Genome Organization of *Sphingobium indicum* B90A: An Archetypal Hexachlorocyclohexane (HCH) Degrading Genotype

Helianthous Verma[1], Abhay Bajaj[1], Roshan Kumar[1], Jasvinder Kaur[1], Shailly Anand[1], Namita Nayyar[1], Akshita Puri[1,2], Yogendra Singh[2], Jitendra P. Khurana[3], and Rup Lal[1,*]

[1]Molecular Biology Laboratory, Department of Zoology, University of Delhi, India

[2]Bacterial Pathogenesis Laboratory, Department of Zoology, University of Delhi, India

[3]Department of Plant Molecular Biology, Interdisciplinary Centre for Plant Genomics, University of Delhi South Campus, India

*Corresponding author: E-mail: ruplal@gmail.com.

Accepted: July 7, 2017

Data deposition: This project has been deposited at NCBI database under the accession CP013070.1, CP013071.1, CP013072.1, and CP013073.1.

Abstract

Among sphingomonads, *Sphingobium indicum* B90A is widely investigated for its ability to degrade a manmade pesticide, γ-hexachlorocyclohexane (γ-HCH) and its isomers (α-, β-, δ-, and ϵ-HCH). In this study, complete genome of strain B90A was constructed using Single Molecule Real Time Sequencing (SMRT) and Illumina platform. The complete genome revealed that strain B90A harbors four replicons: one chromosome (3,654,322 bp) and three plasmids designated as pSRL1 (139,218 bp), pSRL2 (108,430 bp) and pSRL3 (43,761 bp). The study determined the precise location of *lin* genes (genes associated with the degradation of HCH isomers), for example, *linA2*, *linB*, *linDER*, *linF*, *linGHIJ*, and *linKLMN* on the chromosome; *linA1*, *linC*, and *linF* on pSRL1 and *linDEbR* on pSRL3. Strain B90A contained 26 copies of IS*6100* element and most of them (15 copies) was found to be associated with *lin* genes. Duplication of several *lin* genes including *linA*, *linDER*, *linGHIJ*, and *linF* along with two variants of *linE*, that is, *linEa* (hydroquinone 1,2-dioxygenase) and *linEb* (chlorohydroquinone/hydroquinone 1,2-dioxygenase) were identified. This suggests that strain B90A not only possess efficient machinery for upper and lower HCH degradation pathways but it can also act on both hydroquinone and chlorohydroquinone metabolites produced during γ-HCH degradation. Synteny analysis revealed the duplication and transposition of *linA* gene (HCH dehydrochlorinase) between the chromosome and pSRL1, possibly through homologous recombination between adjacent IS*6100* elements. Further, in silico analysis and laboratory experiments revealed that incomplete tyrosine metabolism was responsible for the production of extracellular brown pigment which distinguished strain B90A from other HCH degrading sphingomonads. The precise localization of *lin* genes, and transposable elements (IS*6100*) on different replicons now opens up several experimental avenues to elucidate the functions and regulatory mechanism of *lin* genes acquisition and transfer that were not completely known among the bacterial population inhabiting the HCH contaminated environment.

Key Words: *Sphingobium indicum* B90A, Single Molecule Real Time Sequencing, hexachlorocyclohexane, *lin* genes, IS*6100*.

Introduction

Sphingobium indicum B90A has been extensively studied for the degradation of hexachlorocyclohexane (HCH) isomers (Sahu et al., 1990; Kumari et al., 2002; Lal et al., 2010). Strain B90A was isolated from HCH treated sugarcane rhizosphere soil, Cuttack, India and found to degrade the most recalcitrant HCH isomer, β-HCH (Sahu et al., 1990). Till date many sphingomonads have been isolated that harbor *lin* genes (Nagata et al., 1999; Boltner et al. 2005; Kaur et al., 2013; Tabata et al., 2013). Among these, only strain B90A is

© The Author(s) 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

reported to harbor two copies of *linA* (HCH dehydrochlorinase) with different enantioselectivity towards (+) and (−) enantiomers of α-HCH (Suar et al., 2005). In addition, *linB* gene (haloalkane dehalogenase) of strain B90A mediates an additional step of β-HCH dehalogenation (Sharma et al., 2006). Genetic studies on this strain also provided the first evidence of close association between IS*6100* and *lin* genes (Dogra et al., 2004).

Although, genetics and biochemistry of HCH degradation by strain B90A had been studied in detail (Lal et al., 2010) but, complete genetic repertoire especially the localization of *lin* genes and IS*6100* could not be ascertained. Here, we report the complete genome of *S. indicum* B90A by combining the data obtained from Single Molecule Real Time Sequencing (SMRT) and Illumina platform. As a result, genome was assembled into one chromosome and three plasmids (pSRL1, pSRL2, and pSRL3) and also resolved the controversy of copy number and precise location of *lin* genes and IS*6100* on the replicons. For instance, the analysis revealed the presence of 26 copies of IS*6100* elements instead of 11 (Dogra et al., 2004) or seven (Anand et al. 2012) as reported previously. Further, the data analysis and subsequent laboratory experiments revealed the genetic basis of extracellular brown pigment production by strain B90A which involves accumulation, oxidation, and polymerization of homogentisic acid due to absence of *hmg*A gene (homogentisate dioxygenase) in the complete genome reported in this study.

