Functional characterization of a 28-kb catabolic island from *Pseudomonas* sp. M1 involved in biotransformation of β-myrcene and related plant-derived volatiles

Pedro Soares-Castro¹, Pedro Montenegro-Silva¹, Hermann J. Heipieper² and Pedro M. Santos¹#

¹CBMA – Centre of Molecular and Environmental Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal; ²Helmholtz Centre for Environmental Research - UFZ, Department of Environmental Biotechnology, Leipzig, Germany

#Corresponding author: Pedro M. Santos, CBMA – Centre of Molecular and Environmental Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal, +351253601515, psantos@bio.uminho.pt

Running head: Characterizing a genomic island of *Pseudomonas* sp. M1

Key words: *Pseudomonas*, biotransformation, monoterpenes, genomic island, promoter-probes, β-myrcene hydroxylase
Abstract

*Pseudomonas* sp. M1 is able to mineralize highly hydrophobic and recalcitrant compounds such as benzene, phenol and their methylated/halogenated derivatives, as well as the backbone of several monoterpenes. The ability to use such spectrum of compounds as sole carbon source is, most probably, associated with a genetic background evolved under different environmental constraints. The outstanding performance of M1 strain regarding β-myrcene catabolism was elucidated in this work, with focus on the biocatalytical potential of β-myrcene-associated core-code, comprised in a 28-kb genomic island (GI), predicted to be organized in 8 transcriptional units.

Functional characterization of this *locus* with promoter-probes and analytical approaches validated the genetic organization predicted *in silico* and associated the β-myrcene-induced promoter activity to the production of β-myrcene derivatives. Notably, by using a whole-genome mutagenesis strategy, different genotypes of the 28-kb GI were generated, resulting in the identification of a novel putative β-myrcene hydroxylase, responsible for the initial oxidation of β-myrcene into myrcen-8-ol, and a sensor-like regulatory protein, whose inactivation abolished the *myr*+ trait of M1 cells.

Moreover, it was demonstrated that the range of monoterpene substrates of M1 enzymatic repertoire, besides β-myrcene, also includes other acyclic (e.g. β-linalool) and cyclic molecules (e.g. R-(+)-limonene and (-)-β-pinene). Our findings are the cornerstone for following metabolic engineering approaches and hint a major role of the 28-kb GI in the biotransformation of a broad monoterpene-backbone spectrum for its future biotechnological applications.
Importance

Information regarding microbial systems able to biotransformation monoterpenes, especially β-myrcene, is limited and focused mainly in non-systematic metabolite identification. Full and detailed knowledge at the genetic, protein, metabolite and regulatory level is essential in order to set a model organism or a catabolic system as biotech tool. Moreover, molecular characterization about reported systems is scarce, almost inexistent, limiting advances in the development of optimized cell-factories recurring to new generation of metabolic engineering platforms. This study provide new insights on the intricate molecular functionalities associated with β-myrcene catabolism in *Pseudomonas*, envisaging the production of molecular knowledge-base about the underlying catalytic and regulatory mechanisms of plant-derived volatile catabolic pathways.
Introduction

Plant monoterpenes and their derivatives are widely used in industry processes because of their organoleptic and therapeutic properties (1–4). Among them, β-myrcene (7-Methyl-3-methylene-1,6-octadiene) is one of the most promising monoterpenes, being mainly used as starting material for the chemical synthesis of value-added products with a broad spectrum of applications: pharmaceutics (e.g. antimutagens, analgesics, tyrosinase inhibitors), top selling flavors and fragrances (e.g., linalool, nerol, geraniol, citral, menthol) (4) and also in the production of polymers, biodegradable surfactants, pheromones and agrochemicals (4–6). The intrinsic physicochemical properties of monoterpane molecules (e.g. high hydrophobicity and structural instability) pose challenges regarding their utilization in industrial catalysis. Their catalytic modification to obtain oxidized/hydroxylated monoterpane derivatives usually requires the presence of a second phase of organic solvents or utilization of detergents to aid mass transfer in catalytic processes (5–7).

Strikingly, biological systems can overcome the limited availability of substrate and quickly adapt to the hydrophobic character of monoterpenes, with the advantage of yielding “natural” derivatives with high regio- and stereo-selectivity (6, 8). Moreover, the versatility of microorganisms allows easy and inexpensive maintenance of the biocatalyst, whose catalytic machinery can be improved and optimized by modern metabolic engineering platforms, allowing limitless possibilities of novel products and industrial applications (9, 10).

