Comparative genomics unravels mechanisms of genetic adaptation for the catabolism of the phenylurea herbicide linuron in *Variovorax*

Başak Öztürk#[1, 2] Johannes Werner[3], Jan P. Meier-Kolthoff[4], Boyke Bunk[4], Cathrin Spröer[5], Dirk Springael#[2]

1 Junior Research Group Microbial Biotechnology, Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

2 Division of Soil and Water Management, KU Leuven, Leuven, Belgium

3 Department of Biological Oceanography, Leibniz Institute for Baltic Sea Research, Rostock, Germany

4 Bioinformatics Department, Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

5 Central Services, Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

Running Head: Comparison of linuron-degrading *Variovorax*

#Address correspondence to Başak Öztürk: basak.oeztuerk@dsmz.de, Dirk Springael: dirk.springael@kuleuven.be.

Word counts:

Abstract: 220

Importance: 136

Main text: 5350

Abstract

Biodegradation of the phenylurea herbicide linuron appears a specialization within a specific clade of the *Variovorax* genus. The linuron catabolic ability is likely acquired by horizontal gene transfer but the mechanisms involved are not known. The full genome sequences of six linuron degrading *Variovorax* strains isolated from geographically distant locations were analyzed to acquire insight in the mechanisms of genetic adaptation towards linuron metabolism in *Variovorax*. Whole genome sequence analysis confirmed the phylogenetic position of the linuron degraders in a separate clade within *Variovorax* and indicated their unlikely origin from a common ancestral linuron degrader. The linuron degraders differentiated from non-degraders by the presence of multiple plasmids of 20 to 839 kb, including plasmids of unknown plasmid groups. The linuron catabolic gene clusters showed (i) high conservation and synteny and (ii) strain-dependent distribution among the different plasmids. All were bordered by IS*1071* elements forming composite transposon structures appointing IS*1071* as key for catabolic gene recruitment. Most of the strain carried at least one broad host range plasmid that might have been a second instrument for catabolic gene acquisition. We conclude that clade 1*Variovorax* strains, despite their different geographical origin, made use of a limited genetic repertoire to acquire linuron biodegradation.

Importance

The genus *Variovorax* and especially a clade of strains that phylogenetically separates from the majority of *Variovorax* species, appears to be a specialist in the biodegradation of the phenyl urea herbicide linuron. Horizontal gene transfer (HGT) likely played an essential role in the genetic adaptation of those strain to acquire the linuron catabolic genotype. However, we do not know the genetic repertoire involved in this adaptation both regarding catabolic gene functions as well as gene functions that promote HGT neither do we know

how this varies between the different strains. These questions are addressed in this paper by analyzing the full genome sequences of six linuron degrading *Variovorax* strains. This knowledge is important for understanding the mechanisms that steer world-wide genetic adaptation in a particular species and this for a particular phenotypic trait as linuron biodegradation.

Introduction

Linuron [3-(3,4-dichlorophenyl)-1-methoxy-1-methyl urea] is a phenylurea herbicide that has been widely used for weed control in agriculture. Biodegradation is the major route of linuron dissipation in the environment(1). Bacteria belonging to the genus *Variovorax* were isolated from geographically-distant locations either as single strains (2–4) or as members of consortia (4, 5) that have the ability to mineralize and utilize the herbicide for growth. Single strains convert linuron to CO_2 and cell material while in consortia, *Variovorax* perform particularly the initial hydrolysis of linuron into the primary metabolite 3,4-dichloroaniline (DCA). The metabolic pathway of linuron degradation in *Variovorax* sp. WDL1 and SRS16 are well studied. The linuron hydrolases HylA (identified in WDL1) (6) and LibA (identified in SRS16) (1) perform the hydrolysis of linuron into DCA and *N,O*-dimethylhydroxylamine (*N,O*-DMHA) (1, 6). In both strains, a multicomponent chloroaniline dioxygenase $DcaQTA_1A_2B$ converts DCA to 4,5-dichlorocatechol while chlorocatechol is further metabolized to oxo-adipate by enzymes encoded by the *ccdCFDE* gene cluster(7). PCR analysis has shown that other linuron-degrading *Variovorax* share the same catabolic genes. Interestingly, based on 16S rRNA gene phylogeny, the linuron degrading *Variovorax* strains appear to belong to a clade of *Variovorax* strains that separates from the main bulk of strains, including most of the type strains (4). The ability to degrade and/or grow on linuron is unique for those strains within the *Variovorax* genus, indicating that they must have genetically adapted by acquiring the catabolic genes by horizontal gene transfer (HGT). This is supported by the observation that in strains SRS16 and WDL1, the catabolic genes are physically-linked with mobile genetic elements (MGE). In SRS16, the DCA catabolic genes are bordered by multiple insertion sequence (IS) elements[2]. The same applies to *hylA*, the *dca* cluster and the *ccd* cluster in strain WDL1[9]. Moreover, the three catabolic gene clusters in *Variovorax* sp. WDL1 reside on a large extra-chromosomal element that shows several plasmid features including gene functions for conjugation (8).

However, how the constellation and the genetic context of the catabolic genes and their linkage with MGEs varies between different linuron-degrading *Variovorax* strains and how this relates to the geographic origin of the strains is yet unknown. Such knowledge will provide insight in the mechanisms that govern the functional evolution of genomes and especially those of organic xenobiotic degraders and more specifically of the genus *Variovorax*. This organism inhabits a wide variety of environments suggesting that it is prone to adaptation to new environmental constraints. To this end, we sequenced the complete genomes of six different linuron-degrading *Variovorax* strains isolated from distantly located geographical areas. We (i) re-analyzed the phylogenetic relationship between the strains and their phylogenetic position within the *Variovorax* genus, (ii) examined how their genomes differ with those of non-linuron degrading *Variovorax* strains emphasizing on the occurrence and types of MGEs and (iii) and compared the genetic constellation and context of the gene clusters involved in linuron metabolism to reveal how these traits were acquired among different strains.

Results and discussion

General genome features of linuron-degrading *Variovorax* strains

The full genome sequences of six linuron-degrading *Variovorax* sp. strains, i.e., WDL1 (5), SRS16 (2), PBL-H6, PBL-E5, PBS-H4 (4) and RA8 (3) were obtained. Their general genomic features are listed in Table 1. Strain WDL1 was recently found to consist of two subpopulations that only deviate in the presence of linuron or DCA degradation genes (9). In this study, the genome of one of these two subpopulations, i.e., the one carrying the *hylA* gene cluster was re-sequenced. The new sequence deviated slightly from the one reported by Albers et al. (9) The 5400 kb and 1240 kb replicons reported in (9) formed one chromosome of 6.7 Mbp while the 1380 kbp plasmid-like extrachromosomal replicon consisted of two replicons, i.e., pWDL1-1 (800 kbp) carrying linuron catabolic genes and

pWDL1-2 (540 kbp). Chromosome sizes (ranging from 5.99 to 8.36 Mbp) and GC content (ranging from 66.24 to 66.86%) of the linuron-degrading strains were comparable to those of other *Variovorax* genomes. Sizes of other reported non-linuron degrading *Variovorax* genomes range between 4.31 to 9.24 Mbp (median: 7.2 Mbp) with GC contents of 64.6 to 69.6 (median: 67.4) (Table S1). All linuron- degrading *Variovorax* strains contained a relatively high number of extra-chromosomal elements (two to six), including smaller obvious plasmid replicons (20 to 70 kbp) but also larger replicons of more than 500 kbp). The GC content and codon usage of most of those larger extra-chromosomal elements substantially differed from those of the chromosome (Figure S1). They also did not contain any essential genes for cell viability, categorizing them rather as plasmids than as a second chromosome or chromid (10, 11). pPBL-E5-2 and pSRS16-3 showed GC-contents similar to those of the chromosome but did not contain genes for cell viability and carried plasmid-like replication modules, suggesting they are also plasmids. In contrast to the linuron-degrading *Variovorax* genomes, none of the non-degrading ones *Variovorax* strains for which genome sequences are available, carried plasmids. IncP-1 plasmids were though reported in three not phylogenetically-classified *Variovorax* isolates with unknown genome sequences. pHB44(12) and pBS64(12) were identified in *Variovorax* strains associated with the mycorrhizal fungus *Laccaria proxima,* and carry genes that increase the *Variovorax* host fitness by enabling metal ion transport and bacitracin resistance(13). pDB1(14) in *Variovorax sp.* DB1, carries genes for the biodegradation of the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D). Another feature that distinguishes the linuron-degrading strains from the non-degraders is the occurrence of a high number of IS*1071* elements varying from four to seven copies in the degraders, while non-degraders did not carry any IS*1071*. IS*1071* is an insertion element that was first described bordering the 3-chlorobenzoate catabolic genes of *Pseudomonas* sp. BRC60 plasmid pBRC60(15). Since then it has been frequently associated with primarily catabolic genes in various organisms, especially β-proteobacteria (16). It often flanks the catabolic

genes at both sites, forming a putative composite transposon which has been shown to translocate as a whole (15). The element has been suggested to play a primarily role in the acquisition and subsequent distribution of adaptive genes and especially catabolic functions in bacteria (16–18).

Phylogenetic analysis of linuron-degrading *Variovorax* strains

The phylogenetic relatedness between the linuron-degrading *Variovorax* strains was determined by digital DNA:DNA hybridization values (dDDH) (Table S2). With the exception of PBL-E5 and SRS16, which represent the same species (dDDH value of 86%), all linuron-degrading *Variovorax* were designated as distinct species, and none of them belonged to any type species. Their phylogenetic divergence strongly indicates that the five degraders acquired linuron degradation genes independently as opposed to being derived from one common ancestral linuron degrader. Whole genome-based phylogeny (showed that the *Variovorax* species sequenced to date separated into two clades (clade 1 and clade 2), and that the linuron-degraders are very closely related species, all belonging to clade 1. The tree topology remained the same when only the linuron-degrader chromosomes were used (Figure S2). This separation largely replicated the 16S rRNA gene sequence-based phylogeny (Figure S3) with the exception of *V. soli* and *V. sp.* OV329. The low dDDH values of the *V. soli* genome with the degrader genomes (24.7-25.7%) however confirmed that these are distantly related. In addition to the linuron-degrading *Variovorax* strains, clade 1 included various other isolates but no type species. From those, a closed genome sequence was only available for the lignin-degrading soil isolate strain HW608 (19). Clustering of the linuron degrading strains was independent of either the geographical origin, the capacity to degrade linuron completely or partially to 3, 4-DCA, or the presence of specific catabolic genes involved in linuron biodegradation. Clade 2 contained the majority of the *Variovorax* strains, including the species *V. boronicumulans, V. soli, V. paradoxus, V. gossypi* and *V. guangxiensis*. Interestingly, in contrast to clade 2, non-linuron degrading strains from clade

1 were often associated with the catabolism of natural and anthropogenic organic compounds such as *Variovorax* sp. WS11 (an isoprene-degrading phyllosphere isolate (20)), KK3 (a 2,4-D-degrading freshwater isolate (21)), and JJ 1663 (an *N*-nitroglycine-degrading activated sludge isolate (22)). As such, including the linuron degraders, eight of the fourteen clade 1 strains were degraders of anthropogenic compounds. Clade 2, however, included only one xenobiotic-degrading isolate (one in 55 strains), i.e., *V. boronicumulans* J1(23) that degrades the neonicotinoid thiacloprid but only co-metabolically (24). These results indicate that the linuron-degrading strains belong to a *Variovorax* clade or originates from common ancestor that is/was more prone to genetic adaptation and hence specialization towards the biodegradation of anthropogenic compounds. The clade separation of strains with and without biodegradation capacity has not been observed before in other genera (25, 26).

Plasmids hosted by linuron-degrading *Variovorax sp.*

Phylogenetic analysis of the entire plasmid sequences clustered some of the plasmids identified in the linuron-degrading *Variovorax* sp. strains, into well-known but also novel plasmid groups (Figure 2). Some of the plasmids occur in multiple strains in which they are highly conserved (Figure S2). Table S3 shows overview of the replication and conjugation systems encoded on these plasmids. Twelve plasmids have type IV secretion system genes (T4SS) which facilitate conjugative transfer, but the origin of transfer (*oriT*) could not always be determined.

Known conjugative plasmids and their role in linuron degradation

Plasmids pPBL-E5-3, pRA8-3, and pSRS16-4 were classified as IncP-1δ plasmids. The three plasmids were 99% identical at the nucleotide (nt) level over the entire plasmid sequence and carry the *libA* locus between *trfA* and *oriV*, a known insertion hot spot for accessory genes in IncP-1 plasmids (16). Catabolic IncP-1δ plasmids have been reported before either isolated by means of exogenous isolation (27) or from isolates[17,](29) all carrying 2,4-D

185 degradation genes, between *trfA* and *oriV*. The above mentioned *Variovorax* plasmids
186 pDB1(14), pHB44(12) and pBS64(12) are also IncP-1 plasmids, but belong to the IncP-1 β
187 group.

188 Instead of IncP-1 plasmids, PBL-H6 and PBS-H4 carried self-transmissable broad host range
189 PromA plasmids. pPBS-H4-2 carries *hylA* and the *ccd* gene cluster while pPBL-H6-2 carries
190 an isolated IS*1071* transposase copy with inverted repeats (IR). pPBL-H6-2 and pPBS-H4-2
191 are most closely related to the PromA γ plasmids pSN1104-11 and pSN1104-34 (30) isolated
192 by exogenous isolation, and hence pPBS-H4-2 and pPBL-H6-2 represent the first PromA γ
193 plasmids obtained from isolates. Their presence in *Variovorax* extends the host range of
194 PromA γ plasmids, as for PromA α and PromA β plasmids, to β-proteobacteria. pPBS-H4-2 is
195 the first catabolic PromA plasmid, and is one of the few non-cryptic PromA plasmids(31).
196 The often cryptic character of PromA plasmids has been the subject of a debate since it
197 might harness their stability as they do not benefit the host fitness. It was suggested that
198 they mainly support the conjugative transfer of other mobilizable replicons (31). The finding
199 that PromA plasmids can carry catabolic genes shows that they, as is the case for IncP-1
200 plasmids, can indeed acquire and distribute genes beneficial for the host. The location of
201 cargo genes in both pPBS-H4-2 and pPBL-H6-2 (near *virD2*) differs from this in other PromA
202 plasmids (near *parA*)(32) and identifies the *virD* locus as an alternative hot spot for
203 insertion of accessory genes in PromA plasmids.