Materials and Methods

Genome Sequencing and Assembly of the Strain B90A

Genome of *S. indicum* B90A was re-sequenced using third generation SMRT sequencing from Genome Quebec, Canada. For this, a 20 kb library was constructed with genome coverage of ∼390×. The genomic reads were assembled using Hierarchical Genome Assembly Process (HGAP) at Genome Quebec, Canada. In silico gap filling was performed using the combined data from SMRT and Illumina Hiseq platforms (Anand et al. 2012).

Genome Annotation and Pathways Mapping

Functional annotation of strain B90A was performed with Rapid Annotation Subsystem Technology (RAST) server *v* 4.0 (Aziz et al. 2008) using Glimmer (Delcher et al., 2007). Single copy essential genes were identified by AMPHORA2 (Wu and Scott 2012). Insertion sequences were identified using IS finder (Siguier et al., 2006), and Island Viewer3 (Dhillon et al. 2015) was used for the prediction of genomic islands. PHAST online server (Zhou 2011) was used to identify integrated phages. Gene specific to *lin* system and xenobiotic degradation were mapped on B90A replicons using BRIG (Alikhan et al., 2011).

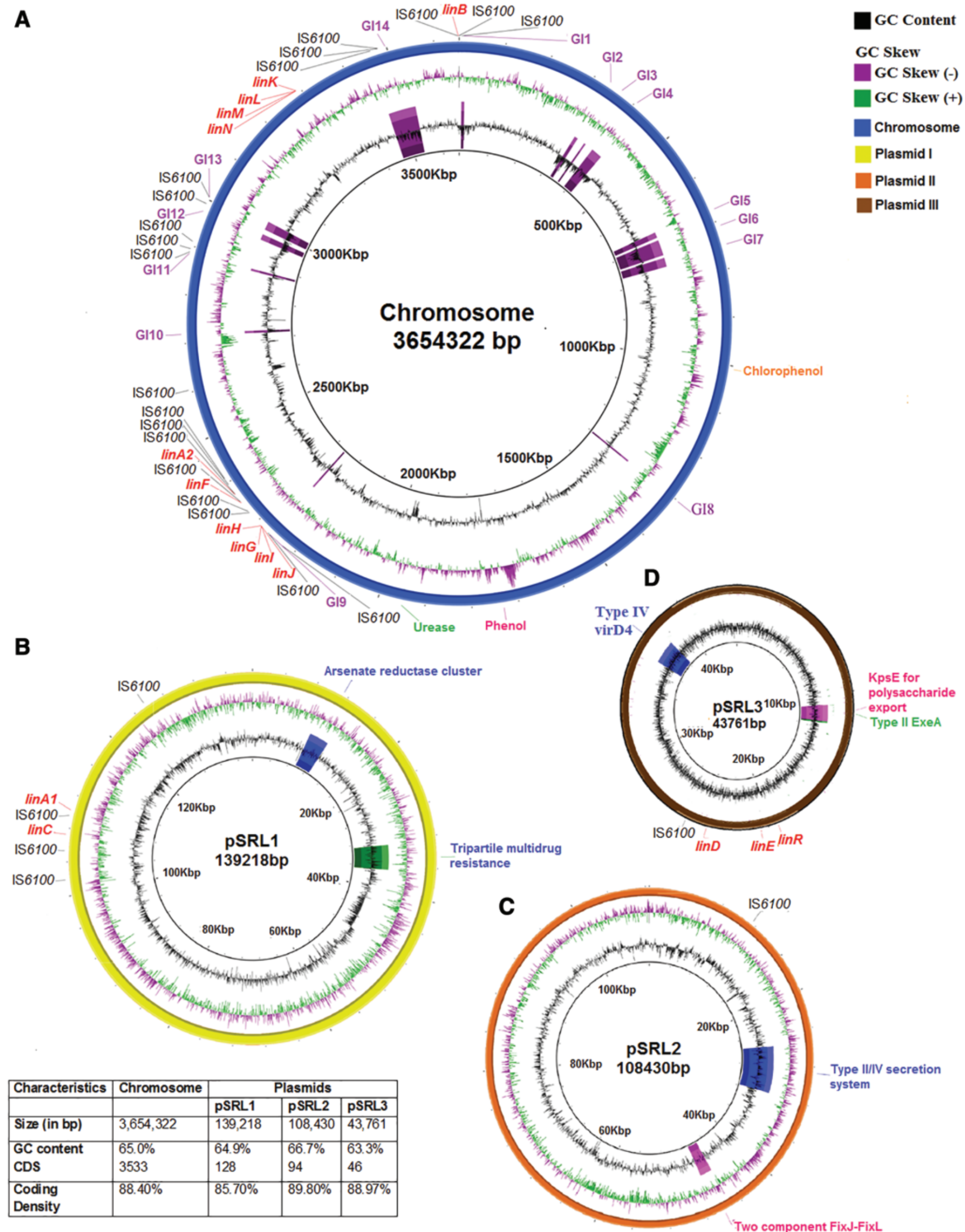
Genome and *lin* System Synteny of HCH Degraders

