Complete sequencing of *Novosphingobium* sp. PP1Y reveals a biotechnologically meaningful metabolic pattern

Valeria D’Argeo
to1,2†, Eugenio Notomista3†, Mauro Petrillo1,2†, Piergiuseppe Cantiello1, Valeria Cafaro3, Viviana Izzo4, Barbara Naso1,2, Luca Cozzuto1, Lorenzo Durante3, Luca Troncone3, Giovanni Paolella1,2, Francesco Salvatore1,5* and Alberto Di Donato3

Abstract

**Background:** *Novosphingobium* sp. strain PP1Y is a marine α-proteobacterium adapted to grow at the water/fuel oil interface. It exploits the aromatic fraction of fuel oils as a carbon and energy source. PP1Y is able to grow on a wide range of mono-, poly- and heterocyclic aromatic hydrocarbons. Here, we report the complete functional annotation of the whole *Novosphingobium* genome.

**Results:** PP1Y genome analysis and its comparison with other Sphingomonadal genomes has yielded novel insights into the molecular basis of PP1Y’s phenotypic traits, such as its peculiar ability to encapsulate and degrade the aromatic fraction of fuel oils. In particular, we have identified and dissected several highly specialized metabolic pathways involved in: (i) aromatic hydrocarbon degradation; (ii) resistance to toxic compounds; and (iii) the quorum sensing mechanism.

**Conclusions:** In summary, the unraveling of the entire PP1Y genome sequence has provided important insight into PP1Y metabolism and, most importantly, has opened new perspectives about the possibility of its manipulation for bioremediation purposes.

**Keywords:** De novo sequencing, *Novosphingobium* sp. PP1Y, Sphingomonads, Next generation sequencing, Aromatic pollutant compounds/bioremediation

**Background**

Aromatic compounds are among the most widespread dangerous pollutants [1]. Petroleum and its derivatives are the main sources of aromatic molecules released into the environment. The aromatic hydrocarbon content of petroleum can range from about 20% to more than 40% [2-4], whereas the aromatic hydrocarbon content of gasoline and diesel oil is about 30% and 25%, respectively [5,6].

*Novosphingobium* sp. strain PP1Y is a recently isolated marine α-proteobacterium that is able to grow on a surprisingly wide spectrum of pure mono-, poly- and heterocyclic aromatic hydrocarbons and on complex mixtures of aromatic hydrocarbons dissolved in paraffin oil phases including gasoline and especially diesel-oil which is an optimal growth substrate. Moreover, PP1Y can emulsify diesel-oil by producing small (<1 mm) regular biofilm-covered oil drops that have been described as spherical colonies harbouring a reservoir of growth substrates [7].

Strain PP1Y belongs to the *Sphingomonadaceae* family, which is characterized by the presence of glycosphingolipids in the outer membrane, instead of the more common lipopolysaccharides. This peculiarity renders the surface of their cells more hydrophobic than those of the other Gram-negative strains and, has probably contributed to the development of the ability to degrade mono- and polycyclic aromatic hydrocarbons (PAHs). Moreover, many *Sphingomonadales* harbour several (up to six) large conjugative plasmids, ranging in length from less than 50 kbp to more than 500 kbp [8]. Thanks to these
megaplasmids, several Sphingomonadales have “collected” genes for the degradation of xenobiotics and continuously exchange them with other bacterial strains [9-11]. Interesting examples are Novosphingobium aromaticivorans F199, which uses alkyl-benzenes as the sole carbon and energy source [12], Novosphingobium pentaromativorans US6-1, which degrades PAHs with 3–5 aromatic rings [13], Novosphingobium sp. TYA-1, which simultaneously degrades biphenol A and 4-alkylphenols [14] Sphingomonas paucimobilis EPA505, which degrades several polycyclic compounds [15], Sphingomonas wittichii RW1, which can grow using dibenzofuran and dibenzo-p-dioxin [16], Sphingomonas sp. TTNP3 which uses alkylphenolic compounds as a source of carbon and energy [17] and Sphingobium chlorophenolicum L-1 which degrades pentachlorophenol [11].

Here, we report the analysis of the genome of Novosphingobium sp. strain PPIY and its comparison with the genomes of N. aromaticivorans F199 (genome accession number NC_007794.1) [18] and S. wittichii RW1 [19], the closest genomes in terms of nucleotide sequence. This comparison has yielded insights into PPIY and its ability to encapsulate and degrade the aromatic fraction of fuel oils.

Results and discussion

Complete genome features and chromosomal architecture

PPIY genome sequence assembly produced four replicons classified according to their size, as we previously reported [20] (Figure 1A–D). Because the coverage of “small” plasmid (Spl) sequences was, on average, about twice that of the other replicons, it is expected that Spl is present as a two-copy object within each bacterial cell. At present, very few complete sequences of bacterial chromosomes and plasmids are available for organisms of the genus Novosphingobium (see Table 1). These sequences have a similar G+C content (about 60%), but PPIY appears to have the largest and most complex genomic organization of the genus.

Various predictive and comparative bioinformatics tools supported by biological databases were used to annotate putative open reading frames (ORFs) and other functional elements [21-27]. As in other bacteria, most of the genome sequence is predicted to be coding depending on the replicon) appear to have TTG or GTG as the starting codon. Most of them (73% of the 4,709 coding sequences predicted in the four replicons taken together) and all rRNA and tRNA genes are located on the Chr molecule. The same applies to other RNA elements; the only exceptions being three RNAs predicted on Lpl (see Table 2, “Other RNA elements” section and Additional file 1: Table S1).

Evaluation of the putative DNA replication origins

DNA replication was investigated by searching for the putative genome replication origins using a bioinformatic tool. This tool, Orifinder [28], locates predicted bacterial replication origins within each DNA sequence by taking into account base composition asymmetry, distribution of DNA-A boxes and the presence of genes frequently located close to the bacterial replication start (Additional file 1: Table S1). This tool revealed a putative Type-III replication origin on Chr, around base 1, where there is a region of base composition asymmetry containing three DNA-A-boxes, close to the hemE gene (as in the N. aromaticivorans DSM 12444 genome) and to a DNA-A gene. Differently, on Mpl, Lpl and Spl replicons, Orifinder failed to locate an acceptable putative replication origin, suggesting that other mechanisms may be involved in DNA replications origins. Interestingly, a typical plasmid replication parA/parB/parS cluster was found on each of these replicons, and Mpl and Lpl contain also a predicted plasmid replication repA gene close to the parA/parB/parS cluster, but in a different orientation to those predicted on the N. aromaticivorans DSM 12444 pNL1 and pNL2 plasmids.