204 **Novel putatively conjugative plasmids in linuron-degrading Variovorax strains**

205 Other plasmids than IncP1 and PromA plasmids were identified that carry homologues of
206 TS44 genes (Table S3), i.e., pWDL1-3 and pWDL1-5, pRA8-1 and pSRS16-2. None of those
207 plasmids categorized into a known plasmid group. Although these plasmids carried T4SS,
208 the origin of transfer (oriT) and the type IV coupling protein (T4CP) could not always be
209 identified, and a relaxase, which is necessary for conjugative transfer (33), could only be
210 identified in pWDL1-3. pWDL1-3 carries a remarkably high number of 41 putative

transposases albeit without IS*1071*, and several gene clusters for xenobiotic degradation. Among these is a gene cluster that encodes for homologues (40-43% identity) of proteins encoded by the *tphA1A2A3BR* –gene cluster for terephthalate degradation in *Comamonas sp.* E6(34) as well as for benzoate 1,2-dioxygenase subunits. pRA8-1 is distantly related to pWDL1-3 and carries the *ccdCFBD* operon. In addition, it contains homologues of genes for the biodegradation of non-chlorinated catechols, as well as cation efflux proteins CusABF (35) and cadmium transport protein CadA (36) flanked by an IS*1071* element. The small pWDL1-5 carries no cargo gene. A highly similar plasmid (99% nt identity, 72% coverage), also without cargo, is present in the chlorobenzene-degrading *Pandoraea pnomenusa* strain MCB032(37). The finding of this plasmid group in two different genera/families of the same bacterial order indicates its transferability within *Burkholderiales*. The Trb homologues encoded by pSRS16-2 as well as the *oriT* are highly similar (75-80%) to those encoded by IncP-1 plasmids. However, unlike IncP-1 plasmids, pSRS16-2 does not carry *trfA*, and its size (560 kbp) is much larger than IncP-1 plasmids. pSRS16-2 encodes for a broad range of functions, including 37 transport-related proteins, 18 proteins related to aromatic degradation and 31 transposases. We conclude that this plasmid represents a novel plasmid group, with conjugative transfer machinery similar to this of IncP-1 plasmids.

Non-transferrable plasmids pSRS16-3 and pPBL-E5-2

The closely related plasmids pPBL-E5-2 and pSRS16-3 carrying the *repB-parAB* replication system do not contain homologues of genes related to conjugal transfer, suggesting that they are not self-transmissable. These plasmids are different to the rest in that they have a GC content and codon usage similar to the chromosomes. Both carry distantly-related homologues of catabolic genes such as *tfdA* encoding conversion of 2,4-D (33% aa identity) in *Cupriavidus necator* JMP 134(38), and the *dmpKLMNOPQBCDEFGHI* gene cluster for phenol degradation (45-66% aa identity) in *Pseudomonas sp.* CF600) (39), as well as the

phnCDEGHIJKLMN gene cluster for phosphonate uptake and degradation (45-62% aa identity) in *Eschericia coli* K12) (40).

t-RNA carrying megaplasmids of linuron-degrading *Variovorax sp.*

Pairwise alignment showed that pPBL-H6-1, pSRS16-1, pPBL-E5-1 and pWDL1-1 are highly identical to one another. These show 99% nt identity over the entire sequence including the putative replication/partitioning module *repB-parAB* (Figure 3A). pPBL-H6-1 carries all three gene clusters required to convert linuron to 3-oxoadipate, while pSRS16-1 and pPBL-E5-1 only carry the *dcaQTA1A2BR* and *ccdCFDE* gene clusters. The pWDL1-1 variant sequenced in this study carries the *ccdCFDE* genes and the *hylA* gene. The proteins encoded by the *repB-parAB* module show only slight similarity to their nearest relatives (27, 48 and 39% aa similarity for RepB, ParA and ParB, respectively) and hence the four mega-plasmids might represent a new plasmid group, potentially specific for *Variovorax*. All four plasmids carry *traALBFHJDNUWG* and *trbCG* homologues, albeit with low similarity at aa level (30-43%) involved in conjugative transfer, suggesting that these may be transferrable plasmids. No putative relaxase was though found.

Strikingly, the four megaplasmids carry tRNA genes that encode for all the proteinogenic aa codons. Unlike the scattered appearance of the tRNA genes located on the chromosome, the plasmid encoded tRNA genes are concentrated in one array (Figure 3B). Except for the tRNA encoding for codons glutamine (CAG) and arginine (CGC), all tRNAs on these plasmids are also present on the chromosome. In all four hosts, these two codons are preferred by both plasmids and the chromosome, however, multiple other tRNAs encode for these aa's on the chromosome, suggesting that although the plasmid-encoded tRNAs may enhance gene expression they are not essential for expression of plasmid genes. pSRS16-1 further lacks the tRNA for valine (CAC), which is preferred by both the chromosome and plasmids; however this tRNA is present in pSRS16-2 as well as in the chromosome.

261 The presence of tRNA genes on large plasmids has been reported before in other bacteria
262 (41–44). None of these plasmids however related to the tRNA-carrying *Variovorax* plasmids.
263 In *Bifidobacterium breve* the tRNA-encoding plasmid improves gene expression from both
264 the chromosome and the plasmid (43) while in *Anabaena sp.* PCC 7120 it is dispensable for
265 growth (44). Other MGEs different from plasmids (45) as well as bacteriophages (46)
266 encode for tRNA genes. In the acidophilic, bioleaching bacterium *Acidithiobacillus*
267 *ferrooxidans*, the tRNA genes are located on an integrative-conjugative element and are
268 likely not essential for growth (45).

269 Another feature that sets these plasmids apart is the presence of a CRISPR3-Cas cassette,
270 which is identical (100% identity on nt level) in all of them. The cassette consists of a CRISPR
271 array with 15 spacers, in addition to the genes encoding for a Cas6/Cse3/CasE-type
272 endonuclease, the Cascade subunits CasA, CasB, CasC and CasD and the CRISPR-associated
273 proteins Cas1 and Cas2, which is similar to the class I-E CRISPR-Cas systems (47). The
274 CRISPR-defense system protects bacteria and archaea against MGEs and phages (48), and
275 can be transferred horizontally (43, 49). Although the exact direct repeats of the CRISPR
276 structure were also found in the *Serpentinomonas mccroryi* strain B1 genome
277 (GCA_000828915.1), the spacer sequences did not have any match in the CRISPR
278 databases. Interestingly, no CRISPRs were found in other *Variovorax* genomes available in
279 public databases, with the exception of *Variovorax sp.* PDC80 (GCF_900115375.1), which
280 carries a class I-F CRISPR-Cas system with spacers unrelated to those of the megaplasmids.

281 Other plasmids that carry tRNA genes in the linuron-degrading strains are plasmids pRA8-2
282 and pWDL1-2. These plasmids are distantly-related to pPBL-H6-1, pSRS16-1, pPBL-E5-1 and
283 pWDL1-1 and also carry the tRNA genes in an array, however, their tRNAs do not encode
284 for all proteinogenic amino acids and are all redundant. The two plasmids neither have
285 functions for conjugal transfer nor carry catabolic genes but encode for putative heavy
286 metal resistance, like the cobalt-zinc-cadmium efflux system encoded by *czcABCD* (50, 51),

the copper-response two-component system encoded by *cusRS* and *cusABRS* (35), and mercury resistance encoded by *merACPTR* (52) . Unlike pWDL1-2, pRA8-2 carries an IS*1071* element adjacent to a gene cluster encoding for homologues of the toxin-antitoxin system proteins DinJ-YafQ (53) and the antirestriction protein KlcA that play a role in ensuring plasmid stability.

Genomic organization of linuron degradation genes among different *Variovorax* strains

We analyzed the presence, location and genomic context of *hylA* and *libA* genes for linuron hydrolysis, *dcaQTA1A2BR* genes for DCA conversion to 4,5-chlorocatechol(54) and *ccdCDEFR* genes for 4,5- dichlorocatechol degradation(1, 6) in each of the degraders genome. These genes were not present in any of the other publicly-available *Variovorax* genomes.

Genomic context of the linuron-hydrolysis genes *hylA* and *libA*

The linuron hydrolysis genes *hylA* and *libA* were highly conserved among the different strains. The six linuron degraders carried either *hylA* or *libA*, but never both. *hylA* is present in one copy in strains PBS-H4 and WDL1. As previously reported for WDL1, *hylA* in strain PBS-H4 makes part of a larger gene cluster of 13 ORFs that is highly conserved among the two hosting strains (99 to 100 % nt identity and complete synteny) and that is flanked at both sites by an IS*1071* composing a putative composite transposon (Figure 4A). The function of the *hylA* associated ORFs, and in particular the downstream ORFs, is currently unclear, but their conservative nature indicates that they play a role in linuron hydrolysis. Albers et al.(55) showed the upregulation of the downstream *luxRI* homologues when WDL1 is degrading linuron within a consortium. The *hylA* carrying composite transposon likely originated by inserting the *hylA* gene together with its downstream ORFs in a precursor composite transposon carrying the *iorAB* and *dca* gene clusters as suggested by

312 the presence of *iorAB* and a *dcaQ* remnant directly upstream of *hylA*. Interestingly, the
313 *dcaQ* gene that directly flanks *hylA*, is truncated at a different residue in WDL1 and in PBS-
314 H4 (PBS-H4: nt position 749, WDL1: 689), which suggests that the *hylA* gene and its
315 associated downstream ORFs were independently acquired by the composite transposons
316 present in the two strains. However, this does not necessarily mean that IS*1071* recruited
317 the *hylA* locus for WDL-1 and PBS-H4. The locus might have existed, though in different
318 forms, before its recruitment by WDL1 and PBS-H4, as integrated into a composite
319 transposon and might have been distributed as such.

320 *libA* is present in SRS16, PBL-E5, RA8 and PBL-H6. This gene also makes part of a larger
321 highly conserved gene cluster of four ORFs flanked at both borders by an IS*1071* element
322 and hence composing a putative composite transposon structure (Figure 4B). In addition to
323 *libA*, this gene cluster contains a *luxR*-family transcription regulator directly adjacent to *libA*
324 followed by an IS*91*-family insertion element. The remarkable conservation of the *libA* locus
325 including its integration into a composite transposon, might suggest that *Variovorax*
326 recruited the *libA* locus rather through an already existing composite transposon structure
327 carrying *libA*, as suggested above for the recruitment of the *hylA* locus, As reported above,
328 the *libA* locus is carried by identical IncP-1 plasmids in strains PBL-E5, RA8 and SRS16. SRS16
329 and PBL-E5 were isolated from the same agricultural field, albeit at different time points,
330 and hence this plasmid seems to play an important role in distributing the *libA* locus in that
331 field. RA8 though was isolated in Japan, indicating that similar plasmids evolved at different
332 locations, or that the BHR plasmid was transferred across a large geographic distance. In
333 contrast, in PBL-H6, which originated from a Belgian agricultural field, the transposon is
334 located on the t-RNA carrying megaplasmid pPBL-H6-1 as two flanking complete copies. As
335 such, in PBL-H6, the *libA* composite transposon appears to be obtained by integrating in a
336 replicon different from IncP-1 plasmids. However, PBL-H6 also harbors the PromA plasmid
337 pPBL-H6-2 that carries a copy of the IS*1071* element. An ancestral version of the PromA

plasmid might have been the initial carrier of the *libA* carrying composite transposon in PBL-H6. After the ancestral plasmid was recruited by PBL-H6, the transposon might have transposed to and duplicated in pPBL-H6-1, after which the *libA* gene cluster was lost from pPBL-H6-2 by homologous recombination of the IS*1071* element, leaving one copy behind. A similar phenomenon was previously shown for an IncP-1 plasmid carrying genes for atrazine biodegradation (56).

The *dca* and *ccd* clusters of linuron-degrading *Variovorax*

Similar to the *hylA* and *libA* genes, MGEs determine the genomic context of the *dca* and *ccd* genes. PBL-E5, PBL-H6 and SRS16 carry the entire *dca* and *ccd* clusters, which is consistent with their ability to degrade linuron and DCA. The *ccd* clusters of strains PBL-E5, PBL-H6 and SRS16 are on the 800 Mbp megaplasmids, directly downstream of the *dca* clusters (Figure 4B). In both strains, the entire *dca/ccd* gene cluster is bordered by IS*1071* at both sides, with two inward-directed IRs and one outward-directed IR missing. A similar composite transposon configuration including the *dca* and *ccd* genes is found in the chloroaniline degrader *Delftia acidovorans* LME1(54) (Figure 4B), except that PBL-H6 and SRS16 have two copies of the *dcaA1A2* genes and PBL-E5 two adjacent copies of the *dcaQTA1A2B* genes with one intact and one truncated *dcaB*. Other chloroaniline degraders like *Comamonas testosteroni* WDL7 and *Delftia acidovorans* B8c(54) carry a similar structure but lacking the *ccd* genes (Figure 4B). Amino acid-level similarity of the proteins encoded by *dcaQTA1A1BR* and *ccdCFDE* between the three *Variovorax* strains and LME1/WDL7/B8c is around 99% (Table S4). As for the *hylA* and *libA* loci, we hypothesize that the *dca/ccd* gene clusters were obtained by being already integrated into an ancestral composite transposon but that afterwards gene rearrangements occurred that explained the observed variations. . This is further supported by the observation that in the chloroanaline degrading *Comamonas/Delftia* strains, the composite transposon structures carrying *dca/ccd* genes are located on IncP-1 plasmids, while in the three *Variovorax* strains they are located on the

364 800 kbp t-RNA carrying megaplasmids. In case of PBL-H6, these are directly adjacent to the
365 composite transposon structure carrying *libA* (see above). In all three strains the location of
366 the composite transposons is the same, marking the corresponding location as a likely
367 accessory gene insertion hot spot in the mega-plasmid.

