammengefasst, wenn sie 90% Sequenzidentität über mindestens 80% der Genlänge aufwiesen.

2.24 Kompetitive Fitnessexperimente

Kompetitive Wachstumsexperimente verschiedener Stämme wurden durchgeführt, um Isolate zu identifizieren, welche einen Wachstumsvorteil gegenüber anderen Stämmen aufweisen. Wurden Gewinner und Verlierer erkannt, konnten anhand der stammspezifischen genomischen Eigenschaften wie Einzelnukleotidaustauschen, Insertionen, Deletionen und stammspezifischen Genen mögliche Gründe für die Wachstumsvor- bzw. -nachteile untersucht werden.

2.24.1 Messung der Pyocinproduktion

Bevor ein kompetitives Fitnessexperiment mehrerer Stämme durchgeführt werden konnte, musste überprüft werden, ob die Stämme kompatibel in ihrer Pyocinproduktion waren. Dazu wurden die Pyocinloci der einzelnen Stämme *in silico* auf die jeweilige Präsenz und die darin auftretenden Mutationen überprüft. Da zum Zeitpunkt der Fitnessexperimente noch nicht alle Genomsequenzierungen abgeschlossen waren, konnten lediglich 47 Klon C Stämme auf Pyocinaktivität getestet werden. Dabei zeigte *in silico* keines der Isolate eine Mutation im Vergleich zu dem Pyocinlocus des Referenzgenoms NN2. Die 47 Isolate wurden in drei Gruppen à zwölf Isolate und eine Gruppe mit elf Isolaten eingeteilt und experimentell auf Pyocinkompatibilität getestet. Innerhalb der PA14 Isolate wurde eine Gruppe mit 20 Isolaten gebildet, deren Stämme laut *in silico* Analyse alle drei der, aus der Referenzsequenz PA14 bekannten Pyocinloci tragen. Zusätzlich wurde eine zweite Gruppe mit acht Isolaten gebildet, welche zwei gleiche Loci tragen.

2,5 ml einer LB-Übernachtskultur (37°C, 150 rpm) jedes Isolats wurden in einem 100 ml Kolben vermischt und inkubiert (24 h, 37°C, 150 rpm). Weitere 100 µl der jeweiligen Einzelkulturen wurden separat in frisches LB-Medium inokuliert und ebenfalls inkubiert (24 h, 37°C, 150 rpm). Nach der Inkubation der Mischungen wurden die Bakterien pelletiert und der Überstand in einer Speedvac-Apparatur bei 37°C um die Hälfte des Volumens reduziert. Währenddessen wurden 100 µl der Einzelkulturen auf Agarplatten ausgestrichen und vorinkubiert. Von den konzentrierten Überständen wurden 40 µl auf die vorinkubierten Agarplatten getropft und nach einem kurzen Antrocknen über Nacht bei 37°C inkubiert. Es wurden diejenigen Stämme verworfen, welche beim Auftropfen einer Grup-

pe kein Wachstum zeigten, da diese sensitiv gegenüber den Pyocinen der aufgetropften Mischung waren.

2.24.2 Kompetitive Wachstumsexperimente

Bakterien aus Glycerolstocks der betreffenden Stämme wurde auf LB-Agarplatten ausgestrichen und über Nacht bei 37°C inkubiert. 5 ml LB-Flüssigmedium wurden mit einer Impföse voller Bakterien von den Agarplatten angeimpft und inkubiert (8 h, 37°C, 150 rpm). Nach Bestimmung der OD-Werte wurde sowohl LB- als auch M9-Minimalmedium mit Aliquots der Vorkulturen inokuliert, so dass a) die initiale CFU-Zahl pro Stamm möglichst gleich war und b) die Gesamt-OD der Konkurrenzkultur 0,1 betrug. Insgesamt wurden 6 Replikate in 25 ml Kolben angelegt. Nach 12, 24, 36 und 48 h Inkubation wurden Aliquots der Bakterienmixturen in frisches Medium überführt (initiale $OD_{578nm}=0,1$) und anschließend für drei weitere Tage ohne Medienwechsel weiter inkubiert. Proben für die DNA-Präparation wurden nach 0, 48 und 120 h genommen und die DNA nach einem optimierten Protokoll für gramnegative Bakterien isoliert (Kap. 2.2).

2.24.3 Bestimmung der Anteile der einzelnen Isolate an der jeweiligen Konkurrenz

Zur Berechnung der prozentualen Anteile der einzelnen Stämme an der Konkurrenz wurden die in Kapitel 2.15 ermittelten stammspezifischen SNPs als Marker verwendet. Nach der Vorhersage von SNPs mit SAMtools wurde aus den vollständigen full-vcf-Dateien, welche die Informationen für alle Positionen einer Kartierung enthielten, diejenigen Informationen für alle stammspezifischen SNP-Positionen mit Hilfe eines Skripts extrahiert (extract_positions.pl). Ein SNP wurde nur dann als stammspezifisch akzeptiert, wenn in allen Isolaten an dieser Position mindestens eine Abdeckung von vier vorhanden war, dabei aber nur ein Stamm an der betreffenden Position einen SNP zeigte.

Die präparierte DNA der separaten Probenahmezeitpunkte der Wachstumsexperimente wurde innerhalb der Arbeitsgruppe nach dem adaptierten Protokoll für *P. aeruginosa* mit der SOLiD-Technologie sequenziert (Kap. 2.4.2) und die Reads gegen die jeweilige Referenz kartiert. Für alle als stammspezifisch identifizierten SNPs wurde aus den .bam-Dateien der Fitnessexperimente die Informationen über das Verhältnis zwischen SNP und Wildtyp der Reads extrahiert und daraus der prozentuale Anteil jedes Stamms berechnet. Diese berechneten Werte wurden anschließend anhand linearer Gleichungssysteme, welche sowohl die prozentualen Anteile der singulären Nukleotidaustausche, als auch die Anteile der in zwei, drei und vier Stämmen vorkommenden Mutationen enthielt, geprüft und ggf. korrigiert.

Kapitel 3

Ergebnisse und Diskussion

3.1 Die Entschlüsselung des Klon C NN2 Genoms

Klon C ist der weltweit am weitesten verbreitete Vertreter der *P. aeruginosa* Population (Römling et al., 1994, 2005; Wiehlmann et al., 2007). Komplette sequenzierte *P. aeruginosa* Genome sind zum Zeitpunkt der Niederschrift für neun verschiedene Klone (Pseudomonas Datenbank, März 2015) vorhanden, jedoch nicht für Klon C. In dieser Arbeit wurde das phänotypisch bereits charakterisierte Klon C Isolat NN2 (Cramer et al., 2011) durch DNA- und RNA-Sequenzierung detaillierter untersucht. Das Isolat NN2 wurde 1985 aus der Lunge eines Patienten mit zystischer Fibrose (eng. Cystic Fibrosis, CF) isoliert. NN2 wurde als Klon C Referenzstamm für eine komplette Sequenzierung ausgewählt, da es das erste Klon C Isolat aus den Atemwegen eines bis dahin *P. aeruginosa* freien Patienten war und seine Mikroevolution im CF-Habitat über 25 Jahre untersucht wurde (Cramer et al., 2011). Ferner wurde in früheren Arbeiten für dieses Isolat eine *SpeI-SwaI-PacI-I-CeuI* basierte physikalische Genomkarte erstellt (Schmidt et al., 1996), die Sequenz genomischer Inseln aufgeklärt (Larbig et al., 2002; Klockgether et al., 2004) und phänotypische Marker wie der Morphotyp, die Möglichkeiten der Zellbewegung, die Virulenz sowie die Fitness untersucht (Cramer et al., 2011).

Abbildung 3.1 zeigt zur Übersicht den Arbeitsablauf zur Entschlüsselung des NN2 Genoms mit Hilfe der verschiedenen Sequenzierungstechniken und der verwendeten Softwarepakete.

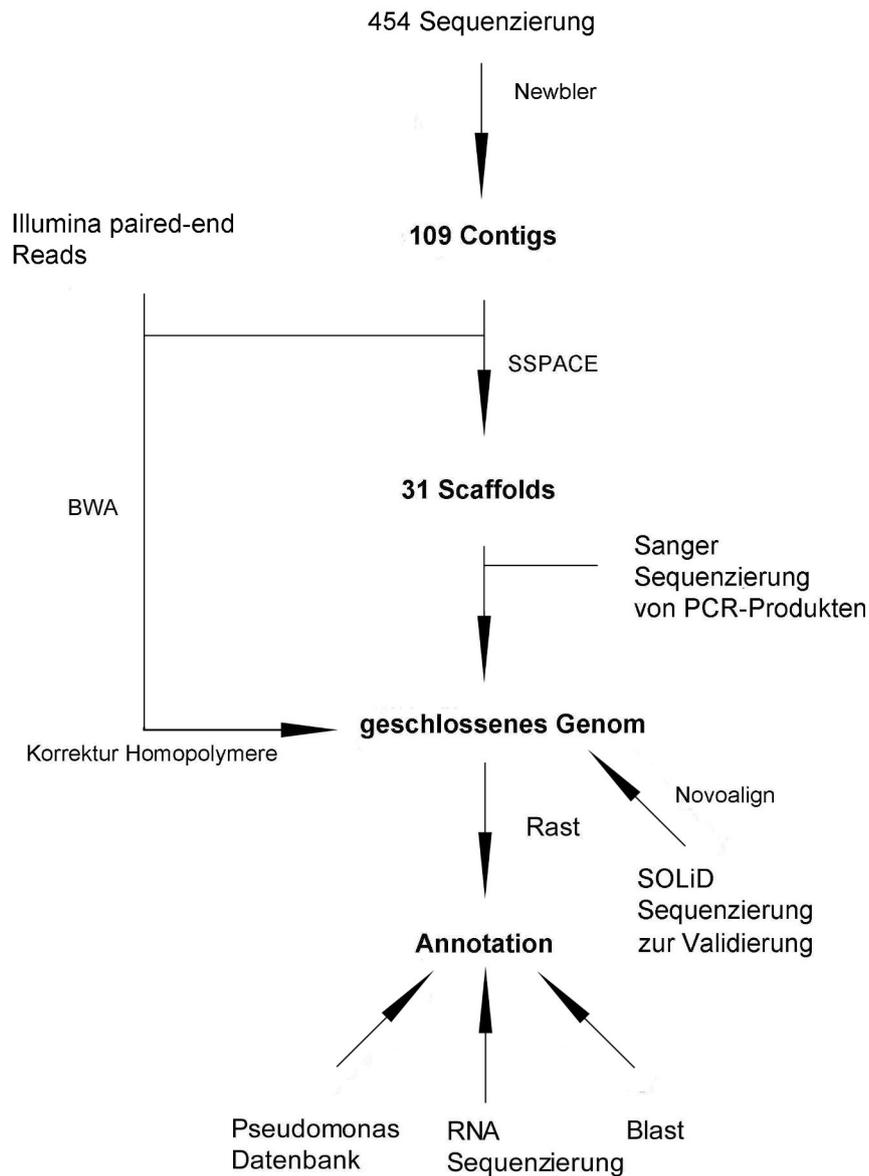


Abbildung 3.1: Das Flussdiagramm zeigt, wie unter Zuhilfenahme der verschiedenen Sequenzierungstechniken und Softwarepakete das NN2 Genom entschlüsselt wurde.

3.1.1 Sequenzierung und *de novo* Assemblierung des NN2 Genoms

Das Isolat NN2 wurde bereits mit 36 bp langen Illumina Reads sequenziert und vor allem auf SNPs hin charakterisiert (Cramer et al., 2011). Jedoch reichte dieser Read-Datensatz nicht für eine verlässliche Sequenzierung des kompletten Genoms aus. Für diese Arbeit wurde NN2 mit der 454 Technik, die Reads mit einer Länge von bis zu 750 bp produziert, und zusätzlich mit der Illumina Technik (paired-end Bibliotheken) sequenziert. Bei Illumina paired-end Bibliotheken werden Enden von größeren DNA-Fragmenten sequenziert, wobei der mittlere Teil unbekannt bleibt (siehe Kap. 2.4.1). Die Kombination der beiden

Sequenzierungstechniken ermöglichte es, die generierten 454 Reads Contigs mit den Informationen aus den Illumina Bibliotheken zu verknüpfen und anschließend verbliebene unbekannte Bereiche mit der Sequenzierung von PCR-Produkten, welche diese Bereiche überspannen, zu entschlüsseln (Abb. 3.2).

Contigs wurden aus überlappenden Reads generiert, wenn die Überlappung eine festgelegte Mindestlänge aufwies. Wurde eine zu kleine Überlappung gewählt, wurden falsche Contigs zusammengesetzt, weil die Wahrscheinlichkeit stieg, dass mehrere Genompositionen die gleiche Sequenz tragen (vgl. Kap. 3.1.8 Tetramerverteilung).

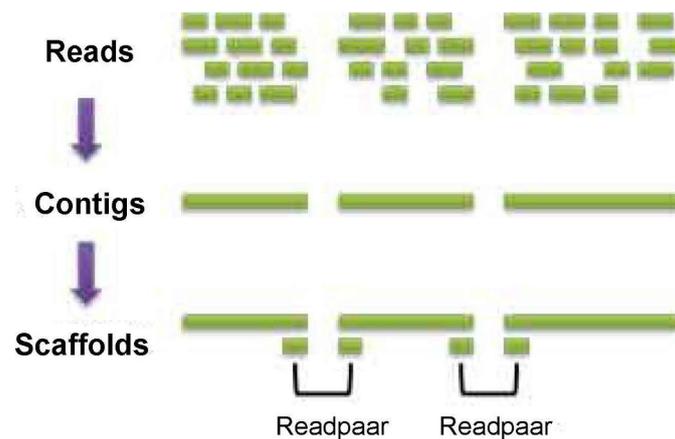


Abbildung 3.2: Das Schema zeigt wie im ersten Schritt aus einer Gruppe überlappender Reads Contigs gebildet werden. Im zweiten Schritt werden die Contigs mit paired-end Readpaaren zu einer konsekutiven Reihenfolge verbunden (Adaptiert von <https://senayoo.com/>).

Die 340.494 454 Reads der NN2 Sequenzierung wiesen einen GC-Gehalt von 65,64% und eine mediane Länge von 401 bp auf (Tab. 3.1 oben). Mit dem Programm Newbler konnten daraus 109 Contigs mit einer Gesamtlänge von 6.872.252 bp (Tab. 3.1 unten) zusammengeführt werden.

Tabelle 3.1: Statistik der 454 Sequenzierung (oben) und der daraus generierten Contigs (unten).

| | |
|-----------------------------|------------|
| Anzahl Reads | 340.494 |
| minimale Readlänge | 40 bp |
| maximale Readlänge | 590 bp |
| durchschnittliche Readlänge | 379 bp |
| mediale Readlänge | 401 bp |
| GC-Gehalt | 65,64% |
| <hr/> | |
| Anzahl Contigs | 109 |
| Basenpaare | 6.872.252 |
| maximale Contiglänge | 495.312 bp |
| minimale Contiglänge | 102 bp |
| durchschnittliche Länge | 63.048 bp |

Die Visualisierung der Contigs gegen das Referenzgenom PAO1 (Stover et al., 2000) mit dem Programm Circoletto lieferte einen ersten Hinweis auf die potentielle Reihenfolge der Contigs (Abb. 3.3) und somit auch über den Genomaufbau von NN2. So ließ sich zum einen feststellen, dass fast alle kurzen Contigs (Contigs 70 - 109) auf die gleiche genomische Region kartiert werden konnten. Vermutlich handelte es sich dabei um einen repetitiven Bereich, welcher entweder Probleme während der 454 Sequenzierung verursachte oder aber den Assemblierungsalgorithmus störte. Zum anderen zeigte sich, dass nahezu das komplette PAO1 Genom durch die Contigs abgedeckt wurde. Einige NN2 Contigs wie Contig 11 zeigten nur in einem kleinen Bereich eine Übereinstimmung mit dem PAO1 Genom (Abb. 3.3). Dieser Befund war ein früher Hinweis darauf, dass das NN2 Genom spezifische Bereiche beinhalten und insgesamt größer als das PAO1 Genom sein würde.

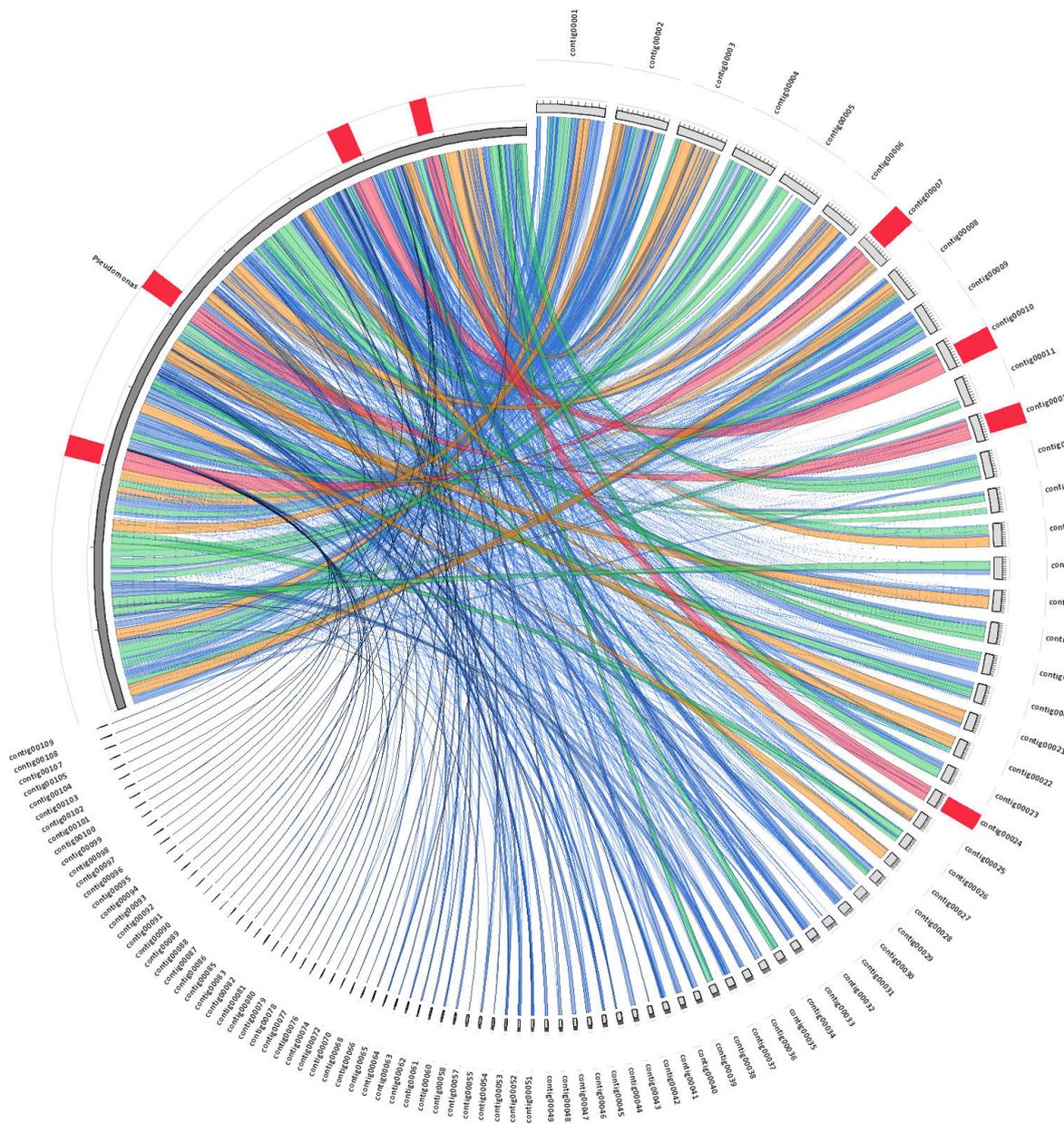


Abbildung 3.3: Visualisierung der 109 454 Contigs gegen das Referenzgenom PAO1.

3.1.2 Sortieren der Contigs zu einer konsekutiven Reihenfolge

Im Wissen über die relativ geringe Anzahl an Contigs und die gute Abdeckung des Referenzgenoms PAO1 konnten die Contigs durch die Implementierung der Illumina paired-end und mate pair Bibliotheken mit dem Programm SSPACE zu Scaffolds zusammengefasst werden. Einen Scaffold kann man als ein Grundgerüst des späteren Genoms bezeichnen, da es die Hauptinformationen über die Reihenfolge der Basen enthält, jedoch noch nicht das komplette Genom darstellt (Abb. 3.2).

Im Fall von NN2 war der Scaffoldingprozess nicht in der Lage, wenigstens eine der Lücken zwischen den Contigs zu schließen. Einzig die Reihenfolge der Contigs relativ zueinander konnte auf diesem Weg bestimmt werden. Dem Contig 65 kam dabei eine besondere Rolle zuteil, da er die Sequenz der vier ribosomalen Operons enthielt. So war es für den Scaffoldingalgorithmus nicht möglich, diesen Contig an alle vier Genompositionen zu reihen. Eine Vervielfachung des Contigs in der Ausgangsdatei konnte dieses Problem ebenfalls nicht lösen. Weiterhin wurde festgestellt, dass der Algorithmus nicht in der Lage war, die große, exklusiv im sequenzierten Referenzstamm PAO1 auftretende Inversion (Stover et al., 2000), welche an den Bruchpunkten zweier ribosomaler Operons auftritt, aufzulösen. Aus diesem Grund wurde Contig 65 aus dem Prozess des Scaffoldings herausgelassen und manuell an die bekannten Positionen der Contigreihenfolge platziert (Stover et al., 2000; Schmidt et al., 1996).

Tabelle 3.2 zeigt im Vergleich die Resultate des Scaffoldings der 3 kb und der 300 bp Bibliotheken. Wie daraus ersichtlich wird, war der Scaffoldingprozess mit den längeren 3

Tabelle 3.2: Vergleich des Scaffoldings der 3 kb mate pair und der 300 bp paired-end Bibliothek.

| | 3 kb | 300bp |
|-----------------------|-----------|-----------|
| Anzahl der Scaffolds | 31 | 91 |
| Basenpaare | 6.888.896 | 6.872.395 |
| maximale Länge [bp] | 2.952.323 | 495.312 |
| minimale Länge [bp] | 118 | 102 |
| durchschn. Länge [bp] | 222.222 | 75.520 |

kb mate pair Bibliotheken sehr viel erfolgreicher. So war die Anzahl der Scaffolds um zwei Drittel kleiner, dabei die durchschnittliche Contiglänge dreifach größer und das längste Contig sogar sechsfach länger, verglichen mit dem Scaffolding der 300 bp paired-end Bibliothek. Dabei galt für die mit der 3 kb Bibliothek generierten Contigs, dass lediglich sieben der 31 Scaffolds aus mehr als zwei Contigs aufgebaut waren. Der längste Scaffold war aus 47 konsekutiven Contigs zusammengesetzt und die Länge der 24 nicht in die Reihenfolge eingeordneten Contigs schwankte zwischen 118 und 366 Basenpaaren. Trotz des Wissens über die Reihenfolge der Contigs zueinander waren innerhalb der Scaffolds jedoch weiterhin unbekannte Bereiche vorhanden. Die abschließende Voraussage über die

Contigreihenfolge und die Größe der Abstände zwischeneinander wurde durch eine manuelle, auf BLAST-Resultaten gegen die Referenzen PAO1 und PA14 basierende Analyse getroffen (Tab. 3.3).

Tabelle 3.3: Übersicht über die manuell vorhergesagte Länge der Lücken zwischen den einzelnen Contigs.

| vorhergesagte Lückengröße [nt] | Anzahl |
|--------------------------------|--------|
| Überlappung | 6 |
| 0 | 42 |
| 1 - 10 | 5 |
| 11 - 100 | 14 |
| 101 - 1000 | 12 |
| unklar | 12 |

3.1.3 Schließen des NN2 Genoms

In den vorangegangenen Kapiteln wurde beschrieben, wie aus einem Set von Reads durch *de novo* Assemblierung und Scaffolding eine Contigreihenfolge generiert wurde. Dabei konnten die Contigs jedoch nicht vereinigt und nur eine Vorhersage über die verbliebenen Lückengrößen zwischen den Contigs getroffen werden.

Um die Contigs zu vereinigen wurden als erstes diejenigen verbunden, deren Vorhersage eine Überlappung der jeweiligen Enden darstellte. Eine Zusammenführung im vorangegangenen Schritt der Assemblierung war in diesen Fällen nicht möglich, da vermutlich die Readabdeckung der jeweiligen Positionen nicht groß genug war oder aber schlechte Qualitätswerte vorlagen.

Für alle Vorhersagen, dass Contigs direkt aneinander anschließen, wurde eine lokale Kartierung von Reads durchgeführt. Dazu wurden die Contigenden *in silico* aneinander gefügt und alle Reads der längeren 300 bp paired-end Bibliotheken mit dem Kartierungsprogramm BWA gegen diesen virtuellen Genomausschnitt kartiert. Alle Kartierungen, welche eine durchgehende Abdeckung zeigten, bestätigten dabei die Annahme der nicht vorhandenen Lücke. War eine deutliche Bruchkante sichtbar, lag entweder ein Contig in der falschen Orientierung vor, oder die Vorhersage über die nicht vorhandene Lücke war falsch (Abb. 3.4). Mit „falscher Orientierung“ ist gemeint, dass der betreffende Contig die Sequenz des Gegen- und nicht des Leitstrangs darstellte, wobei der Leitstrang von NN2 anhand der orthologen *oriC*-Sequenz der Referenzstämme PAO1 und PA14 definiert worden war. Wenn nach der Inversion eines nicht passenden Contigs eine vollständige Abdeckung des untersuchten Bereiches nachgewiesen werden konnte, wurde als bestätigt angesehen, dass die Contigs an dieser Stelle aneinander anschließen. Um diese Ergebnisse weiter zu bestätigen, wurden exemplarisch drei der Bruchbereiche durch PCR amplifiziert und mit der Kettenabbruchmethode nach Sanger sequenziert. Auf diesem Weg konnte in

allen Fällen das *in silico* Ergebnis bestätigt werden.

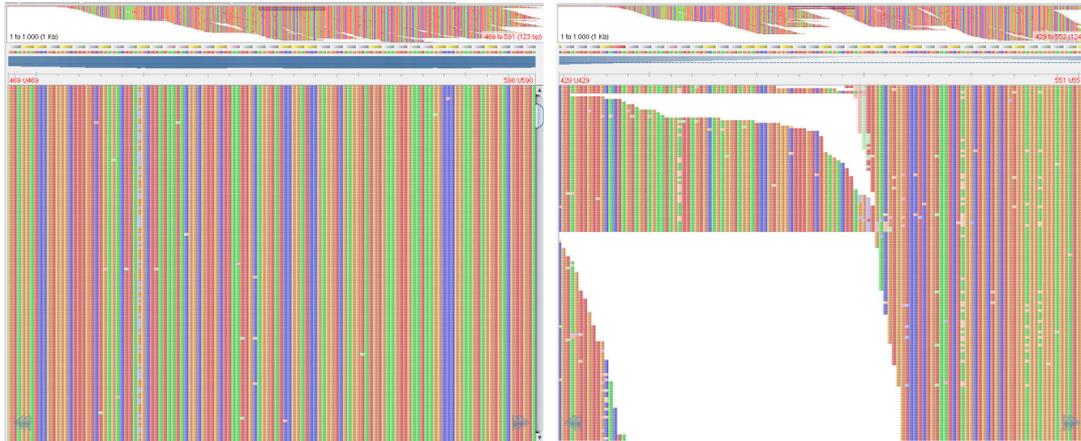


Abbildung 3.4: Darstellung aneinander anschließender (links) und nicht anschließender Contigs nach Validierung durch eine Kartierung der Illumina Reads gegen das zu untersuchende DNA-Fragment. Der Abbruch der durchgängigen Readabdeckung und der Anstieg zu beiden Seiten verdeutlichen die falsche Contigreihenfolge (rechts).

Alle Lücken von, laut Vorhersage, mehreren unbekannt Basen wurden durch ein PCR-Produkt überspannt, das anschließend mit der Kettenabbruchmethode nach Sanger sequenziert wurde.

Durch *in silico* Analysen und Lücken überspannende PCRs konnte die Zahl der Contigs auf zwei reduziert werden. Eine dieser Lücken im Bereich der genomischen Insel PAGI-2 konnte durch Einfügen der bereits bekannten Sequenzdaten für diese Region (Larbig et al., 2002) geschlossen werden.

Die verbliebene unbekannt Sequenz zum Schließen des Genoms war in einem hochrepetitiven Element innerhalb des PAO1-Orthologs PA4625 lokalisiert und ca. 2,6 kb groß. Da PCR-Amplifikationen solcher Bereiche und Sequenzierungen häufig aufgrund der vielen Wiederholungen und den damit einhergehenden Sekundärstrukturen ohne Ergebnis bleiben, konnte die Lücke weder durch das Scaffolding noch durch Sanger-Sequenzierung aufgelöst werden. In einem manuellen *in silico* Ansatz wurden die Bruchpunkte beider Contigenden in den Illuminareads gesucht und, wie im Schema in Abbildung 3.5 dargestellt, verlängert. Dabei wurde in allen Fällen mit mehreren möglichen Varianten die jeweils Häufigste gewählt. Lagen zwei Varianten gleich häufig vor, wurde als weiterer Hinweis die betroffene Base aus den Referenzgenomen PAO1 und PA14 zurate gezogen. Nach mehrmaligem Wiederholen des Vorgangs konnte die Lücke so durch manuelles Ansetzen von Sequenzreads schließlich überspannt und die Genomsequenz geschlossen werden.

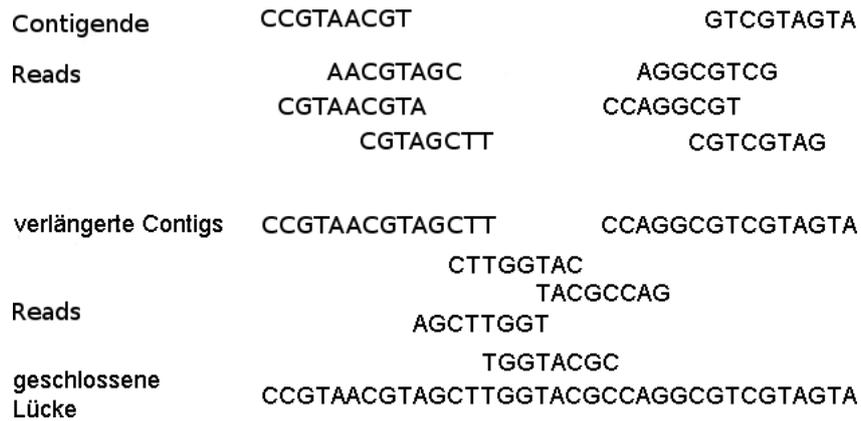


Abbildung 3.5: Schematische Darstellung der Verlängerung der Contigs basierend auf den Illumina Reads. In diesem verkürzten Schema wird der repetitive Charakter der zu überspannenden Lücke vernachlässigt.

Tabelle 3.4 zeigt eine Übersicht über die Größe der geschlossenen Lücken. Vergleicht man diese Einträge mit denen aus Tabelle 3.3 erkennt man, dass sich beide nicht sehr stark unterscheiden. Diese Tatsache ist ein guter Beleg für die Qualität der getroffenen Vorhersagen. Um 454-Sequenzierungsfehler zu eliminieren, wurden alle Illuminareads gegen das geschlossene Genom kartiert und 18 Homopolymerfehler korrigiert. Die finale Genomlänge von NN2 betrug abschließend 6.902.967 bp bei einem GC-Gehalt von 66,12% (Tab. 3.6).

Tabelle 3.4: Übersicht der geschlossenen Lücken zwischen den Contigs.

| Lückengröße [nt] | Anzahl |
|------------------|--------|
| Überlappung | 8 |
| 0 | 47 |
| 1 - 10 | 0 |
| 11 - 100 | 12 |
| 101 - 1000 | 21 |
| > 1000 | 3 |

Um zu überprüfen, inwieweit die NN2 Sequenz mit den früheren Kartierungsergebnissen korreliert, wurde *in silico* nach Restriktionsschnittstellen innerhalb des Genoms gesucht. So wurden unter anderem 47 *SpeI*, 8 *PacI*, 5 *SwaI* und 4 *I-CeuI* Schnittstellen gefunden, was exakt den bereits publizierten Zahlen entsprach (Schmidt et al., 1996). Um die Genomsequenz noch auf einem weiteren Weg zu validieren und gleichzeitig die Genauigkeit der späteren SOLiD Sequenzierung der Isolate der klonalen Komplexe C und PA14 zu überprüfen, wurde die Kompletsequenz mit den SOLiD-Sequenzierdaten für NN2 aus dem Projekt zum intraklonalen Genomvergleich (Kap. 3.2) verglichen. Dabei wurden, bei einer Genomabdeckung von 95,5%, durch die SOLiD-Daten lediglich acht Punktmutationen gegenüber den 454/Illumina-Daten gefunden. Da keine Sequenziermethode eine 100%-ige Genauigkeit garantieren kann, war es unmöglich festzulegen, welche Variante die biologisch korrekte ist und welche einen technischen Fehler darstellte. Tätigt man

die Annahme, dass das NN2 Genom, welches aus 454 und Illumina Reads zusammengesetzt wurde, korrekt war, ergab sich für die SOLiD Sequenzierung eine Genauigkeit von 99,9999%.

3.1.4 Annotation des NN2 Genoms

Bei einer Annotation werden offene Leseraster (engl. open reading frames, ORFs) von potentiell Protein-kodierenden Bereichen in den Genomsequenzen definiert und versucht, den vorhergesagten Genprodukten mögliche Funktionen zuzuweisen. Um auf Ergebnisse anderer sequenzierter Genome zurückzugreifen, wurde die Annotation des NN2 Genoms mit dem schnellen Onlineannotationsserver „RAST“ (Aziz et al., 2008) durchgeführt. Durch manuelle Korrektur wurden die Startcodons aller ORFs, welche ein Ortholog im sehr gut charakterisierten Referenzstamm PAO1 aufwiesen, an das jeweils aus PAO1 bekannte orthologe Startcodon angeglichen. Für die verbliebenen Gene, welche kein PAO1-, jedoch ein PA14-Ortholog aufwiesen, erfolgte eine Angleichung an das entsprechende PA14 Startcodon. Für alle vorhergesagten ORFs ohne gut charakterisiertes Ortholog wurde die ORF-Länge gemäß der RAST-Vorhersage beibehalten. Eine spätere Korrektur der Startcodons wurde auch vorgenommen, wenn die RNA-Sequenzierung die Expression eines Gens über eine abweichende Länge nachwies und ein passendes alternatives Startcodon in der NN2 Sequenz vorhanden war. Ferner wurden Gene zusätzlich annotiert, wenn eine RNA-Expression in einem bislang intergenischen Bereich gefunden wurde und ein ORF von mindestens 150 Basenpaaren identifiziert werden konnte. Insgesamt konnten so durch die automatische Annotation und der anschließenden manuellen Korrektur im NN2 Genom 6601 Gene lokalisiert werden (vgl. Zusatztabelle S1).

