

# The Whole Genome Sequence of *Sphingobium chlorophenicum* L-1: Insights into the Evolution of the Pentachlorophenol Degradation Pathway

Shelley D. Copley<sup>1,\*</sup>, Joseph Rokicki<sup>1</sup>, Pernilla Turner<sup>2</sup>, Hajnalka Daligault<sup>3</sup>, Matt Nolan<sup>4</sup>, and Miriam Land<sup>5</sup>

<sup>1</sup>Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder

<sup>2</sup>The Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder GeoSynFuels LLC, Golden, Colorado

<sup>3</sup>Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

<sup>4</sup>Joint Genome Institute, Walnut Creek, California

<sup>5</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee

\*Corresponding author: E-mail: shelley.copley@colorado.edu.

**Accepted:** 8 December 2011

**Data deposition:** GenBank accession numbers for the replicons of *S. chlorophenicum* are given in table 1.

## Abstract

*Sphingobium chlorophenicum* Strain L-1 can mineralize the toxic pesticide pentachlorophenol (PCP). We have sequenced the genome of *S. chlorophenicum* Strain L-1. The genome consists of a primary chromosome that encodes most of the genes for core processes, a secondary chromosome that encodes primarily genes that appear to be involved in environmental adaptation, and a small plasmid. The genes responsible for degradation of PCP are found on chromosome 2. We have compared the genomes of *S. chlorophenicum* Strain L-1 and *Sphingobium japonicum*, a closely related Sphingomonad that degrades lindane. Our analysis suggests that the genes encoding the first three enzymes in the PCP degradation pathway were acquired via two different horizontal gene transfer events, and the genes encoding the final two enzymes in the pathway were acquired from the most recent common ancestor of these two bacteria.

**Key words:** horizontal gene transfer, biodegradation, enzyme evolution, pentachlorophenol hydroxylase, tetrachlorohydroquinone dehalogenase, tetrachlorobenzoquinone reductase.

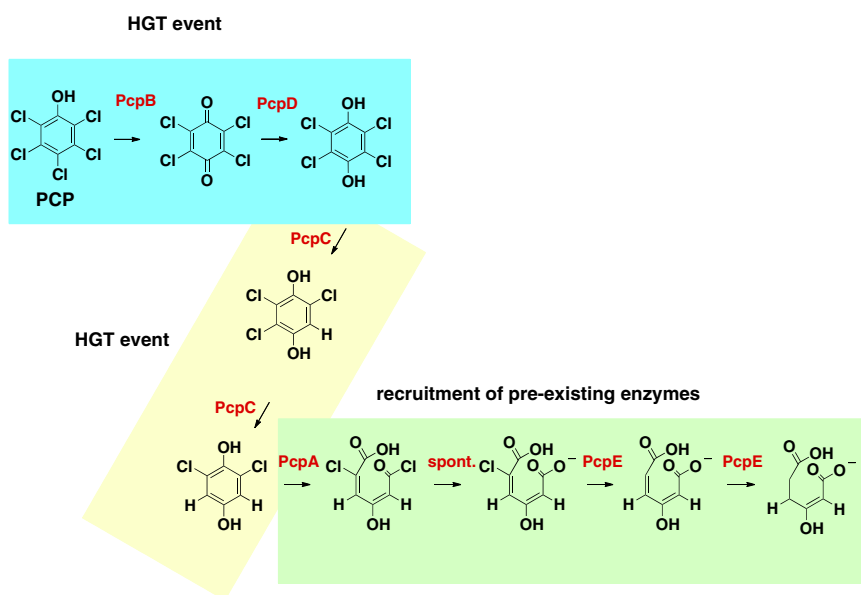
## Introduction

In the early 20th century, consumers embraced the concept of “better living through chemistry.” Massive quantities of chlorinated anthropogenic chemicals such as dichlorodiphenyltrichloroethane (DDT), lindane, pentachlorophenol (PCP), atrazine, tetrachloroethylene, and polychlorinated biphenyls (PCBs) were introduced into the environment with little concern over their ultimate fate. Many of these compounds persist in the environment and have adverse effects on ecosystems. Use of some of these compounds (e.g., DDT and PCP) is now restricted or prohibited in the United States, although some are still used in developing countries.

The introduction of anthropogenic chemicals into the environment imposes selective pressures that can foster

evolution of novel pathways that allow microbes to access new sources of carbon, nitrogen, or phosphorus and/or to detoxify toxic compounds. However, degradation of anthropogenic chemicals is often inefficient because microbes have not yet evolved enzymes that efficiently catalyze the steps needed to convert anthropogenic compounds into metabolites in central carbon metabolism (Copley 2009).

The pesticide PCP is listed as a Priority Pollutant by the US EPA due to its toxicity and persistence in the environment. A number of microbes capable of mineralizing PCP have been isolated from contaminated sites (Saber and Crawford 1985; Apajalahti and Salkinoja-Salonen 1987; Schenk et al. 1990; Uotila et al. 1991; Radehaus 1992; Karn et al. 2010). The best studied of these is *Sphingobium chlorophenicum* (Saber and Crawford 1985; Steiert and Crawford 1986; Crawford



**FIG. 1.**—PCP degradation pathway. PcpB, pentachlorophenol hydroxylase; PcpD, tetrachlorobenzoquinone reductase; PcpC, tetrachlorohydroquinone dehalogenase; PcpA, 2,6-dichlorohydroquinone dehalogenase; PcpE, maleylacetate reductase; HGT, horizontal gene transfer.

and Ederer 1999; Takeuchi et al. 2001). *Sphingobium chlorophenolicum* has patched together a poorly functioning pathway (see fig. 1) that allows mineralization of PCP, although degradation is slow and the bacterium is unable to grow at high concentrations of PCP (Dai and Copley 2004). The first three enzymes in the pathway are unusually ineffective. PCP hydroxylase (PcpB) has a very low  $k_{\text{cat}}$  of  $0.02 \text{ s}^{-1}$  and shows substantial uncoupling; the enzyme frequently fails to hydroxylate the substrate and instead releases  $\text{H}_2\text{O}_2$  as the C4a-hydroperoxyflavin intermediate decomposes (Hlouchova K, Rudolph J, Pietari JMH, Behlen LS, Copley SD, unpublished data). Tetrachlorobenzoquinone (TCBQ) reductase (PcpD) is also rather inefficient, with a turnover rate of only  $0.7 \text{ s}^{-1}$  at  $50 \mu\text{M}$  TCBQ (Dai et al. 2003). Tetrachlorohydroquinone (TCHQ) dehalogenase (PcpC) is subject to profound substrate inhibition; binding of its aromatic substrates to the active site before completion of the catalytic cycle interferes with the final step required to regenerate the free enzyme (Warner and Copley 2007a). These findings are consistent with the hypothesis that the enzymes in the PCP degradation pathway have not yet evolved to the level of function typical of most metabolic enzymes. Thus, *S. chlorophenolicum* provides a window on the process of assembly of a new pathway at a relatively early stage in its evolution.

We recently sequenced the genome of *S. chlorophenolicum* L-1. Analysis of this sequence and comparison with the sequence of the closely related *Sphingobium japonicum*, which degrades lindane (Nagata et al. 2007), provides insights into the origins of the PCP degradation enzymes. The rDNA genes of *S. chlorophenolicum* and *S. japonicum* share 97% identity. The phylogenetic distance between *S. chlorophenolicum* and *S. japonicum* is ideal to facilitate

identification of genes that were present in the most recent common ancestor of these two *Sphingomonads*, as proteins involved in core processes show  $>80\%$  pairwise identities at the amino acid level.

Our analysis suggests that the first three enzymes in the pathway were acquired by *S. chlorophenolicum* by horizontal gene transfer (HGT) after it diverged from *S. japonicum*. In contrast, the last two enzymes in the pathway were present in the most recent common ancestor of *S. chlorophenolicum* and *S. japonicum*. None of the genes encoding the PCP degradation enzymes arose by recent duplication and divergence of genes within *S. chlorophenolicum*. The genes occur in two disparate parts of the genome and have not yet been integrated into a compact and consistently regulated operon.

## Materials and Methods

### Isolation of Genomic DNA from *S. chlorophenolicum* L-1

A sample of *S. chlorophenolicum* L-1 was originally deposited to the American type Culture Collection and designated ATCC 39723. The initial characterization of the enzymes in the PCP degradation pathway was done with the ATCC 39723 strain (Orser and Lange 1994; Copley 2000; Dai et al. 2003; Warner and Copley 2007b). However, the ATCC 39723 strain lost the ability to degrade PCP, so a second sample of *S. chlorophenolicum* L-1 was deposited and designated ATCC 53874. Genomic DNA was isolated from this strain.

A 5 ml culture of ATCC medium 1687 (*Flavobacterium* medium) containing  $5 \mu\text{M}$  PCP was inoculated with a colony from a fresh agar plate of half-strength tryptone soy broth

containing 50  $\mu$ M PCP and incubated with shaking at 30 °C overnight. One milliliter aliquots of cells were harvested by centrifugation at 16,100  $\times$  g for 1 min. Genomic DNA was isolated using the Aquapure Genomic DNA kit (Bio-Rad, Hercules, CA). Cell pellets from 1 ml of culture were resuspended in 300  $\mu$ l Genomic DNA Lysis Solution. The solution was incubated at 80 °C for 5 min to lyse the cells. A solution of RNase A (1.5  $\mu$ l, 4 mg/ml) was added, and tube was inverted 25 times before incubation at 37 °C for 45 min. The temperature of the sample was adjusted to room temperature and 100  $\mu$ l of Protein Precipitation Solution was added. The sample was vortexed for 20 s and then subjected to centrifugation at 13,000  $\times$  g for 3 min to remove cellular debris. The supernatant containing DNA was transferred to a clean 1.5 ml tube, incubated on ice for 5 min, and then subjected to centrifugation again. The supernatant was then transferred to a clean 1.5-ml tube containing 300  $\mu$ l 100% isopropanol to precipitate the DNA. The tube was inverted 50 times. The DNA was pelleted by centrifugation at 13,000  $\times$  g for 1 min. The supernatant was poured off, and the pellet was washed with 300  $\mu$ l 70% ethanol. The ethanol was poured off and residual ethanol removed by draining the tube onto filter paper for 10–15 min. The pellet was dissolved in 100  $\mu$ l 10 mM Tris–HCl, pH 7.5, at 65 °C for 5 min and then at room temperature overnight. The quality of the DNA was checked on a 0.8% agarose 1X TAE gel alongside a 1 kb DNA Extension Ladder (Invitrogen, Carlsbad, CA), and the concentration was measured on a NanoDrop spectrophotometer.