Pairwise Average Nucleotide Identity (ANI) (Konstantinidis and Tiedje, 2004) between all the replicons of three HCH degraders namely *S. indicum* B90A, *S. japonicum* UT26S and *Sphingomonas* sp. MM1 was calculated. Strain UT26S consists of two chromosomes; Chr-1 and Chr-2 and three plasmids; pCHQ1, pUT1, and pUT2. Strain MM1 possesses only one chromosome and five plasmids; pISP0, pISP1, pISP2, pISP3, and pISP4 (supplementary table 2, Supplementary Material online). These replicons were clustered hierarchically using Pearson correlation method and heat map was constructed on MeV4.9.0 (Saeed et al., 2003). Replicons identity was visualized using CIRCOS (Krzywinski et al., 2009). Synteny mapping of *lin* gene clusters among strains and highly identical replicons was done using annotation from BLASTX (Altschul et al., 1990) and Kablammo server (Wintersinger et al., 2015). Recruitment plot between strain B90A plasmids and Illumina platform reads from HCH dumpsite metagenome was constructed using MUMmer3.23 (Kurtz et al., 2004). Syntenic region between chromosome and pSRL1 were visualized using Microscope server (Vallenet et al., 2009).

Results and Discussion

General Genome Features of *Sphingobium indicum* B90A Replicons

The data obtained from SMRT (390X) and Illumina Hiseq (2000X) (Anand et al. 2012) were used to assemble the genome into four replicons: one chromosome (Size: 3,654,322 bp, CDS: 3533, coding density: 88.40%) and three plasmids namely, pSRL1 (Length: 139,218 bp, CDS: 128, coding density: 85.70%), pSRL2 (Length: 108,430 bp, CDS: 128, coding density: 89.90%) and pSRL3 (Length: 43,761 bp, CDS: 46, coding density: 88.97%) (fig. 1). The detection of 31 essential single copy genes indicated that the genome is complete (Wu and Scott 2012). Previously, the draft genome of strain B90A was reported to consist 149 contigs (Anand et al. 2012), with genome size of 4.08 Mb and 65.0% GC content. However, the present study revealed a complete genome size of 3.9 Mb and GC content of 65.0%. These attributes are within the range of sphingomonad genomes (Genome size: 3.4–5.9 Mb, GC content: 62–68%) (Aylward et al., 2013, Verma et al., 2014; Kumar et al., 2017). GC content of plasmids pSRL1, pSRL2, and pSRL3 were 64.9%, 66.7%, and 63.32%, respectively. Almost nearly equal GC content of plasmids and chromosome of strain B90A suggested that they have been acquired either from sphingomonads or closely related species of sphingomonads and should be stable (Nishida 2012). A total of 3602 coding genes were annotated in the complete genome in which 33% codes for hypothetical proteins. Thus, the function of nearly one-third genes is yet to be ascertained.



Characteristics	Chromosome	Plasmids		
		pSRL1	pSRL2	pSRL3
Size (in bp)	3,654,322	139,218	108,430	43,761
GC content CDS	65.0% 3533	64.9% 128	66.7% 94	63.3% 46
Coding Density	88.40%	85.70%	89.80%	88.97%

FIG. 1.—Circular plot of genetic elements of *Sphingobium indicum* B90A (*A*) Chromosome, (*B*) plasmid pSRL1, (*C*) plasmid pSRL2, and (*D*) plasmid pSRL3. From inside to outside; first, second, and third circle show GC content, GC skew, and replicon, respectively. All *lin* pathway genes and IS*6100* elements were mapped over replicons along with Type II/IV secretion pathway and two component system.

In strain B90A, 26 copies of IS*6100* element were identified, of which 20 were present on chromosome, four on pSRL1 and one each on pSRL2 and pSRL3. So far this number is maximum among the reported range from five in *S. japonicum* UT26S (Nagata et al., 2011) to 24 in *S. quisquilarum* P25 (Verma et al., 2014). In the previous studies based on southern blot hybridization and draft genome, only 11 (Dogra et al., 2004) and seven (Anand

et al. 2012) copies of IS*6100* were reported, respectively. This discrepancy might be due to the low resolution of the DNA bands obtained from southern hybridization. Moreover, in the draft genome, these insertion sequences might have been missed as repeat regions are more likely to be mis-assembled in de-novo assembly (Salzberg and Yorke 2005). Nevertheless, the role of IS*6100* in genome shuffling is well known (Sangwan et al., 2014) and high copy number reflected ongoing shuffling and rearrangements in B90A genome especially in relation to *lin* genes.

