Published in "Environmental Microbiology doi: 10.1111/1462-2920.12498, 2014
which should be cited to refer to this work.

Comparative genome analysis of *Pseudomonas knackmussii* B13, the first bacterium known to degrade chloroaromatic compounds

Ryo Miyazaki,[1,9] Claire Bertelli,[2,3] Paola Benaglio,[4] Jonas Canton,[5] Nicoló De Coi,[5] Walid H. Gharib,[6] Bebeka Gjoksi,[5] Alexander Goesmann,[7] Gilbert Greub,[2] Keith Harshman,[8] Burkhard Linke,[7] Josip Mikulic,[5] Linda Mueller,[5] Damien Nicolas,[5] Marc Robinson-Rechavi,[3,6] Carlo Rivolta,[4] Clémence Roggo,[5] Shantanu Roy,[9] Vladimir Sentchilo,[9] Alexandra Von Siebenthal,[5] Laurent Falquet[3] and Jan Roelof van der Meer[9]*

[1]*Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan.*
[2]*Center for Research on Intracellular Bacteria, Institute of Microbiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland.*
[3]*Swiss Institute for Bioinformatics (SIB), 1015 Lausanne, Switzerland.*
[4]*Departments of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland.*
Departments of [6]*Ecology and Evolution,* [9]*Fundamental Microbiology,* [5]*Master in Molecular Life Sciences,*
[8]*Lausanne Genomic Technologies Facility, Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland.*
[7]*Bioinformatics and Systems Biology, Justus-Liebig-University, 35392 Gießen, Germany.*

Summary

***Pseudomonas knackmussii* B13 was the first strain to be isolated in 1974 that could degrade chlorinated aromatic hydrocarbons. This discovery was the prologue for subsequent characterization of numerous bacterial metabolic pathways, for genetic and biochemical studies, and which spurred ideas for pollutant bioremediation. In this study, we determined the complete genome sequence of B13 using next generation sequencing technologies and optical mapping. Genome annotation indicated that B13 has a variety of metabolic pathways for degrading monoaromatic hydrocarbons including chlorobenzoate, aminophenol, anthranilate and hydroxyquinol, but not polyaromatic compounds. Comparative genome analysis revealed that B13 is closest to *Pseudomonas denitrificans* and *Pseudomonas aeruginosa*. The B13 genome contains at least eight genomic islands [prophages and integrative conjugative elements (ICEs)], which were absent in closely related pseudomonads. We confirm that two ICEs are identical copies of the 103 kb self-transmissible element ICE*clc* that carries the genes for chlorocatechol metabolism. Comparison of ICE*clc* showed that it is composed of a variable and a 'core' region, which is very conserved among proteobacterial genomes, suggesting a widely distributed family of so far uncharacterized ICE. Resequencing of two spontaneous B13 mutants revealed a number of single nucleotide substitutions, as well as excision of a large 220 kb region and a prophage that drastically change the host metabolic capacity and survivability.**

Received 23 March, 2014; accepted 28 April, 2014. *For correspondence. E-mail janroelof.vandermeer@unil.ch; Tel. (+41) 21 692 5630; Fax (+41) 21 692 5605.

Introduction

Pseudomonas species are typical opportunistic proteobacteria, which inhabit a wide range of environments. Several species have been characterized in depth for a variety of clinical and infectious properties, as well as for potential plant beneficial or environmentally useful characteristics. For example, *Pseudomonas aeruginosa* species are known to have useful properties in degradation of, e.g., long-chain alkanes (Marin *et al*., 2003) or in production of rhamnolipids (Soberon-Chavez *et al*., 2005), but they also represent the most common aetiology of nosocomial pneumonia and cause severe infections among patients with underlying comorbidity such as cystic fibrosis (CF) (Govan and Deretic, 1996). *Pseudomonas syringae* comprise severe plant pathogens secreting effector proteins into plant cells by the type III secretion system (Lindeberg *et al*., 2012), whereas several *Pseudomonas fluorescens* are known for their plant protection (biocontrol) properties (Haas and Defago, 2005). In contrast, *Pseudomonas putida* strains are mostly non-pathogenic. They show robust metabolic capacities, in some cases coupled to extreme solvent stress (Rojas *et al*., 2001); some interact with plant roots

(Segura and Ramos, 2013), and typically degrade a variety of aromatic hydrocarbons and/or chlorinated compounds (Wu *et al.*, 2011; De Lorenzo *et al.*, 2013). The genomes of a number of *Pseudomonas*-type strains have been deciphered, which has contributed to an improved understanding of the evolution and diversity of the *Pseudomonas* genus (Silby *et al.*, 2011). In particular, more and more individual *P. aeruginosa* and *P. syringae* isolates are being sequenced to follow clonal development and appearance of pathogenicity or virulence characteristics (Mathee *et al.*, 2008; Klockgether *et al.*, 2011; Marcelletti *et al.*, 2011).

Here we focus on the bacterium *P. knackmussii* strain B13, which was described in 1974 and was the first strain to grow on chlorinated aromatic compounds [3- and 4-chlorobenzoate (CBA)] as sole carbon and energy sources (Dorn *et al.*, 1974). At that time, this was an important discovery that spurred numerous further studies, focusing on the biochemical and genetic characterization of metabolic pathways for chlorinated compounds, on genetic adaptation mechanisms or on the possible application for cleanup of contaminated sites (Reineke and Knackmuss, 1979; 1988; Reineke, 1984). Strain B13 subsequently became famous for its use in what was called 'molecular breeding', a natural selection through gene transfer of mutants capable of degrading specific chloroaromatic compounds, which up to then were thought to be non-biodegradable (Oltmanns *et al.*, 1988; Mokross *et al.*, 1990; Ravatn *et al.*, 1998b). This was done in a process of mixing strain B13 with a potential recipient strain, with the idea that the capacity for metabolism of chlorinated catechols would be transferred from B13 into the recipient, leading to its inclusion in a new metabolic pathway. Much later on, it was discovered that the mobile DNA, which is indeed transferred spontaneously from strain B13 into other species is an integrative and conjugative element (ICE). This was first named the '*clc* element' and later ICE*clc* because of the presence of the genes for chlorocatechol degradation (Ravatn *et al.*, 1998a,b). ICE*clc* has a size of 103 kb and seemed to occur in two copies in the B13 genome, as judged from Southern hybridizations (Ravatn *et al.*, 1998a). The sequence of ICE*clc* itself has been determined and shows a composite nature of a region with the *clc* genes and a region with genes for 2-aminophenol degradation, plus a ~ 50 kb region with gene synteny to other known ICE (Gaillard *et al.*, 2006; Reinhard *et al.*, 2013). Subsequent studies could show that genes in- and outside this 50 kb region are important for its self-transfer (Gaillard *et al.*, 2006; Miyazaki and van der Meer, 2011; Miyazaki *et al.*, 2013). It was further demonstrated that ICE*clc* transfer is initiated from a small (1–3%) subpopulation of B13 cells that become transfer competent when they enter stationary phase (Reinhard *et al.*, 2013).

The goal of the present study was to fully sequence and annotate the genome of *P. knackmussii* B13, in order to better understand its specific phylogenetic position among pseudomonads, its potential catabolic properties and the nature of the two ICE*clc* elements. We also sequenced two spontaneous mutants of strain B13, in order to understand the genome stability of an environmental isolate. These mutants differed in their ability to degrade 3-CBA and appeared during selection of recombinants in which specific ICE*clc* genes were deleted (Minoia *et al.*, 2008). Sequencing, assembly and annotation were performed as part of a practical Master class in the Molecular Life Sciences programme of the University of Lausanne during 2010–2011.

Results and discussion

Overview of B13 genome sequence

The strategy for sequencing of *P. knackmussii* B13 consisted of *de novo* assembling a combination of 38 and 76 bp paired-end (PE) reads generated by Illumina GAII technology at an overall average coverage of 286-fold (1.76 Gb). We tested assembly qualities generated by four different assembly programmes, CLCBio, AbySS, Velvet and SOAPdenovo (Supporting Information Table S1), while specifically varying kmer-values. The resulting contigs were then organized by comparing to a separately generated optical map of the B13 genome. Velvet showed the best result in terms of assembly statistics (n50, lowest number of contigs, Supporting Information Table S1), but four large Velvet contigs matched the optical map of the B13 genome at two distinct positions, pointing to highly likely misassemblies (Supporting Information Fig. S1). In contrast, the SOAPdenovo assembly generated slightly more contigs (142) but with overall better fit to the optical map. Assembly was complicated by the presence of two identical copies of the ICE*clc* element (103 kb, see below), but could be resolved by re-analysis of the optical map and focusing on consistent differences in sequence coverage depth (Supporting Information Table S1). Finally, all gaps between SOAPdenovo contigs were closed by a combination of multiplex polymerase chain reaction (PCR) plus Sanger sequencing, and by whole genome PacBio sequencing, which produced longer reads (~ 5 kb) that could bridge large gaps. The final gapless complete sequence of the B13 genome consists of a 6 167 895 bp circular chromosome, containing 5 753 predicted coding sequences (CDSs) (Fig. 1). Strain B13 carries four copies of the genes for 16S and 23S rRNA. The two identical copies of ICE*clc* were found at the 3′ ends of two out of six genes for glycine-accepting tRNA ($tRNA^{Gly}$), interspaced by 240 kb. Cumulative GC-skew analysis of the genome showed the typical 'mountain'