### Genome Sequencing

The draft genome of *S. chlorophenicum* L-1 was generated at the DOE Joint genome Institute (JGI) using a combination of Illumina (Bennett 2004) and 454 technologies (Margulies et al. 2005). Three libraries were constructed and sequenced: an Illumina GAii shotgun library that generated 40,283,193 reads totaling 3,061 Mb; a 454 Titanium standard library that generated 302,660 reads and a paired end 454 library with an average insert size of  $11.26 \pm 2.81$  kb that generated 174,361 reads. In total, 176.8 Mb of 454 data were obtained. General aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. The initial draft assembly contained 79 contigs in one scaffold. The 454 Titanium standard data and the 454 paired end data were assembled with Newbler, version 2.3. The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET, version 0.7.63 (Zerbino 2008), and the consensus sequences were computationally shredded into 1.5 kb shreds. The 454 Newbler consensus shreds, the Illumina VELVET consensus shreds, and the read pairs in the 454 paired end library were integrated using parallel phrap,

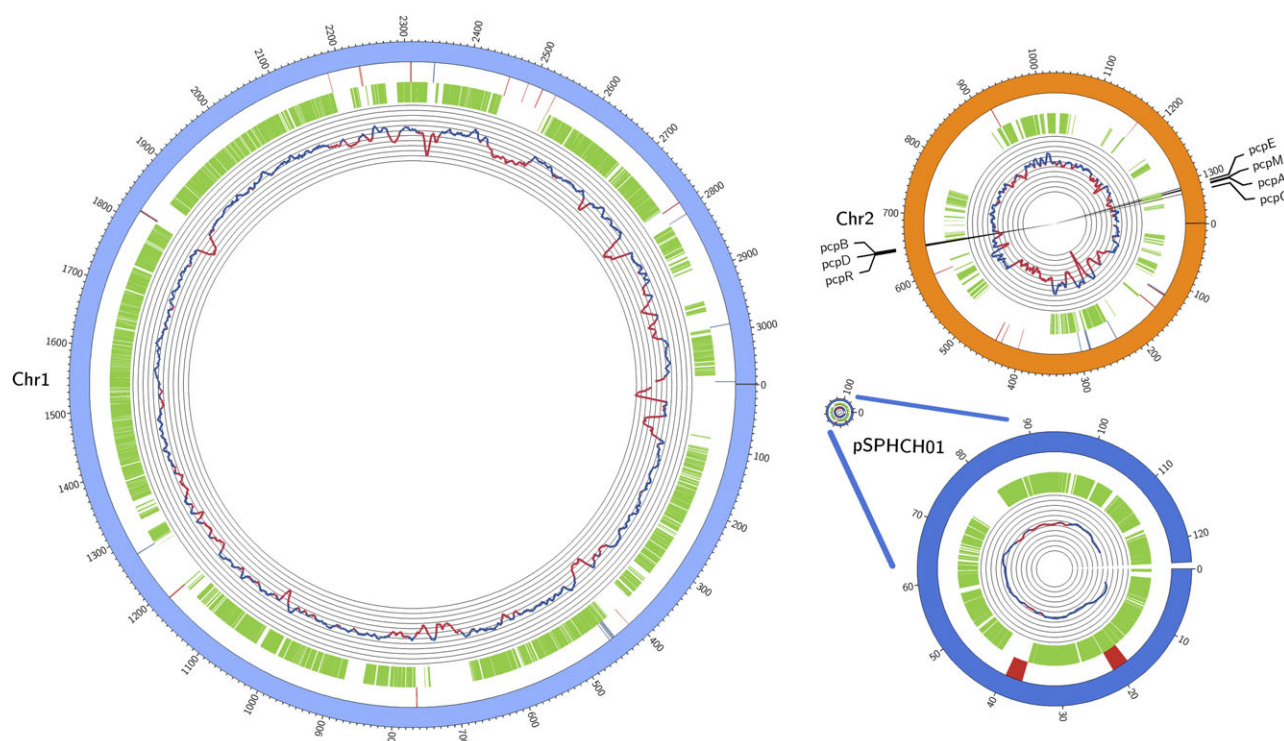
version SPS—4.24 (High Performance Software, LLC). The software Consed (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998) was used in the finishing process. Illumina data were used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Lapidus A, unpublished data). Possible misassemblies were corrected using gapResolution (Han C, unpublished data) or Dupfinisher (Han 2006) or by sequencing cloned bridging PCR fragments. Gaps between contigs were closed by editing in Consed, by PCR, and by Bubble PCR (Cheng J-F, unpublished data) primer walks. A total of 87 additional finishing reactions (either sequencing of bridging PCR fragments or primer walking) were necessary to close gaps and to raise the quality of the finished sequence. The total size of the genome is 4,573,221 bp. The final assembly is based on 176.8 Mb of 454 draft data that provides an average 36 $\times$  coverage of the genome and 3,553 Mb of Illumina draft data that provides an average 790 $\times$  coverage of the genome.

Genes were identified using Prodigal (Hyatt et al. 2010) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline (Pati et al. 2010). The predicted coding sequences were translated and used to search the National Center for Biotechnology Information nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Noncoding genes and miscellaneous features were predicted using tRNAscan-SE (Lowe and Eddy 1997), RNAmmer (Lagesen et al. 2007), Rfam (Griffiths-Jones et al. 2003), TMHMM (Krogh et al. 2001), and signalP (Bendtsen et al. 2004).

## Results and Discussion

### The *S. chlorophenicum* Genome

The *S. chlorophenicum* genome consists of two chromosomes and a plasmid (see fig. 2). A summary of the characteristics of each replicon is provided in table 1. Like many bacteria such as *Burkholderia pseudomallei* (Holden et al. 2004) and *Vibrio cholerae* (Heidelberg et al. 2000) that contain multiple replicons, *S. chlorophenicum* has a dominant chromosome that carries most of the genes for core processes. We generated a list of genes involved in replication, transcription, translation, cell division, peptidoglycan biosynthesis, and core metabolic processes (including glycolysis, the pentose phosphate pathway, the TCA cycle, electron transport, and biosynthesis of amino acids, nucleotides, and cofactors) (see supplementary table 1, Supplementary Material Online). Figure 3 shows the distribution of genes between the two chromosomes. If genes encoding core functions were randomly distributed between the chromosomes, we would expect 354 (72%) to be on chromosome 1



**FIG. 2.**—Circle diagram of the replicons of *Spingobium chlorophenolicum*. The second circle in each replicon indicates the locations of PCP degradation genes in black, phage genes in blue, and transposon-related genes in red. The third circle shows locations of genes that have a close homolog in *Spingobium japonicum* (>80% identity over >90% of the *S. chlorophenolicum* sequence). The fourth circle shows GC content. Red indicates sequences with GC content <64% and blue indicates sequences with GC content >64%. Light gray lines are placed at intervals of one standard deviation from the mean of 63.8%. (One standard deviation = 0.027%; GC content was calculated by averaging over a 10-kb window and sliding that window in 1-kb increments.) Diagram was generated using Circos (Krzywinski et al. 2009).

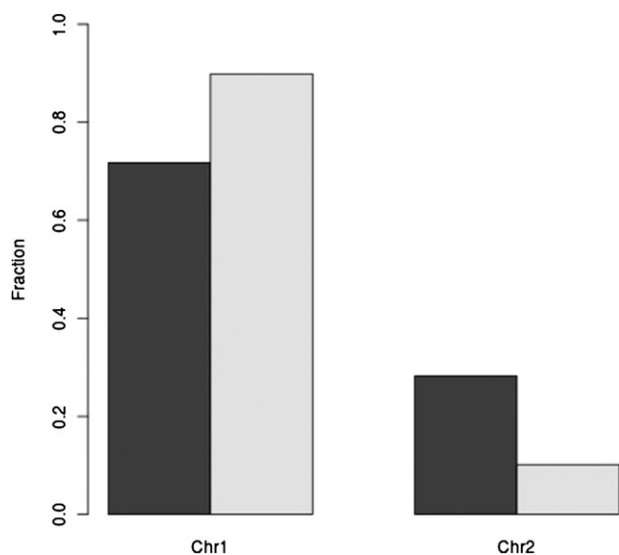
and 138 (28%) to be on chromosome 2. In fact, 442 core genes (90%) are on chromosome 1 and only 50 (10%) are on chromosome 2. This deviation from the expected distribution is highly significant by chi-square analysis ( $P$  value =  $2.2 \times 10^{-16}$ ). An odd collection of essential enzymes, including a few genes for central carbon metabolism, two subunits of RNA polymerase, some components of the electron transfer chain, and a few genes for cofactor biosynthesis are found on chromosome 2 rather than on chromosome 1. However, most of the genes on chromosome 2 appear to be involved in environmental adaptation. Chromosome 2 carries a number of genes predicted to encode transporters for sugars and metal ions and enzymes involved in degradation of various sugars, short-chain fatty acids, and aromatic compounds. All the PCP degradation genes are found on chromosome 2. A secondary chromosome with a low

density of essential genes may serve as a convenient storage depot for genes acquired by HGT, as integration of newly acquired DNA is not likely to disrupt an essential function. Chromosome 2 also contains two genes encoding proteins related to ParB, which is involved in plasmid partitioning. These features of chromosome 2 are consistent with the proposal that bacterial secondary chromosomes have arisen by intragenomic transfer of essential genes, including rRNA genes, to a plasmid (Slater et al. 2009).