Interestingly, along with *linA1* and *linA2* (will be discussed in later section), strain B90A also contained two variants of *linE*, that is, *linEa* (chromosome) and *linEb* (pSRL3) which differed by 53% and encode for hydroquinone 1,2-dioxygenase and chlorohydroquinone/hydroquinone 1,2-dioxygenase, respectively (supplementary table 1, Supplementary Material online). This demarcates its ability to metabolize both chlorohydroquinone and hydroquinone formed during degradation of γ-HCH (Nagata et al., 1999). Also, duplication of *linD*, *linR*, *linF*, *and linGHIJ* (supplementary table 2, Supplementary Material online) with *linE* variants suggested an efficient lower HCH degradation pathway.

Genomic Islands (GIs) and Integrated Phage

In strain B90A and other HCH degrading sphingomonads (Kaur et al., 2013; Negi et al. 2014) accessory genome, especially genomic islands were not described precisely. Genomic islands were predicted using SIGIHMM, IslandPick and IslandPath-DIMOB, and GIs identified by atleast two of the methods were considered. A total of 14 genomic islands (GIs) were predicted and designated as GI1-GI14 (fig. 1). Nitrite-sensitive transcriptional repressor NsrR and nitric oxide dioxygenase (NOD) were identified in GI3. NsrR and NOD are essentially needed for endurance against toxic nitric oxide produced by denitrifying bacteria in the rhizosphere by converting it into nitrous oxide (N_2O) or nitrate (NO_3^-) (Tucker et al., 2008). In GI5 and GI9, genes for type IV secretion (T4SS) system were identified that are known to mediate the extracellular transport of virulence factors (Yeo and Waksman 2004). Out of 11 genes of virB-virD system (Yeo and Waksman 2004), all except *virB2* and *virB7* were found in strain B90A However, *traC* and *traI*, homologs of *virB2* and *virB7* were present in GI6 which suggest complete set of virulence factors are present highlighting its role in combating the stress in rhizospheric environment.

A complete integrated prophage in chromosome (Location: 1255173–1303366 bp) was identified. It contained 67 CDSs in which nearly 50% codes for phage-associated proteins while the other codes for hypothetical proteins. Integration and maintenance of a phage by strain B90A might be an adaptation against environmental stress (Burrus et al. 2004). But the precise role of this phage in the genome of strain B90A is yet to be explored due to the hypothetical proteins.

Evidence of *linA* (HCH Dehydrochlorinase*)* Duplication: Origin of *linA1* from *linA2*

The *linA* gene that encodes for HCH dehydrochlorinase holds its importance in the degradation of HCH isomers as it mediate the first step of dehydrochlorination of α-, γ-, δ-, and ε-HCH (Lal et al., 2010). But, its origin and acquisition is still not known. For instance, unlike HCH degrading sphingomonads, that have been reported to contain a single copy of *linA* gene, strain B90A harbor two variants, *linA1* and *linA2* with enantioselective preference towards (+) and (−) α-HCH (Suar et al., 2005), respectively. The *linA* gene of strain UT26S (Nagata et al. 2001) is 100% identical to *linA2* of strain B90A (Lal et al., 2010; Verma et al., 2014). Previous study by Sangwan et al., (2014) revealed that *linA* is environment specific and was even absent in the last common ancestor of *Sphingobium* spp. Exceptionally, *linA1* has only been reported in strain B90A (Kumari et al., 2002; Suar et al. 2004; Boltner et al. 2005; Ceremonie et al. 2006; Malhotra et al., 2007). *linA1* and *linA2* were identified on pSRL1 and chromosome, respectively, which were aligned to fetch the syntenic regions between the replicons. The analysis showed that flanking sequences of *linA1* (upstream) and *linA2* (downstream) possess sequence similarity of ~15 kb (with >80% identity) (supplementary fig. 1, Supplementary Material online). In addition, 22 bp at C-terminus of *linA1* and N-terminus of IS*6100* were identical, as reported previously (Dogra et al., 2004). Along with the oppositely oriented IS*6100* elements, this 15 kb region reflects the possibility of homologous recombination event resulting into duplication of *linA* in strain B90A. This is further supported by the previous study in which under stress condition, gene that confer adaptation against stress was duplicated between replicons of proteobacterial strains (Slater et al. 2009). Further, genes on the plasmid were found more prone to the sequence variation as compared to the chromosome (Slater et al. 2009). Hence, this suggests that *linA2* was perhaps acquired initially by sphingomonads under HCH stress followed by its duplication and sequence divergence as *linA1*. This gene might have been selected during evolution as it not only enhanced the degradation ability of strain B90A but also possess enantioselectivity towards (+) enantiomer of α-HCH as ancient known *linA* or *linA2* prefers (−)α-HCH as the substrate (Suar et al., 2005). Hence, sequence identity between the flanking regions of both the *linA* and multiple copies of IS*6100* provides an evidence of *linA* duplication followed by sequence variation in *linA1* which led to the current scenario of strain B90A genome. Further, this data and metagenome of HCH dumpsite (Sangwan et al. 2014, Shrivastava et al. 2015) suggest that *linA* is very recent and evolving at a rate beyond scientific imagination.