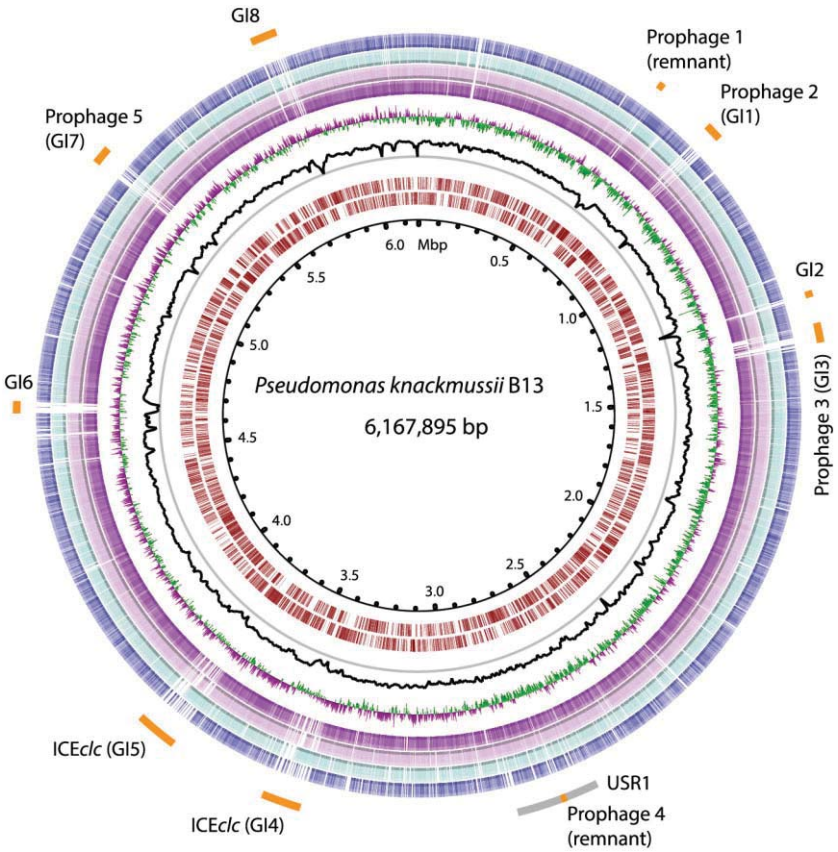

Fig. 1. Whole genome map of *P. knackmussii* B13 compared with related *Pseudomonas* species. From the inner to outer circles: genome coordinates, predicted coding sequences (CDS) of B13 on minus and plus strands (red), GC content plot (black) with a grey circle representing 50%, GC skew plot (green and magenta), orthologous genes found in *P. denitrificans* ATCC 13867 (purple), *P. aeruginosa* PAO1 (light purple), *P. stutzeri* DSM 4166 (light blue) and *P. putida* KT2440 (blue). Genomic island or phage-like regions (orange) and the USR1 (grey) in B13 are indicated on the outermost circle.

shape, but with two unequal replichores because of the presence of the two ICE*clc* copies on one side (Supporting Information Fig. S2).

The closest relative of B13 for which finished genomes are available was *P. denitrificans* ATCC 13867, with 99% identity at the level of 16S rRNA gene sequence (Fig. 2A). Phylogenetic analysis using a concatenated set of amino acid sequences from 20 conserved housekeeping proteins positioned B13 into a single clade with *P. denitrificans* ATCC 13867 and *P. aeruginosa* PAO1 (Fig. 2B), consistent with previous reports mentioning that B13 belongs to the *P. aeruginosa* subgroup (Stolz *et al.*, 2007). Despite this, the gene synteny across the whole genome between B13 and its closest known related *Pseudomonas* species was not very conserved, and numerous genome rearrangements, gene deletions or insertions could be observed (Fig. 3).

Metabolic pathways for aromatic compounds

Pseudomonas knackmussii B13 was the first bacterial strain discovered, which was capable of utilizing 3- and 4-CBA as sole carbon and energy sources under aerobic conditions (Dorn *et al.*, 1974). 3- and 4-CBA degradation is initiated by a multicomponent benzoate or toluate 1,2-dioxygenase (e.g. as characterized for BenABC from *Acinetobacter* sp. ADP1 or XylXYZ from pWW0 plasmid of *P. putida*) with relaxed specificity, leading to conversion of 3-CBA into 3- or 4-chloro-3,5-cyclohexadiene-1,2-diol-1-carboxylic acid. This is subsequently transformed into 3- and 4-chlorocatechol by a dihydrodiol dehydrogenase (e.g. BenD or XylL) (Harayama *et al.*, 1991). We found a single *xylXYZL* locus (PKB_2101-2104) on the B13 chromosome, with 72–85% identities across the whole length to the archetype XylXYZ from pWW0 (sp|P23099|XYLX_PSEPU) and 54–68% to the classical BenABCD system of *Acinetobacter* sp ADP1 (sp|P07769|BENA_ACIAD). The intermediates 3- and 4-chlorocatechol are subsequently degraded in a modified *ortho* cleavage pathway that is encoded by the two *clcABDE* loci, one on each of the ICE*clc* copies (PKB_3275-3279 and PKB_3624-3628) (Schwien *et al.*, 1988; Ravatn *et al.*, 1998a). This leads to β-ketoadipate, which, like in other pseudomonads, can be converted into two intermediates of the TCA cycle, succinyl-CoA and acetyl-CoA (Harwood and Parales, 1996). These reactions are catalysed by the CatFIJ enzymes (encoded by the loci PKB_2951-2953).

The *xylXYZL* locus is preceded by a putative transporter (*pcaK*-like) and a *xylS* homologue (PKB_2099) that

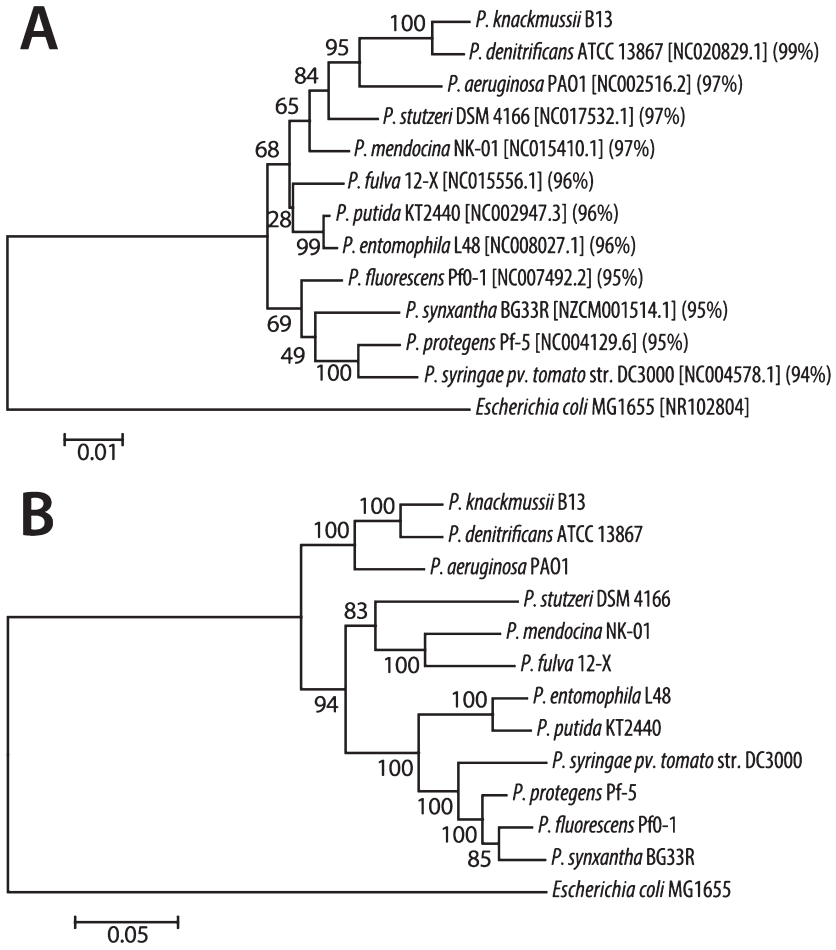

Fig. 2. Phylogenetic analysis of *P. knackmussii* B13 with other species of the *Pseudomonas* genus.
A. A 16S rRNA gene-based maximum-likelihood tree reconstructed by using the Tamura-Nei model. Percentages 16S rRNA gene nucleotide sequence identities between B13 and each of the *Pseudomonas* comparison species shown in parentheses. Accession numbers are given in brackets.
B. Maximum-likelihood tree based on the concatenated amino acid sequences of 20 housekeeping proteins, individually aligned, concatenated and then used as input in the Jones-Taylor-Thornton model. Values shown on tree branches indicate the percentage of trees among 1000 (A) and 100 (B) bootstrap replicates carrying that particular branching.

in analogy to other systems may control *xylXYZL* expression in response to (chloro)benzoate (Spooner *et al.*, 1986; Ramos *et al.*, 1990). We also found a small CDS between *xylS* and *xylXYZL*, PKB_2100, which may encode another transcriptional regulator predicted to contain an AraC-like ligand-binding domain (pfam02311). Downstream of the *xylXYZL* loci one finds genes for a catechol to muconolactone pathway (i.e. *catR*, *catB*, *catC*, *catA*) (McFall *et al.*, 1998).