On the Spl plasmid, a complete protein killer gene system is also found, namely, an operon containing two genes that force the host bacteria to retain the plasmid [29].

Protein genes identified and their significance

The gene products encoded by the 4,709 ORFs were characterized by searching for sequence similarity with known bacterial proteins contained in various collections (Figure 2A). About 94% of the ORFs matched at least one protein stored in the Uniref50 or KEGG GENES databases, although the fraction of matched sequences varies and is significantly lower for Lpl and Spl (about 80%). It is noteworthy that about 20% of the ORFs matched proteins annotated as “hypothetical”, “putative” or “uncharacterized”, and are thus classified as coding for “conserved hypothetical proteins”. When the same search was done against protein sequences stored in the COG database, the fraction of identified gene products was lower. In fact, most of the ORFs coding for “conserved hypothetical proteins” did not show any similarity. About 6% of the ORFs did not match any sequence stored in the three databases and are thus classified as coding for “hypothetical proteins”.

An all-against-all comparison of the protein sequences encoded within each replicon was done using BLAST [30] under very stringent conditions (see Table 3, PPIY-PPIY section, and Additional file 1: Figure S1A) to look for inter-duplicated genes. About 20% of ORFs from Mpl have a counterpart within the main chromosome (Chr), thereby indicating a partial genome duplication. The
The whole complement of protein-coding genes was also compared to the one encoded within other complete genomes and plasmids from bacteria of the *Sphingomonadaceae* family (Table 3 and Additional file 1: Figure S1B–D). A number of ORFs ranging between 1,500 and 1,700, i.e. 45-50% of those encoded within Chr in PP1Y, have a counterpart in the main chromosome of the closest analysed bacterial species, the most similar gene set being that from *N. aromaticivorans*, putatively from the same genus. There is no clear evidence that the three smaller replicons are functionally equivalent to other known plasmids in terms of protein coding genes: many protein genes predicted within Mpl appear to have counterparts in *N. aromaticivorans*, although only 20% of Lpl ORFs have a counterpart in the pNL1 plasmid (Accession: NC_009426.1), while others are in the pNL2 plasmid (Accession: NC_009427.1) and some are scattered along the main chromosome. Half the Spl-encoded

---

**Figure 1** Circular maps and genetic features of PP1Y replicons. The principal genomic features of PP1Y Chr (A) and its plasmids (B, C and D) are shown. For each of them, we report, from outside to the center, genes on the forward and on reverse strands (red and blue), GC content (black) and GC skew (green and violet).
Table 1 Principal features and comparison of the genomic sequences available for bacteria of the genus *Novosphingobium*

| Species                                      | Sequence complete | Sequence available | Size (Mb) | Chrs | Plasmids | Scaffolds | Contigs | CDS 16SrRNA | Identities | Gaps       |
|----------------------------------------------|-------------------|--------------------|-----------|------|----------|-----------|---------|-------------|------------|------------|
| *Novosphingobium* sp. PP1Y                  | yes               | yes                | 5,073     | 1    | -        | -         | yes     | yes         | 1,477/1,477 | 0/1,477    |
| *Novosphingobium* sp. AP12                  | no                | yes                | 5,611     | 1    | 3        | -         | -       | yes         | 1,437/1,480 | 6/1,480    |
| *Novosphingobium* sp. Rr 2-17               | no                | yes                | 4,539     | 1    | -        | -         | 166     | yes         | 1,424/1,480 | 6/1,480    |
| *Novosphingobium pentaromativorans* US6-1<sup>a</sup> | no                | yes                | 5,096     | 2    | -        | -         | 123     | yes         | 1,476/1,477 | 0/1,477    |
| *Sphingobium chlorophenolicum* L-1<sup>b</sup> | yes               | yes                | 4,449     | 2    | 1        | -         | -       | yes         | 1,360/1,480 | 7/1,480    |
| *Sphingobium indcum* B90A                    | no                | yes                | 4,082     | 1    | -        | -         | 149     | yes         | 1,373/1,480 | 7/1,480    |
| *Sphingomonas echinoides* ATCC 14820         | no                | yes                | 4,200     | 1    | 2        | -         | 64      | yes         | 1,372/1,488 | 26/1,488   |
| *Sphingomonas wittichii* DPS8                | no                | yes                | 5,628     | 1    | 2        | -         | 739     | no          | 1,372/1,484 | 16/1,484   |
| *Sphingomonas wittichii* RW1                 | yes               | yes                | 5,382     | 1    | 2        | -         | -       | yes         | 1,373/1,484 | 16/1,484   |
| *Novosphingobium aromaticivorans* DSM 12444  | yes               | yes                | 3,561     | 1    | 2        | -         | -       | yes         | 1,400/1,480 | 10/1,480   |
| *Novosphingobium nitrogenifigens* DSM 19370 B057DRAFT | no                | yes                | 4,148     | 1    | -        | -         | -       | yes         | 1,399/1,478 | 29/1,478   |
| *Novosphingobium nitrogenifigens* DSM 19370 | no                | yes                | 4,140     | 1    | -        | -         | -       | yes         | 1,400/1,479 | 31/1,479   |

<sup>a</sup>no plasmids plA1 and plA2 in the size count; <sup>b</sup>chr1 + chr2 in the size count. PP1Y strain is in boldface type.
proteins are encoded by the main chromosome in *N. aromaticivorans*. *Sphingobium japonicum* [31] and PP1Y share elements of comparable size, although the latter has a 45% similarity within the main chromosome in terms of protein-encoding content, but diverge more extensively in an additional smaller chromosome. The two species have aromaticivorans. *Sphingobium japonicum* N. proteins are encoded by the main chromosome in N.

| Genome  | Chr   | Mpl  | Lpl  | Spl  |
|---------|-------|------|------|------|
| **Size, bp**  | 3,911,486 | 1,161,602 | 192,103 | 48,714 |
| **G + C content, %**  | 63.7 | 62.3 | 60.7 | 60.1 |
| **Copies for each bacterial cell**  | 1 | 1 | 1 | 2 |

**ORFs**

| Number | Minimum length, bp | Average length, bp | Maximum length, bp | ATG initiation codons, % | GTG initiation codons, % | TTG initiation codons, % | % of coding sequence |
|--------|--------------------|--------------------|--------------------|--------------------------|--------------------------|--------------------------|---------------------|
| 3,481  | 975                | 199                | 54                 | 86.8                     | 88.4                     | 4.4                     | 88.1                |

