Deciphering the Genome Repertoire of *Pseudomonas* sp. M1 toward β-Myrcene Biotransformation

Pedro Soares-Castro and Pedro M. Santos*

CBMA—Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus de Gualtar, Braga, Portugal

*Corresponding author: E-mail: psantos@bio.uminho.pt.

Accepted: November 21, 2014

Data deposition: The Whole Genome Shotgun project of *Pseudomonas* sp. M1 has been deposited at DNA Data Bank of Japan/EMBL/GenBank under the accession ANIR00000000. The version described in this paper is ANIR00000000.2.

Abstract

Pseudomonas sp. M1 is able to mineralize several unusual substrates of natural and xenobiotic origin, contributing to its competence to thrive in different ecological niches. In this work, the genome of M1 strain was resequenced by Illumina MiSeq to refine the quality of a published draft by resolving the majority of repeat-rich regions. In silico genome analysis led to the prediction of metabolic pathways involved in biotransformation of several unusual substrates (e.g., plant-derived volatiles), providing clues on the genomic complement required for such biodegrading/biotransformation functionalities. *Pseudomonas* sp. M1 exhibits a particular sensory and biotransformation/biocatalysis potential toward β-myrcene, a terpene vastly used in industries worldwide. Therefore, the genomic responsiveness of M1 strain toward β-myrcene was investigated, using an RNA sequencing approach. M1 cells challenged with β-myrcene(compared with cells grown in lactate) undergo an extensive alteration of the transcriptome expression profile, including 1,873 genes evidencing at least 1.5-fold of altered expression (627 upregulated and 1,246 downregulated), toward β-myrcene-imposed molecular adaptation and cellular specialization. A thorough data analysis identified a novel 28-kb genomic island, whose expression was strongly stimulated in β-myrcene-supplemented medium, that is essential for β-myrcene catabolism. This island includes β-myrcene-induced genes whose products are putatively involved in 1) substrate sensing, 2) gene expression regulation, and 3) β-myrcene oxidation and bioconversion of β-myrcene derivatives into central metabolism intermediates. In general, this locus does not show high homology with sequences available in databases and seems to have evolved through the assembly of several functional blocks acquired from different bacteria, probably, at different evolutionary stages.

Key words: genome sequencing, biocatalysis, terpenes, *Pseudomonas*, RNA-seq, genomic island.

Introduction

Pseudomonas spp. and their genomic features (e.g., multifaceted oxidative enzymatic repertoire) have been extensively explored as cell-factories and functional blocks for biocatalysis (Zeyer et al. 1985; Wubbolts and Timmis 1990; Di Gennaro et al. 1999; Schmid et al. 2001; Poblete-Castro et al. 2012), particularly in the production of fine chemicals for several industries (e.g., terpenoids) (Cantwell et al. 1978; Narushima et al. 1982; van der Werf et al. 1997; Förster-Fromme et al. 2006; Förster-Fromme and Jendrossek 2006; Bicas et al. 2009). The instability of some monoterpenes directed biotechnological research to their derivatives which are known to retain or enhance some properties of the original compound (van der Werf et al. 1997) and are usually easier to incorporate into hydrophilic/aqueous formulations.

The utilization of monoterpenes and their derivatives is widespread in industry because of their unique properties (Hocquemiller et al. 1991; Crowell 1999; Rasooli and Mirmostafa 2003). One of the most promising monoterpenes, displaying a plethora of industrial applications, is β-myrcene (7-Methyl-3-methylene-1,6-octadiene). β-myrcene is mainly used for the synthesis of top selling flavors and fragrances (e.g., linalool, nerol, geraniol, menthol) (Behr and Johnen 2009) and compounds with pharmacological potential (e.g., antimutagenics, analgesics, tyrosinase inhibitors), as well as a starting material in production of polymers, biodegradable surfactants, pheromones, and insect repellents (Kauderer et al. 1991; Matsuura et al. 2006; Behr and Johnen 2009).

In 1999, a *Pseudomonas* sp. (strain M1) was isolated from soil sediments of the Rhine River (Iurescia et al. 1999), which

exhibited the ability to use β-myrcene as sole carbon and energy sources. Transposon random mutagenesis led to the identification of four genes potentially coding for β-myrcene-biotransforming enzymes: An aldehyde dehydrogenase (*myrA*), an alcohol dehydrogenase (*myrB*), an acyl-CoA synthetase (*myrC*), and an enoyl-CoA hydratase (*myrD*) (Iurescia et al. 1999). Later, a proteomic analysis (Santos and Sá-Correia 2009) provided further insights into the β-myrcene catabolic pathway and associated adaptive mechanisms of *Pseudomonas* sp. M1. However, during that proteomic analysis, the authors outlined that a significant number of β-myrcene-responsive proteins were not identified due to the lack of information in the public databases, suggesting that the β-myrcene catabolic apparatus of *Pseudomonas* sp. M1 would be based, at least partially, on enzymes that have evolved divergently with respect to what have been reported in the literature so far. Curiously, in another work using *Pseudomonas* sp. M1, it was reported that its ability to use phenol and benzene as sole carbon sources relied on a unique σ^{54}-dependent gene cluster organization (Santos and Sá-Correia 2007), which, apparently, has evolved differently, when compared with other reported phenol catabolic pathways. Altogether, these hints suggest that *Pseudomonas* sp. M1 harbors, in its genomic repertoire, a unique code for biotechnologically relevant enzymes and functional modules, probably due to particular evolutionary constrains.

Nowadays, *Pseudomonas* sp. M1 is one of the most promising gram-negative bacteria able to metabolize β-myrcene, for which a catabolic pathway draft has been proposed (Iurescia et al. 1999; Santos and Sá-Correia 2009). *Pseudomonas* sp. M1 is also capable of mineralizing other terpenes (e.g., citronellol, citral), as well as several toxic and/or recalcitrant phenolic compounds, such as phenol, 4-chlorophenol, benzene, and toluene (Santos et al. 2002; Santos and Sá-Correia 2009).

The genome reconstruction conducted in this work updated the current available draft reported by Soares-Castro and Santos (2013) for exploitation of *Pseudomonas* sp. M1 full potential. A genomic analysis led to the identification of genomic loci that suggest a versatile metabolism of terpene-backbone compounds. Transcriptome sequencing of β-myrcene-challenged M1 cells revealed that the core-code of β-myrcene metabolism is located in a unique genetic locus and allowed the identification of the full set of M1 genes involved in β-myrcene-dependent molecular mechanisms of catabolism, adaption, and regulation.

Materials and Methods

Bacterial Strains, Culture Conditions, and Sampling

Pseudomonas sp. M1 (Iurescia et al. 1999) was grown in Lysogenic Broth Lennox for DNA extraction. For gene expression analysis, cells were cultivated in 50 ml of mineral medium (MM; 8.9 mM K_2HPO_4, 6.2 mM NaH_2PO_4, 34.2 μM EDTA, 7 μM $ZnSO_4$, 6.8 μM $CaCl_s$, 18 μM $FeSO_4$, 0.8 μM Na_sMoO_4, 0.7 μM $CuSO_4$, 1.7 μM $CoCl_2$, 1.9 μM $MnCl_2$, 15.1 mM $(NH_4)_2SO_4$, and 0.5 mM $MgCl_2$) supplemented with 0.4% lactate or 100 μl of β-myrcene (Merck) (CAS number: 123-35-3) as sole carbon source, in 250-ml Erlenmeyer-sealed flasks, at 30 °C and with orbital shaking (200 rpm). As the water solubility of β-myrcene is very low (density: 0.796 g/ml; log K_{ow}: 4.17; water solubility at 25 °C approximately 6 mg/l (Hansen et al. 2006), the amount (100 μl) of β-myrcene used in the experiments assured a constant water phase saturation with β-myrcene to avoid eventual carbon starvation effects being reflected on the overall transcriptome analysis.

For total RNA sample preparation, after overnight growth, M1 cells were inoculated to an initial OD_{600nm} of 0.08–0.1 in MM supplemented with respective carbon sources (0.4% lactate or 100 μl of β-myrcene) and cultures diluted again to OD_{600nm} of 0.08–0.1 when OD_{600nm} of 0.5 was reached to synchronize cell growth. Samples for RNA extraction were collected in triplicate (lactate: L1, L2, L3; β-myrcene: M1, M2, M3) when fresh cultures reached an OD_{600nm} of about 0.3–0.4 (2 h of exponential growth) and were immediately flash-frozen in liquid nitrogen.

Next-Generation Genome Sequencing, Assembly, and Annotation

The genomic DNA from *Pseudomonas* sp. M1 was extracted and purified using the Wizard Genomic purification kit (Promega). The DNA samples were processed according to Illumina instructions generating Nextera XT paired-end libraries (2 × 250 bp) which were sequenced using the Illumina MiSeq system at Yale Center for Genomic Analysis.

Raw paired-reads obtained from Illumina MiSeq platform (2 × 250 bp with insert size around 470 bp) and the previously obtained data set using Illumina GAIIx (2 × 50 bp paired-end reads with insert size around 320 bp) (Soares-Castro and Santos 2013) were both corrected with QUAKE (Kelley et al. 2010). These QUAKE-corrected read-data sets and the previously obtained data sets: Illumina GAIIx 2 × 50 bp mate-pair reads with insert size around 5,200 bp, and 454 FLX (single reads with average length of 523 bp) technologies (Soares-Castro and Santos 2013) were then all processed using fastq-mcf tool (Aronesty 2011), with which a quality filter was applied on all reads based on Phred quality scores (Ewing and Green 1998) and eventual adapter contamination, low quality, and ambiguous nucleotides were trimmed off from the remaining reads. A comparison of the overall quality of the resulting corrected and filtered data sets with the original data sets was carried out using FastQC (Andrews 2010).

Paired-end reads from MiSeq were de novo assembled using Edena version 3.131028 (Hernandez et al. 2008) and Spades version 2.5.1 (Bankevich et al. 2012) using different parameterization. To analyze the assembly quality, internal

SNVs and indels were estimated from reference assembly mapping of the filtered Illumina GAIIx paired-end reads against the respective de novo genome assemblies, using bowtie2 (Langmead and Salzberg 2012). Samtools version 0.1.19 (Li et al. 2009) and SNVer version 0.5.1 (Wei et al. 2011) were used to convert .SAM files to sorted .BAM files and detection of variant nucleotides, respectively. The tuned assemblies from Edena and Spades assemblers were merged using GAM-NGS assembly merger (Vicedomini et al. 2013), followed by automatically improvement of the resulting de novo genome assembly, using Pilon version 1.6 (http://www.broadinstitute.org/software/pilon/, last accessed October 1, 2014). Afterwards, filtered data sets from Illumina MiSeq and GAIIx (both) were used for scaffolding with SSPACE version 2.3 (Boetzer et al. 2011) and gap closure by using Gapfiller version 1.5 (Boetzer and Pirovano 2012) and GapCloser version 1.12 (Luo et al. 2012).