Downstream of those and in opposite direction (*antCBA*, PKB_2111-2113) are three genes for anthranilate dioxygenase, which may lead to production of catechol from anthranilate (Bundy *et al.*, 1998). Fragments of another multicomponent anthranilate dioxygenase (PKB3281_3282) are also found on ICE*clc*. ICE*clc* further contains genes for a 2-aminophenol degradation pathway via *meta*-cleavage (*amn*) (Gaillard *et al.*, 2006). The locus PKB_1742-1747 is a homologue to the *dmpKLMNOP* encoded phenol multicomponent hydroxylase of *P. putida* CF600 (Shingler *et al.*, 1992), which is preceded by another *catA* gene and by a XylR/DmpR homologue, suggesting that B13 can degrade phenol via catechol to *cis-cis*-muconate and further via the *ortho*-cleavage pathway. B13 also codes for a predicted pathway of 4-OH-benzoate degradation via protocatechuate (PKB_2365-2371) (Romero-Steiner *et al.*, 1994), for hydroxyquinol via maleylacetate to 3-oxoadipate (PKB_3294-3295) (Daubaras *et al.*, 1996), and for a homogentisate pathway (PKB_4439-4444) (Arias-Barrau *et al.*, 2004). No specific homologous gene clusters to well-known operons for polyaromatic hydrocarbon degradation, such as biphenyl, naphthalene or phenanthrene, were found in the B13 genome, except for some genes for dioxygenase (PKB_2666-2667, 3453-3454), oxidoreductase (PKB_2668) and ferredoxin (PBK_2669), with unknown target substrates.

Genomic islands in the B13 genome

To identify unique genomic regions in B13, we compared its genome with those of related pseudomonads, such as *P. denitrificans* ATCC 13867, *P. aeruginosa* PAO1, *Pseudomonas stutzeri* DSM4166 and *P. putida* KT2440. We detected eight specific regions of genome plasticity in the B13 genome, designated here as genomic islands GI1–GI8, which were more than 20 kb in size, and absent

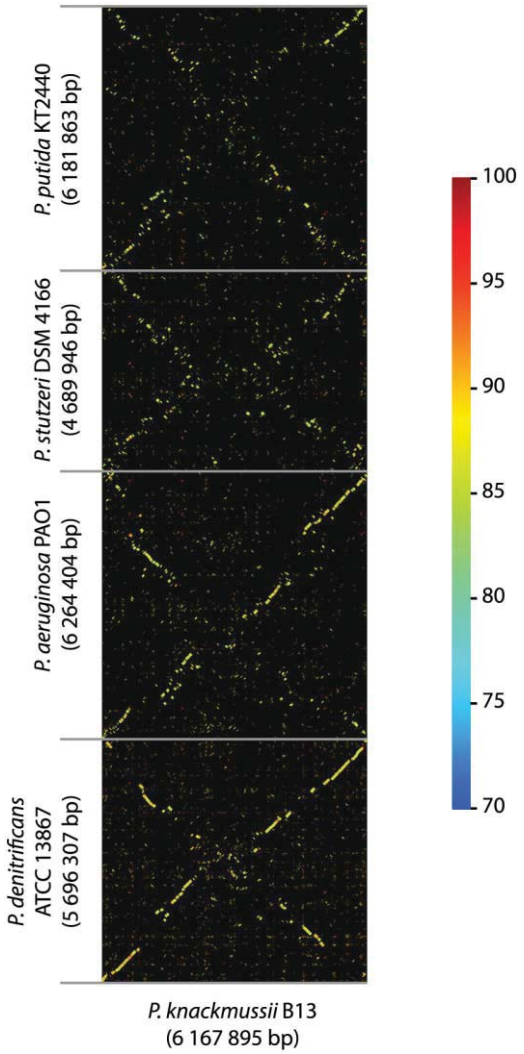

Fig. 3. Genomic synteny among B13 and other closely related *Pseudomonas* species. Plots show the 21 bp (coloured dots) sliding window comparison results in GenomeMatcher of the B13 genome sequence (horizontal axis) versus the four reference pseudomonads. Colours of the dots indicate the identity score in Blastn according to the scale on the right (a score of 70 corresponding to an e-value of 10^{-7}). Accession numbers: *P. denitrificans* ATCC 113867, CP004143; *P. aeruginosa* PAO1, AE004091; *P. stutzeri* DSM 4166, CP002622; *P. putida* KT2440, AE015451.

or different in the other *Pseudomonas* species (Fig. 1). The average GC content of GI regions was relatively low (61.5%) compared with the rest of the B13 genome (66.0%), suggesting their foreign origin and recent acquisition by a B13 ancestor through horizontal gene transfer (Fig. 1). Regions GI4 and GI5 correspond to the two copies of ICE*clc*, whereas bioinformatic analysis by PHAST (Zhou *et al*., 2011) predicted regions GI1, GI3 and GI7 to be prophages (prophage 2, 3 and 5, respectively, in Supporting Information Table S2). Most of the CDSs in prophage 2 were homologous to genes of bacteriophage B3, a Mu-like transposable phage found in *P. aeruginosa* (Braid *et al*., 2004). Conservation of gene synteny between bacteriophage B3 and the B13-prophage 2 region suggests that it is functionally intact. The prophage 3 region is 48 kb in size, with CDSs homologous to phiE125 from *Burkholderia thailandensis* (Woods *et al*., 2002). The typical phiE125 target attachment sites, consisting of the 49 bp 3′-end of the gene for proline-accepting tRNA, were not found flanking prophage 3 in B13. In addition, prophage 3 contained a unique 6 kb region with similar gene organization as the small mobile genetic element NR-II found in a pathogenicity island PAGI-5 of *P. aeruginosa* (Battle *et al*., 2008). NR-II carries its own integrase gene, as well as genes for potential virulence factors and for mercury resistance, and thus has been proposed to contribute to the virulence of its host (Battle *et al*., 2008). In contrast, the related region in prophage 3 only contained CDSs predicted to be involved in polyketide biosynthesis. Since the B13 6 kb NR-II-type region lies within prophage 3, it may have had an evolutionarily distinct origin and may have inserted into the phage-like region. B13-prophage 5 was over 42 kb and contained 56 CDSs, 38 of which were assigned phage functions. Most of them are homologous to genes from *Pseudomonas* phage F10 (Kwan *et al*., 2006), and are predicted to encode all essential components for an intact phage, such as integrase, tail protein, portal protein, protease and lysin. Furthermore, the prophage 5 region is flanked by unique 28 bp direct repeats, which may constitute the right and left attachment sites. Of note, a specific attachment site for phage F10 has not been reported. Further experimental evidence showed that prophage 5 can indeed excise from the B13 genome (see below). PHAST further predicted two other potential phage regions (prophage 1 and 4) in the B13 genome, which are also present in similar form in the other *Pseudomonas* species used for comparison. Both regions missed a number of genes thought to be essential for phage development, such as capsid protein or DNA recombinase. This suggests that the predicted B13-prophage 1 and 4 are remnants of ancient prophages, which were present in pseudomonads before the diversification of the four species examined here.

Two GI regions, GI6 (32 kb) and GI8 (56 kb), clearly did not contain phage-like genes but carried genes for integrases of the tyrosine recombinase family (PKB_4426 and PKB_5453 respectively). These were located downstream of genes for $tRNA^{Gly}$ (for GI6) and threonine-accepting tRNA ($tRNA^{Thr}$ for GI8), suggesting site-specific insertion. GI6 contained a few plasmid-related genes (i.e. PKB_4422 encoding a MobA relaxase protein and PKB_4406-4407 encoding toxin-antitoxin proteins), suggesting that it might be a co-mobilizable element that can use its own integrase and relaxase but depends on

conjugative transfer systems provided *in trans* by other mobile elements, such as ICE*clc*. Most of the other CDSs in GI6 encode hypothetical proteins. Several genes in GI8, on the other hand, encode proteins involved in carbohydrate metabolism (PKB_5461-5465), polysaccharide synthesis and polysaccharide transport (PKB_5474-5475, 5490-5499), whereas we also found a gene encoding a homologue of TrfA plasmid regulatory protein (PKB_5455). This suggests that GI8, like GI6, is a degenerate integrative mobile element that once may have been active and self-transferable.