*Pseudomonas* sp. M1 is so far the best investigated aerobic bacteria able to mineralize β-myrcene (11–13), as well as to oxidatively transform β-citronellol and citral (14). By using a systems biology approach, strain M1 has been previously characterized at
the genome, transcriptome and proteome levels (12, 13). The sequencing of M1 strain genome and characterization of the myrcene-stimulon by RNA-seq identified a novel 28-kb genomic island (GI), whose expression was strongly stimulated in the presence of β-myrcene. This locus includes genes, apparently organized in eight transcriptional units (TUs), whose products are putatively involved in: i) substrate sensing; ii) gene expression regulation; iii) β-myrcene oxidation and bioconversion of β-myrcene derivatives into the central metabolism intermediates (12). Significantly, the majority of the functional modules present in the 28-kb β-myrcene-inducible catabolic island did not show high homology with sequences available in databases and seemed to have evolved through the assembly of several functional blocks acquired from different bacteria, probably at different evolutionary stages and driven by particular niche-related constrains (12). Moreover, the molecular characterization about reported microbial enzymatic systems involved in the catabolism of monoterpenes is scarce, almost inexistent, limiting advances in the development of tailored cell-factories. The work presented here is the first report of the functional characterization of the 28-kb GI associated with the catabolism of β-myrcene, by using molecular and analytical approaches during biotransformation experiments of Pseudomonas sp. M1 strains harboring different mutations within the genomic island. Most importantly, a novel putative β-myrcene hydroxylase and a sensor-like regulatory protein were identified and insights into the substrate range of the 28-kb GI were provided, thus, describing the biotechnological potential of M1 enzymatic machinery towards the biotransformation of a broader range of acyclic and cyclic plant-derived volatiles.
Materials and Methods

Bacterial strains, culture conditions, vectors and primers used

Bacterial strains, plasmids and primers used in this work are listed in Table 1 and Table 2, respectively. For molecular biology procedures, *Escherichia coli* and *Pseudomonas* *sp.* M1 strains were routinely maintained in Luria-Bertani (LB) medium, supplemented with antibiotics: gentamicin (30 μg mL⁻¹) or kanamycin (50 μg mL⁻¹). β-Myrcene, (±)-β-citronellol, linalyl acetate (acetate ester of linalool) and (-)-α-terpineol were purchase from Merck. R-(+)-Limonene, (-)-β-pinene, (-)-carveol and geraniol were purchased from Alfa Aesar. Analytical standards of fatty acid methyl esters for analysis of the membrane lipid content (C16:0, C16:1trans, C16:1cis, C18:0, C18:1trans, C18:1cis) were purchased from Sigma Aldrich.

Construction of promoter-probes to evaluate expression of different transcriptional units from the 28-kb β-myrcene genomic core

The molecular characterization of the transcriptional units from the 28-kb genomic island was performed by using a promoter-probe reporter system. The promoter-probes were constructed by coupling the eight promoter regions predicted *in silico* in the 28-kb GI of M1 strain by Soares-Castro and Santos (12), with the green fluorescence (GFP) reporter gene as transcriptional fusions. The promoter regions were amplified by PCR with primer sequences listed in Table 2, yielding DNA fragments flanked by *Bam*HI and *Hind*III restriction sites. PCR amplification was performed according to the following protocol: initial denaturation step at 96 °C for 3 minutes, 30 cycles of denaturation at 96 °C for 30 seconds, annealing at 50 °C for 30 seconds and extension at 72 °C for 45 seconds, followed by final extension step at 72 °C for 5 minutes. The amplification products were purified from agarose gel and cloned into the broad-host range vector pSEVA637 (16).
Whole-genome mutagenesis targeting inactivation of **PM1_0216320-PM1_0216325**
sensory cluster (TU2)

The mutagenesis strategy to identify putative activators of the sensory cluster **PM1_0216320-PM1_0216325**, comprising the predicted transcriptional unit 2 (TU2) of the 28-kb GI, was performed with the pBAM1 vector (17) harboring the P2::GFP reporter fusion.

The reporter transposon was constructed by cloning the promoter P2 fused to GFP from pSEVA637-P2, flanked by BamHI/EcoRI (Table 2), into pBAM1, which resulted in the vector pBAM1-P2::GFP. The pBAM1-P2::GFP was mobilized by standard biparental mating to M1 strain and colonies were screened for the loss of P2::GFP fluorescence by microscopy. From this mutagenesis approach, two mutant strains were isolated (M1-C19 and M1-C38).