The final genome sequence draft of *Pseudomonas* sp. M1 was obtained following automatic improvement using Pilon, mapping of Sanger sequencing reads of regions evidencing abnormal coverage and/or gaps, mapping of 454 Roche FLX data set and manual curation. Genome annotation was carried out using National Center for Biotechnology Information (NCBI)'s Prokaryotic Genomes Annotation Pipeline 2.0 (Angiuoli et al. 2008) and metabolic pathways were predicted in silico by KAAS (Moriya et al. 2007).

Total RNA Extraction from Pseudomonas sp. M1

Total RNA samples were obtained using the Aurum Total RNA Mini kit (Bio-Rad), as described by the manufacturer's instructions. Prior to RNA sequencing (RNA-seq) and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) experiments, all genomic DNA remaining in RNA samples was digested with RNA-free DNase I. RNA integrity was verified in the Agilent 2100 Bioanalyzer (Agilent Technologies, CA) and only samples with RNA Integrity Number over 8.0 were used in RNA-seq and RT-qPCR experiments.

Differential Expression Analysis of RNA-Seq Samples from M1 Cells Grown in Lactate or β-Myrcene

Total RNA samples extracted from M1 cells were processed according to Illumina instructions generating libraries of 50-bp single reads, which were sequenced by Illumina GAIIx platform at BaseClear (the Netherlands), resulting in three data sets from M1 cells grown in lactate medium and three data sets from M1 cells grown in β-myrcene medium, which ranged from 20 million to 56 million reads.

Raw reads were trimmed and clipped for adaptor sequences and ambiguous nucleotides (N's) using fastq-mcf tool (Aronesty 2011). Expression levels were estimated by mapping and counting filtered RNA-seq data sets against the M1 draft genome using EDGE-pro pipeline (Magoc et al. 2013). Differential expression levels were determined as fold change (FC) by the ratio of normalized β-myrcene-derived read count values per lactate values using the EdgeR Exact Test (Robinson et al. 2010) with false discovery rate (FDR) *P* value correction. Genes were defined as significantly differentially expressed if they evidenced an altered FC ratio of 1.5 (or higher) and FDR smaller than 0.05. Genes having less than five reads mapped were considered to not be expressed.

Characterization of the 28-kb Genomic Island Harboring β-Myrcene Core-Code

In silico prediction of the transcriptional units (TUs), promoter and terminator regions was carried out by FGENESB tool from Softberry server (http://www.softberry.com/, last accessed October 1, 2014). RNA-seq based prediction of operons and transcription start sites (TSS) within this genomic island (GI) was carried out using ReadXplorer (Hilker et al. 2014).

Quantitative Expression of β-Myrcene-Biotransforming Key Genes by RT-qPCR

For RT-qPCR experiments, reverse transcription of RNA to cDNA was performed with the iScript cDNA Synthesis Kit (Bio-Rad). A total of 500 ng of RNA (from the same batch used for RNA-seq) was used for each sample and cDNA synthesized according to manufacturer's instructions (5 min at 25 °C, 30 min at 40 °C, and 5 min at 85 °C).

Expression of *PM1_0216305*, *PM1_0216325*, *PM1_0216335*, *PM1_0216350*, *myrA* (*PM1_0216395*), *myrD* (*PM1_0216400*), and *recA* (*PM1_0217885*) was analyzed in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using SensiFAST SYBR No-ROX Kit (Bioline) as the detecting agent. Primers used were synthesized by Stab Vida with following sequences: *PM1_0216305* (fwd: 5′-TATCGCAGCTTG GTCTTGGC-3′; rev: 5′-CTAACCAGTCACCAGACGCC-3′), *PM1_0216325* (fwd: 5′-AAAGGACTGGGCAAACTGGG-3′; rev: 5′-GCTGACGGCTTTGTCTTTGG-3′), *PM1_0216335* (fwd: 5′-CCGGATTACCGAGCTTCTGG-3′; rev: 5′-GGTGTTGGCTGT GTCAGGC-3′), *PM1_0216350* (fwd: 5′-TGTTCTCAAGGTCG CTGACG-3′; rev: 5′-TCCTGCAAATGCTTCAGGGG-3′), *myrA* (fwd: 5′-CGATACCGCTCTGCAACTCC-3′; rev: 5′-CCGCTGC GATGATTTTGTCG-3′), *myrD* (fwd: 5′-GGCGACTTACAGCGTT TTGC-3′; rev: 5′-GCTGGAGTGGTATGCAGAGC-3′), and *recA* (fwd: 5′-GGCCGAGTTCCAGATCCTCTAC-3′; rev: 5′-GATTTCC GCGTTCTCTTCCAGG-3′). Efficiency of each pair of primers (E) was determined with cDNA dilution curves (from 1 μg to 10 pg) from M1 cultures grown in lactate or β-myrcene, using the equation $E = 10^{(-1/\text{slope})} - 1$ (Rutledge and Côté 2003). Primers used showed efficiencies between 90% and 110%.

Each reaction was set up in a total volume of 20 μl, containing 100 ng of cDNA, SensiFAST SYBR No-ROX mix (Bio-Rad), 2 pmol of each gene-specific primer, and nuclease-free water. Amplification was performed at 95 °C for 2 min followed by 40 cycles at 95 °C for 5 s, 55 °C for 10 s, and 72 °C for 20 s. Specificity of primer annealing and presence of

amplification artifacts were assessed by the dissociation temperatures of amplification products in a melting curve analysis of temperatures between 65 °C and 95 °C (stepwise increase of 0.5 °C/5 s), with a continuous measurement of fluorescence. Negative control without cDNA template and control of genomic DNA contamination in RNA samples were performed for each pair of primers and each batch of RNA extracted. Eight independent RNA samples were used and all reactions were performed in triplicated.

Quantification cycle values (Cq) were determined with Bio-Rad CFX Manager software as the threshold cycle. Relative quantification of gene expression as FC was calculated by the $2^{-\Delta\Delta Cq}$ method taking into account primer efficiencies, as described by Pfaffl (2001) and using *recA* (*PM1_02298*) as reference gene for normalization (Manos et al. 2008; Greenwald et al. 2012). Error propagation was calculated according to Nordgård et al. (2006).

Phylogeny of Genes from the 28-kb GI

Phylogenetic analysis was carried for all 22 genes comprised in the 28-kb GI and for the *GMP synthase* (*guaA*, *PM1_216145*). Each gene was compared with sequences deposited in GenBank by BLASTn (Altschul et al. 1990) and the 20 most homologous entries were chosen for multialignment with MAFFT v7 (Katoh and Standley 2013). Using these alignments, maximum likelihood (ML) trees were inferred with PHYML v3.0 (Guindon and Gascuel 2003), using aproximate likelihood-ratio test nonparametric branch support based on a Shimodaira–Hasegawa-like procedure. For all the genes of the GI, the preferred model of nucleotide substitution was GTR (General Time Reversible) with a proportion of invariable sites. Similarities/dissimilarities between phylogenetic tree topologies may provide important clues on the relatedness of the evolutionary track of each gene of the GI under analysis in this work. Therefore, the phylogenetic tree-based parameters Log-likelihood score and patristic distance between multialigned sequences for each gene were deduced from the phylogenetic tree reconstruction and used to build a multivariate matrix. Hierarchical clustering analysis of the resulting matrix was carried by estimating the squared Euclidean distance between the different objects (gene alignment tree parameters), followed by agglomerative clustering with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method, using R statistical software with the gplots package (Warnes et al. 2009). The number of clusters was estimated by partitioning around mediods clustering algorithm, using the function pamk() in the R package fpc (Hennig 2010).

Results and Discussion

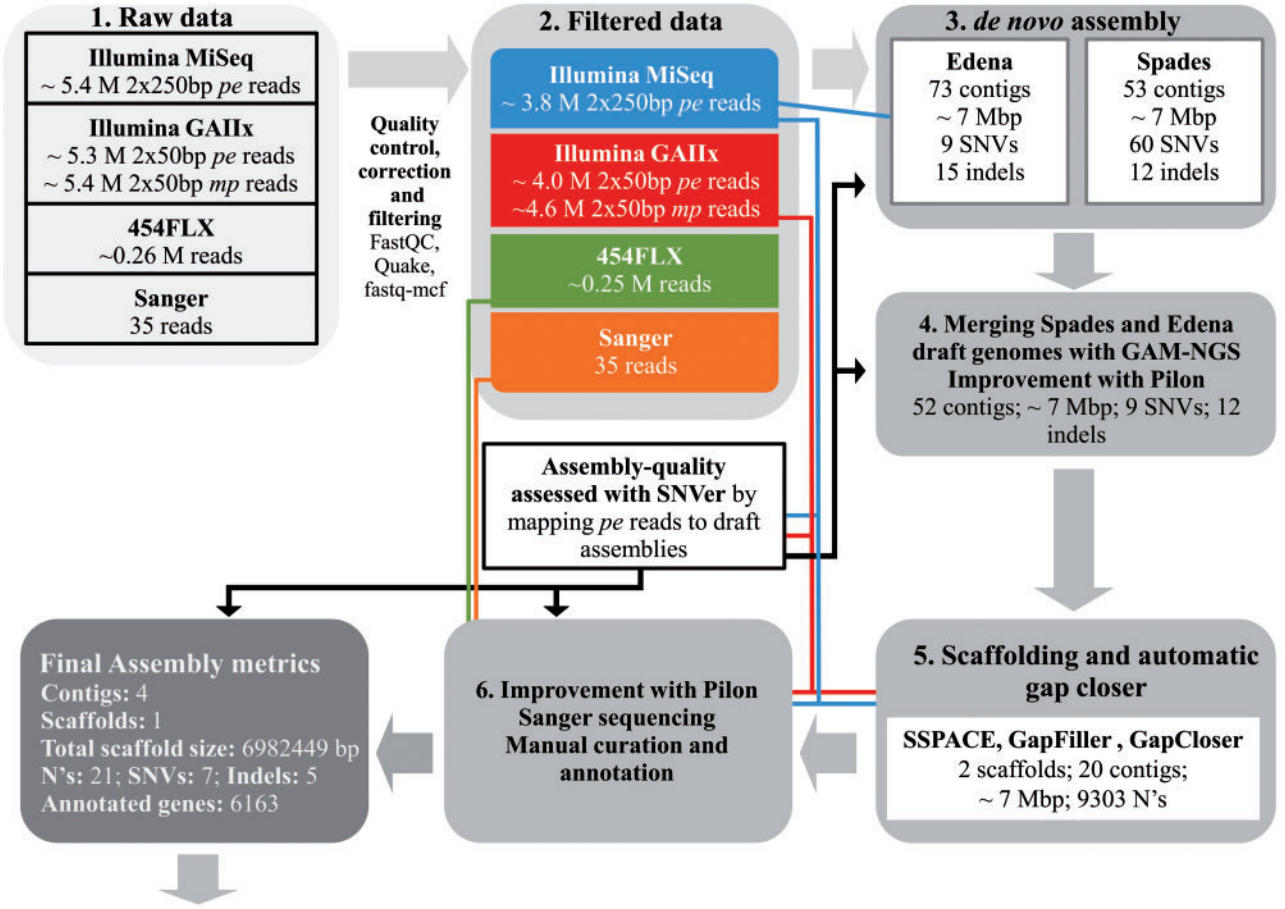
Whole-Genome Sequencing and Assembly

Previous work describing key players of β-myrcene catabolism in *Pseudomonas* sp. M1 (Iurescia et al. 1999; Santos and Sá-Correia 2009) was limited by the absence of reference genomic data. To set the proper background for exploring the biotransformation potential of *Pseudomonas* sp. M1, its genome was resequenced with Illumina MiSeq technology and assembled together with sequencing data sets obtained previously with Illumina GAIIx and 454 Roche FLX technologies (Soares-Castro and Santos 2013). The assembly pipeline is summarized in figure 1.