HGT is rampant among microbes and is known to play a major role in acquisition of resistance to or degradation of toxic compounds, including antibiotics. We examined the *S. chlorophenolicum* genome for features indicative of mobile genetic elements. As mentioned above, *S. chlorophenolicum* contains one plasmid. The genome appears to contain one integrated prophage; a cluster of phage-related

**Table 1**  
Genome Statistics

	Length (bp)	GC Content	ORFs	rRNA Operons
Chromosome 1 (NC_015593)	3080818	0.64	2,940	1
Chromosome 2 (NC_015594)	1368670	0.64	1,159	2
Plasmid (pSPHCH01) (NC_015595)	123733	0.65	125	0



**FIG. 3.**—Distribution of genes between chromosomes 1 and 2 of *Sphingobium chlorophenicum*. Black bars, fraction of all genes (excluding those on pSPHCH01); gray bars; fraction of core genes.

genes (including a major capsid protein, major tail protein, phage portal protein, pro-head peptidase, and some conserved phage proteins of unknown function) is found on chromosome 1 (genes Sc\_00004260–00004380). Curiously, seven isolated genes annotated as “phage integrase family” genes are found on both chromosomes 1 and 2 (Chr1: Sc\_00026840, Sc\_00029370, Sc\_00022480, Sc\_00028520, Sc\_00012030; Chr 2: Sc\_00031410, Sc\_00030440). These genes have anomalously low GC content (0.48–0.58) and are not closely related to each other. Only one (Sc\_00026840) is found in the vicinity of a prophage gene (a CP4-57 regulatory protein). Two (Sc\_00017260 and Sc\_0003440) are adjacent to transposase genes. The genome also carries 27 sequences annotated as “transposase,” “transposase/integrase core domain,” or “transposase and inactivated derivatives” (16 on chromosome 1, 9 on chromosome 2, and 2 on the plasmid). These proteins are related to transposases found in several different families of insertion elements and in Tn3 family transposons. Thus, like most microbial genomes, the genome of *S. chlorophenicum* displays evidence of continual onslaught by mobile genetic elements. However, the PCP degradation genes show no association with any of these elements.

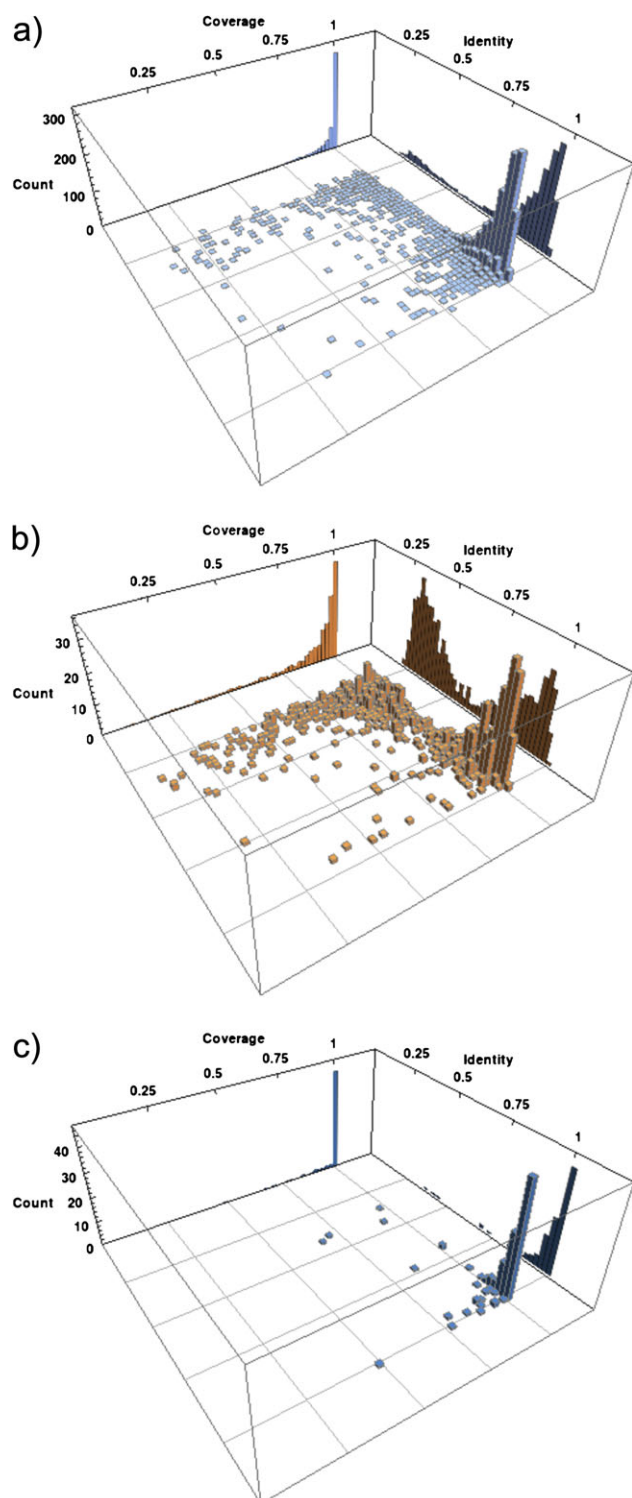
New genes that can help microbes survive and grow in the face of selective pressure from environmental toxins can also arise by gene duplication and divergence (Hughes 1994; Bergthorsson et al. 2007). We carried out a Blast search of the *S. chlorophenicum* genome against itself to identify genes that are nearly identical and may have arisen by recent gene duplication. Only 38 proteins have >90% sequence identity to another protein in *S. chlorophenicum*. Of these, 13 are related to transposases; three

of these genes are found in three identical copies scattered throughout the genome. Although the presence of highly similar transposase genes is not unexpected, there is little rhyme or reason to the identities and locations of the remaining duplicated genes. Seven genes annotated as encoding 2-hydroxychromene-2-carboxylate isomerase, a short-chain alcohol dehydrogenase, a hypothetical protein, a nucleoside-diphosphate-sugar epimerase, an arabinose efflux permease, a Zn-dependent dipeptidase, and an outer membrane receptor protein are present between positions 667462 and 675176 on the—strand of chromosome 2. A second copy of five of these genes are found in a cluster on the + strand of chromosome 2 between positions 1036969 and 1030812. Copies of the first two genes in the cluster are found together in a different location on chromosome 2. Nearly identical copies of the genes encoding the alpha and beta subunits of the E1 component of pyruvate dehydrogenase are found on chromosomes 1 and 2 and nearly identical copies of a gene annotated as an acyl-CoA synthetase/AMP-acid ligase II are found in chromosome 2 but surrounded by completely different genes. The utility of these duplicated genes and the processes leading to their location in distant parts of the genome are not clear. One possibility is that they were not actually duplicated in *S. chlorophenicum* but rather acquired by integration into the genome of DNA fragments taken up from lysed cells of *S. chlorophenicum* or closely related Sphingomonads. Notably, none of the PCP degradation genes is found among the set of highly similar genes.

#### Comparison of the Genomes of *S. chlorophenicum* and *S. japonicum*

The genomes of *S. chlorophenicum* and *S. japonicum* are of similar size (4.57 and 4.46 Mbp, respectively) and contain a similar number of open reading frames (ORFs) (4,159 and 4,460, respectively). Both *S. chlorophenicum* and *S. japonicum* (Nagata et al. 2010) have a primary chromosome containing most of the genes for core processes (including glycolysis, the TCA cycle, amino acid and nucleotide biosynthesis, fatty acid oxidation, DNA replication, transcription, and translation) and a secondary chromosome. *Sphingobium chlorophenicum* has a single plasmid (pSphCh01), whereas *S. japonicum* has three (pUT1, pUT2, and pCHQ1).

