Pohl, S., Klockgether, J., Eckweiler, D., Khaledi, A., Schniederjans, M., Chouvarine, P., Tümmler, B., Häussler, S.

The extensive set of accessory Pseudomonas aeruginosa genomic components

(2014) FEMS Microbiology Letters, 356 (2), pp. 235-241.

The extensive set of accessory *Pseudomonas aeruginosa* genomic components

Sarah Pohl[1,2], Jens Klockgether[3], Denitsa Eckweiler[1,2], Ariane Khaledi[1,2], Monika Schniederjans[1,2], Philippe Chouvarine[3], Burkhard Tümmler[3], Susanne Häussler[1,2*]

[1] Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2] Institute for Molecular Bacteriology, TWINCORE GmbH, Centre for Clinical and Experimental Infection Research, a joint venture of the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany

[3] Clinical Research Group 'Molecular Pathology of Cystic Fibrosis and Pseudomonas Genomics', Clinic for Pediatric Pneumology, Allergology and Neonatology, OE 6710, Hannover Medical School, Hannover, Germany

*Corresponding author:

Susanne Häussler

Address: Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig

Tel: +49 (0)531 61 81 - 30 00

Fax: +49 (0)531 61 81 - 30 99

Email: susanne.haeussler@helmholtz-hzi.de

Keywords

Extended gene pool, horizontal gene transfer

Running Title

Extended Accessory Genome of *Pseudomonas aeruginosa*

Abstract

Up to 20% of the chromosomal *Pseudomonas aeruginosa* DNA belong to the so-called accessory genome. Its elements are specific for subgroups or even single strains and are likely acquired by horizontal gene transfer (HGT). Similarities of the accessory genomic elements to DNA from other bacterial species, mainly the DNA of γ- and β-proteobacteria, indicate a role of interspecies HGT. In this study we analysed the expression of the accessory genome in 150 clinical *P. aeruginosa* isolates as uncovered by transcriptome sequencing and the presence of accessory genes in eleven additional isolates. Remarkably, despite the large number of *P. aeruginosa* strains that have been sequenced to date, we found new strain-specific compositions of accessory genomic elements and a high portion (10 – 20%) of genes without *P. aeruginosa* homologues. Although some genes were detected to be expressed/present in several isolates, individual patterns regarding the genes, their functions and the possible origin of the DNA were widespread among the tested strains. Our results demonstrate the unaltered potential to discover new traits within the *P. aeruginosa* population and underline that the *P. aeruginosa* pangenome is likely to increase with increasing sequence information.

Introduction

Pseudomonas aeruginosa bacteria inhabit various aquatic and terrestrial environments and are a frequent cause of opportunistic infections in animal and human hosts. Colonisation of such a broad spectrum of habitats goes along with broad metabolic versatility and a high potential to adapt to new environments

(Ramos, 2004). In addition to many common features, *P. aeruginosa* isolates often display individual phenotypic traits. This phenotypic diversity is reflected in the composition of the genomes: The major part of a *P. aeruginosa* genome, the core genome, is found in all strains and, with the exception of a few loci, is usually highly conserved with only 0.5 – 0.7% sequence diversity (Spencer *et al.*, 2003; Cramer *et al.*, 2011). Less conserved loci are the so-called 'replacement islands'. Flagellin glycosylation genes, O-antigen and pyoverdine biosynthesis gene clusters and the major pilin gene *pilA* are present in all genomes. However, they are under diversifying selection (Kung *et al.*, 2010) and can be individually shaped with regards to size and gene composition. Comparative analyses revealed different types and subtypes of these gene clusters (Klockgether *et al.*, 2011).

Besides that, up to 20% of a genome can be composed of DNA blocks specific for subgroups of strains or even single strains. Such DNA blocks appear as genomic islands or islets inserted between core genome parts at various chromosomal loci. Their sizes range from a few hundred bp up to 200 kbp, and together they make up a strain's accessory genome. The individual composition causes variable genome sizes, usually between six and seven Mbp (Klockgether *et al.*, 2011). Accessory DNA blocks often display features of mobile elements such as phages, transposons or integrative and conjugative elements (Kung *et al.*, 2010), indicating the importance of mobile DNA in shaping the accessory genome. These mobile elements serve as vehicles for associated 'cargo' DNA that integrates into the host genome as well. Several large genomic islands and smaller insertions have been described so far (Klockgether *et al.*, 2011). They were either sequenced separately, such as PAGI-1 (Liang *et al.*, 2001) or PAGI-5 to -11 (Battle *et al.*, 2008; Battle *et al.*, 2009), or they were identified from fully sequenced strains upon comparison with existing genomes,

such as the large genomic islands and prophages of strain LESB58 (Winstanley *et al*., 2009).