**Repetitive elements**

| Prophage elements | IS elements |
|-------------------|-------------|
| 3                 | 5           |

**RNAs**

| rRNA (16S-23S-5S) | tRNAs | Other RNA elements* |
|-------------------|-------|---------------------|
| 3                 | 0     | 0                   |
| 58                | 0     | 0                   |

| Cobalamin riboswitches | suhB | SRP_bact | Rnase P clA | tmRNA | TPP riboswitch | ctRNA p42d | Glycine riboswitch | LR PK1 | GroupII catalytic introns | ROSE |
|-----------------------|------|----------|------------|-------|---------------|------------|-------------------|--------|--------------------------|------|
| 5                    | 3    | 1        | 0          | 0     | 1             | 0          | 1                 | 0      | 0                        | 0    |

*Predicted RNA elements, besides tRNAs and rRNAs.

### Characterization of the PP1Y genes involved in aromatic hydrocarbon degradation

The degradation of aromatic hydrocarbons requires activation of the aromatic ring. This generally occurs by dihydroxylation of the aromatic ring catalyzed by pairs of monooxygenases or dioxygenases/dehydrogenases that constitute the upper pathways. Ring activation is followed by ring cleavage catalyzed by specialized dioxygenases (intra- and extradiol dioxygenases) that start the lower pathways. In the case of methylated aromatic compounds, the initial step can be a monooxygenation reaction of a methyl group followed by oxidation to carboxylate. These reactions can be catalyzed by soluble dioxygenases or by membrane monooxygenases related to xylene monooxygenase XylM. The arylcarboxylate eventually undergoes ring dihydroxylation and cleavage [34]. Analysis of the PP1Y genome revealed at least 81 ORFs (Table 4) that potentially code for the enzymes of both the upper (ring activation) and lower (ring cleavage) pathways.

No soluble multicomponent monooxygenase that resembled the well characterized methane monooxygenases and toluene/o-xylene monooxygenase [35] was found in the present study. Thirty-eight ORFs, which were predicted to code for 34 different multicomponent aromatic hydroxylating dioxygenases [36], were identified – a number clearly higher than in the closely related strains *N. aromaticivorans* F199 and *N. pentaromativorans* US6-1 (27 and 18 dioxygenases, respectively) (Figure 3). PP1Y has a close counterpart of each F199 dioxygenase: three of these are present in double copy with a 100% identity, which is indicative of a very recent duplication event; and four others have a 90-95% identity, which suggests a less recent duplication event followed by divergence. All duplicated ORFs are closely related to seven ORFs coding for hydroxylating dioxygenases found on plasmid pNL1 from strain F199. Indeed, replicon A of strain PP1Y contains two copies of a region of plasmid pNL1 probably derived by multiple fusion/duplication events (Additional file 1: Figure S3A). Six PP1Y oxygenases from the megaplasmids (Mpl6792, Mpl2166, Mpl5621, Mpl5540, Mpl5477, Mpl5466) do not have homologues in strains F199 and US6-1 but are closely related to predicted oxygenases from strain RW1 (Additional file 1: Figure S4A and B), suggesting that strain PP1Y combined the
Figure 2 Open reading frame (ORF) annotation. Fraction of ORFs that resemble another protein based on BLAST matches with the Uniref50 and KEGG databases and that have a predicted function according to COG categories (A). The distribution of Chr and Mpl genes is reported with respect to COG functional categories (B) and according to Kegg pathways (C). D, Cell cycle control, cell division, chromosome partitioning; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; O, Posttranslational modification, protein turnover, chaperones; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V, Defense mechanisms; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; C, Energy production and conversion; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; P, Inorganic ion transport and metabolism; Q, Secondary metabolite biosynthesis, transport and catabolism; R, General function prediction only; S, Function unknown.
Table 3 Predicted protein genes comparison

| Genome | Replicon | Total ORFs | ORFs in PP1Y replicons |
|--------|----------|------------|------------------------|
|        |          | Chr        | Mpl        | Lpl        | Spl        |
| PP1Y   | Chr      | 3,481      | 203        | 30         | 4          |
|        | Mpl      | 975        | 240        | 975        | 10         |
|        | Lpl      | 199        | 43         | 10         | 199        |
|        | Spl      | 54         | 5          | -          | 54         |
| Naro   | pNL1     | 182        | 203        | 32         | 8          |
|        | pNL2     | 431        | 204        | 232        | 5          |
| E.lit  | Chr      | 3,012      | 1,513      | -          | -          |
| Sala   | Chr      | 3,166      | 1,536      | -          | -          |
| S.wit  | Chr      | 4,851      | 1,543      | -          | -          |
| Z.mob1 | Chr      | 1,801      | 885        | -          | -          |
| Z.mobZ | Chr      | 1,729      | 865        | -          | -          |
|        | Chr1     | 3,530      | 1,573      | -          | -          |
|        | Chr2     | 590        | 234        | 147        | 0          |
|        | pCHQ1    | 225        | 102        | 11         | 23         |
|        | pUT1     | 45         | 8          | -          | -          |
|        | pUT2     | 9          | -          | -          | -          |
| S.jap  | pCHQ1    | 225        | 102        | 11         | 23         |
|        | pUT1     | 45         | 8          | -          | -          |
|        | pUT2     | 9          | -          | -          | -          |

N. aro = Novosphingobium aromaticivorans DSM 12444; E. lit = Erythrobacter littoralis; S. ala = Sphingopyxis alaskensis; S. wit = Sphingomonas wittichii; Z. mob = Zymomonas mobilis 11163; Z. mob2 = Zymomonas mobilis ZM4; S. jap = Sphingonium japonicum.

The protein-coding genes identified in PP1Y are compared with those of several genomes and plasmids.

Dioxygenase pools of strains F199 and RW1 and later expanded the pool by duplication events. This strategy enabled PP1Y to expand the pathway for the degradation of naphthalene and methyl substituted naphthalenes, and to degrade larger PAHs. The predicted pathway is shown in Additional file 1: Figure S3B.