MiSeq-sequencing yielded 5,431,166 paired-end reads of 250 bp. Raw reads were then corrected for sequencing errors, trimmed, quality-filtered, and approximately 70% of the original data set (3,847,594 reads) was de novo assembled using Edena (Hernandez et al. 2008) and Spades (Bankevich et al. 2012). Assembly-quality and number of SNVs and indels were assessed by mapping-filtered Illumina GAIIx paired-end data set (4,052,990 reads) against de novo draft genome assemblies using SNver (Wei et al. 2011). Merging Spades and Edena draft genomes with GAM-NGS assembly merger (Vicedomini et al. 2013) and improvement with Pilon (http://www.broadinstitute.org/software/pilon/, last accessed October 1, 2014) yielded 52 contigs (longest contig size of 506,988 bp and N50 of 252,542 bp) with 9 SNVs and 12 indels. Scaffolding and gap closure were performed using the filtered data sets from Illumina MiSeq (3,847,594 paired-end reads) and GAIIx (4,052,990 paired-end reads and 4,698,623 mate-pair reads), which resulted in 20 contigs organized in two scaffolds with 9,303 unknown nucleotides (N's). After improvement using Pilon, mapping Sanger sequencing reads of regions evidencing abnormal coverage and/or gaps, mapping of the filtered 454 Roche FLX data set (253,532 single reads) and manual curation, the final improved genome draft of *Pseudomonas* sp. M1 was obtained in 4 contigs organized in 1 scaffold, with 21 unknown nucleotides (N's), 7 SNVs, 5 indels, and a total size of 6,982,449 bp. The resulting sequence was annotated with PGAP (Prokaryotic Genome Annotation Pipeline, NCBI) and deposited at DDBJ/EMBL/GenBank (accession number: ANIR00000000.2).

The improved draft of *Pseudomonas* sp. M1 genome includes 6,163 genes (6,023 coding sequences, 58 pseudogenes, 5 copies of each 5 S, 16 S, 23 S ribosomal RNA [rRNA] genes, 66 transfer RNA [tRNA] genes, and 1 noncoding RNA), with an average GC content of 67.2%. The major improvements of the current draft, when compared with the previous version, are reflected in the number of annotated genes and, importantly, in the reliability of base call along the genomic sequence, as deduced from the very low number of internal SNVs. Such improvement may have a significant impact on downstream functional analyses such as reference genome read mapping and protein identification.

The genome of *Pseudomonas* sp. M1 comprises a large set of oxygenases/hydroxylases-coding genes (99). Several enzymes belonging to this family have been shown to play an essential role on the initial oxidation steps of several natural and anthropogenic recalcitrant compounds. In silico prediction



Comparison of the current assembly with the previously released one

Genome assembly	version	Total contig size (bp)	Average coverage	Total number of genes	tRNAs	rRNAs	SNVs	Indels
ANIR00000000.1	v1 (previous)	6956853	150	6134	70	12	242	126
ANIR00000000.2	v2 (this study)	6982428	340	6163	66	15	7	5

Fig. 1.—Schematic representation summarizing the custom assembly pipeline used to reconstruct the genome of *Pseudomonas* sp. M1. Gray arrows indicate the pipeline stepwise flow. The differently colored lines associate each sequencing dataset with the phase(s) of the pipeline in which they were used.

of M1 versatile metabolism by KAAS (Moriya et al. 2007) suggested that the genomic repertoire of *Pseudomonas* sp. M1 may include the ability to engage in biotransformation/mineralization of recalcitrant compounds such as cresol and xylene, styrene, chloroalkanes and chloroalkenes, polycyclic aromatic compounds, naphtalene, benzoates, bisphenol, and atrazine (supplementary table S1, Supplementary Material online). Additionally, M1 nutritional versatility is probably prompted by respiratory flexibility in using N-oxides as alternative electron donors, as suggested by the presence of denitrifying clusters in the genome. Such genomic features enhance the potential application of M1 for bioremediation at contaminated oxygen-depleted foci and as a biosensor for in situ monitorization of aliphatic and aromatic hydrocarbons.

Genome sequencing also allowed identification of several genetic features putatively involved in the metabolism of plant-derived volatiles other than β-myrcene (supplementary table S1, Supplementary Material online), such as the acyclic terpene utilization cluster "atuRABCDEFGH" (*PM1_0225865–PM1_0225910*) that is described to be involved in

citronellol and related terpenes catabolism in *Pseudomonas citronellolis* (Förster-Fromme and Jendrossek 2006) and *Pseudomonas aeruginosa* strain PAO1 (Förster-Fromme et al. 2006), and some oxygenases and dehydrogenases putatively involved in the metabolism of limonene- and alpha-pinene-related terpenes. Thus, *Pseudomonas* sp. M1 seems to harbor an attractive set of enzymes for biocatalysis of some high-value terpenoids.

RNA-Seq of M1 Cells Challenged with β-Myrcene

Identification of genes involved in β-myrcene-stimulon was performed by comparative RNA-seq analysis of the transcriptomes of M1 cells grown in lactate (L) (control) and β-myrcene (M). Raw reads were trimmed and clipped for adaptor sequences and ambiguous nucleotides (N's) and mapped to a pool of rRNA and tRNA gene sequences from M1 draft genome to filter RNA-seq data sets exclusively for mRNA reads. To assess expression levels, mRNA reads were 99% uniquely mapped against M1 draft genome using EDGE-pro pipeline (Magoc et al. 2013) and the Bioconductor package EdgeR (Robinson et al. 2010). Among the large number of currently available tools for RNAS-seq analysis, this strategy was chosen based on: 1) EDGE-pro has been designed toward analysis of bacterial RNA-seq samples (in contrast with most of other read count applications which are directed toward the analysis of eukaryotic samples), and 2) EdgeR has been reported as one of the most used and reliable methods for data set normalization and statistical inference (Dillies et al. 2013).

As shown in table 1 (and more detailed in supplementary table S2, Supplementary Material online), the comparative transcriptome analysis revealed a number of different functions being altered when cells are cultivated with β-myrcene. In M1 challenged cells, 627 genes were found to be upregulated at least 1.5-fold (all evidencing an FDR value lower than 0.05) and 1,246 genes were found to be downregulated at least 1.5-fold (all evidencing an FDR value of 0.05 or lower), suggesting an extensive molecular reprogramming of cell physiology. Globally, the cultivation of M1 cells in β-myrcene induced a change in the expression levels of over 30% of the genes in its genomic repertoire.

Having in mind the assigned Clusters of Orthologous Groups (COG) categories, the most extensive alterations were registered in genes assigned to energy production and conversion; amino acid transport and metabolism; carbohydrate transport and metabolism; lipid transport and metabolism; transcription; inorganic ion transport and metabolism; and secondary metabolites biosynthesis, transport and catabolism. Importantly, β-myrcene metabolism seems to be dependent on expression of specific genetic loci, as deduced by a registered FC ratio greater than 10 for 52 of the upregulated genes.

A Unique 28-kb Genomic Island Harbors the myr+ *Genetic Trait*

β-myrcene supplementation induced the expression of 22 genes of a 28-kb genomic locus (*PM1_0216305–PM1_0216410*), from which 15 registered the highest expression levels from all β-myrcene-dependent transcriptome (fig. 2 and supplementary table S2, Supplementary Material online). Interestingly, the *myrDABC* cluster (*PM1_0216385–PM1_0216400*), previously identified by Iurescia et al. (1999) as essential for the initial steps of β-myrcene catabolism, is located in this 28-kb locus. Elements involved in conjugative transfer were found to be flanking this locus, which has an average GC content of 61.7% (genome average is 67.2%), indicating a hypothetical foreign nature for this genomic region.

To characterize this GI, TU were defined in silico by predicting operon-like organization of genes from the GI, the respective promoter regions and terminator sequences. The 22 genes seem to be organized in eight TUs: TU1, TU2, TU3, TU4, TU5, TU6, TU7, and TU8 (fig. 2*A*). Transcriptional organization was validated by mapping RNA-seq reads and de novo assembled contigs against the 28-kb GI, suggesting the polycistronic transcription of five of the eight TUs (fig. 2*B*). Moreover, RNA-seq based analysis of this GI also allowed the prediction of seven TSS, six of which overlapping putative promoter regions deduced from in silico analysis (fig. 2*A*). The detection of two TSS within TU3 may suggest that this TU may actually lead to different transcripts, depending on the growth conditions.

The first predicted TU comprises two acyl-CoA dehydrogenases (*PM1_0216305* and *PM1_0216310*) and an acetyl-CoA acetyltransferase, (*PM1_0216315*) putatively involved in β-myrcene catabolism, whose expression was increased approximately 365-, 370-, and 160-fold in β-myrcene-grown cells. A second cluster comprises two genes coding for a putative chemotaxis sensory transducer protein and a putative membrane protein (*PM1_0216320* and *PM1_0216325*). In the presence of β-myrcene, expression of this TU increased over 15-fold. From all 172 genes annotated on M1 genome as being involved in chemosensory systems and in signal transduction, 42 genes registered altered expression of at least 2-fold, from which only *PM1_0216320* and other signal protein-coding gene (*PM1_0226790*) were found to be upregulated in β-myrcene grown-cells, suggesting that TU2, together with *PM1_0226790*, might play essential role in recognition, transduction of the initial signal triggered by β-myrcene stimulus and possible uptake of β-myrcene (or derivative) molecules. OmpW family porins have been described to transport hydrophobic molecules across the membrane in some *Pseudomonas* spp. and *Escherichia coli* (Hong et al. 2006; Neher and Lueking 2009; Touw et al. 2010). Moreover, Neher and Lueking (2009) showed that mutation in OmpW of *Pseudomonas fluorescens* resulted in loss of ability to catabolize polycyclic aromatic molecules, such as naphtalene.