The presence of genomic islands in the accessory genome of *P. aeruginosa* goes along with the acquisition of additional functional traits. Genes typical for mobile elements and ORFs encoding for hypothetical or uncharacterized proteins have been found, as well as genes encoding catabolic traits or antibiotic resistance features (Kung *et al.*, 2010). Contribution of genomic islands to a strain's pathogenic potential has also been reported (He *et al.*, 2004). While accessory genome elements generally display homology to DNA from other bacteria, especially the appearance of mobile elements indicates horizontal gene transfer (HGT). These mobile elements are often related in various species, mainly β- and γ-proteobacteria (Klockgether *et al.*, 2006), implying HGT across species borders. Thus, *P. aeruginosa* has in principle access to an extended pool of DNA which can contribute to a strain's accessory genome. Questions on size and possible restrictions of this DNA pool and to what extent it is actually exploited by single *P. aeruginosa* strains are difficult to answer, but increasing insights are expected with every additional genome being sequenced. Recent accessory genome analyses still revealed dozens of genes without *P. aeruginosa* homologues, but homologues from other species – other Pseudomonads as well as other genera (Bezuidt *et al.*, 2013; Klockgether *et al.*, 2013).

In this study we screened for the presence and the expression of accessory genes in a large collection of *P. aeruginosa* isolates. Searches for high sequence identities to genes from previously sequenced *P. aeruginosa* strains, other Pseudomonads and other genera were performed, and the nature and distribution of acquired genes within our strain pools were evaluated. Furthermore, in order to gain broader knowledge on the principles of accessory genome composition, predicted functions of

100 the acquired genes were assessed. Analysis was done on full transcriptome datasets
101 for 150 individual clinical isolates and on genome sequence data for eleven *P.*
102 *aeruginosa* strains isolated from airways of cystic fibrosis patients.
103

Materials & Methods

P. aeruginosa isolates

The 150 transcriptome sequenced clinical isolates were provided by different clinics or research institutions; 87 from Hannover Medical School (MHH), 40 from a strain collection of the University of Freiburg, 13 from the Robert-Koch-Institute in Wernigerode and 10 from the National Reference Laboratory for multidrug-resistant Gram-negative bacteria in Bochum.

Eleven genome sequenced isolates were chosen from a collection of isolates from the airways of CF patients at Hannover Medical School. Genotyping with a custom-made microarray (Wiehlmann *et al.*, 2007) indicated that the selected isolates represent clonal lineages that persisted in the respective patient airways for seven years or longer. Analysis was performed on the sequenced genomes of the early isolates.

Sequencing

Whole transcriptome-sequencing was performed using a custom-made protocol with barcoded RNA-libraries to enable pooled sequencing of several samples (Dötsch *et al.*, 2012). Briefly, bacterial cultures were grown under standard conditions (LB broth, 37 °C) and harvested in RNAprotect (Qiagen) at OD_{600}= 2. Total RNA was extracted (RNeasy plus, Qiagen) and the ribosomal RNA was depleted (MICROBExpress, Ambion). After mechanical fragmentation of the mRNA and ligation with RNA-adapters containing 6 nt barcode sequences for multiplexing, strand-specific cDNA libraries were generated. Up to 12 libraries were pooled and sequenced, after additional rRNA removal, on an Illumina GenomeAnalyzer-II_x generating paired-end

reads of 76 or 110 bp. The raw sequence reads were sorted by their expected barcodes, which were subsequently removed. Reads with more than one error in their barcode were discarded; the others were trimmed using the fastq-mcf script of the ea-utils package (Aronesty, 2011), removing adapters and low quality sequences.