Two potential membrane monoxygenases are predicted in PP1Y; they show a 96% identity with each other and a 71-75% identity with the sole membrane monoxygenase found in strain F199, which suggests another recent event of gene duplication. The two PP1Y monoxygenases (Additional file 1: Figure S5) mainly differ in the substrate-binding region, possibly to allow different substrate specificity. No membrane monoxygenase is present in the genome of Sphingobium sp. MM-1, Sphingobium japonicum UT26, Sphingobium chlorophenolica L-1, Sphingobium sp. SYK-6, Sphingobium wittichii RW1 or Novosphingobium pentaromativorans US6-1. This suggests that, also in this case, the PP1Y enzymatic repertoire was expanded by horizontal gene transfer and duplication events.

Ten potential ORFs code for seven homomultimeric extradiol ring cleavage dioxygenases (RCDs) [37], related to Pseudomonas putida MT2 catechol 2,3-dioxygenase (Additional file 1: Figure S6). ORF AT15671/AT31688 codes for a putative classic catechol 2,3-dioxygenase; homology modeling and substrate docking have an active site pocket slightly larger than the family benchmark P. putida MT2 catechol 2,3-dioxygenase (not shown). AT15599/AT31688 code for a putative dihydroxynaphthalene dioxygenase, and Mpi3065 for a putative 2,3-dihydroxybiphenyl-1,2-dioxygenase. AT32663 is a divergent member of the extradiol dioxygenase family and no closely related sequence is present in the protein databases.

Homology models of these three RCDs are shown in Additional file 1: Figure S7. The four RCDs were cloned, expressed in Escherichia coli and their cleavage activity was assayed on 3-methylcatechol, 2,3-dihydroxybiphenyl (2,3-DHBP), and 4-hydroxyestradiol (4-OHE). The latter was used as analogue of dihydroxy PAHs because these compounds are unstable, difficult to synthesize and not commercially available. The protein coded in AT15599/AT31688 is a versatile enzyme, able to cleave substrates with 1 to 4 rings (Table 5). Enzyme AT32663 is only active on polycyclic substrates, while Mpi3065 is active on 2,3-DHBP, as predicted. Finally, AT15671/AT31685 is very active on monocyclic catechols, even though its substrate specificity is wider than that of P. putida MT2 catechol 2,3-dioxygenase. Taken together, these four enzymes are able to cleave all classes of 3- and/or 4-substituted catechols in complex mixtures. The other three PP1Y RCDs, AT33026, Mpi10251 and Mpi4329/Mpi4634, are poorly characterized. Preliminary modelling studies suggest that they are dioxygenases specialized in cleaving catechols bearing substituents at positions 3,5 and/or 4,5 and/or 3,6. Therefore, these dioxygenases have a substrate specificity complementary to the four described above.

The Neighbor-joining tree of RCDs (Additional file 1: Figure S6) shows a great heterogeneity among sphingomonads both in the number of potential RCDs (from 1, in the case of strain L-1, to 8 in the case of strain RW1) and in the distribution of the proteins among the RCD subfamilies. Only strains F199 and PP1Y have at least one representative for each subfamily. This particular set of RCDs could allow strain PP1Y to metabolize complex mixtures of catechols deriving from the simultaneous oxidation of several mono- and polycyclic-aromatic hydrocarbons (Additional file 1: Figure S8), which are the preferred substrates for growing this strain.

Besides the seven homomultimeric estradiol RCDs, the PP1Y genome contains also four potential ORFs for heterodimeric extradiol RCDs that are able to cleave catechol rings bearing substituents with carboxylate groups like protocatechuate (see also Additional file 2: Supplementary Results and Discussion). The genome of strain PP1Y contains several other ORFs coding for hypothetical mono- and dioxygenases whose involvement in the degradation of xenobiotics is less clear. Among these, CDS AT10830 is particularly interesting as it codes for a 2-oxoglutarate-dependent oxygenase. These oxygenases
Cleave different substrates, namely alkyl-sulphonates and fenoxy-acids, by catalyzing monooxygenation reactions of CH bonds adjacent to good leaving groups. Interestingly, no sphingomonad contains a homologous enzyme. Moreover, AT10830 is a member of a group of adjacent ORFs coding for: (i) a hydroxylating dioxygenase (AT10866) that is only distantly related to RW1 and F199 dioxygenases (Additional file 1: Figure S4A); (ii) a heterodimeric extradiol ring cleavage dioxygenase related to 3,4-dihydroxybenzoate dioxygenases; and (iii) a hypothetical acetamidase (AT10838). This cluster of ORFs is present in several distantly related strains including some beta and gamma proteobacteria, thus suggesting a horizontal gene transfer event. At present, nothing is known about the physiological role of this pathway, but its wide diffusion suggests a potentially important ecological role. The data related to Additional file 1: Figures S9–S11 are reported under “Additional file 2: Supplementary Results and Discussion”.

Stress response genes and their functions

The PP1Y genome contains several ORFs potentially coding for the so-called resistance-nodulation-cell division (RND)-type efflux pumps [38] that actively excrete toxic molecules, and have thus been implicated in the capacity of PP1Y to grow in close contact with a diesel oil phase. (RND)-type efflux pumps are constituted by three subunits: the inner membrane, the outer membrane and the membrane fusion component. The PP1Y genome contains eight potential ORFs for the inner membrane membrane-bound monooxygenases. The PP1Y genome also contains eight potential ORFs for P-type ATPases (Additional file 1: Figure S12B), which are ATPases (Additional file 1: Figure S12B), which are membrane ATP-dependent efflux pumps specialized in the excretion of metal cations [40]. For comparison, C. metallidurans CH34 genome codes for 9 P-type ATPases.

On the basis of these findings, we assayed the ability of PP1Y to grow in liquid medium containing high concentrations of metal cations. Figure 4 shows that PP1Y can grow in the presence of millimolar concentrations of nickel (2.5 mM), lead (10 mM), copper (10 mM) and zinc (5 mM). At higher concentrations, the growth rate steeply decreases to zero (not shown). Interestingly, all the metals increase the carbohydrates/proteins ratio with respect to the control culture, thus suggesting that modification of the cell envelope could contribute to resistance to metals. These results show that the ability of PP1Y to tolerate heavy metals is comparable to that of heavy metal-tolerating strains like C. metallidurans CH34 [40], which suggests that PP1Y could play a role in the bioremediation of hydrocarbons in environments polluted by heavy metals.