Das NN2 Genom hatte mit dem Referenzgenom PAO1 5379 der 5570 Gene gemeinsam (96,5%) und wies bekannte Geninseln bzw. inserierte DNA in 48 der 89 bekannten Regionen genomischer Plastizität (RGP) auf (Klockgether et al., 2011). Dazu zählen z.B. die genomischen Inseln PAGI-1, PAGI-2 (zwei Varianten), pKLC102, PAGI-4 und der LES Prophage 6. Das NN2 Genom kodierte insgesamt für 1222 ORFs, welche nicht in dem PAO1-Referenzgenom vorhanden waren, wovon laut Annotation jedoch 60% lediglich hypothetische Proteine kodierten. Dieser Umstand gab einen Hinweis darauf, dass NN2 ein großes, bislang jedoch kaum erforschtes Set von klon- bzw. stammspezifischen Genen kodierte, welche den globalen Phänotyp des Bakteriums beeinflussen konnten. Diese Gene kodierten für Elemente des horizontalen Gentransfers, der DNA-Reparatur und des Schutzes der genomischen Integrität, Enzyme, Transkriptionsregulatoren und Schwermetallresistenzen. Diese Zusatzeigenschaften der klonspezifischen Gene könnten Ursachen für die dominierende Stellung von Klon C in der Pseudomonaspopulation darstellen. So kodierte beispielsweise das Operon *mukBEF* für ein Kondensin, welches in *E. coli* die korrekte Faltung des Chromosoms und dessen Ringarchitektur ermöglichte (She et al., 2013;

Badrinarayanan et al., 2012). Das DNA-Reparaturgen *radC*, welches als Einzelkopie im Kerngenoms vorlag, wurde durch fünf zusätzlich im akzessorischen Genom zu findende Homologe ergänzt.

Tabelle 3.5: Leserasterverschiebungen in NN2 gegenüber dem Referenzgenom PAO1.

| Funktion | PAO1-Ortholog | Mutation |
|--|---------------|--|
| hypothetisches Protein | PA0055 | SNP führt zu früherem STOP-Codon |
| Nukleosid 2-Deoxyribosyltransferase | PA0145 | SNP führt zu früherem STOP-Codon |
| Phosphoenolpyruvat Phosphotransferase PtsP | PA0337 | Leserasterverschiebung |
| hypothetische Hydrolase | PA0562 | Startregion unterschiedlich zu PAO1 |
| hypothetisches Protein | PA0573 | Startregion unterschiedlich zu PAO1 |
| organisches Lösungsmitteltoleranzprotein OstA | PA0595 | viele Mutationen mit div. Verschiebungen |
| Kalium Transporter | PA0614 | Genende verschieden zu PAO1 |
| 4-Hydroxyphenylazetat-3-Monooxygenase | PA4092 | Genende verschieden zu PAO1 |
| P-Pilus Assemblierungsprotein CupB3 | PA4084 | Startcodon mutiert |
| hypothetisches Protein | PA4060 | Leserasterverschiebung |
| UDP-N-acetylmuramat:L-alanyl-gamma-D-glutamyl-meso-diaminopimelat Ligase | PA4020 | SNP führt zu späterem STOP-Codon |
| Adhärenzfaktor LadS | PA3974 | Leserasterverschiebung |
| Zellantwortregulator WspR | PA3702 | Leserasterverschiebung |
| Chemotaxis Methyltransferase CheR | PA3348 | Leserasterverschiebung |
| hypothetisches Protein | PA3298 | Leserasterverschiebung |
| hypothetisches Protein | PA3198 | SNP führt zu späterem STOP-Codon |
| hypothetisches Protein | PA3076 | verschiedene Indels |
| hypothetisches Protein | PA2928 | Leserasterverschiebung |
| hypothetisches Protein | PA2778 | Startcodon mutiert |
| Porin d. äußeren Membran, koexprimiert mit Pyoverdinbiosynthese Regulon | PA2420 | Startregion unterschiedlich zu PAO1 |
| Protein mit CinA Domäne | PA2141 | Genende verschieden zu PAO1 |
| hypothetisches Protein | PA2036 | Genende verschieden zu PAO1 |
| hypothetisches Protein | PA1914 | Genfragment |
| hypothetisches Protein | PA1889 | Startregion unterschiedlich zu PAO1 |
| hypothetisches Protein | PA1749 | Leserasterverschiebung |
| GMP Synthase - Glutamin-Amidotransferase-Domäne | PA1729 | Leserasterverschiebung |
| Nukleosid-Diphosphat-Zucker-Epimerase | PA1560 | Leserasterverschiebung |
| Cytochrome c Oxidase | PA1554 | Leserasterverschiebung |
| DNA/RNA Endonuklease G, NUC1 | PA1327 | SNP führt zu früherem STOP-Codon |
| periplasmatischer Nitratreduktase Vorläufer | PA1174 | Leserasterverschiebung |
| hypothetisches Membranprotein | PA1134 | Leserasterverschiebung |
| Peptid-Deformylase | PA1122 | SNP führt zu früherem STOP-Codon |
| Protein des generellen Sekretionsweges L | PA0683 | Leserasterverschiebung |
| hypothetisches Protein | PA4360 | Leserasterverschiebung |
| Protein für Stäbchenstruktur MreC | PA4480 | Leserasterverschiebung |
| Phenazinbiosynthese Protein, PhzF ähnlich | PA4716 | Leserasterverschiebung |
| hypothetisches Protein | PA4929 | Leserasterverschiebung |

Das NN2 Genom trägt 36 Verschiebungen des Leserasters bzw. Änderungen in Start- und Stopp-Positionen im Vergleich zum PAO1-Referenzgenom (Tab. 3.5). Diese funktionsbeeinflussenden Mutationen betreffen unter anderem das Chemotaxis-Gen *cheR1*, Gene für die Toleranz gegenüber organischen Lösungsmitteln (*ostA*) sowie Faktoren für den Phenylalaninabbau durch den 3-Hydroxyphenylazetat-Stoffwechselweg. Von den im Rahmen dieser Arbeit sequenzierten 58 sequenzierten Klon C Stämmen (siehe Kapitel 3.2) wies lediglich NN2 exklusiv Leserasterverschiebungen in einer funktionellen Cbb3-1 Cytochromoxidase auf, welche das Wachstum bei hohen Sauerstoffkonzentrationen reguliert (Comolli und Donohue, 2004), sowie in der hybriden Diguanilatzyklase WspR, welche die Biofilmbildung induziert (Huangyutitham et al., 2013).

3.1.5 Bestimmung von Transkriptionsstartpunkten und nicht-kodierenden RNAs (ncRNAs)

Mit Hilfe einer RNA-Sequenzierung lassen sich nicht nur Aussagen über die Anwesenheit von kodierenden und nicht-kodierenden Transkripten treffen, sondern auch deren Genomkoordinaten bestimmen. Dazu wurden zwei Bibliotheken verwendet, eine unbehandelte, welche bis auf die ribosomalen RNAs (rRNAs) alle RNAs enthielt, und eine zweite Bibliothek, in welcher primäre Transkripte angereichert wurden. Als ein primäres Transkript bezeichnet man dabei die erste RNA-Form, welche unmittelbares Ergebnis der Transkription im Rahmen der Genexpression ist.

Um die Transkriptionsstartpunkte der einzelnen ORFs zu untersuchen, wurde eine RNA-Bibliothek mit der Terminator 5'-Phosphat abhängigen Exonuklease (TEX) verdaut. Dieses Enzym degradiert alle Transkripte mit nur einem 5'-Monophosphat, welches in sekundären oder bereits degradierten Transkripten vorkommt. Somit konnte auf primäre Transkripte selektiert und damit einhergehend die 5'-untranslatierten Bereiche (5'-UTRs) der ORFs bestimmt werden. Aus den unbehandelten RNA-Bibliotheken wurde die Expression der einzelnen Gene des Isolats NN2 untersucht, indem die Anzahl der kartierten Reads auf die ORF-Länge und Gesamtanzahl der Reads der Bibliothek normiert wurde (RPKM-Wert, Mortazavi et al. 2008).

Zur Bestimmung der 5'-UTRs und ncRNAs wurden Bakterien bei konstanten Wachstumsbedingungen im Fermenter in nährstoffreichem TSB-Medium bis zur mittleren exponentiellen, sowie zur frühen stationären Phasen kultiviert. Aus der isolierten RNA und den mit TEX verdauten Bibliotheken konnten die Transkriptionsstartpunkte identifiziert werden, da in diesen Bibliotheken primäre Transkripte angereichert wurden. Abbildung 3.6 zeigt exemplarisch die typischen Readakkumulationen der primären Transkripte im 5'-untranslatierten Bereich eines Gens.

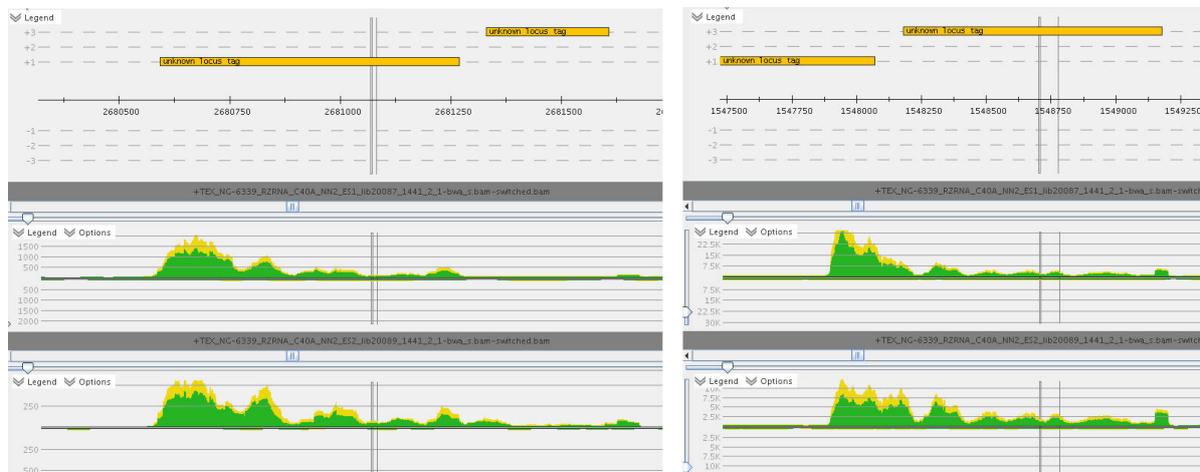


Abbildung 3.6: Beispiele zweier 5'-UTRs in den mit TEX behandelten Bibliotheken. Die Expression der 5'-UTRs konnte jeweils in zwei unabhängigen Proben nachgewiesen werden und zeigt einen reproduzierbaren Fingerabdruck.

NN2 wurde durch Scrollen des gesamten Genoms im Genombrowser ReadXplorer (Hilker et al., 2014) manuell auf die Anwesenheit solcher Readstapel untersucht. Dieser Aufwand war notwendig, da die *in silico* Vorhersage lediglich den Anstieg der Readabdeckung zwischen zwei Basen berechnete. Abbildung 3.6 deutet jedoch an, dass in den vorliegenden Experimenten der Anstieg stetig über mehrere Basen verteilt war und dadurch die gesetzten Kriterien für die Bestimmung der Transkriptionsstartpunkte nicht mehr erfüllt wurden. Auf diese Weise wurden für 1.800 ORFs Transkriptionsstartpunkte ermittelt (Tab. S2). Der Vergleich mit einer PA14 Transkriptom-Studie (Wurtzel et al., 2012) zeigte, dass NN2 weniger Transkripte ohne 5'-UTR (engl. leaderless) synthetisiert. Eine Subgruppe mit kurzen 5'-UTRs einer Länge bis zu 17 bp war hingegen in den exprimierten Genen von NN2 deutlich stärker vertreten als im Transkriptom des Stamms PA14. Das Auftreten von 5'-UTRs mit mehr als 33 bp Länge unterschied sich zwischen PA14 und NN2 nicht. Solche längeren 5'-UTRs wurden für ca. 60% der exprimierten Gene detektiert (Abb. 3.7).

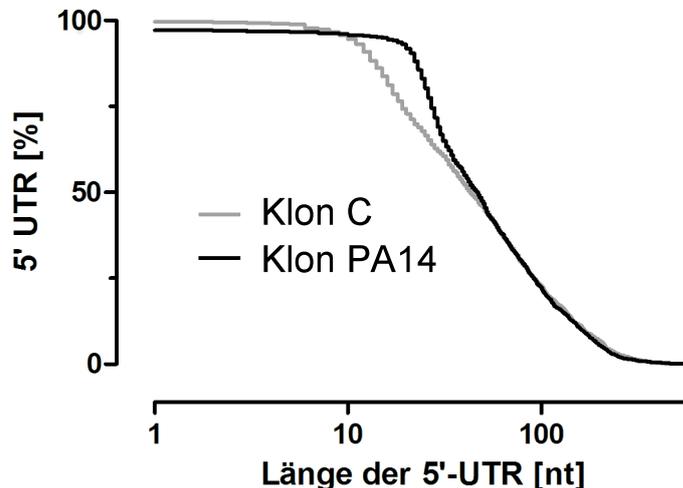


Abbildung 3.7: Längenverteilung der 5'-untranslatierten Regionen in den Stämmen NN2 (grau) und PA14 (schwarz).

Zusätzlich zu den 5'-UTRs fanden sich auch Readakkumulationen innerhalb einzelner ORFs oder in intergenischen Bereichen (Abb. 3.8). Diese Bereiche wurden gesondert auch in den unbehandelten RNA-Sequenzierungen untersucht. Zeigten diese Regionen ebenfalls eine erhöhte Expression, welche nicht das komplette Gen überspannte, wurden diese Fragmente als intragenische nicht-kodierende RNA annotiert.

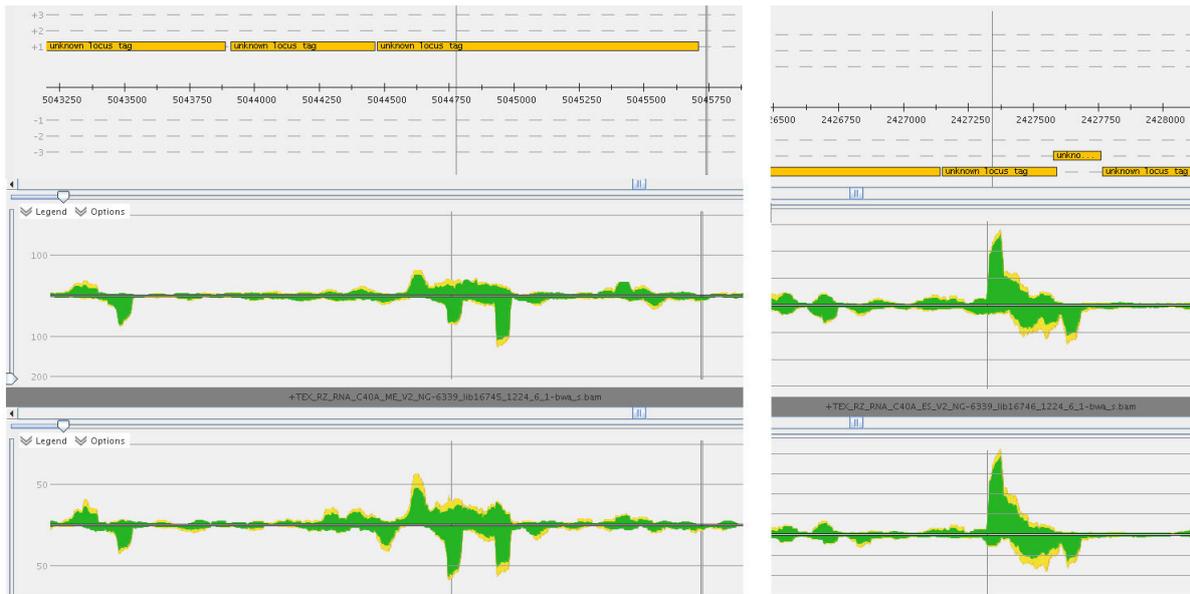


Abbildung 3.8: Beispiele für detektierte nicht-kodierende RNAs. Das linke Bild zeigt mindestens drei, das rechte Bild eine ncRNA. Die rechte ncRNA könnte man auch als 5'-UTR interpretieren, jedoch zeigen Transkriptionsstartpunkte (vgl. Abb. 3.6) im Gegensatz zu ncRNAs keine Readakkumulation auf beiden Strängen in der Kartierung.

Aufgrund einer unvollständigen Enzymreaktion des dUTP Abbaus im Prozess der Probenvorbereitung zur strangspezifischen Sequenzierung (Kap. 2.9.3) war es jedoch nicht möglich, eine Aussage darüber zu treffen, auf welchem Strang im Genom die RNA kodiert ist.

Analog wurden Readstapel in einem intergenischen Bereich als ncRNA identifiziert, wenn kein potentieller kodierender ORF vorhanden war. Auf diese Weise wurden in dieser Arbeit 327 (288 intragenisch, 39 intergenisch) bislang unbekannte ncRNAs annotiert (Tab. 3.6; Tab. S1).

Die intragenischen ncRNAs setzten sich aus einer kleinen Population kurzer (51-87 Nukleotide) Transkripte ($n=46$, Median 66 Nukleotide) und einer größeren Population mit einer Gaussverteilung der Transkriptlängen ($n=242$, Längen 92–604 nt, Median 257 nt) zusammen (Abb. 3.9). 42 Operons und 28 Loci wurden identifiziert, welche zwei bis vier ncRNAs akkumulierten. Intragenische ncRNAs waren in Genen, welche für Motilität, metabolische Enzyme und DNA-assoziierte Prozesse kodieren, überrepräsentiert. Dagegen waren ncRNAs in transkriptionellen Regulatoren kaum exprimiert.

Bereits im Jahr 2012 wurden für den Referenzstamm PAO1 513 ncRNAs beschrieben (Gómez-Lozano et al., 2012). Für 453 von diesen konnte die entsprechende Sequenz im NN2-Genom nachgewiesen werden (88,3%). Jedoch war lediglich für 87 ncRNAs der RPKM-Wert größer als 100 und nur für drei ncRNAs größer als 1.000. Zusätzlich wurde die Sequenz von 165 der 200 in PA14 beschriebenen intergenischen ncRNAs (Wurtzel et al., 2012) in NN2 nachgewiesen. Davon hatten 57 ncRNAs einen RPKM größer 100 und 14 einen RPKM größer 1.000. Die stärkste Expression zeigten die ncRNAs *rnpB*, *crcZ*

und eine transfer-messenger-RNA. Dabei zeigte die RNA der RNase P, *rnpB*, die stärkste Expression, was mit der Bedeutung des regulierten Proteins als essentielles Enzym in der Prozessierung der 5'-Enden stabiler RNA Moleküle wie tRNA und 4.5S RNA (Wu et al., 2014) einhergeht. Hingegen erlaubt eine starke Expression der ncRNA *crcZ* den Bakterien, sich an verschiedene Kohlenstoffquellen zu adaptieren (Sonnleitner et al., 2009). Zusammen mit den in PAO1 bekannten ncRNAs (Gómez-Lozano et al., 2012; Sonnleitner und Haas, 2011), den in PA14 beschriebenen (Wurtzel et al., 2012) und denjenigen dieser Arbeit sind im NN2 Genom mehr als 900 RNAs, wobei jedoch nur für 481 ncRNAs eine Expression nachgewiesen wurde (Tab. 3.6).

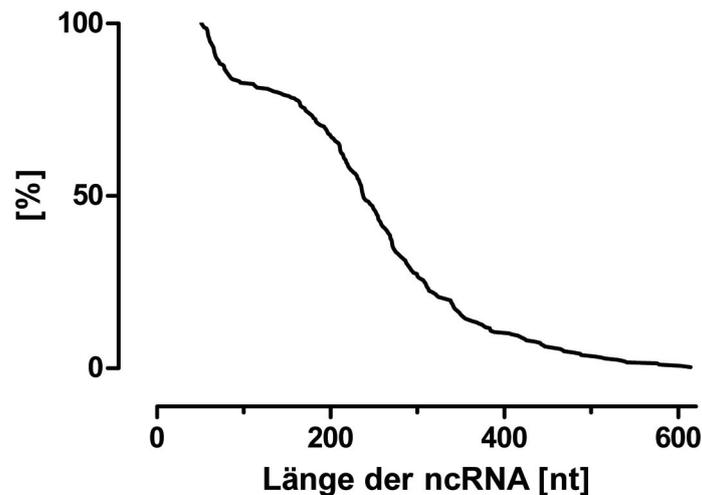


Abbildung 3.9: Kaplan-Meier-Darstellung der Längenverteilung der detektierten intragenischen ncRNAs. Es lassen sich deutlich eine Subpopulation kurzer ncRNAs und eine zweite mit längeren Vertretern unterscheiden.

Tabelle 3.6: Eckdaten des NN2 Genoms.

| | |
|--------------------------|-----------|
| Genomlänge [bp] | 6.902.967 |
| GC-Gehalt [%] | 66,12 |
| Anzahl | |
| ORFs | 6601 |
| aller exp. RNA | 557 |
| neuer intergenischer RNA | 39 |
| neuer intragenischer RNA | 288 |
| tRNA | 62 |
| rRNA | 13 |
| bekannter ncRNA | 154 |
| tmRNA | 1 |

3.1.6 Genomkarte des Isolats NN2

Mit Hilfe der in Kapitel 3.1.3 beschriebenen Referenzsequenz, der in Kapitel 3.1.4 dazu dargestellten Annotationsliste und den in Kapitel 3.1.5 angeführten ncRNAs wurde mit

dem Online-Werkzeug CGView (Stothard und Wishart, 2005) ein Genomatlas von NN2 erstellt (Abb. 3.10). Der äußere Ring zeigt die Verteilung der in dieser Arbeit detektierten

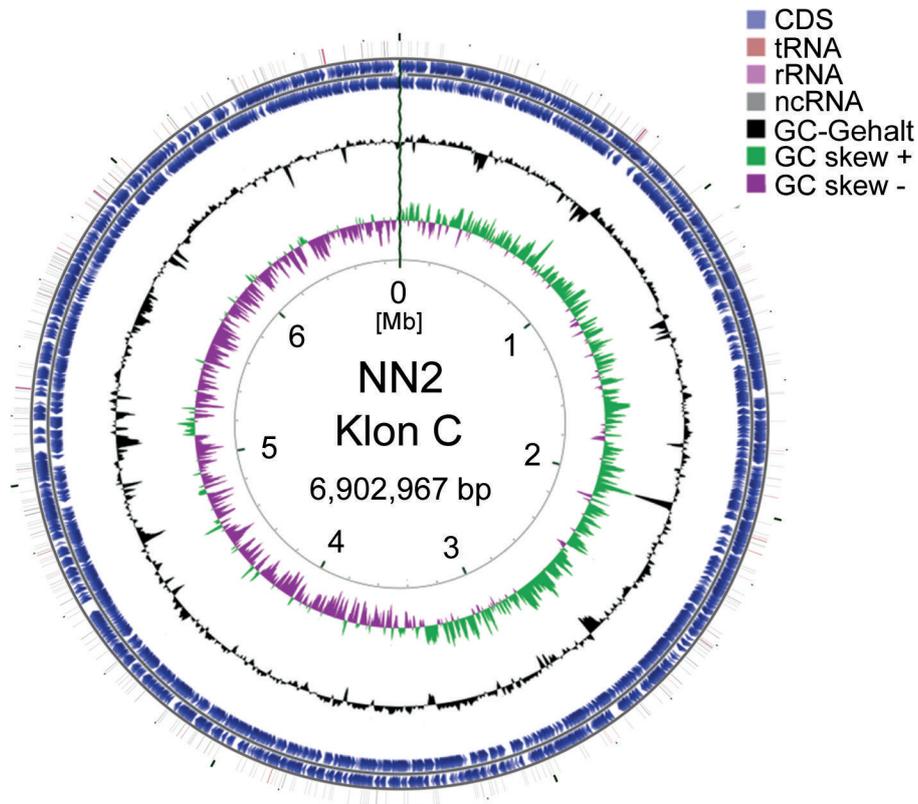


Abbildung 3.10: Genomatlas des NN2 Genoms. Die Ringe zeigen von außen nach innen die RNAs, die annotierten ORFs, den GC-Gehalt, die GC-Nutzung (GC-Skew) und im Inneren die Genomkoordinaten.

ncRNAs. Zusätzlich zeigt der äußere Ring die bereits in der Pseudomonas Datenbank annotierten ncRNAs des Referenzgenoms PAO1. Die ncRNAs sind relativ gleichmäßig über das gesamte Genom verteilt. Dabei gab es nur wenige Genomregionen, die entweder kaum ncRNAs (um Genomposition 0,2 Mbp) oder ein vermehrtes Auftreten (Genomposition 5,6 Mbp) der ncRNAs zeigten. Der zweite Ring von außen zeigt die Lage der ORFs innerhalb des NN2 Genoms. Aufgrund der hohen Anzahl von 6601 ORFs scheinen diese mit nur wenigen Unterbrechungen über das gesamte Genom auf beiden Strängen gleichverteilt angeordnet zu sein. Der dritte Ring stellt die Verteilung des GC-Gehalts in NN2 dar. Deutlich sichtbar ist die geringe Schwankung des GC-Gehalts. Abweichungen stellen die genomischen Inseln und Regionen genomischer Plastizität (RGP) dar. Diese Bereiche des akzessorischen Genoms haben einen geringeren GC-Gehalt als das Kerngenom und können deutlich im Genomatlas abgelesen werden. Der innere Ring des GC skews zeigt die Ungleichverteilung der GC-Nutzung zwischen dem Leit- und dem Folgestrang, welche Chargaff allgemeingültig schon vor mehr als 50 Jahren postuliert hat (Chargaff, 1950). An den kurzen Unterbrechungen wie im Bereich um Genomposition 5,2 Mbp lassen sich deutlich die Einschübe der genomischen Inseln ins Kerngenom visualisieren, von denen

bekannt ist, dass sie sich, wie im GC-Gehalt, auch im GC-Skew vom umgebenden Kerngenom des Wirtsstammes unterscheiden.

Der GC-Gehalt des geschlossenen NN2 Genoms beträgt 66.12%. Im Vergleich dazu betrug der GC-Gehalt der 454 Reads lediglich 65,64% (Tab. 3.1). Hier trat zum wiederholten Male ein typisches Problem von Sequenzierungen zu Tage: die Unterrepräsentation von Sequenzen mit hohem GC-Gehalt. Um diese Nachteile auszugleichen, wurde für die Sequenzierung der GC-reichen *P. aeruginosa* Isolate eine Optimierung des Sequenzierungsprotokolls durchgeführt, damit das gesamte Genom möglichst gleichmäßig abgedeckt werden kann (vgl. Kap. 2.4.2).

3.1.7 Differentielle Genexpressionsanalyse des Isolats NN2

Aus den nicht mit TEX behandelten Bibliotheken, welche, bis auf rRNAs, alle RNA-Moleküle enthielten, ließ sich nach der Kartierung auf das Genom die Anzahl der kartierten Reads jedes Gens bestimmen, welche die Transkriptmenge repräsentierten. Zur Berücksichtigung unterschiedlicher Genlängen wurde dabei aus der Readanzahl und der Genlänge der RPKM-Wert (engl. reads per kilobase per million) errechnet. Darin stellen sich die Genlänge, die Anzahl der Reads innerhalb eines Gens und die Gesamtzahl aller Reads des Datensatzes dar. Anhand des RPKM-Werts lässt sich die differentielle Genexpression sowohl zwischen verschiedenen Isolaten als auch zwischen verschiedenen Probenahmezeitpunkten untersuchen. Es war nicht ausreichend, die kartierten RNA-Reads der ORFs zu zählen, da beispielsweise 1.000 kartierte Reads auf einem 500 bp langem ORF die doppelte Expression im Vergleich zu derselben Kartierung auf einen 1.000 bp langen ORF darstellen. Auch war die Normierung auf die Gesamtanzahl der Reads des Datensatzes notwendig, um verschiedene Proben mit unterschiedlichen Gesamtreadzahlen vergleichen zu können. Im Fall von NN2 ließ sich mit Hilfe des RPKM-Wertes die differentielle Genexpression exprimierter ORFs zwischen der mittleren exponentiellen und der frühen stationären Phase untersuchen und vergleichen.

Dabei zeigte sich überraschenderweise, dass keines der Gene zwischen den beiden Wachstumsphasen differentiell exprimiert wurde. Vergleichsweise dazu wurde in einem anderen Projekt der Arbeitsgruppe das gleiche Experiment für weitere *P. aeruginosa* Isolate (B420, F469) durchgeführt. Dabei weist F469 ebenfalls keine und B420 lediglich ein differentiell exprimiertes Gen auf. Möglicherweise war der Zeitpunkt der Probenahme der frühen stationären Phase zu zeitig gewählt, sodass die Bakterien noch nicht signifikant in Nährstoffmangel geraten waren und somit keine Änderung der Genexpression gefunden werden konnte.

3.1.8 Tetranukleotidzusammensetzung des NN2 Genoms

Die Analyse von Oligonukleotiden ist ein gebräuchlicher Ansatz, um die bakterielle Genomorganisation zu untersuchen (Abe et al., 2003; Pride et al., 2003). Dabei wird in den meisten Fällen die Varianz und das Abweichen von erwarteten Verhältnissen betrachtet. Zur vergleichenden Analyse wurden die 256 möglichen Tetranukleotide in 39 Äquivalenzklassen gruppiert (Baldi und Baisnée, 2000). Innerhalb der Klassen weisen die Tetranukleotide jeweils dieselben Werte der Basenstapelenergie, des Verdrillungswinkels, der Verformbarkeit sowie der Zugänglichkeit und der Biegsamkeit auf.

Zur Untersuchung der Tetranukleotidzusammensetzung des NN2 Genoms wurden die Tetranukleotide in einem um jeweils ein Basenpaar verschobenen Fenster gezählt, den 39 Gruppen zugeordnet und die Gruppen nach absteigender Basenstapelenergie in einer Übersicht angeordnet (Abb. 3.11). Dabei hatten die Vertreter von Gruppe 9 (CGCG, GCGC) die höchste und die von Gruppe 32 (ATAT, TATA) die niedrigste Basenstapelenergie. NN2 zeigte nach der Normalisierung auf den GC-Gehalt (vgl. Kap. 2.14) in neun der 13 Gruppen mit höchster Stapelenergie eine Überrepräsentation der entsprechenden Tetranukleotide. Andererseits waren die Gruppen 20, 21 und 22, welche alle eine relativ hohe Basenstapelenergie gemeinsam haben, kaum besetzt. Gruppen mit Vertretern niedriger Basenstapelenergie sind zumeist unterrepräsentiert. Eine Ausnahme dazu bildeten lediglich die Gruppen 3, 9 und 33.

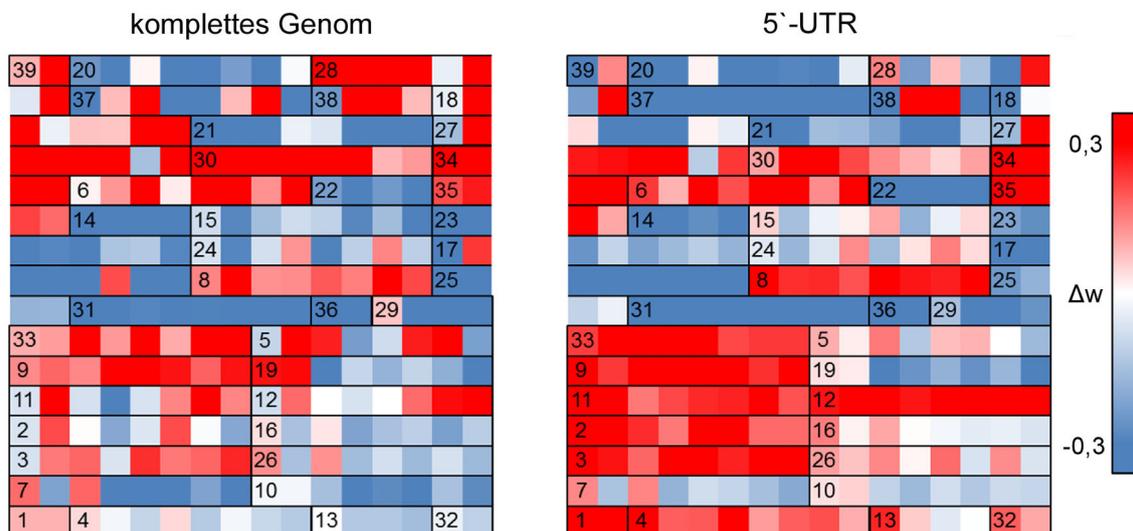


Abbildung 3.11: Vergleich der Tetranukleotidzusammensetzung zwischen dem kompletten NN2 Genom (links) und dem Bereich 250 bp vor und 50 bp nach dem Startcodon von Genen mit experimentell identifiziertem Transkriptionsstartpunkt. Die Tetramere wurden in 39 Gruppen gleicher Basenstapelenergie eingeteilt und absteigend von links oben nach rechts unten sortiert.

Eine vergleichende Analyse wurde auch für die Gruppe von Genen durchgeführt, für die in dieser Arbeit 5'-UTRs identifiziert werden konnten (vgl. Kap. 3.1.5). Dazu wurden für die-

se Gene in den Regionen 250 bp vor und 50 bp nach dem Startcodon die Tetranukleotide gezählt, wobei die tatsächliche Länge der 5'-UTRs vernachlässigt wurde. In diesen Panel zeigte sich eine deutliche Verschiebung hin zu den Tetrameren mit geringerer Basenstapelenergie, wodurch die biologische Funktion dieser genomischen Region verdeutlicht wurde. So führen die geringeren Bindungsenergien zwischen den beiden DNA-Strängen zu einem erleichterten Auflösen des Doppelstranges, was die Anlagerung von Promotoren und der Polymerase im Verlauf der Transkription erleichtert und beschleunigt. Markant für das NN2 Genom ist auch der Fingerabdruck der Gruppen 20, 27, 34 und 36, welcher im Vergleich der globalen und der 5-UTR Analyse nahezu identisch war.

3.1.9 Diskussion der Ergebnisse zur Entschlüsselung des NN2 Genoms

In dieser Arbeit wurde ein Isolat des häufigsten *P. aeruginosa* Klons C, NN2, mit DNA- und RNA-Sequenzierung charakterisiert und auf diese Weise ein geschlossenes Referenzgenom entschlüsselt.

Mit einer Genomgröße von 6.902.967 Basenpaaren wies NN2 innerhalb der Spezies *P. aeruginosa* ein verhältnismäßig großes Genom auf, welches knapp 700 kb größer als der Referenzstamm PAO1 (6,2 Mbp) (Stover et al., 2000) und 400 kb größer als der Referenzstamm PA14 (6,5 Mbp) (Lee et al., 2006) war. Jedoch wurde auch schon von unvollständigen Genomen in *P. aeruginosa* berichtet, welche mehr als 7 Mbp kodierende Sequenz aufwiesen (Hilker et al., 2015). Verglichen mit anderen Bakterien kodiert *P. aeruginosa* ein mittelgroßes Genom. So wurde innerhalb der Prokaryoten das Genom von *Tremblaya princeps* mit einer Länge von weniger als 0,2 Mbp beschrieben (McCutcheon und von Dohlen, 2011). Das andere Extrem stellt das mehr als 13 Mbp große Genom von *Ktedonobacter racemifer* dar (Chang et al., 2011).