The third circle in each chromosome map in figure 2 shows in green the positions of genes that encode proteins in *S. chlorophenicum* that have close homologs in *S. japonicum* that exhibit >80% identity over >90% of the length of the *S. chlorophenicum* sequence. A total of 2,324 *S. chlorophenicum* genes (1,931 on chromosome 1, 285 on chromosome 2, and 108 on pSphCh01) have close homologs in *S. japonicum* (see supplementary table 2, Supplementary Material Online). In both chromosomes, regions with close homologs in *S. japonicum* show a typical GC

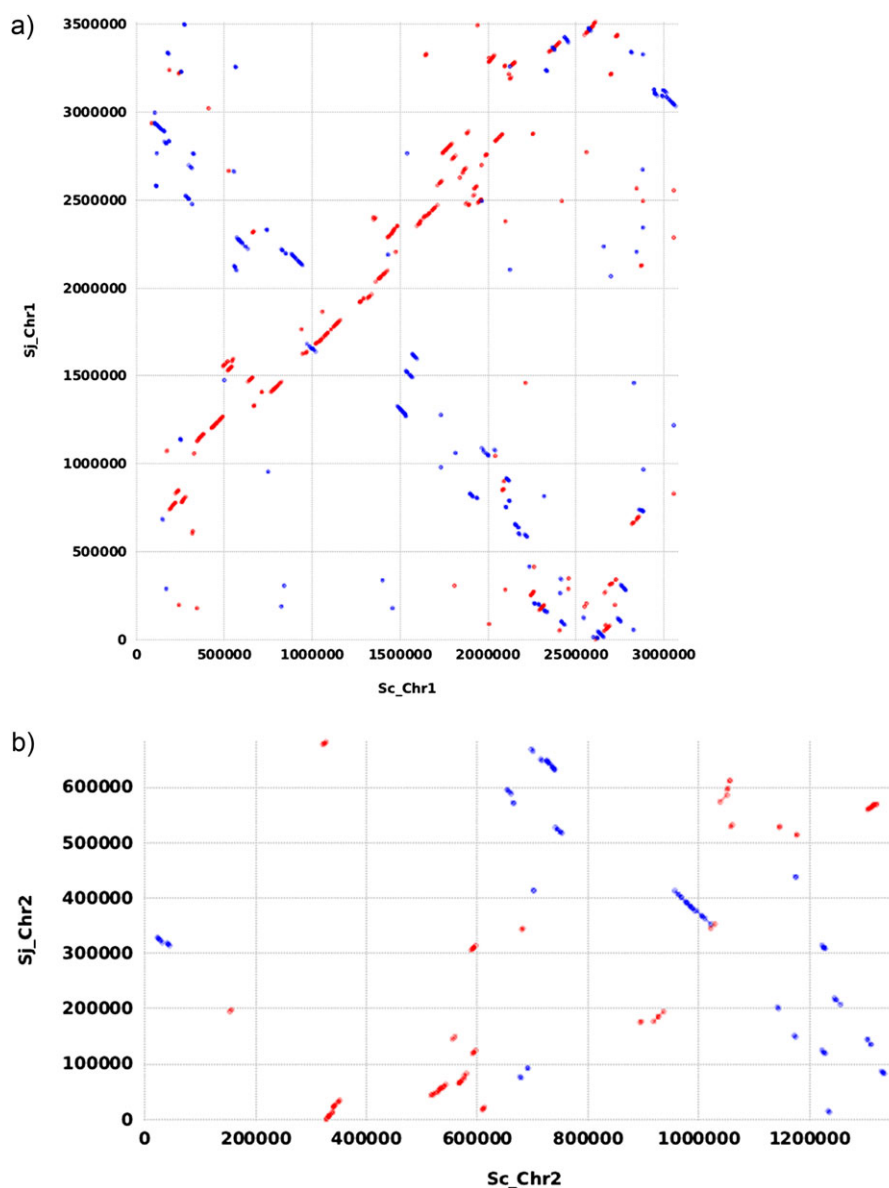


**FIG. 4.**—Relationships between proteins encoded by (a) chromosome 1, (b) chromosome 2, and (c) the plasmid of *Sphingobium chlorophenolicum* and the best hits in the *Sphingobium japonicum* genome found by a Blast search using each *S. chlorophenolicum* protein as a query sequence. The number of pairs is plotted as a function of % identity and % coverage of the query sequence for each replicon. (Only pairs for which  $E < 0.0001$  are shown.)

content of about 64%. Regions without close homologs (Sc islands) might have resulted from either loss of genes in *S. japonicum* or acquisition of genes by HGT in *S. chlorophenolicum*. Some of the Sc islands show a lower GC content, suggesting that these regions may have been acquired by HGT. Most of the genes in the Sc islands are hypothetical proteins, but there are several predicted glycosyltransferases (some predicted to be involved in cell wall biosynthesis), as well as some O-antigen ligases, some ABC transporters, cellobiose phosphorylase, a K<sup>+</sup>-transporting ATPase, and Type IV secretory pathway components. Notably, almost all the sequences associated with transposons and phage genes are found in Sc islands.

A more detailed analysis of the relationships between proteins found in both *S. chlorophenolicum* and *S. japonicum* is shown in figure 4, which shows plots of sequence identity versus coverage for the top hit in the *S. japonicum* genome for each *S. chlorophenolicum* protein. (Coverage is defined as the length of the *S. chlorophenolicum* query sequence that is aligned to a sequence in *S. japonicum* divided by the total length of the query sequence. The data were filtered to remove pairs for which the  $e$  value was  $>0.0001$ .) On this plot, homologs cluster in three regions: 1) close homologs that share high sequence identity over most of the query sequence; 2) more distant homologs that share moderate sequence identity over most of the query sequence; and 3) homologs that share sequence identity only over part of the query sequence. Notably, chromosome 1 is highly enriched in close homologs, and chromosome 2 is modestly enriched in distant homologs. Together with the observation that most of the genes for core metabolic processes are present on chromosome 1, this observation suggests that chromosome 2 may preferentially collect horizontally transferred genes.

Figure 4a shows that most of the proteins encoded on chromosome 1 share  $>80\%$  identity with proteins in *S. japonicum*; indeed, 50% share  $>90\%$  identity. We posit that close homologs with  $>80\%$  identity were present in the most recent common ancestor of these two species; most are likely to be orthologs. The more distant homologs in region 2 are unlikely to be orthologs derived from the most recent common ancestor, since they are much more divergent than the large number of close homologs in region 1. The genes encoding these proteins may have been acquired by HGT independently from different sources in the two species; they may serve the same or different functions. Alternatively, the *S. chlorophenolicum* protein may indeed have derived from the most recent common ancestor, but the ortholog in *S. japonicum* may have been lost so that the best hit in the *S. japonicum* genome is actually a paralog. Finally, there are 112 proteins for which significant sequence identity is seen over only part of the query sequence ( $<75\%$ ); these include proteins in which conserved domains have been utilized in different structural contexts.



**FIG. 5.**—Scatter plot of gene conservation in chromosomes 1 and 2 of *Sphingobium chlorophenolicum* and *Sphingobium japonicum*. Red dots correspond to pairs of ORFs found when the two chromosomes are compared directly and blue dots to pairs of ORFs found when one genome is aligned to the reverse complement of the other.

Although the dominant chromosomes of *S. chlorophenolicum* and *S. japonicum* share a common core of genes, there has been considerable rearrangement of genes since the common ancestor of these two bacteria. Figure 5 shows a scatter plot of gene conservation between the chromosomes of *S. chlorophenolicum* and *S. japonicum* made using nucmer in the MUMmer package with the default parameters (Delcher et al. 1999, 2002; Kurtz et al. 2004). This plot shows a pattern known as an “x-alignment” (Eisen et al. 2000) that is often seen in closely related bacteria. Many genes in chromosome 1 of *S. chlorophenolicum* are found

in comparable positions in chromosome 1 of *S. japonicum*, as indicated by the red dots along the diagonal. However, a number are found in an inverted orientation (see blue dots) in positions that lie close to a diagonal perpendicular to the red diagonal. This pattern is believed to result from multiple inversions centered around either the origin or terminus of replication. There has evidently been considerable remodeling of the genome via movement of blocks of genes within chromosome 1 since *S. chlorophenolicum* and *S. japonicum* diverged from a common ancestor. The x-alignment pattern is less distinct for chromosome 2,

suggesting greater plasticity in this replicon. This is consistent with the lower number of genes for core processes on chromosome 2 described above. Furthermore, only 12 of the 30 genes encoding core processes found on chromosome 2 of *S. chlorophenolicum* are found on chromosome 2 of *S. japonicum*, suggesting that movement of genes between the two chromosomes has continued to occur since the divergence of *S. chlorophenolicum* and *S. japonicum*.

Figure 6 depicts the correspondence between the positions of very close homologs (>90% identity over >80% of the query length) in the entire genomes of *S. chlorophenolicum* and *S. japonicum*. Light blue, orange, and dark blue lines connect the positions of genes in chromosome 1, chromosome 2, and the plasmid, respectively, of *S. chlorophenolicum* with the positions of homologs in the two chromosomes and three plasmids of *S. japonicum*. As noted above, chromosome 1 in both species is densely populated with shared genes. A number of genes found on chromosome 2 of *S. chlorophenolicum* have homologs on chromosome 1 of *S. japonicum*, although the converse is not true. Notably, genes found in *S. chlorophenolicum* but not in *S. japonicum* are more heavily represented on chromosome 2 than on Chromosome 1. Since chromosome 2 appears to carry many of the genes for degradation of organic compounds, this difference may be due to the availability of different carbon sources in the environmental niches occupied by the two bacteria. Notably, the *S. chlorophenolicum* plasmid, pSPHCH01, is comprised of a large region that is syntenic with a region of *S. japonicum* chromosome 1 (with the exception of a few small indels) and a smaller region that is syntenic with a region around the origin of *S. japonicum* pCHQ1. This region encodes several proteins, including two chromosome partitioning proteins (ParA and ParB homologs) and the plasmid replication initiation protein (RepA). Thus, the *S. chlorophenolicum* pSphCh01 and *S. japonicum* pCHQ1 share an origin of replication and the associated genes, but the genes carried on the two plasmids are not closely related. The two smallest plasmids in *S. japonicum* (pUT1 and pUT2) carry genes with no homologs in *S. chlorophenolicum*.

One hundred and eight of 125 genes on pSPHCH01 have >90% identity to genes on plasmid pSLGP in *Sphingobium* SYK-6 (a more distantly related Sphingomonad that is the closest relative of *S. chlorophenolicum* and *S. japonicum* for which a whole genome sequence is available), suggesting that these genes may have been present on a plasmid in the ancestor of *S. chlorophenolicum* and *S. japonicum* and may have been incorporated into chromosome 1 of *S. japonicum* after divergence of the two species. Movement of blocks of genes among the plasmids and chromosomes in these organisms may have been facilitated by transposases, as TN3-family transposase elements are present in both plasmids and in the *S. japonicum* chromosome near one end of the integrated region.