Tellurite anion is highly toxic to microorganisms (much more than arsenate and arsenite) thanks to its ability to catalyze the oxidation of cell thiols and produce radical oxygen species [41]. Therefore, the wide diffusion of tellurite-resistance mechanisms among bacteria is not surprising, and they might include an aspecific increase of the radical scavenger systems and specific tellurite anion transporters [42]. The PP1Y genome contains three ORFs potentially coding for proteins belonging to three different tellurite-resistance mechanisms:

Table 4 Number of ORFs coding for potential upper (ring activation) and lower (ring cleavage) pathway enzymes in strains PP1Y and F199

| Pathway Enzymes                                      | Novosphingobium sp. PP1Y | Novosphingobium aromaticivorans F199 |
|-------------------------------------------------------|---------------------------|--------------------------------------|
| Ring hydroxylating dioxygenases                       | 38a                       | 27                                   |
| Molybdopteryn-dependent monooxygenases                | 2                         | 1                                    |
| Membrane-bound monooxygenases                         | 5                         | 3                                    |
| 2-oxoglutarate-dependent oxygenases (taurine-dioxygenase like) | 19                        | n.d.                                 |
| 2-oxoglutarate-dependent oxygenases                   | 1                         | 0                                    |
| Extradiol ring-cleavage dioxygenases                  | 10f                       | 6                                    |
| Protocatechuate 4,5 dioxygenases (extradiol ring-cleavage) | 4                         | 3                                    |

n.d., not determined.

*Four ORFs coding for ring hydroxylating dioxygenases and 3 ORFs coding for extradiol ring-cleavage dioxygenases are duplicated, therefore, the genome of PP1Y codes for 34 potential ring hydroxylating dioxygenases and 7 potential extradiol ring-cleavage dioxygenases.
telA (from the *E. coli* kilA/telA/telB system), tehB (from the *E. coli* tehA/tehB system) and terC from *Proteus mirabilis* [43]. Due to the scarce knowledge about these systems, it is difficult to predict their role in tellurite resistance. However, all these ORFs are located in a cluster of ORFs coding for proteins probably involved in detoxification. Interestingly, a similar cluster of ORFs is present in the genome of strain RW1, but not in other sphingomonads (data not shown).

The importance of glutathione as a radical scavenger and mediator of detoxification systems varies among...
bacteria. However, several bacteria use glutathione and glutathione-dependent enzymes to detoxify reactive organic compounds (like epoxides), halogenated compounds or alkylhydroperoxides, and reactive oxygen species (ROS) such as oxygen radicals [44].

In addition to genes involved in glutathione synthesis and in the reduction of oxidized glutathione, the PP1Y genome codes for 18 glutathione S-transferases (Additional file 1: Table S3). This number is about double that of \textit{E. coli} and suggests that glutathione could play an important role in detoxification of toxic diesel oil components and of toxic metabolites produced by the oxidation of aromatic hydrocarbons, like epoxides and ROS.

The PP1Y genome also codes for six members of a peculiar family of very small (about 100 amino acids) monooxygenases known as “antibiotic biosynthesis monooxygenases” [45]. These enzymes are the only known monooxygenases not containing any metal or flavin cofactors [46], and that prevalently oxidize phenolic groups to quinones. They are involved in at least two very different physiological processes: (i) the synthesis of the polyketide antibiotics (e.g. the products of ActVA-Orf6 of \textit{Streptomyces coelicolor}), and (ii) the quinol redox cycle (e.g. quinol monooxygenase YgiN from \textit{E. coli}). In particular, \textit{E. coli} YgiN could prevent the accumulation of the semiquinone intermediate formed during the oxidation of quinols to quinones thus minimizing the formation of free radical species [47]. At least some of the six PP1Y antibiotic biosynthesis monooxygenases could have similar functions. However, some of them could be also involved in the synthesis of secondary metabolites. It is noteworthy that PP1Y is able to inhibit the growth of molds (unpublished results), which suggests it secretes antifungal compounds.

\textbf{Identification of genes involved in extracellular polymer secretion and biofilm formation}

The analysis of the PP1Y genome has revealed potential regulatory mechanisms (quorum sensing, QS) and secretion systems for extracellular polymers, including polysaccharides and poly-gamma-glutamate, which may play a role in the complex “social” behavior of PP1Y, a strain able to form different types of multicellular amorphous aggregates and ordered biofilm

### Table 5 PP1Y RCD specific activity assayed on 3-methylcatechol (3-MC), 2,3-dihydroxybiphenyl (2,3-DHBP), and 4-hydroxy-oestradiol (4-OHE)

| Substrate   | AT15599/AT31688 | Mpl3065 | AT32663 | AT15671/AT31616 |
|-------------|-----------------|---------|---------|-----------------|
| 3-MC        | 13.54           | 1.72    | 0.453   | 372.9           |
| 2,3-DHBP    | 32.73           | 12.05   | 23.5    | 48              |
| 4-OHE       | 10.33           | 0       | 44.62   | 3 x 10^{-3}     |

AT numbers refer to the chromosome DNA.

---

![Figure 4](image_url)
(see also Additional file 2: Supplementary Results and Discussion). Quorum sensing is a simple molecular mechanism that results in coordinated behavior in response to cell density [48]. The presence in PPIY of two QS systems is interesting since they could work simultaneously in response to two different cell densities or, could be activated alternatively under specific conditions. Both possibilities could account for PPIY’s complex behavior.

Although several ORFs for sphingan synthesis have distantly related homologues in the PPIY genome (identity <30-40%), a gene cluster similar to those present in other Sphingomonas does not exist in PPIY. Therefore, it is unlikely that PPIY could produce a sphingan-like polysaccharide. However, several clusters potentially coding for the synthesis of extracellular polysaccharides are distributed among the larger replicons (chromosome and Mpl), as shown in Additional file 1: Table S4, Table S5 and Figure S14A. Lpl contains two regions that are probably involved in the synthesis of exopolysaccharides (Additional file 1: Figure S14B), and are widely distributed among sphingomonads. The closest sequences can be found in S. japonicum UT26 with an identity of 70-90% at protein level. Interestingly, the region between these two couples of ORFs in Lpl contains five ORFs coding for hypothetical glycosyl transferases and four ORFs coding for the subunits of an ABC-type polysaccharide transport system with high homology in several sphingomonads (Additional file 1: Figure S15 A-B). Lpl651 is particularly interesting as it codes for a large protein containing three glycosyl transferase-like domains. No other sphingomonad contains a representative of this subfamily of glycosyl transferases that can be found in distantly related bacteria, suggesting another case of horizontal gene transfer. Taken together these findings suggest that Lpl codes for the synthesis and export of one or more capsular polysaccharide(s) that probably contains mannose and rhamnose, like sphingans, but whose structures could differ from those produced by other sphingomonads.