Genome sequencing was done on a SOLiD 5500XL system (Life Technologies). Bead-coupled fragment libraries for high-throughput sequencing were generated from genomic DNA of bacteria grown at 37°C in LB-medium. DNA fragmentation and processing was done using kits and standard protocols provided by Life Technologies (Fragment Library Preparation: 5500 Series SOLiD™ Systems User Guide (Part no. 4460960), 5500 SOLiD™ Fragment Library Core Kit (Part no. 4464412) and EZ Bead™ E20 System Consumables (Part no. 4453094)). The resulting datasets consisted of 75 bp reads, which were trimmed by removing 3' ends with bad base call qualities (< 20). Duplicated reads were removed before further analysis. The reads for each isolate were aligned to the PA14 reference genome (NC_008463.1) using NovoalignCS. Reads not aligned to the PA14 reference were then used for accessory genome analysis.

De novo **Analysis**

The transcriptome reads were mapped against five *P. aeruginosa* reference genomes (PA14, PAO1, LESB58, PACS2 and PA7) using Stampy (Lunter & Goodson, 2011). Sequencing reads that did not map to any of the references were used in a *de novo* transcriptome assembly approach using Velvet (Zerbino, 2008) with a wide range of k-mers (27 to 37) and a minimal contig length of 100 bp. 202,907 contigs were blasted against all microbial genes downloaded from the Microbial Genome Database (MBGD) (Uchiyama *et al.*, 2010) using a minimal hit

length of 100 bp and sequence similarity higher than 90% as cut-off values. An initial accessory gene list containing 9242 entries was established by extracting always the first coding sequence that neither belonged to nor was an ortholog of the *P. aeruginosa* reference genomes used. Previously unmapped sequencing reads were mapped against this list using Stampy, and gene coverage was calculated based on the mapping results. Final analysis was performed only on genes with at least 90% coverage in the respective isolate(s). All figures were produced using R for statistical computing (R Core Team, 2013) and the pheatmap package (Kolde, 2013).

De novo assembly of genomic reads was done using ABySS (Birol *et al.*, 2009) with k-mer size spaces optimised for each dataset to maximise the N50 value. Contigs smaller than 200 bps were removed, while contigs ≥ 200 bps were assembled into supercontigs using the software CAP3 (Huang & Madan, 1999). Supercontigs and unassembled regular contigs were then aligned to all bacterial entries of the UniProt database (UniProt Consortium, 2014) using blastx to identify known proteins/genes from other *P. aeruginosa* strains or other species. The results were filtered for the best hit of a given region in a contig in order to remove multiple descriptions for the same coding region.

Results

Accessory Genome Analysis from Transcriptome Sequencing

When the transcriptome sequencing reads of 150 clinical *Pseudomonas aeruginosa* isolates were mapped against five *P. aeruginosa* reference genomes (PA14, PAO1, LESB58, PACS2 and PA7), around 57% of the reads mapped to all five references, and around 20% mapped to at least one of them. Sequencing reads that did not map to any of the references were used in a *de novo* transcriptome assembly approach resulting in an accessory gene collection containing 9242 entries. Mapping of the previously unmapped sequencing reads to this accessory gene collection revealed the expression of 1164 genes with at least 90% coverage. 106 of these genes have previously been found in other *P. aeruginosa* strains and 1058 genes showed homologues in a variety of non-*P. aeruginosa* species (Table S1).

The vast majority of the latter genes were found in only few isolates: 752 (71%) genes were well-covered in one to five, another 129 (12%) in six to ten isolates. Furthermore, the number and combination of new genes varied between the isolates (Figure 1). A group of three clinical isolates contained no well-covered non-*P. aeruginosa* homologues, and another 53 contained less than 20 new genes. Nevertheless, most isolates displayed several dozens of non-*P. aeruginosa* homologues (up to 162) in their accessory genomes. While subgroups of isolates shared small sets of genes, no large common gene clusters were found. Apparently, most of the 150 clinical isolates harbour an individual set of accessory genes.