Finally, GI2 had an estimated size of 22 kb and comprises 21 CDSs, none of which showed similarities to genes known from prophages or ICE. The gene PKB_1177, located at one end of GI2, is homologous to the ISPpu*10* transposase found in *P. putida* (Ramos-Gonzalez *et al*., 2006), suggesting that GI2 might have been inserted as a composite transposon. Interestingly, GI2 contains a large gene cluster (PKB_1181-1197) predicted to encode lipopolysaccharide synthesis and export.

*ICE*clc*-related chromosomal elements in other bacteria*

The B13 genome contains two perfectly identical copies of ICE*clc*, a mobile DNA element that had been previously applied to transfer the *clc* genes for chlorocatechol degradation to new recipient species (Oltmanns *et al*., 1988; Mokross *et al*., 1990; Ravatn *et al*., 1998a,b). Earlier comparisons had shown that ICE*clc* has a large (~ 50 kb) syntenic gene region to other ICE and genomic islands in other *Proteobacteria* (Gaillard *et al*., 2006; Miyazaki *et al*., 2013), which is important for its self-transfer (Miyazaki and van der Meer, 2011). Two ICE with almost 100% identical sequence and synteny conservation to ICE*clc* have so far been found in independently isolated *Betaproteobacteria*. This suggests ongoing and active distribution of ICE*clc*, and further microevolution by gene insertions. One of these is a 124 kb ICE*clc*^{LB400} in the polychlorinated biphenyl degrading bacterium *Burkholderia xenovorans* LB400 (Chain *et al*., 2006). The ICE*clc*^{LB400} element has acquired two insertions compared with ICE*clc*, one of which encodes the genes for *o*-halobenzoate degradation, and aids LB400 to enlarge the spectrum of metabolizable chlorinated compounds (Denef *et al*., 2006; Gaillard *et al*., 2006). The second ICE*clc*-like element was found in a *Ralstonia* sp. strain J705, which was isolated from a contaminated groundwater at Tyndall Air Force base (van der Meer *et al*., 1998). This ICE*clc*^{JS705} element has a 10 kb insertion of a gene cluster for a multicomponent dioxygenase and dihydrodiol dioxygenase, which enables strain JS705 to metabolize monochlorobenzene in addition to 3-chlorobenzoate (Müller *et al*., 2003). We compared the ICE*clc* sequence further to recently finished and draft genome sequences and found a large (> 100) number of genomes with gene regions carrying significant homology to the ICE*clc* 'core' region. Since putative ICE are difficult to distinguish *a priori* in a genome of interest, we searched manually in the vicinity of the regions homologous to the ICE*clc* core for major ICE hallmarks, such as a gene for an integrase, potential ICE insertion sites (e.g. gene for tRNA^{Gly} or tRNA^{Phe}) or attachment site duplication. A comparison of ICE*clc* to a small selection of suspected ICE regions from different host species is presented in Fig. 4 to illustrate the remarkable conservation of the core regions. In general, the integrase gene and core region are clearly separated by a variable region of between 20 and 60 kb with highly diverse gene content (Supporting Information Table S3). The core regions mostly retain gene synteny without much inversion or deletion, but with some smaller insertions or deletions (Fig. 4). Analysis of genes within the ICE*clc* core showed their importance for ICE self-transfer, such as the relaxase (Miyazaki and van der Meer, 2011) or distant homologues to VirB4, VirD4 and PilL (Gaillard *et al*., 2010). Interestingly, the alignment revealed a small gene cluster located in between the integrase gene and the core that was conserved among most of the ICE analysed here except for the putative ICE in *Nitrosomonas europaea* C91. Recent work suggested that this gene cluster may play a key regulatory role in ICE*clc* transfer (N. Pradervand, in revision).

Several hosts were found with two or three different ICE related to ICE*clc* within a single genome, such as *Bordetella petrii* (Gross *et al*., 2008), *Achromobacter xylosoxidans* NH44784-1996 (NC_021285.1) and *Acidovorax* sp. JS42 (CP000539.1). In two different and independently isolated species, namely *P. aeruginosa* CF18, a strain isolated from a CF patient in the United States (BioSample: SAMN02360661), and *A. xylosoxidans* NH44784-1996, isolated from a CF patient in Denmark (Jakobsen *et al*., 2013), an almost identical (> 99.9%) 91 kb-sized ICE was discovered (Fig. 4A). This ICE carries bleomycin resistance (Supporting Information Table S3) and its presence in two different clinically relevant species suggests that it is self-transmissible, and may play a role in chronic infections in CF patients. Another *A. xylosoxidans* strain A8, isolated from soil contaminated with polychlorinated biphenyls and able to utilize 2-chlorobenzoate and 2,5-dichlorobenzoate as carbon and energy sources (Pavlu *et al*., 1999), has an ICE*clc*-related element carrying a gene cluster for heavy metal resistance but no *clc* genes (Fig. 4B, Supporting Information Table S3). We also found other variants of ICE*clc* that may encode mercury resistance (putative ICE in *Bordetella bronchiseptica*

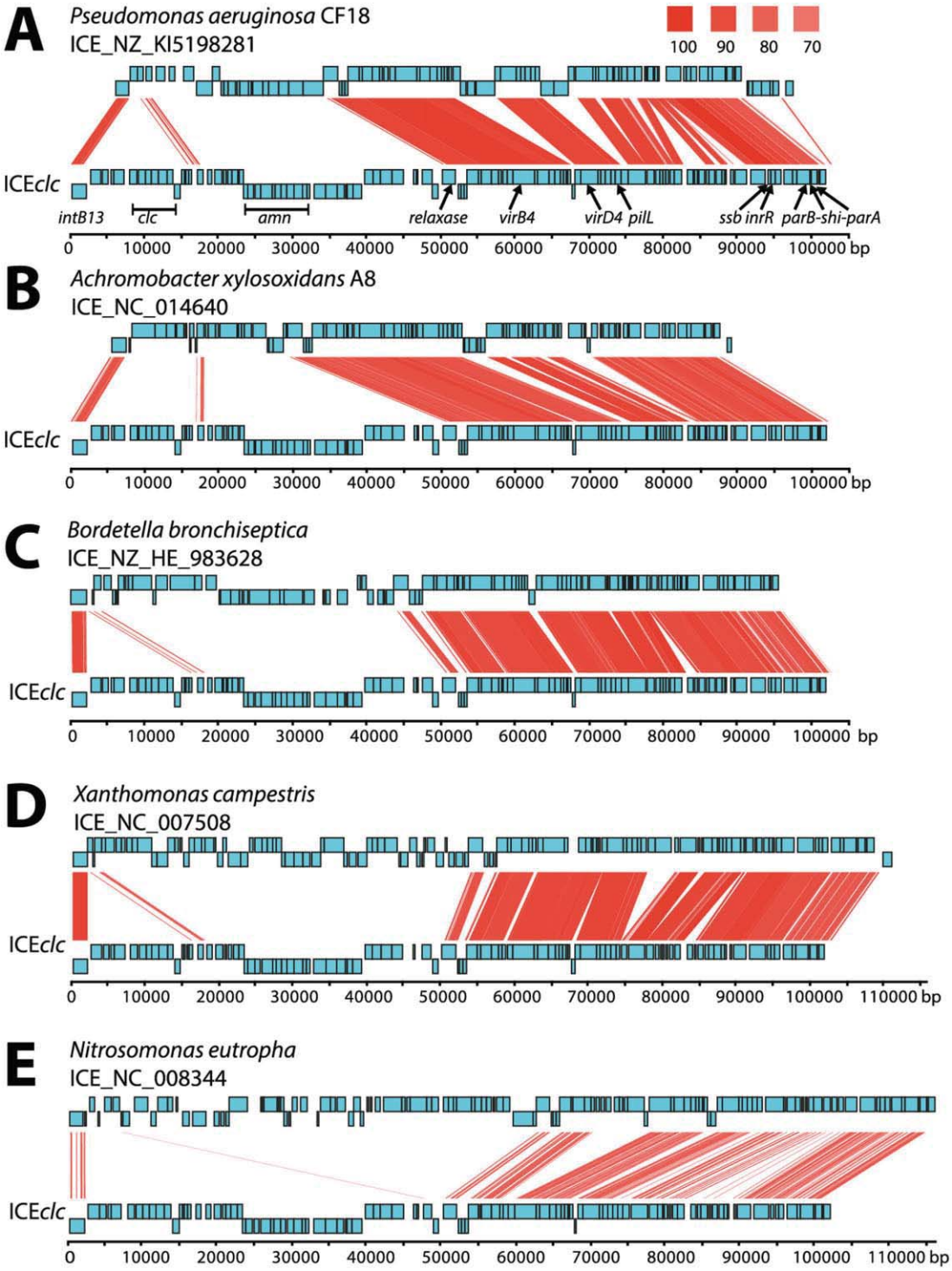

Fig. 4. Detailed comparisons of ICE*clc* with putative ICE regions in five other bacterial genomes.
A. Comparison to an ICE identified in the draft genome of *P. aeruginosa* CF18.
B. Comparison to an ICE identified in the draft genome of *A. xylosoxidans* A8.
C. Idem to *B. bronchiseptica*.
D. Idem to *X. campestris*.
E. Idem to *N. europaea*. Genes on top and bottom strands are represented by blue rectangles. Regions with nucleotide identity above 70% are connected by red windows using a colour intensity gradient based on identity scores of Blastn comparison in WebACT. Hallmark genes of ICE*clc* are indicated in panel A. Accession numbers are given behind each ICE-name. For a full list of predicted gene functions in the variable regions, see Supporting Information Table S3.