Replicon Based Identity among HCH Degraders

Due to the absence of complete genome, comparative studies on HCH degrading sphingomonads were limited to the draft

genomes (Verma et al., 2014; Pearce et al. 2015; Kumar et al., 2017) hence, could not infer replicon based phylogeny. Here, replicons of *S. indicum* B90A (fig. 1) and two other HCH degraders with complete genomes, *S. japonicum* UT26S (Nagata et al., 2011), *Sphingomonas* sp. MM1 (Tabata et al., 2013) (supplementary table 2, Supplementary Material online) were clustered hierarchically with Pearson Correlation method using average nucleotide identiy (ANI) (fig. 2*B*, Supplementary Material online). The replicons were clustered into two major clades of which one clade consists of all chromosomes with pISP0 and pSRL2, while the other clustered the remaining nine plasmids. Interestingly, both pSRL1 and pSRL2 shared high ANI ($\geq$90%) with chromosome 2 of strain UT26S (fig. 2*A*). Previously, many species belonging to the phylum proteobacteria have been reported to mediate inter-genomic transfers for the construction of secondary chromosome from primary chromosome and plasmids (Slater et al. 2009). This possibility cannot be ruled out as plasmids of strain B90A (pSRL1 and pSRL2) and chromosome 2 of strain UT26S showed high identity which provides a sign of probable inter-genomic transfer between replicons of B90A and UT26S.

Further, pairwise ANI values of whole genome (chromosome/s and plasmids) showed that both strains B90A and UT26S shared ~83% identity with strain MM1, 5% more than their chromosomal identity, that is, ~78%. Generally, plasmids due to their role in horizontal gene transfer (HGT) are considered as the vehicles of genome variability (Bergstrom et al. 2000; Gogarten et al. 2002). But here, inclusion of plasmid sequences has increased the identity between the whole genome of HCH degraders which highlights their acquisition by the last common ancestor (LCA). High identity (~90%) between pISP3 of strain MM1 and pSRL3 of strain B90A which harbored *linD* (reductive dechlorinase), *linE* (ring cleavage oxygenase), *linR* (transcriptional regulator) and an IS*6100* element (supplementary fig. 2*A*, Supplementary Material online) further supports acquisition of pSRL3 or pISP3 by LCA of HCH degraders. Metagenomic recruitment plot of pSRL3 using the data obtained from the HCH dumpsite, India, also determined the co-occurrence of this plasmid at the site from where different HCH degrading sphingomonads have been isolated including *Sphingomonas* sp. MM1 (Tabata et al. 2013). Also, this suggest the active role of pSRL3 in acquisition of lower HCH degradation pathway in sphingomonads (supplementary fig. 2*B*, Supplementary Material online).