Protein descriptions were obtained from the Microbial Genome Database (Table S1) and grouped into the predicted protein function categories shown in Table 1. Overall, 40% of the 1058 transcribed genes are described to encode uncharacterised or hypothetical proteins, followed by proteins related to DNA integration and

196 transposition (10%) and transcriptional regulators (6%). Other frequent functional
197 categories are metal resistance (35 proteins) and antibiotic resistance (38 proteins).
198 The metal resistance proteins, often described as copper resistance proteins,
199 originated from species like *P. mendocina* and *P. stutzeri*, but also from *Ralstonia*
200 *metallidurans*, *Achromobacter xylosoxidans* and *Stenotrophomonas maltophilia*. The
201 antibiotic resistance proteins also came from various species like *Acinetobacter*
202 *baumannii*, *E. coli*, *Klebsiella pneumoniae* and *Salmonella enterica*. Some of the
203 proteins are β-lactamases, other confer resistance to gentamicin, streptomycin or
204 tetracycline. Since the genes encoding for these proteins were found via
205 transcriptome sequencing, it cannot be differentiated whether they integrated into the
206 genomes or are localized on resistance plasmids.

207 Overall, the 1058 transcribed genes were annotated in 98 potential donor species
208 other than *P. aeruginosa*, and Table 2 presents the most common ones. Predominant
209 are other Pseudomonads like *P. stutzeri* and *P. putida*, of which 114 and 69 genes
210 were found, but also *K. pneumoniae* provided many genes to this pool. Other γ- and
211 β-proteobacteria like *R. metallidurans*, *S. enterica* and *E. coli* appeared multiple times
212 as well.

213 Table 2 also displays the number of isolates which have acquired genes annotated in
214 the respective bacterial species. Not only was a great variety of different genes from
215 other Pseudomonads found in general, but also a large number of clinical isolates
216 expressed at least one gene of *P. stutzeri* (111 isolates) or *P. putida* (52 isolates).
217 Additionally, although a more limited number of genes was annotated in the
218 Enterobacteriaceae *E. coli* (25 genes) and *S. enterica* (42 genes), homologues from
219 these species were found in as many as 58 and 38 of the 150 isolates included in
220 this approach.

Figure 2 illustrates that homologues from a maximum of 33 different species were discovered in single clinical isolates. Interestingly, some pairs of isolates shared almost identical sets of species, and these pairs were mostly phylogenetically closely related. However, since not all of them share the same set of individual genes, this finding indicates that rather than having a common ancestor they shared a common habitat with access to a similar gene pool.

Accessory Genome Analysis from Genome Sequencing

After sequencing the genomes of eleven *P. aeruginosa* isolates from different cystic fibrosis patients, a *de novo* assembly was performed with all reads that did not map to the PA14 reference genome. Blastx comparison of the resulting accessory genome contigs against the UniProt database led to the identification of 69 – 192 loci with homologues in different species. For a majority of the accessory ORFs, the closest homologues were found in other *P. aeruginosa* strains, whereas 10 – 20% were from other bacterial species (Table 3). Similar to the transcriptome results, these other species contained Pseudomonads as well as bacteria from other genera (Table 3). In total, similarities to Pseudomonads such as *P. putida*, *P. stutzeri* or *P. fluorescens* were more frequent (Table 4), but homologues of other proteobacteria were detected for each isolate (Table S2). Most of them were γ-, few were β-proteobacteria (e. g. *Ralstonia* or *Polaromonas* sp.), but also species from other classes were found in individual cases (Table S2).

Analysis of the protein names of the identified ORFs of all analysed isolates revealed that more than 50% code for uncharacterized or hypothetical proteins, which is a slightly higher proportion than in the transcriptomic approach (40%). Some functional categories were found for many isolates (Table S3A), like genes from replacement

islands and integrase/transposase as well as phage- and plasmid-related genes, all reflecting the role of such elements in shaping the accessory genome. Other categories appearing more often were transcriptional regulators and transporter components. Some features were seen for few isolates only, such as copper resistance or Type I restriction modification system genes, and some function predictions appeared in single strains exclusively, e.g. type VI secretion system components in isolate NA1. When focussing only on non-*P. aeruginosa* hits (Table S3B), most were again uncharacterized and hypothetical proteins. Phage components or other proteins typical for mobile elements were also in these shorter lists. However, these lists were all highly individual: no mobile element component appeared as best homologue for ORFs from many isolates.