Von den 6.601 annotierten ORFs in NN2 kodierte das Genom für mehr als 90% der in PAO1 enthaltenen Gene. Beim Vergleich der funktionellen Kategorien der annotierten Gene in NN2 mit denen der Referenzgenome PAO1 und PA14 zeigten sich die größten Unterschiede bei ORFs ohne Funktionsspezifikation. So waren 48% der ORFs in NN2 als hypothetisch oder putativ beschrieben, 53% in PAO1, jedoch nur 33% in PA14. Die Tatsache, dass NN2 viele Proteine ohne Funktionscharakterisierung kodiert, verdeutlicht, dass es viele unbekannte potentielle Effektoren dafür gibt, warum Klon C eine dominierende Stellung innerhalb der *P. aeruginosa* Population einnimmt. Die zusätzlichen Gene von NN2 im Vergleich zur Referenz PAO1 mit Funktionsannotation kodierten neben Elementen des horizontalen Gentransfers und der Schwermetallresistenzen auch für Gene der DNA-Reparatur und -Integrität (*radC*, *mukBEF*). Zusätzlich zu den funktional charakterisierten ORFs der zusätzlichen Gene sind mehr als 60% bislang nur als hypothetisch

annotiert. Beide Fakten zeigen, dass Klon C ein großes, teilweise kaum erforschtes akzessorisches Genom kodiert, welches bedeutende Komponenten für die bakterielle Lebensweise enthält. Die Ansammlung an zusätzlichen Eigenschaften kann eine mögliche Erklärung für die dominierende Position von Klon C innerhalb der *P. aeruginosa* Population darstellen. Eine Besonderheit von NN2 ist das Exoenzym S (ExoS) (Hauser et al., 1998), welches auch vom Referenzstamm PAO1 kodiert wird (Stover et al., 2000). Von ExoS ist bekannt, dass es die Immunität des Wirts modulieren und Lungenödeme induzieren kann (Sawa, 2014). Im Gegensatz dazu kodiert der Referenzstamm PA14 das Exotoxin U (ExoU). Von ExoU, welches über das Typ-III Sekretionssystem sezerniert wird (Yahr et al., 1996), wurde berichtet, dass es *P. aeruginosa* befähigt, akute Verletzungen des Lungenepithels zu induzieren und nachfolgend eine Sepsis zu induzieren (Allewelt et al., 2000). Die Expression von *exoU* wird dabei durch den Transkriptionsfaktor ExsA reguliert (Finck-Barbancon et al., 1997) und es wurde berichtet, dass ExoU von allen virulenten *P. aeruginosa* kodiert wird (Sawa, 2014). Nach dieser Definition kann NN2 nicht zu den virulenten Stämmen gezählt werden, was auch durch Infektionsexperimente bestätigt werden konnte (Hilker et al., 2015).

Bei der Betrachtung der annotierten Gene gilt zu beachten, dass NN2 96,5% der PAO1-Orthologen kodiert, was für einen hohen konservierten Rahmen spricht. Dieses chromosomale Rückgrat wird in NN2 durch die Anwesenheit mehrerer genomischer Inseln ergänzt. So kodiert NN2 die genomischen Inseln PAGI-2, pKLC102 und die PAGI-2 ähnliche Insel SpB (RGP5), welche alle in tRNA Gene integriert wurden und sich in der Tetranukleotidzusammensetzung von Genen des Kerngenoms unterscheiden (Klockgether et al., 2007). Von den Inseln pKLC102 und PAGI-2 wurde berichtet, dass sie in verschiedenen Taxa vorkommen und eine evolutionär alte Familie von mobilen Elementen bilden (Mohd-Zain et al., 2004).

Wie jedes Isolat einer bakteriellen Population zeigte auch NN2 charakteristische stammspezifische Mutationen und Leserasterverschiebungen. Dazu zählten Leserasterverschiebungen in den Genen einer funktionellen Cbb3-1 Cytochromoxidase und für die Diguanylatzyklase WspR. Die Leserasterverschiebung der Cytochromoxidase ist dahingehend bemerkenswert, als dass dieses Enzym im Verlauf der Infektion einer CF-Lunge ein Markergen für die metabolische Adaptation an dieses neue Habitat darstellt (Eichner et al., 2014). Um zu überprüfen, ob diese Leserasterverschiebung im Verlauf der Kolonisierung des Patienten NN kompensiert oder beibehalten wird, müsste die entsprechende Sequenz in weiteren Isolaten aus diesem Patienten analysiert werden. Bislang wurden jedoch erst drei Isolate auf Einzelnukleotidaustausche hin untersucht (Cramer et al., 2011), nicht aber auf Leseraster verschiebende Insertionen oder Deletionen.

Kurze, nicht-kodierende RNA Moleküle (ncRNAs) werden in der Regel durch Hybridisierungsreaktionen in Mikroarrays oder Northern-Blotting-Analysen detektiert (Hernández-Prieto et al., 2012). Erstmals in *P. aeruginosa* konnten in NN2 intragenische ncRNAs

ermittelt, jedoch technisch bedingt keine Aussage über die Lokalisierung auf einem der Stränge getroffen werden. Von den in PAO1 beschriebenen intergenischen ncRNAs (Gómez-Lozano et al., 2012), konnte die Sequenz von 453 ncRNAs in NN2 nachgewiesen werden. Von den ncRNAs aus PAO1 zeigten nur 3 Vertreter einen RPKM-Wert von mehr als 1.000 und 84 weitere einen RPKM-Wert größer 100. Zusätzlich wurden von den 200 in PA14 beschriebenen intergenischen ncRNAs (Wurtzel et al., 2012) die Sequenz von 165 Vertretern in NN2 nachgewiesen. Davon hatten 57 einen RPKM-Wert größer 100 und 14 einen RPKM-Wert größer 1.000. Die stärkste Expression zeigten die ncRNAs *rnpB*, *crcZ* und eine transfer-messenger-RNA. Dabei zeigte die RNA der RNase P, *rnpB*, die stärkste Expression, was mit der Bedeutung des regulierten Proteins als essentielles Enzym in der Prozessierung von 5'-Enden stabiler RNA Moleküle wie tRNA und 4.5S RNA einhergeht (Wu et al., 2014). Hingegen erlaubt eine starke Expression von *crcZ* dem Bakterium, sich an verschiedene Kohlenstoffquellen zu adaptieren (Sonnleitner et al., 2009). Im Vergleich mit der Literatur konnte im Rahmen dieser Arbeit festgestellt werden, dass die Stämme PAO1, PA14 und NN2 die Sequenz für einen Großteil der ncRNAs teilten. Jedoch zeigte jeder der Stämme eine stammspezifische Signatur in der Expression der ncRNAs, was die Schlussfolgerung nahelegt, dass die Regulation der Genexpression innerhalb von *P. aeruginosa* klon- oder sogar stammspezifisch ist.

3.2 Intraklonaler Genomvergleich der dominanten Klone C und PA14

Der Vergleich von Isolaten oder Individuen einer Spezies ist seit vielen Jahrzehnten Gegenstand von Studien (Park und Williams, 1905). Häufig werden dabei Isolate aus verschiedenen Habitaten oder geografischen Regionen verglichen, welche innerhalb ihrer Spezies verschiedenen klonalen Linien angehören. So wurde auch für *P. aeruginosa* bereits ein Vergleich zwischen Vertretern von 20 unverwandten Klonen durchgeführt (Hilker et al., 2015).

In der vorliegenden Arbeit wurde im Gegensatz zu der bereits untersuchten interklonalen Genomdiversität die intraklonale Diversität der am weitesten verbreiteten Stämme in der *P. aeruginosa* Population (Wiehlmann et al., 2007), Klon PA14 und Klon C, untersucht. Für diese Analyse wurden die Genome von 57 Klon C und 42 Klon PA14 Isolaten auf einer SOLiD-Maschine sequenziert, welche in einem Zeitraum von 30 Jahren aus unabhängigen geografischen Regionen isoliert wurden. Die Isolate stammten sowohl aus der Umwelt, als auch aus akuten Augeninfektionen (Keratitis) und aus Brandwunden. Zusätzlich wurden Isolate sequenziert, die aus chronisch infizierten Atemwegen von Patienten mit Zystischer Fibrose (CF) oder chronisch obstruktiver Lungenerkrankung (engl. chronic obstructive pulmonary disease, COPD) gewonnen wurden (Tab. 3.7).

Tabelle 3.7: Herkunft und Isolationsdatum der sequenzierten Klon PA14 (oben) und Klon C (unten) Isolate. Alle als „erstes Isolat“ bezeichneten Stämmen waren die Erstisolate aus bis dahin Pseudomonas-freien Patienten.

| Isolat | Beschreibung | Habitat | Isolationsdatum |
|--------|---------------------------------|---------|-----------------|
| 4 | erstes PA14 Isolat | CF | unbekannt |
| 6 | erstes PA14 Isolat, Ulm | CF | 11/09/2006 |
| 37 | CF Isolat München | CF | 08/01/2001 |
| 113 | erstes PA14 Isolat, Oldenburg | CF | 12/10/2006 |
| 158 | CF Isolat München | CF | 18/10/2006 |
| 202 | erstes PA14 Isolat, Göttingen | CF | 03/09/2002 |
| 209 | erstes PA14 Isolat, Gifhorn | CF | 02/11/2002 |
| 212 | erstes PA14 Isolat; Bremen | CF | 04/11/2002 |
| 228 | erstes PA14 Isolat, Celle | CF | 16/05/2002 |
| 250 | erstes PA14 Isolat, Osnabrück | CF | 28/03/2002 |
| 324 | erstes PA14 Isolat, Lüneburg | CF | 16/06/2002 |
| 338 | CF Isolat München | CF | 20/12/2006 |
| 487 | erstes PA14 Isolat, Alfeld | CF | 25/02/2005 |
| 110_E8 | Fluss Oker/Aller, Niedersachsen | Umwelt | 10/2007 |
| 12_D2 | Fluss Weser, Niedersachsen | Umwelt | 10/2007 |
| 12_SA1 | Fluss Weser, Niedersachsen | Umwelt | 10/2007 |
| 120B6 | Fluss Oker/Aller, Niedersachsen | Umwelt | 10/2007 |
| 55F6 | Fluss Weser, Niedersachsen | Umwelt | 10/2007 |

| Isolat | Beschreibung | Habitat | Isolationsdatum |
|--------|---|------------|-----------------|
| 61Pf | erstes PA14 Isolat, Leer | CF | 04/1997 |
| 86Pa2 | Aberdeen | CF | 04/08/1997 |
| AL159 | CF Isolat München | CF | unbekannt |
| CF1 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF12 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF13 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF4 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF6 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| K10 | London | Keratitis | 2003/2004 |
| K12 | Liverpool | Keratitis | 2003/2004 |
| K14 | Birmingham | Keratitis | 2003/2004 |
| K2 | Manchester | Keratitis | 2003/2004 |
| K3 | London | Keratitis | 2003/2004 |
| K4 | Bristol | Keratitis | 2003/2004 |
| K6 | Manchester | Keratitis | 2003/2004 |
| K7 | Bristol | Keratitis | 2003/2004 |
| K8 | London | Keratitis | 2003/2004 |
| K9 | Newcastle | Keratitis | 2003/2004 |
| PA14 | Kalifornien, USA | Brandwunde | vor 1990 |
| PT2 | Fluss Ruhr, Mülheim | Umwelt | 1992 |
| RN3 | erstes PA14 Isolat, Oldenburg | CF | 25/02/1987 |
| RZ2 | Krankenhaus Necker, Paris | CF | unbekannt |
| SS8 | sporadischer Patientenklon PA14, Lüneburg | CF | 02/04/1987 |
| WU5 | sporadischer Patientenklon PA14, Bückebug | CF | 08/06/1988 |
| Zw85 | Aberdeen | CF | 1998 |
| 17 | CF Isolat München | CF | 08/09/2006 |
| 36 | COPD, Buffalo, USA | COPD | 16/05/1996 |
| 69 | CF Isolat München | CF | 15/09/2006 |
| 84 | COPD, Buffalo, USA | COPD | 17/04/2001 |
| 87 | CF Isolat München | CF | 20/09/2006 |
| 97 | CF Isolat München | CF | 28/09/2006 |
| 110 | CF Isolat München | CF | 28/09/2006 |
| 115 | erstes Klon C Isolat, Hannover | CF | 08/01/2001 |
| 124 | CF Isolat München | CF | 29/09/2006 |
| 142 | erstes Klon C Isolat | CF | 18/05/2001 |
| 159 | erstes Klon C Isolat, Herford | CF | 18/09/2001 |
| 179 | CF Isolat München | CF | 19/10/2006 |
| 214 | erstes Klon C Isolat, Bremen | CF | 28/08/2002 |
| 237 | erstes Klon C Isolat, Jever | CF | 16/09/2002 |
| 240 | CF Isolat München | CF | 26/10/2006 |
| 343 | CF Isolat München | CF | 22/01/2003 |
| 426 | CF Isolat München | CF | 09/02/2007 |
| 584 | erstes Klon C Isolat, Dortmund | CF | 09/05/2006 |
| 642 | CF Isolat München | CF | 28/03/2007 |
| 747 | CF Isolat München | CF | 26/04/2007 |

| Isolat | Beschreibung | Habitat | Isolationsdatum |
|---------|------------------------------------|-----------------|-----------------|
| 800 | CF Isolat München | CF | 06/06/2007 |
| 10_C1 | Fluss Oker/Aller, Niedersachsen | Umwelt | 10/2007 |
| 10SA1 | Fluss Oker/Aller, Niedersachsen | Umwelt | 07/2008 |
| 10SB1 | Fluss Oker/Aller, Niedersachsen | Umwelt | 07/2008 |
| 110_D4 | Fluss Oker/Aller, Niedersachsen | Umwelt | 10/2007 |
| 12_F8 | Fluss Weser, Niedersachsen | Umwelt | 10/2007 |
| 12_SB1 | Fluss Weser, Niedersachsen | Umwelt | 07/2008 |
| 28PA3 | Innsbruck, Österreich | CF | 27/07/1997 |
| 702_D3 | Fluss Unterweser, Niedersachsen | Umwelt | 10/2007 |
| 902_D1 | Fluss Unterweser, Niedersachsen | Umwelt | 10/2007 |
| AL962 | CF Isolat München | CF | 02/01/2008 |
| B6470 | Ohrinfektion, Heidelberg | acute infection | unbekannt |
| C17 | Langzeitstudie MHH | CF | 2006 |
| CF10 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF11 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF2 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF3 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF5 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF7 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| CF8 | Liverpool Erwachsenenlinik | CF | 2000-2005 |
| K11 | London | Keratitis | 2003/2004 |
| K13 | Bristol | Keratitis | 2003/2004 |
| K15 | Liverpool | Keratitis | 2003/2004 |
| K16 | Liverpool | Keratitis | 2003/2004 |
| K17 | Bristol | Keratitis | 2003/2004 |
| K18 | Manchester | Keratitis | 2003/2004 |
| K19 | London | Keratitis | 2003/2004 |
| K20 | Birmingham | Keratitis | 2003/2004 |
| K5 | Bristol | Keratitis | 2003/2004 |
| NN2 | erstes Klon C Isolat, Burgdorf | CF | 05/12/1985 |
| PJC1 | erstes Klon C Isolat, Gifhorn | CF | 16/07/1985 |
| SG1 | erstes Klon C Isolat, Bückeberg | CF | 22/01/1986 |
| SG17M | Fluss Ruhr, Deutschland | Umwelt | 1993 |
| SG50M | Schwimmbad, Ruhrtal | Umwelt | 1993 |
| SS30 | chron. Atemwegsinfektion, Lüneburg | CF | 01/1989 |
| W5Aug28 | Woluwe, Fluss, Belgien | Umwelt | 2001 |
| WI1 | erstes Klon C Isolat, Wolfsburg | CF | 09/01/1985 |
| WU2 | erstes Klon C Isolat, Bückeberg | CF | 10/11/1987 |

3.2.1 Erhalt des chromosomalen Grundgerüsts

Der Erhalt des chromosomalen Grundgerüsts eines klonalen Komplex gibt an, ob die Nukleotidsequenz im Vergleich zwischen einzelnen Isolaten tendenziell eher gleich oder stark unterschiedlich ist. Um einen Überblick über den Erhalt des chromosomalen Rahmens

in *P. aeruginosa* zu bekommen, wurde aus den Sequenzierungen der einzelnen Isolate in paarweisen Vergleichen die Länge der syntenischen Segmente mit 100%-iger Sequenzidentität des Kerngenoms zwischen Stammpaaren von entweder Klon C, Klon PA14 oder klonal unverwandten Stämmen (Hilker et al., 2015) berechnet. Als Referenzen wurden dazu die konservierten Kerngenome der komplett sequenzierten Referenzstämmen NN2 (Klon C Isolate), PA14 (Klon PA14 Isolate) und PAO1 (nicht-klonale Isolate) herangezogen. In $\binom{n}{2}$ Vergleichen (Klon C: n=58; Klon PA14: n=42; unverwandte Klone: n=21) wurden die Fragmentlängenpaare des Kerngenoms ermittelt. Abbildung 3.12 zeigt die normalisierte Frequenzverteilung der Fragmentlängenpaare, welche eine identische Sequenz enthielten.

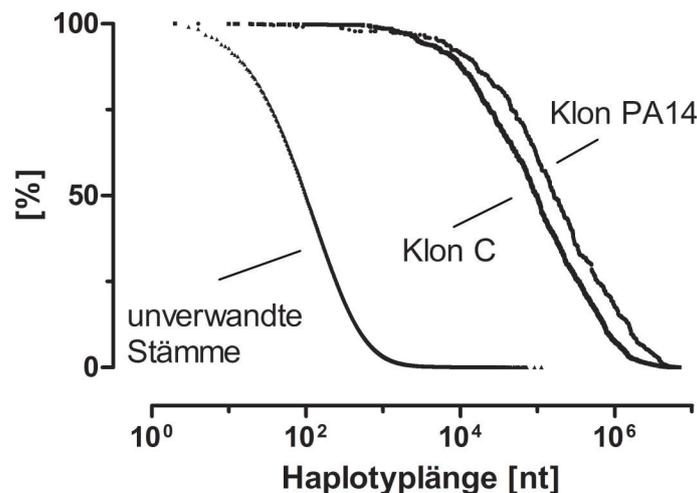


Abbildung 3.12: Normalisierte Verteilung der Fragmentlängenpaare aus paarweisen Vergleichen der syntenischen Sequenz des Kerngenoms von 58 Klon C (n = 33.800 Fragmentlängenpaare), 42 Klon PA14 (n = 9.510) und 21 klonal nicht verwandten *P. aeruginosa* Stämmen (n = 3.779.224).

Beim Vergleich der Fragmentlängenpaare klonal nicht verwandter Stämme (Hilker et al., 2015) ließ sich feststellen, dass hauptsächlich kurze identische DNA-Stücke einer medianen Länge von 100 Nukleotiden vorhanden waren. Im Gegensatz dazu zeigte der Vergleich zweier Klon C oder Klon PA14 Stämme große Abschnitte identischer DNA zwischen den Stämmen. Die mediane Länge der Sequenzidentität zweier Klon C Isolate (99 kb) wurde dabei von den PA14 Stämmen (Median 163 kb) sogar noch übertroffen (Abb. 3.12). Somit waren die Fragmentlängenpaare innerhalb klonaler Komplexe um bis zu 1000-fach länger als in unverwandten Stämmen. Diese Daten gaben einen Hinweis darauf, dass der chromosomale Rahmen eines klonalen Komplexes innerhalb der *P. aeruginosa* Population stark konserviert ist und innerhalb eines klonalen Komplexes nur wenige Mutationen im Kerngenom auftreten. Diese Tatsache wurde auch durch die geringe Mutationsfrequenz innerhalb des Kerngenoms (Abb. 3.19) und die sternförmige Topologie der phylogenetischen Bäume bestätigt (Abb. 3.13). Andererseits zeigten unverwandte klonale Komplexe, deren Haplotypen im Median kürzer als die Länge eines durchschnittlichen Gens waren (NN2: 948 bp), einen unbegrenzten, rekombinationsgetriebenen Genaustausch innerhalb

der bakteriellen Population. Die Anzahl der gefundenen Fragmentlängenpaare in 21 unverwandten Klonen (vier Millionen Paare) im Vergleich zu den Stämme der klonalen Komplexe C und PA14 (C: 33.800 Fragmentlängenpaare, PA14: 9.510) verdeutlichte das Kopplungsgleichgewicht zwischen den unverwandten Klonen (Tab. S3).

Im Folgenden wurden die SNPs des konservierten Kerngenoms, welche für die Fragmentlängenanalyse ausgewählt wurden, genutzt, um das intraklonale Verwandtschaftsverhältnis der Stämme in einem phylogenetischen Baum darzustellen (Abb. 3.13). Deutlich erkenn-

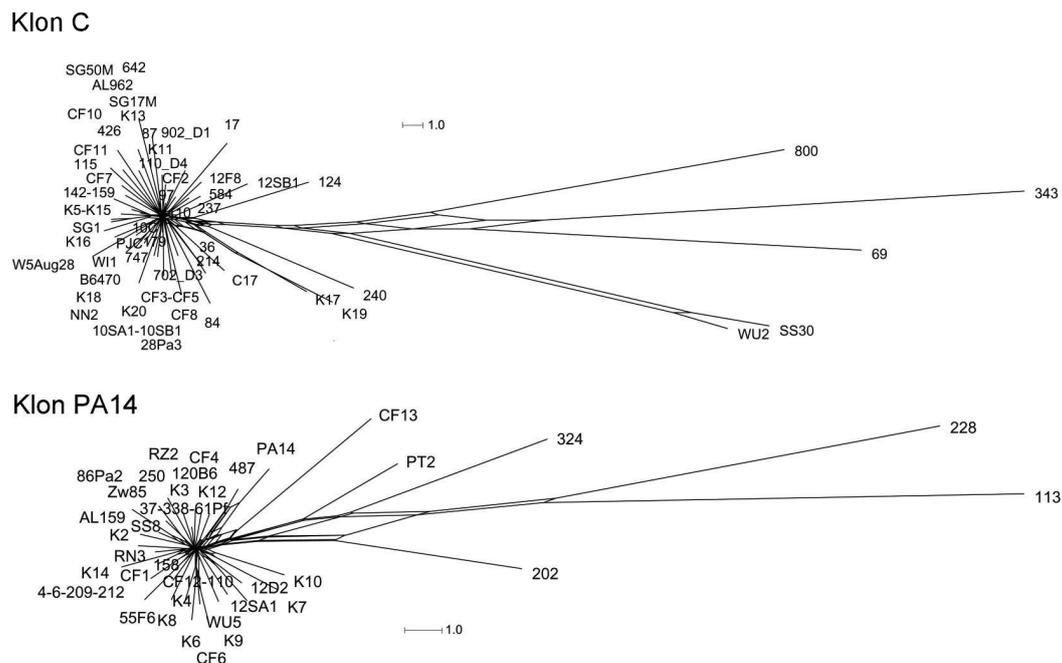


Abbildung 3.13: SNP-basierte phylogentische Bäume der klonalen Komplexe C (oben) und PA14 (unten). Die sternähnliche Topologie in den Bäumen beider Klone und die Tatsache, dass sich nur in seltenen Fällen mehrere Isolate einen Ast teilen, zeigt, dass die Stämme aufgrund weniger, unabhängiger *de novo* Mutationen von einem gemeinsamen Vorfahren abstammen.

bar war, dass sich für beide klonalen Komplexe die Topologien der Baumstrukturen sehr ähnelten. Die Mehrzahl der Stämme formte eine sternförmige Struktur nah verwandter, unabhängiger Stämme. Da die Mehrheit der Isolate auf separaten Ästen des phylogenetischen Baumes lokalisiert war, konnte davon ausgegangen werden, dass sie durch wenige unabhängige *de novo* Mutationen von einem gemeinsamen Vorfahren abstammen. Dabei waren die Stämme unabhängig sowohl hinsichtlich ihres geografischen Ursprungs als auch des Probenahmezeitpunktes. In seltenen Fällen waren mehrere Isolate innerhalb der sternähnlichen Struktur des phylogenetischen Baums auf gleichen Ästen lokalisiert. Von den PA14-Stämmen teilten sich je einmal zwei, drei und vier Stämme und von den Klon C Stämmen vier mal zwei Isolate einen Ast. Dadurch wurde gezeigt, dass einige wenige Stämme innerhalb der untersuchten Kohorte mit einer hohen Genomkonservierung in ihrem chromosomalen Aufbau fast identisch sind. Hingegen wurden fünf (Klon C) bzw. sechs (Klon PA14) Stämme anhand der Einzelnukleotidaustausche des Kerngenoms als

Außenseiter identifiziert.

3.2.2 Intraklonale Genomdiversität des Kern- und des akzessorischen Genoms

Das Kerngenom enthält in der Regel Gene des Stoffwechsels und der Replikation und weist nur wenige Mutationen zwischen Individuen eines Organismus auf (Valot et al., 2015). Das akzessorische Genom hingegen kodiert für Gene des Molekültransports oder des Stoffwechsels sekundärer Metaboliten und zeigt dabei eine hohe Variabilität (Kung et al., 2010). Um die Diversität innerhalb eines klonalen Komplexes zu untersuchen, wurden sowohl die Einzelnukleotidaustausche (SNPs) gegenüber der jeweiligen Referenz NN2 und PA14, als auch das genetische Repertoire von Regionen hoher Genomplastizität untersucht.

Die mediane intraklonale Genomdiversität betrug 0,372‰ (Bereich von 0,002‰ - 0,789‰) in Klon C und 0,024‰ (0,008‰ - 0,973‰) in Klon PA14 Stämmen (Tab. 3.10). Untersuchte man nur das Kerngenom, welches in allen *P. aeruginosa* konserviert ist, konnte eine intraklonale Sequenzdiversität von $8 \cdot 10^{-6}$ für Klon C und $2 \cdot 10^{-5}$ für Klon PA14 berechnet werden. Wie in Tabelle 3.10 dargestellt, verhalten sich, wenn man die bekannten Bereiche mit einer hohen Mutationsfrequenz (pKLC102-, SpB-Insel (RGP5) in Klon C; PAPI-1-Insel in Klon PA14) aus der Analyse ausschließt, beide klonalen Komplexe im Median sehr ähnlich. Anhand der breiten Streuung der Anzahl an Einzelnukleotidaustauschen lassen sich allerdings Abweichungen in beide Richtungen erkennen. Auch bei Anwesenheit der genomischen Inseln pKLC102 und der PAPI-2 ähnlichen RGP5 in den Klon C Isolaten gab es Stämme, welche mit einer, zu den PA14 Stämmen vergleichbar geringen Anzahl an SNPs ausgestattet waren. Folglich liegen die beiden genomischen Inseln entweder in verschiedenen Varianten innerhalb des klonalen Komplexes vor oder unterliegen aufgrund ihrer kodierenden Bereiche (u.a. Typ-VI-Sekretionssysteme) einer hohen Mutationsfrequenz. Diese Tatsache lässt sich damit erklären, dass die genomischen Inseln PAPI-2 und PAPI-1 als mobile Genominselfen über Gattungsgrenzen hinweg ausgetauscht werden können (Klockgether et al., 2007).

Das Verhältnis von intra- zu intergenischen bzw. synonymen zu nicht-synonymen SNPs mit und ohne die Einzelnukleotidaustausche der beiden Inseln zeigte, dass viele Mutationen hauptsächlich innerhalb von ORFs lokalisiert waren und nur in wenigen Fällen zu Aminosäureaustauschen führten (Tab. 3.10). Somit war die Diversität innerhalb der klonalen Komplexe im Vergleich zu unverwandten Stämmen (Hilker et al., 2015) um mehr als das 100-fache geringer. Anders formuliert war das Kerngenom innerhalb eines klonalen Komplexes hoch konserviert und unterschied sich nur durch ein paar Dutzend SNPs zwischen den einzelnen Stämmen.

Anhand der SNP-Statistiken (Tab. 3.11, 3.12) ließen sich für die klonalen Komplexe eini-

ge Habitat-spezifische Charakteristika erkennen. So wurde für jedes Klon C Isolat einer akuten Infektion mehr als 850 Einzelnukleotidaustausche gefunden. Selbst bei Vernachlässigung der Bereiche hoher Mutationsfrequenz wiesen die Isolate der akuten Infektion in Klon C im Median die höchste Zahl an SNPs auf. Zusätzlich wiesen nur die Isolate der akuten Infektion mehr synonyme als nicht-synonyme Mutationen auf. Im Gegensatz dazu hatten die Umweltisolate des PA14 Komplexes die größte mediane Zahl an SNPs akkumuliert. Die Analyse der Isolate chronischer Infektionen ergab, dass intragenische Mutationen um das 3,5-fache häufiger auftreten als intergenische Einzelnukleotidaustausche. Diese kamen in Isolaten aus akuten Infektionen und der Umwelt mehr als sechsfach weniger vor. Im Vergleich zu Isolaten der akuten Infektion und der Umwelt kodierten die chronischen Isolate mehr Aminosäureaustausche. Blendete man alle Einzelnukleotidaustausche der PAPI-1 Insel aus der Analyse aus, näherten sich, wie in den Isolaten des klonalen Komplexes C auch in den Stämmen des Klons PA14 die Mutationszahlen der drei Habitats an. Dabei gilt zu beachten, dass lediglich sechs Umweltisolate untersucht wurden und schon wenige Stämme mehr den Gesamteindruck verändern können.

Tabelle 3.8: Übersicht über die Abdeckung des Referenzgenoms PA14 und der SNP-Charakteristika innerhalb der untersuchten PA14 Isolate.

| Isolat | Abdeckung | | Verhältnis | | | Verhältnis [§] | | | |
|--------|-----------|-------|-----------------------|-------------------------|-------------------|-------------------------|-----------------------|-------------------------|-------------------|
| | PA14 [%] | SNPs | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ | SNPs [§] | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ |
| 4 | 97,54 | 159 | 3,08 | 0,58 | 1,27 | 157 | 3,03 | 0,59 | 1,28 |
| 6 | 96,50 | 142 | 3,30 | 0,60 | 1,49 | 139 | 3,21 | 0,63 | 1,48 |
| 37 | 94,90 | 146 | 2,84 | 0,59 | 1,39 | 142 | 2,74 | 0,60 | 1,41 |
| 113 | 91,73 | 519 | 5,92 | 0,71 | 5,57 | 511 | 6,00 | 0,72 | 5,47 |
| 158 | 99,17 | 1.587 | 7,97 | 2,46 | 2,68 | 1.573 | 7,89 | 2,48 | 2,70 |
| 202 | 81,46 | 132 | 7,25 | 0,58 | 4,28 | 131 | 4,04 | 0,59 | 4,24 |
| 209 | 85,64 | 99 | 2,30 | 0,92 | 2,09 | 98 | 2,27 | 0,94 | 2,06 |
| 212 | 88,62 | 76 | 2,04 | 0,65 | 2,17 | 75 | 2,00 | 0,67 | 2,13 |
| 228 | 85,70 | 355 | 5,12 | 0,49 | 6,89 | 338 | 5,15 | 0,43 | 7,67 |
| 250 | 89,16 | 240 | 2,93 | 1,13 | 2,16 | 240 | 2,93 | 1,13 | 2,16 |
| 324 | 92,47 | 226 | 4,65 | 0,36 | 4,38 | 219 | 4,48 | 0,36 | 4,34 |
| 338 | 97,79 | 144 | 3,80 | 0,50 | 1,36 | 141 | 3,70 | 0,52 | 1,35 |
| 487 | 98,85 | 116 | 3,46 | 0,53 | 1,32 | 113 | 3,35 | 0,55 | 1,31 |
| 110_E8 | 95,18 | 2.160 | 10,74 | 2,32 | 1,95 | 180 | 1,86 | 0,54 | 1,47 |
| 12_D2 | 93,14 | 136 | 2,89 | 0,60 | 1,78 | 132 | 2,88 | 0,63 | 1,81 |
| 120B6 | 95,59 | 165 | 4,00 | 0,57 | 1,62 | 162 | 3,91 | 0,59 | 1,61 |
| 12SA1 | 92,67 | 127 | 3,23 | 0,52 | 1,44 | 125 | 3,17 | 0,53 | 1,40 |
| 55F6 | 93,51 | 2.399 | 15,54 | 2,77 | 1,89 | 161 | 3,74 | 0,53 | 1,64 |
| 61Pf | 94,27 | 132 | 3,00 | 0,74 | 1,75 | 128 | 2,88 | 0,79 | 1,72 |
| 86PA2 | 97,31 | 153 | 2,40 | 0,57 | 1,47 | 149 | 2,31 | 0,60 | 1,44 |
| AL159 | 92,31 | 94 | 3,48 | 0,52 | 1,54 | 94 | 3,48 | 0,52 | 1,54 |
| CF1 | 94,92 | 138 | 3,45 | 0,78 | 1,71 | 135 | 3,35 | 0,82 | 1,70 |
| CF12 | 92,23 | 2.096 | 9,43 | 2,39 | 1,93 | 148 | 2,02 | 0,59 | 1,51 |

| Isolat | Abdeckung | | Verhältnis | | | Verhältnis [§] | | | |
|--------|-----------|-------|-----------------------|-------------------------|-------------------|-------------------------|-----------------------|-------------------------|-------------------|
| | PA14 [%] | SNPs | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ | SNPs [§] | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ |
| CF13 | 84,56 | 1.288 | 10,50 | 2,41 | 1,98 | 115 | 1,61 | 0,37 | 1,45 |
| CF4 | 93,13 | 2.041 | 10,94 | 2,53 | 1,95 | 179 | 1,56 | 0,82 | 1,49 |
| CF6 | 81,11 | 1.762 | 10,37 | 2,60 | 1,96 | 102 | 1,13 | 0,74 | 1,76 |
| K10 | 93,76 | 479 | 9,19 | 2,66 | 2,15 | 476 | 9,13 | 2,73 | 2,15 |
| K12 | 92,09 | 120 | 2,53 | 0,83 | 1,93 | 120 | 2,53 | 0,83 | 1,93 |
| K14 | 95,51 | 127 | 3,88 | 0,60 | 1,27 | 126 | 4,04 | 0,60 | 1,25 |
| K2 | 95,38 | 120 | 3,80 | 0,64 | 1,61 | 118 | 3,72 | 0,63 | 1,57 |
| K3 | 93,45 | 128 | 3,92 | 0,85 | 1,72 | 125 | 3,81 | 0,87 | 1,72 |
| K4 | 97,53 | 1.922 | 11,56 | 2,36 | 1,86 | 171 | 1,76 | 0,51 | 1,44 |
| K6 | 99,34 | 2.033 | 11,55 | 2,34 | 1,86 | 185 | 2,19 | 0,51 | 1,43 |
| K7 | 98,57 | 502 | 9,04 | 2,40 | 2,06 | 499 | 8,98 | 2,45 | 2,06 |
| K8 | 92,83 | 132 | 3,55 | 0,72 | 1,64 | 129 | 3,45 | 0,75 | 1,63 |
| K9 | 91,42 | 1.224 | 11,36 | 3,11 | 2,26 | 401 | 5,57 | 2,95 | 3,05 |
| PT2 | 92,93 | 6.364 | 10,20 | 3,87 | 2,83 | 4.275 | 8,13 | 4,59 | 3,41 |
| RN3 | 95,80 | 265 | 5,97 | 1,44 | 1,91 | 262 | 5,89 | 1,49 | 1,91 |
| RZ2 | 92,93 | 166 | 3,61 | 0,83 | 2,25 | 140 | 3,52 | 0,73 | 1,92 |
| SS8 | 93,93 | 115 | 3,42 | 0,65 | 1,74 | 112 | 3,87 | 0,65 | 1,67 |
| WU5 | 84,92 | 55 | 2,67 | 0,48 | 2,67 | 55 | 2,67 | 0,48 | 2,67 |
| Zw85 | 95,75 | 140 | 3,24 | 0,55 | 1,91 | 136 | 3,12 | 0,58 | 1,57 |

Die Abdeckung gibt den mit der SOLiD Sequenzierung abgedeckten Anteil des Referenzgenoms PA14 an.