Bbr77), streptomycin resistance (ICE in *Xanthomonas campestris* pv. vesicatoria str. 85–10) and mercury plus copper resistance (ICE in *N. europaea* C91, Supporting Information Table S3). This small selection already shows the wide distribution of ICE*clc*-like elements within *Proteobacteria* and their divergence in terms of variable gene content, which may have provided selective benefit to their host.

Genome-wide comparison of B13 wild-type and spontaneous mutants

Previously, we showed that excision and transfer of ICE*clc* is a consequence of a bistable switch that culminates in the activation of the *intB13* gene promoter (P_{int}) in 3–5% of cells during stationary phase (Reinhard *et al.*, 2013). The expression level of *intB13* from P_{int} and transfer frequency of ICE*clc* are significantly reduced by the disruption of a protein named InrR, encoded in a small four-gene operon on ICE*clc* (Fig. 4A) (Minoia *et al.*, 2008). Strangely, a B13 mutant with a double *inrR* deletion (named B13-2061, one in each of the two copies of ICE*clc*) is not able to grow with 3-CBA and cultures develop a strong black colour as a result of photopolymerization of 3-chlorocatechol (Minoia *et al.*, 2008). Mutants from B13-2061 that grow on 3-CBA as sole carbon and energy source arise spontaneously after prolonged incubation of the culture flask (> 36 h) (Minoia *et al.*, 2008). One such spontaneous mutant (named B13-2201) displays growth rates on 3-CBA similar as B13 wild-type and no longer produces black colour (Minoia *et al.*, 2008). We assumed that B13-2201 contains mutations suppressing the growth defect caused by double *inrR* deletion. Resequencing and mapping of the B13-2201 reads on the B13 wild-type genome revealed two unique nucleotide substitutions in B13-2201. The first was an A-G transition at position 1 648 784 in the gene *rsmA*, resulting in a His-Arg substitution at residue 43 in the predicted RNA-binding protein RsmA (Table 1). The second mutation was a T-A transversion at position 456 462 in the gene *gshB*, resulting in a Leu-His substitution at residue 237 in the predicted glutathione synthetase GshB. To identify the possible moment of appearance of these mutations, we sequenced both genes in two predecessors in the B13-2201 ancestry. The first was strain B13-2061 itself, lacking both *inrR* copies but incapable of growth with 3-CBA (see above), and the second was B13-2058, a strain with a deletion of only one *inrR* copy (the other intact) and still able to grow on 3-CBA. Interestingly, no mutations were found in *rsmA*, whereas the transversion in *gshB* had already occurred in both strains. These results indicate that the mutation in *gshB per se* has no effect on 3-CBA metabolism. In contrast, it suggests that the double *inrR* deletion may indeed lead to a growth defect on 3-CBA, which might be suppressed by the mutation in *rsmA* (that had occurred only in B13-2201). It is known that RsmA/CsrA family proteins act as global post-transcriptional regulators by binding to target mRNAs, and thereby control several physiological processes in bacteria, including motility, pathogenicity and carbon metabolism (Heeb *et al.*, 2006). RsmA amino acid residues 42–44 constitute a highly conserved beta-sheet and R44 is essential for binding to target mRNAs (Heeb *et al.*, 2006). The H43R substitution in RsmA of B13-2201 might thus have changed its RNA-binding capacity or affinity, somehow leading to the restoration of 3-CBA metabolism. The exact mechanisms of interplay between InrR and RsmA in 3-CBA metabolism may be interesting to investigate further.

A second B13-ICE*clc* mutant strain was also resequenced and mapped to the B13 wild-type genome. This mutant (named B13-2811) has a double deletion of gene *parB* (Fig. 4A), which has been suspected of being involved in integrase expression control (Sentchilo *et al.*, 2003). During the course of producing the double *parB* deletion by allelic exchange from a single *parB* deletion mutant, we noticed that several different colony morpholo-

Table 1. Spontaneous mutations in derivatives of *P. knackmussii* B13.

	Gene[a]					
Strain	*inrR*	*parB*	*rsmA*	*gshB*	Prophage 5[b]	Growth on 3-CBA[c]
B13-2058	+/–	+/+	WT	L237H	+	+
B13-2061	–/–	+/+	WT	L237H	+	–
B13-2201	–/–	+/+	H43R	L237H	+	+
B13-4493	+/+	±	WT	WT	+	+
B13-4494	+/+	±	WT	WT	+	+
B13-2705	+/+	±	WT	WT	–	+
B13-2811	+/+	–/–	WT	WT	–	+

a. +/+, both copies present; ±, one copy interrupted; –/–, both copies interrupted.
b. +, present; –, absent.
c. +, growth; –, no growth.

gies appeared (not shown). This suggested that a complete *parB* disruption somehow influences growth of B13, and thus that B13-2811 may be a spontaneous suppressor mutant. Remapping of the B13-2811 genome surprisingly revealed that it completely lacked prophage 5, but kept a copy of the putative 28 bp attachment site (Table 1). These results suggested that the F10-like prophage 5 excised and disappeared during the process of *parB* deletion without killing the host. To get more clues for causality between *parB* deletion and phage excision, we investigated the presence of prophage 5 in strains B13-2705, -4493 and -4494, which are single *parB*-deleted mutants obtained from independent experiments. PCR analysis revealed that prophage 5 had disappeared in B13-2705 but was still present in the B13-4493 and -4494 genomes (Table 1). Indeed, the double *parB* deletion mutant B13-2811 had been generated from B13-2705, whereas double *parB* mutants had never been obtained from B13-4493 nor -4494. Interestingly, the *parB* gene could be easily knocked-out in *P. putida* UWC1, a rifampicin-resistance derivative of *P. putida* KT2440, carrying one copy of ICE*clc* but which does not carry the complete prophage 5 region in its genome (Fig. 1). These results suggest that complete *parB* deletion is possible only in cells that do not carry the prophage. ParB of ICE*clc* has a predicted centromere-binding ParB domain and a winged helix–turn–helix domain (Reinhard *et al*., 2013). The centromere-binding ParB protein is essential for proper distribution of duplicated chromosomes to dividing daughter cells, but also controls expression of several genes by functioning as a transcriptional regulator (Hayes and Barilla, 2006). We thus hypothesize that ParB$^{ICE\mathit{clc}}$ may directly or indirectly repress phage excision and that ParB$^{ICE\mathit{clc}}$ deficiency may lead to phage propagation and host killing.

Plasticity in the B13 genome

Apart from the unique mutations presented above, the alignment of the genome sequences of *P. knackmussii* B13 wild-type with both B13-2201 and B13-2811 mutants also revealed a large region of about 220 kb (from 2625116 to 2848501), which was present only in the wild-type but was absent in both mutants (Fig. 1). Intriguingly, both ends of this region [hereafter designated 'unstable region 1' (USR1)] carry an identical copy of an IS*110*-family transposase gene (PKB_2523 and PKB_2736). Both mutants lacked USR1 but kept one copy of the transposase gene, strongly suggesting that USR1 had looped out via homologous recombination between the two identical copies of the insertion sequence (IS). To investigate the instability of USR1 during cell cultivation, we assayed by PCR the presence of USR1 and of the junction region formed upon IS recombination in B13 wild-type grown with 3-CBA for up to 35 generations. At all time points tested, the USR1-specific region was present in the cultures, but the junction was also detected at a low level, indicating that some cells in the population spontaneously lose USR1 through recombination (Supporting Information Fig. S3).

The USR1 region contains a complete *paa* gene cluster (PKB_2682 to PKB_2697; *paaZEDCBAKJIHGFY*) predicted for phenylacetate degradation and its regulation (*paaX*), plus two unknown genes (PKB_2683 and PKB_2689). Since the *paa* cluster would permit metabolism of phenylacetate (Teufel *et al*., 2010), we hypothesized that B13-2201 and B13-2811 that lacked USR1 could not grow with phenylacetate as sole carbon and energy substrate. Yet both mutants and B13 wild-type grew equally well in minimal medium (MM) with 5 mM phenylacetate, suggesting that metabolism of phenylacetate may proceed via a pathway distinct from the *paa*-encoded phenylacetyl-CoA-mediated pathway. *Pseudomonas putida* strain U metabolizes phenylacetate via phenylacetyl-CoA as well as via 4-hydroxyphenylacetate by hydroxylation, and then converts the intermediate via homoprotocatechuate (Olivera *et al*., 1994). The B13 genome carries several hydroxylase genes belonging to the same family as phenylacetate hydroxylase, and also carries the *hpa* genes for homoprotocatechuate degradation (PKB_2453-2462), suggesting that the mutants might use this pathway for growth on phenylacetate.