Discussion

Data of a transcriptomic and a genomic sequencing approach were used independently in this study to analyse the accessory genomes of 150 and eleven clinical *Pseudomonas aeruginosa* isolates, respectively. In the transcriptome analysis, between none and 162 genes without *P. aeruginosa* homologues were detected to be transcribed in individual isolates (mean: 40); the genomic sequencing revealed the presence of 7 – 40 (mean: 20) 'foreign' genes per isolate. Among the non-*P. aeruginosa* species for which the accessory genes were already described, *P. stutzeri* and *P. putida* were detected most frequently in both approaches. However, genetic elements from genera other than *Pseudomonas* were found more often within the transcriptome sequencing results. Most interestingly, independent of the chosen approach, no large groups of isolates shared the same pattern of so far in *P. aeruginosa* unannotated genes within their accessory genomes. In the same line,

many functional elements were found specifically within subgroups or even within single isolates, so that, as for the new genes, no common patterns were seen in the chosen collection. Contrary to this, the transcriptomic approach showed that phylogenetically closely related isolates often express different genes from the same species, implying that they had access to specific gene pools.

In summary, the extended analysis of two independent collections of clinical *P. aeruginosa* isolates indicates that a broad gene pool from various bacterial species serves as a possible source of accessory genome elements that can be accepted by *P. aeruginosa*. Despite the already large number of *P. aeruginosa* genomes deposited in the databases, almost all isolates analysed in this study exhibited genes that were not known to belong to the accessory genome of this species so far. A great variety of potential donor species as well as function predictions could be detected by our approaches, demonstrating the ongoing possibility to find individual genomic traits whenever the DNA or the RNA of an independent *P. aeruginosa* strain is sequenced. This would imply an extended pangenome for this species, the limits of which could not be unequivocally estimated.

Acknowledgements

We thank Iris F. Chaberny (Hannover Medical School, Hannover, Germany), Daniel Jonas (Freiburg University Medical Centre, Freiburg, Germany), Wolfgang Witte and Yvonne Pfeifer (Robert-Koch-Institute, Wernigerode, Germany) and Martin Kaase and Sören Gatermann (National Reference Laboratory for multidrug-resistant Gram-negative bacteria, Bochum, Germany) for providing us with clinical *P. aeruginosa* isolates.

This work was supported by the Helmholtz International Graduate School for Infection Research, the Deutsche Forschungsgemeinschaft (CRC 900 ‘Chronic Infections: Microbial Persistence and its Control’, projects A2 and A3) and the European Research Council (Starting Grant 260276 ‘RESISTOME’).

References

Aronesty E (2011) ea-utils: command-line tools for processing biological sequencing data http://code.google.com/p/ea-utils.

Battle SE, Rello J & Hauser AR (2009) Genomic islands of Pseudomonas aeruginosa. *FEMS Microbiol Lett* **290**: 70-8.

Battle SE, Meyer F, Rello J, Kung VL & Hauser AR (2008) Hybrid pathogenicity island PAGI-5 contributes to the highly virulent phenotype of a Pseudomonas aeruginosa isolate in mammals. *J Bacteriol* **190**: 7130-40.

Bezuidt OK, Klockgether J, Elsen S, Attree I, Davenport CF & Tümmler B (2013) Intraclonal genome diversity of Pseudomonas aeruginosa clones CHA and TB. *BMC Genomics* **14**: 416.

Birol I, Jackman SD, Nielsen CB*, et al.* (2009) De novo transcriptome assembly with ABySS. *Bioinformatics* **25**: 2872-7.

Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF & Tümmler B (2011) Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* **13**: 1690-704.

Dötsch A, Eckweiler D, Schniederjans M, Zimmermann A, Jensen V, Scharfe M, Geffers R & Häussler S (2012) The Pseudomonas aeruginosa transcriptome in planktonic cultures and static biofilms using RNA sequencing. *PLoS One* **7**: e31092.

He J, Baldini RL, Deziel E, Saucier M, Zhang Q, Liberati NT, Lee D, Urbach J, Goodman HM & Rahme LG (2004) The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A* **101**: 2530-5.

Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868-77.

Klockgether J, Reva ON & Tümmler B (2006) Spread of genomic islands between clinical and environmental strains. *Prokaryotic diversity: mechanisms and significance,*(Logan NA, Lapin-Scott HM & Oyston PCF, eds.), 187-200. Cambridge University Press, Cambridge, United Kingdom.