[§] ohne die genomische Insel PAPI-1. $\frac{intra}{inter}$ ist das Verhältnis intra- zu intergenischen SNPs. $\frac{syn}{nicht-syn}$ ist das Verhältnis synonymmer zu nicht-synonymen SNPs. $\frac{T_s}{T_v}$ ist das Verhältnis von Transitionen zu Transversionen.

Tabelle 3.9: Übersicht über die Abdeckung des Referenzgenoms NN2 und der SNP-Charakteristika innerhalb der untersuchten Klon C Isolate.

| Isolat | Abdeckung | | Verhältnis | | | Verhältnis* | | | |
|--------|-----------|-------|-----------------------|-------------------------|-------------------|-------------|-----------------------|-------------------------|-------------------|
| | NN2 [%] | SNPs | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ | SNPs* | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{T_s}{T_v}$ |
| 17 | 92,20 | 2.928 | 11,62 | 2,90 | 1,81 | 77 | 5,42 | 0,44 | 2,08 |
| 36 | 94,83 | 2.568 | 13,11 | 3,24 | 1,76 | 58 | 3,46 | 0,41 | 4,80 |
| 69 | 94,66 | 1.987 | 12,07 | 2,21 | 2,30 | 536 | 9,94 | 0,63 | 8,08 |
| 84 | 93,77 | 2.961 | 11,49 | 3,22 | 1,68 | 529 | 8,98 | 2,78 | 1,65 |
| 87 | 93,96 | 2.274 | 12,07 | 3,15 | 1,65 | 112 | 4,33 | 0,54 | 2,39 |
| 97 | 94,46 | 3.114 | 12,42 | 3,05 | 1,83 | 42 | 3,67 | 0,43 | 2,50 |
| 110 | 96,65 | 2.766 | 13,04 | 3,12 | 1,65 | 41 | 4,13 | 0,50 | 2,73 |
| 115 | 93,26 | 2.360 | 8,37 | 3,66 | 1,65 | 736 | 4,49 | 2,91 | 1,94 |
| 124 | 94,54 | 2.938 | 12,79 | 2,88 | 1,80 | 79 | 4,64 | 0,38 | 3,39 |
| 142 | 94,36 | 1.336 | 11,49 | 3,64 | 1,66 | 73 | 9,43 | 1,54 | 2,65 |
| 159 | 89,73 | 929 | 10,61 | 3,85 | 1,62 | 29 | 4,80 | 0,85 | 1,64 |
| 179 | 92,56 | 2.728 | 13,06 | 3,05 | 1,80 | 43 | 7,60 | 0,46 | 2,91 |
| 214 | 89,96 | 2.226 | 11,37 | 3,41 | 1,83 | 44 | 3,40 | 0,55 | 3,89 |
| 237 | 81,20 | 1.642 | 9,53 | 2,94 | 1,81 | 28 | 3,67 | 0,47 | 2,11 |
| 240 | 95,32 | 3.350 | 8,94 | 3,39 | 1,70 | 1.411 | 5,95 | 3,01 | 1,84 |
| 343 | 90,22 | 3.476 | 10,36 | 2,11 | 2,30 | 603 | 6,44 | 0,38 | 15,75 |

| Isolat | Abdeckung | | Verhältnis | | | Verhältnis* | | | |
|---------|-----------|-------|-----------------------|-------------------------|-----------------|-------------|-----------------------|-------------------------|-----------------|
| | NN2 [%] | SNPs | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{Ts}{Tv}$ | SNPs* | $\frac{intra}{inter}$ | $\frac{syn}{nicht-syn}$ | $\frac{Ts}{Tv}$ |
| 426 | 85,80 | 1.950 | 9,77 | 3,61 | 1,82 | 1.075 | 8,03 | 3,47 | 2,05 |
| 584 | 91,36 | 2.920 | 12,84 | 3,15 | 1,75 | 56 | 6,00 | 0,37 | 3,00 |
| 642 | 94,80 | 2.708 | 12,21 | 2,81 | 1,66 | 63 | 6,00 | 0,46 | 2,32 |
| 747 | 94,20 | 2.572 | 11,19 | 3,38 | 1,70 | 316 | 10,70 | 2,25 | 2,29 |
| 800 | 96,29 | 3.495 | 11,76 | 2,32 | 2,01 | 449 | 7,47 | 0,52 | 4,91 |
| 10_C1 | 90,02 | 2.733 | 11,54 | 3,16 | 1,82 | 57 | 3,38 | 0,57 | 2,35 |
| 10SA1 | 87,90 | 44 | 3,89 | 0,40 | 2,14 | 44 | 3,89 | 0,40 | 2,14 |
| 10SB1 | 84,11 | 132 | 9,15 | 1,90 | 1,81 | 33 | 3,13 | 0,56 | 2,67 |
| 110_D4 | 84,40 | 1.812 | 9,72 | 2,99 | 1,76 | 49 | 11,25 | 0,73 | 2,77 |
| 12_F8 | 90,86 | 3.030 | 11,89 | 3,09 | 1,80 | 49 | 4,44 | 0,67 | 2,27 |
| 12_SB1 | 90,89 | 3.351 | 13,32 | 3,32 | 1,77 | 211 | 18,18 | 1,63 | 1,43 |
| 28PA3 | 87,32 | 1.571 | 10,90 | 3,78 | 1,69 | 39 | 3,88 | 0,35 | 3,33 |
| 702_D3 | 92,00 | 2.716 | 11,13 | 2,92 | 1,76 | 36 | 4,14 | 0,32 | 2,00 |
| 902D1 | 89,22 | 1.588 | 12,23 | 3,86 | 1,57 | 48 | 4,33 | 0,86 | 3,36 |
| AL962 | 89,25 | 1.858 | 9,74 | 3,05 | 1,79 | 134 | 4,58 | 1,56 | 1,53 |
| B6470 | 91,50 | 1.376 | 12,62 | 3,83 | 1,62 | 32 | 5,40 | 0,42 | 3,57 |
| C17 | 94,94 | 2.612 | 9,08 | 3,29 | 1,85 | 610 | 4,45 | 2,95 | 2,24 |
| CF10 | 86,72 | 1.075 | 10,81 | 3,78 | 1,57 | 28 | 4,60 | 0,28 | 3,67 |
| CF11 | 89,42 | 1.029 | 10,43 | 3,58 | 1,69 | 41 | 4,13 | 0,83 | 5,83 |
| CF2 | 92,90 | 1.839 | 12,04 | 3,66 | 2,00 | 166 | 9,38 | 2,26 | 2,46 |
| CF3 | 93,86 | 1.330 | 11,79 | 3,64 | 1,60 | 52 | 5,50 | 0,57 | 3,73 |
| CF5 | 86,24 | 1.042 | 9,42 | 4,06 | 1,66 | 29 | 2,67 | 0,71 | 8,67 |
| CF7 | 90,36 | 2.182 | 11,61 | 3,38 | 1,84 | 55 | 4,00 | 0,52 | 1,89 |
| CF8 | 92,78 | 1.309 | 11,83 | 3,89 | 1,59 | 47 | 5,71 | 0,43 | 3,70 |
| K11 | 85,61 | 897 | 7,63 | 3,66 | 1,64 | 347 | 6,71 | 3,03 | 1,94 |
| K13 | 91,56 | 5.447 | 12,76 | 3,23 | 1,70 | 2.900 | 11,66 | 3,25 | 1,94 |
| K15 | 91,18 | 1.175 | 11,63 | 3,96 | 1,55 | 19 | 3,75 | 0,36 | 2,80 |
| K16 | 96,08 | 1.646 | 12,72 | 3,88 | 1,56 | 32 | 3,57 | 0,32 | 3,57 |
| K17 | 96,31 | 4.439 | 8,95 | 3,42 | 1,85 | 2.461 | 7,82 | 3,29 | 2,10 |
| K18 | 95,52 | 2.004 | 12,82 | 3,16 | 1,75 | 60 | 5,67 | 0,55 | 2,16 |
| K19 | 83,31 | 3.033 | 8,88 | 3,22 | 1,96 | 2.222 | 8,22 | 3,08 | 2,14 |
| K20 | 95,88 | 3.305 | 10,93 | 3,81 | 1,58 | 718 | 2,36 | 3,00 | 2,00 |
| K5 | 82,38 | 862 | 13,37 | 4,38 | 1,61 | 16 | 2,20 | 0,57 | 4,33 |
| PJC | 94,94 | 3.116 | 12,61 | 3,13 | 1,79 | 41 | 4,13 | 0,38 | 2,73 |
| SG1 | 96,98 | 15 | 0,88 | 0,40 | 2,75 | 15 | 1,14 | 0,40 | 2,75 |
| SG17m | 94,37 | 4.069 | 10,06 | 3,10 | 1,78 | 751 | 4,82 | 2,94 | 2,04 |
| SG50m | 93,27 | 2.842 | 12,60 | 3,07 | 1,75 | 47 | 3,27 | 0,33 | 2,13 |
| SS30 | 95,16 | 3.335 | 12,50 | 2,37 | 2,11 | 402 | 8,57 | 0,41 | 19,10 |
| W5Aug28 | 89,79 | 2.945 | 11,97 | 3,04 | 1,80 | 50 | 2,85 | 0,32 | 1,63 |
| WI1 | 94,33 | 2.970 | 13,01 | 3,07 | 1,80 | 50 | 6,14 | 0,54 | 3,55 |
| WU2 | 94,42 | 3.126 | 14,25 | 2,37 | 2,12 | 342 | 10,03 | 0,40 | 15,29 |

Die Abdeckung gibt den mit der SOLiD Sequenzierung abgedeckten Anteil des Referenzgenoms NN2 an.

* ohne die genomische Inseln PAGI-2 und RGP5. $\frac{intra}{inter}$ ist das Verhältnis intra- zu intergenischen SNPs. $\frac{syn}{nicht-syn}$ ist das Verhältnis synonymmer zu nicht-synonymen SNPs. $\frac{Ts}{Tv}$ ist das Verhältnis von Transitionen zu Transversionen.

Es konnte gezeigt werden, dass das Kerngenom nur eine geringe Anzahl an Mutationen aufweist und die Anzahl und Art der Austausche in beiden untersuchten klonalen Komplexen vergleichbar war. Die höchste Sequenzdiversität wird somit durch genomische Inseln (GI) und einige Regionen genomischer Plastizität (RGP) des akzessorischen Genoms verursacht (Abb. 3.14, 3.15). In der Zusammensetzung des akzessorischen Genoms konnten 26 der 100 Isolate mit raumzeitlich gleichem Isolationsursprung als gruppiert erkannt werden. So konnten beispielweise je drei Paare, vier Trios und ein Quintett von Isolaten, welche innerhalb von 3 Jahren in einem Gebiet vergleichbar mit der Größe Niedersachsens gesammelt wurden, einem Cluster zugeordnet werden.

Aus den Abbildungen 3.14 und 3.15 ließ sich für beide klonalen Komplexe ablesen, dass einige genomische Inseln und RGPs wie PAP1-1 (PA14) oder PAGI-4 (beide Komplexe) in allen untersuchten Stämmen vorhanden waren. Zusätzlich war es jedoch auch möglich, RGPs zu identifizieren, welche in bereits sequenzierten *P. aeruginosa* annotiert wurden, jedoch in keinem Isolat der beiden klonalen Komplexe vorkommen (z.B. RGP13 aus PAO1). Zwischen den beiden Extremen der Abwesenheit von genomischen Inseln und RGPs in allen Isolaten eines klonalen Komplexes und der Anwesenheit in allen Stämmen, ließen sich auch Subgruppen von Isolaten definieren, die ein bestimmtes Set an genomischen Inseln und RGPs teilten. So wurden in beiden klonalen Komplexen je vier Untergruppen definiert, welche untereinander in der Zusammensetzung des akzessorischen Genoms verschieden waren. Als nächstes wurden die Untergruppen in der Zusammensetzung der genomischen Inseln mit denen der RGP-Insertionen verglichen. Dabei wurde gezeigt, dass die meisten Stämme ihre nächsten Nachbarn behielten. Einzelne Stämme wurden allerdings in anderen Clustern wiedergefunden und hatten neue nächste Nachbarn (z.B. 37, CF12 in Klon PA14; 237, K19 in Klon C). Daraus folgte, dass das Repertoire der genomischen Inseln für die Isolate der klonalen Komplexe mit der Anwesenheit der RGPs zwar korrelierte, jedoch nicht strikt auf die RGPs übertragen werden konnte.

Tabelle 3.10: Übersicht zur Zahl der Einzelnukleotidaustausche und Indels pro Stamm.

| | Klon C | Klon C* | Klon PA14 | Klon PA14 [§] |
|------------------------------------|----------------------|---------------------|---------------------|------------------------|
| Zahl der Stämme | 57 | 57 | 42 | 42 |
| Häufigkeit der SNPs pro Stamm | | | | |
| gesamt | 2.568 (15 - 5.447) | 56 (15 - 2.900) | 156 (55 - 6.364) | 141 (55 - 4.275) |
| davon | | | | |
| intragenisch (intra) | 2.353 (7 - 5.051) | 45 (8 - 2.671) | 118 (40 - 5.796) | 105 (40 - 3.807) |
| intergenisch (inter) | 194 (8 - 446) | 11 (4 - 279) | 37 (15 - 568) | 34 (15 - 468) |
| synonym (syn) | 1.805 (2 - 3.856) | 16 (2 - 2.042) | 47 (13 - 4.605) | 41 (13 - 3.126) |
| nicht-synonym (nicht-syn) | 548 (5 - 1.195) | 28 (5 - 629) | 74 (27 - 1.191) | 65 (27 - 681) |
| Transition (Ts) | 1.620 (11 - 3.426) | 41 (11 - 1.912) | 97 (40 - 4.703) | 87 (40 - 3.305) |
| Transversion (Tv) | 916 (4 - 2.021) | 17 (3 - 988) | 59 (15 - 1.661) | 55 (15 - 970) |
| Verhältnis $\frac{intra}{inter}$ | 11,62 (0,88 - 14,24) | 4,72 (1,14 - 18,18) | 3,84 (2,04 - 15,54) | 3,35 (1,13 - 9,13) |
| Verhältnis $\frac{syn}{nicht-syn}$ | 3,21 (0,4 - 4,38) | 0,55 (0,28 - 3,47) | 0,73 (0,36 - 3,87) | 0,63 (0,36 - 4,59) |
| Verhältnis $\frac{Ts}{Tv}$ | 1,76 (1,55 - 2,75) | 2,5 (1,43 - 19,1) | 1,91 (1,27 - 6,89) | 1,68 (1,25 - 7,67) |
| Indels | 5 (1 - 64) | 3 (1 - 62) | 4 (0 - 58) | 2 (0 - 56) |
| davon Leserasterverschiebung | 2 (0 - 44) | 2 (0 - 44) | 1 (0 - 34) | 0 (0 - 33) |

Angegeben ist jeweils der mediane Wert, sowie das Minimum und das Maximum.

* ohne die genomischen Inseln pKLC102 und RGP5.

§ ohne die genomische Insel PAPI-1.

Tabelle 3.11: Vergleich der Mutationsfrequenz zwischen den Habitaten der Klon C Stämme.

| Habitat | Klon C | | | Klon C* | | |
|------------------------------------|----------------------|----------------------|----------------------|---------------------|--------------------|---------------------|
| | Umwelt | Akut | Chronisch | Umwelt | Akut | Chronisch |
| Zahl der Stämme | 11 | 10 | 36 | 11 | 10 | 36 |
| SNPs | | | | | | |
| gesamt | 2.733 (44 - 4.069) | 1.825 (862 - 5.447) | 2.464 (15 - 3.495) | 49 (33 - 751) | 204 (16 - 2.900) | 61 (15 - 1.411) |
| davon | | | | | | |
| intragenisch (intra) | 2.515 (35 - 3.701) | 1.693 (793 - 5.051) | 2.231 (7 - 3.221) | 39 (25 - 622) | 177 (11 - 2.671) | 51 (8 - 1.208) |
| intergenisch (inter) | 218 (9 - 368) | 133 (60 - 446) | 188 (8 - 337) | 9 (4 - 129) | 27 (4 - 279) | 11 (5 - 203) |
| synonym (syn) | 1.910 (10 - 2.799) | 1.313 (623 - 3.856) | 1.731 (2 - 2.326) | 16 (7 - 464) | 123 (4 - 2.042) | 17 (2 - 907) |
| nicht-synonym (nicht-syn) | 635 (25 - 902) | 380 (149 - 1.195) | 544 (5 - 1.018) | 26 (16 - 158) | 54 (7 - 629) | 34 (5 - 378) |
| Transition (Ts) | 1.763 (30 - 2.604) | 1.139 (532 - 3.426) | 1.544 (11 - 2.422) | 34 (24 - 504) | 135 (13 - 1.912) | 46 (11 - 915) |
| Transversion (Tv) | 985 (14 - 1.465) | 686 (330 - 2.021) | 904 (4 - 1.240) | 15 (9 - 247) | 69 (3 - 988) | 19 (3 - 496) |
| Verhältnis $\frac{intra}{inter}$ | 11,54 (3,89 - 13,32) | 12,13 (7,63 - 13,37) | 11,62 (0,88 - 14,24) | 4,14 (2,85 - 18,18) | 5,53 (2,2 - 11,66) | 5,11 (1,14 - 10,70) |
| Verhältnis $\frac{syn}{nicht-syn}$ | 3,07 (0,34 - 3,86) | 3,74 (3,15 - 4,38) | 3,18 (0,4 - 4,06) | 0,57 (0,32 - 2,94) | 1,79 (0,32 - 3,29) | 0,52 (0,28 - 3,47) |
| Verhältnis $\frac{Ts}{Tv}$ | 1,78 (1,57 - 2,14) | 1,63 (1,55 - 1,96) | 1,79 (1,57 - 2,75) | 2,14 (1,43 - 3,36) | 2,15 (1,94 - 4,33) | 2,74 (1,53 - 19,1) |
| Indels | 5 (3 - 12) | 5 (3 - 19) | 6 (1 - 64) | 2 (1 - 5) | 3 (2 - 17) | 4 (1 - 62) |
| davon Leserasterversch. | 1 (1 - 3) | 2 (0 - 4) | 2 (0 - 44) | 1 (0 - 2) | 2 (0 - 4) | 2 (0 - 44) |

Angegeben ist jeweils der mediane Wert, sowie das Minimum und das Maximum.

* ohne die genomischen Inseln pKLC102 und RGP5. $\frac{intra}{inter}$ ist das Verhältnis intra- zu intergenischen SNPs. $\frac{syn}{nicht-syn}$ ist das Verhältnis synonymmer zu nicht-synonymen SNPs. $\frac{Ts}{Tv}$ ist das Verhältnis von Transitionen zu Transversionen.

Tabelle 3.12: Vergleich der Mutationsfrequenz zwischen den Habitaten der Klon PA14 Stämme.

| Habitat | Klon PA14 | | | Klon PA14§ | | |
|------------------------------------|---------------------|---------------------|---------------------|--------------------|--------------------|--------------------|
| | Umwelt | Akut | Chronisch | Umwelt | Akut | Chronisch |
| Zahl der Stämme | 6 | 10 | 26 | 6 | 10 | 26 |
| SNPs | | | | | | |
| gesamt | 1.163 (127 - 6.364) | 306 (120 - 2.033) | 150 (55 - 2.096) | 180 (125 - 4.275) | 150 (118 - 499) | 140 (55 - 1.573) |
| davon | | | | | | |
| intra-genisch (intra) | 1.054 (97 - 5.796) | 267 (86 - 1.871) | 115 (40 - 1.895) | 122 (95 - 3.807) | 105 (86 - 449) | 104 (40 - 1.396) |
| inter-genisch (inter) | 90 (30 - 568) | 41 (25 - 162) | 37 (15 - 201) | 34 (30 - 468) | 41 (25 - 62) | 38 (15 - 177) |
| synonym (syn) | 714 (33 - 4.605) | 181 (37 - 1.310) | 43 (13 - 1.340) | 43 (33 - 3.126) | 43 (36 - 319) | 39 (13 - 995) |
| nicht-synonym (nicht-syn) | 340 (63 - 1.191) | 91 (47 - 561) | 70 (27 - 559) | 79 (60 - 681) | 68 (47 - 130) | 64 (27 - 401) |
| Transition (Ts) | 765 (75 - 4.703) | 205 (71 - 1.322) | 90 (40 - 1.380) | 100 (73 - 3.305) | 91 (70 - 336) | 84 (40 - 1.148) |
| Transversion (Tv) | 398 (49 - 1.661) | 104 (41 - 711) | 54 (15 - 716) | 62 (47 - 970) | 63 (41 - 163) | 50 (15 - 425) |
| Verhältnis $\frac{intra}{inter}$ | 7,1 (2,89 - 15,54) | 6,48 (2,52 - 11,56) | 3,47 (2,04 - 10,94) | 3,45 (1,86 - 8,13) | 3,76 (1,76 - 9,13) | 3,21 (1,13 - 7,89) |
| Verhältnis $\frac{syn}{nicht-syn}$ | 1,46 (0,52 - 3,87) | 1,59 (0,60 - 3,11) | 0,65 (0,38 - 2,60) | 0,57 (0,53 - 4,59) | 0,79 (0,51 - 2,95) | 0,62 (0,36 - 2,48) |
| Verhältnis $\frac{T_s}{T_v}$ | 1,83 (1,44 - 2,83) | 1,86 (1,27 - 2,26) | 1,94 (1,27 - 6,89) | 1,63 (1,40 - 3,41) | 1,68 (1,25 - 3,05) | 1,71 (1,28 - 7,67) |
| Indels | 6 (0 - 34) | 3 (1 - 14) | 4 (0 - 58) | 2 (0 - 22) | 3 (1 - 4) | 3 (0 - 56) |
| davon Leserasterversch. | 1 (0 - 7) | 1 (0 - 3) | 2 (0 - 34) | 0 (0 - 4) | 0 (0 - 2) | 1 (0 - 33) |

Angegeben ist jeweils der mediane Wert, sowie das Minimum und das Maximum.

§ ohne die genomische Insel PAPI-1. $\frac{intra}{inter}$ ist das Verhältnis intra- zu intergenischen SNPs. $\frac{syn}{nicht-syn}$ ist das Verhältnis synonym zu nicht-synonymen SNPs. $\frac{T_s}{T_v}$ ist das Verhältnis von Transitionen zu Transversionen.

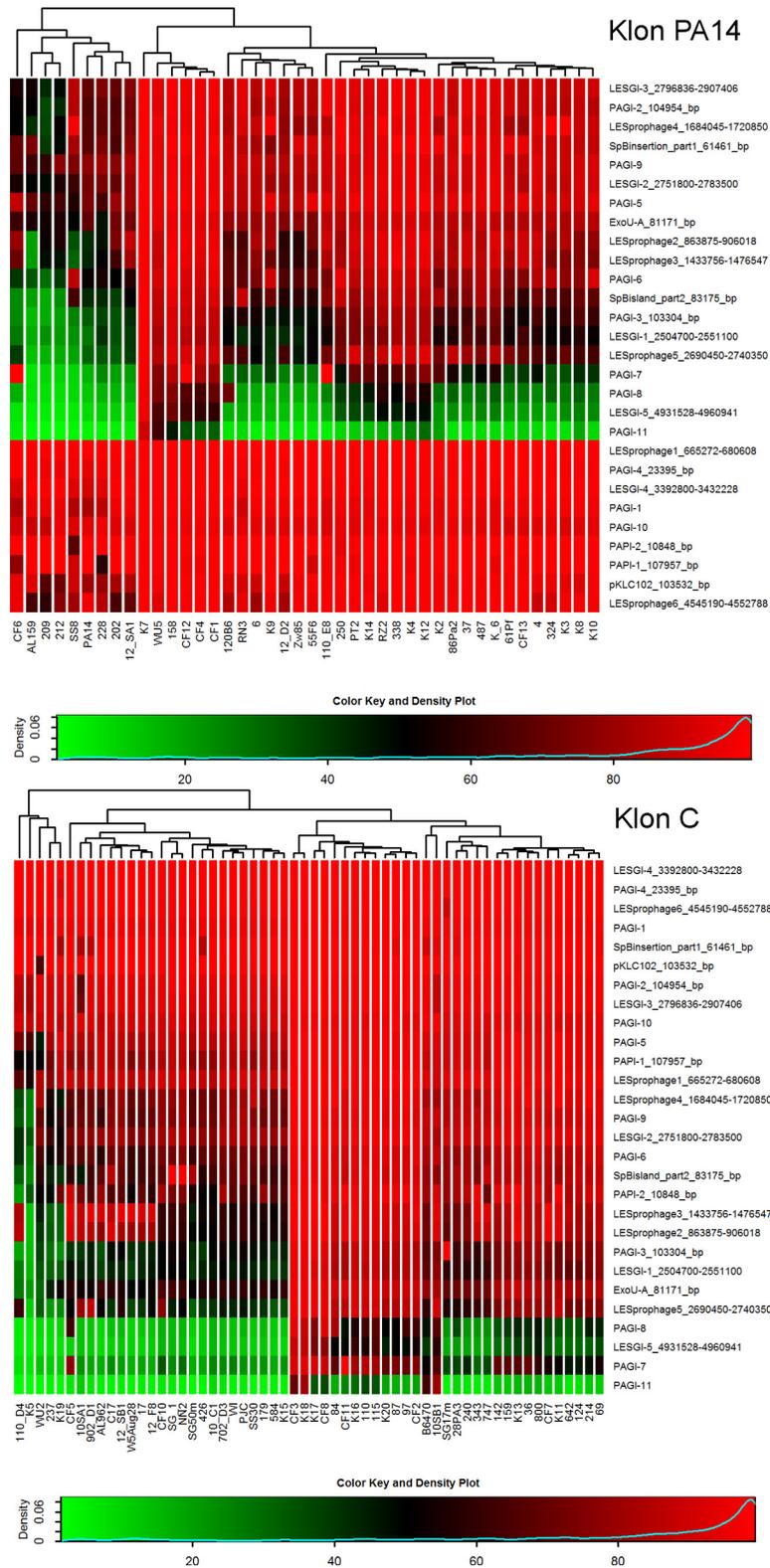


Abbildung 3.14: Diversität des Repertoires an genomischen Inseln im akzessorischen Genom der klonalen Komplexe C und PA14. Die Heatmaps zeigen die An- (rot) oder Abwesenheit (grün) genomischer Inseln in Klon PA14 (oben) und Klon C (unten) Isolaten. Die Stämme sind nach ihrer Ähnlichkeit in der Zusammensetzung der genomischen Inseln angeordnet.

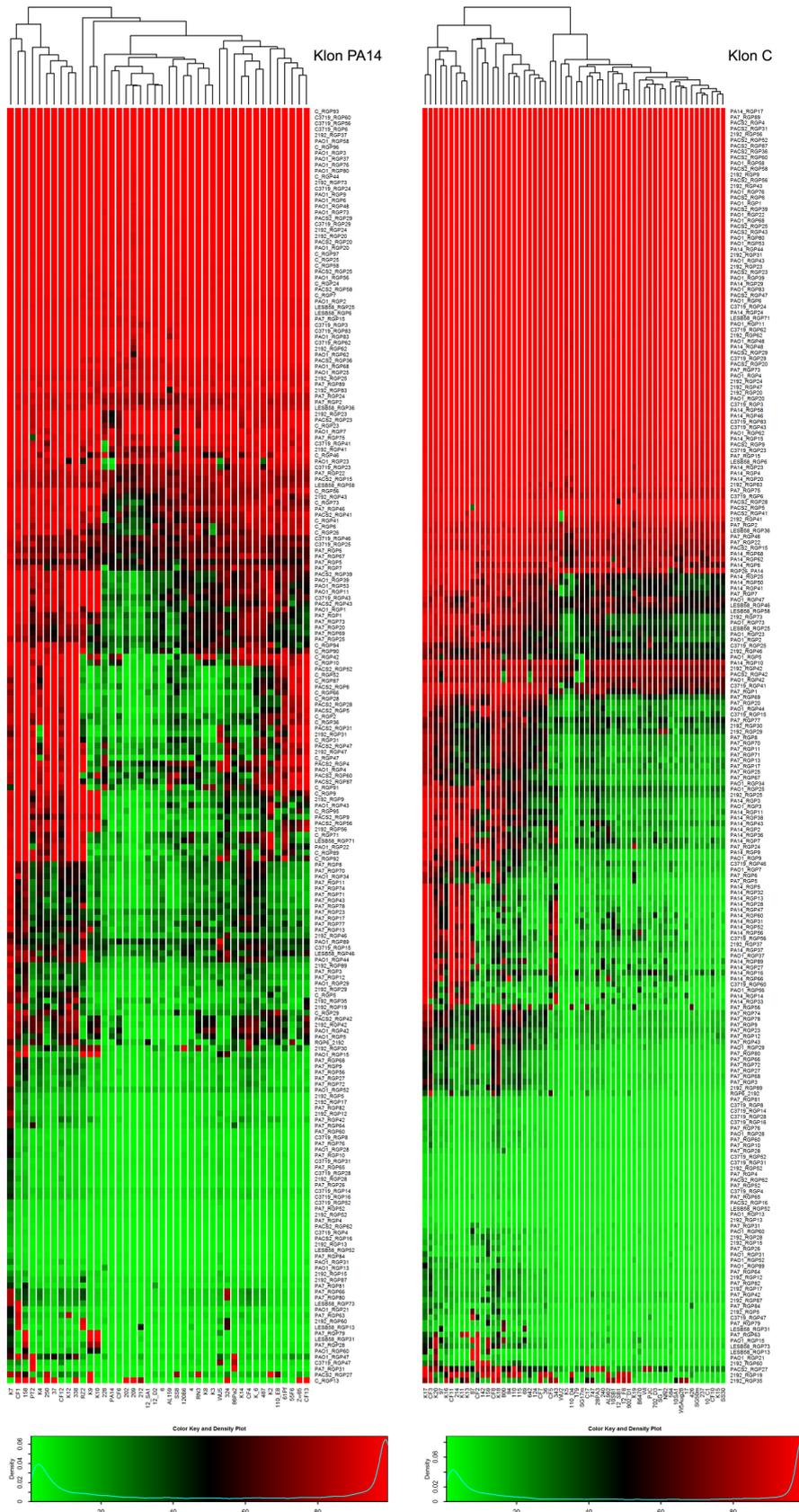


Abbildung 3.15: Diversität des Repertoires an RGP im akzessorischen Genom der klonalen Komplexe C und PA14. Die Heatmaps zeigen die An- (rot) oder Abwesenheit (grün) von Regionen genomischer Plastizität (RGP) in Klon PA14 (links) und Klon C (rechts) Isolaten. Die Stämme sind nach ihrer Ähnlichkeit in der Zusammensetzung der Regionen genomischer Plastizität angeordnet.

Die intraklonale Diversität wurde über paarweise Stammvergleiche des Repertoires an genomischen Inseln und RGPs bestimmt (Abb. 3.16). Ein additiver binärer Score diente als Maß für die Ähnlichkeit des akzessorischen Genoms eines Stammpaares. Die Diversität des akzessorischen Genoms zeigte eine hohe Streuung innerhalb beider klonaler Komplexe in der Zusammensetzung aus genomischen Inseln (GI) und RGPs. Dabei erwies sich die inter- und intraklonale Verteilung der genomischen Inseln und RGPs innerhalb der Stammpaare als nicht signifikant voneinander verschieden. Mit anderen Worten war die Zusammensetzung der mobilen Elemente ebenso variabel zwischen individuellen Klon C oder Klon PA14 Stämmen wie zwischen unverwandten Stämmen (Abb. 3.16). Die geringste Gemeinsamkeit im Repertoire von GI und RGP (Score < 0) zeigten innerhalb der Analyse einige Klon PA14 Stämme.

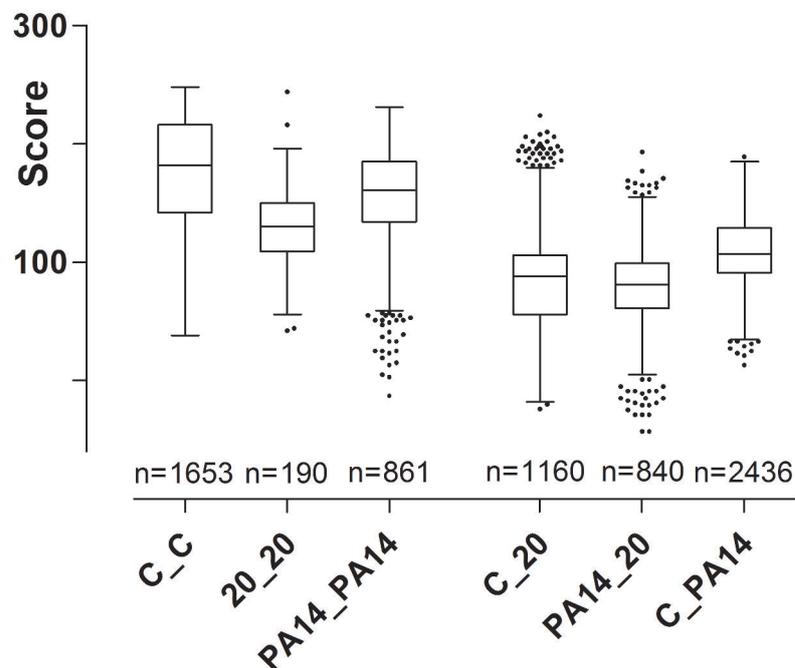


Abbildung 3.16: Boxplot-Darstellung der Ähnlichkeit des akzessorischen Genoms innerhalb und zwischen klonalen Komplexen. Für jeden Stamm wurde ein globaler Wert (Score) berechnet, wobei in paarweisen Vergleich überprüft wurde, ob die Stämme konkordant (Wert +1) oder diskordant (Wert -1) in der Ab- bzw. Anwesenheit jeder genomischen Insel und RGP von acht sequenzierten *P. aeruginosa* Genomen sind. Man beachte dabei die große Überlappung von Werten der intraklonalen Vergleiche (C_C; PA14_PA14) mit denen der interklonalen Vergleiche von 20 nicht miteinander verwandten Stämmen (20_20). n gibt dabei die Anzahl der durchgeführten Paarvergleiche an.

Beim Vergleich der Gruppen untereinander ließ sich feststellen, dass die Vergleiche der klonalen Komplexe C und PA14 mit den unverwandten Stämmen sehr ähnlich waren und lediglich wenige Ausnahmen unter den Klon C Stämmen eine etwas erhöhte Gemeinsamkeit mit den unverwandten Klonen und wenige PA14 Stämme mehr Unterschiede zu den

unverwandten Klonen zeigten. Der direkte Vergleich von Klon C und Klon PA14 Isolaten illustrierte zudem eine leichte erhöhte, jedoch nicht signifikante Gemeinsamkeit zwischen den klonalen Komplexen und eine geringe Anzahl an Ausreißerwerten verglichen mit dem Test gegen die unverwandten Stämme (Abb. 3.16).

Zusammenfassend ließ sich feststellen, dass die höchste Genomdiversität in mobilen integrativen und konjugativen Elementen (engl. integrative and conjugative elements, ICE) der Typen pKLC102, PAGI-2 und PAPI-1 nachgewiesen wurde, welche auch als extrachromosomale Elemente ko-existieren können. Von diesen Elementen ist bekannt, dass sie durch horizontalen Gentransfer über Spezies- und Genusgrenzen hinaus ausgetauscht werden können (Klockgether et al., 2007; Qiu et al., 2006; Pradervand et al., 2014). Übereinstimmend mit der Zugänglichkeit zu einem großen Pool proteobakterieller Wirtsgenome wurden mehr SNPs den ICEs zugeordnet als jeder anderen genomischen Insel oder RGP der Klon C bzw. Klon PA14 Genome. Blendete man die SNPs der genannten ICEs aus, näherten sich die Mutationsstatistiken der beiden Klone an, was darauf hindeutet, dass ähnliche Mutations- und Reparaturmechanismen in beiden Klonen genutzt wurden (Tab. 3.10).