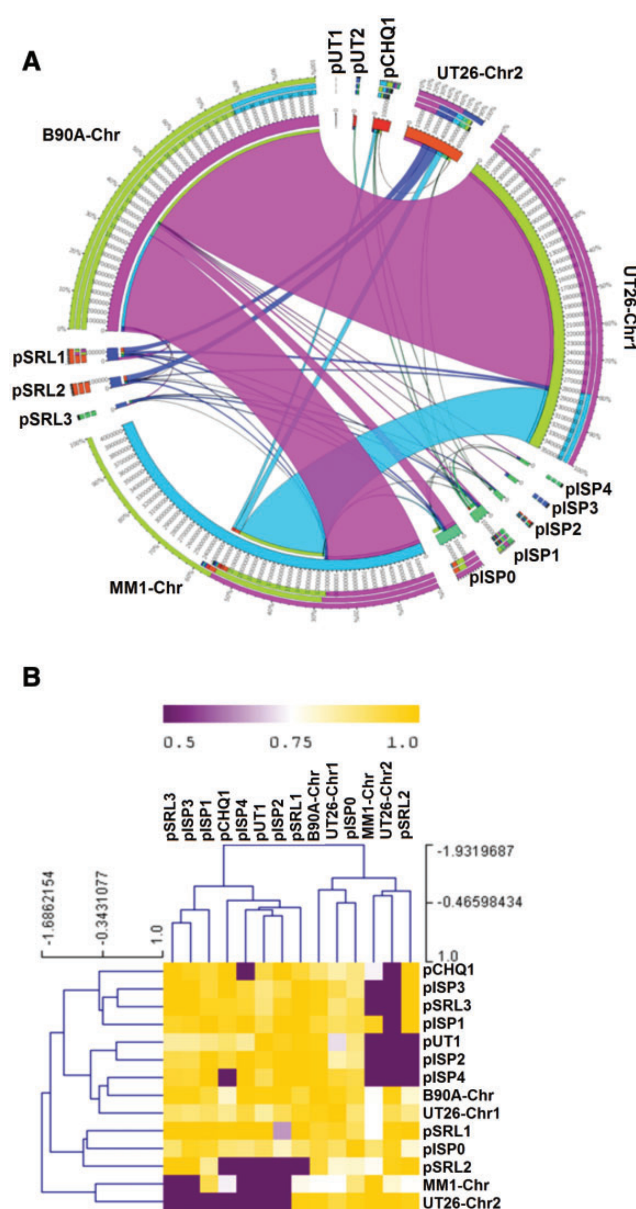


Fig. 2.—Comparison between genetic elements of *Sphingobium indicum* B90A, *Sphingobium japonicum* UT26S and *Sphingomonas* sp. MM1. (*A*) Similarity plot with B90A genetic elements: B90A-CHR (chromosome), pSRL1, pSRL2, and pSRL3; UT26 genetic elements: UT26-CHR1 (chromosome 1), UT26-CHR2 (chromosome 2), pCHQ1, pUT1, and pUT2; MM1 genetic elements: MM1-CHR (chromosome), pISP0, pISP1, pISP2, PISP3, and pISP4. (*B*) Heat map on pairwise Average Nucleotide Identity between the replicons (chromosomes and plasmids) of the strains.

lin Gene Clusters and IS Elements: Southern Hybridization versus SMRT Sequence Analysis

Previously, copy number of *lin* genes and IS*6100* in strain B90A, were predicted by southern hybridization which reported the presence of at least 11 copies of IS*6100* element (Dogra et al. 2004) while 26 copies of IS*6100* were identified in the complete genome. To resolve this ambiguity, in silico restriction mapping was done using *BamHI* and compared with the previous gel electrophoresis profile (Dogra et al. 2004) (supplementary fig. 3, Supplementary Material online). In silico profile predicted 21 fragments (containing one or more IS*6100* elements) instead of 11 fragments obtained previously on the gel electrophoresis. A similar comparison was drawn for the *lin* genes using previous studies (Kumari et al.

2002; Dogra et al. 2004; Malhotra et al. 2007) and in silico restriction digestion with *BclI*, *HindIII*, *BamHI*, and *BglII* (supplementary fig. 4 and table 3, Supplementary Material online). Here, in silico and agarose gel profiles of *lin* genes were similar which might corresponds to less ambiguity due to the presence of one or two copies of *lin* genes as compared to 26 copies of IS*6100*.

Pigment Biosynthesis by Strain B90A

Production of extracellular brown pigment is a characteristic feature of strain B90A (Sahu et al., 1990) that was not observed in other HCH degrading sphingomonad. Genome analysis of strain B90A revealed the presence of gene cluster encoding incomplete pathway for tyrosine and phenylalanine metabolism. It was also found from the complete genome analysis of strain B90A that this strain lacked homogentisate dioxygenase gene, *hmg*A of the pathway thereby making it incapable of utilizing homogentisic acid (HGA) as substrate. HGA is a stable metabolite formed during metabolism of tyrosine and phenylalanine. Recently, Wang et al., (2015) reported that in *Aeromonas* spp. due to absence of *hmg*A gene, HGA got accumulated and further converted to pyomelanin like brown pigment. Genome analysis of strain B90A revealed the presence of four essential genes for biosynthesis of HGA such as, *phhA* (phenylalanine-4-hydroxylase), *phh*B (4-alpha-carbinolamine dehydratase), *hpp*D (4-hydroxyphenyl pyruvate dioxygenase), and *asp*C (aspartate aminotransferase). In general, this pathway further extends with *hmgABC* that catalyzes the conversion of homogentisic acid into fumarate and acetoacetate. But due to the absence of *hmgA* gene in B90A, HGA accumulated and underwent auto-oxidation and polymerization resulting in pyomelanin production (supplementary fig. 5, Supplementary Material online). Further, comparative in silico annotation using the functional genes involved in HGA based pigment synthesis reported previously from *P. aeruginosa* UCBPPPA14 (Hunter and Newman, 2010) as reference; nearly 16 functional genes were identified in the genome of strain B90A (supplementary table 4, Supplementary Material online). The analysis also revealed the absence of tyrosinase gene in B90A genome indicating that the synthesis of pigment is not through the oxidation of tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) and dopaquinone (Plonka and Grabacka 2006). This suggested that strain B90A does not synthesize eumelanin or phaeomelanin like pigment (Donlon et al., 1983). Also, genes encoding laccases and polyketide synthases which may contribute to melanin synthesis were not identified in the genome of strain B90A. This confirmed that HGA biosynthesis in strain B90A is responsible for the production of extracellular brown pigment. Moreover, comparative analysis of tyrosine metabolism has also revealed that phylogenetically related strains of *Sphingobium* that harbored *hmgA* gene do not produce brown pigment.