368 The *dca* cluster and associated ORFs present on pWDL1-1 are also bordered by two IS*1071*
369 elements, but the composite transposon does not contain a *ccd* cluster and the *dca* genes
370 are rather related (99% nt identity) to the *tadQTA1A2B* encoding for conversion of aniline
371 to catechol in *Delftia tsuruhatensis* AD9 (57). Also the *ccd* cluster, present on pWDL1-1,
372 differs from those of PBL-E5, PBL-H6 and SRS16. This cluster is also bordered by IS*1071*
373 elements and its location is in the direct vicinity of either *hylA* or the *dca* genes, depending
374 on the WDL1 subpopulation (9). These genes are relatively distantly-related to other known
375 catechol degradation genes (Table S4).

376 The *ccd* gene clusters in the non-DCA degraders PBS-H4 and RA8 are identical to those of
377 WDL1. In PBS-H4, the *ccd* cluster is on the PromA plasmid pPBS-H4-2, that also bears the
378 *hylA* carrying composite transposon. The *ccd* cluster is flanked by IS*1071* elements at both
379 ends composing a putative composite transposon. Unlike pWDL1-1, where the putative
380 transposons carrying the *hylA* and *ccd* cluster are directly adjacent to each other, on pPBS-
381 H4-2 they are separated by other IS*1071* elements that don't carry catabolic genes. In RA8,
382 the *ccd* cluster is located on pRA8-1. This cluster is flanked by one truncated $tnpA_{IS1071}$
383 (Figure 4A), without an IR, indicating that IS1071 also played a role in recruitment of the ccd
384 cluster by pRA8-1. Overall, as for the *hylA*, *libA* and *dca* loci, the apparent strong
385 conservation of the *ccd* genes in a composite transposon structure in different plasmid
386 vehicles indicates again that these genes were recruited as a composite transposon
387 structure that already contained the respective loci.

388 Overall, the analysis of the genetic context of the genes involved in linuron catabolism,
389 either upstream or downstream functions in the pathway, indicates that IS*1071* insertion
390 element play an essential role in the plasmid-associated acquisition of the catabolic genes
391 and genetic adaptation of *Variovorax* toward the ability to degrade linuron. The
392 phenomenon that IS*1071* elements are associated with genes involved in xenobiotic
393 degradation was previously observed via both cultivation-dependent (7, 27, 58, 59) and
394 cultivation-independent(16) methods. Subsequent inter- and intramolecular transposition
395 of IS*1071* is thought to lead to the assembly of catabolic genes into a composite transposon
396 structure with the recruited genes flanked by IS*1071* at both sites(60, 61). The new
397 composite transposon structure can then move on itself between different replicons as
398 shown for instance for Tn*5271*, the composite transposon consisting of the chlorobenzoate
399 catabolic genes *cbaABC* flanked by IS*107*1, in *Alcaligines* sp BRC60(62). The strong
400 conservative nature of the catabolic cargo in the putative composite transposon structures
401 suggest that the recruitment of the different catabolic clusters in the linuron degrading
402 *Variovorax* strains is rather due to the recruitment of already existing catabolic composite
403 transposons rather than by the assembly process itself. Broad host range plasmids such as
404 IncP-1 are known to distribute catabolic clusters in communities and often contain IS*1071*
405 associated catabolic composite transposons (7, 54). Interestingly, each of the linuron
406 degraders with the exception of WDL1, carries at least one plasmid of a well-known
407 promiscuous plasmid group (IncP-1 or PromA). Their involvement in distributing linuron
408 catabolic genes in the linuron degrading strains is suggested from the fact that these
409 plasmids all carry at least one of the involved catabolic gene functions with the exception of
410 pPBL-H6-2. However, as mentioned above, the latter carries an IS*1071* copy which might be
411 obtained from internal recombination between two IS*1071* elements bordering a
412 composite transposon as explained above. In addition, other plasmids of yet unknown type,
413 carrying signs of conjugative features, were present. Likely, plasmids move around in a

414 community, and pick up the composite transposons, for further transfer to their final hosts.
415 As such, the BHR plasmids found in the linuron degrading strains might have, next to
416 IS*1071*, functioned as a second crucial vehicle for acquisition of the catabolic genes.
417 Interestingly, regarding catabolic functions, only a limited reservoir of catabolic loci was
418 utilized for integration in the linuron catabolic pathways indicating that the choice of
419 suitable catabolic functions for integration into a functional linuron catabolic pathway in
420 *Variovorax* is limited, even on a worldwide scale. Curiously, these genes seem only to be
421 recruited by a specific clade 1 of *Variovorax*, whose members, in addition to linuron, seem
422 to be prone to genetic adaptation and hence specialization towards the biodegradation of
423 other anthropogenic compounds. Apparently, for some reason, this clade is able to recruit
424 and express foreign genes for xenobiotic biodegradation. Sequence analysis of other
425 genomes within this clade, in addition to the linuron degraders, and comparison with the
426 genome sequences of clade 2, might unravel the mechanisms involved in this special ability
427 of catabolic gene recruitment.

428 # Materials and methods

429 ## Genome sequencing, assembly and annotation

430 The details of the biomass and library preparation for genome sequencing are given in the
431 supplementary text S1. Genome assembly was performed based on the PacBio reads by
432 means of the RS_HGAP_Assembly.3 protocol included in SMRT Portal version 2.3.0 applying
433 target genome sizes of 5 Mbp (PBL-E5), 15 Mbp (WDL1) and 10 Mbp (others). All assemblies
434 showed one chromosomal contig, several extra-chromosomal contigs and several artificial
435 contigs. Artificial contigs were removed from the assembly. The remaining contigs were
436 circularized and assembly redundancies at the ends of the contigs were removed. ORFs on
437 the replicons were ordered using *dnaA* (chromosome) or *repA*/*parA* (plasmids) as the first
438 ORF. Error correction was performed by mapping the Illumina short reads onto finished

genomes using bwa v. 0.6.2 in paired-end (sampe) mode using default settings(63) with subsequent variant and consensus calling using VarScan v. 2.3.6 (Parameters: mpileup2cns --min-coverage 10 --min-reads2 6 --min-avg-qual 20 --min-var-freq 0.8 --min-freq-for-hom 0.75 --p-value 0.01 --strand-filter 1 --variants 1 --output-vcf 1)(63). A consensus concordance of QV60 was reached. Automated genome annotation was performed using Prokka 1.8(64). Genomes were deposited at EMBL/ENA under the accession numbers LR594659-LR594661 (PBL-H6), LR594662-LR594665 (RA8), LR594666-LR594670 (SRS16), LR594671-LR594674 (PBL-E5), LR594675-LR594677 (PBS-H4), and LR594689-LR594694 (WDL1).

Phylogenomic analysis of *Variovorax sp.* plasmids and genomes

The accession numbers of the genome and plasmid sequences used to construct the phylogenetic trees are listed in Table S1.

First, a phylogenomic analysis of the whole genome dataset was conducted at the nucleotide level using the truly whole-genome-based Genome-BLAST Distance Phylogeny method (GBDP)(65, 66)(67). Briefly, GBDP infers accurate intergenomic distances between pairs of genome sequences and subjects resulting distances matrices to a distance-based phylogenetic reconstruction under settings recommended for the comparison of prokaryotic genomes(65). The method is used by both the Genome-to-Genome Distance Calculator 2.1(65) and the Type Strain Genome Server(67).

A second phylogenetic analysis based on the plasmids' amino acid sequences was conducted using GBDP as well, except that GBDP distance calculations were done under settings recommended for the analysis of bacteriophage sequences(68). The reason is that the sequence lengths of the plasmid sequences were in the same order of magnitude than bacteriophage sequences(68) and thus promised an equally good performance of the GBDP method when applied to plasmid data. The publicly-available plasmid sequences included in

464 this study were selected based on their relatedness to the newly-sequenced plasmids, in
465 order to allocate them into known plasmid groups.

466 Regarding both analyses, a balanced minimum evolution tree was inferred using FastME
467 v2.1.4 with SPR postprocessing each(69). 100 replicate trees were reconstructed in the
468 same way and branch support was subsequently mapped onto the respective tree(66).

469 For the 16S rRNA gene sequence-based phylogeny, the whole 16S rRNA gene sequences
470 were retrieved from the SILVA database(70), and aligned with the SINA aligner(70). The
471 phylogenies were inferred on the GGDC web server(65) using the DSMZ phylogenomics
472 pipeline (https://ggdc.dsmz.de/phylogeny-service.php). Maximum likelihood (ML) and
473 maximum parsimony (MP) trees were inferred from the alignment with RAxML(70) and
474 TNT(70), respectively. For ML, rapid bootstrapping in conjunction with the autoMRE
475 bootstopping criterion(70) and subsequent search for the best tree was used; for MP, 1000
476 bootstrapping replicates were used in conjunction with tree-bisection-and-reconnection
477 branch swapping and ten random sequence addition replicates. The sequences were
478 checked for a compositional bias using the X^2 test as implemented in PAUP*(70). All
479 phylogenetic trees were visualized with iTOL(71).

480 Analysis of *Variovorax sp.* plasmids

481 The codon usage of chromosomes and plasmids was calculated with CompareM v.
482 0.0.23(72). Subsequently, a PCA was conducted and the principle components were
483 hierarchically clustered using the Ward's criterion with FactoMineR FactoMineR v. 1.36(73).
484 The replication origins as well as the type IV secretion systems were predicted with
485 oriTfinder v. 1.1(70).

486 The COG categories on pPBL-H6-1 was determined with eggNOG(74). The circular
487 representation was drawn with CGview(75) (Figure 3A), and the BLAST-based comparative
488 analysis illustration was generated with Easyfig(76) (Figure 3B). SimpleSynteny(77) was

used to determine the positions of the catabolic clusters and associated ORFs, and draw the catabolic cluster illustrations.

References

1. Bers K, Leroy B, Breugelmans P, Albers P, Lavigne R, Sørensen SR, Aamand J, De Mot R, Wattiez R, Springael D. 2011. A Novel Hydrolase Identified by Genomic-Proteomic Analysis of Phenylurea Herbicide Mineralization by Variovorax sp. Strain SRS16. Appl Environ Microbiol 77:8754–8764.

2. Sørensen SR, Rasmussen J, Jacobsen CS, Jacobsen OS, Juhler RK, Aamand J. 2005. Elucidating the key member of a linuron-mineralizing bacterial community by PCR and reverse transcription-PCR denaturing gradient gel electrophoresis 16S rRNA gene fingerprinting and cultivation. Appl Environ Microbiol 71:4144–4148.

3. Satsuma K. 2010. Mineralisation of the herbicide linuron by Variovorax sp. strain RA8 isolated from Japanese river sediment using an ecosystem model (microcosm). Pest Manag Sci.

4. Breugelmans P, D'Huys PJ, De Mot R, Springael D. 2007. Characterization of novel linuron-mineralizing bacterial consortia enriched from long-term linuron-treated agricultural soils. FEMS Microbiol Ecol.

5. Dejonghe W, Berteloot E, Goris J, Boon N, Crul K, Maertens S, Höfte M, De Vos P, Verstraete W, Top EM. 2003. Synergistic degradation of linuron by a bacterial consortium and isolation of a single linuron-degrading Variovorax strain. Appl Environ Microbiol 69:1532–1541.

6. Bers K, Batisson I, Proost P, Wattiez R, De Mot R, Springael D. 2013. Hyla, an alternative hydrolase for initiation of catabolism of the phenylurea herbicide linuron in variovorax sp. strains. Appl Environ Microbiol 79:5258–5263.

513 7. Boon N, Goris J, De Vos P, Verstraete W, Top EM. 2001. Genetic Diversity among 3-
514 Chloroaniline- and Aniline-Degrading Strains of the Comamonadaceae. Appl Environ
515 Microbiol 67:1107–1115.

516 8. Albers P, Lood C, Özturk B, Horemans B, Lavigne R, van Noort V, De Mot R, Marchal K,
517 Sanchez-Rodriguez A, Springael D. 2018. Catabolic task division between two near-isogenic
518 subpopulations co-existing in a herbicide-degrading bacterial consortium: consequences
519 for the interspecies consortium metabolic model. Environ Microbiol.

520 9. Albers P, Lood C, Özturk B, Horemans B, Lavigne R, van Noort V, De Mot R, Marchal K,
521 Sanchez-Rodriguez A, Springael D. 2018. Catabolic task division between two near-isogenic
522 subpopulations co-existing in a herbicide-degrading bacterial consortium: consequences
523 for the interspecies consortium metabolic model. Environ Microbiol 20:85–96.

524 10. Petersen J, Frank O, Göker M, Pradella S. 2013. Extrachromosomal, extraordinary and
525 essential - The plasmids of the Roseobacter clade. Appl Microbiol Biotechnol 97:2805–
526 2815.

527 11. Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial “chromid”:
528 not a chromosome, not a plasmid. Trends Microbiol 18:141–148.