3.2.3 Natürliche Selektion von Einzelnukleotidaustauschen innerhalb klonaler Komplexe

Im vorherigen Kapitel wurde die Zusammensetzung des Kern- und des akzessorischen Genoms beschrieben und ein Überblick über die Einzelnukleotidaustausche in den beiden klonalen Komplex angegeben. Das folgende Kapitel beschreibt die Untersuchung der gefundenen Mutationen detaillierter. Dazu wurde als erstes das Verhältnis $\frac{s}{n}$ von synonymen zu nicht-synonymen Nukleotidaustauschen in den kodierenden Regionen der Isolate untersucht. Der Anteil synonyme Austausche war in allen Stämmen höher als der GC-normalisierte Erwartungswert $\frac{s}{n} = \frac{2,9}{9,1}$, welcher durch statistisches Auszählen aller Austauschmöglichkeiten unter Vernachlässigung thermodynamischer Effekte ermittelt wurde. Dabei konnte deutlich zwischen zwei Gruppen unterschieden werden (Abb. 3.17). In dem größeren, 70 Stämme umfassenden Cluster lag die Überrepräsentation der synonymen Austausche im Vergleich zur Nullhypothese noch innerhalb des oberen Konfidenzintervalles des Erwartungswertes. Allerdings waren mehr Aminosäureaustausche verursachende als neutrale Mutationen detektiert worden ($\frac{s}{n} < 1$). In den weiteren 30 Stämmen, wozu auch diejenigen mit der größten intraklonalen Genomdiversität zählen, nahm der Anteil synonyme SNPs mit steigender Gesamtmutationszahl zu (Abb. 3.17). Die logarithmische Regressionslinie der klonalen Komplexe beschreibt treffend das Verhältnis $\frac{s}{n}$ innerhalb der unverwandten Klone. Diese Daten gaben einen Hinweis auf die Evolution der kodierenden Genombereiche in *P. aeruginosa*. Beim Vergleich der intraklonalen Genomdiversität gegen eine klonale Referenz (Klon C: NN2; Klon PA14: PA14), zeigte sich, dass neutra-

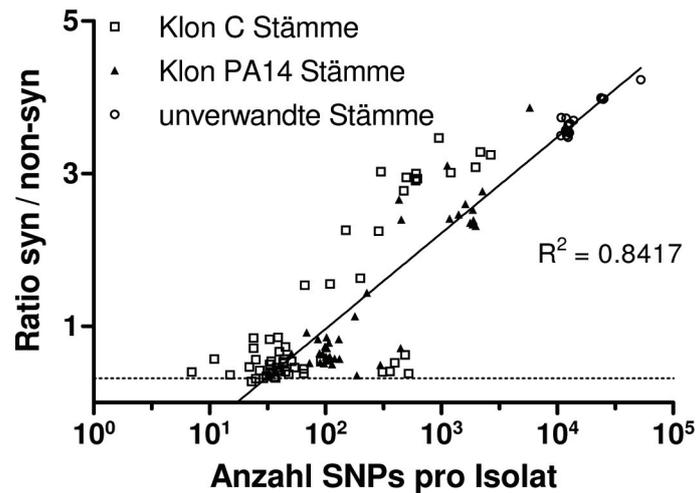


Abbildung 3.17: Vergleich intra- gegen interklonale Sequenzdiversität. Darstellung des Verhältnisses von synonymen zu nicht-synonymen SNPs ($\frac{s}{n}$) gegen die Gesamtzahl der Einzelnukleotidaustausche pro Stamm. Klon C (Referenz NN2), Klon PA14 (Referenz PA14) und klonal nicht verwandte Stämme (Referenz PAO1) sind durch Symbole dargestellt. Die gepunktete Linie gibt den Erwartungswert für zufällig auftretende Mutationen an.

le Mutationen mit einer höheren Wahrscheinlichkeit toleriert werden als nicht-synonyme Einzelnukleotidaustausche und somit bei steigender Gesamtmutationszahl akkumulieren. Dieser Trend der „reinigenden“ Selektion von nicht-synonymen Austauschen erhöhte sich mit der Gesamtmutationszahl, bis ein $\frac{s}{n}$ -Verhältnis von ungefähr vier erreicht wurde. Dieser Wert war typisch für die interklonale Genomdiversität zwischen unverwandten Klonen (Hilker et al., 2015).

Der vergleichsweise höhere Anteil von Aminosäureaustauschen innerhalb eines klonalen Komplexes spiegelte sich zusätzlich auch in der abweichenden Frequenz der Art des Austausches und der funktionellen Kategorie des betroffenen Proteins wider (Tab. S4). So unterschied sich das intraklonale Spektrum der Aminosäureaustausche signifikant ($p < 10^{-6}$) für die klonalen Komplexe C und PA14 von dem des *P. aeruginosa* Pangenoms (Hilker et al., 2015). Sortierte man die Austausche nach der Dayhoff-Ähnlichkeitsmatrix (Dayhoff, 1978), zeigte sich, dass der Anteil drastischer Aminosäureaustausche (Dayhoff < 6) innerhalb der klonalen Komplexe höher war als zwischen unverwandten Klonen (Tab. 3.13, Abb. 3.18).

Zudem wurde ein signifikanter Unterschied in der Besetzung der Hälfte der Dayhoff-Klassen im Vergleich zwischen Klon C und Klon PA14 erkannt (Tab. 3.13). Die inter- und intraklonale Häufigkeitsverteilung unterschied sich nicht nur in der Art und Frequenz der Aminosäureaustausche, sondern auch in den funktionellen Kategorien der betroffenen Proteine (Tab. 3.14). Die relative Häufigkeit von Aminosäuresequenzvarianten unterschied sich in 16 der 26 funktionellen Klassen für Klon PA14 und allen 26 Klassen für Klon C gegenüber dem Kerngenom, welches von allen *P. aeruginosa* Stämmen geteilt wird. DNA

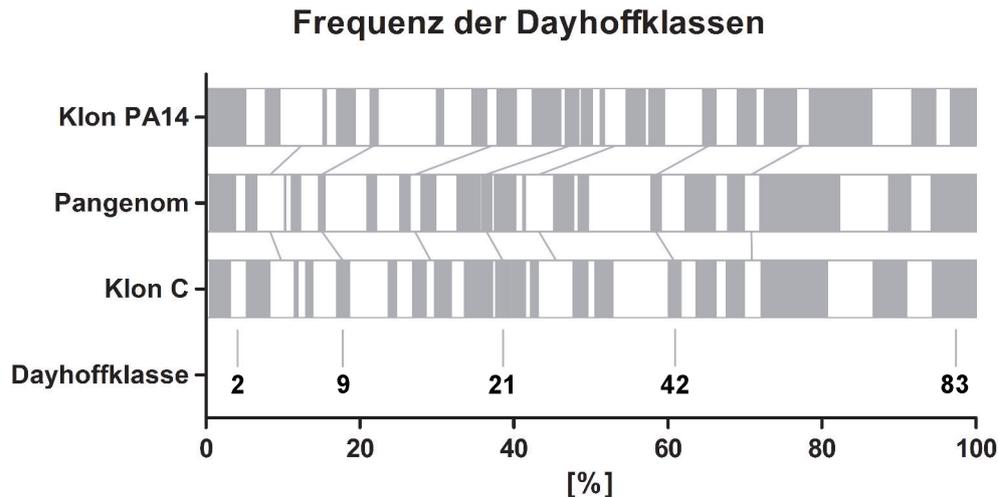


Abbildung 3.18: Verteilung der Aminosäureaustausche nach Dayhoffklassen in den Klonen C und PA14, sowie klonal unverwandten Stämmen (Pangenom).

modifizierende Enzyme und Proteine des Fettsäuremetabolismus, der Adhärenz und Motilität wurden als bevorzugte Ziele intraklonaler Proteinsequenzvarianten identifiziert (Tab. 3.14). Zusammenfassend ließ sich feststellen, dass sowohl das Zielspektrum der Aminosäureaustausche, als auch deren Charakter innerhalb der Klone C und PA14 unterschiedlich im Vergleich zum *P. aeruginosa* Kerngenom war.

Vergleich man die Besetzung der Dayhoffklassen zwischen den klonalen Komplexen C und PA14, konnte man für 20 der 40 Gruppen signifikante Unterschiede feststellen. Dabei waren im klonalen Komplex C vor allem die Dayhoffklassen mit geringem Einfluss auf die Proteinfunktion (Dayhoffklassen >58) überrepräsentiert. Innerhalb der drastischen (Dayhoffklassen < 6) und mittelschweren Aminosäureaustausche (Dayhoffklassen 6 - 30) ließ sich kein Trend zur signifikanten Überrepräsentation zwischen beiden klonalen Komplexen erkennen.

Tabelle 3.13: Besetzung der Dayhoffklassen in Klon C, Klon PA14 oder klonal unverwandten Stämmen (Pangenom).

| Anzahl in | | | | Anzahl in | | | |
|-----------|----------|--------------|-------------|-----------|----------|---------------|-------------|
| Klasse | Pangenom | Klon C | PA14 | Klasse | Pangenom | Klon C | PA14 |
| 0 | 315 | 162* | <u>22</u> | 23 | 251 | <u>72</u> | 36 |
| 1 | 1.931 | <u>730</u> | 409* | 24 | 1.589 | <u>507</u> | 116 |
| 2 | 795 | 632 | 217 | 25 | 587 | <u>195</u> | 91 |
| 3 | 829 | 870* | 152 | 26 | 191 | 276* | 37 |
| 4 | 2.133 | <u>938</u> | 472* | 27 | 2.202 | 1.319* | <u>241</u> |
| 5 | 70 | 108 | 32 | 30 | 1.457 | <u>537</u> | 203* |
| 6 | 473 | 309 | 114 | 32 | 390 | 286 | 43 |
| 7 | 658 | <u>258</u> | 205* | 35 | 739 | 658 | 165 |
| 8 | 1.441 | 922* | <u>160</u> | 37 | 4.848 | 2.100* | <u>424</u> |
| 9 | 462 | 461* | 87 | 42 | 745 | 460 | 140 |
| 10 | 3.234 | <u>1.474</u> | 636* | 43 | 1.858 | <u>591</u> | 233* |
| 11 | 692 | <u>275</u> | 75 | 45 | 2.269 | 738 | 194 |
| 12 | 1.830 | <u>621</u> | 309* | 48 | 957 | <u>402</u> | <u>98</u> |
| 13 | 755 | 499 | 157 | 53 | 1.275 | 645 | 348* |
| 14 | 869 | <u>344</u> | 117 | 58 | 1.220 | 654* | 142 |
| 15 | 1.093 | 602 | 208 | 59 | 6.079 | <u>2.482</u> | <u>677</u> |
| 16 | 1.674 | <u>516</u> | <u>174</u> | 66 | 3.785 | <u>1.731*</u> | <u>437</u> |
| 17 | 1.709 | 1.046 | 307 | 70 | 1.648 | 1.240* | 255 |
| 20 | 178 | 111 | 51 | 77 | 1.589 | 981* | <u>169</u> |
| 21 | 697 | 546 | 142 | 83 | 3.423 | 1.625* | <u>272</u> |

signifikant ($p < 0,05$) über- (fett) oder unterrepräsentiert (fett, unterstrichen) im Klon verglichen mit dem Pangenom.

*=signifikanter ($p < 0,05$) Unterschied zwischen Klon C und Klon PA14.

Alle p -Werte wurden korrigiert durch multiples Vergleichen nach dem Bonferroni Algorithmus.

Tabelle 3.14: Häufigkeit der Aminosäureaustausche in den funktionellen Kategorien der klonalen Komplexen C und PA14.

| Funktionelle Kategorie | Gene des Kerngenoms | | Frequenz von Aminosäureaustauschen | | Frequenz von nicht-konservativen Aminosäureaustauschen (Dayhoff < 6) | |
|---|---------------------|------|------------------------------------|---------------|--|-----------|
| | n | % | Klon PA14 | Klon C | Klon PA14 | Klon C |
| Adaptation, Schutz | 92 | 1,90 | <u>30</u> | <u>49</u> | 4 | 12 |
| Aminosäurebiosynthese und -metabolismus | 186 | 3,84 | 165 | 69 | 16 | 10 |
| Antibiotikaresistenz und -suszeptibilität | 9 | 0,18 | <u>2</u> | <u>2</u> | 2 | 0 |
| Biosynthese von Kofaktoren und prosthetischen Gruppen | 128 | 2,64 | <u>6</u> | <u>37</u> | 3 | 8 |
| Kohlenstoffkatabolismus | 111 | 2,29 | 5 | <u>36</u> | 6 | 9 |
| Zellteilung | 22 | 0,45 | 19 | 228 | 4 | 9 |
| Zellwand, LPS, Kapsel | 114 | 2,35 | 75 | <u>149</u> | 8 | 21 |
| zentraler Intermediärstoffwechsel | 178 | 3,68 | <u>37</u> | <u>29</u> | <u>4</u> | <u>3</u> |
| Chaperone | 16 | 0,33 | 6 | 261 | 0 | 6 |
| Chemotaxis | 21 | 0,43 | 14 | <u>16</u> | 0 | 3 |
| DNA Replikation, Modifikation, Reparatur | 76 | 1,57 | 98 | 4.632 | 21 | 83 |
| Energiemetabolismus | 159 | 3,28 | 94 | <u>114</u> | 8 | 10 |
| Fettsäure- und Phospholipidmetabolismus | 47 | 0,97 | 81 | 397 | 7 | 15 |
| Hypothetische und unklassifizierte Proteine | 1.657 | 34,2 | <u>812</u> | <u>10.749</u> | 158 | <u>55</u> |
| Membranproteine | 267 | 5,52 | <u>72</u> | <u>327</u> | 17 | 25 |
| Motilität und Anheftung | 44 | 0,91 | 286 | 492 | 15 | 31 |
| Nukleotidmetabolismus | 68 | 1,40 | <u>6</u> | <u>25</u> | 0 | 3 |
| Proteinsekretion/Exportapparat | 53 | 1,09 | 35 | <u>1.149</u> | 8 | 30 |
| putative Enzyme | 367 | 7,59 | <u>51</u> | <u>2.455</u> | 32 | <u>13</u> |
| Phage, Transposon, Plasmid | 6 | 0,12 | <u>124</u> | 3.332 | 20 | 50 |
| sekretierte Faktoren (Toxine, Enzyme, Alginate) | 40 | 0,82 | 80 | <u>17</u> | 2 | 4 |
| Transkription, RNA-Prozessierung und -Degradation | 49 | 1,01 | 49 | 28 | 5 | 3 |
| Transkriptionelle Regulatoren | 398 | 8,23 | <u>37</u> | <u>617</u> | <u>12</u> | <u>18</u> |
| Translation, posttranslationale Modifikation | 143 | 2,95 | 96 | <u>1.086</u> | 11 | 28 |
| Transport kleiner Moleküle | 527 | 10,8 | 64 | 623 | 42 | 20 |
| Zwei-komponenten Regulationssystem | 57 | 1,17 | 387 | 68 | 12 | 11 |

signifikant ($p < 0,05$) über- (**fett**) oder unterrepräsentiert (**fett, unterstrichen**)

innerhalb des Klons verglichen mit den Werten für das Pangenom.

Alle p -Werte wurden korrigiert durch multiples Vergleichen nach dem Bonferroni Algorithmus.

3.2.4 Intraklonale Bereiche hoher Mutationsfrequenz

Im Wissen über die globale Frequenz der Mutationen innerhalb der Genome wurde im nächsten Schritt nach Bereichen mit hoher Mutationsfrequenz (Hot-Spots) im Kerngenom der sequenzierten Stämme gesucht. Neben zahlreichen Phagen- und Plasmid-assoziierten Proteinen in beiden Klonen wurde das Schwermetallion-Efflux-Protein CusA, die zyklische-di-GMP Phosphodiesterase BifA (Kuchma et al., 2007) und der Hauptregulator des Quorum Sensings, LasR (Williams und Camara, 2009), in Klon C als Ziele wiederkehrender, nicht-konservativer Aminosäureaustausche identifiziert. Dabei war davon auszugehen, dass einige dieser Austausche die Struktur sowie die Proteinfunktion modifizieren können. Im Falle von LasR betrafen die Aminosäureaustausche den Helixbrechenden Einbau eines Prolins oder waren in konservierten Positionen der Alpha-Helices 8 und 10 und in der Bindungstasche des Homoserin-Lakton-Autoinducers lokalisiert (Bottomley et al., 2007).

Zusätzlich zur Suche nach Genen mit akkumulierten Aminosäureveränderungen wurden die Genome auf Segmente mit signifikant erhöhten SNP-Raten ($FDR < 0,05$) untersucht. Die Klon C Stämme zeigten zwölf Mutations-Hot-Spots, wovon zehn intergenische Regionen oder Phagen- und Plasmid-bezogene Gene betrafen (Tab. S5). Die zwei betroffenen Gene des Kerngenoms waren der Transkriptionsregulator PA2020 und erneut *lasR*. Von den 60 in Klon PA14 gefundenen Segmenten waren 44 im Kern- und 16 im akzessorischen Genom lokalisiert. Bereiche hoher Mutationsfrequenz waren vorwiegend Phagen- und Plasmidbezogene Gene sowie funktionell noch nicht charakterisierte offene Lesebereiche (ORFs). Der Stamm PT2 hatte nahezu alle SNPs in den 22 Genen akkumuliert, welche die RGP31 flankieren (Tab. S5). Diese lokale Akkumulation kann durch Rekombination mit einem anderen Klon hervorgerufen worden sein.

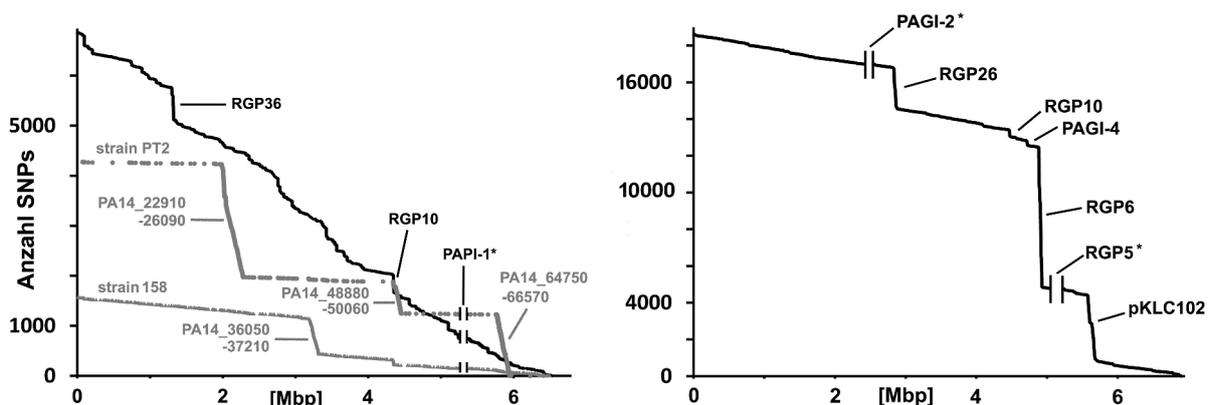


Abbildung 3.19: Kumulative Kaplan-Meier-Plots der SNP-Frequenz innerhalb der Klon C Isolate ($n=57$, rechts) und der Klon PA14 Isolate ($n=40$, sowie zwei separate Ausreißer, links). Regionen mit hoher Mutationsfrequenz sind durch RGP- oder ORF-Nummern gekennzeichnet. SNPs in den, mit einem '*' markierten RGPs wurden aufgrund ihrer hohen Zahl von vornherein ausgeschlossen.

Im Gegensatz zu diesem Einzelfall trugen viele PA14 Stämme Mutationen in funktionell charakterisierten Genen, was Ausdruck für eine diversifizierende Selektion war. Beispiele dafür waren die Gene *pchF*, *rocs2*, und *pelA*, deren Genprodukte in die Kommunikation von *P. aeruginosa* mit seiner Umwelt involviert sind. PchF ist an der nicht-ribosomalen Biosynthese des Siderophors Pyochelin beteiligt (Patel und Walsh, 2001); der Transkriptionssensor RocS2 kontrolliert die Biogenese von CupC-Fimbrien und den Antibiotika Transport (Sivaneson et al., 2011) und PelA deazetyliert die Pel Exopolysaccharide, welche essentiell für die Biofilmbildung sind (Colvin et al., 2013).

Die Visualisierung der Lokalisation von Nukleotidaustauschen gegen die Genomkoordinaten des jeweiligen Referenzstamms zeigte, dass in dem klonalen Komplex C fast alle Mutationen in Komponenten des akzessorischen Genoms akkumulierten, was sich in der Treppenform des Graphs darstellte (Abb. 3.19). 40 Stämme des PA14 Komplexes verhielten sich vergleichbar zu Klon C, jedoch zeigten die Ausreißer PT2 (höchste SNP-Anzahl) und 158 (mehr als 1.000 SNPs nach Abzug der Mutationen der PAPI-1 Insel; Tab. 3.8) auch in Bereichen des Kerngenoms Akkumulationen von SNPs, was als ein Indiz für eine Rekombination gedeutet wurde.

3.2.5 Sequenzdiversität der nicht-kodierenden RNAs

Sequenzdiversität wurde nicht nur in den kodierenden Genen, sondern auch in nichtkodierenden RNAs untersucht. Basierend auf den ncRNAs von NN2 (diese Arbeit) und PA14 (Wurtzel et al., 2012) wurden SNPs in 21% (NN2) bzw. 14% (PA14) der ncRNAs detektiert (Tab. S6). In Klon PA14 wurden 95 SNPs nur jeweils in einem und 35 SNPs in zwei bis neun Isolaten gefunden. Klon C zeigte eine höhere Sequenzdiversität sowohl im Auftreten, als auch der Häufigkeit der SNPs (Tab. S6, Abb. 3.20). So wurden 573 SNPs in 68 ncRNAs identifiziert (PA14: 130 SNPs in 28 ncRNAs). Dabei muss erwähnt werden, dass die untersuchten ncRNAs von PA14 allesamt intergenisch lokalisiert sind und für beide Klone festgestellt wurde, dass der größte Anteil an Nukleotidaustauschen innerhalb von ORFs stattfindet (Tab. 3.10, Tab. S6, S7). Der Einfluss der SNPs auf die Sekundärstruktur der RNAs wurde *in silico* (Gruber et al., 2008) für alle diejenigen SNPs untersucht (Tab. S8), die in mehr als zwei Stämmen auftraten. Die vorhergesagte Topologie von Haarnadel- und Stammstrukturen und deren globale thermodynamische Stabilität unterschied sich zwischen dem Referenzgenom und der Mutante in zehn (Klon C) respektive sechs (PA14) der 19 untersuchten ncRNAs (Tab. S8). Diese Daten legen nahe, dass die intraklonale Sequenzvariation einen Einfluss auf die Struktur, die Halbwertszeit und die Funktion von ungefähr 10% der ncRNAs eines klonalen Komplexes in *P. aeruginosa* haben kann.

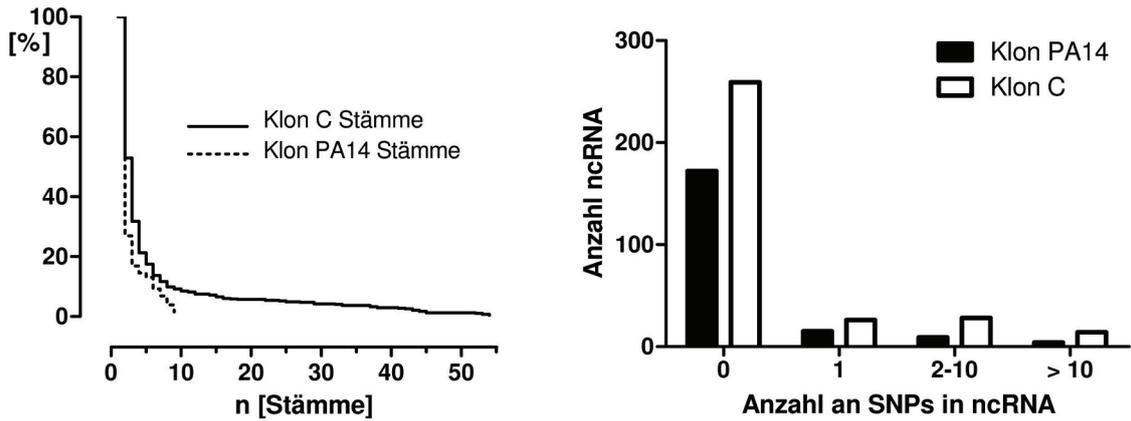


Abbildung 3.20: Sequenzdiversität von nicht-kodierenden RNAs. Die Anzahl der Stämme mit Nukleotidaustauschen in ncRNAs (links) und die Anzahl von SNPs in den ncRNAs (rechts).

3.2.6 Insertionen, Deletionen und Leserasterverschiebungen

In den vorangegangenen Kapiteln wurde beschrieben, wie die klonalen Komplexe hinsichtlich ihrer Zusammensetzung des akzessorischen Genoms aufgebaut sind und wie sich Einzelnukleotidaustausche im Genom darstellen. Die genomische Diversität wurde jedoch auch durch das Auftreten von kleinen und großen Insertionen und Deletionen moduliert.

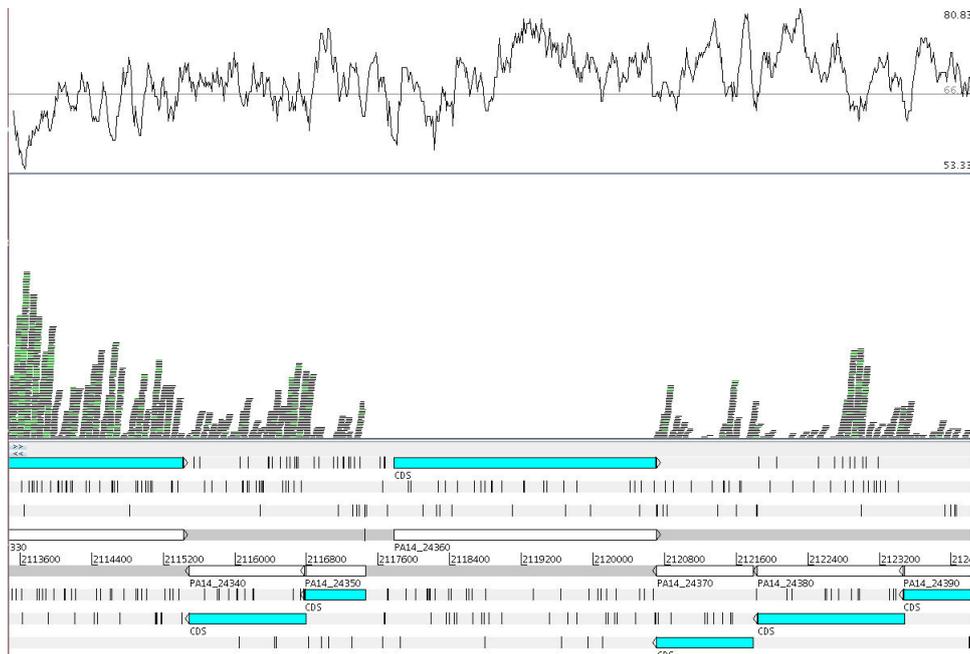


Abbildung 3.21: Exemplarischer Ausschnitt einer großen Deletion im Stamm PT2. Der obere Bereich zeigt den GC-Gehalt, der mittlere die Verteilung der Reads und der untere Teil die Annotation des betrachteten Genomausschnitts. Deutlich erkennbar fehlen Reads für das Gen PA14_24360 in der Bildmitte. Im Bereich der ORFs PA14_24370 und PA14_24380 lässt sich zudem die Korrelation zwischen erhöhtem GC-Gehalt und verringerter Sequenzabdeckung erkennen.

Große Deletionen wurden identifiziert, indem nach Regionen ohne Abdeckung zwischen kartierten Reads gesucht wurde. Da Bereiche mit hohem GC-Gehalt, genauso wie tatsächliche Deletionen, zu einer Verringerung der Abdeckung führten (Abb. 3.21), wurden alle ermittelten Kandidatenpositionen manuell daraufhin untersucht, ob auch tatsächlich eine Deletion vorlag oder nur ein Sequenzierungsartefakt. Da vereinzelt Reads aufgrund von homologen Sequenzen in deletierte Bereiche kartiert wurden, war es nicht ausreichend, Regionen mit einer Abdeckung von null Reads als Deletionen zu definieren und in den Kartierungsdateien zu suchen.

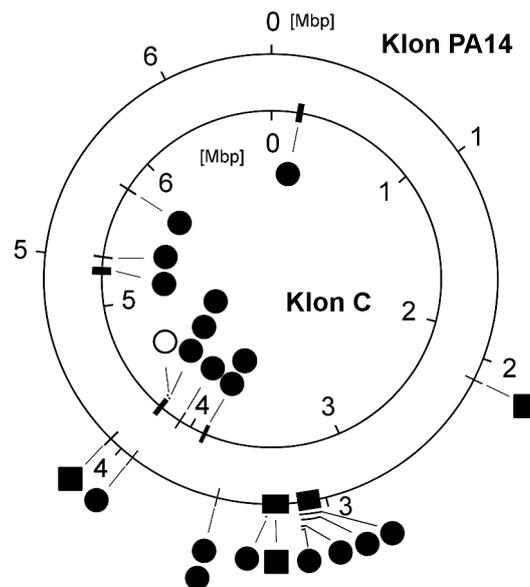


Abbildung 3.22: Deletionen des Kerngenoms für Klon C (innerer Ring) und Klon PA14 Isolate (äußerer Ring) in verschiedenen Habitaten (geschlossener Kreis: chronische Infektion; offener Kreis: akute Infektion; Quadrat: Umweltisolat).

In 18 Stämmen, verteilt auf beide klonalen Komplexe, wurde das konservierte Kerngenom durch Deletionen von jeweils mehr als 1 kb Länge unterbrochen (Abb. 3.22). Dabei akkumulierten in Stämmen des klonalen Komplexes PA14 die Deletionen in der Nähe des Replikationsterminus, wobei vier 9 - 144 kb große Deletionen einen gemeinsamen Bruchpunkt teilten. Die fehlenden Gene dieser großen Blöcke betrafen Gene des Siderophors Pyoverdin, Stoffwechselwege des Kohlenhydrat- und Aminosäuremetabolismus und die extrazelluläre Matrixkomponente Psl, welche die Abwehrlinie gegen Antibiotika während der initialen Phase der Biofilmbildung darstellt (Colvin et al., 2012). Das Las quorum-sensing System fehlte in einem PA14-Umweltisolat und vier klinischen Klon C Isolaten. Andere Deletionen entfernten das Adhesin CdrA, die c-di-GMP Synthetase SiaD sowie Cytochrom C Oxidase-, Chemotaxis- oder Typ-III-Sekretionsoperons (Tab. S9).

Neben den großen Deletionen, welche komplette Genblöcke aus den Genomen entfernten, waren auch zahlreiche kleine Insertionen und Deletionen (Indels) zu finden. Diese Indels mit einer Länge kürzer als zehn Nukleotide akkumulierten in den nicht-kodierenden Re-

gionen des Klon C oder Klon PA14 Chromosoms ($p < 10^{-6}$) (Abb. 3.23). Eine Insertion oder Deletion, welche das Leseraster der betroffenen kodierenden Region nicht verschob, wurde in 25% (C) bzw. 30% (PA14) der Stämme beobachtet. Die Mehrheit dieser Indels sollte keine funktionellen Konsequenzen nach sich ziehen und somit als neutral klassifiziert werden. Im Gegensatz dazu wird von Leserasterverschiebungen angenommen, dass sie zu verkürzten und/oder nicht-funktionalen Genprodukten führen. Leserasterverschiebungen betrafen vor allem ORFs von bislang nicht charakterisierter Funktion (Abb. 3.23, $p = 0,029$; Exakter Test nach Fisher). Bemerkenswert war, dass Isolate mit fünf oder mehr Leserasterverschiebungen einzig in einer Subgruppe von CF-Isolaten auftraten, wovon acht Isolate 18 - 44 Leserasterverschiebungen zeigten. Dabei ist davon auszugehen, dass die Verschiebungen einige funktionell charakterisierte Gene inaktivieren, welche für Elemente der Motilität, der Adhärenz, der Biofilmbildung, des Antibiotikatransports, der Signalübermittlung oder der Virulenz kodieren (Tab. S10).

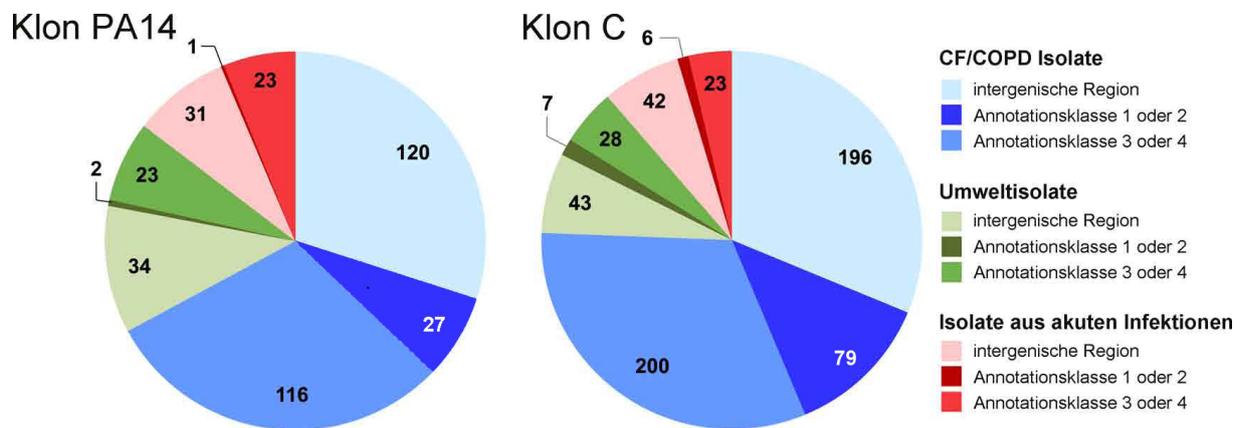


Abbildung 3.23: Verteilung der Indels sortiert nach Habitat und Annotationsklassen.

Die Hot-Spots der Leserasterverschiebungen in CF-Isolaten waren das multiple Antibiotikaresistenz Efflux Operon *mexR-mexA-mexB-oprM* (6x), der Polyamintransporter *PotA* (4x) und das Schlüsseltransportprotein der Fettsäuresynthese *FabG* (4x). Fünf Klon C und alle Klon PA14 Isolate trugen Leserasterverschiebungen in *ladS*. Diese hybride Sensorkinase kontrolliert die reziproke Expression für Gene der Typ-III-Sekretion und von Biofilm-fördernden Polysacchariden (Ventre et al., 2006). Von der Leserasterverschiebung in *ladS* konnte gezeigt werden, dass sie die Aktivität des Typ-III-Sekretionssystems in Klon PA14 erhöht (Mikkelsen et al., 2011).