328 Klockgether J, Cramer N, Wiehlmann L, Davenport CF & Tümmler B (2011) Pseudomonas
329 aeruginosa Genomic Structure and Diversity. *Front Microbiol* **2**: 150.
330 Klockgether J, Miethke N, Kubesch P*, et al.* (2013) Intraclonal diversity of the Pseudomonas
331 aeruginosa cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures
332 of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost
333 identical genome sequence. *Environ Microbiol* **15**: 191-210.
334 Kolde R (2013) pheatmap: Pretty Heatmaps http://CRAN.R-project.org/package=pheatmap.
335 Kung VL, Ozer EA & Hauser AR (2010) The accessory genome of Pseudomonas
336 aeruginosa. *Microbiol Mol Biol Rev* **74**: 621-41.
337 Liang X, Pham XQ, Olson MV & Lory S (2001) Identification of a genomic island present in
338 the majority of pathogenic isolates of Pseudomonas aeruginosa. *J Bacteriol* **183**: 843-53.
339 Lunter G & Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping
340 of Illumina sequence reads. *Genome Res* **21**: 936-9.
341 R Core Team (2013) R: A Language and Environment for Statistical Computing
342 http://www.R-project.org/.
343 Ramos J-L (ed.) (2004) *Pseudomonas Volume 1: Genomics, Life Style and Molecular*
344 *Architecture*. New York: Kluwer Academics / Plenum publishers.
345 Spencer DH, Kas A, Smith EE, Raymond CK, Sims EH, Hastings M, Burns JL, Kaul R &
346 Olson MV (2003) Whole-genome sequence variation among multiple isolates of
347 Pseudomonas aeruginosa. *J Bacteriol* **185**: 1316-25.
348 Uchiyama I, Higuchi T & Kawai M (2010) MBGD update 2010: toward a comprehensive
349 resource for exploring microbial genome diversity. *Nucleic Acids Res* **38**: D361-5.
350 UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic*
351 *Acids Res* **42**: D191-8.
352 Wiehlmann L, Wagner G, Cramer N*, et al.* (2007) Population structure of Pseudomonas
353 aeruginosa. *Proc Natl Acad Sci U S A* **104**: 8101-6.
354 Winstanley C, Langille MG, Fothergill JL et al. (2009) Newly introduced genomic prophage
355 islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of
356 Pseudomonas aeruginosa. *Genome Res* **19**: 12–23.
357 Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de
358 Bruijn graphs. *Genome Res* **18**: 821–829.

Table 1. Functional profiling of proteins encoded in the accessory genome of 150 clinical isolates without *P. aeruginosa* homologues in the Microbial Genome Database.

Functional group	Number of proteins
uncharacterised/hypothetical	428
integrase/transposase	111
transcriptional regulator	59
antibiotic resistance	38
metal resistance	35
transporter component	26
phage protein	21
Type I restriction modification system	17
helicase	15

Table 2. Species distribution of non-*P. aeruginosa* homologues within the accessory genome of 150 clinical isolates.

Species	No. of genes from the species	No. of isolates with at least one gene from the species
Pseudomonas stutzeri	114	111
Pseudomonas putida	69	52
Klebsiella pneumoniae	64	79
Acidovorax sp.	57	93
Ralstonia metallidurans	51	41
Burkholderia vietnamiensis	50	42
Pseudomonas mendocina	48	82
Alicycliphilus denitrificans	44	89
Salmonella enterica	42	38
Azotobacter vinelandii	29	75
Delftia acidovorans	29	29
Ralstonia pickettii	29	56
Escherichia coli	25	58
Stenotrophomonas maltophilia	25	64
Verminephrobacter eiseniae	24	37
Nitrosomonas europaea	19	38
Achromobacter xylosoxidans	17	61
Dechlorosoma suillum	15	49
Methylophaga sp.	15	29

79 additional species containing less than 15 non-*P. aeruginosa* homologues are not listed.

368 **Table 3.** Numbers of proteins not known from the PA14 reference detected by
369 genome sequencing.