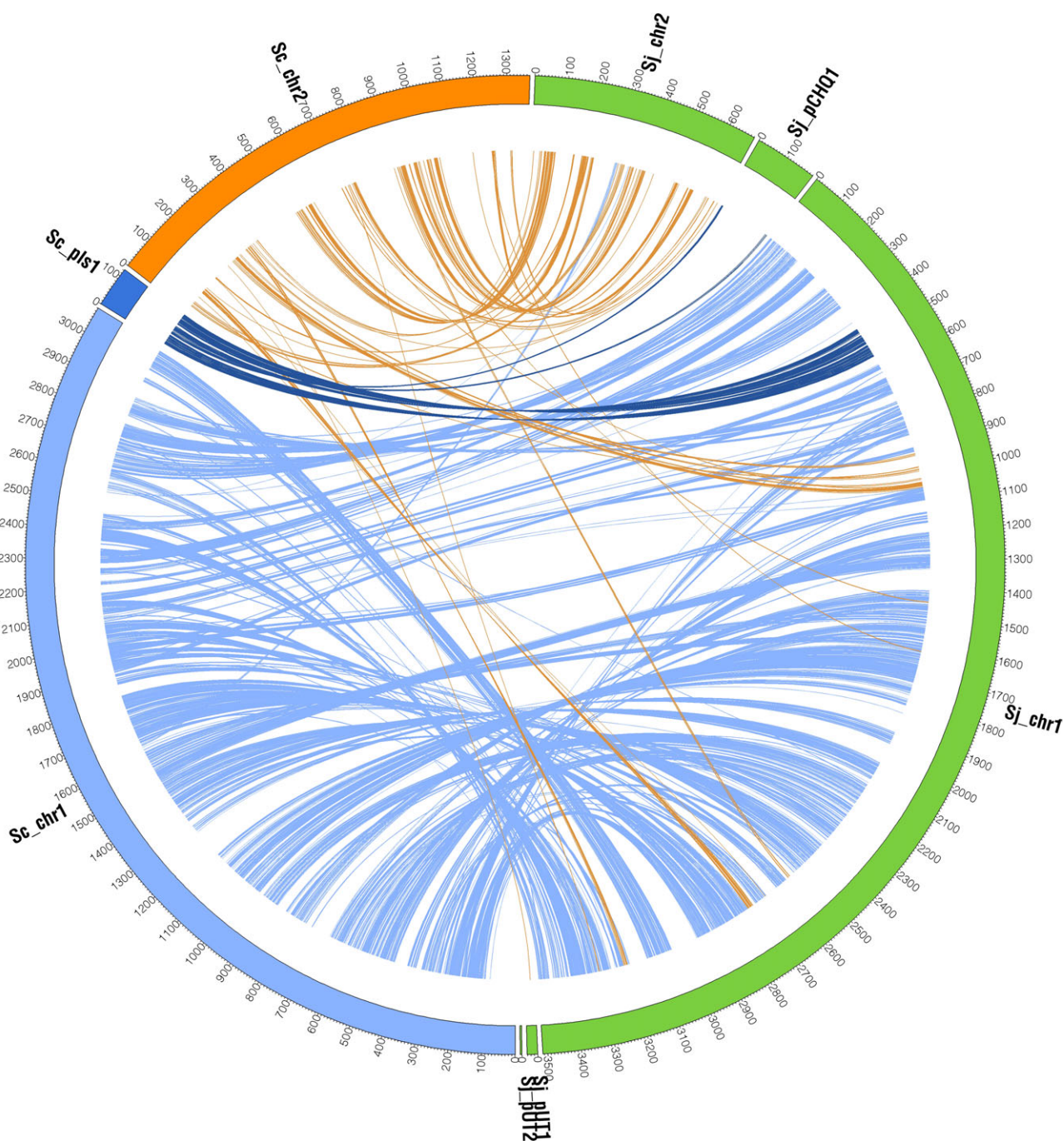
### Insights into the Origin of the PCP Degradation Genes

The PCP degradation genes are found on chromosome 2. *pcpA*, *pcpC*, and *pcpE* are found in proximity to each other, although not within an operon, whereas *pcpB* and *pcpD* are found in an operon on the opposite side of the chromosome (see figs. 2 and 7). Expression of all the genes except *pcpC* (Orser et al. 1993) is induced in the presence of PCP, although it is not known whether the genes are regulated by PCP itself or by downstream products. The constitutive expression of *pcpC* suggests that a mechanism for its regulation by PCP has not yet arisen.

In theory, the genes encoding the PCP degradation enzymes might have originated by 1) recruitment of an ancestral enzyme without gene duplication, requiring sharing of the enzyme between the new and original functions; 2) recruitment of a pre-existing enzyme, followed by duplication and divergence of the original gene to provide one copy that is specialized, although possibly not optimized, for its function in PCP degradation; and 3) HGT of genes encoding either enzymes that have already specialized for their functions in PCP degradation or enzymes with inefficient promiscuous activities that became useful when *S. chlorophenolicum* encountered PCP. Comparison of the genomes of *S. chlorophenolicum* and *S. japonicum* allows us to determine which of these possibilities is most likely for each of the PCP degradation genes. If an enzyme originated by recruitment of an enzyme present in the most recent common ancestor of *S. chlorophenolicum* and *S. japonicum*, then a close homolog is likely to be found in *S. japonicum*. If gene duplication and divergence has resulted in an enzyme specialized for its role in PCP degradation, we should find a close homolog in *S. chlorophenolicum*. If a gene has been acquired by HGT, we would likely find no close homologs in either *S. chlorophenolicum* or *S. japonicum* and might find other signatures of HGT, as well.

PCP hydroxylase (PcpB) and PcpD (TCBQ reductase) are encoded by an operon on chromosome 2 that also includes *pcpR*, which encodes a LysR-transcriptional regulator that has been reported to be essential for induction of expression of *pcpB*, *pcpA*, and *pcpE* (Cai and Xun 2002). PCP hydroxylase belongs to a family of flavin monooxygenases that hydroxylate phenols. Such enzymes are common in soil bacteria because phenolic compounds derived from lignin provide an important source of carbon. The lack of a close homolog in *S. japonicum* (see table 2) suggests that *pcpB* was not present in the most recent common ancestor of *S. chlorophenolicum* and *S. japonicum*. Close homologs of *pcpB* are also not found in *S. wittichii*, so it is unlikely that *pcpB* arose from an ancestral gene that was lost in *S. japonicum*. The low pairwise sequence identities between PCP hydroxylase and other flavin monooxygenases in *S. chlorophenolicum* (see table 2) suggest that *pcpB* did not arise by recent duplication and divergence of a previously existing

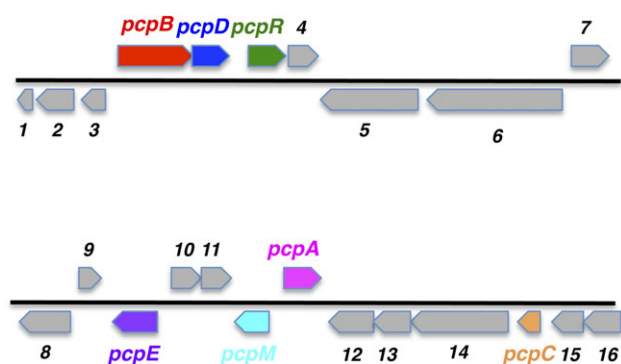




**FIG. 6.**—Plot of gene mappings between replicons of *Sphingobium chlorophenicum* (blue, orange, and dark blue) and *Sphingobium japonicum* (green). Blue, orange, and dark blue lines connect the positions of genes in chromosome 1, chromosome 2, and the plasmid, respectively, of *S. chlorophenicum* with the positions of homologs in *S. japonicum* with >80% identity over >90% of the *S. chlorophenicum* sequence. Diagram was generated using Circos (Krzywinski et al. 2009).

*S. chlorophenicum* gene. These findings, along with the observation that *pcpB* and *pcpD* fall in a region of chromosome 2 with relatively low GC content (see fig. 2), suggest that *pcpB* was acquired by HGT from an unknown source. This hypothesis is supported by the observation that enzymes with 72–98% identity to PcpB are found in *Novos-*

*phingobium lentum* (Tirola, Mannisto, et al. 2002), several polychlorophenol-degrading Sphingomonads from Finland (Tirola, Wang, et al. 2002), *Sphingomonas* sp. UG30 (Cassidy et al. 1999), and in several uncultured bacteria from environmental samples collected from PCP-contaminated soils (Beaulieu et al. 2000). The high sequence



**Fig. 7.**—Gene neighborhoods surrounding the PCP degradation genes. *pcpR* and *pcpM* encode transcriptional regulators that control expression of *pcpBD*, *pcpA*, and *pcpE*. (a) 1, predicted transcriptional regulator; 2 putative NADP-dependent oxidoreductase; 3, hypothetical protein; 4, glycosyl transferase family 2; 5, outer membrane receptor protein, mostly Fe transport; 6, outer membrane receptor for ferrienterochelin and colicins; 7, glucose/sorbose dehydrogenase; (b) 8, outer membrane cobalamin acceptor protein; 9, predicted esterase; 10, formyltetrahydrofolate deformylase; 11, methenyltetrahydrofolate cyclohydrolase/methylene tetrahydrofolate dehydrogenase; 12, Kef-type K<sup>+</sup> transport systems, membrane components; 13, metal-dependent hydrolase,  $\beta$ -lactamase superfamily III; 14, outer membrane receptor proteins, mostly Fe transport; 15, predicted glutathione S-transferase; 16, putative LysR-type transcriptional regulator.

identities between these genes suggest that *pcpB* has been transferred among a number of bacteria in PCP-contaminated environments, a conclusion previously reached by Tirola, Wang, et al. (2002). Strikingly, PcpB has <35% identity to most flavin monooxygenases from non-PCP degrading bacteria. The high level of sequence divergence makes it difficult to discern what the original substrate for PCP hydroxylase might have been before its recruitment into

the PCP degradation pathway. (Careless annotation transfers have led to a proliferation of genes that are annotated as encoding PCP hydroxylase. The proteins encoded by these genes are homologous to PcpB, but have only about 35% identity to PcpB, a level at which substrate specificity cannot be confidently assigned. Furthermore, the ability of the corresponding microbes to degrade PCP has not been assessed. These assignments are probably wrong in the majority of cases.)