Conclusion

The study revealed the relevance of SMRT technique in constructing the complete genome of *Sphingobium indicum* B90A and has resolved several issues and ambiguity that existed previously. The number of plasmids and chromosomes in strain B90A was determined. Strain B90A consisted of four replicons, one chromosome and three plasmids; pSRL1, pSRL2, and pSRL3. The analysis determined the precise location of *lin* genes on the replicons, in order, such as *linA2*, *linB*, *linDER*, *linF*, *linGHIJ*, and *linKLMN* on the chromosome; *linA1*, *linC*, and *linF* on pSRL1 and *linDEbR* on pSRL3. Hence, along with *linA*, two variant/copy of *linDER*, *linGHIJ*, and *linF* were also identified. Complete genome analysis also led to filling of the gaps in genomic information from the previous studies that resulted in identification of 26 copies of IS*6100*; maximum among HCH degraders reported till date. The sequence similarity between adjacent regions of *linA* variants on chromosome and pSRL1 provided the strong evidence of its duplication *via.* homologous recombination events rather than their independent acquisition. At last, characteristic brown pigment production by strain B90A was identified as pyomelanin which formed by the accumulation and oxidation of homogentisic acid due to the absence of *hmgA* gene.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by research grants from DBT, Government of India and DST-PURSE. The authors; H.V., A.B., R.K., N.N., and A.P. gratefully acknowledge the CSIR, UGC's Dr D.S. Kothari Postdoctoral Fellowship (DSK-PDF), NBAIM-AMASS/2006–07/NBAIM/CIR and UGC for providing the research fellowships, respectively.

Literature Cited

Alikhan NB, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402.

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

Anand S, et al., 2012. Genome sequence of *Sphingobium indicum* B90A, a hexachlorocyclohexane-degrading bacterium. *J Bacteriol*. 194:4471–4472.

Aylward FO, et al. 2013. Comparison of 26 Sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Appl Environ Microbiol*. 79:3724–3733.

Aziz RK, et al., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.

Bergstrom CT, Lipsitch M, Levin BR. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* 155:1505–1519.

Boltner D, Moreno-Morillas S, Ramos JL. 2005. 16S rDNA phylogeny and distribution of *lin* genes in novel hexachlorocyclohexane-degrading *Sphingomonas* strains. *Environ Microbiol*. 7:1329–1338.

Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol*. 155:376–386.

Ceremonie H, Boubakri H, Mavingui P, Simonet P, Vogel TM. 2006. Plasmid-encoded γ-hexachlorocyclohexane degradation genes and insertion sequences in *Sphingobium francense* (ex-*Sphingomonas paucimobilis* Sp+). *FEMS Microbiol Lett*. 257:243–252.

Delcher AL, Bratake K, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.

Dhillon BK, et al., 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res*. 43:W104–W108.

Dogra C, et al. 2004. Organization of *lin* genes and IS*6100* among different strains of hexachlorocyclohexane-degrading *Sphingomonas paucimobilis*: evidence for horizontal gene transfer. *J Bacteriol*. 186:2225–2235.

Donlon J, McGettigan S, O'Brien P, Ã Carra P. 1983. Re-appraisal of the nature of the pigment produced by *Aeromonas salmonicida*. *FEMS Microbiol Lett*. 19:285–290.

Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol*. 19:2226–2238.

Hunter RC, Newman DK. 2010. A putative ABC transporter, hatABCDE, is among molecular determinants of pyomelanin production in *Pseudomonas aeruginosa*. *J Bacteriol*. 192:5962–5971.

Kaur J, Verma H, Tripathi C, Khurana JP, Lal R. 2013. Draft genome sequence of a hexachlorocyclohexane-degrading bacterium, *Sphingobium baderi* strain LL03T. *Genome Announc*. 1:e00751–e00713.

Konstantinidis K, Tiedje J. 2004. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*. 102:2567–2572.

Krzywinski M, et al., 2009. Circos: an information aesthetic for comparative genomics. *Genome Res*. 19:1639–1645.

Kumari R, et al., 2002. Cloning and characterization of *lin* genes responsible for the degradation of hexachlorocyclohexane isomers in *Sphingomonas paucimobilis* strain B90. *Appl Environ Microbiol*. 68:6021–6028.

Kumar R, et al., 2017. Comparative genomic analysis reveals habitat specific genes and regulatory hubs within the genus Novosphingobium. *mSystems*. 10.1128/mSystems.00020-17.

Kurtz S, et al., 2004. Versatile and open software for comparing large genomes. *Genome Biol*. 5:12.