Several biofilm-forming strains secrete cellulose as a matrix component. Lpl from PPIY shares with Sphingobium japonicum UT26 a cluster of ORFs coding for a two-subunit cellulose synthase (Additional file 1: Figure S16A), which implicates Lpl in both biofilm synthesis and remodelling. Another CDS coding for a hypothetical cellulase is located on chromosome (AT36325) not far from a CDS coding for an exo-1,3/1,4-beta-glucanase which could act downstream the cellulase (endo-1,4-beta-glucanase) (Additional file 1: Figure S16B). Interestingly, PPIY has the largest number of glycosyl hydrolases and glycosyl transferases among sphingomonadaceae and related groups of alpha proteobacteria (Additional file 1: Table S6).

The PPIY genome contains three ORFs coding for γ-PGA polymerases (Additional file 1: Figure S16C), which are involved in the synthesis of poly-gamma-glutamate, a strongly anionic homopolymer composed of glutamate residues linked by amide bonds between α-amino and γ-carboxyl groups [49]. This polymer can perform different functions, including the stabilization of the extracellular matrix, glutamate storage and toxic metals binding (Additional file 2: Supplementary Results and Discussion).

Conclusions

This analysis of the annotated Novosphingobium sp. PPIY genome has revealed peculiar biochemical and biotechnological properties, namely, the metabolic pathways specifically involved in: (i) the degradation of a vocabulary of aromatic hydrocarbons, (ii) the resistance to toxic compounds and (iii) the QS social behavior mechanism. This detailed functional evaluation opens new translational perspectives regarding the possible manipulation of the PPIY genome for bioremediation purposes. Moreover, the comparison between the enzymatic machinery of PPIY and those of the other sphingomonads able to degrade environmental pollutants suggests that each sphingomonad has independently evolved its own repertoire of degradative enzymes through a complex combination of vertical heredity, horizontal gene transfers, duplications and rearrangements. This process is still ongoing as demonstrated by the presence of multiple copies of pNL1-like regions at different locations of the PPIY chromosome. As a consequence, even closely related strains like PPIY, F199 and US6-1, which belong to the genus Novosphingobium, have unique features and adaptations to specific, also polluted, environments. The analysis reported in this paper strongly supports the general belief that sphingomonads are very adaptable bacteria with extraordinary genomic plasticity. It also raises biotechnological perspectives of using sphingomonads in bioremediation processes.

Methods

Bacterial growth and DNA extraction
Novosphingobium sp. strain PPIY was routinely grown and genomic DNA was extracted as previously described [7].

Genome sequencing and assembly
The de novo whole-genome shotgun sequencing of Novosphingobium sp. PPIY was carried out as described in a preliminary report (EMBL database under accession numbers: FR 856862, FR 856861, FR 856860 and FR 856859 for Chr, Mpl, Lpl and Spl, respectively) [20].

Sequence annotation
Sequence annotation includes predicted ORFs, tRNAs, rRNAs and other ncRNAs, identified by using the following tools:
• ORFs were predicted by Grc [20] combined with the Uniref50 [21] and KEGG GENES [22] databases;
• tRNA genes and tRNAs, genes were identified by using RNAmmer [23] and tRNAscan-SE [24] respectively;
• Other predicted ncRNA elements were found by Infernal using the RFAM database records as models [25,26].

In-house developed pipelines guided the whole annotation process, scheduling and running single applications on a 56-blade cluster. ORFs on chromosome, mega-, large- and small plasmids are identified by a number preceded by “AT”, “Mpl”, “Lpl” and “Spl” respectively. All the PPIY ORFs and their protein sequences discussed in the text and/or included in the trees are available on the “Gene” database at http://www.ncbi.nlm.nih.gov/gene/.

Phylogenetic analysis
The sequences included in this study were selected by searching public protein databases with BLAST and PSI-BLAST [50]. Clustal Omega [http://www.ebi.ac.uk/Tools/msa/clustalo/] was used to obtain multiple alignments. Alignments were visualized and examined using JalView [51] and MEGA5.1 [52]. Phylogenetic trees were obtained, visualized and manipulated using MEGA5.1. Bootstrap confidence analysis was performed on 1,000 replicates using the Neighbor-Joining method [53]. The evolutionary distances were computed using the Poisson correction method [54] and were expressed as the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated.

Subcloning, expression and activity analysis of RCDs
Open reading frames coding for RCDs were amplified by PCR using genomic DNA as template. Gene sequences were engineered to introduce an NdeI site at the 5’-end and a HindIII site at the 3’-end. PCRs were performed in a total reaction volume of 50 μl, containing 50 ng of genomic DNA, 1 μM of each primer, 0.2 mM dNTPs (Roche, Basel, Switzerland), 1× PCR buffer and 2.5 U of Platinum pfX polymerase from Pyrococcus sp. (Invitrogen). The amplification program was optimized as follows: initial denaturation at 95°C for 2 min, amplification for 20 cycles of denaturation at 92°C for 1 min, annealing at 56°C for 1 min, extension at 68°C for 1 min. The amplified fragments cut with NdeI and HindIII were cloned into pET22b (+) expression vector (Novagen) previously cut with the same enzymes. RCDs were expressed in E. coli strain BL21(DE3), transformed with the appropriate expression vector, purified by ion-exchange chromatography on Q-Sepharose FF resin and analyzed for quality as described previously [55]. Assays were performed at 25°C in 50 mM Tris/HCl (pH 7.5) in a final volume of 500 μl by spectrophotometric determination of the product of the reaction as described elsewhere [55]. The amount of the products was measured using their extinction coefficients: ε_{388} = 13,800 M⁻¹ cm⁻¹ for the product of 3-methylcatechol (3-MC) [55]; ε_{434} = 13,200 M⁻¹ cm⁻¹ for the product of 2,3-dihydroxybiphenyl (2,3-DHBP) [56]; ε_{298} = 9,100 M⁻¹ cm⁻¹ for the product of 4-hydroxy-oestradiol (4-OHE). One unit of enzyme activity was defined as the amount of enzyme required to form 1 μmol of the product per minute under the assay conditions. Specific activity is given as units per milligram of protein.