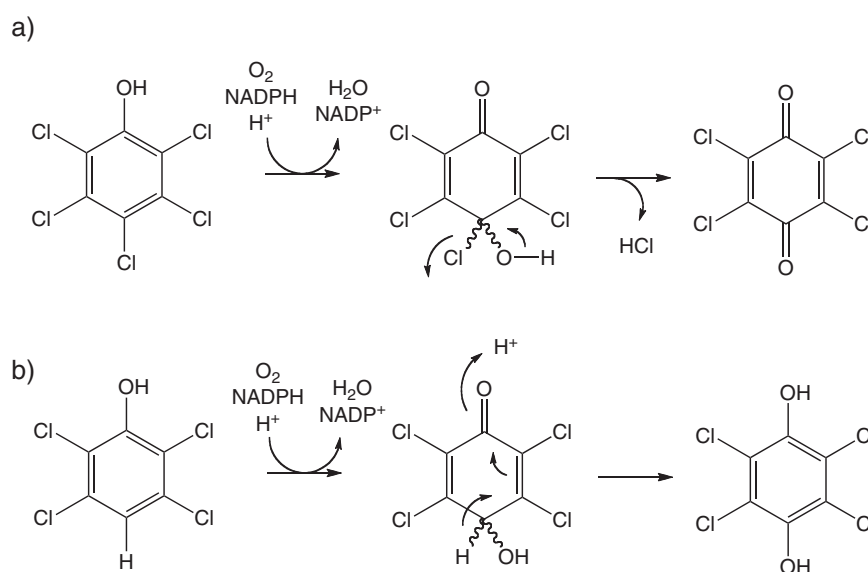
TCBQ reductase (PcpD) catalyzes the second step in the PCP degradation pathway (Dai et al. 2003). Reduction of a benzoquinone is an uncommon step in aromatic degradation pathways. Benzoquinones are formed as a result of hydroxylation of a phenol only when there is a leaving group, such as a chlorine or nitro group, at the position of hydroxylation. Most commonly, there is a hydrogen at the position of hydroxylation, and the product is a hydroquinone (see fig. 8).

TCBQ reductase is not closely related to any other enzyme in *S. chlorophenolicum* or *S. japonicum* (see table 2), suggesting that it, like *pcpB*, was acquired by HGT. Enzymes that carry out comparable reductions of benzoquinones produced by phenol monooxygenases are found in three other microbes, *Pseudomonas* sp. Strain WBC-3 (Zhang et al. 2009), *Alcaligenes* sp. strain NyZ215 (Xiao et al. 2007), and *Cupriavidus necator* JMP134 (formerly *Ralstonia eutropha*) (Belchik and Xun 2008), which degrade *p*-nitrophenol, *o*-nitrophenol, and trichlorophenol, respectively. Notably, these three benzoquinone reductases and TCBQ reductase belong to three different superfamilies (see fig. 9). The recruitment of enzymes from three different superfamilies to catalyze reduction of quinones in pathways for degradation of four different anthropogenic compounds suggests that, in each case, a bacterium has taken

**Table 2**

Closest Homologs of PCP Degradation Genes in *Sphingobium chlorophenolicum*, *Sphingobium japonicum*, and *Sphingobium* sp. SYK-6

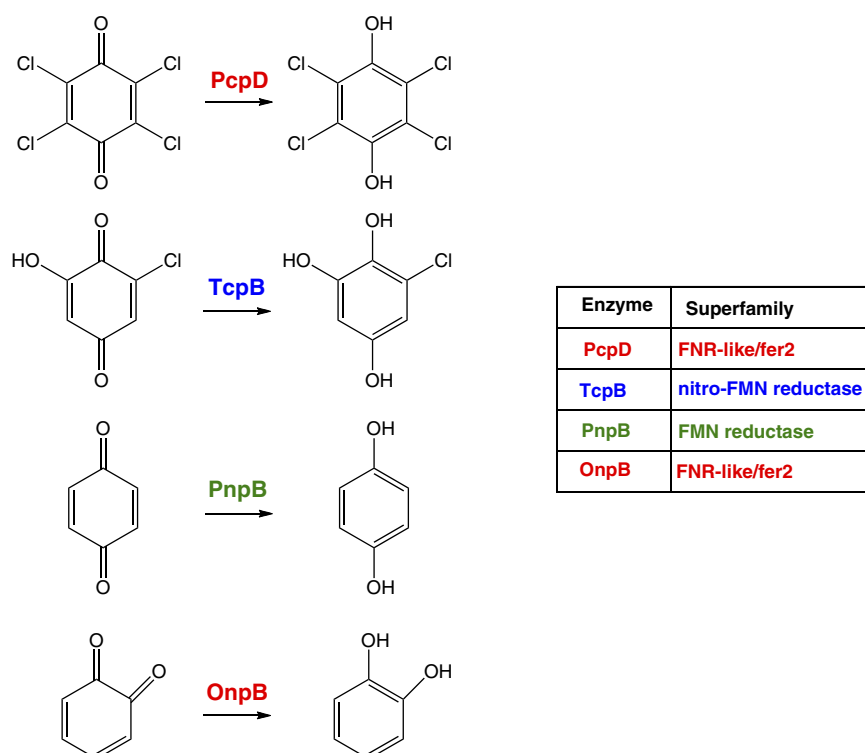
Protein	Closest Homolog in <i>S. chlorophenolicum</i>		Closest Homolog in <i>S. japonicum</i>		Closest Homolog in <i>Sphingobium</i> sp. SYK-6	
	Protein	% Identity	Protein	% Identity	Protein	% Identity
PcpB (gi21362858)	FAD-binding monooxygenase (gi10740769)	28	FAD-binding monooxygenase (gi294146540)	29	<i>p</i> -hydroxy-benzoate hydroxylase (gi347528071)	21
PcpD (gi21362854)	Phthalate 4,5-dioxygenase subunit (gi10723764)	42	Vanillate monooxygenase oxidoreductase subunit (gi294023794)	42	Putative flavodoxin reductase (gi347430572)	27
PcpC (gi 22417110)	Glutathione S-transferase domain-containing protein (gi10724309)	25	2,5-dichlorohydro-quinone dehalogenase (LinD) (gi294023795)	25	Glutathione S-transferase (gi347527108)	25
PcpA (gi 3760223)	None		2,6-dichlorohydro-quinone 1,2-dioxygenase (LinEb) (gi294146914)	93	None	
PcpE (gi 22417101)	Maleylacetate reductase (gi334342570)	53	Maleylacetate reductase (LinF) (gi294146911)	91	Alcohol dehydrogenase (gi347529189)	29



**FIG. 8.**—Products formed by hydroxylation at a position bearing a hydrogen (a) and a chlorine (b).

advantage of a different pre-existing protein when faced with the need to degrade an aromatic compound that carries a leaving group at the site of the initial hydroxylation reaction.

While PCP hydroxylase catalyzes a reaction similar to that of its homologs, the progenitor of TCBQ reductase may have served a different function. Although there are several families of known quinone reductases, TCBQ reductase is not



**FIG. 9.**—Benzoquinone reductases have been recruited from three different superfamilies in microorganisms degrading PCP (*Spingobium chlorophenolicum*), 2,4,6-trichlorophenol (*Cupriavidus necator* JMP134), *p*-nitrophenol (*Pseudomonas* sp. Strain WBC-3), and *o*-nitrophenol (*Alcaligenes* sp. strain NyZ215). The benzoquinones shown (from top to bottom) result from hydroxylation of a phenol at a position bearing a chlorine atom (for PCP and 2,4,6-trichlorophenol) or a nitro group (for *p*- and *o*-nitrophenol).

related to any of these. Rather, TCBQ reductase is most closely related to proteins that serve as the reductase component of two-component dioxygenases that initiate aerobic degradation of aromatic compounds that lack hydroxyl or amine substituents (Dai et al. 2003). These proteins contain a flavin that accepts electrons from NAD(P)H and transfers them one at a time to an iron-sulfur cluster, which then donates electrons to the active site of the oxygenase component of the enzyme. TCBQ reductase is also more distantly related to ferredoxins that restore activity of extradiol dioxygenases that have been inactivated during substrate turnover (Polissi and Harayama 1993; Hugo et al. 1998, 2000; Tropel et al. 2002). (This inactivation occurs after O<sub>2</sub> reacts with Fe(II) at the active site of the enzyme-substrate complex to form superoxide and Fe(III). If the superoxide diffuses out of the active site before reacting with the substrate, the iron atom is left in the inactive Fe(III) state.) The reaction catalyzed by TCBQ reductase could certainly take advantage of this protein architecture; donation of hydride to a flavin, followed by transfer of single electrons to the iron-sulfur cluster and thence to bound TCBQ would form first a semiquinone and then TCHQ after transfer of a second electron. Thus, adaptation of the progenitor protein to this new role may have required little more than enhancement of a small molecule-binding site adjacent to the iron-sulfur cluster.