3.2.7 Identifikation von stammspezifischen Genen

Alternativ zu Einzelnukleotidaustauschen sowie kleinen und großen Insertionen und Deletionen, kann die stammspezifische Akquise von Genen die Fitness, die Lebensweise und die metabolische Kompetenz eines klonalen Komplexes modulieren. Ein umfangreiches Repertoire an zusätzlichen Genen wurde in den individuellen Stämmen des Klon C und

auch des Klon PA14 Komplexes detektiert (Tab. 3.14). Im Median hatte ein Klon PA14 oder Klon C Isolat 170 bzw. 103 Gene aufgenommen ($p < 0,001$ für den Vergleich zusätzlicher Gene zwischen PA14 und C) (Tab. 3.16, 3.17). Dabei wurden 80% der nächsten Orthologen in anderen, bereits annotierten *P. aeruginosa* Stämmen oder Mitgliedern des Genus *Pseudomonas* identifiziert. Der Ursprung der restlichen 20% der zusätzlich erworbenen Gene wurde phylogenetisch entfernteren Taxa zugeordnet.

Tabelle 3.15: Übersicht über stammspezifische Gene in den klonalen Komplexen C und PA14.

| Funktionelle Kategorie | stammspezifische Gene | |
|---|-----------------------|--------------|
| | Klon PA14 | Klon C |
| Adaptation, Schutz | 101 | 127 |
| Aminosäurebiosynthese und -metabolismus | 88* | <u>6</u> |
| Antibiotikaresistenz und -suszeptibilität | 19 | 40 |
| Biosynthese von Kofaktoren und prosthetischen Gruppen | <u>14</u> | <u>8</u> |
| Kohlenstoffkatabolismus | 62 | 275* |
| Zellteilung | 37 | 13 |
| Zellwand, LPS, Kapsel | 15 | 62* |
| zentraler Intermediärstoffwechsel | <u>0</u> | <u>18*</u> |
| Chaperone | <u>0</u> | 7 |
| Chemotaxis | <u>0</u> | 4 |
| DNA Replikation, Modifikation, Reparatur | 88 | 187* |
| Energiemetabolismus | <u>41</u> | 132* |
| Fettsäure- und Phospholipidmetabolismus | 45* | <u>0</u> |
| Hypothetische und unklassifizierte Proteine | 2.177 | 2.138 |
| Membranproteine | <u>11</u> | <u>0</u> |
| Motilität und Anheftung | 56 | 47 |
| Nukleotidmetabolismus | 101 | 53 |
| Proteinsekretion/Exportapparat | <u>4</u> | 11 |
| putative Enzyme | <u>0</u> | <u>0</u> |
| Phage, Transposon, Plasmid | 450* | 210 |
| sekretierte Faktoren (Toxine, Enzyme, Alginate) | 46* | <u>0</u> |
| Transkription, RNA-Prozessierung und -Degradation | <u>0</u> | <u>0</u> |
| Transkriptionelle Regulatoren | <u>170</u> | <u>95</u> |
| Translation, postranslationale Modifikation | 58 | <u>50</u> |
| Transport kleiner Moleküle | <u>149</u> | <u>201</u> |
| Zwei-komponenten Regulationssystem | 18 | <u>11</u> |

signifikant ($p < 0,05$) über- (**fett**) oder unterrepräsentiert (**fett, unterstrichen**) im Klon verglichen mit dem Pangenom.

*=signifikanter ($p < 0,05$) Unterschied zwischen Klon C und Klon PA14.

Alle p -Werte wurden korrigiert durch multiples Vergleichen nach dem Bonferroni Algorithmus.

Tabelle 3.15 fasst das gesamte Repertoire der stammspezifischen Gene der beiden klonalen Komplexe sortiert nach deren funktioneller Kategorie zusammen. Gene der mobilen genetischen Elemente wie Phagen und Plasmide sowie hypothetische Gene mit Produkten unbekannter Funktion waren signifikant überrepräsentiert innerhalb der stammspezifischen Gene beider klonalen Komplexe. Diese Erkenntnis entsprach der Erwartung, da Phagen, Transposons und Plasmide die klassischen Vehikel des horizontalen Gentransfers darstellen, welcher Gene in einen individuellen Stamm oder einen klonalen Komplex integriert. Im Gegensatz dazu waren Gene, welche für Transkriptionselemente oder Elemente des intermediären Metabolismus kodieren, selten oder gar nicht innerhalb der stammspezifischen Gene vertreten.

Nichtsdestotrotz konnte ein unterschiedliches Repertoire an Genen identifiziert werden, welche die metabolische Kompetenz der Bakterien, Substrate umzusetzen, fördern. Die Stämme des klonalen Komplexes PA14 trugen im Vergleich zu Klon C eine signifikant größere Anzahl von Genen des Aminosäure- oder Fettsäuremetabolismus und der sezernierten Faktoren. Auch kodierten die PA14 Isolate zusätzlich mehr mobile Gene, welche ihren Ursprung in Phagen, Transposons und Plasmiden hatten. Eine mögliche Begründung dafür ist die etwas geringere Genomgröße der PA14 Referenz (6,5 Mbp) verglichen mit der Klon C Referenz (6,9 Mbp). Da die unterschiedliche Genomgröße mit zusätzlichen Genen des akzessorischen Genoms erklärt werden kann, haben die Klon PA14 Stämme noch das Potential, mehrere Gene zusätzlich in ihr Genom zu inkorporieren. Die Stämme des klonalen Komplexes C kodierten, verglichen mit PA14, ein größeres Repertoire an Elementen des Kohlenstoffmetabolismus, des Intermediärstoffwechsels und des Energiemetabolismus (Tab. 3.15). Diese zusätzlichen Gene der Kohlenstoffverwertung können eine Begründung dafür sein, warum Klon C eine dominierende Stellung innerhalb der weltweiten Population einnimmt. Eine detaillierte Auflistung der stammspezifischen Gene befindet sich in Tabelle S11.

Tabelle 3.16: Stammspezifische Gene im klonalen Komplex PA14. Taxonomische Herkunft des orthologen Gens höchster Homologie anhand von BLAST-Resultaten.

| Isolat | Zahl an stammspez. Genen | <i>P. aeru- ginosa</i> | Pseudo- monaden | γ -Proteo- bakterien | Andere Taxa |
|------------|-----------------------------|----------------------------|--------------------|--------------------------------|----------------|
| 6 | 114 | 52 | 42 | 4 | 16 |
| 12_D2 | 249 | 112 | 93 | 22 | 22 |
| 12_SA1 | 175 | 92 | 45 | 18 | 20 |
| 37 | 286 | 127 | 116 | 9 | 34 |
| 55F6 | 107 | 62 | 33 | 2 | 10 |
| 61Pf | 213 | 105 | 81 | 7 | 20 |
| 86PA2 | 107 | 75 | 25 | 3 | 4 |
| 110_E8 | 284 | 97 | 117 | 28 | 42 |
| 120B6 | 165 | 93 | 49 | 10 | 13 |
| 158 | 493 | 272 | 126 | 46 | 49 |
| 209 | 109 | 50 | 43 | 5 | 11 |
| 212 | 116 | 53 | 45 | 3 | 15 |
| 228 | 96 | 37 | 44 | 3 | 12 |
| 338 | 230 | 129 | 69 | 8 | 24 |
| 487 | 17 | 6 | 1 | 0 | 10 |
| AL159 | 37 | 27 | 9 | 0 | 1 |
| CF6 | 376 | 143 | 157 | 39 | 37 |
| CF12 | 470 | 186 | 172 | 42 | 70 |
| CF13 | 498 | 161 | 254 | 30 | 53 |
| K2 | 286 | 147 | 81 | 21 | 37 |
| K6 | 162 | 97 | 48 | 4 | 13 |
| K9 | 532 | 271 | 159 | 37 | 65 |
| K10 | 202 | 115 | 46 | 20 | 21 |
| K12 | 247 | 162 | 52 | 14 | 19 |
| PA14 | 0 | 0 | 0 | 0 | 0 |
| PT2 | 420 | 275 | 89 | 16 | 40 |
| RN3 | 251 | 145 | 59 | 15 | 32 |
| RZ2 | 320 | 155 | 152 | 11 | 2 |
| SS8 | 165 | 113 | 38 | 7 | 7 |
| Zw85 | 79 | 47 | 26 | 1 | 5 |
| 202 | 112 | 51 | 40 | 6 | 15 |
| 250 | 187 | 130 | 43 | 9 | 5 |
| 324 | 78 | 51 | 19 | 3 | 5 |
| CF1 | 198 | 104 | 68 | 5 | 21 |
| CF4 | 87 | 44 | 27 | 10 | 6 |
| K3 | 83 | 57 | 18 | 2 | 6 |
| K4 | 215 | 128 | 55 | 8 | 24 |
| K7 | 159 | 83 | 61 | 8 | 7 |
| K8 | 162 | 70 | 46 | 4 | 42 |
| K14 | 147 | 89 | 35 | 7 | 16 |
| WU5 | 178 | 151 | 22 | 2 | 3 |
| Median | 170 | 97 | 47 | 8 | 16 |
| Mittelwert | 203 | 105 | 66 | 12 | 21 |
| Summe | 8.526 | 4.412 | 2.753 | 494 | 867 |

Tabelle 3.17: Stammspezifische Gene im klonalen Komplex C. Taxonomische Herkunft des orthologen Gens höchster Homologie anhand von BLAST-Resultaten.

| Isolat | Zahl an stammspez. Genen | <i>P. aeru- ginosa</i> | Pseudo- monaden | γ -Proteo- bakterien | Andere Taxa |
|------------|-----------------------------|----------------------------|--------------------|--------------------------------|----------------|
| 10SA1 | 403 | 104 | 124 | 171 | 4 |
| 12_F8 | 96 | 61 | 16 | 4 | 15 |
| 12_SB1 | 216 | 108 | 36 | 22 | 50 |
| 17 | 79 | 61 | 4 | 2 | 12 |
| 28PA3 | 243 | 88 | 142 | 5 | 8 |
| 36 | 109 | 70 | 24 | 3 | 12 |
| 69 | 70 | 49 | 6 | 6 | 9 |
| 84 | 248 | 97 | 70 | 20 | 61 |
| 87 | 128 | 71 | 28 | 8 | 21 |
| 97 | 89 | 78 | 5 | 0 | 6 |
| 110 | 188 | 133 | 31 | 8 | 16 |
| 110_D4 | 156 | 98 | 36 | 8 | 14 |
| 115 | 86 | 59 | 18 | 1 | 8 |
| 124 | 64 | 53 | 0 | 0 | 11 |
| 142 | 128 | 80 | 29 | 6 | 13 |
| 159 | 227 | 120 | 71 | 20 | 16 |
| 179 | 64 | 51 | 4 | 1 | 8 |
| 214 | 104 | 65 | 25 | 6 | 8 |
| 237 | 48 | 41 | 1 | 0 | 6 |
| 240 | 16 | 8 | 3 | 1 | 4 |
| 343 | 77 | 60 | 5 | 2 | 10 |
| 426 | 89 | 65 | 8 | 4 | 12 |
| 584 | 63 | 53 | 4 | 1 | 5 |
| 642 | 110 | 60 | 25 | 5 | 20 |
| 702_D3 | 86 | 58 | 10 | 1 | 17 |
| 747 | 165 | 96 | 33 | 17 | 19 |
| 800 | 57 | 50 | 1 | 0 | 6 |
| 902_D1 | 214 | 106 | 28 | 69 | 11 |
| AL962 | 197 | 83 | 77 | 12 | 25 |
| C17 | 140 | 78 | 23 | 14 | 25 |
| CF2 | 350 | 213 | 88 | 26 | 23 |
| CF3 | 189 | 115 | 58 | 4 | 12 |
| CF5 | 177 | 110 | 53 | 6 | 8 |
| CF7 | 91 | 71 | 11 | 2 | 7 |
| CF8 | 193 | 124 | 50 | 5 | 14 |
| CF10 | 121 | 83 | 20 | 7 | 11 |
| CF11 | 115 | 88 | 14 | 4 | 9 |
| K5 | 56 | 45 | 2 | 2 | 7 |
| K11 | 96 | 74 | 12 | 2 | 8 |
| K13 | 100 | 68 | 12 | 1 | 19 |
| K15 | 57 | 44 | 1 | 2 | 10 |
| K16 | 155 | 86 | 21 | 23 | 25 |
| K17 | 516 | 244 | 193 | 31 | 48 |
| K18 | 188 | 127 | 35 | 6 | 20 |
| K19 | 199 | 89 | 82 | 11 | 17 |
| K20 | 104 | 63 | 13 | 10 | 18 |
| PJC | 70 | 57 | 4 | 1 | 8 |
| SG1 | 1 | 0 | 0 | 1 | 0 |
| SG17m | 90 | 72 | 5 | 1 | 12 |
| SG50m | 59 | 47 | 3 | 1 | 8 |
| SS30 | 103 | 87 | 3 | 5 | 8 |
| W5Aug28 | 90 | 67 | 7 | 2 | 14 |
| WI1 | 79 | 65 | 4 | 1 | 9 |
| WU2 | 126 | 108 | 4 | 7 | 7 |
| B6470 | 55 | 46 | 2 | 1 | 6 |
| 10_C1 | 58 | 53 | 1 | 1 | 3 |
| 10SB1 | 84 | 69 | 14 | 1 | 0 |
| Median | 103 | 71 | 14 | 4 | 11 |
| Mittelwert | 131 | 79 | 28 | 10 | 14 |
| Summe | 7.482 | 4.519 | 1.599 | 581 | 783 |

3.2.8 Diskussion der Ergebnisse des intraklonalen Genomvergleiches

Im Hauptteil dieser Arbeit wurden 57 Klon C und 42 Klon PA14 Isolate mit der SOLiD Technologie sequenziert. Basierend auf den so gewonnenen Genomdaten konnte die Diversität und das genetische Repertoire innerhalb der beiden klonalen Komplexe untersucht und mit unverwandten Stämmen verglichen werden (Hilker et al., 2015). Die Isolate wurden sowohl aus akuten und chronischen Patienteninfektionen als auch aus der Umwelt isoliert. Die meisten Isolate der chronischen CF-Infektion waren Erstisolate von - bis dahin - Pseudomonas-freien Patienten, und es ließ sich nicht mit abschließender Sicherheit klären, ob die Bakterien aus der Umwelt stammten oder direkt von anderen Patienten übertragen worden waren. Zu vermuten ist, dass die Infektionen in den meisten Fällen von Umweltisolaten ausgingen (Gibson et al., 2003) und nicht von bereits stark an das individuelle Lungenhabitat einzelner Patienten angepassten Bakterien. Bei Festlegung der Stämme wurde darauf geachtet, aus vorliegenden sequentiellen Isolaten nur das Erstisolat auszuwählen, um die Diversität unabhängig von Evolutionsereignissen während der Adaptation an die CF-Lunge, die bereits ausgiebig untersucht wurde, zu betrachten (Cramer et al., 2011, 2012; Rau et al., 2012; Jørgensen et al., 2015; Williams et al., 2015). Ein Großteil der Isolate wurde in Deutschland isoliert (71%), knapp ein Drittel der Stämme stammt aus anderen Ländern (29%). Isolate aus Südamerika, Afrika, Asien und Australien waren in der Analyse nicht vertreten. Trotz dieser Überrepräsentation einiger Regionen in den Klon C und Klon PA14 Sammlungen konnte jedoch, mit Ausnahme weniger Gemeinsamkeiten in der Zusammensetzung des akzessorischen Genoms, keine Gruppierung von bestimmten Genomcharakteristika innerhalb eines Habitats festgestellt werden. Dieser Umstand kann als Beweis dafür angesehen werden, dass die Isolate unabhängig voneinander evolviert sind.

Für die Sequenzierungen galt es, die bekannten Probleme, die sich aus dem hohen GC-Gehalt von *P. aeruginosa* (66%) ergeben, zu minimieren. Dazu wurde für die Sequenzierung der GC-reichen *P. aeruginosa* Isolate eine Optimierung des Sequenzierungsprotokolls durchgeführt, um sicherzustellen, dass auch besonders GC-reiche Abschnitte ausreichend durch Sequenzierungsreads repräsentiert werden und so möglichst das gesamte Genom der einzelnen Stämme abgedeckt wird (vgl. Kap. 2.4.2). Die Tabellen 3.8 und 3.9 zeigen, dass bei nur wenigen Ausnahmen konstant mehr als 90% des jeweiligen Referenzgenoms NN2 oder PA14 abgedeckt wurden. Selbst die Isolate mit der geringsten Anzahl an Reads zeigten mehr als 80% der Sequenz der Referenzgenome. Eine komplette Abdeckung der Referenzsequenz in einem Großteil der Stämme war nicht zu erwarten, da jedes Isolat eine stammspezifische Signatur des akzessorischen Genoms aufwies, die sich von der des Referenzgenoms unterschied. Hier stellt sich die Frage, ob eine höhere Genomabdeckung bei den damit verbundenen Kosten tatsächlich erstrebenswert wäre, da nicht sichergestellt

ist, dass die fehlenden Bereiche zur vollständigen Abdeckung des Referenzgenoms überhaupt in den Stämmen kodiert sind. So trugen die Isolate andere genomische Inseln und RGP als die Referenzstämme NN2 und PA14 oder zumindest andere Varianten dieser mobilen Elemente. Ferner kann es sich bei den nicht durch Reads abgedeckten Bereichen ebenso um hochrepetitive Elemente handeln, die bei Sequenzierungen mit relativ kurzen Reads (75 bp) oft nicht rekonstruierbar sind. In diesem Fall wäre es erstrebenswert, die Sequenzierung mit langen Reads (> 1.000 bp) durchzuführen, die z.B. bei der PacBio-Sequenziermethode (Firma Pacific Biosciences) möglich sind (Au et al., 2012).

Im Rahmen dieser Arbeit wurde gezeigt, dass in beiden klonalen Komplexen das Kerngenom stark konserviert ist und nur wenige Mutationen in diesen Bereichen lokalisiert sind. Diese Aussage wurde sowohl durch die sehr langen Bereiche gemeinsamer DNA („Haplotypen“), deren Länge im Median 99 kb in Klon C und 163 kb in Klon PA14 betrug, als auch durch die geringe Anzahl der Einzelnukleotidaustausche (SNPs) innerhalb der klonalen Komplexe verdeutlicht (Tab. 3.10). Bei der Bestimmung der Haplotypen werden Genominversionen, -insertionen und -deletionen der Stämme im Vergleich zur Referenzsequenz außer Acht gelassen. Dieser Fehler ist zu tolerieren, da Genomrearrangements, die einen Einfluss auf die Länge syntentischer DNA-Blöcke haben, zu den seltenen Ereignissen innerhalb der klonalen Komplexe zählen.

Zeigten unverwandte Stämme im Vergleich zur Referenz PAO1 mindestens 20.000 Einzelnukleotidaustausche (Hilker et al., 2015), wiesen die Stämme der klonalen Komplexe C (2.568) und PA14 (156) im Median weniger SNPs gegenüber den jeweiligen Referenzgenomen NN2 bzw. PA14 auf. Das Klon PA14 Isolat PT2 zeigte von allen untersuchten Isolaten den Maximalwert von 6.364 SNPs, was jedoch nur einem Viertel der SNPs im Vergleich zu der Anzahl zwischen unverwandten Klonen entsprach. In Fällen mit mehr als 1.000 SNPs pro Isolat wurde festgestellt, dass die Punktmutationen hauptsächlich in den mobilen Elementen lokalisiert waren. Diese Tatsache deutete darauf hin, dass in der *P. aeruginosa* Population verschiedene Varianten einiger genomischer Inseln präsent sind. Substrahiert man die SNPs in den Bereichen der genomischen Inseln, reduziert sich die Anzahl der SNPs in beiden klonalen Komplexen im Median auf unter 150 Einzelnukleotidaustausche pro Stamm. Das bedeutet, dass nach vielen Millionen Bakteriengenerationen innerhalb eines klonalen Komplexes ein chromosomales Rückgrat des Kerngenoms geschaffen wurde, welches kaum noch Veränderungen durchläuft. Auch der sternförmige phylogenetische Baum verdeutlichte diese Tatsache. So waren die Stämme unabhängig voneinander, und es ließ sich keine habitatbezogene Segregation innerhalb der Einzelnukleotidaustausche des Kerngenoms feststellen. Nur in wenigen Fällen wurden mehrere Isolate auf einem gemeinsamen Ast positioniert und teilen sich somit ein zu 100% identisches Kerngenom. Die Mehrzahl der Isolate war verstreut um einen gemeinsamen Ursprung angeordnet, was ein weiterer Hinweis darauf war, dass das Kerngenom der klonalen Komplexe gefestigt ist und nicht in eine Richtung evolviert. Dies bestätigt, dass in der Auswahl der Stämme mit der

Überrepräsentation von Isolaten aus Deutschland kein entscheidender Fehler begangen worden war.

Studien von CF-Langzeitverläufen hatten gezeigt, dass in den CF-Isolaten während des Verlaufs nur wenige zusätzliche Mutationen hinzukommen, solange sich kein Hypermutatorphänotyp herausbildet (Cramer et al., 2012; Markussen et al., 2014). Untersuchte man die Einzelnukleotidaustausche der Stämme getrennt nach Habitaten, ließen sich beim Vergleich der intraklonalen Diversität zwischen den klonalen Komplexen C und PA14 sowohl Gemeinsamkeiten als auch Unterschiede finden. Dabei galt zu beachten, dass lediglich sechs Isolate des klonalen Komplexes PA14 aus der Umwelt stammen. Schon wenige Isolate mehr wären in der Lage, das Bild nachhaltig zu verändern oder zu festigen. Sollte sich die Möglichkeit ergeben, weitere Umweltstämme zu untersuchen, sollte diese Möglichkeit wahrgenommen werden, um den gewonnenen Gesamteindruck weiter zu festigen.

Die Suche nach Genomregionen mit einer hohen Mutationsfrequenz zeigte, dass die meisten Mutationen in mobilen Elementen oder Phagen- und Plasmid-bezogenen Genen lokalisiert waren. Eine Ausnahme davon war der Stamm PT2, welcher eine große SNP-Zahl in den 22 Genen des Kerngenoms akkumulierte, welche die kleine mobile Insel RGP31 des akzessorischen Genoms flankierten (Tab. S5). Diese lokale Akkumulation wurde vermutlich durch homologe Rekombination eines Teils des Kerngenoms mit einem anderen Klon verursacht.

Einen Kontrast zum hohen Konservierungsgrad des Kerngenoms stellte die Diversität des akzessorischen Genoms dar, welches sowohl eine hohe Variabilität in der Zusammensetzung von genomischen Inseln und Regionen genomischer Plastizität (RGP), als auch in den Einzelnukleotidaustauschen dieser Bereiche aufwies. Diese hohe Diversität konnte schon in unverwandten Stämmen nachgewiesen werden (Hilker et al., 2015). Bemerkenswert dabei war die geringe Ähnlichkeit innerhalb der Zusammensetzung des akzessorischen Genoms. So konnte in dieser Arbeit durch paarweise Vergleiche der Zusammensetzung des akzessorischen Genoms von Vertretern der klonalen Komplexe C und PA14 gezeigt werden, dass diese nur eine minimal höhere Ähnlichkeit zueinander aufwiesen als unverwandte Stämme (vgl. Kap. 3.2.2). Somit weist jedes Isolat eine stammspezifische Zusammensetzung des akzessorischen Genoms auf. Trotz dieser hohen Plastizität konnte für das akzessorische Genom in einigen Fällen eine habitatbezogene geografische Segregation festgestellt werden. Ergebnisse für Isolate, welche innerhalb von drei Jahren aus einer Fläche vergleichbar mit Niedersachsen isoliert wurden, verdeutlichten, dass es einen vom Kerngenom unabhängigen DNA-Pool des akzessorischen Genoms gibt, welcher innerhalb lokal begrenzter Gebiete zwischen den Isolaten ausgetauscht wird.

Die Analyse von Einzelnukleotidaustauschen demonstrierte, dass für kein Isolat der statistische, auf den GC-Gehalt normalisierte Erwartungswert $\frac{s}{n} = \frac{2,9}{9,1}$ des Verhältnisses von synonymen zu nicht-synonymen Austauschen ($\frac{syn}{nicht-syn}$) erreicht wurde. Für den Großteil der Stämme lagen die Werte jedoch im oberen Konfidenzintervall des Erwartungswertes,

und die Stämme zeigten mehr Aminosäureaustausche als stille Mutationen. Zusätzlich konnte festgestellt werden, dass sich mit steigender SNP-Zahl das Verhältnis $\frac{syn}{nicht-syn}$ dem Wert vier annäherte, welcher auch beim Vergleich von unverwandten Stämmen gefunden wurde (Hilker et al., 2015). Somit nähern sich bei Isolaten, welche sich im phylogenetischen Baum aufgrund ihrer hohen Zahl an Einzelnukleotidaustauschen weit vom Ursprung entfernen, die Werte $\frac{syn}{nicht-syn}$ denen unverwandter Stämme an. Es ist somit nur eine Frage der Zeit, bis aus diesen Isolaten ein neuer Klon entsteht.

Mutationen in nicht-kodierenden RNAs (ncRNAs) wurden in höherer Anzahl in Vertretern des klonalen Komplexes C identifiziert. Diese Tatsache ist sofern nicht verwunderlich, da für den klonalen Komplex PA14 bislang lediglich intergenische RNAs annotiert sind (Wurtzel et al., 2012). Da im Rahmen dieser Arbeit festgestellt wurde, dass viele Mutationen in kodierenden Bereichen lokalisiert sind, kann dies die geringere Zahl der Mutationen in den ncRNAs von PA14 erklären. So ist es nicht möglich zu postulieren, dass der klonale Komplex C generell mehr Nukleotidaustausche in ncRNAs zeigt als PA14. Abbildung 3.20 zeigt, dass die Einzelnukleotidaustausche in ncRNAs der beiden klonalen Komplexe in höchstens 15 bis 20 Isolaten auftraten. Als Ausnahme zeigten Vertreter von Klon C einige wenige Mutationen, die in nahezu allen Isolaten präsent waren. Bei allen SNPs, welche in mehr als 40 Stämmen auftraten, konnte es sich dabei allerdings auch um Sequenzierfehler oder stammspezifische Nukleotidaustausche des Referenzgenoms NN2 handeln.

Betrachtet man die Gesamtheit der Aminosäureaustausche in Klon C, unterscheidet sich die Verteilung der betroffenen Proteine auf die funktionellen Kategorien komplett von den entsprechenden Ergebnissen für das *P. aeruginosa* Pangenom (Hilker et al., 2015). Hingegen zeigten die Ergebnisse für den klonalen Komplex PA14 lediglich in 16 von 26 Kategorien signifikante Unterschiede. Untersuchte man nur die drastischen Aminosäureaustausche (Dayhoffklasse < 5), zeigte der klonale Komplex C für die Hälfte der Kategorien einen signifikanten Unterschied im Vergleich zum Pangenom, der klonale Komplex PA14 jedoch nur in sechs Kategorien. So lässt sich feststellen, dass das Spektrum der betroffenen Aminosäureaustausche innerhalb der Klone C und PA14 unterschiedlich im Vergleich zum Kerngenom aller *P. aeruginosa* war. Diese Erkenntnis war erklärbar, wenn man die evolutionäre Zeitskala der Sequenzvariationen innerhalb einer Spezies und seiner klonalen Komplexe verglich. Ein klonaler Komplex in *P. aeruginosa* war durch einen konservierten chromosomalen Rahmen und eine geringe Diversität der Aminosäuresequenz charakterisiert. Die wenigen nicht-synonymen Mutationen waren in den meisten Fällen kladenspezifisch, und die betroffenen Proteine sind an der Kommunikation der Bakterien mit ihrer Umwelt beteiligt. Dieser Umstand deutet auf ein habitatbezogenes Auftreten eines Vorfahren im Zeitraum von Tagen bis hin zu Jahrzehnten hin. Andererseits stellen die konservierten Proteinvariationen zwischen unverwandten Stämmen verschiedener klonaler Linien ein Resultat von vielen Millionen Jahren negativer Selektion dar.

Ein typisches *P. aeruginosa* Isolat zeigte im Vergleich zum jeweiligen Referenzgenom NN2

bzw. PA14 im Median fünf (Klon C) bzw. vier (Klon PA14) kleine Insertionen oder Deletionen (Indels). Ein wiederkehrendes Ziel verschiedener Indels war das Gen *ladS*. Die in der Population weit verbreitete Anwesenheit der Leserasterverschiebungen dieses Gens in allen untersuchten PA14-Stämmen legt nahe, dass die ansteigende Zytotoxizität durch die Mutationen in *ladS* als ein phänotypisches Charakteristikum des klonalen Komplexes PA14 gesehen werden kann. Der Bruch des Leserasters von *ladS* scheint, wie die in der Literatur beschriebenen Mutationen der Gene *gacS* und *retS*, ein paradigmatisches Beispiel dafür zu sein, wie das Ausschalten eines einzelnen Gens den kompletten Phänotyp verändern kann (Bezuidt et al., 2013; Sall et al., 2014). Mutationen, die zum Funktionsverlust führen, haben demnach einen großen Einfluss auf die phänotypische Diversität eines klonalen Komplexes.

Zusätzlich zu den kurzen Indels wurden in einigen Stämmen auch große Deletionen innerhalb des Kerngenoms detektiert, die sich in den PA14 Stämmen um den Replikationsterminus anhäuften und Gene des Siderophors Pyoverdin oder die extrazelluläre Matrixkomponente Psl betrafen. Da Psl in der Biofilmbildung im Rahmen der Antibiotikaresistenz eine wichtige Rolle einnimmt (Colvin et al., 2012), sind Isolate mit fehlendem Protein vermutlich anfälliger für bakterizide Stoffe und können in CF-Patienten ggf. keine chronische Infektion induzieren.

Trotz der für die 100 Genome erzielten Sequenziertiefe von mehr als einer Million Reads pro Isolat war eine Aussage über große Insertionen nur möglich, wenn diese in Form bekannter genomischer Inseln oder RGP's auftraten und die Insertionsstellen bekannt waren. Daher war eine Aussage über die potentielle Lokalisation nur vergleichend mit den Ursprungsstämmen der genomischen Inseln und RGP's möglich, da angenommen wurde, dass mobile Elemente stets in dieselben Regionen des Kerngenoms inserieren. Um diese Annahme zu verifizieren und die Lokalisation der Insertionen nachzuweisen, müssten die Genome aller Stämme geschlossen oder mittels Makrorestriktionskartierungen und Pulsfeldgelelektrophorese untersucht werden. Damit ließen sich zusätzlich auch freie Plasmide und Genominversionen nachweisen. Der Versuch, Genominversionen anhand von chimären Reads zu identifizieren, von denen Fragmente auf auseinanderliegenden Genompositionen kartiert worden waren, und somit einen Inversionsbruchpunkt überspannten, wurde nicht weiter verfolgt. Für den Stamm B6470 waren aus Makrorestriktionskarten bereits Inversionen größerer Bereiche des Genoms bekannt (Römling et al., 1997). Im *in silico* Ansatz der Analyse chimärer Reads konnte jedoch keine dieser Inversionen nachgewiesen werden. Vielmehr wurde eine große Anzahl angeblicher zusätzlicher großer Inversionen angezeigt. Aus diesem Grund wurde die Analyse verworfen, da die Länge und Qualität der SOLiD-Sequenzierungsreads keine verlässliche Auflösung solcher Inversionsbruchpunkte gewährleistet und zusätzlich viele falsch positive Ergebnisse anzuzeigen scheint. Somit müssten die Inversionen über Restriktionsverdau und nachfolgender Pulsfeldgelelektrophorese oder PacBio-Sequenzierungen nachgewiesen werden.

Zusätzlich zu den bekannten genomischen Inseln und RGP's hatten die Isolate beider klonaler Komplexe im Schnitt 100 Gene zusätzlich aufgenommen, wobei die meisten dieser ORFs die größte Ähnlichkeit zu Genen aus Pseudomonaden hatten. Das ist insofern nicht verwunderlich, da zu erwarten ist, dass der Genaustausch eher zwischen Vertretern derselben Spezies als mit Vertretern fremder Spezies stattfindet. Da die meisten, der aus nicht gegen die Referenz kartierten Reads generierten Contigs nur den Bruchteil einer typischen Genlänge beinhalteten (Contigs aus 200 - 300 Nukleotiden), wurden alle ORFs, welche 90% Sequenzidentität über 80% der Genlänge aufwiesen, zusammengefasst. Das schloss jedoch nicht aus, dass einige Genvarianten mit geringerer Sequenzübereinstimmung als verschiedene Gene interpretiert und somit eine zu hohe Zahl zusätzlicher Gene angegeben wurde. Für viele der stammspezifischen Gene konnten keine Hinweise auf die potentiellen Funktionen der kodierten Proteine gefunden werden. Bemerkenswert war jedoch, dass Klon C Isolate signifikant mehr Gene des Kohlenstoffmetabolismus akquiriert hatten. Da *P. aeruginosa* in der Regel Glukose nur über den Entner-Douderoff-Weg verstoffwechselt und bevorzugt auf Succinat oder Citrat als Kohlenstoffquelle wächst, könnten diese zusätzlichen Gene die dominante Stellung von Klon C innerhalb der *P. aeruginosa* Population erklären. So kann es für Klon C Isolate einen Wachstumsvorteil darstellen, über einen vielseitigeren Kohlenstoffmetabolismus als andere klonale Komplexe zu verfügen. Die Entdeckung, dass stammspezifische Gene mit Ausnahme der Gene des Kohlenstoffmetabolismus in Klon C nur wenige weitere Metabolismus-Proteine kodierten, deutet darauf hin, dass das genetische Repertoire des *P. aeruginosa* Kerngenoms bereits ausreicht, um für ein breites Spektrum an Umweltbedingungen einen funktionierenden Metabolismus und Zellwachstum zu gewährleisten.

Basierend auf Studien von Referenzstämmen wie PAO1 wird es mittlerweile als Lehrbuchwissen vermittelt, dass *P. aeruginosa* bevorzugt Amino- und Fettsäuren als Kohlenstoffquelle des Intermediär- und des Energiestoffwechsels nutzt (Katabolische Repressionskontrolle, engl. „catabolic repression control“) (Linares et al., 2010). Die Genomanalysen dieser Arbeit deuten jedoch darauf hin, dass die Repression der Aufnahme und des Katabolismus von Zuckern nicht in allen *P. aeruginosa* ausgeprägt ist. Einige klonale Linien, wie die des häufigsten Klons C, scheinen die durch das Kerngenom gegebene Limitierung der Zuckerverwertung durch die Aufnahme von Genen des Kohlenhydratmetabolismus über horizontalen Gentransfer oder Mutationen von Schlüsselregulatoren zu überwinden. Diese Tatsache wurde bereits für *P. aeruginosa* Isolate der CF-Lunge berichtet (Silo-Suh et al., 2005). Hierzu steht eine Validierung durch Enzymaktivitätsmessung noch aus.

3.3 Kompetitive Wachstumsexperimente

Die bisherigen Kapitel behandelten die genomische Zusammensetzung und die intraklonale Diversität der einzelnen Klone. Dieses Kapitel beschreibt Untersuchungen dazu, inwieweit sich intraklonale Genomvariationen in der Fitness der einzelnen Stämme eines klonalen Komplexes niederschlagen. Dazu wurden Subgruppen von Klon C oder Klon PA14 Stämmen gemeinsam in planktonischen Kulturen im nährstoffreichen LB- sowie nährstoffarmen Mineralmedium kultiviert und die Zusammensetzung der Kulturen zu definierten Probenahmezeitpunkten (0 h, 48 h, 120 h) untersucht (Abb. 3.24).