The USR1 further contained genes encoding components for flagella synthesis (*fliDSKLA*, *motAB*, *fliC*, *flgABCDEFGHIJKL*), which represent only part of all genes known to be required for functional flagella (Fig. 5). These genes mainly encode the flagella motor protein and the flagella structural parts located outside of the cytoplasmic membrane, such as subunits of the proximal rod, the P and L rings, and proteins of the hook and flagellin (Pallen and Matzke, 2006). On the other hand, genes for internal flagellar proteins, such as the export apparatus of flagellin and the rotor-mounted switch complex (C ring), are missing from this region. The B13 genome, however, contains at least four other distinct loci with genes for flagellar components (PKB_1634-1697, PKB_2507-2522, PKB_3936-3947, and PKB_5267-5268), which can theoretically provide all essential components for functional flagella (Fig. 5). Thus, a loss of the gene clusters in USR1 may not to be critical for cellular motility, although this was not tested. All of the flagellar associated genes in USR1 were also found in the complete genomes of other pseudomonads, although their synteny was more conserved in *P. stutzeri* DSM 4166 than in *P. denitrificans* ATCC 13867 or *P. aeruginosa* PAO1 (Supporting Information Fig. S4), which are phylogenetically closer to B13 (Fig. 2).

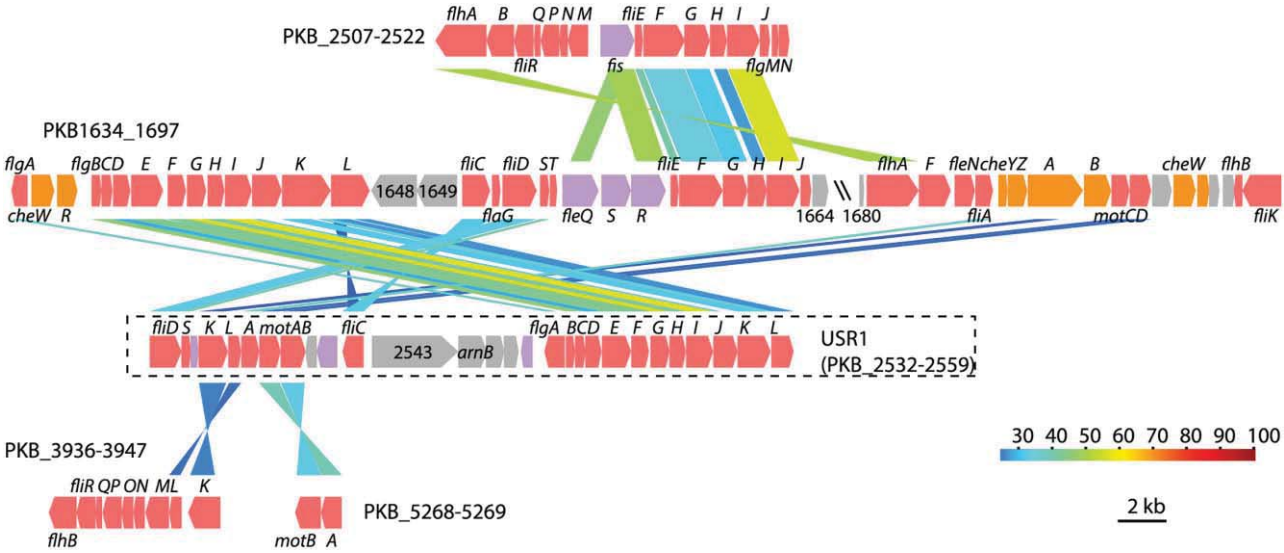

Fig. 5. The five gene regions in the B13 genome encoding flagella components. Regions named according to their gene number: PKB_1634-1697, PKB_2507-2522, PKB_2532-2559 (within USR1), PKB_3936-3947 and PKB_5267-5268. Genes for flagellar biosynthesis are indicated in red, for chemotaxis in orange, regulatory functions in purple and other functions in grey. Coloured windows show identity scores of individual Blastp comparisons in GenomeMatcher. Note how all flagellar associated genes in USR1 have paralogs in the other loci.

Conclusions

The genome sequence of *P. knackmussii* B13 provides an independent high-quality genome for the phylogeny and evolution of pseudomonads, illustrating a variety of aspects of genome flexibility and adaptability. Furthermore, 40 years after the description of strain B13 as the first bacterium capable of metabolizing chlorinated aromatic compounds, we can begin to appreciate the extent of distribution of the unique class of mobile elements that enabled the evolution of this metabolic pathway: ICE*clc*. Finally, our study demonstrates that high-quality genome sequence and analysis can be generated with dedicated master student classes.

Experimental procedures

Growth conditions and genomic DNA isolation

Pseudomonas knackmussii B13 and its derivatives were cultured in type 21C MM (Gerhardt *et al*., 1981) supplemented with either 3-CBA (5 mM), or phenylacetate (5 mM) as sole carbon and energy source. The genomic DNAs of B13, B13-2201 and B13-2811 were isolated from cultures at exponential phase in MM with 3-CBA using the method described previously (Sentchilo *et al*., 2009).

Sequencing

Pseudomonas knackmussii B13 wild-type and mutant DNA sequence libraries with a mean insert size of 0.5 kb were prepared according to the method of Aird *et al*. (2011) using Accuprime HF polymerase and betaine in the PCR amplification buffer. Each library was sequenced using 38- and 76-nucleotide PE protocols in a single lane of a flowcell using the Illumina GAII sequencing platform. The raw data were processed using v1.60 of the Illumina data processing pipeline and exported as fastq files. For the B13 wild-type genome, an additional 5 kb insert library was prepared and sequenced on the Pacific Biosciences RS on two SMRT cells according to methods of the manufacturer (Pacific Biosciences).

Genome assembly

Exported 38- and 76-nucleotide PE reads were quality controlled using fastqc (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Non-B13 contaminant reads were filtered out using an in-house Perl script. Where necessary, the reads were trimmed either on the 5′ or on the 3′-end or both to remove low-quality base callings.

Cleaned reads were assembled using either CLCBio (www.clcbio.com), AbySS (Simpson *et al*., 2009), Velvet (Zerbino and Birney, 2008) or SOAPdenovo (Li *et al*., 2010), using different kmer-settings (Supporting Information Table S1).

OpGen optical map

An optical map of B13 genomic DNA for the restriction enzyme KpnI was generated by OpGen (Maryland, USA). The map was used in the program MapSolver (OpGen) to correctly position the contigs generated by the four different assembly programs. We chose contigs generated by SOAPdenovo as the best quality assembly and scaffolded the contigs according to the map (Supporting Information Fig. S2).

Genome finishing

A draft B13 wild-type genome was assembled by combining SOAPdenovo scaffolds and PacBio reads, and further verified by multiplex PCR. Primers to be tested in multiplex PCR using B13 genomic DNA were selected from within 400 bp of both contig and scaffold ends by using Consed (Gordon *et al*., 1998), R (http://www.R-project.org) and an in-house script. Positive PCR products were purified and sequenced using standard Sanger technology, and sequences were used to close all gaps. The draft complete genome sequence was then verified by remapping all individual reads and removing final inconsistencies using PrInSeS (Massouras *et al*., 2010). The final gapless B13 genome sequence was submitted to the European Nucleotide Archive and is accessible under accession number HG322950.

Annotation

The *P. knackmussii* B13 genome sequence was automatically annotated using GenDB (Meyer *et al*., 2003). Functional annotation was derived from BLAST (Altschul *et al*., 1997) sequence similarity searches to Swiss-Prot/UniProtKB (UniProt, 2012) as well as RefSeq (Pruitt *et al*., 2012). In case no strong similarity was detected, BLAST results were manually analysed by comparison to the non-redundant NCBI database, to KEGG (Kanehisa *et al*., 2012) or to COG (Tatusov *et al*., 2000), as well as through hidden Markov model searches against the Pfam (Finn *et al*., 2008) and TIGRFAM databases (Haft *et al*., 2003). Manual annotations were focused in particular on potential GI regions, the USR1 region, on flagellar genes and on metabolic pathways for the degradation of aromatic compounds.

Phylogenetic analysis

The nucleotide sequences of 16S rRNA genes from 12 *Pseudomonas* species plus *Escherichia coli* MG1655 as an outgroup were retrieved from the NCBI database. The sequences were aligned with the program MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/), and a maximum-likelihood (ML) tree was reconstructed using MEGA 5.2.2 with the Tamura-Nei model and the Nearest-Neighbor-Interchange (NNI) move, further applying 1000 bootstrap replicates (Tamura *et al*., 2011). The amino acid sequences of 20 conserved housekeeping genes (Supporting Information Table S4), which are considered not to have been horizontally transferred (Ciccarelli *et al*., 2006), were selected from Microscope Genoscope (http://www.genoscope.cns.fr/agc/microscope/home/). Amino acid sequences were aligned with MUSCLE and concatenated with the help of an in-house script. A phylogenetic tree was reconstructed by using the ML method with Jones-Taylor-Thornton model, NNI move and 100 bootstrap replicates.