TCHQ dehalogenase (PcpC) catalyzes the third and fourth steps in the PCP degradation pathway. Each reductive dehalogenation step results in oxidation of two molecules of glutathione to glutathione disulfide. Similar reductive dehalogenation steps occur during lindane degradation in *S. japonicum* (Miyachi et al. 1998) and during degradation of PCBs in *Burkholderia xenovorans* LB400 (Batels et al. 1999; Tocheva et al. 2006). All of these dehalogenases are members of the glutathione *S*-transferase family. Most enzymes in this superfamily catalyze the nucleophilic attack of glutathione upon an electrophilic substrate to form a glutathione conjugate. However, a few, such as TCHQ dehalogenase (Warner et al. 2005; Warner and Copley 2007b), maleylacetoacetate isomerase (Polekhina et al. 2001), and maleylpyruvate isomerase (Marsh et al. 2008), catalyze more complicated reactions in which additional steps occur before and/or after the canonical attack of glutathione upon an electrophilic intermediate.

Although there are numerous members of the glutathione *S*-transferase superfamily in *S. chlorophenolicum*, TCHQ dehalogenase has no more than 25% identity to any of them, so *pcpC* likely did not arise by divergence from another gene in *S. chlorophenolicum*. Close homologs of *pcpC* are not found in the genome of *S. japonicum*, so *pcpC* was probably not present in the ancestor of *S. chlorophenolicum* and *S. japonicum*. The GC content of *pcpC* is 61%, and it falls in a region of chromosome 2 with unusually low GC content. Thus, *pcpC* was most likely acquired by HGT. Strikingly, *S. japonicum* LinD, which cata-

lyzes the reductive dehalogenation of 2,5-dichlorohydroquinone to chlorohydroquinone (Miyachi et al. 1998), has only 25% identity to TCHQ dehalogenase. TCHQ dehalogenase and LinD are the only known reductive dehalogenases in aerobic bacteria that act on aromatic compounds. The high divergence between these enzymes suggests that they were recruited independently from different members of the GST superfamily, possibly due to selective pressure to degrade PCP and lindane, respectively. The function of the progenitors of these enzymes is not clear. Residues in the active sites of TCHQ dehalogenase and LinD resemble those in the active sites of some maleylpyruvate isomerases, and TCHQ dehalogenase has low activity with maleylacetone, an analogue of maleylpyruvate (Anandarajah et al. 2000). Promiscuous activities can provide clues to the activity of an ancestral enzyme, so TCHQ dehalogenase may have originated from a maleylpyruvate isomerase. However, this cannot be stated with certainty.

2,6-Dichlorohydroquinone dioxygenase (PcpA) catalyzes the cleavage of 2,6-dichlorohydroquinone (DCHQ). DCHQ dioxygenase is 93% identical to LinEb in *S. japonicum*, which also cleaves DCHQ. (Despite its name, LinEb is not involved in degradation of lindane [Endo et al. 2005].) Thus, this enzyme was likely present in the most recent common ancestor of *S. chlorophenolicum* and *S. japonicum*. Whether cleavage of DCHQ is the natural function of this enzyme is not clear, as studies of its substrate range and catalytic efficiency have not been reported beyond an early report that shows that hydroquinone and chlorohydroquinone also serve as substrates (Ohtsubo et al. 1999). *Sphingobium japonicum* does not degrade PCP, so the role of LinEb in this bacterium is not known. (*Sphingobium japonicum* contains a homolog of LinEb called LinE that cleaves chlorohydroquinone in the lindane degradation pathway. LinE and LinEb share 53% identity and DCHQ dioxygenase and LinE share 52% identity.)

The final two steps in degradation of PCP are catalyzed by maleylacetate reductase (PcpE), which carries out the successive reductions of 2-chloromaleylacetate to maleylacetate and of maleylacetate to  $\beta$ -keto adipate. PcpE has 91% identity to maleylacetate reductase from *S. japonicum* (LinF), indicating that maleylacetate reductase was found in the progenitor of *S. chlorophenolicum* and *S. japonicum*. Thus, PcpE has been probably recruited to carry out the reduction of 2-chloromaleylacetate in addition to the more typical reduction of maleylacetate. The *S. chlorophenolicum* genome also encodes two other maleylacetate reductase homologs, Sc\_00030750 and Sc\_00027830, which have 53% and 55% pairwise sequence identities to PcpE, respectively. These enzymes have not been characterized, but one or both may play a role in reduction of 2-chloromaleylacetate based upon the observation that substantial 2-chloromaleylacetate reductase activity is present in a strain in which *pcpE* has been knocked out (Cai and Xun 2002).

## Conclusions

Comparison of the genome sequences of *S. chlorophenolicum* and *S. japonicum* suggests that the first three genes in the PCP degradation pathway, *pcpB*, *pcpD*, and *pcpC*, were acquired by HGT. Since these genes were incorporated into two different places in the genome, at least two different HGT were involved. The documented spread of nearly identical *pcpB* genes among bacteria in PCP-contaminated soils suggests that transfer of *pcpB*, and possibly *pcpD* as well, may have occurred within the last century since the introduction of PCP into the environments.

The last two genes in the PCP degradation pathway, *pcpA* and *pcpE*, were inherited from the progenitor of *S. chlorophenolicum* and *S. japonicum*. Pathways for degradation of aromatic compounds typically begin with specialized enzymes that convert the initial compounds to catechol or hydroquinone intermediates that are then cleaved by intradiol or extradiol dioxygenases. (Intradiol dioxygenases cleave catechols between the two hydroxyl groups. Extradiol dioxygenases cleave catechols or hydroquinones adjacent to one of the hydroxyl groups.) The ring cleavage products are then processed to C3 and C4 carboxylates that feed into central metabolism. Thus, numerous “upper” pathways funnel into a small number of “lower” pathways. The enzymes encoded by *pcpA* and *pcpE* likely existed as part of a standard lower pathway for degradation of naturally occurring compounds. The close proximity of *pcpA* and *pcpE* in the genome is consistent with this interpretation.

Duplication and divergence is believed to be the most common mechanism for evolution of enzymes with novel capabilities. This process allows exploration of sequence space while the original activity of the encoded enzyme is provided by a copy that deviates little, if at all, from the original sequence. Ultimately, extraneous copies of the gene are eliminated, leaving one copy encoding the original enzyme and one copy encoding the novel enzyme. Analysis of the *S. chlorophenolicum* genome sequence does not reveal evidence for recent duplication and divergence of any of the PCP degradation genes within *S. chlorophenolicum* itself. Duplication and divergence may have occurred in the unknown soil bacterium in which the genes originated, and only the genes encoding enzymes useful for degradation of PCP may have been transferred to *S. chlorophenolicum* and other PCP-degrading Sphingomonads.

Gene amplification is commonly observed when inefficient enzymatic activity limits the ability of a bacterium to degrade a carbon source. For example, the benzoate degradation genes in a strain of *Acinetobacter* ADP-1 that lacks the transcriptional regulators that normally induce these genes are amplified up to 20-fold when benzoate was supplied as a sole carbon source (Reams and Neidle 2003). Given the inefficiency of the initial enzymes in the PCP degradation pathway, it is intriguing that there is no evidence

for amplification of genomic segments containing the genes encoding these enzymes. In this case, cells carrying duplicated or amplified genes may be at a disadvantage. Amplification of a segment of the genome carrying *pcpB* and *pcpD* would indeed increase flux through the initial two steps but would probably not improve fitness. The waste of NADPH and the concomitant production of H<sub>2</sub>O<sub>2</sub> by the uncoupled PCP hydroxylase reaction would increase in proportion to the enzyme concentration. Additionally, the increased concentration of TCHQ formed would inhibit TCHQ dehalogenase, leading to a buildup of TCHQ, which would be expected to cause redox cycling and depletion of cellular reductants. On the other hand, duplication of a segment of the genome carrying *pcpC* would provide no selective advantage because the initial step, conversion of PCP to TCBCQ, limits the flux through the pathway (McCarthy et al. 1997). Thus, duplication of this region might simply cause a metabolic burden without providing any benefit. If *pcpB*, *pcpD*, and *pcpC* were colocalized, then duplication of all the genes together might be beneficial. However, given the incorporation of *pcpBD* and *pcpC* into two distant regions of chromosome 2, gene duplication does not provide a pathway for further evolution at this point.

## Supplementary Material

Supplementary tables 1 and 2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We thank Prof. Robin Dowell-Dean and Drs Johannes Rudolph and Itamar Yadid for helpful discussions. The contributions of the following toward sequencing, assembling, and annotating the genome are gratefully acknowledged: David Bruce, Chris Detter, Roxanne Tapia, Shunsheng Tan, and Lynne Goodwin (Los Alamos National Laboratory) and James Han, Tanja Woyke, Sam Pitluck, and Len Pennacchio (Joint Genome Institute, Walnut Creek). This work was supported by the National Institutes of Health (GM078554 to S.C.) The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors declare that they have no competing interests.

## Literature Cited

- Anandarajah K, Kiefer PM, Copley SD. 2000. Recruitment of a double bond isomerase to serve as a reductive dehalogenase during biodegradation of pentachlorophenol. *Biochemistry* 39:5303–5311.
- Apajalahti JH, Salkinoja-Salonen MS. 1987. Complete dechlorination of tetrachlorohydroquinone by cell extracts of pentachlorophenol-induced *Rhodococcus chlorophenolicus*. *J Bacteriol.* 169:5125–5130.