529 12. Zhang M, Warmink J, Pereira e Silva MC, Brons J, Smalla K, van Elsas JD. 2015. IncP-1β
530 Plasmids Are Important Carriers of Fitness Traits for Variovorax Species in the
531 Mycosphere—Two Novel Plasmids, pHB44 and pBS64, with Differential Effects Unveiled.
532 Microb Ecol 70:141–153.

533 13. Zhang M, Brons JK, van Elsas JD. 2016. The complete sequences and ecological roles of two
534 IncP-1ß plasmids, pHB44 and pBS64, isolated from the mycosphere of Laccaria proxima.
535 Front Microbiol 7:1–11.

536 14. Kim DU, Kim MS, Lim JS, Ka JO. 2013. Widespread occurrence of the tfd-II genes in soil

bacteria revealed by nucleotide sequence analysis of 2,4-dichlorophenoxyacetic acid degradative plasmids pDB1 and p712. Plasmid 69:243–248.

15. Fulthorpe RR, Wyndham RC. 1992. Involvement of a chlorobenzoate-catabolic transposon, Tn5271, in community adaptation to chlorobiphenyl, chloroaniline, and 2,4-dichlorophenoxyacetic acid in a freshwater ecosystem. Appl Environ Microbiol 58:314–325.

16. Dunon V, Bers K, Lavigne R, Top EM, Springael D. 2018. Targeted metagenomics demonstrates the ecological role of IS *1071* in bacterial community adaptation to pesticide degradation. Environ Microbiol 00.

17. Dunon V, Sniegowski K, Bers K, Lavigne R, Smalla K, Springael D. 2013. High prevalence of IncP-1 plasmids and IS1071 insertion sequences in on-farm biopurification systems and other pesticide-polluted environments. FEMS Microbiol Ecol 86:415–431.

18. Providenti MA, Shaye RE, Lynes KD, McKenna NT, O'Brien JM, Rosolen S, Wyndham RC, Lambert LB. 2006. The locus coding for the 3-nitrobenzoate dioxygenase of Comamonas sp strain JS46 is flanked by IS1071 elements and is subject to deletion and inversion events. Appl Environ Microbiol 72:2651–2660.

19. Woo HL, DeAngelis KM, Teshima H, Davenport K, Daligault H, Erkkila T, Goodwin L, Gu W, Lo C-C, Munk C, Scholz M, Xu Y, Chain P, Bruce D, Detter C, Tapia R, Han C, Simmons BA, Hazen TC. 2017. High-Quality Draft Genome Sequences of Four Lignocellulose-Degrading Bacteria Isolated from Puerto Rican Forest Soil: Gordonia sp., Paenibacillus sp., Variovorax sp., and Vogesella sp. Genome Announc. 1752 N St., N.W., Washington, DC.

20. Crombie AT, Larke-Mejia NL, Emery H, Dawson R, Pratscher J, Murphy GP, McGenity TJ, Murrell JC. 2018. Poplar phyllosphere harbors disparate isoprene-degrading bacteria. Proc Natl Acad Sci U S A 115:13081–13086.

561 21. Wang Y, Hatt JK, Tsementzi D, Rodriguez-R LM, Ruiz-Perez CA, Weigand MR, Kizer H,
562 Maresca G, Krishnan R, Poretsky R, Spain JC, Konstantinidis KT. 2017. Quantifying the
563 Importance of the Rare Biosphere for Microbial Community Response to Organic
564 Pollutants in a Freshwater Ecosystem. Appl Environ Microbiol 83:1–19.

565 22. Mahan KM, Zheng H, Fida TT, Parry RJ. 2017. Iron-Dependent Enzyme Catalyzes the Initial
566 Step in Biodegradation of N-Nitroglycine by Variovorax sp. Strain JS1663 Kristina 83:1–10.

567 23. Satola B, Wübbeler JH, Steinbüchel A. 2013. Metabolic characteristics of the species
568 Variovorax paradoxus. Appl Microbiol Biotechnol 97:541–560.

569 24. Zhang HJ, Zhou QW, Zhou GC, Cao YM, Dai YJ, Ji WW, Shang GD, Yuan S. 2012.
570 Biotransformation of the neonicotinoid insecticide thiacloprid by the bacterium variovorax
571 boronicumulans strain J1 and mediation of the major metabolic pathway by nitrile
572 hydratase. J Agric Food Chem 60:153–159.

573 25. Tabata M, Ohhata S, Nikawadori Y, Kishida K, Sato T, Kawasumi T, Kato H, Ohtsubo Y,
574 Tsuda M, Nagata Y. 2016. Comparison of the complete genome sequences of four γ-
575 hexachlorocyclohexane-degrading bacterial strains: insights into the evolution of bacteria
576 able to degrade a recalcitrant man-made pesticide. DNA Res 23:581–599.

577 26. Kim DW, Lee K, Lee DH, Cha CJ. 2018. Comparative genomic analysis of pyrene-degrading
578 Mycobacterium species: Genomic islands and ring-hydroxylating dioxygenases involved in
579 pyrene degradation. J Microbiol 56:798–804.

580 27. Sen D, Yano H, Suzuki H, Król JE, Rogers L, Brown CJ, Top EM. 2010. Comparative genomics
581 of pAKD4, the prototype IncP-1δ plasmid with a complete backbone. Plasmid 63:98–107.

582 28. Vedler E, Vahter M, Heinaru A. 2004. The Completely Sequenced Plasmid pEST4011
583 Contains a Novel IncP1 Backbone and a Catabolic Transposon Harboring
584 tfd Genes for 2,4-Dichlorophenoxyacetic Acid Degradation. J

585 Bacteriol 186:7161 LP-7174.

586 29. Xia X-S, Aathithan S, Oswiecimska K, Smith ARW, Bruce IJ. 1998. A Novel Plasmid pIJB1
587 Possessing a Putative 2,4-Dichlorophenoxyacetate Degradative Transposon
588 Tn5530inBurkholderia cepaciaStrain 2a. Plasmid 39:154–159.

589 30. Yanagiya K, Maejima Y, Nakata H, Tokuda M, Moriuchi R, Dohra H, Inoue K, Ohkuma M,
590 Kimbara K, Shintani M. 2018. Novel Self-Transmissible and Broad-Host-Range Plasmids
591 Exogenously Captured From Anaerobic Granules or Cow Manure. Front Microbiol 9:1–12.

592 31. Mela F, Fritsche K, Boersma H, Van Elsas JD, Bartels D, Meyer F, De Boer W, Van Veen JA,
593 Leveau JHJ. 2008. Comparative genomics of the pIPO2/pSB102 family of environmental
594 plasmids: Sequence, evolution, and ecology of pTer331 isolated from Collimonas
595 fungivorans Ter331. FEMS Microbiol Ecol 66:45–62.

596 32. Van der Auwera GA, Król JE, Suzuki H, Foster B, Van Houdt R, Brown CJ, Mergeay M, Top
597 EM. 2009. Plasmids captured in C. metallidurans CH34: Defining the PromA family of
598 broad-host-range plasmids. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 96:193–
599 204.

600 33. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. 2010. Mobility of
601 Plasmids. Microbiol Mol Biol Rev 74:434 LP-452.

602 34. Sasoh M, Masai E, Ishibashi S, Hara H, Kamimura N, Miyauchi K, Fukuda M. 2006.
603 Characterization of the Terephthalate Degradation Genes of. Society 72:1825–1832.

604 35. Franke S, Grass G, Rensing C, Nies DH. 2003. Molecular analysis of the copper-transporting
605 efflux system CusCFBA of Escherichia coli. J Bacteriol 185:3804–3812.

606 36. Tsai KJ, Yoon KP, Lynn AR. 1992. ATP-dependent cadmium transport by the cadA cadmium
607 resistance determinant in everted membrane vesicles of Bacillus subtilis. J Bacteriol

608 174:116 LP-121.

609 37. Baptista IIR, Zhou NY, Emanuelsson EAC, Peeva LG, Leak DJ, Mantalaris A, Livingston AG.
610 2008. Evidence of species succession during chlorobenzene biodegradation. Biotechnol
611 Bioeng 99:68–74.

612 38. Plumeier I, Pérez-Pantoja D, Heim S, González B, Pieper DH. 2002. Importance of Different
613 tfd Genes for Degradation of Chloroaromatics by Ralstonia eutropha JMP134. J
614 Bacteriol 184:4054 LP-4064.

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

618 40. Jiang W, Metcalf WW, Lee K, Wanner BL. 1995. Molecular Cloning , Mapping , and
619 Regulation of Pho Regulon Genes for Phosphonate Breakdown by the Phosphonatase
620 Pathway of Salmonella typhimurium LT2 177:6411–6421.

621 41. Sakai Y, Ogawa N, Shimomura Y, Fujii T. 2018. A 2 , 4-dichlorophenoxyacetic acid
622 degradation plasmid pM7012 discloses distribution of an unclassified megaplasmid group
623 across bacterial species 525–536.

624 42. Morgado SM, Vicente ACP. 2018. Beyond the Limits: tRNA Array Units in Mycobacterium
625 Genomes . Front Microbiol .

626 43. Bottacini F, Motherway MOC, Casey E, McDonnell B, Mahony J, Ventura M, van Sinderen
627 D. 2015. Discovery of a conjugative megaplasmid in Bifidobacterium breve. Appl Environ
628 Microbiol 81:166–176.

629 44. Puerto-Galan L, Vioque A. 2012. Expression and processing of an unusual tRNA gene
630 cluster in the cyanobacterium Anabaena sp. PCC 7120. FEMS Microbiol Lett 337:10–17.

631 45. Alamos P, Tello M, Bustamante P, Gutiérrez F, Shmaryahu A, Maldonado J, Levicán G,
632 Orellana O. 2018. Functionality of tRNAs encoded in a mobile genetic element from an
633 acidophilic bacterium. RNA Biol 15:518–527.

634 46. Pope WH, Anders KR, Baird M, Bowman CA, Boyle MM, Broussard GW, Chow T, Clase KL,
635 Cooper S, Cornely KA, DeJong RJ, Delesalle VA, Deng L, Dunbar D, Edgington NP, Ferreira
636 CM, Weston Hafer K, Hartzog GA, Hatherill JR, Hughes LE, Ipapo K, Krukonis GP, Meier CG,
637 Monti DL, Olm MR, Page ST, Peebles CL, Rinehart CA, Rubin MR, Russell DA, Sanders ER,
638 Schoer M, Shaffer CD, Wherley J, Vazquez E, Yuan H, Zhang D, Cresawn SG, Jacobs-Sera D,
639 Hendrix RW, Hatfull GF. 2014. Cluster M Mycobacteriophages Bongo, PegLeg, and Rey with
640 Unusually Large Repertoires of tRNA Isotypes. J Virol 88:2461 LP-2480.

641 47. Koonin E V, Makarova KS, Zhang F. 2017. Diversity, classification and evolution of CRISPR-
642 Cas systems. Curr Opin Microbiol 37:67–78.

643 48. Pothier JF, Vorholter F-J, Blom J, Goesmann A, Puhler A, Smits THM, Duffy B. 2011. The
644 ubiquitous plasmid pXap41 in the invasive phytopathogen Xanthomonas arboricola pv.
645 pruni: complete sequence and comparative genomic analysis. FEMS Microbiol Lett 323:52–
646 60.

647 49. Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their
648 associated genes: Evidence of horizontal transfer among prokaryotes. J Mol Evol 62:718–
649 729.

650 50. Legatzki A, Grass G, Anton A, Rensing C, Nies DH. 2003. Interplay of the Czc System and
651 Two P-Type ATPases in Conferring Metal Resistance to Ralstonia metallidurans. J Bacteriol
652 185:4354 LP-4361.

653 51. Rensing C, Pribyl T, Nies DH. 1997. New Functions for the Three Subunits of the CzcCBA
654 Cation-Proton Antiporter 179:6871–6879.

655 52. Munson GP, Lam DL, Outten FW, O'Halloran T V. 2000. Identification of a copper-
656 responsive two-component system on the chromosome of Escherichia coli K-12. J Bacteriol
657 182:5864–5871.

658 53. Armalyte J, Jurenaite M, Beinoraviciute G, Teiserskas J, Suziedeliene E. 2012.
659 Characterization of Escherichia coli dinJ-yafQ toxin-antitoxin system using insights from
660 mutagenesis data. J Bacteriol 194:1523–1532.

661 54. Król JE, Penrod JT, McCaslin H, Rogers LM, Yano H, Stancik AD, Dejonghe W, Brown CJ,
662 Parales RE, Wuertz S, Top EM. 2012. Role of IncP-1β plasmids pWDL7::rfp and pNB8c in
663 chloroaniline catabolism as determined by genomic and functional analyses. Appl Environ
664 Microbiol 78:828–838.

665 55. Albers P, Weytjens B, De Mot R, Marchal K, Springael D. 2018. Molecular processes
666 underlying synergistic linuron mineralization in a triple-species bacterial consortium
667 biofilm revealed by differential transcriptomics. Microbiologyopen 7:1–14.

668 56. Devers M, Rouard N, Martin-Laurent F. 2007. Genetic rearrangement of the atzAB
669 atrazine-degrading gene cassette from pADP1::Tn5 to the chromosome of Variovorax sp.
670 MD1 and MD2. Gene 392:1–6.

671 57. Liang Q, Takeo M, Chen M, Zhang W, Xu Y, Lin M. 2005. Chromosome-encoded gene
672 cluster for the metabolic pathway that converts aniline to TCA-cycle intermediates in
673 Delftia tsuruhatensis AD9. Microbiology 151:3435–3446.

674 58. Sota M, Kawasaki H, Tsuda M. 2003. Structure of Haloacetate-Catabolic IncP-1β Plasmid
675 pUO1 and Genetic Mobility of Its Residing Haloacetate-Catabolic Transposon. J Bacteriol
676 185:6741 LP-6745.