Synthesis of 4-OHE was achieved by Dr. Pezzella (Department of Chemistry, University of Naples Federico II) via the o-iodoxybenzoic acid (IBX)-mediated phenolic oxygenation procedure as previously described [57]. All chemicals were of the highest grade available and were from Amersham Biosciences, Promega, New England Biolabs, Sigma, ABCR GmbH, Fluka, or AppliChem. Escherichia coli strain BL21 (DE3) and plasmid pET22b (+) were purchased from Novagen (Madison, WI, USA). DNA sequences and oligonucleotide synthesis were performed by Eurofins MWG Operon (Germany).

Heavy metal resistance
The resistance of bacteria to heavy metals was evaluated by measuring bacterial growth according to Notomista et al. [7] in a minimal medium containing 20 mM MOPS pH 6.9, 100 mM NaCl, 1 g/L NH₄Cl and 1.0% glutamic acid as sole carbon and energy source, plus trace amounts of four heavy metal salts: NiCl₂, CuCl₂, ZnCl₂, and PbCl₂, plus four heavy metal salts: NiCl₂ (2.5 mM), CuCl₂ (10 mM), ZnCl₂ (5 mM), and PbCl₂ (10 mM) (Sigma–Aldrich, St Louis, MO, USA).

Availability of supporting data
The following additional data are available with the online version of this paper: Additional file 2, which includes Supplementary Results and Discussion; and Additional file 1, which includes Tables S1 to S6 and Figures S1 to S16. Phylogenetic tree newick files are available online as Additional file 3. PPIY genomic sequences are available in the EMBL database (http://www.ebi.ac.uk/ena/) under accession numbers: FR856862, FR856861, FR856860 and FR856859 for Chr, Mpl, Lpl and Spl, respectively (http://www.ebi.ac.uk/ena/data/view/Taxon:Novosphingobium%20sp.%20PPIY).

Additional files
Additional file 1: Supplementary Tables and Figures.
Additional file 2: Supplementary Tables and Figures.
Additional file 3: Phylogenetic tree newick files.
Abbreviations
PAHs: Polycyclic aromatic hydrocarbons; Chr: Chromosome; Mpl: Mega plasmid; Lpl: Large plasmid; Spl: Small plasmid; ORF: Open reading frames.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
VD, PC, BN and FS designed and performed the sequencing experiments; MP, LC and GP carried out the genome assembly and annotation; EN, VC, VI, LD, LT and ADD characterized and functionally evaluated PPIY metabolic features; VD, EN, GP, FS and ADD analyzed the final data and wrote the manuscript. All the authors read and approved the final manuscript.

Acknowledgment
This work was supported by grant LS95 (to F.S.) from Regione Campania; grant PS 35-126/hnd and grant PON01_02589 (MICROMAP) 2012 from the Ministry of University and Research (both to F.S.); grant RF-2010-2318372 from the Ministry of Health (to F.S.). The authors thank Jean Ann Gilder (Scientific Communication srl, Naples) for editing the text, and Vittorio Lucignano, CEINGE-Biotechonologia Avanzate for technical assistance.

Author details
1CEINGE-Biotechonologia Avanzate, Napoli, Italy. 2Dipartimento di Medicina Molecolare e Biotechonologia Mediche, Università di Napoli Federico II, Napoli, Italy. 3Dipartmento di Biologia, Università di Napoli Federico II, Napoli, Italy. 4Dipartimento di Medicina e Chirurgia, Università degli Studi di Salerno, Salerno, Italy. 5IRCCS-Fondazione SDN, Naples, Italy.

Received: 5 November 2013 Accepted: 8 May 2014
Published: 19 May 2014

References
1. Baird WM, Hooven LA, Mahadevan B: Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. Environ Mol Mutagen 2005, 45:106–114.
2. Henny JA: Composition and toxicity of petroleum products and their additives. Hum Exp Toxicol 1998, 17:111–123.
3. King RW: Petroleum: its composition, analysis and processing. Occup Med 1988, 3:409–430.
4. Wang Z, Fangas M, Blenkinsop S, Sergy G, Landiault M, Sigouin L, Fogli T, Semple K, Westlake DW: Comparison of oil composition changes due to biodegradation and physical weathering in different oils. J Chromatogr A 1998, 809:89–107.
5. Sawyer RF: Trends in auto emissions and gasoline composition. Trends in auto emissions and gasoline composition. Sawyer RF: Trends in auto emissions and gasoline composition.
6. Petrov V, Petriillo M, Cantillo P, Niso B, Cozzuto L, Notomista E, Paola E, Di Donato A: Characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. Appl Environ Microbio 1991, 57:796–803.
7. John DH, Kwon KK, Kang JH, Jung HB, Kim SJ: Norovirionovirus sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. Int J Syst Evol Microbio 2004, 54:1463–1467.
8. Toriyama T, Kinomura Y, Nakuchi S, Morii K: Biodegradation of biphenyl A and 4-alkylphenols by Norovirionovirus sp. strain TYA-1 and its potential for treatment of polluted water. Water Sci Technol 2012, 66:2022–2028.
9. Desai AM, Autenrieth RL, Dimitriu-Christidis P, McDonald TJ: Biodegradation kinetics of select polycyclic aromatic hydrocarbon (PAH) mixtures by Sphingomonas paucimobilis EPAS05. Biodegradation 2008, 19:223–233.
10. Corato A, Errico E, Roggo C, Johnson DR, van der Meer JR: Genome-wide analysis of salicylate and dibenzofuran metabolism in Sphingomonas wittichii RW1. Front Microbiol 2012, 3:300.
11. Varca R, Di Donato A: Complete genome sequence of the representative strain of a novel bacterial species, Sphingobium chlorophenolicum strain MM-1 isolated from a contaminated site. Microb Ecol 2009, 58:210–218.
12. Varca R, Di Donato A: Complete genome sequence of the dioxin-mineralizing bacterium Sphingomonas wittichii RW1. J Bacteriol 2010, 192:6101–6102.
13. Varca R, D’Argenio V, Petrillo M, Cantillo P, Niso B, Cozzuto L, Notomista E, Paola E, Di Donato A, Salvatore F: Development of a whole genome sequencing strategy for assembly of the genome of Novosphingobium sp. strain PPIY. J Bacteriol 2011, 193:2496.
14. Warren AS, Setulab JC: The Genome Reverse Compiler: an explorative annotation tool. BMC Bioinform 2009, 10:35.
15. Varca R, D’Argenio V, Petrillo M, Cantillo P, Niso B, Cozzuto L, Notomista E, Paola E, Di Donato A, Salvatore F: De novo sequencing and assembly of the whole genome of Novosphingobium sp. strain PPIY. J Bacteriol 2011, 193:2496.
16. Warren AS, Setulab JC: The Genome Reverse Compiler: an explorative annotation tool. BMC Bioinform 2009, 10:35.
17. Varca R, D’Argenio V, Petrillo M, Cantillo P, Niso B, Cozzuto L, Notomista E, Paola E, Di Donato A, Salvatore F: De novo sequencing and assembly of the whole genome of Novosphingobium sp. strain PPIY. J Bacteriol 2011, 193:2496.
18. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
19. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
20. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
21. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
22. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
23. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
24. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
25. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
26. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
27. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
28. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
29. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
30. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
31. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
32. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
33. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
34. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
35. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
36. Genomic organization and genomic structural analysis of salicylate and dibenzofuran metabolism in Sphingomonas sp. strain TTNP3 - a review on seven years of research. J Mol Biol 2003, 335:1000–1010.
o-xylene, toluene, and ethylbenzene by Rhodococcus sp. Strain DK17. Appl Environ Microbiol 2004, 70:7086–7092.