Sequence comparison

The B13 genomes were compared in silico to other pseudomonads by using the softwares BRIG (Alikhan *et al*., 2011), ARTEMIS (Rutherford *et al*., 2000) and GenomeMatcher (Ohtsubo *et al*., 2008). Potential ICE*clc* homologue regions in other bacterial genomes were detected by using MegaBlast (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome). Individual hits were then retrieved and manually searched for the nearby presence of ICE*clc* hallmark genes such as the integrase, a gene for $tRNA^{Gly}$ and the genes *parB-shi-parA-alpA*, which are located nearby the other end of ICE*clc* (Fig. 4A). Putative ICE regions were isolated in silico from the host genome and pair-wise compared with ICE*clc* using WebACT (http://www.webact.org/WebACT/home). Regions with nucleotide sequence similarity above 80% were exported and displayed on the local gene map using DNAPlotter (Carver *et al*., 2009). The final display was edited for clarity in Adobe Illustrator CS6. Nucleotide polymorphisms in the B13-2201 and B13-2811 genomes were detected by remapping their Illumina reads onto the B13 reference genome using SAMtools (Li *et al*., 2009), and visually verified by using IGV (Robinson *et al*., 2011). Large INDELs were detected by MAUVE (Darling *et al*., 2010) using contigs of assembled mutant reads.

PCR analysis

In order to analyse instability of the USR1 region in the B13 genome, B13 was cultured in MM with 5 mM of 3-CBA as a sole carbon and energy source for up to 35 generations. After approximately 7, 14, 21, 28 and 35 generations, cells were harvested from the culture, their DNA was released and used as templates in the PCR. The 1.6 kb left border of USR1 (between PKB_2522 and PKB_2524) and the 1.5 kb junction between PKB_2522 and PKB_2737 formed by the loss of USR1 was amplified using two primer sets PKB2522.d (TCTATTCGCCCGCCTACTG) plus PKB2524.u (CCTGG TCAGGCCAATAATGAG), and PKB2522.d plus PKB2737.u (GTTCAAGCGCGCCTGAAATC) respectively.

Acknowledgement

Computations were performed at the Vital-IT (http://www.vital-it.ch) centre for high-performance computing of the Swiss Institute of Bioinformatics. We thank Sandra Sulser and Nicolas Pradervand for their assistance in PCR gap closure.

References

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., *et al*. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12** (2)**:** R18.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., and Beatson, S.A. (2011) BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12:** 402.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Arias-Barrau, E., Olivera, E.R., Luengo, J.M., Fernandez, C., Galan, B., Garcia, J.L., *et al*. (2004) The homogentisate

pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. *J Bacteriol* **186:** 5062–5077.

Battle, S.E., Meyer, F., Rello, J., Kung, V.L., and Hauser, A.R. (2008) Hybrid pathogenicity island PAGI-5 contributes to the highly virulent phenotype of a *Pseudomonas aeruginosa* isolate in mammals. *J Bacteriol* **190:** 7130–7140.

Braid, M.D., Silhavy, J.L., Kitts, C.L., Cano, R.J., and Howe, M.M. (2004) Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. *J Bacteriol* **186:** 6560–6574.

Bundy, B.M., Campbell, A.L., and Neidle, E.L. (1998) Similarities between the *antABC*-encoded anthranilate dioxygenase and the *benABC*-encoded benzoate dioxygenase of *Acinetobacter* sp. strain ADP1. *J Bacteriol* **180:** 4466–4474.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25:** 119–120.

Chain, P.S., Denef, V.J., Konstantinidis, K.T., Vergez, L.M., Agullo, L., Reyes, V.L., *et al*. (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci USA* **103:** 15280–15287.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283–1287.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5:** e11147.

Daubaras, D.L., Saido, K., and Chakrabarty, A.M. (1996) Purification of hydroxyquinol 1,2-dioxygenase and maleylacetate reductase: the lower pathway of 2,4,5-trichlorophenoxyacetic acid metabolism by *Burkholderia cepacia* AC1100. *Appl Environ Microbiol* **62:** 4276–4279.

De Lorenzo, V., Pieper, D., and Ramos, J.L. (2013) From the test tube to the environment – and back. *Environ Microbiol* **15:** 6–11.

Denef, V.J., Klappenbach, J.A., Patrauchan, M.A., Florizone, C., Rodrigues, J.L., Tsoi, T.V., *et al*. (2006) Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400. *Appl Environ Microbiol* **72:** 585–595.

Dorn, E., Hellwig, M., Reineke, W., and Knackmuss, H.J. (1974) Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Arch Microbiol* **99:** 61–70.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., *et al*. (2008) The Pfam protein families database. *Nucleic Acids Res* **36:** D281–D288.

Gaillard, M., Vallaeys, T., Vorholter, F.J., Minoia, M., Werlen, C., Sentchilo, V., *et al*. (2006) The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. *J Bacteriol* **188:** 1999–2013.

Gaillard, M., Pradervand, N., Minoia, M., Sentchilo, V., Johnson, D.R., and van der Meer, J.R. (2010) Transcriptome analysis of the mobile genome ICEclc in *Pseudomonas knackmussii* B13. *BMC Microbiol* **10:** 153.

Gerhardt, P., Murray, R.G.E., Costilow, R.N., Nester, E.W., Wood, W.A., Kreig, N.R., and Briggs Phillips, G. (1981) *Manual of Methods for General Bacteriology*. Washington, DC: American Society for Microbiology.

Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8:** 195–202.

Govan, J.R., and Deretic, V. (1996) Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* **60:** 539–574.

Gross, R., Guzman, C.A., Sebaihia, M., dos Santos, V.A., Pieper, D.H., Koebnik, R., *et al*. (2008) The missing link: *Bordetella petrii* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. *BMC Genomics* **9:** 449.

Haas, D., and Defago, G. (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol* **3:** 307–319.

Haft, D.H., Selengut, J.D., and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Res* **31:** 371–373.

Harayama, S., Rekik, M., Bairoch, A., Neidle, E.L., and Ornston, L.N. (1991) Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal *benABC* and *Pseudomonas putida* TOL pWW0 plasmid *xylXYZ*, genes encoding benzoate dioxygenases. *J Bacteriol* **173:** 7540–7548.

Harwood, C.S., and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* **50:** 553–590.

Hayes, F., and Barilla, D. (2006) The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat Rev Microbiol* **4:** 133–143.

Heeb, S., Kuehne, S.A., Bycroft, M., Crivii, S., Allen, M.D., Haas, D., *et al*. (2006) Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site. *J Mol Biol* **355:** 1026–1036.

Jakobsen, T.H., Hansen, M.A., Jensen, P.O., Hansen, L., Riber, L., Cockburn, A., *et al*. (2013) Complete genome sequence of the cystic fibrosis pathogen *Achromobacter xylosoxidans* NH44784-1996 complies with important pathogenic phenotypes. *PLoS ONE* **8:** e68484.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40:** D109–D114.

Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C.F., and Tummler, B. (2011) *Pseudomonas aeruginosa* genomic structure and diversity. *Front Microbiol* **2:** 150.

Kwan, T., Liu, J., Dubow, M., Gros, P., and Pelletier, J. (2006) Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J Bacteriol* **188:** 1184–1187.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Li, Y., Hu, Y., Bolund, L., and Wang, J. (2010) State of the art de novo assembly of human genomes from massively parallel sequencing data. *Hum Genomics* **4:** 271–277.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012) *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol* **20:** 199–208.

McFall, S.M., Chugani, S.A., and Chakrabarty, A.M. (1998) Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene* **223:** 257–267.

Marcelletti, S., Ferrante, P., Petriccione, M., Firrao, G., and Scortichini, M. (2011) *Pseudomonas syringae* pv. *actinidiae* draft genomes comparison reveal strain-specific features involved in adaptation and virulence to Actinidia species. *PLoS ONE* **6:** e27297.

Marin, M.M., Yuste, L., and Rojo, F. (2003) Differential expression of the components of the two alkane hydroxylases from *Pseudomonas aeruginosa*. *J Bacteriol* **185:** 3232–3237.

Massouras, A., Hens, K., Gubelmann, C., Uplekar, S., Decouttere, F., Rougemont, J., *et al*. (2010) Primer-initiated sequence synthesis to detect and assemble structural variants. *Nat Methods* **7:** 485–486.

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J.M., Koehrsen, M., *et al*. (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc Natl Acad Sci USA* **105:** 3100–3105.

van der Meer, J.R., Werlen, C., Nishino, S.F., and Spain, J.C. (1998) Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. *Appl Environ Microbiol* **64:** 4185–4193.

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., *et al*. (2003) GenDB – an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187–2195.

Minoia, M., Gaillard, M., Reinhard, F., Stojanov, M., Sentchilo, V., and van der Meer, J.R. (2008) Stochasticity and bistability in horizontal transfer control of a genomic island in *Pseudomonas*. *Proc Natl Acad Sci USA* **105:** 20792–20797.