Strain	No. of proteins	From other *P. aeruginosa*	From other Pseudomonads	From other genera
AA2	69	56	5	8
HK3	120	99	12	9
KB1	77	64	12	1
KK1	164	133	24	7
MF1	157	132	12	13
NA1	81	60	12	9
NM1	192	152	29	11
RP1	191	174	9	8
ST1	109	102	5	2
TR1	143	127	12	4
WU2	138	123	7	8

370

371 **Table 4.** Species distribution of non-*P. aeruginosa* homologues in eleven genome
372 sequencing datasets.

Species	**No. of genes from the species**	**No. of isolates with at least one gene from the species**
Pseudomonas sp.	33	9
Pseudomonas putida	27	9
Pseudomonas stutzeri	24	8
Pseudomonas syringae	17	9
Pseudomonas fluorescens	14	8
Pseudomonas chlororaphis	6	2
Acidovorax sp.	5	2
Polaromonas sp.	4	2
Providencia burhodogranariea	4	2
Pseudomonas mendocina	4	3
Ralstonia metallidurans	4	3
Stenotrophomonas maltophilia	4	2
Pseudomonas fulva	3	3
Pseudomonas pseudoalcaligenes	3	3
Salmonella enterica	3	2

373

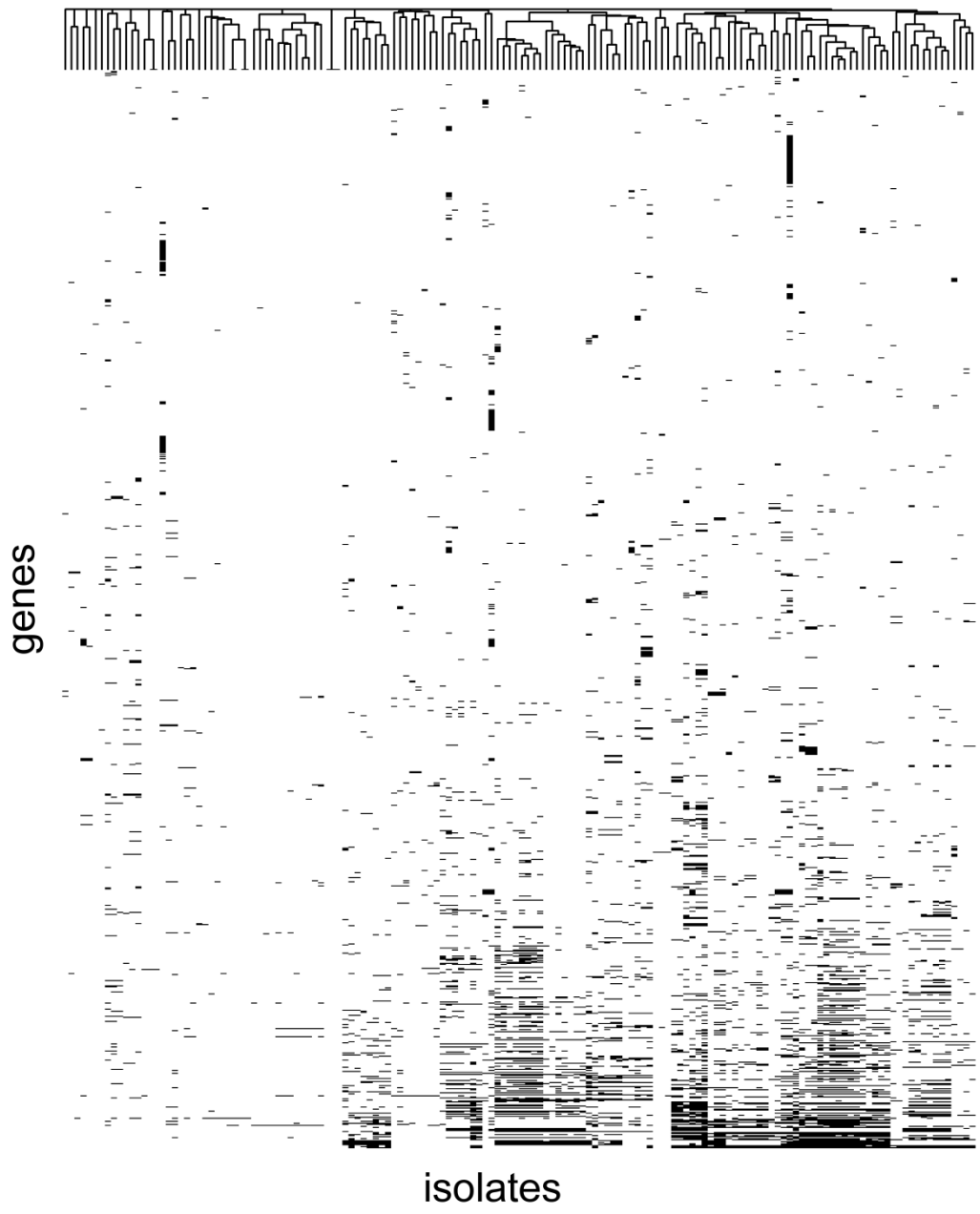
374 55 additional species are not listed. In nine of those two and in 46 of those one non-
375 *P. aeruginosa* homologues were detected.