677 59. Martinez B, Tomkins J, Wackett LP, Wing R, Sadowsky MJ. 2001. Complete nucleotide
678 sequence and organization of the atrazine catabolic plasmid pADP-1 from Pseudomonas

679 sp. strain ADP. J Bacteriol 183:5684–5697.

680 60. Nakatsu C, Ng J, Singh R, Straus N, Wyndham C. 1991. Chlorobenzoate catabolic
681 transposon {Tn}5271 is a composite class-{I} element with flanking class-{Ii} insertion
682 sequences. Proc Natl Acad Sci U S A 88:8312–8316.

683 61. Di Gioia D, Peel M, Fava F, Wyndham RC. 1998. Structures of Homologous Composite
684 Transposons Carrying cbaABC genes from Europe and North America. Appl Environ
685 Microbiol 64:1940 LP-1946.

686 62. Ng J, Wyndham RC. 1993. IS1071-mediated recombinational equilibrium in Alcaligenes sp.
687 BR60 carrying the 3-chlorobenzoate catabolic transposon Tn5271. Can J Microbiol 39:92–
688 100.

689 63. Goesmann A, Bertelli C, Ernst C, Kreis J, Spänig S, Juhre T, Blom J. 2016. EDGAR 2.0: an
690 enhanced software platform for comparative gene content analyses. Nucleic Acids Res
691 44:W22–W28.

692 64. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–
693 2069.

694 65. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species
695 delimitation with confidence intervals and improved distance functions. BMC
696 Bioinformatics 14:60.

697 66. Meier-kolthoff JP, Auch AF, Klenk H, Göker M. 2014. Highly parallelized inference of large
698 genome-based phylogenies. Concurr Comput Pract Exp 26:1715–1729.

699 67. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for
700 state-of-the-art genome-based taxonomy. Nat Commun 10:2182.

701 68. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of

prokaryotic viruses. Bioinformatics2017/07/07. 33:3396–3404.

69. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol Biol Evol2015/06/30. 32:2798–2800.

70. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, Ou H-Y. 2018. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. Nucleic Acids Res 46:W229–W234.

71. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259.

72. D. P. 2017. CompareM (https://github.com/dparks1134/CompareM).

73. Le S, Josse J, Husson F, Ois. 2008. FactoMineR: An R Package for Multivariate Analysis. J Stat Software, Artic 25:1–18.

74. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286-93.

75. Stothard P, Wishart DS. 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21:537–539.

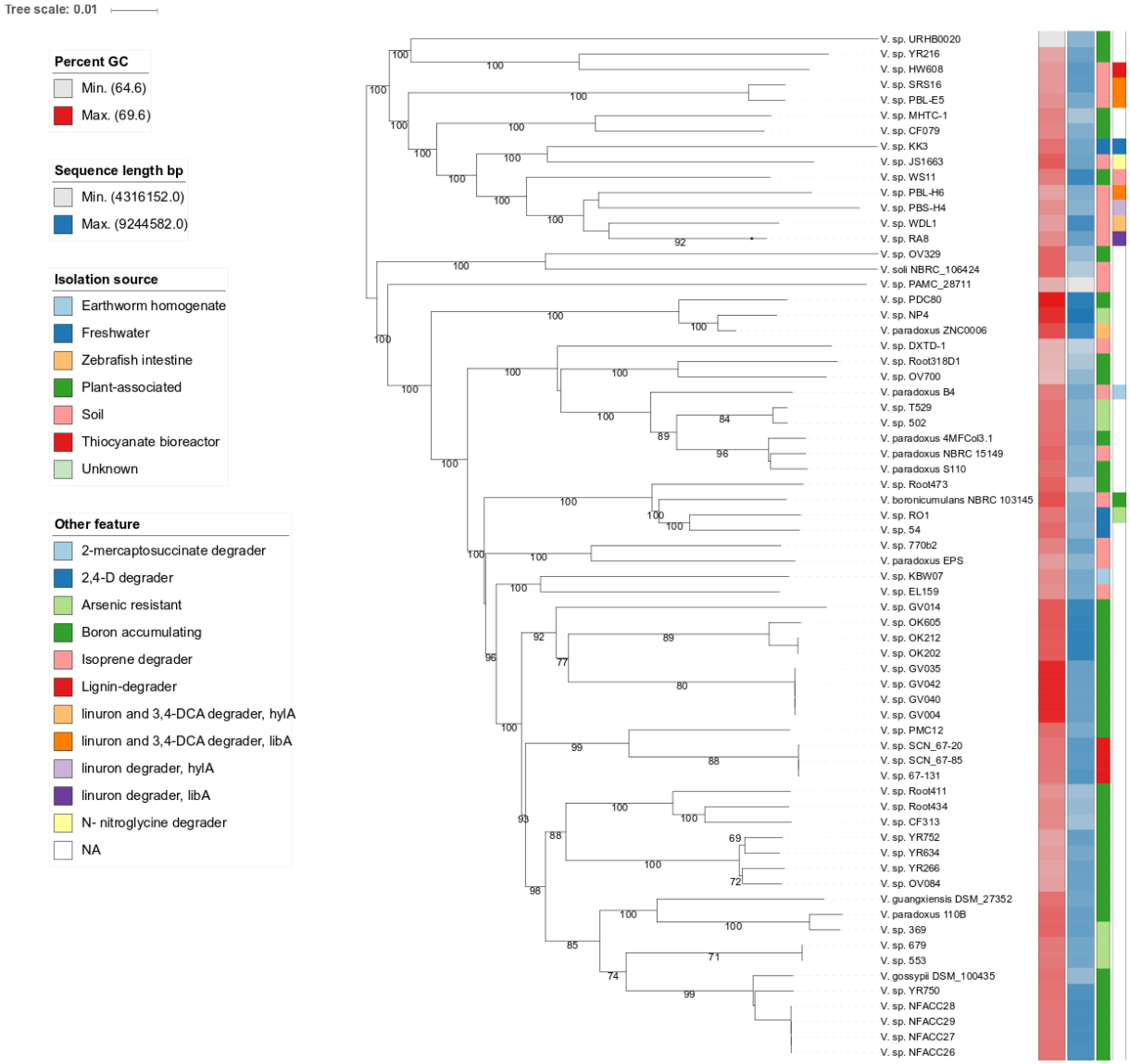
76. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics2011/01/28. 27:1009–1010.

77. Veltri D, Wight MM, Crouch JA. 2016. SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. Nucleic Acids Res 44:W41-5.

Table 1: General properties of the newly sequenced *Variovorax* genomes

Strain	Replicon	Size (kbp)	GC%	Number of ORFs	Number of tRNA genes	Number of Transposases	Number of IS1071	Degradation genes
WDL1	chromosome	6724.9	67.2	6376	56	64	5	
(linuron=>CO2)	pWDL1-1	825.1	62.5	862	44	29		*hylA* or *dca* and *ccd*
	pWDL1-2	565.9	63.5	543	27	24		
	pWDL1-3	207.9	63.7	229		42		
	pWDL1-5	20.1	62.6	25				
	pWDL1-4	25.0	62.6	23		8	1	
	Total	**8368.8**		**8058**	**127**	**167**	**6**	
PBL-H6	chromosome	5990.2	66.8	5557	54	6		
(linuron=>CO2)	pPBL-H6-1	839.2	62.4	883	50	36	5	*libA, dca, ccd*
	pPBL-H6-2 (PromA)	42.1	63.5	49		1	1	
	Total	**6871.6**		**6489**		**43**	**6**	
PBS-H4	chromosome	6429.8	66.9	6031	57	7	2	
(linuron=>DCA)	pPBS-H4-1	117.4	64.9	95		2		
	pPBS-H4-2 (PromA)	104.9	62.7	112		14	5	*hylA, ccd*
	Total	**6652.1**		**6238**	**161**	**23**	**7**	
RA8	chromosome	6501.6	67.2	6129	52	7		
(linuron=>DCA)	pRA8-1	429.0	64.9	443		8	2	*ccd*
	pRA8-2	425.3	64.2	419	23	19	2	
	pRA8-3 (IncP-1)	68.4	61.2	63		30	3	*libA*
	Total	**7424.2**		**482**	**23**	**64**	**7**	
SRS16	chromosome	5763.0	67.3	5469	50	13		
(linuron=>CO2)	pSRS16-1	801.4	62.5	852	50	19	2	*dca, ccd*
	pSRS16-2	560.6	64.5	555		31		
	pSRS16-3	478.8	67	469		0		
	pSRS16-4 (IncP-1)	71.1	61.9	66		12	3	*libA*
	Total	**7674.9**		**7411**	**100**	**75**	**4**	
PBL-E5	chromosome	5660.8	67.3	5421	47	10		
(linuron=>CO2)	pPBL-E5-1	801.5	62.5	844	50	20	1	*dca, ccd*
	pPBL-E5-2	553.0	67.1	550		4		
	pPBL-E5-3 (IncP-1)	71.0	61.3	66		11	3	*libA*
	Total	**7086.4**		**6881**	**97**	**45**	**4**	

724

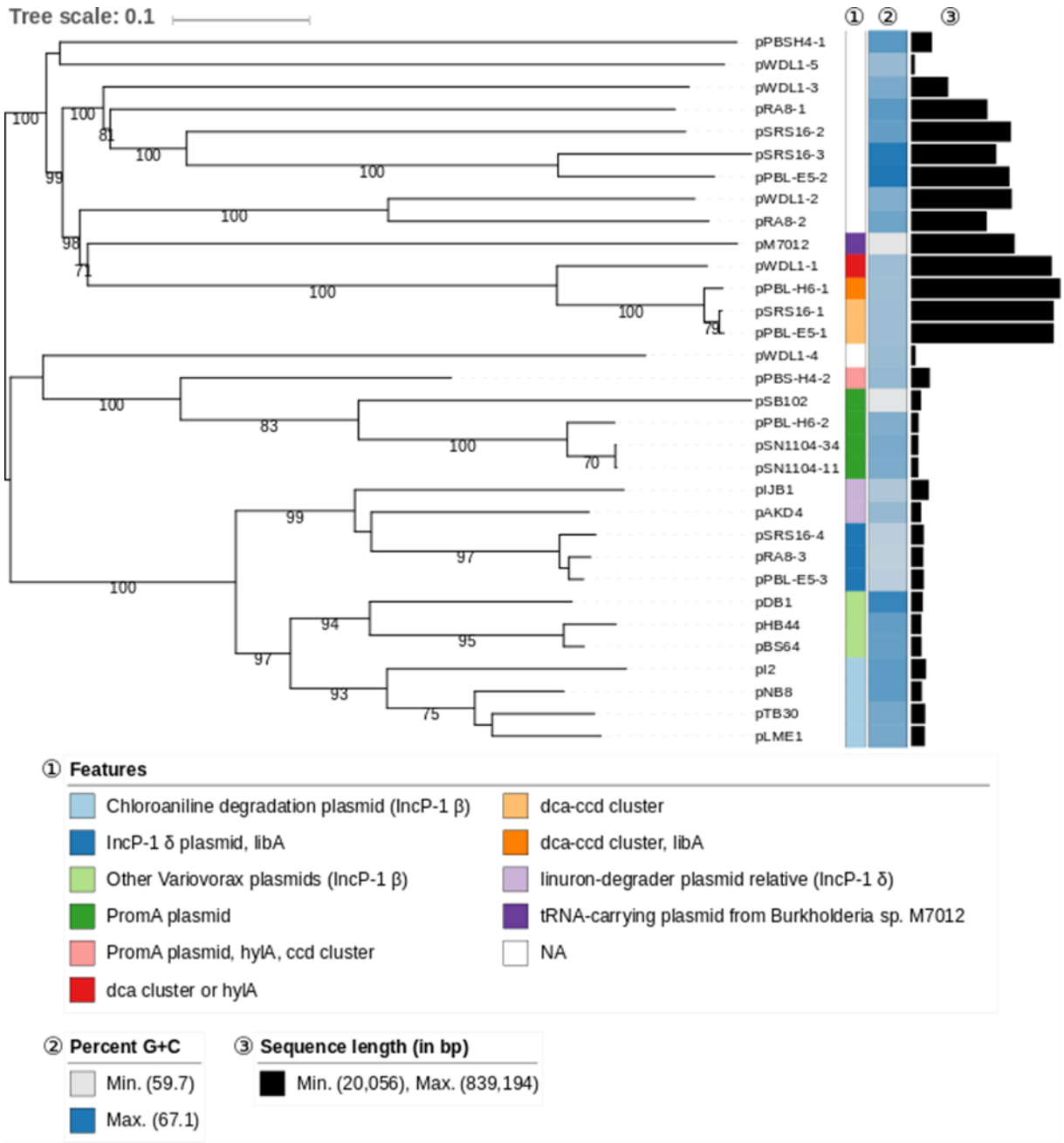


725

726 Figure 1: GBDP phylogenomic analysis of the *Variovorax* whole genome dataset. The branch
727 lengths are scaled in terms of GBDP distance formula d_5. The numbers above branches are
728 GBDP pseudo-bootstrap support values from 100 replications, with an average branch
729 support of 80.6%. Leaf labels are further annotated by their genomic G+C content, genome
730 sequence length, phenotypic attributes and origin of isolation.