Table 1

Summary of Differential Expressed Genes in Myrcene-Grown *Pseudomonas* sp. M1 Cells, Organized in COG Categories

Functional Categories	Genome	RNA-Seq	
		Number of Genes	
		Up ≥ 1.5-Fold	Down ≥ 1.5-Fold
[A] RNA processing and modification	1	1	0
[B] Chromatin structure and dynamics	5	0	2
[C] Energy production and conversion	359	62	108
[D] Cell cycle control, cell division, chromosome partitioning	39	4	0
[E] Amino acid transport and metabolism	543	33	202
[F] Nucleotide transport and metabolism	100	6	15
[G] Carbohydrate transport and metabolism	211	10	76
[H] Coenzyme transport and metabolism	189	26	35
[I] Lipid transport and metabolism	243	51	67
[J] Translation, ribosomal structure and biogenesis	184	12	15
[K] Transcription	462	44	63
[L] Replication, recombination and repair	144	8	7
[M] Cell wall/membrane/envelope biogenesis	225	25	40
[N] Cell motility	178	6	50
[O] Posttranslational modification, protein turnover, chaperones	172	21	19
[P] Inorganic ion transport and metabolism	292	35	90
[Q] Secondary metabolites biosynthesis, transport and catabolism	160	18	51
[R] General function prediction only	490	40	101
[S] Function unknown	420	46	45
[T] Signal transduction mechanisms	301	21	52
[U] Intracellular trafficking, secretion, and vesicular transport	164	22	41
[V] Defense mechanisms	62	13	17
[W] Extracellular structures	1	0	1
Total genes with COGs[a]	4,463	469	974
No COG assigned	1,700	158	272
Total	6,163	627	1,246

[a]In total, 482 genes of the whole-genome annotation have multi-COG assignment. Further details are provided in supplementary table S2, Supplementary Material online.

TU3 codes for six products that might also be involved in β-myrcene oxidation: Another acetyl-CoA acetyltransferase (*PM1_0216330*) with increased expression of approximately 320-fold, a monooxygenase (*PM1_0216335*) increased by 431-fold, an enoyl-CoA hydratase (*PM1_0216340*) increased by 412-fold, an acyl-CoA dehydrogenase family protein (*PM1_0216345*) upregulated by 263-fold, an epoxide hydrolase (*PM1_0216350*) overexpressed by 102-fold, and a hypothetical protein (*PM1_0216355*) with increased expression by 51-fold. Moreover, TU4 is composed by a single gene, coding for a putative LuxR-family transcriptional regulator (*PM1_0216360*). This putative regulator registered increased expression of 190-fold in β-myrcene-grown cells. The fifth transcriptional unit is composed by two genes which code a rubredoxin (*PM1_0216365*), whose expression was increased by 260-fold, and a fatty acid desaturase (*PM1_0216370*), whose expression was increased by 202-fold. The sixth TU, comprises five genes which code for another LuxR-family transcriptional regulator (*PM1_0216375*), an oxidoreductase (*PM1_0216380*), the acyl-CoA synthase *myrC* (*PM1_0216385*), the alcohol dehydrogenase *myrB* (*PM1_0216390*), and the aldehyde dehydrogenase *myrA* (*PM1_0216395*). Expression levels were upregulated by 5-fold, 71-fold, 129-fold, 86-fold, and 44-fold, respectively. Gene coding for the enoyl-CoA hydratase *myrD* (*PM1_0216400*) comprises a seventh transcriptional unit and registered an increased expression of 220-fold. The eighth transcriptional unit predicted is composed by two genes which code for a putative LysR-family transcriptional regulator (*PM1_0216405*) and a hypothetical protein (*PM1_0216410*), whose expression was increased by 95- and 70-fold, respectively.

The high expression of such set of catabolic enzymes strongly suggests that this GI is the main genetic trait that provides *myr*+ phenotype to *Pseudomonas* sp. M1. Particularly, the monooxygenase-coding gene *PM1_0216335* showed the highest expression registered in all transcriptome and might be involved in initial oxidation of β-myrcene into a more soluble substrate. The understanding of its substrate range would be of utmost interest for terpene-based industry and

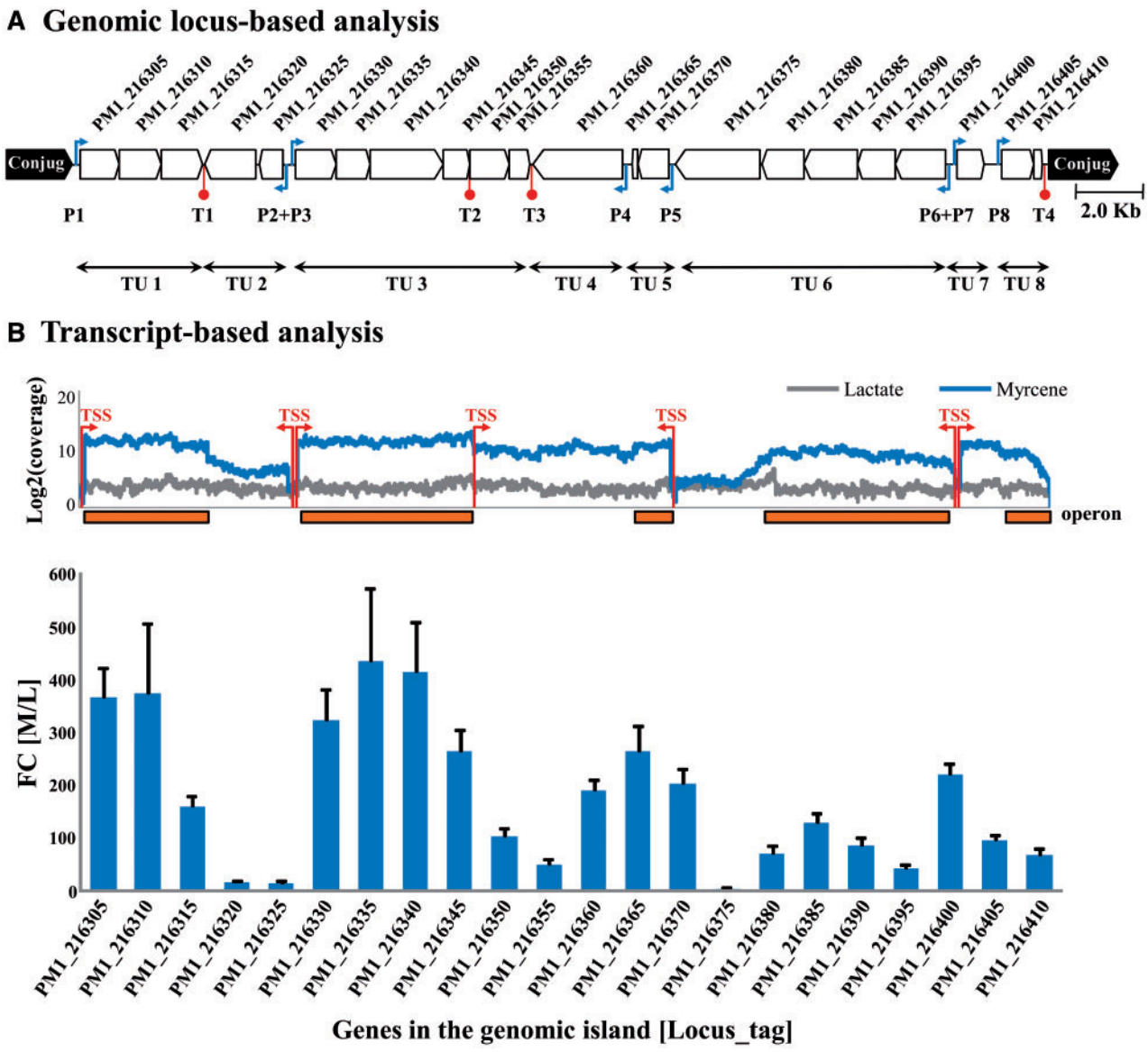


Fig. 2.—Global organization and functionality of the newly identified 28-kb GI harboring β-myrcene core-code. (*A*) Genomic locus organization and functional elements detected using in silico prediction. Genes transcribed in forward orientation are represented by rightward arrows, whereas genes located in antisense strand are represented by leftward arrows. Genes are organized in TU, promoter regions (P) are illustrated by blue arrows, and terminator sites indicated by red circles. (*B*) Detailed analysis of the RNA-seq data mapping with the GI. In the top panel, read mapping coverage per base was plotted for samples derived from M1 growth in lactate (gray line) and in β-myrcene (blue line). Prediction of the operon-like transcripts (horizontal orange bars) and transcription starting sites (TSS) (red arrows) was carried out with ReadXplorer (Hilker et al. 2014). Expression levels for each gene are shown, in the bottom panel, as the ratio (FC) of normalized transcripts between cells grown in lactate (L) and cells grown in β-myrcene (M).

future fine-tuning of its activity for biotechnological applications. Moreover, the differential expressed genes encoding for putative regulators (*PM1_0216360, PM1_0216405*, and *PM1_0216375*) suggest that they may be key players in the modulation of expression of the catabolic enzymes encoded within the GI.

Expression of predicted TUs was validated by RT-qPCR according to Pfaffl (Pfaffl 2001) and using *recA* (*PM1_02298*) as reference gene for normalization (Manos et al. 2008; Greenwald et al. 2012) (fig. 3). Similar to expression levels measured by RNA-seq, key genes *PM1_0216305, PM1_0216325*, *PM1_0216335*, *PM1_0216350*, *myrA,* and *myrD* were highly expressed in M1 cells challenged with β-myrcene, as showed by the FC ratio in figure 3. In general, the FC ratios obtained by RT-qPCR were higher when compared with RNA-seq-based estimated FC ratios.

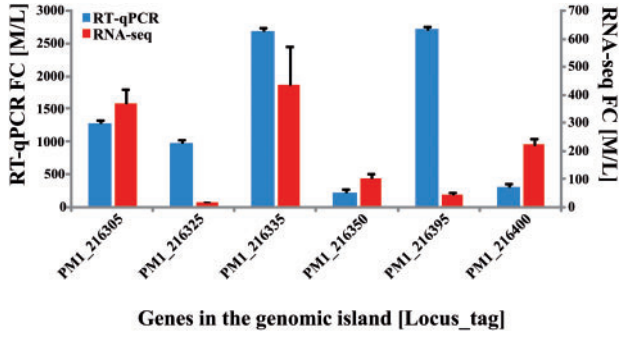


Fig. 3.—Expression levels of key genes from the 28-kb GI detected by RNA-seq and validated by RT-qPCR, shown as the ratio of normalized transcripts between cells grown in lactate and cells grown in β-myrcene (FC).

Although both methods have been proved to provide significant information regarding the evaluation of mRNA levels among different conditions, they rely on different methodological approaches which can differently affect the estimation of FC ratios (e.g., library preparation, data normalization, amount of starting template, and template amplification procedures). Nonetheless, both approaches clearly show that the expression of the selected genes is strongly stimulated by β-myrcene.