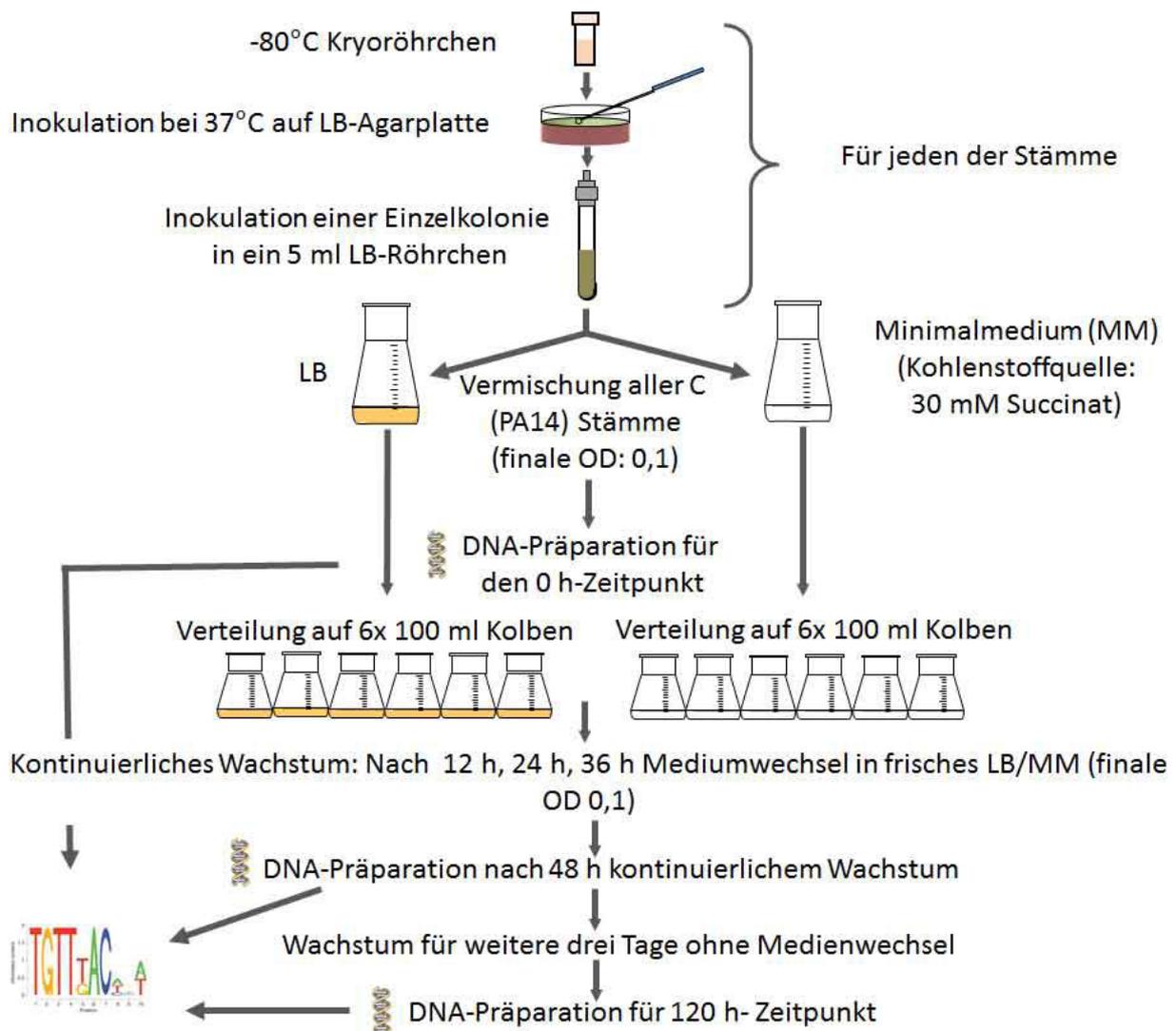


Abbildung 3.24: Flussdiagramm der kompetitiven Wachstumsexperimente in LB- und Minimalmedium.

Um sicherzustellen, dass die zu erwartenden Fitnessunterschiede auf individuelle Stammeigenschaften und nicht auf die Aktivität bakterizider Pyocine zurückzuführen war, wurde vor dem eigentlichen Wachstumsexperiment die Toleranz eines Stamms gegenüber den Pyocinen der anderen Stämme getestet. Pyocine sind bekannt dafür, selbst innerhalb

nah verwandter Stämme bakterizide Wirkungen zu entfalten (Michel-Briand und Baysse, 2002). Die Bakterienstämme sind jedoch in der Lage, sich durch neutralisierende Antitoxine vor Pyocinen zu schützen.

Zur Vermeidung Pyocin-vermittelter Wachstumsinhibition wurden die einzelnen Stämme zuerst *in silico* auf die Anwesenheit und nachfolgend auf Mutationen in den Pyocinloci untersucht. Das NN2 Referenzgenom weist einen Pyocinlocus (PANN43030) aus. Dieser war in allen Stämmen vorhanden, und es wurden keine Mutationen in diesem Locus gefunden. Das Referenzgenom PA14 beinhaltet drei Pyocinloci (PA14_13940, PA14_49520, PA14_59220). Dabei war ein Locus in allen Stämmen vorhanden, welcher jedoch einen Aminosäureaustausch in einem Stamm zeigte. Ein zweiter Locus fehlte in einem Stamm und zeigte einen Aminosäureaustausch in einem zweiten Stamm. Der dritte Locus war nur in 31 Stämmen vorhanden und enthielt ebenfalls eine zusätzliche Mutation (Tab. 3.18).

Tabelle 3.18: Übersicht über die Pyocinloci der klonalen Komplexe C und PA14.

| klonaler Komplex | Locus | nicht präsent in | Mutationen |
|------------------|------------|---|--------------|
| Klon C | PANN43030 | - | - |
| Klon PA14 | PA14_13940 | - | CF12 (E189V) |
| | PA14_49520 | PT2, K9 | 146 (M424V) |
| | PA14_59220 | 110_E8, 250, 55F6, CF12, CF13, CF4, K4, K9, PT2, RZ2, SS8 | 113 (C200P) |

Alle Stämme mit einer mutierten Pyocinvariante wurden im Vorhinein von den Fitnessexperimenten ausgeschlossen. Für nachfolgende Wachstumskontrollen wurden die Stämme dann in Gruppen zusammengefasst, kultiviert und die Überstände auf Platten mit Einzelstämmen pipettiert (siehe Kapitel 2.24.1). Alle Stämme, die kein Wachstum zeigten und daher sensitiv gegenüber den Pyocinen in den Überständen zu sein schienen, wurden ebenfalls nicht in den Fitnessexperimenten verwendet. Als gegenseitig Pyocin-tolerierende Stämme verblieben schließlich zehn Klon PA14 und 32 Klon C Stämme. Die reduzierte Anzahl zeigt auch, dass keine feste Korrelation zwischen der *in silico* Vorhersage der Pyocine und deren tatsächlicher Wirkung in Wachstumskulturen bestand.

Die verbleibenden Stämme wurden in gleichen Verhältnissen aus separaten Übernachtskulturen zu einem anfänglichem Inokulum mit einer OD von 0,1 vermischt und bei 37°C inkubiert. Innerhalb der ersten 48 h nach Inokulation wurde alle 12 Stunden ein Medienwechsel durchgeführt, bei dem die Kultur wieder auf eine OD von 0,1 verdünnt wurde. Anschließend wurde die Bakterienkultur für weitere 72 h ohne weiteren Medienwechsel inkubiert. Zu den Zeitpunkten 0 h, 48 h und 120 h nach Beginn des Experiments wurde ein Bakterienaliquot abgenommen und daraus die DNA isoliert (vgl. Kap. 2.24.2). Nach anschließender SOLiD-Sequenzierung der DNA konnte das Verhältnis der Bakterien untereinander anhand der in Kapitel 3.2.2 identifizierten SNPs berechnet werden. Die Analyse

der 0 h-Werte zeigte, dass in den Konkurrenzkulturen zu Beginn nicht alle Stämme eines Klons in gleicher Menge vertreten waren. Da zur Inokulation für alle Stämme gleich große Aliquots mit gleichen OD-Werten verwendet worden waren, lässt sich folgern, dass die optische Dichte kein verlässliches Maß für die initiale Zusammensetzung einer Mischung war. Streuungs- und Verklumpungseffekte schienen zu ungenauen OD-Werten geführt zu haben. In den Konkurrenzkulturen waren daher zum 0 h-Zeitpunkt drei Stämme (Klon C: 84, 179, 584) über- und sechs Stämme (Klon PA14: 202, Klon C: 10SB1, 702_D3, 902_D1, K11, SS30) unterrepräsentiert. Aufgrund dessen wäre eine Analyse absoluter Werte stark fehlerbehaftet. Stattdessen mussten die relativen Veränderungen gegenüber den jeweiligen Startwerten ermittelt werden. Die relative Fitness der einzelnen Stämme stellte sich daher nicht als Anteil eines Stammes in der Konkurrenzkultur nach 48 h bzw. 120 h dar, sondern als Veränderung gegenüber dem 0 h-Wert.

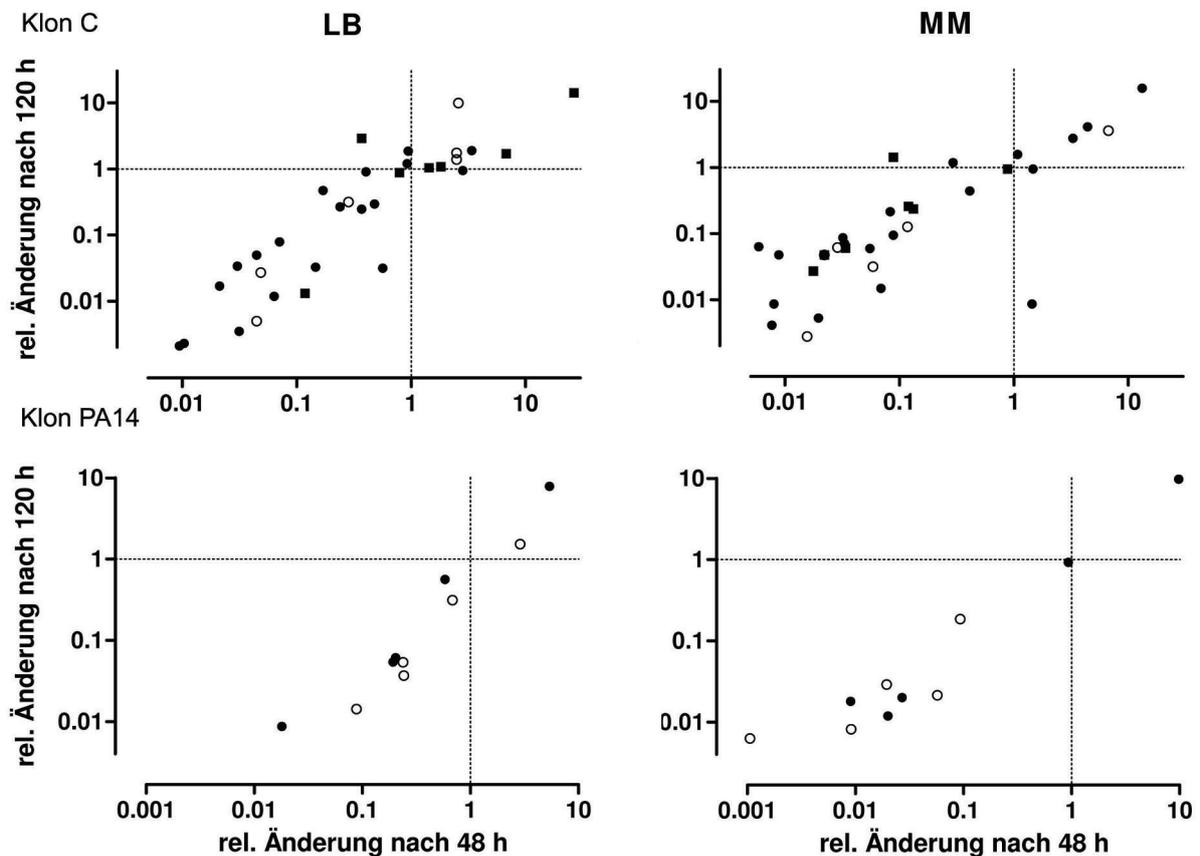


Abbildung 3.25: Fitnessexperimente in LB (linke Spalte) und Minimalmedium (MM, rechts) von 32 Klon C (obere Reihe) und 10 Klon PA14 Stämmen (unten). Die Abbildung zeigt die relative Änderung des Anteils der individuellen Stämme in der Konkurrenzkultur nach zwei und fünf Tagen Kokultivierung verglichen mit dem Startwert des Experiments (offener Kreis: Patientenisolat akuter Infektion; geschlossener Kreis: Patientenisolat chronischer Infektion; Quadrat: Umweltisolat).

Abbildung 3.25 zeigt die relative Veränderung der einzelnen Stämme innerhalb der Bakterienkultur nach zwei bzw. fünf Tagen Kokultivierung. Dabei konnten alle Werte innerhalb

eines Detektionsrahmens von einer Anteilsverminderung um das tausendfache bis hin zu einer mehr als zehnfachen Vergrößerung des relativen Anteils ermittelt werden. Diese Entdeckung wurde unabhängig vom Wachstumsmedium und der Änderung der Wachstumsphase vom exponentiellen (48 h, alle 12 h Medienwechsel) zum stationären Wachstum (72 h ohne Mediumwechsel) beobachtet. Die prozentualen Anteile der einzelnen Isolate sind in den Tabellen 3.19 und 3.20 aufgelistet. Daran ist zu erkennen, dass von den acht (PA14) bzw. elf (C) Stämmen, welche am stärksten über- oder unterrepräsentiert waren, pro Klon zwei Isolate als Verlierer (C: 324, K14; PA14: 97, B6470) sowie je ein Stamm als Gewinner (C: CF1; PA14: K16) in beiden Medien identifiziert werden konnte. Besonders hervorzuheben ist das CF-Isolat CF1 aus der Klon PA14-Kompetition, welches im Mineralmedium zu beiden untersuchten Zeitpunkten mehr als 95% der Bakterienkultur ausmachte.

Tabelle 3.19: Relative Anteile der PA14 Isolate zu den verschiedenen Zeitpunkten des Wachstumsexperiments.

| Isolat | Anteil [%] der einzelnen Isolate in | | | | | |
|------------|-------------------------------------|---------------------|---------------------|--------------|---------------------|---------------------|
| | Minimalmedium | | | LB-Medium | | |
| | 0 h | 48 h | 120 h | 0 h | 48 h | 120 h |
| 202 | 0,43 ± 0,32 | 0,40 ± 0,17 | 0,40 ± 0,14 | 0,74 ± 0,15 | 0,43 ± 0,11 | 0,41 ± 0,14 |
| 250 | 14,97 ± 1,78 | 0,40 ± 0,16 | 0,30 ± 0,15 | 13,35 ± 0,68 | 2,57 ± 0,27 | 0,72 ± 0,17 |
| 324 | 11,17 ± 1,46 | <u>0,10</u> ± 0,04 | <u>0,20</u> ± 0,08 | 11,88 ± 0,56 | <u>0,21</u> ± 0,07 | <u>0,10</u> ± 0,05 |
| CF1 | 9,87 ± 3,1 | 96,80 ± 6,21 | 96,25 ± 5,72 | 10,09 ± 1,29 | 54,34 ± 3,15 | 79,44 ± 4,60 |
| CF4 | 7,59 ± 0,46 | <u>0,15</u> ± 0,03 | <u>0,09</u> ± 0,03 | 8,41 ± 0,20 | 1,71 ± 0,09 | 0,52 ± 0,06 |
| K14 | 9,54 ± 4,23 | <u>0,01</u> ± 0,01 | <u>0,06</u> ± 0,06 | 14,51 ± 1,91 | <u>1,29</u> ± 0,56 | <u>0,21</u> ± 0,21 |
| K3 | 14,10 ± 6,87 | 0,81 ± 0,81 | 0,30 ± 0,30 | 9,57 ± 1,69 | 6,54 ± 1,94 | 3,00 ± 1,50 |
| K4 | 11,06 ± 0,63 | <u>0,10</u> ± 0,03 | <u>0,09</u> ± 0,03 | 8,41 ± 0,20 | 2,04 ± 0,10 | 0,31 ± 0,05 |
| K7 | 10,41 ± 0,64 | 0,20 ± 0,05 | 0,30 ± 0,05 | 13,46 ± 0,27 | 3,22 ± 0,12 | 0,72 ± 0,07 |
| K8 | 10,85 ± 3,24 | 1,01 ± 0,73 | 2,01 ± 1,03 | 9,57 ± 1,31 | 27,65 ± 2,83 | 14,57 ± 2,47 |

Die Werte der als Gewinner identifizierten Stämme sind **fett**, die der als Verlierer identifizierten Stämme unterstrichen markiert.

Die Gewinner- und Verliererstämme wurden im folgenden auf genomische Eigenschaften untersucht, welche die Vor- und Nachteile in den Wachstumseigenschaften erklären könnten. Für mehrere Gewinner der Klon C Stämme waren seltene Aminosäureaustausche oder Leserasterverschiebungen in Schlüsselregulatoren der Lebensweise (*ladS*: 84, K16; *gacA*: 84; *cbrAB*: 110_D4, 747, 84) gefunden worden. Zusätzlich dazu trugen einzelne Isolate mobile Elemente wie einen lytischen Phagen (110_D4), ein Plasmid (747) oder zusätzliche Kopien von Genen der sogenannten stringenten Antwort (*uvrD*, *relA*; 84, vgl. Tab. S11). Als stringente Antwort wird dabei die Anpassung des Stoffwechsels auf eine Unterversorgung mit Nährstoffen bezeichnet.

Die Verliererstämme zeigten Aminosäureaustausche in Enzymen der DNA-Topologie, LPS-

Biosynthese (Wbpm, ArnA) oder Proteinen, welche die Reaktionen auf Umwelteinflüsse steuern (WspR). Im Gegensatz dazu hatten zwei (CF4, K4) der vier Verlierer der PA14-Stämme eine Variante der Geninsel PAPI-1 gemeinsam. Zusätzlich hatten beide Isolate eine hohe Zahl weiterer Aminosäureaustausche akkumuliert. Die zwei Gewinnerstämme (CF1, K8) wiesen entweder die höchste Ähnlichkeit mit dem PA14 Referenzstamm auf (CF1) oder kodierten konträr dazu eine große Menge zusätzlicher Gene wie Transkriptionsregulatoren, NADPH-abhängige Oxidoreduktasen und Schwermetallionen-Efflux-Proteine (K8).

Zusammenfassend konnte gezeigt werden, dass im Einklang mit der sternförmigen Populationsstruktur (Abb. 3.13) die Gewinner und Verlierer keine gruppenspezifische Signatur trugen, sondern nur individuelle genomische Merkmale. Ebenso konnte auch keine Assoziation zwischen Habitat und geografischem Ursprung festgestellt werden.

Tabelle 3.20: Relative Anteile der Klon C Isolate zu den verschiedenen Zeitpunkten des Wachstumsexperiments.

| Isolat | Anteil [%] der einzelnen Isolate in | | | | | |
|------------|-------------------------------------|---------------------|---------------------|--------------|---------------------|---------------------|
| | Minimalmedium | | | LB-Medium | | |
| | 0 h | 48 h | 120 h | 0 h | 48 h | 120 h |
| 17 | 2,38 ± 0,58 | 0,20 ± 0,13 | 0,21 ± 0,17 | 4,21 ± 0,61 | <u>0,04</u> ± 0,04 | <u>0,01</u> ± 0,01 |
| 36 | 1,78 ± 0,69 | 0,01 ± 0,01 | 0,11 ± 0,11 | 3,08 ± 0,58 | 1,72 ± 0,41 | 0,10 ± 0,10 |
| 84 | 10,70 ± 1,81 | 44,84 ± 1,99 | 41,52 ± 3,06 | 5,64 ± 0,91 | 5,17 ± 0,64 | 6,73 ± 0,90 |
| 87 | 3,57 ± 0,44 | 4,98 ± 0,23 | 3,19 ± 0,37 | 3,08 ± 0,31 | 10,34 ± 0,45 | 5,77 ± 0,35 |
| 97 | 1,31 ± 1,29 | <u>0,01</u> ± 0,01 | <u>0,01</u> ± 0,01 | 4,62 ± 1,58 | <u>0,04</u> ± 0,04 | <u>0,01</u> ± 0,01 |
| 110 | 2,14 ± 1,04 | <u>0,04</u> ± 0,04 | <u>0,01</u> ± 0,01 | 2,87 ± 0,86 | 0,09 ± 0,09 | 0,10 ± 0,10 |
| 179 | 10,70 ± 2,75 | 10,96 ± 1,56 | 15,97 ± 2,99 | 2,77 ± 1,03 | <u>0,09</u> ± 0,09 | <u>0,01</u> ± 0,01 |
| 214 | 1,78 ± 0,69 | 0,70 ± 0,70 | 0,75 ± 0,75 | 3,08 ± 3,08 | 8,62 ± 5,36 | 2,88 ± 2,88 |
| 426 | 6,30 ± 1,20 | 0,50 ± 0,18 | 1,28 ± 0,44 | 3,28 ± 0,66 | 1,55 ± 0,39 | 0,96 ± 0,29 |
| 584 | 7,13 ± 1,59 | 1,99 ± 0,37 | 7,99 ± 1,22 | 2,36 ± 0,62 | 0,86 ± 0,24 | 0,58 ± 0,20 |
| 747 | 0,95 ± 0,95 | 11,96 ± 1,83 | 14,16 ± 3,28 | 2,77 ± 1,28 | 2,59 ± 1,05 | 5,09 ± 1,46 |
| 10SB1 | 0,12 ± 0,12 | 0,01 ± 0,01 | 0,16 ± 0,16 | 6,67 ± 1,16 | 9,48 ± 0,89 | 6,92 ± 0,89 |
| 110_D4 | 5,59 ± 1,41 | 0,18 ± 0,13 | 0,32 ± 0,25 | 0,82 ± 0,42 | 21,55 ± 1,87 | 11,53 ± 1,36 |
| 12_F8 | 2,38 ± 0,56 | 0,05 ± 0,05 | 0,11 ± 0,11 | 4,41 ± 0,22 | 0,52 ± 0,18 | 0,06 ± 0,06 |
| 12_SB1 | 4,16 ± 0,93 | 0,07 ± 0,06 | 0,11 ± 0,11 | 4,41 ± 0,71 | 3,45 ± 0,54 | 3,84 ± 0,53 |
| 702_D3 | 0,12 ± 0,12 | 0,10 ± 0,10 | 0,11 ± 0,11 | 0,51 ± 0,33 | 3,45 ± 0,77 | 0,86 ± 0,38 |
| 902_D1 | 0,48 ± 0,42 | 0,06 ± 0,06 | 0,11 ± 0,11 | 0,72 ± 0,41 | 1,29 ± 0,44 | 0,77 ± 0,31 |
| B6470 | 4,04 ± 1,96 | <u>0,06</u> ± 0,06 | <u>0,01</u> ± 0,01 | 1,95 ± 1,00 | <u>0,09</u> ± 0,09 | <u>0,01</u> ± 0,01 |
| CF2 | 1,31 ± 1,31 | 1,79 ± 1,09 | 0,01 ± 0,01 | 2,15 ± 1,74 | 0,86 ± 0,86 | 1,92 ± 1,67 |
| CF3 | 1,90 ± 1,01 | 0,10 ± 0,10 | 0,11 ± 0,11 | 2,87 ± 0,84 | <u>0,06</u> ± 0,06 | <u>0,05</u> ± 0,05 |
| CF8 | 1,31 ± 0,74 | 4,09 ± 0,71 | 3,41 ± 1,08 | 3,28 ± 0,92 | 0,78 ± 0,37 | 0,86 ± 0,41 |
| K11 | 0,36 ± 0,36 | 0,02 ± 0,02 | 0,01 ± 0,01 | 0,31 ± 0,30 | 0,09 ± 0,09 | 0,10 ± 0,10 |
| K13 | 5,47 ± 0,59 | 0,15 ± 0,05 | 0,32 ± 0,12 | 3,49 ± 0,36 | 8,62 ± 0,39 | 4,80 ± 0,35 |
| K16 | 2,50 ± 0,88 | 15,94 ± 1,16 | 8,52 ± 1,29 | 2,36 ± 0,65 | 6,03 ± 0,78 | 23,26 ± 1,75 |
| K17 | 5,35 ± 0,30 | 0,60 ± 0,05 | 0,64 ± 0,09 | 3,49 ± 0,18 | 8,62 ± 0,24 | 6,05 ± 0,21 |
| K18 | 2,38 ± 0,96 | 0,05 ± 0,05 | 0,11 ± 0,11 | 3,59 ± 0,92 | 0,17 ± 0,17 | 0,10 ± 0,10 |
| PJC | 3,80 ± 3,76 | 0,25 ± 0,25 | 0,05 ± 0,05 | 10,26 ± 4,89 | 1,72 ± 1,42 | 4,80 ± 2,72 |
| SG1 | 2,73 ± 2,05 | <u>0,02</u> ± 0,02 | <u>0,01</u> ± 0,01 | 1,23 ± 1,07 | 0,09 ± 0,09 | 0,10 ± 0,10 |
| SG50m | 1,31 ± 0,80 | 0,15 ± 0,15 | 0,32 ± 0,32 | 4,00 ± 1,04 | 1,47 ± 0,53 | 11,53 ± 1,29 |
| SS30 | 0,95 ± 0,33 | 0,03 ± 0,03 | 0,06 ± 0,06 | 0,82 ± 0,23 | 0,05 ± 0,05 | 0,01 ± 0,01 |
| WI1 | 2,62 ± 1,25 | 0,08 ± 0,08 | 0,21 ± 0,21 | 1,95 ± 0,81 | 0,09 ± 0,09 | 0,10 ± 0,10 |
| WU2 | 2,38 ± 0,84 | 0,02 ± 0,02 | 0,11 ± 0,11 | 2,97 ± 0,70 | 0,43 ± 0,22 | 0,10 ± 0,10 |

Die Werte der als Gewinner identifizierten Stämme sind **fett**, die der als Verlierer identifizierten Stämme unterstrichen markiert.

3.3.1 Diskussion der kompetitiven Wachstumsexperimente

Im abschließenden Teil der Arbeit wurden kompetitive Wachstumsexperimente durchgeführt, um mögliche Korrelationen zwischen den identifizierten stammspezifischen genomischen Eigenschaften und der Durchsetzungsfähigkeit innerhalb der Population zu untersuchen.

Nachdem die Isolate auf bakterizide Pyocine hin untersucht und inkompatible Isolate von der Analyse ausgeschlossen worden waren, verblieben für die Experimente in PA14 zehn und für Klon C 32 Isolate. Der Umstand, dass von 42 Klon PA14 Isolaten nur 10 und von 57 Klon C Isolaten nur 32 den Test auf Pyocinkompatibilität bestanden, zeigte, dass es in planktonischen Kulturen noch weitere Effektoren des Zellwachstums geben muss. Eine andere mögliche Erklärung ist, dass nicht alle Stämme die Pyocine und zugehörigen Antitoxine konstitutiv gleich exprimierten.

Für den Test auf Pyocinkompatibilität wurden Isolate gepoolt, kultiviert und die Pyocin-enthaltenden Überstände auf Platten getropft, auf denen die einzelnen Isolate ausgestrichen worden waren (vgl. Kap. 2.24.1). Dieses Vorgehen wurde gewählt, da die paarweise Überprüfung aller Stämme gegeneinander nicht in einem budget- und ressourcenschonenden Maßstab hätten durchgeführt werden können, auch wenn die paarweisen Tests eventuell eine höhere Anzahl an Isolaten für die kompetitiven Wachstumstests zugelassen hätten.

Zu spezifizierten Zeiten des Wachstums (Startwert, 48 h, 120 h) wurde aus einem Bakterienaliquot die DNA isoliert und mit Hilfe einer SOLiD Maschine sequenziert. Anhand der Readabdeckung von, für die einzelnen Isolate der Konkurrenz spezifischen SNPs, wurde der Anteil jedes Stamms innerhalb des Experiments berechnet und für PA14 ein klarer Gewinner (CF1) identifiziert, welcher alle anderen Stämme zurückdrängte und zum Ende des Experimentes mehr als 70% in LB-Medium und sogar mehr als 90% der Population im Mineralmedium ausmachte. Da innerhalb der siebentägigen Ko-Kultivierung lediglich 60 Bakteriengenerationen entstehen, ist dieses Resultat umso eindrucksvoller. Auch können die prozentualen Unterschiede nicht mit langsamerem Wachstum einzelner Stämme erklärt werden, da im separaten Wachstum der Vorkulturen keine auxotrophe Mutante erkennbar war. Im Gegensatz zu CF1 kodierte das weitere Gewinnerisolat im LB-Medium der PA14-Konkurrenz (K8) zusätzliche Gene der Schwermetallresistenzen.

Zwei der Verliererstämme aus der PA14-Konkurrenz (CF4, K4), welche beide in England isoliert worden waren, wiesen neben vielen Mutationen und Aminosäureaustauschen auch eine andere Variante der genomischen Insel PAPI-1 auf. Aufgrund der Vielzahl der Mutationen war es für diese Isolate nicht möglich, den Fitnessnachteil auf einzelne Genomvariationen zurückzuführen. Vermutlich führte die Summe der Mutationen zu den beobachteten Wachstumsnachteilen.

Unter den untersuchten Klon C Stämmen waren die Gewinner im nährstoffreichen Medi-

um mit der Ausnahme des Keratitisisolats K9 nicht auch die Gewinner im nährstoffarmen Mineralmedium. Überraschenderweise setzten sich die Umweltisolate im nährstoffreichen LB-Medium durch und Isolate aus chronisch infizierten Patienten im nährstoffarmen Medium. Daraus ließ sich schlußfolgern, dass die Adaptation an das jeweilige Habitat noch nicht endgültig war, da eigentlich erwartet worden war, dass sich die Umweltisolate aufgrund der noch nicht erfolgten Adaptation an das Patientenhabitat im nährstoffarmen Medium besser durchsetzen können. Dabei sollte berücksichtigt werden, dass die CF-Isolate aus der frühen Phase der Kolonisierung der Patienten stammten und somit eventuell noch nicht komplett an das neue Habitat angepasst waren.

Bemerkenswert war auch die breite Streuung des Spektrums der Anwesenheit einzelner Stämme innerhalb der Kompetitionskultur um fünf Größenordnungen, von einer 1000-fachen Unterrepräsentierung bis zu einer zehnfachen Überrepräsentierung einzelner Stämme. Eine solch hohe Auflösung einer Analyse lässt sich nur mit einem Sequenzieransatz erzielen. Eine weitere Möglichkeit zur Bestimmung der Zellzahl wäre, die Bakterienstämme mit Fluoreszenzfarbstoffen (z.B. GFP) zu markieren. In diesem Ansatz ergibt sich aber sowohl eine Limitierung der möglichen Stammzahl aufgrund der Zahl verschiedener, einsetzbarer Fluoreszenzmoleküle als auch der Auflösung der Anteile der einzelne Stämme an der Kompetitionskultur.

Die Suche nach Ursachen für diese Stammdominanz ergab, dass CF1 die wenigsten Veränderungen gegenüber dem Referenzstamm PA14 aufwies und darunter keine stark funktionsverändernden Mutationen detektiert werden konnten. Der Ansatz, das Verhältnis der Stämme anhand von spezifischen Reads zu ermitteln, war besonders für die Berechnung der Anteile der unterrepräsentierten Isolate fehleranfällig, weil einige Isolate nur ein bis zwei spezifische Reads zeigten. Diese Tatsache bedeutete jedoch, dass ein seltener Sequenzierfehler (Genauigkeit der SOLiD Sequenzierung für Einzelreads von 99,94%) in diesem Fall einen starken Einfluss auf das Ergebnis des Stamms hat. Sollte eine Unkorrektheit trotzdem vorkommen, so kann es im Falle eines Reads, der einen Sequenzierfehler beinhaltete, anstatt eine echte Mutation anzuzeigen, bedeuten, dass der Anteil eines Isolats deutlich überschätzt wird. Andererseits ist auch möglich, dass ein Anteil angezeigt wird, obwohl der Stamm in der Kompetitionskultur nicht mehr präsent ist. Es ist aus der Literatur bekannt, dass DNA über einen längeren Zeitpunkt sehr stabil ist und noch nach vielen Jahren und teilweise Jahrhunderten nachgewiesen werden konnte (Nerlich et al., 1997; Papagrigorakis et al., 2006). Deshalb ließ es sich nicht zweifelsfrei feststellen, ob die verdrängten Isolate in geringen Individuenzahlen im Experiment vorhanden waren, oder die DNA von bereits abgestorbenen Zellen sequenziert wurde, die mit in die Bakterienpellets für die DNA-Präparation gelangt ist. Gegen einen systematischen Fehler sprechen dabei jedoch der Verdünnungsfaktor des Medienwechsels (Re-Inokulation mit einer OD von 0,1) und ein vergrößerter Anteil der Isolate innerhalb der Kompetitionskultur nach 120 h verglichen mit dem Wert nach 48 h. Da nahezu alle Isolate in der logarithmischen

Darstellung auf einer Gerade liegen (Abb. 3.25) und nur wenige Isolate davon abweichen, lässt sich lediglich für ein Isolat (CF2) der Klon C Kompetition dieser Fehler nicht vollständig ausschließen.

Mögliche Wachstumsvorteile der Gewinnerisolate beider klonaler Komplexe wurden *in vitro* untersucht, ließen sich jedoch nicht validieren. So wurden die Isolate auf Zink-, Cadmium- und Nickel-haltigen LB-Agarplatten kultiviert. Dabei zeigte sich, dass das Isolat K8, welches zusätzliche Gene der Schwermetallresistenzen im akzessorischen Genom kodierte, auf den Schwermetall-haltigen Agarplatten keinen Wachstumsvorteil hat. Wie in allen Isolaten war vermutlich nicht ein einziger Aminosäureaustausch, nur eine Deletion oder lediglich ein zusätzliches, stammspezifisches Gen für den Fitnessvor- bzw. -nachteil verantwortlich, sondern das Zusammenspiel mehrerer Ereignisse. Weiterhin besteht die Möglichkeit, dass nicht-synonyme Mutationen und Indels in Genen mit bislang unbekannter Funktion die Fitness der einzelnen Stämme nachhaltig beeinflussen können. Neue Funktionszuweisungen für Proteine, die zukünftig in der Fachliteratur und den Datenbanken erscheinen, könnten eine tiefere Interpretation der hier beschriebenen Kompetitionsergebnisse ermöglichen.

Kapitel 4

Schlussfolgerungen und Ausblick

Im Rahmen dieser Arbeit war es möglich, das Genom eines Vertreters des häufigsten Klons der *P. aeruginosa* Population, Klon C, ohne verbleibende Lücken zu entschlüsseln und erstmals innerhalb der Spezies intragenische nicht-kodierende RNAs zu annotieren. Zusätzlich wurden, basierend auf SOLiD-Sequenzierungsdaten, *in silico* erstmals Isolate der beiden dominanten Klone der *P. aeruginosa* Population, Klon C und Klon PA14, hinsichtlich ihrer genomischen Diversität untersucht. Es wurde gezeigt, dass das Kerngenom innerhalb eines klonalen Komplexes hochkonserviert ist und die wenigen Mutationen singuläre Ereignisse darstellen. Daher zeigte der phylogenetische Baum eine sternförmige Topologie um einen gemeinsamen Vorfahren. Auch waren die Fragmentlängenpaare, welche durch paarweise Vergleiche jeweils zweier Stämme der klonalen Komplexe C und PA14 gewonnen wurden, im Median 1000-fach länger, als wenn unverwandte Klone verglichen wurden. Im Gegensatz dazu zeigte das akzessorische Genom eine hohe Plastizität und war bei paarweisen Vergleichen zweier Vertreter des gleichen Klons nicht von unverwandten Klonen zu unterscheiden. In kompetitiven Wachstumsexperimenten konnte Stämme identifiziert werden, welche sich im Wachstum gegenüber anderen Stämmen durchsetzen konnten. Jedoch war es nicht möglich, die Wachstumsvorteile auf einzelne genomische Faktoren zurückzuführen.