731

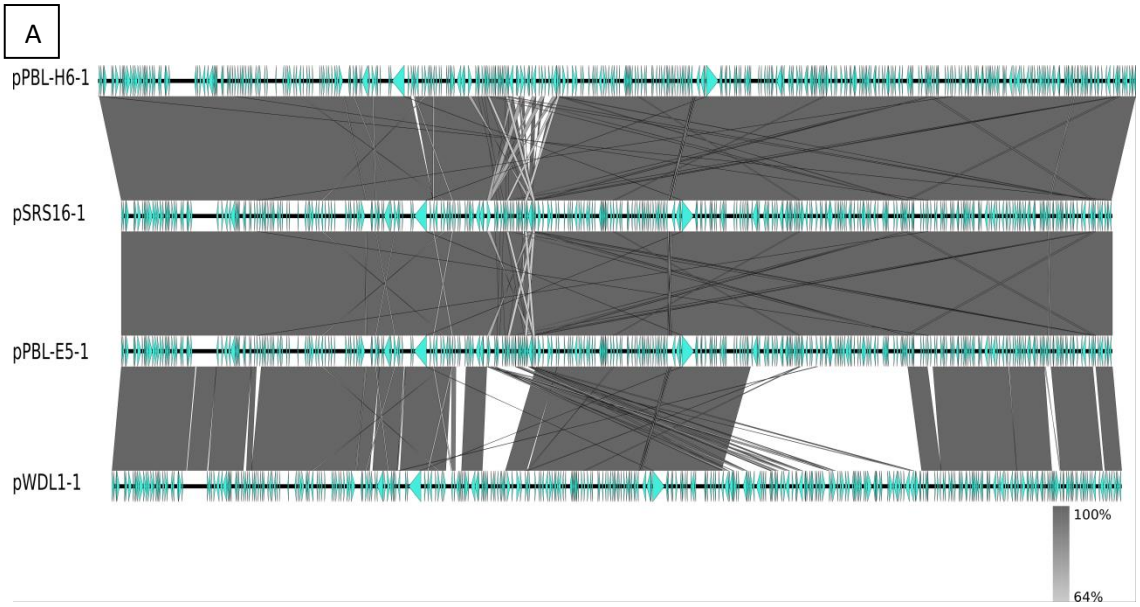
732



733

734 Figure 2: GBDP phylogenomic analysis of *Variovorax* plasmids and relevant relatives. The
735 branch lengths are scaled in terms of GBDP distance formula d_6. The numbers above
736 branches are GBDP pseudo-bootstrap support values from 100 replications, with an
737 average branch support of 91.3%. Leaf labels are further annotated by their genomic G+C
738 content, length as well as special attributes. Previously-sequenced plasmids included in the
739 study and their NCBI accession numbers are: pM7012 (NC_022995.1), pSB102 (AJ304453.1),
740 pSN1104-34 (AP018708.1), pSN1104-11 (AP018707.1), pIJB1 (JX847411.1), pAKD4

(GQ983559.1), pDB1 (JQ436721.1), pHB44 (KU356988.1), pBS64 (KU356987.1), pL2 (JF274989.1), pNB8 (NC_019264.1), pLME1 (NC_019263.1) and pTB30 (NC_016968.1).



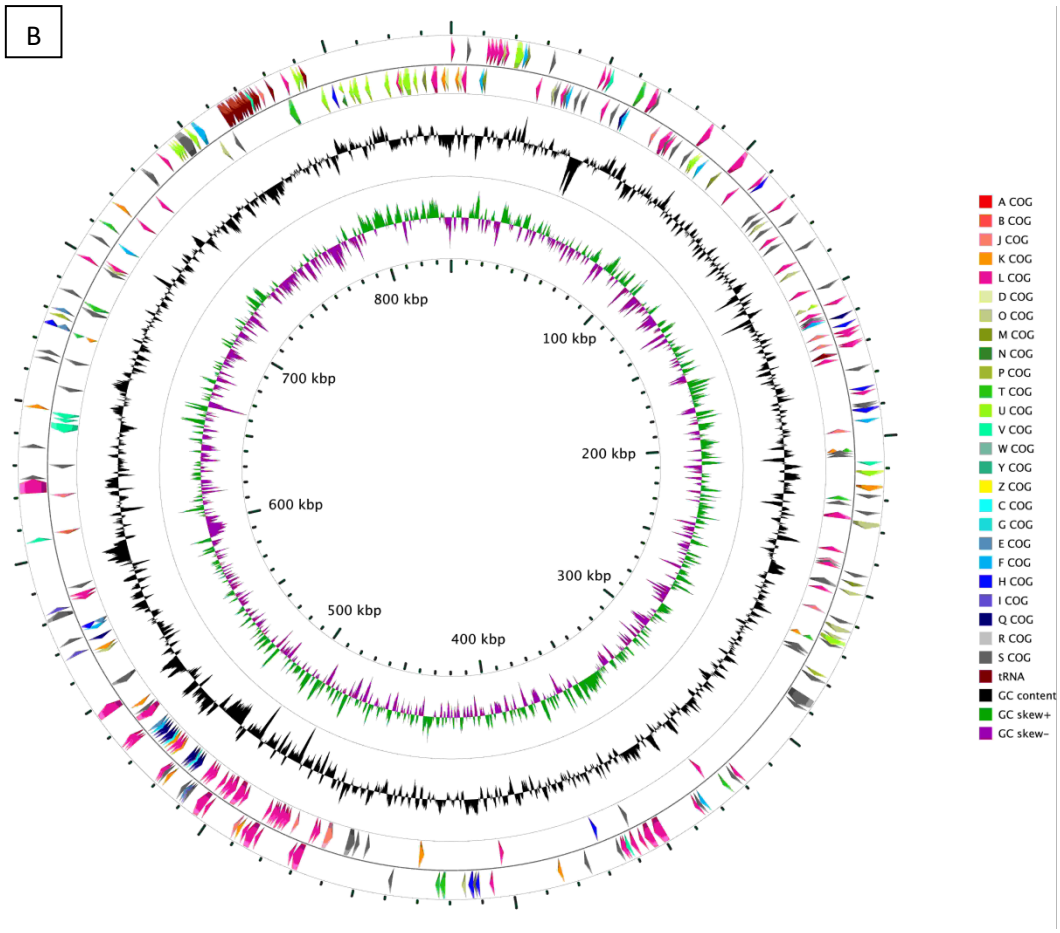
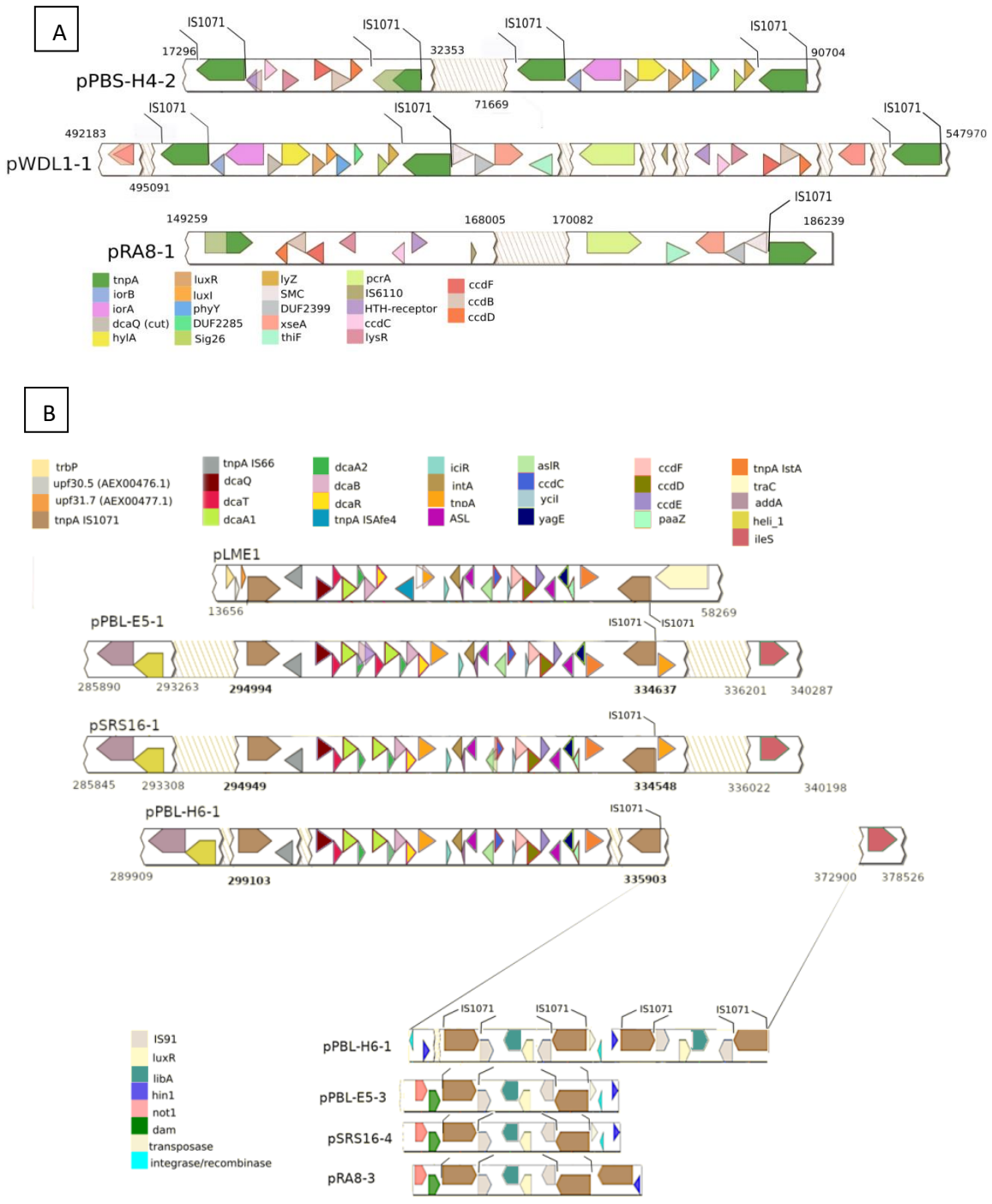


Figure 3: (A) Synteny and BLAST identity of the four Variovorax tRNA carrying megaplasmids. Shaded regions indicate the BLAST identity at the nucleotide level. (B) Circular representation of the pPBL-H6-1 as an example of a Variovorax tRNA carrying megaplasmid. The colored arrows represent the main COG categories that the proteins were assigned to, as well as the tRNA genes. G+C content and skew are represented in the inner circles.



781 Figure 4: Catabolic clusters and their synteny. The genes with the same color code share 99-
782 100% identity at the aa level. Shaded arrows indicate truncated genes, with the lighter-
783 shade arrow indicating the full gene size. Broken likes indicate IS*1071* IRs.

784 (A): Catabolic genes of WDL1, RA8 and PBS-H4. The *hylA* and *ccd* clusters on the plasmids
785 pPBS-H4-2, pWDL1-1 and pRA8-1, as well as flanking genes are illustrated. Catabolic cluster
786 locations are depicted for each plasmid in base pairs.

787 (B) Catabolic genes of SRS16, PBL-H6 and PBL-E5. On the top panel, the *dca* and *ccd* clusters
788 on plasmids pPBL-E5-1, pSRS16-1 and pPBL-H6-1 as well as the *Delftia acidovorans* plasmid
789 pLME1 are illustrated, together with flanking genes and genes directly neighboring each
790 IS*1071* element. The catabolic cluster locations and IS*1071* insertion positions are given for
791 each plasmid in base pairs. In the lower panel, the *libA* gene-associated IS*1071* elements
792 are illustrated for pPBL-H6-1, pPBL-E5-2, pSRS16-4 and pRA8-3. The insertion site of the
793 *libA*- associated IS*1071* elements on pPBL-H6-1 is depicted with fading lines. For the other
794 three plasmids, genes directly flanking the IS*1071* elements are given as well.

795 ## Acknowledgements

796 We thank Sebastian S. Sørensen and Koji Satsuma for the donation of strains SRS16 and RA8,
797 Simone Severitt for technical assistance, and Markus Göker, Jörn Petersen, Kornelia Smalla and
798 Isabel Schober for valuable discussions regarding the manuscript. JMK was supported by Deutsche
799 Forschungsgemeinschaft within “Sonderforschungsbereich TRR 51”. The authors acknowledge the
800 use of de.NBI cloud and the support by the High Performance and Cloud Computing Group at the
801 Zentrum für Datenverarbeitung of the University of Tübingen and the Federal Ministry of
802 Education and Research (BMBF) through grant no 031 A535A.

803

804

805

Supplementary information

Supplementary Text S1: Library preparation for PacBio and Illumina sequencing:

Biomass for genome sequencing of the six linuron-degrading Variovorax was obtained from 10 mL cultures grown in R2B medium supplemented with 10 mg/L linuron (Sigma-Aldrich, analytical standard) at 20 °C till an OD600 of 0.8-1. The linuron and/or DCA degradation phenotype of the cultures was assessed by monitoring the compounds concentration as previously described(1). DNA was isolated from the cultures using Qiagen Genomic-tip 100/G (Qiagen, Hilden Germany) kit according to the manufacturer's instructions. For long read sequencing, 15 kbp libraries were prepared according to the SMRTbell™ template preparation protocol of PacificBiosciences (Menlo Park, USA), following the Procedure & Checklist – Greater Than 10 kbp Template Preparation. Briefly, 8 µg genomic DNA was sheared using g-tubes™ from Covaris (Woburn, USA). DNA was end-repaired and ligated overnight to hairpin adapters using the DNA/Polymerase Binding Kit P6 (Pacific Biosciences, Menlo Park, USA). BluePippin™ Size-Selection gel cassettes were used to select for DNA fragments greater than 4 kbp according to the manufacturer´s instructions (Sage Science, Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell™ template were assessed with the Calculator in RS Remote (PacificBiosciences, Menlo Park, USA). One SMRT cell for sequencing was used per strain on a PacBio RSII apparatus (PacificBiosciences, Menlo Park, USA) taking one 240-minutes movie with exception of strain RA8, for which three SMRT Cells were used. For short read sequencing, short insert libraries were created using the Illumina Nextera XT DNA Library Prep Kit (Illumina, San Diego, USA) and pair-end (2 X 151 bp) sequenced on an Illumina NextSeq 550 (Illumina, San Diego, USA).