The 28-kb GI Evolved from the Assembly of Functional Blocks

As the GC content percentage within this GI has an average of 61.7% which is significantly different of the whole-genome average (67.2%), it was hypothesized that this GI may have been acquired as a whole DNA sequence segment through horizontal gene transfer. However, a thorough inspection of the public databases for DNA sequence deposition indicated that this GI was not present in any records besides the ones related with *Pseudomonas* sp. M1. In fact, the closest homologs of each predicted gene of the GI indicated a wide range of percentage of identity variation ranging from 56% to 87% (for the top hits), *PM1_261405* evidencing a closest similarity with the matched homologs (87%) and *PM1_216375* evidencing a lowest similarity (56%). To obtain clues on the origin of this GI, a phylogenetic analysis of all the genes of the 28-kb GI identified in this work was carried out. In this context, and due to the registered fairly low identity percentage between the genes of the GI and the respective homologs, only the 20 closest related sequences for all the genes (except *PM1_216355* for which only ten sequences resembling this gene were found) were selected for this analysis. Additionally, the analysis of the *guaA* gene of M1 strain and the respective 20 closest sequences was also performed. This gene was selected as a representative gene of M1 core genome as it evidences a GC content % similar to the average of the whole genome of *Pseudomonas* sp. M1 and it often used in multilocus sequence typing of *Pseudomonas* strains. In figure 4, and in more detail in supplementary table S3, Supplementary Material online, the results of the comparison of each gene of GI with their homologs are presented. The comparison of each gene of the GI with the respective 20 homologs indicates 1) a significantly lower average percentage of identity (57–77%) with respect to the 93% of identity for the reference gene *guaA* and 2) a significant difference in the GC% content of M1 genes with respect with the closest homologs (ranging 1–12% difference). These differences are supported, at least partially, by the diversified taxonomic origin of the closest homolog sequences, as depicted in figure 4. As expected, the results for the representative *guaA* gene are in agreement with previously published phylogenetic analysis of *Pseudomonas* sp. M1 based on 16 S rRNA sequencing (Santos et al. 2007) and groups M1 within the Gammaproteobacteria class (and *Pseudomonas* genus). In contrast, most of the genes of the GI, with the exception of *PM1_216320*, share significant identity with homolog sequences from organisms belonging to a taxonomic class not related with *Pseudomonas* sp. M1. In the case of *PM1_216305*, *PM1_216310*, *PM1_216315*, *PM1_216320*, *PM1_216325*, *PM1_216330*, *PM1_216335*, *PM1_216340*, *PM1_216355*, *PM1_216360*, *PM1_216365*, *PM1_216375*, *PM1_216405*, and *PM1_216410* genes, 75% (or more) of their homolog sequences belong to organisms of a specific taxonomic class. Distinctively, in the case of *PM1_216345*, *PM1_216350*, *PM1_216370*, *PM1_216380*, *PM1_216385*, *PM1_216390*, *PM1_216395,* and *PM1_216400* genes, the taxonomic distribution of their homolog sequences is spread among different classes. This asymmetric taxonomic distribution of the homolog sequences of the GI may suggest that the genetic reservoirs of some of the genes within the GI are confined to specific taxonomic groups whereas homologs of the other genes are widely distributed across different taxonomic classes. Moreover, this asymmetry also suggests that, probably, the GI resulted from acquisition of DNA segments from different environmental reservoirs which were assembled together evolving toward gene specialization for β-myrcene utilization. This hypothesis is in agreement with the multiple parallel evolutionary origins of GIs outlined before (Juhas et al. 2008).

Each gene of the GI, the *guaA,* and the most the selected most homolog sequences (analyzed above) were aligned using MAFFT. The resulting multialignments were used to construct a phylogenetic tree for each gene (GI and *guaA*) using an ML method. The respective −Log likelihood (−ln *L*) scores and divergence distance with respect to each *Pseudomonas* sp. M1 gene (estimated as the average of the patristic distances between tree tips and the respective M1 gene) were extracted. The −ln *L* score reflects the overall phylogenetic tree topology and branch lengths as a consequence of the relatedness of the sequences under analysis, being higher values an indication of higher phylogenetic distances.

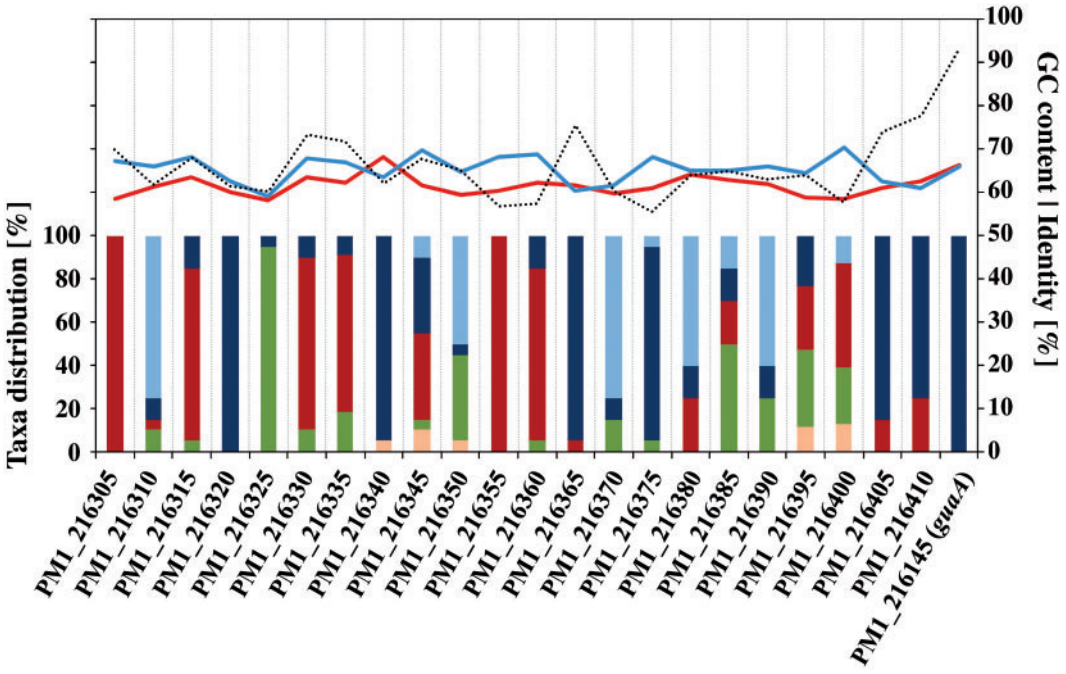


Fig. 4.—Graphical representation of the analysis of the results of multialignments of each gene of the GI implicated in β-myrcene catabolism in *Pseudomonas* sp. M1 and PM1_216145 (*guaA*), and the respective most homolog genes, retrieved using BLASTn. The dotted line indicates the average percentage of identity with respect to M1 genes. The average of GC percentage of each M1 gene or of the homologs sequences is represented by blue or red lines, respectively. The taxonomic origin of each homolog sequence was assessed and grouped according the respective taxonomic class. The bars in graph represent relative distribution (in percentage of the total number homolog sequences) of Alphaproteobacteria (green), Actinobacteria (light blue), Betaproteobacteria (red), Gammaproteobacteria (dark blue), and other classes (light orange).

In fact, in this study it was verified that, in general, the −ln *L* increased when the number of taxonomic genus represented in the multialignment sequences increased (data not shown). Presumably, M1 genes evidencing closer −ln *L* scores and divergence distances, based on the comparison with the selected homolog sequences, may share more similar evolutionary histories. Therefore, a multivariable matrix including the −ln *L* scores and divergence distances for all the genes of the GI and *guaA* was constructed. In figure 5, the agglomerative clustering analysis of the resulting matrix is shown. As expected from the results presented in figure 4, the reference gene *guaA* clustered distantly from the genes within the GI as the homolog sequences share a higher degree of identity. Remarkably, this analysis suggests that most of the genes within TU3 and TU6 (fig. 2) may share a close evolutionary history as suggested by their inclusion in the same cluster partition, C9 and C6 of figure 5, respectively. Additionally, *PM1_216320*/*PM1_216325* and *PM1_216405*/*PM1_216410* pairs of genes cluster close to each other and correspond to the predicted TU2 and TU8, respectively. These observations are in agreement with the correlation between coexpression and coevolution of genes encoding enzymes with coupled metabolic fluxes previously reported (Seshasayee et al. 2009). Moreover, *PM1_216320, PM1_216360,* and *PM1_216375* genes, that putatively code for three of the four transcriptional regulators present in the GI, cluster closer (clusters C1, C2, and C3 of fig. 5, respectively) and apart from the catabolic genes. Significantly, the highest values for the −ln *L* and/or average patristic distance parameters were registered for these genes. Our results may indicate that the control of the catabolic genes of the GI requires an intricate tuning of the regulatory sequences towards an effective responsiveness to β-myrcene (and possibly other terpenes), probably due to the unusual nature of the substrates. Thus, it suggests that the regulatory and the catabolic components of the GI may have evolved separately. This hypothesis is in agreement with a modular assembly concept of emerging catabolic pathways that includes evolution from constitutive to regulated expression of the catabolic enzymes, as previously described (Cases and de Lorenzo 2001).

β-Myrcene-Induced Genetic Repertoire Might Code for Enzymatic Redundancy toward a Broader Range of Terpene-Substrates

Transcriptome data not only support the pathway described by Iurescia et al. (1999) but also strongly suggest the involvement of several other gene products in β-myrcene oxidation and raise new hypotheses about β-myrcene metabolic