Transposon insertion sites were amplified by arbitrary PCR, adapted from Das *et al.* (18).

Initially, the insertion site was amplified from the ME-I transposon end with the primers ARB6/ME-I-extR and from the ME-O transposon end with primers ARB6/gfp-extF. The second PCR was performed with primers ARB2/ME-I-intR targeting the ME-I end and primers ARB2/gfp-intF targeting the ME-O end. The ARB6-related PCR was performed by an initial denaturation step at 96 °C for 3 minutes, 6 cycles of denaturation at 96 °C for 30 seconds, annealing at 30 °C for 30 seconds and extension at 72 °C for 90 seconds, followed by 30 cycles of denaturation at 96 °C for 30 seconds, annealing at 45 °C for 30 seconds and extension at 72 °C for 90 seconds, finalizing with extension step at 72 °C for 5 minutes. The ARB2-related PCR was performed by an initial denaturation step at 96 °C for 3 minutes, 30 cycles of denaturation at 96 °C for 30 seconds, annealing at 52 °C for 30 seconds and extension at 72 °C for 90 seconds,
followed by final extension step at 72 °C for 5 minutes. PCR products were purified and sequenced by ABI 3730 XLs sequencer (Beckman Coulter Genomics).

**Monitoring the activity of promoter reporters from the 28-kb GI in M1 strains**

The pSEVA637 promoter-probe constructs were mobilized into M1 strains (M1 wt, M1-N22, M1-C19 and M1-C38) by standard biparental mating on membranes, using *E. coli* S17.1λpir as donor. Myrcene-dependent activity levels of the assayed promoter regions were monitored by i) fluorometric detection in Qubit® 3.0 Fluorometer and expressed as fold induction of GFP fluorescence normalized by culture optical density measured at 600 nm (OD$_{600nm}$), between myrcene-induced and non-induced cells, ii) and/or confirmed by microscopy in a fluorescence microscope Leica DM5000B-CTR5000 (total amplification of 400× used).

To validate the predicted promoter regions, overnight cultures of M1 wt strain were refreshed in MM (19) supplemented with 48 mM lactate or 100 μL of β-myrcene (130 mg L$^{-1}$, 47.8 μmol in solution) and cultures diluted again at mid-exponential phase. Similarly to the RNA-seq experiment in Soares-Castro and Santos (12), when cultures reached an OD$_{600nm}$ value around 0.3-0.4, cells were collected to monitor the GFP reporter expression by microscopy, in both carbon sources.

To monitor the reporter expression in the M1 mutant strains (M1-N22, M1-C19, M1-C38), cells were grown as described above in MM supplemented with 48 mM lactate. The promoter regions were induced by challenging M1 wt and mutant strains, harboring the GFP-based promoter reporters, with saturating 20 μL of β-myrcene (26 mg L$^{-1}$, 9.6 μmol in solution) and measured 5 hours after medium supplementation with the inducer. The β-myrcene-dependent activity levels were expressed as fold induction as described above and
plotted as heatmap by using the function heatmap.2() from R statistical software with the "gplots" package (20).

**Phenotype complementation of M1-C38 and M1-C19 strains**

To complement the transposon interruption of PM1_0216375 and PM1_0216370 in M1-C19 and M1-C38 strains, respectively, the promoterless PM1_0216375 and TU5 (PM1_0216365-PM1_0216370) segments were amplified by PCR (Table 2). The PM1_0216375 and TU5 products were digested with BamHI/ SalI and EcoRI/ SalI, respectively, and cloned into pSEVA638, which harbors the expression system xylS-Pm (16).

The resulting vectors (pSEVA638-PM1_0216375 and pSEVA638-TU5) were mobilized into M1-C19 and M1-C38 strains, respectively, by standard biparental mating as described above. Phenotype complementation was evaluated by assessing the ability of M1-C19 (pSEVA638-PM1_0216375) and M1-C38 (pSEVA638-TU5) strains to grow in MM supplemented with β-myrcene as sole carbon source, in the presence of the inducer 3-methylbenzoate.