831 Table S1: *Variovorax* genomes used in this study.

Accession nr	Species	Sequence length (bp)	Percent GC	Other feature
GCA 001591345.1	V. boronicumulans NBRC_103145	6724210	68.2885	Boron accumulating, soil
GCA 003965815.1	V. gossypii DSM_100435	6300928	67.4438	Cotton root
GCA 003952165.1	V. guangxiensis DSM_27352	7184833	67.4395	Banana root
GCA 000382045.1	V. paradoxus 110B	7515245	67.7348	Rhizosphere of Arabidopsis thaliana
GCA 000377585.1	V. paradoxus 4MFCol3.1	7011330	67.5227	Rhizosphere of Arabidopsis thaliana
GCA 000463015.1	V. paradoxus B4	7148516	67.1562	Soil, 2-mercaptosuccinate degrader
GCA 000184745.1	V. paradoxus EPS	6550056	66.4797	Soil
GCA 001591365.1	V. paradoxus NBRC_15149	6664268	67.7407	Soil
GCA 000023345.1	V. paradoxus S110	6754997	67.5362	Endophyte
GCA 000807585.2	V. paradoxus ZNC0006	8473132	68.3596	Zebrafish intestine
GCA 001591385.1	V. soli NBRC_106424	5600425	67.787	Greenhouse soil
GCA 003955655.1	V. sp. 369	7487985	67.7807	n.k.
GCA 003951285.1	V. sp. 502	6765052	67.4001	n.k.
GCA 002754375.1	V. sp. 54	6642361	67.6846	Lake sediment
GCA 003950725.1	V. sp. 553	7212420	67.2617	n.k.

GCA 001899795.1	V. sp. 67-131	7946830	67.2828	Thiocyanate bioreactor
GCA 003952185.1	V. sp. 679	7212419	67.2617	n.k.
GCA 900115685.1	V. sp. 770b2	7449424	67.0516	Forest soil
GCA 900101545.1	V. sp. CF079	6846381	66.9727	Populus root
GCA 000282635.1	V. sp. CF313	6028886	66.8402	Populus root
GCA 003984625.1	V. sp. DXTD-1	5254730	65.8435	Desert soil
GCA 900100965.1	V. sp. EL159	7001541	66.6979	Forest soil
GCA 003253535.1	V. sp. GV004	7466671	69.3112	Populus root
GCA 003096925.1	V. sp. GV014	8606402	68.1098	Populus root
GCA 003217395.1	V. sp. GV035	7466290	69.3105	Populus root
GCA 003053685.1	V. sp. GV040	7459457	69.3174	Populus root
GCA 003208625.1	V. sp. GV042	7458914	69.3169	Populus root
GCA 900090195.1	V. sp. HW608	7733463	66.5307	lignin degrader, soil
GCA 002157355.1	V. sp. JS1663	7324361	68.0306	N- Nitroglycine degrader, soil
GCA 003852515.1	V. sp. KBW07	7174338	66.8329	Earthworm homogenate
GCA 001984055.1	V. sp. KK3	7190799	67.4707	2,4-D degrader, freshwater
GCA 003984645.1	V. sp. MHTC-1	5882526	67.0643	Ephedra rhizosphere

GCA 900112125.1	V. sp. NFACC26	8206836	67.3845	Endophyte
GCA 900113295.1	V. sp. NFACC27	8233050	67.3685	Endophyte
GCA 900107915.1	V. sp. NFACC28	8243820	67.3351	Endophyte
GCA 900108265.1	V. sp. NFACC29	8213461	67.3833	Endophyte
GCA 002729445.1	V. sp. NP4	9244582	69.127	n.k.
GCA 900109235.1	V. sp. OK202	8802000	68.0234	Populus root
GCA 900112425.1	V. sp. OK212	8802773	68.0231	Populus root
GCA 900115445.1	V. sp. OK605	8712396	68.1283	Populus root
GCA 900111625.1	V. sp. OV084	7387450	66.2934	Populus root
GCA 900114785.1	V. sp. OV329	6313206	67.8224	Populus root
GCA 900099805.1	V. sp. OV700	6483567	65.691	Populus root
GCA 001577265.1	V. sp. PAMC_28711	4316152	65.9735	Antarctic soil
GCA 900115375.1	V. sp. PDC80	8962376	69.5908	Populus root
GCA 003019815.1	V. sp. PMC12	7015237	67.615	Tomato rhizosphere
GCA 002849325.1	V. sp. RO1	6871164	67.3696	Arsenic resistant
GCA 001424835.1	V. sp. Root318D1	5745530	65.8025	Arabidopsis root
GCA 001425205.1	V. sp. Root411	6063830	66.6364	Arabidopsis root

GCA 001426595.1	V. sp. Root434	6321673	66.9158	Arabidopsis root
GCA 001426505.1	V. sp. Root473	5744754	67.842	Arabidopsis root
GCA 001725775.1	V. sp. SCN_67-20	7722457	67.384	Thiocyanate bioreactor
GCA 001725035.1	V. sp. SCN_67-85	7712539	67.3794	Thiocyanate bioreactor
GCA 003863415.1	V. sp. T529	6695648	67.4047	n.k.
GCA 000620225.1	V. sp. URHB0020	6629371	64.5764	Grassland
GCA 003014875.1	V. sp. WS11	8562068	67.2245	Isoprene degrader, phyllosphere
GCA 900107745.1	V. sp. YR216	7371494	66.2216	Populus root
GCA 900107295.1	V. sp. YR266	7461239	66.2025	Populus root
GCA 900106655.1	V. sp. YR634	7157607	66.3985	Populus root
GCA 900109805.1	V. sp. YR750	8099115	67.4261	Populus root
GCA 900215425.1	V. sp. YR752	7549133	66.2729	Populus root
AY738708	V. sp. MD2	n.a.	n.a.	maize-cultivated soil (only 16S tree)
AY738709	V. sp. MD3	n.a.	n.a.	maize-cultivated soil (only 16S tree)

Table S2: Digital DNA-DNA hybridization values calculated by the type strain genome server (tygs.dsmz.de)(1). A value of 70% of higher indicates that the two microorganisms belong to the same species(2). C.I.: confidence interval.

837

Query strain	Subject strain	dDDH (%)	C.I. (%)
PBL-E5	SRS16	86	[83.3 - 88.3]
RA8	WDL-1	51.2	[48.6 - 53.9]
PBL-H6	WDL-1	44.9	[42.3 - 47.4]
PBL-H6	RA8	43.1	[40.6 - 45.7]
PBS-H4	RA8	40.1	[37.7 - 42.7]
PBL-H6	PBS-H4	39.9	[37.4 - 42.4]
PBS-H4	WDL-1	39.3	[36.8 - 41.8]
PBL-E5	PBL-H6	32.2	[29.8 - 34.7]
PBL-H6	SRS16	32.1	[29.7 - 34.6]
PBL-E5	WDL-1	30.6	[28.2 - 33.1]
SRS16	WDL-1	30.4	[28.0 - 32.9]
PBL-E5	RA8	26.9	[24.6 - 29.4]
RA8	SRS16	26.8	[24.5 - 29.3]
PBL-E5	PBS-H4	25.3	[23.0 - 27.8]
PBS-H4	SRS16	25.3	[22.9 - 27.8]

1. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182.
2. Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z. 2015. Genomic insights into the taxonomic status of the Bacillus cereus group. Sci Rep 5:14082.

838

839

840

841

842

843

844

845

846

847

848 Table S3: Overview of plasmid replication and conjugation systems. T4SS= type IV secretion
849 system, T4CP= type IV coupling protein, n.f.= not found. P=present. For the relaxase, T4CP and
850 T4SS. The % identity on aa level to the nearest relative, as well as the accession number are given.

851

Plasmid	Relaxase	T4CP	T4SS	oriT	Replication genes
pPBL-E5-1 pWDL1-1 pSRS16-1 pPBL-H6-1	n.f.	34% YP_001911165	TraALBFHJDNUW, TrbC, VirB4	n.f.	RepB-ParAB
pSRS16-2	TraI 49% YP_195891	72% NP_990928.1	TrbLJIHGFEDCB, TraJ, VirB1	P	RepB-ParAB
pSRS16-3 pPBL-E5-2	n.f.	n.f.	n.f.	n.f.	RepB-ParAB
pSRS16-4 pRA8-2 pPBL-E5-3	TraI 100% YP_006965894.1	100% NP_990928.1	TrbBCDEFGHIJKLN. TraXF	P	IncP-1
pPBL-H6-2 pPBS-H4-2	TraS 79% CAC79161	67% YP_001672044	VirB123456891011, VirD4	P	RepA
pRA8-1	Rel 81% SDZ72275.1	45% WP_010895213	VirB23456891011, VirD4	n.f.	RepB-ParAB
pRA8-2	n.f.	n.f.	n.f.	n.f.	RepB-ParAB
pPBS-H4-1	TraA 37% AAV52093	n.f.	n.f.	n.f.	RepA
pWDL1-2	n.f.	n.f.	n.f.	n.f.	RepB-ParAB
pWDL1-3	Rel 84% WP_093180082.1	45% WP_010895213	VirB23456891011, VirD4, TrbB	n.f.	RepAB
pWDL1-4	n.f.	n.f.	n.f.	n.f.	ParAB
pWDL1-5	n.f.	38% AEY63616	VirB568, TraD	n.f.	repA
chrWDL1			TrbBCDEJLFGI, VirD4, TraF		
chrSRS16			TrbBCDEJLFGI, VirD4		
chrRA8			n.f.		
chrPBL-H6			TrbBCDEJLFGI, VirD4		
chrPBL-E5			TrbBCDEJLFGI, VirD4		
chrPBS-H4			TrbBCDEJLFGI, VirD4		

852

853

854

Table S4: Presence of the catabolic proteins in each strain, their aa level identity to the nearest relative, and the accession number of the nearest relative.

	PBL-H6	PBL-E5	SRS16	WDL1	PBS-H4	RA8
HylA	no	no	no	yes	yes	no
LibA	yes	yes	yes	no	no	yes
DcaQ	yes (100% identical to AEX00653.2 DcaQ [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (99% identical to WP_047349924.1 DcaQ [*Delftia tsuruhatensis*])	no	no
DcaT	yes (100% identical to AEX00654.1 DcaT [Delftia acidovorans])	yes (see PBL-H6)	yes (see PBL-H6)	yes (98% identical to AAX47240.1 DcaT [Delftia tsuruhatensis])	no	no
DcaA1	yes (100% identical to AEX00655.1 DcaA1 [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (98% identical to AAX47241.1 DcaA1 [Delftia tsuruhatensis])	no	no
DcaA2	yes (AEX00656.1 DcaA2 [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (99% identical to AAX47242.1 DcaA2 [*Delftia tsuruhatensis*])	no	no
DcaB	yes (AEX00657.1 DcaB [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (99% identical to AAX47243.1 DcaT [*Delftia tsuruhatensis*])	no	no
DcaR	yes (100% identical to AEX00658.1 DcaT [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	no	no	no
CcdC	yes (100% identical to WP_015060619.1 chlorocatechol 1,2-dioxygenase [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (76% identical to WP_102776360.1 catechol 1,2-dioxygenase [*Achromobacter pulmonis])*	yes (see WDL1)	yes (see WDL1)
CcdF	yes (100% identical to WP_015060622.1 maleylacetate reductase [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (76% identical to AYE88726.1 chloromaleylacetate reductase [*Diaphorobacter sp.*])	yes (see WDL1)	yes (see WDL1)
CcdD	yes (100% identical to WP_015060623.1	yes (see PBL-H6)	yes (see PBL-H6)	yes (78% identical to AYE88727.1	yes (see	yes (see

	chloromuconate cycloisomerase [*Delftia acidovorans*])			dichloromuconate cycloisomerase [*Diaphorobacter sp.*])	WDL1)	WDL1)
CcdE	yes (100% identical to WP_015060624.1 dienelactone hydrolase [*Delftia acidovorans*])	yes (see PBL-H6)	yes (see PBL-H6)	yes (74% identical to AYE88728.1 chlorodienelactone hydrolase [*Diaphorobacter sp.*])	yes (see WDL1)	yes (see WDL1)