376

Figure Legends

Figure 1. The heat map depicts the combination of individual accessory genes that are expressed by the 150 clinical isolates. The dendrogram is a visualisation of the distance between the isolates based on the accessory gene composition.

Figure 2. The heat map depicts the probable source of the genes of the accessory genomes of the 150 clinical isolates. The dendrogram is a visualisation of the distance between the isolates based on the species to which their accessory genes are annotated.



387

388 **Figure 1**

389

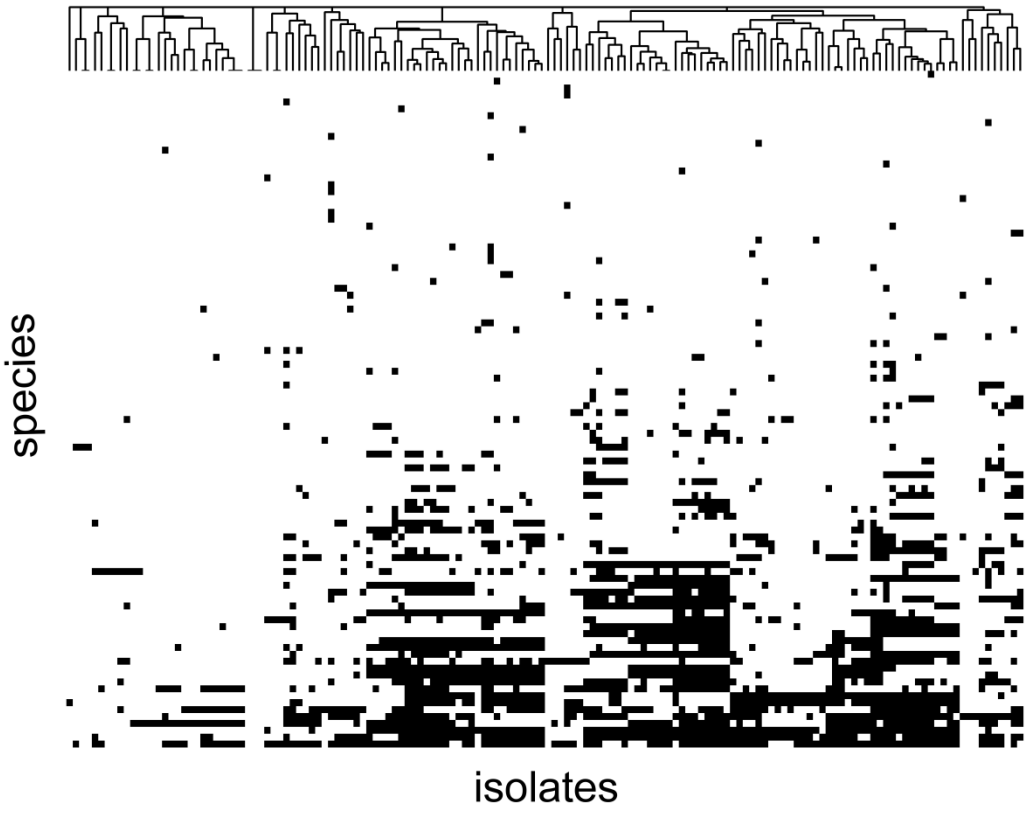


Figure 2