Lal R, et al., 2010. The biochemistry of microbial degradation of hexachlorocyclohexane (HCH) and prospects for bioremediation. *Microbiol Mol Biol Rev*. 74:58–80.

Malhotra S, Sharma P, Kumari H, Singh A, Lal R. 2007. Localization of HCH catabolic genes (*lin*genes) in *Sphingobium indicum*B90A. *Indian J Microbiol*. 47:271–275.

Nagata Y, Miyauchi K, Takagi M. 1999. Complete analysis of genes and enzymes for γ-hexachlorocyclohexane degradation in *Sphingomonas paucimobilis*UT26. *J Ind Microbiol Biotechnol*. 23:380–390.

Nagata Y, Mori K, Takagi M, Murzin AG, Damborsky J. 2001. Identification of protein fold and catalytic residues of γ-hexachlorocyclohexane dehydrochlorinase LinA. *Proteins* 45:471–477.

Nagata Y, Natsui S, et al., 2011. Genomic organization and genomic structural rearrangements of *Sphingobium japonicum* UT26S, an archetypal γ-hexachlorocyclohexane-degrading bacterium. *Enzyme Microb Technol*. 49:499–508.

Negi, et al. 2014. Draft genome sequence of hexachlorohexane (HCH)-degrading *Sphingobium lucknowense* strain F2^T, isolated from an HCH dumpsite. *Genome Announc*. 2:e00788-14.

Nishida H. 2012. Evolution of genome base composition and genome size in bacteria. *Front Microbiol*. 3:420.

Pearce SL, Oakeshott JG, Pandey G. 2015. Insights into ongoing evolution of the hexachlorocyclohexane catabolic pathway from comparative genomics of ten Sphingomonadaceae strains. *G3* 5:1081–1094.

Plonka PM, Grabacka M. 2006. Melanin synthesis in microorganisms-biotechnological and medical aspects. *Acta Biochim Pol*. 53:429–443.

Saeed AI, et al., 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.

Sahu S, Patnaik KK, Sharmila M, Sethunnathan N. 1990. Mineralization of α, γ, and β-isomers of hexachlorocyclohexane by a soil bacterium under aerobic conditions. *Bull Environ ContamToxicol*. 48:265–268.

Salzberg SL, Yorke JA. 2005. Beware of mis-assembled genomes. *Bioinformatics* 21: 4320–4321.

Sangwan N, et al., 2014. Reconstructing an ancestral genotype of two hexachlorocyclohexane degrading *Sphingobium* species using metagenomic sequence data. *ISME J* 8:398–408.

Sharma P, et al., 2006. Haloalkane dehalogenase LinB is responsible for β- and δ-hexachlorocyclohexane transformation in *Sphingobium indicum*B90A. *Appl Environ Microbiol*. 72:5720–5727.

Shrivastava N, Prokop Z, Kumara A. 2015. Identification and characterization of a novel δ-hexachlorocyclohexane-dehydrochlorinase LinA-type 3. *Appl Environ Microbiol*. doi:10.1128/AEM.01683-15.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 34:D32–D36.

Slater, et al. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J Bacteriol*. 191:2501–2511.

Suar M, van der Meer JR. 2004. Dynamics of multiple *lin* gene expression in *Sphingomonaspaucimobilis* B90A in response to different hexachlorocyclohexane isomers. *Appl Environ Microbiol*. 70:6650–6656.

Suar M, et al., 2005. Enantioselective transformation of α-hexachlorocyclohexane by the dehydrochlorinases LinA1 and LinA2 from the soil bacterium *Sphingomonas paucimobilis* B90A. *Appl Environ Microbiol*. 71:8514–8518.

Tabata M, et al., 2013. Complete genome sequence of the gamma-hexachlorohexane-degrading bacterium *Sphingomonas* sp. Strain MM-1. *Genome Announc*. 1:e00247–e00213.

Tucker NP, et al., 2008. The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster. *PLoS One* 3:e3623.

Vallenet D, et al., 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database*. doi:10.1093/database/bap021.

Verma H, et al., 2014. Comparative genomic analysis of nine *Sphingobium* strains: insights into their evolution and hexachlorocyclohexane (HCH) degradation pathways. *BMC Genomics* 15:1014.

Wang H, Qiao Y., Chai B, Qiu C, Chen X. 2015. Identification and molecular characterization of the homogentisate pathway responsible for pyomelanin production, the major melanin constituents in *Aeromonasmedia* WS. *PLoS One* 10:e0120923.

Wintersinger JA, Wasmuth JD. 2015. Kablammo: an interactive, web-based BLAST results visualize. *Bioinformatics* 31:1305–1306.

Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28:1033–1034.

Yeo HJ, Waksman G. 2004. Unveiling molecular scaffolds of the type IV secretion system. *J Bacteriol*. 186:1919–1926.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res*. 39:W347–W352.

Associate editor: Howard Ochman