**Evaluation of the monoterpene substrate spectrum of the 28-kb GI**

In order to identify the potential substrates of the 28-kb β-myrcene core-code, M1 wt cells harboring the transcriptional fusion of promoters expressing the GFP reporter were cultured in MM supplemented with 48 mM lactate, as described above. Culture medium was supplemented with monoterpenes structurally-related to β-myrcene and the promoter activity was measured at 5 hours after stimuli. Molarity of the different monoterpenes was calculated as described in Supplemental Material (section of Table S1) (21, 22), taking into account the different physicochemical properties (e.g. solubility in aqueous medium and vapor pressure) and their influence in the bioavailability of each monoterpene in the assay. Saturating amounts of β-myrcene were used as reference (20 μL corresponding to 9.6 μmol in solution) and the promoter activity levels were expressed as fold induction of normalized GFP fluorescence,
described previously. Statistical significance of the differences in fold change was analyzed by two-way ANOVA test, comparing to the fold change of the lactate control.

Qualitative GC-MS analysis of terpene-derivatives from cultures of *Pseudomonas sp.*

**M1 and derived strains, supplemented with plant-derived volatiles**

*Pseudomonas* sp. M1 and derived strains were cultured as described for the evaluation of the reporter activity, in 50 mL of MM using lactate as carbon source and supplemented with 5 μL of the monoterpenic compound for following identification of the respective oxidative products. Cultures of the wild-type strain of *Pseudomonas* sp. M1 were supplemented with β-myrcene, (±)-β-citronellol, linalyl acetate, R-(+)-limonene and (-)-β-pinene, whereas biotransformation experiments with the M1-derived strains were performed only with β-myrcene.

To extract terpenes-derivatives from biotransformation experiments, cultures were prepared for each time-point analyzed, in which 50 mL of the supernatant (the whole culture) was collected, mixed with 10 mL ethyl acetate and vigorously shaken for 5 min. The ethyl acetate extracts were dried with nitrogen gas stream, dissolved in 1 mL of hexane and concentrated with a nitrogen gas stream to a final volume around 500 μL, for subsequent gas chromatography analysis coupled to mass spectrometry. Control metabolite profiles were prepared by using i) pure terpene standards extracted in hexane, ii) culture supernatants from cells grown in lactate as sole carbon source without supplementation with terpenes, at different time-points, to rule out the metabolite profile of different growth stages, and iii) abiotic control of MM with lactate and terpene supplementation.

The gas chromatography analysis of the supernatants from biotransformation experiments was performed on a 7890A GC-5975C inert XL MSD system, coupled to a 7693 Automatic Liquid Sampler (Agilent Technologies) and equipped with an Agilent J&W HP-INNOWAX.
polar capillary column (polyethylene glycol; 30 m x 0.25 mm I.D. x 0.25 μm film thickness). Helium was used as carrier gas at a flow rate of 1.6 mL min⁻¹ and 1 μL of sample was injected in splitless mode. The column temperature was initially kept at 60 °C for 1 min and then gradually increased to 250 °C at a rate of 10 °C min⁻¹, with a final hold time of 2 min (adapted from 23, 24). The mass spectra were acquired with an electron impact ionization voltage of 70 eV, scanned in the m/z range of 40–500 atomic mass unit. Cultures of β-myrcene biotransformation with M1 strains were assayed in quadruplicates (wt strain) and triplicates (mutant strains), up to 3.5 hours after addition of the monoterpenes. Biotransformations of the other monoterpens by the wt strain were assayed at two time-points (4 hours and 20 hours). The baseline of the obtained chromatograms was defined with the statistics-sensitive non-linear iterative peak-clipping algorithm (25, 26), which showed to be the best fit for the shape of different peak-free regions throughout the chromatogram.

**Metabolite identification**

The total ion chromatograms were deconvoluted with the AMDIS tool version 2.71 (27) to generate the ion spectra for every integer m/z component in the data acquisition range of atomic mass units. The deconvoluted m/z spectra whose intensity values increase, maximize and decrease together, are derived from the same chemical species, thus comprise its ion profile used for compound identification. The filtered ion profile of each chemical species (here defined as the query peak) was obtained by subtracting the intensity of the m/z components from the matrix (region of the GC-MS spectrum without peaks), and/or from adjacent peaks on either side of the query peak. The ion profile of overlapped peaks, corresponding to co-eluted compounds, was manually filtered by subtracting the intensity of the m/z components of the contaminant peak from the ion profile of the query peak.
The identification of terpene-derivatives was performed based on i) the NIST MS Search tool version 2.0 libraries (mainlib comprising 163198 spectra and wiley comprising 228996 spectra), ii) by comparing calculated retention indices (RI) with those reported in the literature and/or from the National Institute of Standards and Technology (NIST) Mass Spec Data Center (28) and iii) validation with commercial compounds when available. Retention indices of each query compound were determined by sampling a mixture of n-alkanes (C10–C28) under the same GC-MS conditions (29).