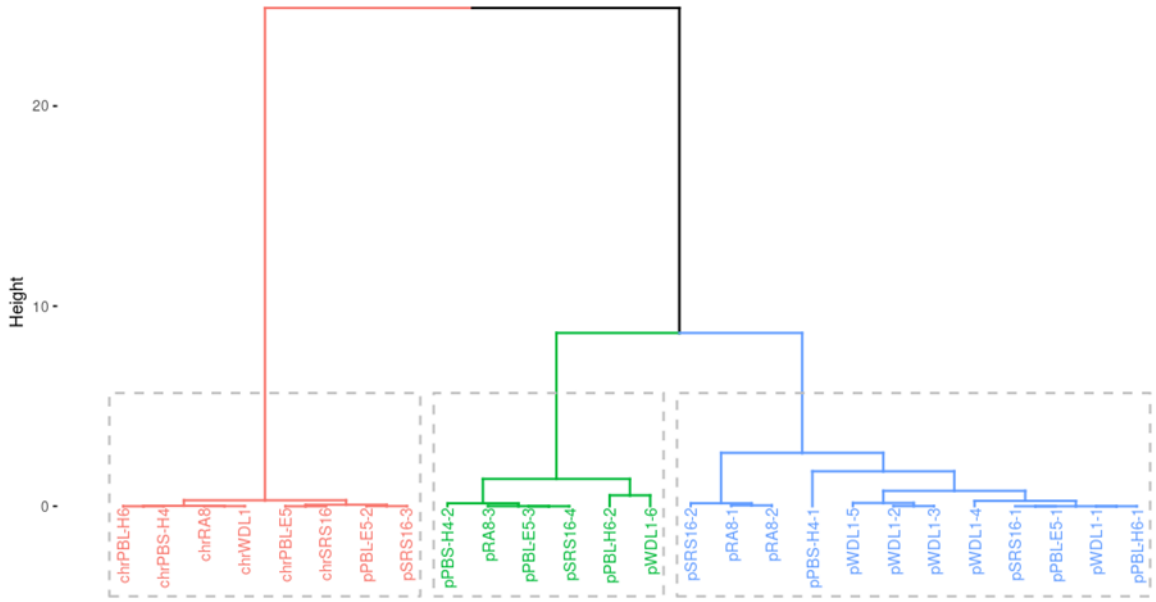


Figure S1: Hierarchical clustering of codon usage of all genomic elements

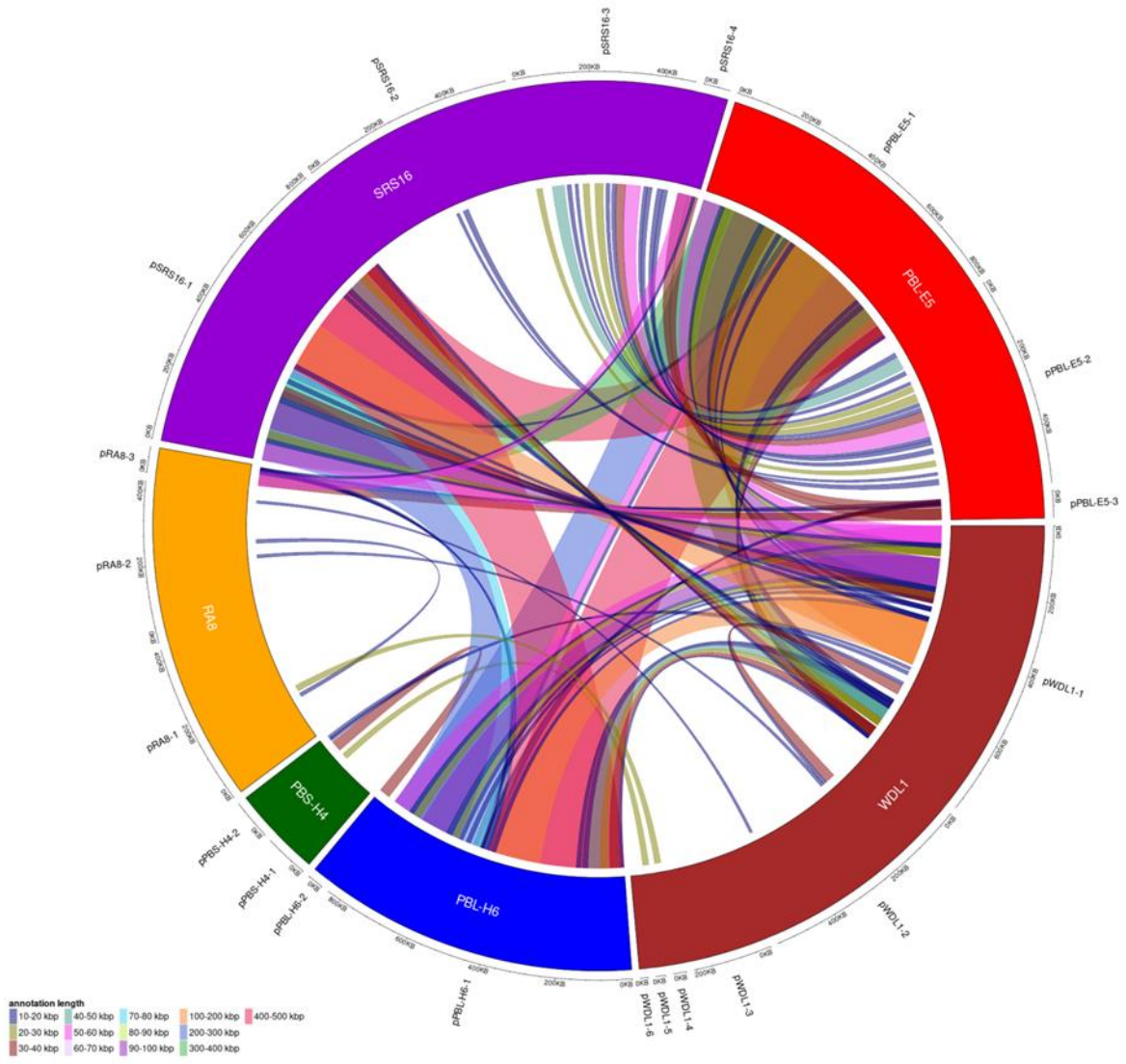
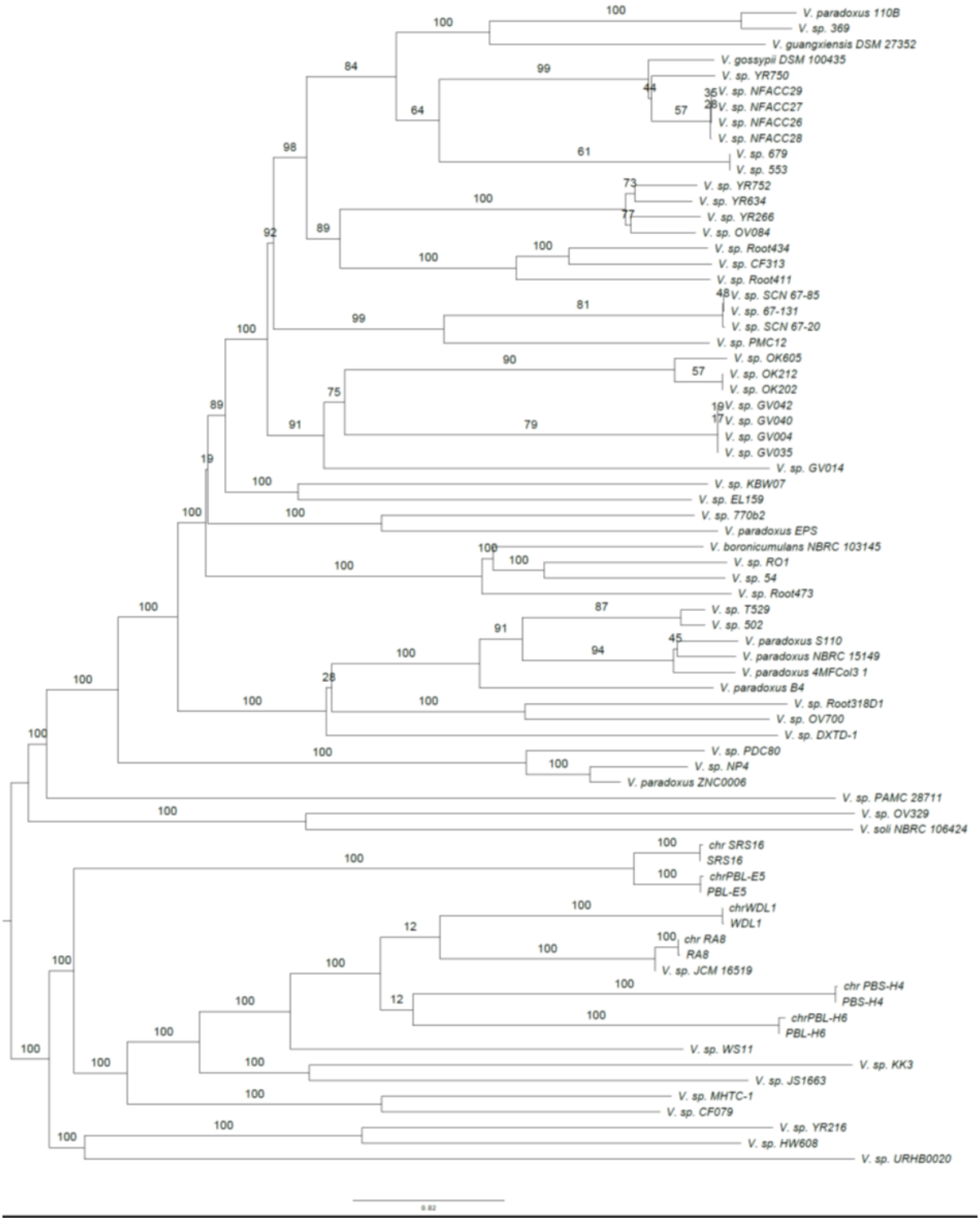


Figure S2: Circular representation of highly-similar regions between the plasmids of the degrading Variovorax genomes. On the ring, the different plasmids are shown, that are linked according to the length of the matching regions. The alignment was calculated with BlastN(2) (evalue < 1e-10, identity >= 95%), only alignment length of at least 10 kbp were taken into account. The circular plasmid comparison was generated with R v. 3.5.2 (https://www.R-project.org/) and the Bioconductor package circlize v. 0.4.5(3).



874

875 Figure S3: GBDP phylogenomic analysis of the Variovorax chromosome dataset. The branch
876 lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP
877 pseudo-bootstrap support values from 100 replications, with an average branch support of 80.6%.
878 For the linuron-degrading Variovorax, whole genomes were included in the calculation for
879 comparison.

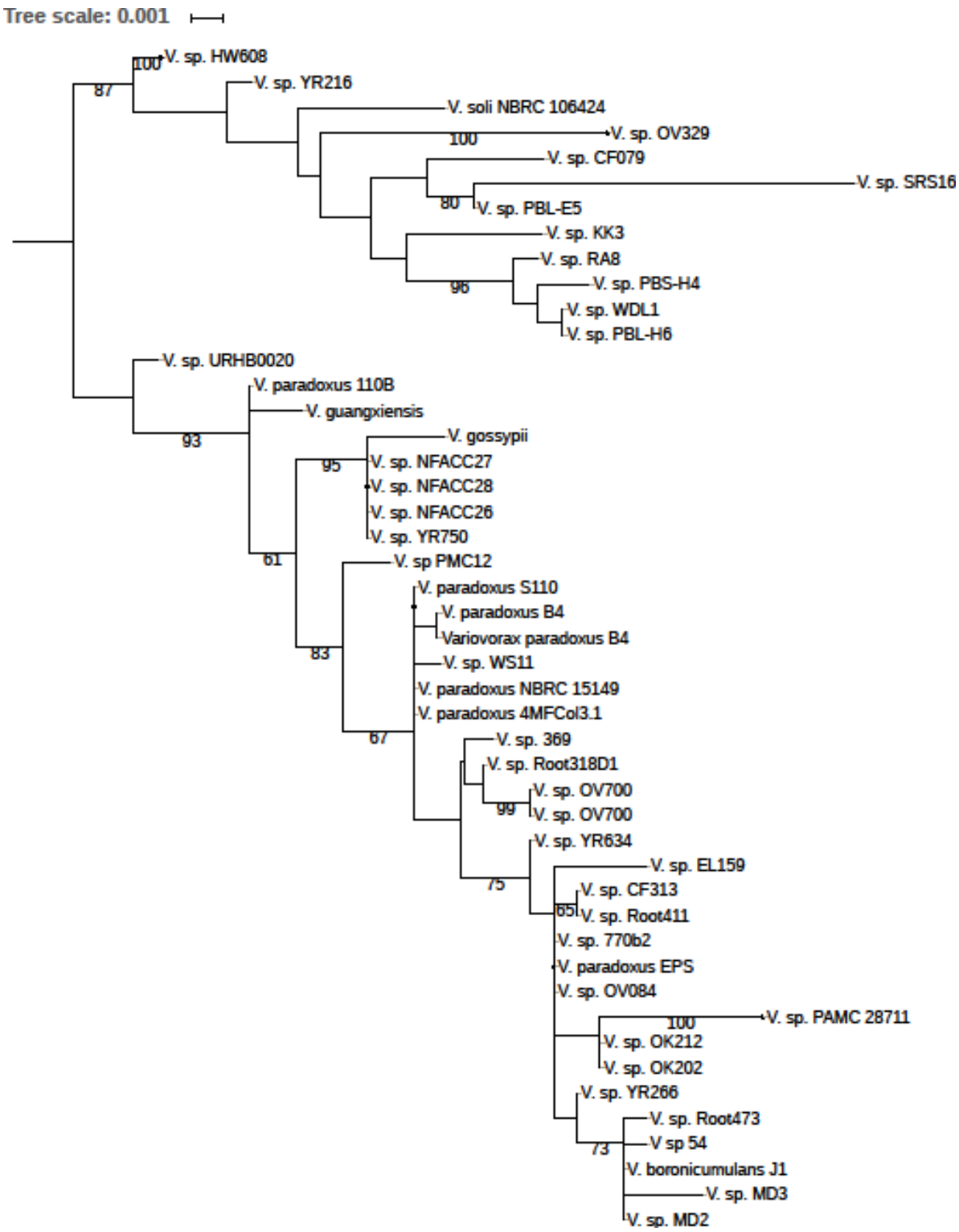


Figure S4: 16S-based phylogeny of Variovorax. ML tree inferred under the GTR+CAT model and rooted by midpoint-rooting. The branches are scaled in terms of the expected number of substitutions per site. The numbers above the branches are support values when larger than 60% from ML bootstrapping.

886 1. Springael D, Smolders E, Horemans B, Hofkens J. 2014. Biofilm formation of a bacterial
887 consortium on linuron at micropollutant concentrations in continuous flow chambers and
888 the impact of dissolved organic matter. FEMS Microbiol Ecol 88:184–194.

889 2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
890 tool. J Mol Biol 215:403–410.

891 3. Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize Implements and enhances circular
892 visualization in R. Bioinformatics 30:2811–2812.

893

894