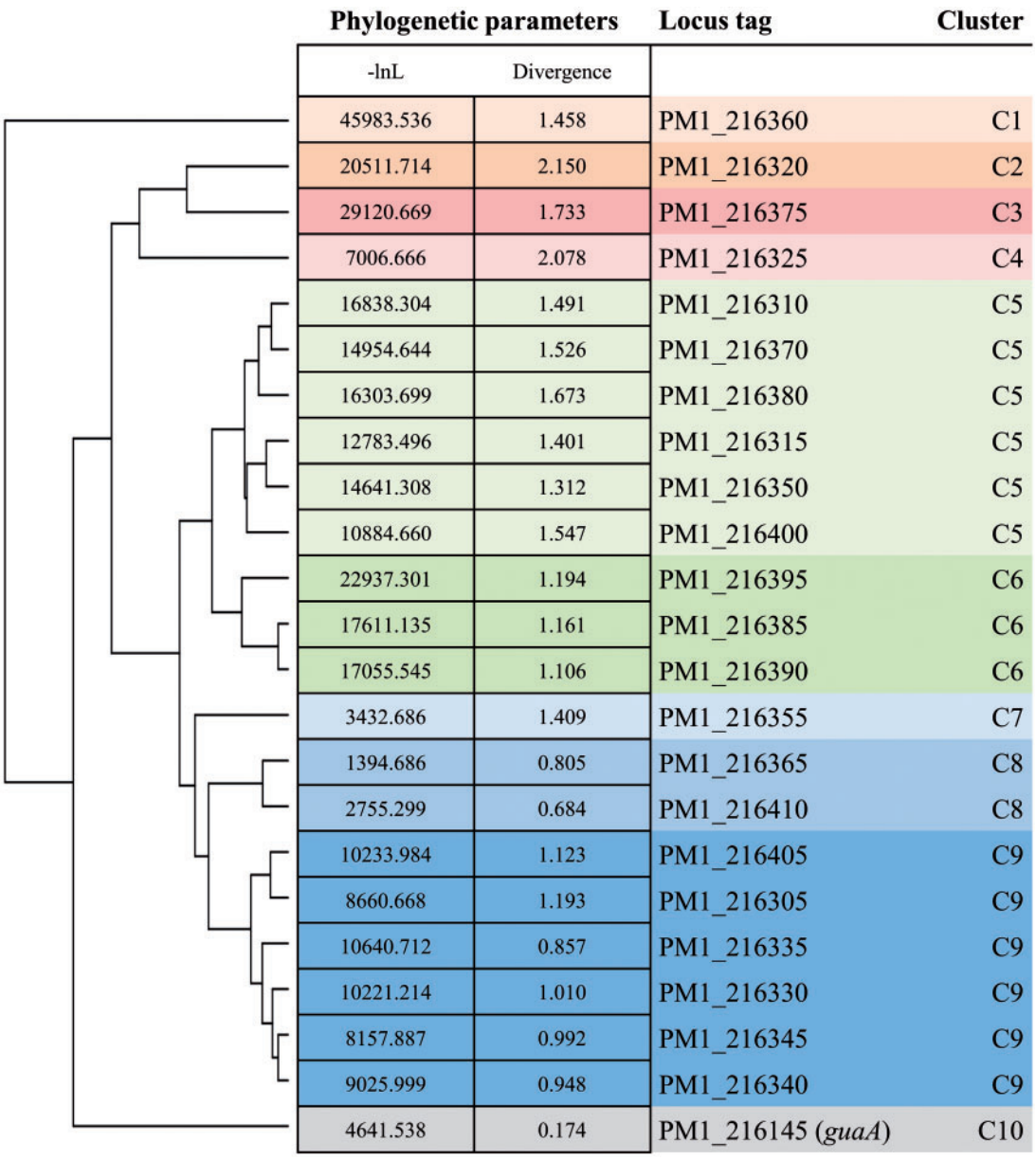


Fig. 5.—Multivariate clustering analysis of the inferred phylogenetic tree-based parameters −log likelihood score (−ln *L*) and average patristic distance (Divergence) for each gene of the GI. For a comparison purpose, the parameters for the representative gene *guaA* were also included in the analysis. Each M1 gene, represented in the figure, was multialigned together with the most homolog sequences and clustering analysis of the inferred parameters was carried by estimating the squared Euclidean distance between the different sequences in each multialignment, followed by agglomerative clustering with UPGMA method. The number of clusters was estimated by partitioning around mediods clustering algorithm. Each cluster is outlined by a different color.

pathway that will be the focus of following studies. At this moment, it is unclear whether all 22 genes from the 28-kb GI are actually necessary for the metabolism of the aliphatic C8-chain of β-myrcene or whether the high expression levels of the island are induced by a chemosensory signal common to a broader metabolic potential toward several terpene-backbone compounds. The expression of the putative epoxide hydrolase (*PM1_0216350*) when *Pseudomonas* sp. M1 is grown in β-myrcene medium may indicate a possible epoxidation of β-myrcene for subsequent hydrolysis by PM1_0216350 into a β-myrcene alcohol. However, terpene-epoxide microbial biotransformation has been reported mainly as result of monooxygenase activity toward cyclic molecules (e.g., alpha-pinene and limonene) (van der Werf et al. 1997).

Therefore, it is also possible that the putative epoxide hydrolase encoded by the GI may be actually involved in epoxidation of substrates other than β-myrcene.

β-myrcene biotransformation has been reported in other gram-negative bacteria. *Rhodococcus erythropolis* MLT1 was described to biotransform aerobically β-myrcene into geraniol (Thompson et al. 2010). In *Castellaniella defragans* 65Phen β-myrcene is hydratated to linalool, following isomerization to geraniol and oxidation to geranic acid, under anaerobic conditions (Brodkorb et al. 2010; Lüddeke and Harder 2011; Lüddeke, Dikfidan, et al. 2012; Lüddeke, Wülfing, et al. 2012), which indicates a cross-talk between catabolic pathways of different terpenes. In M1 cells challenged with β-myrcene, the majority of genes from the acyclic terpene utilization cluster *atuRABCDEFGH* and the leucine/isovalerate degradation cluster *liuRABCDE* (*PM1_0223355–PM1_0223380*) were found to be upregulated (average cluster expression induced by 3.8-and 4.1-fold, respectively). The overexpression of these two clusters suggests that, similarly to *R. erythropolis* and *C. defragans* (Förster-Fromme et al. 2006; Förster-Fromme and Jendrossek 2006), a citronellol/geraniol-like compound might be an intermediate/byproduct of β-myrcene metabolism and channel β-myrcene-derivatives through propionyl-CoA into central metabolism (fig. 6). Another hypothesis is that the sensory system present in the GI may be involved in the recognition of a spectrum of terpene-backbone molecules, including β-myrcene, and trigger expression of several genetic loci with complementary/redundant degradative functions (fig. 6).

Several other genes located outside of the 28-kb locus and coding for beta-oxidation-like enzymes had their expression altered in M1 cells grown in β-myrcene medium. As shown in supplementary table S2, Supplementary Material online, increased expression by at least 2-fold was detected for 8 acyl-CoA-like dehydrogenases (*PM1_0206520, PM1_0208500, PM1_0211935, PM1_0220850, PM1_0221270, PM1_0224730, PM1_0224740,* and *PM1_0224920*), a putative acetyl-CoA acetyltransferase (*PM1_0224735*), a putative acetyl-CoA synthetase (*PM1_0208785*), 2 putative acyl-CoA thiolases (*PM1_0221265* and *PM1_0212440*), 2 putative enoyl-CoA hydratases (*PM1_0221070* and *PM1_0221075*), a putative enoyl-CoA reductase (*PM1_0225900*), and 2 putative aldehyde dehydrogenases (*PM1_0226785* and *PM1_0230170*). In β-myrcene-grown *C. defragans* and *R. erythropolis*, overexpression of additional acyl-CoA dehydrogenases and oxidoreductases was detected (Thompson et al. 2010; Lüddeke, Wülfing, et al. 2012). Furthermore, kinetic assays showed that geraniol dehydrogenase GeoA of *C. defragans* was able to catalyze either cyclic (e.g., perillyl-alcohol) and acyclic monoterpenes (e.g., geraniol) (Lüddeke et al. 2012), and its deletion did not impair completely growth on β-myrcene medium, suggesting the existence of secondary enzyme system acting on acyclic terpene-backbone molecules (Lüddeke, Dikfidan, et al. 2012). Similarly, mutagenesis trials in *Pseudomonas* sp. M1 (e.g., *ΔmyrB* mutant) did not abolish completely the ability of M1 to grow in β-myrcene medium (data not shown), supporting the redundancy of beta-oxidation-like enzymes overexpressed during growth in β-myrcene medium and suggesting a cross-talk between metabolic fluxes of terpene catabolism and a complementary substrate range for these set of enzymes (fig. 6).

Central Metabolic Flux Is Changed in β-Myrcene-Grown Cells

Concomitant with the upregulation of the 28-kb, expression of almost all genes coding enzymes of the Entner–Doudoroff pathway was downregulated (supplementary table S2, Supplementary Material online). On the other hand, genes coding for enzymes of the tricarboxylic acid cycle (TCA) show similar expression levels in M1 cells grown in lactate or β-myrcene (fig. 6 and supplementary table S2, Supplementary Material online), with exception of the citrate conversion to isocitrate as the *acnA* gene (*PM1_0223060*) coding for aconitate hydratase registered 2.4-fold increase in expression. Subsequently, isocitrate might be channeled to the glyoxylate shunt, mediated by a 2.6-fold increase in expression for *PM1_0221765* (coding for a isocitrate lyase AceA-like). The glyoxylate bypass might increase levels of succinate for energy production, surpassing the metabolic branching of TCA intermediaries (e.g., 2-oxoglutarate) for amino acid biosynthesis. Also, glyoxylate shunt might increase levels of malate, which is further converted to oxaloacetate, a key branching step in the central metabolism. *Pseudomonas* sp. M1 genome harbors *mqoB* gene (*PM1_0214580*) coding for a malate:quinone oxidoreductase and a gene coding for a putative malate dehydrogenase (*PM1_0229630*), being the former upregulated 2.0-fold in M1 cells grown in β-myrcene. In *P. citronellolis*, *mqoB* expression has been described to be essential for the metabolism of the acyclic terpenes citronellol and citronellic acid (Förster-Fromme and Jendrossek 2005). A possible bypass of isocitrate toward the glyoxylate shunt in β-myrcene-grown cells supports previous proteomic data (Santos and Sá-Correia 2009), which suggested that TCA flux might be directed toward oxaloacetate synthesis as this intermediate might be used in condensation reaction with a β-myrcene derivative, such as propionyl-CoA (Textor et al. 1997). This would channel the final intermediates of β-myrcene catabolism to the central metabolism through 2-methycitrate cycle (e.g., in form of succinate and pyruvate) (fig. 6). In fact, the expression of all five genes coding for 2-methycitrate cycle enzymes (*PM1_0215080–PM1_0215100*) was also increased in β-myrcene-grown cells (average upregulation of 10.6-fold).

β-Myrcene Catabolism Induces an Anaerobiose-Like Physiology

RNA-seq data also suggested that M1 cells may use alternative pathways for energy production during β-myrcene