**Membrane fatty acid profile and viability of cells challenged with β-myrcene**

Cultures of *Pseudomonas* sp. M1 strains, prepared as described for the evaluation of the expression of promoter-probes, were grown in MM supplemented with lactate and exposed to 20 μL of β-myrcene. Cells were collected after 0.5 hour, 1.5 hour and 3 hours of monoterpene addition for lipid extraction. Cell viability was evaluated in parallel after 3 hours of the 20 μL β-myrcene pulse, by diluting and plating cells in LB agar plates for colony forming units (CFU) counting. Cultures without β-myrcene supplementation were used as control. The specific growth rate for each strain growing in lactate or lactate with β-myrcene pulse, during the assayed time-points, was calculated (μ) and expressed as ratio of \( \frac{\mu_{\text{myr}}}{\mu_{\text{lac}}} \) to correlate changes in cell envelope with cell viability and growth rate during β-myrcene biotransformation.

For the analysis of the fatty acid profile, the collected cell pellets were washed with 10 mM potassium nitrate and the fatty acid methyl esters (FAME) were prepared according to Morrison and Smith (30, 31). Lipids were extracted from cells with methanol and chloroform. The lipid-containing chloroform phase was dried in nitrogen gas stream, the transesterification of fatty acids performed with boron trifluoride/methanol for 15 min at 96
°C and FAME extracted in hexane. The FAME were analyzed in a Agilent 6890N GC system with a flame ionization detector and equipped with a capillary column CP-SIL 88, 50 m (Agilent Technologies). Authentic reference compounds were coinjected with samples. Fatty acid methyl esters were separated at an initial temperature of 40 °C by 2 minutes, followed by an increase rate of 8 °C min⁻¹ up to 220 °C, with a final hold of 10 minutes. The peak areas were used to determine the relative amounts of each fatty acid. The degree of saturation of membrane fatty acids was defined as the ratio of saturated C16:0 and C18:0 to unsaturated C16:1 and C18:1 fatty acids. The trans/cis ratio of unsaturated fatty acids was defined as the ratio between the amounts of the two trans unsaturated fatty acids (16:1trans, 18:1trans) and the two cis unsaturated fatty acids (16:1cis, 18:1cis) of Pseudomonas strains (31).

In silico analysis of protein sequences from PM1_0216375 and PM1_0216370

The protein sequence of PM1_0216375 was compared with the sequence of eight major families of transcriptional regulators involved in the catabolism of acyclic and cyclic/aromatic hydrocarbons, described by Tropel and van der Meer (32). Sequences were retrieved from Uniprot, aligned with MAFFT L-INS-i method (33) and a similarity tree generated with PhyML version 2.4 (34), with 100 bootstrap sets, the best-fit model of a.a. substitution for maximal likelihood predicted by Prottest version 3 (35), kappa estimated, gamma distribution parameter estimated, BIONJ starting tree, with optimization of topology, branch lengths and rate parameter. The functional domains of PM1_0216375 sequence were predicted by Interpro (36) and used for a detailed analysis of the identity of amino acid residues with the respective homologous regions in other characterized MalT-like regulators: MalT from Escherichia coli (Uniprot accession: P06993), AcoK from Klebsiella pneumoniae (Uniprot accession: Q48411), AlkS from P. putida GPo1, former P. oleovorans (Uniprot accession: P17051), AlkS from Pseudomonas putida P1 (Uniprot accession: Q9L4M7).
The protein sequence of PM1_0216370 was compared with the sequence of membrane non-heme di-iron proteins by homology search with blastp (37) and prediction of functional domains with Interpro (36). For a detailed analysis, transmembrane domains in PM1_0216370 protein sequence were predicted by a Kyte–Doolittle hydropathy plot, obtained from the Protoscale tool of the Expasy server (38, 39), using a window size of 19 amino acids. For the homology-based clustering analysis, the protein sequences from PM1_0216370 and reported membrane non-heme hydroxylases, monoxygenases and fatty acid desaturases were retrieved from Uniprot and aligned with MAFFT E-INS-i method. The region comprising the conserved histidine-rich motifs HX1,4H, HX2,3HH and HXXHH, described by Shanklin et al. (40), were used to generate the similarity tree as described above. The accession numbers of the protein sequences used are the following: Q8YQ27, Nostoc sp. PCC7120; AAP99312, Prochlorococcus marinus CCMP1375; A3WE62, Erythrobacter sp. NAP1; O85832, Novosphingobium aromaticivorans; P94792, Flavobacterium sp. ATCC21588; Q44262, Paracoccus sp. PC1; Q01332, Escherichia vulneris; P21688, Pantoea ananas; Q88HV7, P. putida KT2440; A6N7G2, Psychrobacter uratvorans; Q12SQ0, Shewanella denitrificans OS217; P12691, P. oleovorans (current P. putida GPo1); O31250, Acinetobacter bayly ADP1; Q9AEN3, Burkholderia cepacia RR10; E5G6V9, Rhodococcus aetherivorans BCP1; P21395, P. putida KT2440; O33456, P. putida F1.