Die in dieser Querschnittsstudie untersuchten klonalen Komplexe C und PA14 stellen die dominanten Klone in der *P. aeruginosa* Population dar und sind verteilt über den Globus sowohl in Umwelt- als auch in klinischen Isolaten zu finden. Im Vergleich dazu sind die nächsthäufigeren Vertreter der *P. aeruginosa* Population geografisch ungleichmäßig verteilt und tendenziell entweder mehr mit der Umwelt oder dem klinischen Habitat assoziiert. Die Tatsache, dass die klonalen Komplexe C und PA14 sowohl geografische als auch habitatbezogene Grenzen scheinbar spielend überwinden können, deutet darauf hin, dass beide Komplexe innerhalb ihres genetischen Repertoires Besonderheiten gegenüber den anderen Stämmen aufweisen. Mit anderen Worten ausgedrückt bedeutet dies, dass die seltenen Klone stärker mit einem bestimmten Habitat assoziiert sind.

Eine vergleichbare intraklonale Analyse anderer Bakterien fand bislang nicht statt. So

wurden bislang zwar beispielsweise von *E. coli* mehrere hundert Stämme sequenziert, jedoch nie eine große Kohorte des gleichen Klons untersucht (Huja et al., 2015; Toval et al., 2014). Eine solche Analyse könnte zeigen, ob die in dieser Arbeit beobachteten Befunde spezifisch für *P. aeruginosa* oder generell in bakteriellen Spezies verbreitet sind. So ließe sich ein Zusammenhang zwischen der Lebensweise, der Fitness und der Segregation nach Geografie und Habitat untersuchen.

Auch eine vertiefte sequentielle Analyse von Umweltisolaten, wie sie bislang schon für serielle Isolate der Mukoviszidose-Lunge durchgeführt wurde, würde erlauben die Evolution von *P. aeruginosa* in der Umwelt zu verfolgen. Ziel dieser Studien wäre es, die Frage zu beantworten, ob ähnlich wie in der CF-Lunge eine positive Selektion in Genen erfolgt, die die Fitness in wässrigen Habitaten verbessert.

Ein Novum stellen die in dieser Arbeit detektierten intragenischen ncRNAs dar. Von den drei bislang untersuchten Isolaten PAO1, PA14 und NN2 exprimiert jeder ein stammspezifisches Repertoire an ncRNAs, obwohl nahezu alle Sequenzen in den Genomen kodiert sind. In weiterführenden Studien wird zu prüfen sein, ob wirklich jeder Stamm ein individuelles Repertoire an ncRNAs exprimiert. Da für die Gewinner der Fitnessexperimente keine genomischen Eigenschaften den Phänotyp abschließend erklären konnten, stellt die stammspezifische Expression von regulierenden ncRNAs einen interessanten Ansatz dar, die Fitnessunterschiede zu erklären.

Die vorliegenden Analysen konnten nur durchgeführt werden, da sich Sequenzierungstechniken in den letzten Jahren stark verbessert haben. So ist es heutzutage kein Problem mehr, in großem Maßstab bakterielle Genome zu sequenzieren. Auch die Genauigkeit der Sequenzierungstechniken wurde bedeutend weiterentwickelt. Bedauerlicherweise wird die SOLiD-Technologie auf absehbare Zeit kommerziell nicht weiterentwickelt. Lange Sequenzierzeiten, die Produktion von vergleichsweise kurzen Reads und das Versagen der Technik in GC-reichen Regionen sind nachvollziehbare Gründe, warum die SOLiD-Technologie gegenüber den Illumina-Geräten sich nicht hat durchsetzen können.

Diese Studie zur intraklonalen Genomdiversität der *P. aeruginosa* Klone C und PA14 erbrachte als zentrales Ergebnis, dass ein hochkonserviertes Kerngenom mit einem hochvariablen akzessorischen Genom assoziiert ist. Zumindest für die beiden Klone C und PA14 gibt es kein klonspezifisches Repertoire an genomischen Inseln und RGPs. Vertreter der Klone können anhand der aktuellen Erfordernisse RGPs aus dem *Pseudomonas* Genus aufnehmen und wieder abgeben. Im Falle der Geninseln des PAGI-2-Typs ist ein horizontaler Gentransfer sogar über Gengrenzen hinaus möglich (Klockgether et al., 2007).

Diese Plastizität des akzessorischen Genoms steht im Gegensatz zu dem streng konservierten Kerngenom. So waren die Fragmentlängenpaare gleicher Sequenz des Kerngenoms bei Vergleich zweier Klone C oder Klon PA14 Stämme im Median 100 kb groß. Sortierte man die Stämme anhand der Einzelnukleotide zu einem phylogenetischen Baum, zeigte sich eine sternförmige Struktur um einen gemeinsamen Vorfahren. Auch teilten sich nur

in seltenen Fällen mehrere Isolate einen Ast und somit das gemeinsame Kerngenom. Diese Tatsachen sprechen dafür, dass sich das Kerngenom der klonalen Komplexe im Laufe von vielen Millionen Generationen evolutionär optimiert hat und kaum noch Veränderungen stattfinden.

Die wenigen *de novo* auftretenden Aminosäureaustausche unterliegen der purifizierenden Selektion, so dass im interklonalen Vergleich die synonymen Austausche zwölfmal häufiger Vorkommen als statistisch zu erwarten wäre. Diese kosmopolitischen Bakterien sind somit nicht auf ein spezifisches Habitat festgelegt, wie es typischerweise bei obligat pathogenen Bakterien beobachtet wird.

Literaturverzeichnis

- T. Abe, S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, und T. Ikemura. Informatics for unveiling hidden genome signatures. Genome Res., 13(4):693–702, Apr 2003.
- M. Allewelt, F. T. Coleman, M. Grout, G. P. Priebe, und G. B. Pier. Acquisition of expression of the *Pseudomonas aeruginosa* ExoU cytotoxin leads to increased bacterial virulence in a murine model of acute pneumonia and systemic spread. Infect. Immun., 68(7):3998–4004, Jul 2000.
- C. Alvarez-Ortega und C. S. Harwood. Responses of *Pseudomonas aeruginosa* to low oxygen indicate that growth in the cystic fibrosis lung is by aerobic respiration. Mol. Microbiol., 65(1):153–165, Jul 2007.
- K. F. Au, J. G. Underwood, L. Lee, und W. H. Wong. Improving PacBio long read accuracy by short read alignment. PLoS ONE, 7(10):e46679, 2012.
- F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith, und K. Struhl. Current protocols in molecular biology. JohnWiley and Sons, Inc, 1994.
- R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, und O. Zagnitko. The RAST Server: rapid annotations using subsystems technology. BMC Genomics, 9:75, 2008.
- A. Badrinarayanan, C. Lesterlin, R. Reyes-Lamothe, und D. Sherratt. The *Escherichia coli* SMC complex, MukBEF, shapes nucleoid organization independently of DNA replication. J. Bacteriol., 194(17):4669–4676, Sep 2012.
- P. Baldi und P. F. Baisnée. Sequence analysis by additive scales: DNA structure for sequences and repeats of all lengths. Bioinformatics, 16(10):865–889, Oct 2000.
- A. Berger, K. Dohnt, P. Tielen, D. Jahn, J. Becker, und C. Wittmann. Robustness and plasticity of metabolic pathway flux among uropathogenic isolates of *Pseudomonas aeruginosa*. PLoS ONE, 9(4):e88368, 2014.

- O. K. Bezuidt, J. Klockgether, S. Elsen, I. Attree, C. F. Davenport, and B. Tümmler. Intraclonal genome diversity of *Pseudomonas aeruginosa* clones CHA and TB. BMC Genomics, 14:416, 2013.
- M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics, 27(4):578–579, Feb 2011.
- M. J. Bottomley, E. Muraglia, R. Bazzo, and A. Carfi. Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. J. Biol. Chem., 282(18):13592–13600, May 2007.
- K. Botzenhardt und G. Döring. Etiology and epidemiology of *Pseudomonas aeruginosa* in *Pseudomonas aeruginosa* as an opportunistic pathogen. Plenum Press New York., 1993.
- D. Büttner. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. Microbiol. Mol. Biol. Rev., 76(2):262–310, Jun 2012.
- Y. J. Chang, M. Land, L. Hauser, O. Chertkov, T. G. Del Rio, M. Nolan, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, C. Han, L. Goodwin, S. Pitluck, N. Ivanova, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, K. Mavromatis, K. Liolios, T. Brettin, A. Fiebig, M. Rohde, B. Abt, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk, and A. Lapidus. Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21). Stand Genomic Sci, 5(1): 97–111, Oct 2011.
- E. Chargaff. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia, 6(6):201–209, Jun 1950.
- K. Cheng, R. L. Smyth, J. R. Govan, C. Doherty, C. Winstanley, N. Denning, D. P. Heaf, H. van Saene, und C. A. Hart. Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. Lancet, 348(9028):639–642, Sep 1996.
- P. Cingolani, A. Platts, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, und D.M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6(2):80–92, 2012.
- F. M. Cohan. Concepts of Bacterial Biodiversity for the Age of Genomics. Humana Press Inc, 2004.

- K. M. Colvin, Y. Irie, C. S. Tart, R. Urbano, J. C. Whitney, C. Ryder, P. L. Howell, D. J. Wozniak, und M. R. Parsek. The Pel and Psl polysaccharides provide *Pseudomonas aeruginosa* structural redundancy within the biofilm matrix. Environ. Microbiol., 14(8):1913–1928, Aug 2012.
- K. M. Colvin, N. Alnabelseya, P. Baker, J. C. Whitney, P. L. Howell, und M. R. Parsek. PelA deacetylase activity is required for Pel polysaccharide synthesis in *Pseudomonas aeruginosa*. J. Bacteriol., 195(10):2329–2339, May 2013.
- J. C. Comolli und T. J. Donohue. Differences in two *Pseudomonas aeruginosa* cbb3 cytochrome oxidases. Mol. Microbiol., 51(4):1193–1203, Feb 2004.
- T. R. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, und G. Waksman. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol., 13(6):343–359, May 2015.
- J.W. Costerton. *Pseudomonas aeruginosa* in nature and disease in *Pseudomonas aeruginosa: the organism, diseases it causes, and their treatment*. Hans Huber Publishers, Bern., 1980.
- N. Cramer, J. Klockgether, K. Wrasman, M. Schmidt, C. F. Davenport, und B. Tümmler. Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. Environ. Microbiol., 13(7):1690–1704, Jul 2011.
- N. Cramer, L. Wiehlmann, O. Ciofu, S. Tamm, N. Høiby, und B. Tümmler. Molecular epidemiology of chronic *Pseudomonas aeruginosa* airway infections in cystic fibrosis. PLoS ONE, 7(11):e50731, 2012.
- N. Darzentas. Circoletto: visualizing sequence similarity with Circos. Bioinformatics, 26(20):2620–2621, Oct 2010.
- M.O. Dayhoff. Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington D.C., 1978.
- T. R. De Kievit, M. D. Parkins, R. J. Gillis, R. Srikumar, H. Ceri, K. Poole, B. H. Iglewski, und D. G. Storey. Multidrug efflux pumps: expression patterns and contribution to antibiotic resistance in *Pseudomonas aeruginosa* biofilms. Antimicrob. Agents Chemother., 45(6):1761–1770, Jun 2001.
- D. De Vos, A. Lim, J. P. Pirnay, L. Duinslaeger, H. Revets, A. Vanderkelen, R. Hamers, und P. Cornelis. Analysis of epidemic *Pseudomonas aeruginosa* isolates by isoelectric focusing of pyoverdine and RAPD-PCR: modern tools for an integrated anti-nosocomial infection strategy in burn wound centres. Burns, 23(5):379–386, Aug 1997.

- G. Döring, R. Knight, , and G Bellon. Immunology of cystic fibrosis. In Cystic fibrosis. Arnold Publishers. London, United Kingdom., 2000.
- A. Eichner, N. Günther, M. Arnold, M. Schobert, J. Heesemann, und M. Hogardt. Marker genes for the metabolic adaptation of *Pseudomonas aeruginosa* to the hypoxic cystic fibrosis lung environment. Int. J. Med. Microbiol., 304(8):1050–1061, Nov 2014.
- M. S. Favero, L. A. Carson, W. W. Bond, und N. J. Petersen. *Pseudomonas aeruginosa*: growth in distilled water from hospitals. Science, 173(3999):836–838, Aug 1971.
- E. J. Feil und B. G. Spratt. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol., 55:561–590, 2001.
- V. Finck-Barbancon, J. Goranson, L. Zhu, T. Sawa, J. P. Wiener-Kronish, S. M. Fleiszig, C. Wu, L. Mende-Mueller, und D. W. Frank. ExoU expression by *Pseudomonas aeruginosa* correlates with acute cytotoxicity and epithelial injury. Mol. Microbiol., 25(3): 547–557, Aug 1997.
- B. Ghysels, B. T. Dieu, S. A. Beatson, J. P. Pirnay, U. A. Ochsner, M. L. Vasil, und P. Cornelis. FpvB, an alternative type I ferripyoverdine receptor of *Pseudomonas aeruginosa*. Microbiology (Reading, Engl.), 150(Pt 6):1671–1680, Jun 2004.
- R. L. Gibson, J. L. Burns, und B. W. Ramsey. Pathophysiology and management of pulmonary infections in cystic fibrosis. Am. J. Respir. Crit. Care Med., 168(8):918–951, Oct 2003.
- J. R. Govan und V. Deretic. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. Microbiol. Rev., 60(3):539–574, Sep 1996.
- A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neubock, und I. L. Hofacker. The Vienna RNA websuite. Nucleic Acids Res., 36(Web Server issue):W70–74, Jul 2008.
- M. Gómez-Lozano, R. L. Marvig, S. Molin, und K. S. Long. Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. Environ. Microbiol., 14(8):2006–2016, Aug 2012.
- J. Hacker und J. B. Kaper. Pathogenicity islands and the evolution of pathogenic microbes. Springer, Berlin, 2002.
- A. R. Hauser, P. J. Kang, und J. N. Engel. PepA, a secreted protein of *Pseudomonas aeruginosa*, is necessary for cytotoxicity and virulence. Mol. Microbiol., 27(4):807–818, Feb 1998.

- M. A. Hernández-Prieto, V. Schon, J. Georg, L. Barreira, J. Varela, W. R. Hess, und M. E. Futschik. Iron deprivation in *Synechocystis*: inference of pathways, non-coding RNAs, and regulatory elements from comprehensive expression profiling. G3 (Bethesda), 2(12):1475–1495, Dec 2012.
- R. Hilker, K. B. Stadermann, D. Doppmeier, J. Kalinowski, J. Stoye, J. Straube, J. Winnebold, und A. Goesmann. ReadXplorer—visualization and analysis of mapped sequences. Bioinformatics, 30(16):2247–2254, Aug 2014.
- R. Hilker, A. Munder, J. Klockgether, P. M. Losada, P. Chouvarine, N. Cramer, C. F. Davenport, S. Dethlefsen, S. Fischer, H. Peng, T. Schönfelder, O. Türk, L. Wiehlmann, F. Wolbeling, E. Gulbins, A. Goesmann, und B. Tümmler. Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome from disease and environment. Environ. Microbiol., 17(1):29–46, Jan 2015.
- V. Huangyutitham, Z. T. Guvener, und C. S. Harwood. Subcellular clustering of the phosphorylated WspR response regulator protein stimulates its diguanylate cyclase activity. MBio, 4(3):e00242–00213, 2013.
- S. Huja, Y. Oren, E. Trost, E. Brzuszkiewicz, D. Biran, J. Blom, A. Goesmann, G. Gottschalk, J. Hacker, E. Z. Ron, und U. Dobrindt. Genomic avenue to avian colisepticemia. MBio, 6(1), 2015.
- D. H. Huson und D. Bryant. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol., 23(2):254–267, Feb 2006.
- K. M. Jørgensen, T. Wassermann, H. K. Johansen, L. E. Christiansen, S. Molin, N. Høiby, und O. Ciofu. Diversity of Metabolic Profiles of Cystic Fibrosis *P. aeruginosa* during the Early Stages of Lung Infection. Microbiology (Reading, Engl.), Apr 2015.
- C. Kiewitz und B. Tümmler. Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. J. Bacteriol., 182(11):3125–3135, Jun 2000.
- J. Klockgether, O. Reva, K. Larbig, und B. Tümmler. Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. J. Bacteriol., 186(2):518–534, Jan 2004.
- J. Klockgether, D. Würdemann, O. Reva, L. Wiehlmann, und B. Tümmler. Diversity of the abundant pKLC102/PAGI-2 family of genomic islands in *Pseudomonas aeruginosa*. J. Bacteriol., 189(6):2443–2459, Mar 2007.
- J. Klockgether, N. Cramer, L. Wiehlmann, C. F. Davenport, und B. Tümmler. *Pseudomonas aeruginosa* Genomic Structure and Diversity. Front Microbiol, 2:150, 2011.

- S. L. Kuchma, K. M. Brothers, J. H. Merritt, N. T. Liberati, F. M. Ausubel, und G. A. O’Toole. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. J. Bacteriol., 189(22): 8165–8178, Nov 2007.
- V. L. Kung, E. A. Ozer, und A. R. Hauser. The accessory genome of *Pseudomonas aeruginosa*. Microbiol. Mol. Biol. Rev., 74(4):621–641, Dec 2010.
- A. Köhler, H. Karch, T. Beikler, T. F. Flemmig, S. Suerbaum, und H. Schmidt. Multi-locus sequence analysis of *Porphyromonas gingivalis* indicates frequent recombination. Microbiology (Reading, Engl.), 149(Pt 9):2407–2415, Sep 2003.
- K. D. Larbig, A. Christmann, A. Johann, J. Klockgether, T. Hartsch, R. Merkl, L. Wiehlmann, H. J. Fritz, und B. Tümmler. Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. J. Bacteriol., 184(23): 6665–6680, Dec 2002.
- D. G. Lee, J. M. Urbach, G. Wu, N. T. Liberati, R. L. Feinbaum, S. Miyata, L. T. Diggins, J. He, M. Saucier, E. Deziel, L. Friedman, L. Li, G. Grills, K. Montgomery, R. Kucherlapati, L. G. Rahme, und F. M. Ausubel. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol., 7(10):R90, 2006.
- H. Li und R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14):1754–1760, Jul 2009.
- H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, und R. Durbin. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16):2078–2079, Aug 2009.
- J. F. Linares, R. Moreno, A. Fajardo, L. Martinez-Solano, R. Escalante, F. Rojo, und J. L. Martinez. The global regulator Crc modulates metabolism, susceptibility to antibiotics and virulence in *Pseudomonas aeruginosa*. Environ. Microbiol., 12(12):3196–3212, Dec 2010.
- Y. Liu und B. Schmidt. Long read alignment based on maximal exact match seeds. Bioinformatics, 28(18):i318–i324, Sep 2012.
- M. Madigan und J. Martinko. Brock Mikrobiologie. 11., überarbeitete Auflage. Pearson Education Deutschland, 2006.
- M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H.

- Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437(7057):376–380, Sep 2005.
- T. Markussen, R. L. Marvig, M. Gomez-Lozano, K. Aanæs, A. E. Burleigh, N. Høiby, H. K. Johansen, S. Molin, and L. Jelsbak. Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. MBio, 5(5):e01592–01514, 2014.
- L. B. Martin, Z. Fei, J. J. Giovannoni, and J. K. Rose. Catalyzing plant science research with RNA-seq. Front Plant Sci, 4:66, 2013.
- T. Massingham und N. Goldman. Error-correcting properties of the SOLiD Exact Call Chemistry. BMC Bioinformatics, 13:145, 2012.
- K. Mathee, G. Narasimhan, C. Valdes, X. Qiu, J. M. Matewish, M. Koehrsen, A. Rokas, C. N. Yandava, R. Engels, E. Zeng, R. Olavarietta, M. Doud, R. S. Smith, P. Montgomery, J. R. White, P. A. Godfrey, C. Kodira, B. Birren, J. E. Galagan, und S. Lory. Dynamics of *Pseudomonas aeruginosa* genome evolution. Proc. Natl. Acad. Sci. U.S.A., 105(8):3100–3105, Feb 2008.
- J. Maynard Smith. Population Genetics of Bacteria: Symposium 52. Cambridge University Press, 1995.
- J. P. McCutcheon und C. D. von Dohlen. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol., 21(16):1366–1372, Aug 2011.
- Y. Michel-Briand und C. Baysse. The pyocins of *Pseudomonas aeruginosa*. Biochimie, 84(5-6):499–510, 2002.
- H. Mikkelsen, R. McMullan, und A. Filloux. The *Pseudomonas aeruginosa* reference strain PA14 displays increased virulence due to a mutation in ladS. PLoS ONE, 6(12):e29113, 2011.
- R. Milkman und M. M. Bridges. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics, 126(3):505–517, Nov 1990.
- Z. Mohd-Zain, S. L. Turner, A. M. Cerdeno-Tarraga, A. K. Lilley, T. J. Inzana, A. J. Duncan, R. M. Harding, D. W. Hood, T. E. Peto, und D. W. Crook. Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary

- origin with a diverse family of syntenic genomic islands. J. Bacteriol., 186(23):8114–8122, Dec 2004.
- A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods, 5(7):621–628, Jul 2008.
- A. G. Nerlich, C. J. Haas, A. Zink, U. Szeimies, and H. G. Hagedorn. Molecular evidence for tuberculosis in an ancient Egyptian mummy. Lancet, 350(9088):1404, Nov 1997.
- M. J. Papagrigorakis, C. Yapijakis, P. N. Synodinos, and E. Baziotopoulou-Valavani. DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. Int. J. Infect. Dis., 10(3):206–214, May 2006.
- W. H. Park und A. W. Williams. A study of pneumococci: A comparison between the pneumococci found in the throat secretions of healthy persons living in both city and country and those obtained from pneumonic exudates and diseased mucous membranes. J. Exp. Med., 7(5):403–419, Aug 1905.
- D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, and A. Soldatov. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res., 37(18):e123, Oct 2009.
- H. M. Patel und C. T. Walsh. In vitro reconstitution of the *Pseudomonas aeruginosa* non-ribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. Biochemistry, 40(30):9023–9031, Jul 2001.
- S. Pellett, D. V. Bigley, und D. J. Grimes. Distribution of *Pseudomonas aeruginosa* in a riverine ecosystem. Appl. Environ. Microbiol., 45(1):328–332, Jan 1983.
- J. P. Pirnay, S. Matthijs, H. Colak, P. Chablain, F. Bilocq, J. Van Eldere, D. De Vos, M. Zizi, L. Triest, und P. Cornelis. Global *Pseudomonas aeruginosa* biodiversity as reflected in a Belgian river. Environ. Microbiol., 7(7):969–980, Jul 2005.
- J. P. Pirnay, F. Bilocq, B. Pot, P. Cornelis, M. Zizi, J. Van Eldere, P. Deschaght, M. Vanechoutte, S. Jennes, T. Pitt, und D. De Vos. *Pseudomonas aeruginosa* population structure revisited. PLoS ONE, 4(11):e7740, 2009.
- N. Pradervand, S. Sulser, F. Delavat, R. Miyazaki, I. Lamas, und J. R. van der Meer. An operon of three transcriptional regulators controls horizontal gene transfer of the integrative and conjugative element ICEclc in *Pseudomonas knackmussii* B13. PLoS Genet., 10(6):e1004441, Jun 2014.

- D. T. Pride, R. J. Meinersmann, T. M. Wassenaar, und M. J. Blaser. Evolutionary implications of microbial genome tetranucleotide frequency biases. Genome Res., 13(2):145–158, Feb 2003.
- X. Qiu, A. U. Gurkar, und S. Lory. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U.S.A., 103(52):19830–19835, Dec 2006.
- A. R. Quinlan und I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6):841–842, Mar 2010.
- L. G. Rahme, E. J. Stevens, S. F. Wolfort, J. Shao, R. G. Tompkins, und F. M. Ausubel. Common virulence factors for bacterial pathogenicity in plants and animals. Science, 268(5219):1899–1902, Jun 1995.
- J. Ramos. Pseudomonas. New York: Kluwer Academic/Plenum Publishers, 2004.
- D. Ramsköld, E. T. Wang, C. B. Burge, und R. Sandberg. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol., 5(12):e1000598, Dec 2009.
- M. H. Rau, R. L. Marvig, G. D. Ehrlich, S. Molin, und L. Jelsbak. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. Environ. Microbiol., 14(8):2200–2211, Aug 2012.
- O. N. Reva und B. Tümmler. Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. BMC Bioinformatics, 5:90, Jul 2004.
- U. Römling, J. Wingender, H. Muller, und B. Tümmler. A major *Pseudomonas aeruginosa* clone common to patients and aquatic habitats. Appl. Environ. Microbiol., 60(6):1734–1738, Jun 1994.
- U. Römling, K. D. Schmidt, und B. Tümmler. Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. J. Mol. Biol., 271(3):386–404, Aug 1997.
- U. Römling, A. Kader, D. D. Sriramulu, R. Simm, und G. Kronvall. Worldwide distribution of *Pseudomonas aeruginosa* clone C strains in the aquatic environment and cystic fibrosis patients. Environ. Microbiol., 7(7):1029–1038, Jul 2005.
- K. M. Sall, M. G. Casabona, C. Bordini, P. Huber, S. de Bentzmann, I. Attree, und S. Elsen. A *gacS* deletion in *Pseudomonas aeruginosa* cystic fibrosis isolate CHA shapes its virulence. PLoS ONE, 9(4):e95936, 2014.

- T. Sawa. The molecular mechanism of acute lung injury caused by *Pseudomonas aeruginosa*: from bacterial pathogenesis to host response. J Intensive Care, 2(1):10, 2014.
- K. D. Schmidt, B. Tümmler, und U. Römling. Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. J. Bacteriol., 178(1):85–93, Jan 1996.
- K. Selezska, M. Kazmierczak, M. Musken, J. Garbe, M. Schobert, S. Häussler, L. Wiehlmann, C. Rohde, und J. Sikorski. *Pseudomonas aeruginosa* population structure revisited under environmental focus: impact of water quality and phage pressure. Environ. Microbiol., 14(8):1952–1967, Aug 2012.
- W. She, E. Mordukhova, H. Zhao, Z. M. Petrushenko, und V. V. Rybenkov. Mutational analysis of MukE reveals its role in focal subcellular localization of MukBEF. Mol. Microbiol., 87(3):539–552, Feb 2013.
- L. Silo-Suh, S. J. Suh, P. V. Phibbs, und D. E. Ohman. Adaptations of *Pseudomonas aeruginosa* to the cystic fibrosis lung environment can include deregulation of *zwf*, encoding glucose-6-phosphate dehydrogenase. J. Bacteriol., 187(22):7561–7568, Nov 2005.
- M. Sivaneson, H. Mikkelsen, I. Ventre, C. Bordi, und A. Filloux. Two-component regulatory systems in *Pseudomonas aeruginosa*: an intricate network mediating fimbrial and efflux pump gene expression. Mol. Microbiol., 79(5):1353–1366, Mar 2011.
- E. E. Smith, E. H. Sims, D. H. Spencer, R. Kaul, und M. V. Olson. Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. J. Bacteriol., 187(6):2138–2147, Mar 2005.
- J. M. Smith, N. H. Smith, M. O’Rourke, und B. G. Spratt. How clonal are bacteria? Proc. Natl. Acad. Sci. U.S.A., 90(10):4384–4388, May 1993.
- E. Sonnleitner und D. Haas. Small RNAs as regulators of primary and secondary metabolism in *Pseudomonas* species. Appl. Microbiol. Biotechnol., 91(1):63–79, Jul 2011.
- E. Sonnleitner, L. Abdou, und D. Haas. Small RNA as global regulator of carbon catabolite repression in *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U.S.A., 106(51):21866–21871, Dec 2009.
- D. H. Spencer, A. Kas, E. E. Smith, C. K. Raymond, E. H. Sims, M. Hastings, J. L. Burns, R. Kaul, und M. V. Olson. Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. J. Bacteriol., 185(4):1316–1325, Feb 2003.
- B. G. Spratt und M. C. Maiden. Bacterial population genetics, evolution and epidemiology. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 354(1384):701–710, Apr 1999.

- P. Stothard und D. S. Wishart. Circular genome visualization and exploration using CGView. Bioinformatics, 21(4):537–539, Feb 2005.
- C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, und M. V. Olson. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature, 406(6799):959–964, Aug 2000.
- H. Thorvaldsdottir, J. T. Robinson, und J. P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinformatics, 14(2):178–192, Mar 2013.
- F. Toval, C. D. Köhler, U. Vogel, F. Wagenlehner, A. Mellmann, A. Fruth, M. A. Schmidt, H. Karch, M. Bielaszewska, und U. Dobrindt. Characterization of *Escherichia coli* isolates from hospital inpatients or outpatients with urinary tract infection. J. Clin. Microbiol., 52(2):407–418, Feb 2014.
- B. Tümmler, L. Wiehlmann, J. Klockgether, und N. Cramer. Advances in understanding *Pseudomonas*. F1000Prime Rep, 6:9, 2014.
- B. Valot, C. Guyeux, J. Y. Rolland, K. Mazouzi, X. Bertrand, und D. Hocquet. What It Takes to Be a *Pseudomonas aeruginosa*? The Core Genome of the Opportunistic Pathogen Updated. PLoS ONE, 10(5):e0126468, 2015.
- I. Ventre, A. L. Goodman, I. Vallet-Gely, P. Vasseur, C. Soscia, S. Molin, S. Bleves, A. Lazdunski, S. Lory, und A. Filloux. Multiple sensors control reciprocal expression of *Pseudomonas aeruginosa* regulatory RNA and virulence genes. Proc. Natl. Acad. Sci. U.S.A., 103(1):171–176, Jan 2006.
- L. Wiehlmann, G. Wagner, N. Cramer, B. Siebert, P. Gudowius, G. Morales, T. Köhler, C. van Delden, C. Weinel, P. Slickers, und B. Tümmler. Population structure of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U.S.A., 104(19):8101–8106, May 2007.
- D. Williams, B. Evans, S. Haldenby, M. J. Walshaw, M. A. Brockhurst, C. Winstanley, und S. Paterson. Divergent, coexisting *Pseudomonas aeruginosa* lineages in chronic cystic fibrosis lung infections. Am. J. Respir. Crit. Care Med., 191(7):775–785, Apr 2015.
- P. Williams und M. Camara. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. Curr. Opin. Microbiol., 12(2):182–191, Apr 2009.

- G. L. Winsor, D. K. Lam, L. Fleming, R. Lo, M. D. Whiteside, N. Y. Yu, R. E. Hancock, und F. S. Brinkman. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res., 39 (Database issue):596–600, Jan 2011.
- S. Wu, Y. Chen, G. Mao, S. Trobro, M. Kwiatkowski, und L. A. Kirsebom. Transition-state stabilization in *Escherichia coli* ribonuclease P RNA-mediated cleavage of model substrates. Nucleic Acids Res., 42(1):631–642, Jan 2014.
- O. Wurtzel, D. R. Yoder-Himes, K. Han, A. A. Dandekar, S. Edelheit, E. P. Greenberg, R. Sorek, und S. Lory. The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLoS Pathog., 8(9):e1002945, Sep 2012.
- T. L. Yahr, J. Goranson, und D. W. Frank. Exoenzyme S of *Pseudomonas aeruginosa* is secreted by a type III pathway. Mol. Microbiol., 22(5):991–1003, Dec 1996.
- D. R. Zerbino und E. Birney. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res., 18(5):821–829, May 2008.

Lebenslauf

Name: Sebastian Fischer

Geboren: 05.11.1986 in Frankenberg/Sachsen

Staatsangehörigkeit: deutsch

Familienstand: ledig

Ausbildung

- 09/2011 - 06/2015 Bearbeitung der vorliegenden Dissertationsschrift unter der Leitung von Prof. Dr. Dr. Burkhard Tümmler;
Medizinische Hochschule Hannover
- 05/2011 - 09/2011 Sprach- und Kulturreise durch Europa
- 10/2009 - 05/2011 Masterstudium an der Medizinischen Hochschule Hannover, Masterarbeit unter der Leitung von Prof. Dr. Dr. Burkhard Tümmler mit dem Titel
„Mikroevolution eines *Pseudomonas aeruginosa* Klons im atypischen Habitat der Mukoviszidoseleunge“
- 10/2007-09/2009 Bachelorstudium an der Leibniz Universität Hannover;
Bachelorarbeit unter der Leitung von Dr. Clemens Bockmeyer mit dem Titel „Mikro-RNA Profil basaler und luminaler Mammaepithelien“
- 05/2006 - 09/2006 Laborassistent im Trink- und Abwasserlabor des ZWA
„Mittleres Erzgebirgsvorland“
- 08/2005 - 04/2006 Zivildienst im Kindergarten „Villa Zwergenland“
- 1997 - 2005 Abitur am Christian-Fürchtegott-Gellert-Gymnasium in Hainichen
- 1993 - 1997 Grundschule Eduard Feldner in Hainichen

Praktika

August 2004 zweiwöchiges Praktikum am Wehrwissenschaftlichen Institut für Schutztechnologien, Munster

Anhang

Der Ordner „Skripte“ enthält:

| | |
|-------------------------------|--|
| extract_positions.pl | Abdeckung ausgewählter Positionen aus .vcf-Datei |
| ExonicGC_fromTranscriptome.pl | Exon-Länge und GC-Gehalt für Transkriptomanalyse |
| gather.py | Zusammenfassung stammspezifischer Gene hoher Homologie |
| GC_DESeq.R | Bestimmung differentiell exprimierter ORFs |
| haplotypes_identification.pl | Berechnung der Haplotypmatrix |
| heatmap.R | Berechnung der Heatmaps des akzessorischen Genoms |
| LengthCoverage.pl | Bestimmung Readabdeckung über Genlänge |
| PrepareCountsForDESeq.pl | Hilfsskript für die Berechnung differentiell exprimierter Gene |
| SequenceReplacer.java | Inkorporiert SNPs in Referenzgenom |

Der Ordner „Tabellen“ enthält:

| | |
|-------------|---|
| Tabelle S1 | Annotation NN2 Genom |
| Tabelle S2 | 5'-untranslatierte Bereiche NN2 |
| Tabelle S3 | Haplotyplängenverteilung |
| Tabelle S4 | Aminosäureaustausche |
| Tabelle S5 | Gene hoher Mutationsfrequenz |
| Tabelle S6 | Einzelnukleotidaustausche in ncRNAs pro Isolat |
| Tabelle S7 | Zusammenfassung der Einzelnukleotidaustausche in ncRNAs |
| Tabelle S8 | Thermodynamische Stabilität ncRNAs |
| Tabelle S9 | Große Deletionen |
| Tabelle S10 | Indels pro Isolat |
| Tabelle S11 | zusätzliche Gene der klonalen Komplexe |
| Tabelle S12 | Fitnessresultate |

Der Ordner „SNPs“ enthält für jedes Isolat eine Datei mit allen SNPs und die zugehörigen funktionellen Charakteristika.

Der Ordner „Annotation“ enthält die Datei KlonC_NN2.gbk.