37. Andújar E, Hernández MJ, Kaschabek SR, Reineke W, Santero E: Identification of an extraadiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product. J Bacteriol 2000, 182:790–795.

38. Paulson JT, Park JH, Choi PS, Saier MH: A family of Gram-negatively outer membrane factors that function in the export of protein, carbohydrates, drugs, and heavy metals from Gram-negative bacteria. FEMS Microbiol Lett 1997, 156:1–8.

39. Janssen PJ, Van Houwt R, Moors H, Monsieurs P, Morin N, Michaux A, Benomtame MA, Leys N, Vallaeys T, Ladiges W, Monchy S, Médigue C, Taghavi S, McCorkle S, Dunn J, van der Leike D, Merges M: The complete genome sequence of Cupriavidus metallidurans strain CH36, a master survivalist in harsh and anthropogenic environments. PLoS One 2010, 5:e10433.

40. Legatzi A, Grass G, Anton A, Rensing C, Nies DH: Interplay of the Czc system and two P-type ATPases in conferring metal resistance to Ralstonia metallidurans. J Bacteriol 2003, 185:4354–4361.

41. Chasteen TG, Fuentes DE, Tantaleán JC, Vásquez CC: Tellurite: history, oxidative stress, and molecular mechanisms of resistance. FEMS Microbiol Rev 2009, 33:820–832.

42. Joseph S, Desai P, Ji Y, Cummings CA, Shih R, Brzoska P, Hambly SE, Masood N, Harri S, Sonbol H, Chuzhanova N, McClelland M, Furtado MR, Fuentes DE, Tantaleán JC, Vásquez CC: A family of Gram-negative bacterial proteins. J Biol Chem 2008, 283:10297–10306.

43. Taylor DE, Rooker M, Keanan M, Ng UK, Martin I, Perna NT, Buford NT, Blattner FR: Genomic variability of O islands encoding tellurite resistance in Enterohemorrhagic Escherichia coli O157:H7 isolates. J Bacteriol 2002, 184:4690–4698.

44. McLaggan D, Ruffino H, Rufino H, Jaspars M, Booth IR: Glutathione-dependent conversion of N-ethylmaleimide to the maleimonic acid by Escherichia coli: an intracellular detoxification process. Appl Environ Microbiol 2000, 66:1393–1399.

45. Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, Benomtame MA, Leys N, Vallaeys T, Ladiges W, Monchy S, Médigue C, Taghavi S, McCorkle S, Dunn J, van der Leike D, Merges M: The complete genome sequence of Cupriavidus metallidurans strain CH36, a master survivalist in harsh and anthropogenic environments. PLoS One 2010, 5:e10433.

46. Lemon KP, Earl AM, Vlamakis HC, Aguilar C, Kolter R: The ArcBA two-component system of Ralstonia metallidurans controls iron acquisition and expression of the two-component FMN-dependent monooxygenase ActVA-ActVB from Streptomyces coelicolor. J Biotechnol 2009, 146:759–769.

47. Bekker M, Alexeeva S, Laan W, Sawers G, Teixeira De Mattos J, Hellingwerf K: Evolutionary divergence and convergence in the Conserved Residues His-246, His-199, and Tyr-255 in the Catalysis of Catechol 2,3-Dioxygenase from Pseudomonas putida. J Biol Chem 2008, 283:10297–10306.

48. Joseph S, Desai P, Ji Y, Cummings CA, Shih R, Brzoska P, Hambly SE, Masood N, Harri S, Sonbol H, Chuzhanova N, McClelland M, Furtado MR, Fuentes DE, Tantaleán JC, Vásquez CC: A family of Gram-negative bacterial proteins. J Biol Chem 2008, 283:10297–10306.

49. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389–3402.

50. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: Jalview Version 2: a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009, 25:1189–1191.

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

52. Saitou N, Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 1987, 4:406–425.

53. Zuckerkandl E, Pauling L: Evolutionary divergence and convergence in proteins. In Evolving Genes and Proteins. Edited by Blyson V, Vogel HJ. New York: Academic Press; 1965:166–166.

54. Taylor DE, Rooker M, Keanan M, Ng UK, Martin I, Perna NT, Buford NT, Blattner FR: Genomic variability of O islands encoding tellurite resistance in Enterohemorrhagic Escherichia coli O157:H7 isolates. J Bacteriol 2002, 184:4690–4698.

55. Viggiani A, Siani L, Notomista E, Biolo L, Pucci P, Di Donato A: The Role of the Conserved Residues His-246, His-199, and Tyr-255 in the Catalysis of Catechol 2,3-Dioxygenase from Pseudomonas stutzeri OX1. J Biol Chem 2004, 279:48630–48639.

56. Wesche J, Hammer E, Becker D, Burchhardt G, Schauer F: The bphC gene-encoded 2,3-dihydroxybiphenyl-1,2-dioxygenase is involved in complete degradation of dibenzofuran by the biphenyl-degrading bacterium Ralstonia sp. SBUG 290. J Appl Microbiol 2005, 98:635–645.

57. Pozzeula A, Lista L, Napolitano A, D’Ischia M: An expedient one-pot entry to catecholosserogen and other catechol compounds via IBX-mediated phenolic oxydgenation. Tetrahedron Lett 2005, 46:3341–3344.

Cite this article as: D’Argenio et al.: Complete sequencing of Novosphingobium sp. PP1Y reveals a biotechnologically meaningful metabolic pattern. BMC Genomics 2014 15:384.