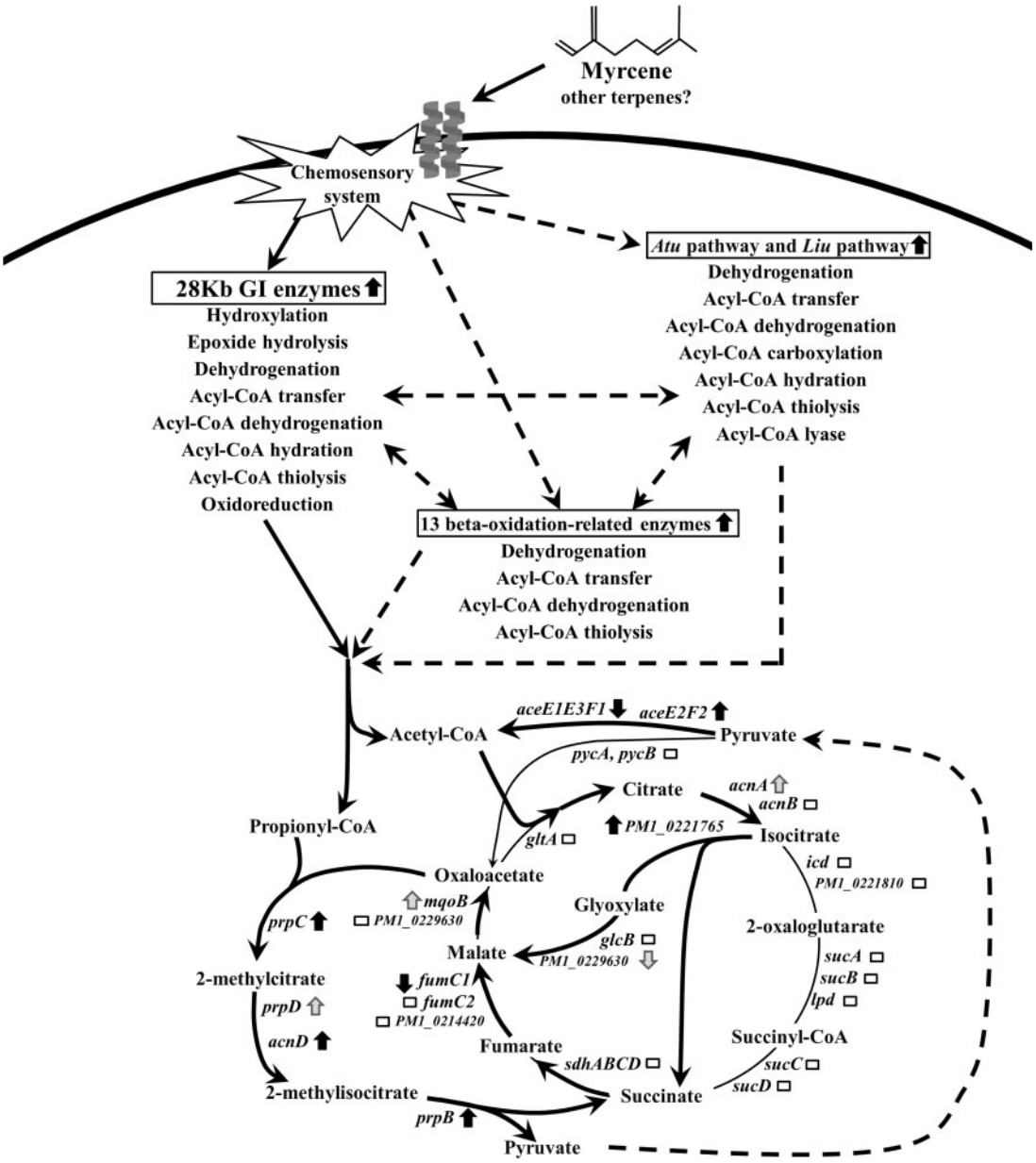


Fig. 6.—Schematic prediction of β-myrcene metabolic pathway and redistribution of central metabolic fluxes in M1 cells challenged with β-myrcene. Upregulated genes are shown by upward black (>2-fold) or gray (<2-fold) arrows and downregulated genes are shown by downward black (>2-fold) or gray (<2-fold) arrows. Genes with similar expression levels in *Pseudomonas* sp. M1 cells grown in lactate medium and β-myrcene medium are indicated by white squares.

catabolism, resembling an oxygen-depletion-like physiology (supplementary table S2, Supplementary Material online). Expression levels of four clustered genes coding for Dnr-like regulators (*PM1_0227665, PM1_0227670, PM1_0227660,* and *PM1_0227675*) were upregulated 7.7-, 2.5-, 2.1-, and 1.9-fold, respectively. Furthermore, Dnr-like regulators were probably responsible for inducing the transcription of several denitrification operons: Average 2.3-fold induction of the *narK1K2GHJI* (*PM1_0215670–PM1_0215695*), an average 38.5-fold induction of the *norCBD* cluster (*PM1_0227680–PM1_0227690*), an average induction of 6.2-fold for the *nirQOP* cluster (*PM1_0227695–PM1_0227705*), 7.1-fold increase in expression of *nirM* (*PM1_0227710*), and an average induction of 3.3-fold for *nosRZDFYL* cluster (*PM1_0227715–PM1_0227740*), coding the set of reductases and cytochromes necessary for N-oxides respiration (Arai et al.

2003). Moreover, expression of genes from the Anr-dependent operon *arcDABC* (*PM1_0202140–PM1_0202155*), coding for enzymes of the arginine deiminase pathway, was also induced at least 4-fold in β-myrcene-grown M1 cells. This operon is responsible for arginine fermentation with the concomitant formation of ATP (Gamper et al. 1991). Arginine fermentation might have been induced at earlier stages as a source of energy production that would be gradually replaced by nitrate respiration as Anr–Dnr transcriptional cascade activated their target genes and as β-myrcene metabolism demanded higher yields of energy production. Catabolism of some terpenes in nitrate-respiring conditions has already been reported for *C. defragrans* (e.g., alpha-pinene, β-myrcene, and limonene) (Foss et al. 1998; Heyen and Harder 2000; Brodkorb et al. 2010) and *Thauera* spp. (e.g., linalool, menthol, and eucalyptol) (Foss and Harder 1998), mainly based on metabolite identification. In this study, a molecular snapshot shed some light in what might be a common mechanism for terpene-backbone metabolism under denitrifying conditions.

M1 Cells Reprogram Their Physiology during β-Myrcene Catabolism

As described above, β-myrcene stimulus activates specific pathways on *Pseudomonas* sp. M1, which in turn appears to switch off genes that might become a metabolic burden for growing cells. M1 cells grown in β-myrcene show several secondary metabolic pathways and respective transport systems downregulated when compared with cells grown in lactate (supplementary table S2, Supplementary Material online), mainly those related to amino acids biosynthesis (e.g., glutamate, glutamine, serine, histidine, lysine, arginine, proline), utilization of alternative carbon sources (e.g., aromatic compounds, sugars), degradation of urea, amines, and polyamines.

The altered expression of genes involved in lipopolysaccharide (LPS) and exopolysaccharide synthesis suggested a reorganization of cell envelope in β-myrcene grown cells. Expression of a cluster harboring the *wzm-wzt-wbpXYZ* genes (*PM1_0228115–PM1_0228170*) (Rocchetta et al. 1999; Greenfield and Whitfield 2012) involved in LPS biogenesis was downregulated over 1.5-fold, as well as a gene coding for the LptA protein (*PM1_0228970*), an element of the LPS transport system across the outer membrane (Tran et al. 2008), whose expression was downregulated by 1.4-fold. The expression of the main cluster associated with alginate synthesis, *algD-8-44-KEGXLIJFA* (*PM1_0215545–PM1_0215600*) (Chitnis and Ohman 1993) was downregulated about 2-fold, suggesting a reduction in the synthesis of this polymer. Changes in outer cell envelope were, apparently, followed by changes in membrane lipid composition, membrane stability and permeability as many genes coding for the synthesis of lipid polar head group, transporters/facilitators, porins and membrane proteins had their expression changed when comparing with lactate grown-cells. In β-myrcene-grown cells, expression of two genes coding for OprD homologous (*PM1_0203585* and *PM1_0218170*) was increased by 70.8-fold and decreased by 10.0-fold, respectively. Moreover, expression of a set of efflux pumps was increased more than 2-fold, including six genes coding for Resistance-Nodulation-Cell Division family efflux systems (*PM1_0222160, PM1_0210745, PM1_0222165, PM1_0217125, PM1_0217120,* and *PM1_0202020*), a gene homologous to fusaric acid-inducible resistance pumps (*PM1_0210750*) and two genes coding for a multidrug resistance transporter system (*PM1_0202030* and *PM1_0202025*) with homology to *emrAB* cluster of *E. coli*. On the other hand, OmpQ-family protein-coding genes *PM1_0230840* and *PM1_0213495* were downregulated to 1.8- and 2.0-fold, respectively; *opdE* (*PM1_0220290*) was downregulated to 1.8-fold; and *opdH* (*PM1_0216660*) expression was reduced to 2.0-fold.

Abundance of porins changes in response to nutritional requirements of the cell and control membrane permeability toward stressors (Yoneyama et al. 1995; Tamber et al. 2006; Rivera et al. 2008). Additionally, efflux systems are usually described to be upregulated in stressful conditions, being responsible for cell homeostasis by exporting reactive and toxic compounds (Furukawa et al. 1993; Ramos et al. 2002; Hu et al. 2012). Furthermore, such cell envelope dynamics might account for the adaptive mechanisms toward β-myrcene hydrophobicity, which might affect overall membrane integrity and induce cellular stress (Sikkema et al. 1995; Turina et al. 2006; Cox and Markham 2007). Increased membrane hydrophobicity was previously described in *Pseudomonas* strains challenged by other hydrophobic compounds, such as n-alkanols and phenanthrene, as a mechanism to enhance biotransformation (Heipieper et al. 2007; Baumgarten et al. 2012). Overall, M1 cells grown in β-myrcene, apparently, leads to a reorganization of membrane structure and functions, most probably to cope with the highly hydrophobic character of β-myrcene and eventual generation of β-myrcene-derived toxic intermediates, as previously evidenced (Santos and Sá-Correia 2009). Additionally, changes in M1 cell envelope structure and functions may account for a strategy to increase substrate availability/contact surface at the oil–water interface by changing cell surface polarity and control levels of stress by changing membrane permeability and stiffness.

Detailed characterization of expression kinetics and regulation are required to fully understand the β-myrcene-stimulon. The sensory system of β-myrcene shows high sensitivity in detecting β-myrcene highly hydrophobic molecules, being an interesting candidate for biosensory applications. Currently, three enzymes, the monooxygenase PM1_0216335, the epoxide hydrolase PM1_0216350 and the 2-methylcitrate synthase PrpC (PM1_0215095), are in protein

crystallization trials (data not shown) for future fine-tuning and source of novel biomolecules. Nevertheless, key players involved in the initial steps of β-myrcene metabolism by *Pseudomonas* sp. M1 and the range of plant-derived volatiles that may be biotransformed by the 28-kb GI, as a functional block and by M1 as a cell-factory, still remain to be elucidated and will be focus of following work.

Supplementary Material

Supplementary tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by FEDER through POFC—COMPETE and by national funds from Foundation for Science and Technology (Portugal) through the projects PEst-C/BIA/UI4050/2011, PTDC/EBB-BIO/104980/2008 and PTDC/BIA-MIC/113733/09, and through a PhD grant (grant number SFRH/BD/76894/2011) to P.S.-C.

Literature Cited

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

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. [cited 2014 Oct 1]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Angiuoli SV, et al. 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS 12:137–141.

Arai H, Mizutani M, Igarashi Y. 2003. Transcriptional regulation of the *nos* genes for nitrous oxide reductase in Pseudomonas aeruginosa. Microbiology 149:29–36.

Aronesty E. 2011. Command-line tools for processing biological sequencing data. [cited 2014 Oct 1]. Available from: http://code.google.com/p/ea-utils.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19: 455–477.

Baumgarten T, et al. 2012. Alkanols and chlorophenols cause different physiological adaptive responses on the level of cell surface properties and membrane vesicle formation in *Pseudomonas putida* DOT-T1E. Appl Microbiol Biotechnol. 93:837–845.

Behr A, Johnen L. 2009. β-myrcene as a natural base chemical in sustainable chemistry: a critical review. ChemSusChem. 2:1072–1095.

Bicas JL, Dionísio AP, Pastore GM. 2009. Bio-oxidation of terpenes: an approach for the flavor industry. Chem Rev. 109:4518–4531.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579.

Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol. 13:R56.

Brodkorb D, Gottschall M, Marmulla R, Lüddeke F, Harder J. 2010. Linalool dehydratase-isomerase, a bifunctional enzyme in the anaerobic degradation of monoterpenes. J Biol Chem. 285:30436–30442.

Cantwell SG, Lau EP, Watt DS, Fall RR. 1978. Biodegradation of acyclic isoprenoids by *Pseudomonas* species. J Bacteriol. 135:324–333.

Cases I, de Lorenzo V. 2001. The black cat/white cat principle of signal integration in bacterial promoters. EMBO J. 20:1–11.

Chitnis CE, Ohman DE. 1993. Genetic analysis of the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* shows evidence of an operonic structure. Mol Microbiol. 8:583–590.

Cox SD, Markham JL. 2007. Susceptibility and intrinsic tolerance of *Pseudomonas aeruginosa* to selected plant volatile compounds. J Appl Microbiol. 103:930–936.