- Batels F, Backhaus S, Moore ERB, Timmis KN, Hofer B. 1999. Occurrence and expression of glutathione *S*-transferase-encoding *bphK* genes in *Burkholderia* sp. strain LB400 and other biphenyl-utilizing bacteria. *Microbiology* 145:2821–2834.
- Beaulieu M, Becaert V, Deschenes L, Villemur R. 2000. Evolution of bacterial diversity during enrichment of PCP-degrading activated soils. *Microb Ecol.* 40:345–356.
- Belchik SM, Xun L. 2008. Functions of flavin reductase and quinone reductase in 2,4,6-trichlorophenol degradation by *Cupriavidus necator* JMP134. *J Bacteriol.* 190:1615–1619.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: signalP 3.0. *J Mol Biol.* 340:783–795.
- Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* 5:433–438.
- Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci U S A.* 104:17004–17009.
- Cai M, Xun L. 2002. Organization and regulation of pentachlorophenol-degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. *J Bacteriol.* 184:4672–4680.
- Cassidy MB, Lee H, Trevors JT, Zabolotowicz RB. 1999. Chlorophenol and nitrophenol metabolism by *Sphingomonas* sp UG30. *J Ind Microbiol Biotechnol.* 23:232–241.
- Copley SD. 2000. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem Sci.* 25:261–265.
- Copley SD. 2009. Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat Chem Biol.* 5:559–566.
- Crawford RL, Ederer MM. 1999. Phylogeny of *Sphingomonas* species that degrade pentachlorophenol. *J Ind Microbiol Biotechnol.* 23:320–325.
- Dai M, Copley SD. 2004. Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by *Sphingobium chlorophenolicum* ATCC 39723. *Appl Environ Microbiol.* 70:2391–2397.
- Dai M, Rogers JB, Warner JR, Copley SD. 2003. A previously unrecognized step in pentachlorophenol degradation in *Sphingobium chlorophenolicum* is catalyzed by tetrachlorobenzoquinone reductase (PcpD). *J Bacteriol.* 185:302–310.
- Delcher AL, et al. 1999. Alignment of whole genomes. *Nucleic Acids Res.* 27:2369–2376.
- Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30:2478–2483.
- Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1:RESEARCH0011.
- Endo R, et al. 2005. Identification and characterization of genes involved in the downstream degradation pathway of gamma-hexachlorocyclohexane in *Sphingomonas paucimobilis* UT26. *J Bacteriol.* 187:847–853.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
- Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res.* 31:439–441.
- Han C, Chain, P. 2006. Finishing repeat regions automatically with Dupfinisher. In: HR Arabnia, H Valafar, editors. International conference on bioinformatics & computational biology; 2006 Jun 26–29; Las Vegas, NV. Las Vegas (NV): CSREA Press. p. 141–146.
- Heidelberg JF, et al. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406:477–483.
- Holden MT, et al. 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A.* 101:14240–14245.
- Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci.* 256:119–124.
- Hugo N, Armengaud J, Gaillard J, Timmis KN, Jouanneau Y. 1998. A novel -2Fe-2S- ferredoxin from *Pseudomonas putida* mt2 promotes the reductive reactivation of catechol 2,3-dioxygenase. *J Biol Chem.* 273:9622–9629.
- Hugo N, et al. 2000. Characterization of three XylT-like [2Fe-2S] ferredoxins associated with catabolism of cresols or naphthalene: evidence for their involvement in catechol dioxygenase reactivation. *J Bacteriol.* 182:5580–5585.
- Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
- Karn SK, Chakrabarty SK, Reddy MS. 2010. Pentachlorophenol degradation by *Pseudomonas stutzeri* CL7 in the secondary sludge of pulp and paper mill. *J Environ Sci (China).* 22:1608–1612.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305:567–580.
- Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645.
- Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- Marsh M, et al. 2008. Structure of bacterial glutathione-*S*-transferase maleyl pyruvate isomerase and implications for mechanism of isomerisation. *J Mol Biol.* 384:165–177.
- McCarthy DL, Claude A, Copley SD. 1997. In vivo levels of chlorinated hydroquinones in a pentachlorophenol-degrading bacterium. *Appl Environ Microbiol.* 63:1883–1888.
- Miyachi K, Suh SK, Nagata Y, Takagi M. 1998. Cloning and sequencing of a 2,5-dichlorohydroquinone reductive dehalogenase gene whose product is involved in degradation of gamma-hexachlorocyclohexane by *Sphingomonas paucimobilis*. *J Bacteriol.* 180:1354–1359.
- Nagata Y, Endo R, Ito M, Ohtsubo Y, Tsuda M. 2007. Aerobic degradation of lindane (gamma-hexachlorocyclohexane) in bacteria and its biochemical and molecular basis. *Appl Microbiol Biotechnol.* 76:741–752.
- Nagata Y, et al. 2010. Complete genome sequence of the representative gamma-hexachlorocyclohexane-degrading bacterium *Sphingobium japonicum* UT26. *J Bacteriol.* 192:5852–5853.
- Ohtsubo Y, et al. 1999. PcpA, which is involved in the degradation of pentachlorophenol in *Sphingomonas chlorophenolica* ATCC39723, is a novel type of ring-cleavage dioxygenase. *FEBS Lett.* 459:395–398.
- Orser CS, Lange CC. 1994. Molecular analysis of pentachlorophenol degradation. *Biodegradation* 5:277–288.
- Orser C, Lange CC, Xun L, Zahrt TC, Schneider BJ. 1993. Cloning, sequence analysis, and expression of the *Flavobacterium* pentachlorophenol-4-monooxygenase gene in *Escherichia coli*. *J Bacteriol.* 175:411–416.

- Pati A, et al. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods*. 7:455–457.
- Polekhina G, Board PG, Blackburn AC, Parker MW. 2001. Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity. *Biochemistry* 40:1567–1576.
- Polissi A, Harayama S. 1993. In vivo reactivation of catechol 2,3-dioxygenase mediated by a chloroplast-type ferredoxin: a bacterial strategy to expand the substrate specificity of aromatic degradative pathways. *EMBO J*. 12:3339–3347.
- Radehaus PM, Schmidt SK. 1992. Characterization of a novel *Pseudomonas* sp. that mineralizes high concentrations of pentachlorophenol. *Appl Env Microbiol*. 58:2879–2885.
- Reams AB, Neidle EL. 2003. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol*. 47:1291–1304.
- Saber DL, Crawford RL. 1985. Isolation and characterization of *Flavobacterium* strains that degrade pentachlorophenol. *Appl Environ Microbiol*. 50:1512–1518.
- Schenk T, Müller R, Lingens F. 1990. Mechanism of enzymatic dehalogenation of pentachlorophenol by *Arthrobacter* sp. Strain ATCC 33790. *J Bacteriol*. 172:7272–7274.
- Slater SC, et al. 2009. Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. *J Bacteriol*. 191:2501–2511.
- Steiert JG, Crawford RL. 1986. Catabolism of pentachlorophenol by a *Flavobacterium* sp. *Biochem Biophys Res Commun*. 1986:825–830.
- Takeuchi M, Hamana K, Akira H. 2001. Proposal of the genus *sphingomonas* sensu stricto and three new genera, *sphingobium*, *novosphingobium* and *sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int J Syst Evol Microbiol*. 51:1405–1417.
- Tiirola MA, Mannisto MK, Puhakka JA, Kulomaa MS. 2002. Isolation and characterization of *Novosphingobium* sp. strain MT1, a dominant polychlorophenol-degrading strain in a groundwater bioremediation system. *Appl Environ Microbiol*. 68:173–180.
- Tiirola MA, Wang H, Paulin L, Kulomaa MS. 2002. Evidence for natural horizontal transfer of the *pcpB* gene in the evolution of polychlorophenol-degrading sphingomonads. *Appl Environ Microbiol*. 68:4495–4501.
- Tocheva EI, Fortin PD, Eltis LD, Murphy ME. 2006. Structures of ternary complexes of BphK, a bacterial glutathione S-transferase that reductively dechlorinates polychlorinated biphenyl metabolites. *J Biol Chem*. 281:30933–30940.
- Tropel D, Meyer C, Armengaud J, Jouanneau Y. 2002. Ferredoxin-mediated reactivation of the chlorocatechol 2,3-dioxygenase from *Pseudomonas putida* GJ31. *Arch Microbiol*. 177:345–351.
- Uotila JS, Salkinoja-Salonen MS, Apajalahti JH. 1991. Dechlorination of pentachlorophenol by membrane bound enzymes of *Rhodococcus chlorophenolicus* PCP-I. *Biodegradation* 2:25–31.
- Warner JP, Lawson SL, Copley SD. 2005. A mechanistic investigation of the thiol-disulfide exchange step in the reductive dehalogenation catalyzed by tetrachlorohydroquinone dehalogenase. *Biochemistry* 44:10360–10368.
- Warner JR, Copley SD. 2007a. Mechanism of the severe inhibition of tetrachlorohydroquinone dehalogenase by its aromatic substrates. *Biochemistry* 46:4438–4447.
- Warner JR, Copley SD. 2007b. Pre-steady state kinetic studies of the reductive dehalogenation catalyzed by tetrachlorohydroquinone dehalogenase. *Biochemistry* 46:13211–13222.
- Xiao Y, Zhang JJ, Liu H, Zhou NY. 2007. Molecular characterization of a novel ortho-nitrophenol catabolic gene cluster in *Alcaligenes* sp. strain NyZ215. *J Bacteriol*. 189:6587–6593.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.
- Zhang JJ, Liu H, Xiao Y, Zhang XE, Zhou NY. 2009. Identification and characterization of catabolic para-nitrophenol 4-monooxygenase and para-benzoquinone reductase from *Pseudomonas* sp. strain WBC-3. *J Bacteriol*. 191:2703–2710.

**Associate editor:** Richard Cordaux