Miyazaki, R., and van der Meer, J.R. (2011) A dual functional origin of transfer in the ICEclc genomic island of *Pseudomonas knackmussii* B13. *Mol Microbiol* **79:** 743–758.

Miyazaki, R., Minoia, M., Pradervand, N., Sentchilo, V., Sulser, S., Reinhard, F., and van der Meer, J.R. (2013) The clc element and related genomic islands in Proteobacteria. In *Bacterial Integrative Mobile Genetic Elements*. Roberts, A.P., and Mullany, P. (eds). Austin, TX, USA: Landes Bioscience, pp. 261–272.

Mokross, H., Schmidt, E., and Reineke, W. (1990) Degradation of 3-chlorobiphenyl by in vivo constructed hybrid pseudomonads. *FEMS Microbiol Lett* **59:** 179–185.

Müller, T.A., Werlen, C., Spain, J., and van der Meer, J.R. (2003) Evolution of a chlorobenzene degradative pathway among bacteria in a contaminated groundwater mediated by a genomic island in *Ralstonia*. *Environ Microbiol* **5:** 163–173.

Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y., and Tsuda, M. (2008) GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinformatics* **9:** 376.

Olivera, E.R., Reglero, A., Martinez-Blanco, H., Fernandez-Medarde, A., Moreno, M.A., and Luengo, J.M. (1994) Catabolism of aromatics in *Pseudomonas putida* U. Formal evidence that phenylacetic acid and 4-hydroxyphenylacetic acid are catabolized by two unrelated pathways. *Eur J Biochem* **221:** 375–381.

Oltmanns, R.H., Rast, H.G., and Reineke, W. (1988) Degradation of 1,4-dichlorobenzene by constructed and enriched strains. *Appl Microbiol Biotechnol* **28:** 609–616.

Pallen, M.J., and Matzke, N.J. (2006) From the origin of species to the origin of bacterial flagella. *Nat Rev Microbiol* **4:** 784–790.

Pavlu, L., Vosahlova, J., Klierova, H., Prouza, M., Demnerova, K., and Brenner, V. (1999) Characterization of chlorobenzoate degraders isolated from polychlorinated biphenyl-contaminated soil and sediment in the Czech Republic. *J Appl Microbiol* **87:** 381–386.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012) NCBI Reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40:** D130–D135.

Ramos, J.L., Rojo, F., Zhou, L., and Timmis, K.N. (1990) A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. *Nucleic Acids Res* **18:** 2149–2152.

Ramos-Gonzalez, M.I., Campos, M.J., Ramos, J.L., and Espinosa-Urgel, M. (2006) Characterization of the *Pseudomonas putida* mobile genetic element ISPpu10: an occupant of repetitive extragenic palindromic sequences. *J Bacteriol* **188:** 37–44.

Ravatn, R., Studer, S., Springael, D., Zehnder, A.J., and van der Meer, J.R. (1998a) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J Bacteriol* **180:** 4360–4369.

Ravatn, R., Zehnder, A.J., and van der Meer, J.R. (1998b) Low-frequency horizontal transfer of an element containing the chlorocatechol degradation genes from *Pseudomonas* sp. strain B13 to *Pseudomonas putida* F1 and to indigenous bacteria in laboratory-scale activated-sludge microcosms. *Appl Environ Microbiol* **64:** 2126–2132.

Reineke, W. (1984) Microbial degradation of halogenated aromatic compounds. In *Microbial Degradation of Organic Compounds*. Gibson, D.T. (ed.). New York: Marcel Dekker, pp. 319–360.

Reineke, W., and Knackmuss, H.-J. (1979) Construction of haloaromatics utilizing bacteria. *Nature (London)* **277:** 385–386.

Reineke, W., and Knackmuss, H.-J. (1988) Microbial degradation of haloaromatics. *Annu Rev Microbiol* **42:** 263–287.

Reinhard, F., Miyazaki, R., Pradervand, N., and van der Meer, J.R. (2013) Cell differentiation to 'mating bodies' induced by an integrating and conjugative element in free-living bacteria. *Curr Biol* **23:** 255–259.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26.

Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J.L., and Segura, A. (2001) Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **183:** 3967–3973.

Romero-Steiner, S., Parales, R.E., Harwood, C.S., and Houghton, J.E. (1994) Characterization of the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of p-hydroxybenzoate. *J Bacteriol* **176:** 5771–5779.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

Schwien, U., Schmidt, E., Knackmuss, H.J., and Reineke, W. (1988) Degradation of chlorosubstituted aromatic compounds by *Pseudomonas* sp. strain B13: fate of 3,5-dichlorocatechol. *Arch Microbiol* **150:** 78–84.

Segura, A., and Ramos, J.L. (2013) Plant-bacteria interactions in the removal of pollutants. *Curr Opin Biotechnol* **24:** 467–473.

Sentchilo, V., Zehnder, A.J., and van der Meer, J.R. (2003) Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. *Mol Microbiol* **49:** 93–104.

Sentchilo, V., Czechowska, K., Pradervand, N., Minoia, M., Miyazaki, R., and van der Meer, J.R. (2009) Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol Microbiol* **72:** 1293–1306.

Shingler, V., Powlowski, J., and Marklund, U. (1992) Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *J Bacteriol* **174:** 711–724.

Silby, M.W., Winstanley, C., Godfrey, S.A., Levy, S.B., and Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* **35:** 652–680.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* **19:** 1117–1123.

Soberon-Chavez, G., Lepine, F., and Deziel, E. (2005) Production of rhamnolipids by *Pseudomonas aeruginosa*. *Appl Microbiol Biotechnol* **68:** 718–725.

Spooner, R.A., Lindsay, K., and Franklin, F.C. (1986) Genetic, functional and sequence analysis of the *xylR* and *xylS* regulatory genes of the TOL plasmid pWW0. *J Gen Microbiol* **132:** 1347–1358.

Stolz, A., Busse, H.J., and Kampfer, P. (2007) Pseudomonas knackmussii *sp. nov*. *Int J Syst Evol Microbiol* **57:** 572–576.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28:** 33–36.

Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., *et al*. (2010) Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci USA* **107:** 14390–14395.

UniProt, C. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **40:** D71–D75.

Woods, D.E., Jeddeloh, J.A., Fritz, D.L., and DeShazer, D. (2002) *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for Burkholderia mallei. *J Bacteriol* **184:** 4003–4017.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2011) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* **35:** 299–323.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821–829.

Zhou, Y., Liang, Y., Lynch, K., Dennis, J.J., and Wishart, D.S. (2011) PHAST: A fast phage search tool. *Nucl Acids Res* **39:** 347–352.

Supporting information

Additional Supporting Information may be found in the online version of this article

Fig. S1. Verification of the de novo assembly of the B13 genome by optical mapping. Contigs generated by SOAPdenovo (kmer = 43) or Velvet (kmer = 35) were aligned on the optical map (prepared for KpnI). Three misassembled contigs by Velvet are indicated by black arrowheads. The final B13 genome was verified for the same restriction enzyme with the optical map. Note how the two ICE*clc* copies are collapsed in the original map (red), but appear in duplicate on the final corrected optical map.

Fig. S2. Cumulative GC-skew plot of the B13 genome. The position +1 corresponds to the predicted origin of replication (*oriC*), whereas the predicted termination site of replication (*ter*) is indicated by an arrow. Note how the right replichore is longer than the left because of the presence of two ICE*clc* copies.

Fig. S3. PCR analysis of instability of USR1 in B13. DNA from B13 cultures in MM with 5 mM of 3-CBA after 7, 14, 21, 28 and 35 generations was used in the PCR. Pictures show amplicons of the 1.6 kb left border of USR1 (between PKB_2522 and PKB_2524) and of the 1.5 kb junction between PKB_2522 and PKB_2737 formed by the loss of USR1. B13-4492 and B13-4493 are two mutants with single *parB* deletions. N, *P. putida* KT2440 culture used as negative control; M, Mass-ruler DNAladder (Fermentas). Note how all cultures except *P. putida* amplify both the left border (in high amounts) and the junction (in lower amounts), indicating that a small proportion of cells in culture lost USR1 through recombination.

Fig. S4. Comparisons of flagellar associated genes in USR1 of B13 with orthologous genes in genomes of other Pseudomonads. Maps show relevant genome regions (coordinates displayed) with location of predicted genes as open or coloured pentagons (indicating the direction of the gene). Coloured bars linking genes show the percentage amino acid similarity calculated by Blastp comparison in GenomeMatcher, according to the colour scale. Colours of genes in B13 region USR1 refer to gene functions mentioned in Fig. 5.

Table S1. Assembly statistics of the genomes of *Pseudomonas knackmussii* B13 and its mutants.

Table S2. In silico prediction of potential prophage regions in the B13 genome by PHAST.

Table S3. Predicted gene functions of ICE similar to ICE*clc* in five Proteobacterial genomes.

Table S4. Housekeeping genes used in the phylogenetic analysis.