Crowell PL. 1999. Prevention and therapy of cancer by dietary monoterpenes. Nutrition 129:775S–778S.

Di Gennaro P, et al. 1999. A new biocatalyst for production of optically pure aryl epoxides by styrene monooxygenase from *Pseudomonas fluorescens* ST. Appl Environ Microbiol. 65: 2794–2797.

Dillies MA, et al. 2013. French StatOmique Consortium. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 14: 671–683.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.

Förster-Fromme K, et al. 2006. Identification of genes and proteins necessary for catabolism of acyclic terpenes and leucine/isovalerate in *Pseudomonas aeruginosa*. Appl Environ Microbiol. 72: 4819–4828.

Förster-Fromme K, Jendrossek D. 2005. Malate:quinone oxidoreductase (MqoB) is required for growth on acetate and linear terpenes in *Pseudomonas citronellolis*. FEMS Microbiol Lett. 246:25–31.

Förster-Fromme K, Jendrossek D. 2006. Identification and characterization of the acyclic terpene utilization gene cluster of *Pseudomonas citronellolis*. FEMS Microbiol. Lett. 264:220–225.

Foss S, Harder J. 1998. *Thauera linaloolentis* sp. nov. and *Thauera terpenica* sp. nov., isolated on oxygen-containing monoterpenes (linalool, menthol, and eucalyptol) nitrate. Syst Appl Microbiol. 21: 365–373.

Foss S, Heyen U, Harder J. 1998. *Alcaligenes defragrans* sp. nov., description of four strains isolated on alkenoic monoterpenes ((+)-menthene, alpha-pinene, 2-carene, and alpha-phellandrene) and nitrate. Syst Appl Microbiol. 21:237–244.

Furukawa H, Tsay JT, Jackowski S, Takamura Y, Rock CO. 1993. Thiolactomycin resistance in *Escherichia coli* is associated with the multidrug resistance efflux pump encoded by *emrAB*. J Bacteriol. 175:3723–3729.

Gamper M, Zimmermann A, Haas D. 1991. Anaerobic regulation of transcription initiation in the *arcDABC* operon of *Pseudomonas aeruginosa*. J Bacteriol. 173:4742–4750.

Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. Carbohydr Res. 356: 12–24.

Greenwald JW, Greenwald CJ, Philmus BJ, Begley TP, Gross DC. 2012. RNA-seq analysis reveals that an ECF sigma factor, AcsS, regulates achromobactin biosynthesis in *Pseudomonas syringae* pv. syringae B728a. PLoS One 7:e34804.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696–704.

Hansen PL, Pommer K, Malmgren B, Hansen OC, Poulsen M. 2006. Survey of chemical substances in consumer products (Denmark): Danish Environmental Protection Agency. p. 66.

Heipieper HJ, Neumann G, Cornelissen S, Meinhardt F. 2007. Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. Appl Microbiol Biotechnol. 74:961–973.

Hennig C. 2010. Methods for merging Gaussian mixture components. Adv Data Anal Classif. 4:3–34.

Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. 2008. *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 18:802–809.

Heyen U, Harder J. 2000. Geranic acid formation, an initial reaction of anaerobic monoterpene metabolism in denitrifying *Alcaligenes defragrans*. Appl Environ Microbiol. 66:3004–3009.
Hilker R, et al. 2014. ReadXplorer-visualization and analysis of mapped sequences. Bioinformatics 30:2247–2254.
Hocquemiller R, et al. 1991. Isolation and synthesis of espintanol, a new antiparasitic monoterpene. J Nat Prod. 54:445–452.
Hong H, Patel DR, Tamm LK, van den Berg B. 2006. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. J Biol Chem. 281:7568–7577.
Hu RM, Liao ST, Huang CC, Huang YW, Yang TC. 2012. An inducible fusaric acid tripartite efflux pump contributes to the fusaric acid resistance in *Stenotrophomonas maltophilia*. PLoS One 7:e51053.
Iurescia S, et al. 1999. Identification and sequencing of beta-β-myrcene catabolism genes from *Pseudomonas* sp. strain M1. Appl Environ Microbiol. 65:2871–2876.
Juhas M, et al. 2008. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 33:376–393.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.
Kauderer B, Zamith H, Paumgartten FJR, Speit G, Holden HE. 1991. Evaluation of the mutagenicity of beta-β-myrcene in mammalian cells *in vitro*. Environ Mol Mutagen. 18:28–34.
Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. Genome Biol. 11:R116.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Li H, et al. 2009. Subgroup 1000 Genome Project Data Processing. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.
Lüddeke F, Dikfidan A, Harder J. 2012. Physiology of deletion mutants in the anaerobic beta-β-myrcene degradation pathway in *Castellaniella defragrans*. BMC Microbiol. 12:192.
Lüddeke F, Harder J. 2011. Enantiospecific (S)-(+)-linalool formation from beta-β-myrcene by linalool dehydratase-isomerase. Z Naturforsch C. 66:409–412.
Lüddeke F, Wülfing A, et al. 2012. Geraniol and geranial dehydrogenases induced in anaerobic monoterpene degradation by *Castellaniella defragrans*. Appl Environ Microbiol. 78:2128–2136.
Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.
Magoc T, Wood D, Salzberg SL. 2013. EDGE-pro: estimated degree of gene expression in prokaryotic genomes. Evol Bioinformatics. 9: 127–136.
Manos J, et al. 2008. Transcriptome analyses and biofilm-forming characteristics of a clonal *Pseudomonas aeruginosa* from the cystic fibrosis lung. J Med Microbiol. 57:1454–1465.
Matsuura R, Ukeda H, Sawamura M. 2006. Tyrosinase inhibitory activity of citrus essential oils. J Agric Food Chem. 54:2309–2313.
Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182–W185.
Narushima H, Omori T, Minoda Y. 1982. Microbial transformation of alpha-pinene. Eur J Appl Microbiol Biotechnol. 16:174–178.
Neher TM, Lueking DR. 2009. *Pseudomonas fluorescens ompW*: plasmid localization and requirement for naphthalene uptake. Can J Microbiol. 55:553–563.
Nordgård O, Kvaløy JT, Farmen RK, Heikkilä R. 2006. Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models—the balance between accuracy and precision. Anal Biochem. 356:182–193.
Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29:e45.
Poblete-Castro I, Becker J, Dohnt K, dos Santos VM, Wittmann C. 2012. Industrial biotechnology of *Pseudomonas putida* and related species. Appl Microbiol Biotechnol. 93:2279–2290.
Ramos JL, et al. 2002. Mechanisms of solvent tolerance in gram-negative bacteria. Annu Rev Microbiol. 56:743–768.
Rasooli I, Mirmostafa SA. 2003. Bacterial susceptibility to and chemical composition of essential oils from *Thymus kotschyanus* and *Thymus persicus*. J Agric Food Chem. 51:2200–2205.
Rivera SL, Vargas E, Ramírez-Díaz MI, Campos-García J, Cervantes C. 2008. Genes related to chromate resistance by *Pseudomonas aeruginosa* PAO1. Antonie Van Leeuwenhoek 94:299–305.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
Rocchetta HL, Burrows LL, Lam JS. 1999. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. Microbiol Mol Biol Rev. 63: 523–553.
Rutledge RG, Côté C. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Res. 31: e93.
Santos PM, Mignogna G, Heipieper HJ, Zennaro E. 2002. Occurrence and properties of glutathione S-transferases in phenol-degrading *Pseudomonas* strains. Res Microbiol. 153:89–98.
Santos PM, Sá-Correia I. 2007. Characterization of the unique organization and co-regulation of a gene cluster required for phenol and benzene catabolism in *Pseudomonas* sp. M1. J Biotechnol. 131: 371–378.
Santos PM, Sá-Correia I. 2009. Adaptation to β-myrcene catabolism in *Pseudomonas* sp. M1: an expression proteomics analysis. Proteomics, 9: 5101–5111.
Schmid A, et al. 2001. Industrial biocatalysis today and tomorrow. Nature 409:258–268.
Seshasayee AS, Fraser GM, Babu MM, Luscombe NM. 2009. Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*. Genome Res. 19:79–91.
Sikkema J, de Bont JA, Poolman B. 1995. Mechanisms of membrane toxicity of hydrocarbons. Microbiol Rev. 59:201–222.
Soares-Castro P, Santos PM. 2013. Towards the description of the genome catalogue of *Pseudomonas* sp. strain M1. Genome Announc. 1: 11–12.
Tamber S, Ochs MM, Hancock RE. 2006. Role of the novel OprD family of porins in nutrient uptake in *Pseudomonas aeruginosa*. J Bacteriol. 188: 45–54.
Textor S, et al. 1997. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. Arch Microbiol. 168: 428–436.
Thompson ML, Marriott R, Dowle A, Grogan G. 2010. Biotransformation of beta-myrcene to geraniol by a strain of *Rhodococcus erythropolis* isolated by selective enrichment from hop plants. Appl Microbiol Biotechnol. 85:721–730.
Touw DS, Patel DR, van den Berg B. 2010. The crystal structure of OprG from *Pseudomonas aeruginosa*, a potential channel for transport of hydrophobic molecules across the outer membrane. PLoS One 5: e15016.
Tran AX, Trent MS, Whitfield C. 2008. The LptA protein of *Escherichia coli* is a periplasmic lipid A-binding protein involved in the lipopolysaccharide export pathway. J Biol Chem. 283: 20342–20349.
Turina AV, Nolan MV, Zygadlo JA, Perillo MA. 2006. Natural terpenes: self-assembly and membrane partitioning. Biophys Chem. 122: 101–113.
van der Werf J, de Bont JAM, Leak DJ. 1997. Opportunities in microbial biotransformation of monoterpenes. Adv Biochem Eng Biotechnol. 55:147–177.

Vicedomini R, Vezzi F, Scalabrin S, Arvestad L, Policriti A. 2013. GAM-NGS: genomic assemblies merger for next generation sequencing. BMC Bioinformatics 14(Suppl 7):S6.

Warnes GR, et al. 2009. gplots: various R programming tools for plotting data. [cited 2014 Oct 1]. Available from: http://cran.r-project.org/web/packages/gplots/index.html.

Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H. 2011. SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. Nucleic Acids Res. 39:e132.

Wubbolts MG, Timmis KN. 1990. Biotransformation of substituted benzoates to the corresponding cis-diols by an engineered strain of *Pseudomonas oleovorans* producing the TOL plasmid-specified enzyme toluate-1,2-dioxygenase. Appl Environ Microbiol. 56: 569–571.

Yoneyama H, Yamano Y, Nakae T. 1995. Role of porins in the antibiotic susceptibility of *Pseudomonas aeruginosa*: construction of mutants with deletions in the multiple porin genes. Biochem Biophys Res Commun. 213:88–95.

Zeyer J, Lehrbach PR, Timmis KN. 1985. Use of cloned genes of *Pseudomonas* TOL plasmid to effect biotransformation of benzoates to cis-dihydrodiols and catechols by *Escherichia coli* cells. Appl Environ Microbiol. 50:1409–1413.

Associate editor: José Pereira-Leal