Results

_Pseudomonas sp._ M1 is able to mineralize β-myrcene with the catabolic genetic core-code being located on a 28-kb GI (12). The biotechnological exploitation of its functional modules requires detailed understanding of which genes are directly involved in β-myrcene catabolism, their interactions and regulation. As described by Soares-Castro and
Santos (12), the β-myrcene-inducible 28-kb GI was predicted in silico to be organized in eight transcriptional units (TU), whose putative coded products are summarized in Table 3. The TU1, TU3, TU5, TU6 and TU7 are predicted to mainly comprise catabolic genes, TU4 and TU8 are predicted to code regulatory elements, and the TU2 is predicted to code a chemosensory system.

**Reporter systems validate the promoter regions controlling β-myrcene-dependent operons from the 28-kb genomic island**

The promoter regions from each TU, listed in Table 3, were cloned into the promoter-probe vector pSEVA637 envisaging the validation of the previously predicted promoter segments (12), therefore, detailing the operonic organization of the genomic island (Figure 1).

As depicted in Figure 1, promoter regions P1, P2, P3, P5, P6 and P7 showed a β-myrcene-dependent production of GFP in M1 cells, validating these DNA segments as functional elements and the overall modular organization of the island. The promoter fusions of P4 and P8 presented low inducible activity (< 1.5-fold increase) under the tested conditions and, thus, were excluded from the following analysis.

To better understand the role and functional interplay between the enzymatic systems coded by the 28-kb GI of M1 strain and obtain insights regarding underlying regulatory mechanism, the expression of the putative chemosensory genetic cluster TU2 was inactivated with a whole-genome mutagenesis strategy, by using the random transposon pBAM1 harboring the reporter fusion P2::GFP to targeted its potential activators. Both genomic and transposon-born promoters P2 would be activated and the fluorescent reporter expressed if the transposon harboring P2::GFP was inserted into a neutral locus. Alternatively, if a transposition event interrupted a locus essential for the...
activation of promoter P2 (and TU2), the reporter would not be induced, resulting in non-
fluorescent cells. The mutagenesis approaches created different M1-derived genotypes,
targeting genes within the 28-kb GI: the M1-C19 strain, showing the insertion of the P2::GFP
reporter transposon cassette in \textit{PM1\_0216375}; the M1-C38 strain showing the insertion of the
P2::GFP reporter transposon cassette in \textit{PM1\_0216370}.

Mutant strains show different growth kinetics when β-myrcene is supplemented as
sole carbon source

In this work, three mutants of M1 strain showing phenotypic alterations regarding
β-myrcene catabolism were characterized: M1-N22 (11), M1-C19 and M1-C38. M1 \textit{wt} and
derived strains were grown in minimal medium supplemented with β-myrcene as sole
carbon source to assess the effect of the different 28-kb GI genotypes in the catabolic
pathway of β-myrcene (Figure 1).

In the \textit{wt} strain, β-myrcene supports the increase of the culture optical density and is
accompanied by a decrease in culture pH, derived from the oxidation of β-myrcene in
acidic derivatives and channeling of intermediates to the central metabolism (13). Thus, the
decrease of pH observed in the initial time-points for M1-N22 suggested a partial
catabolism of the β-myrcene backbone in this mutant, despite the residual increase in the
optical density values. The M1-N22 strain harbors an insertion of mini-\textit{Tn5-Km} transposon
in \textit{myrB} gene (\textit{PM1\_0216390}), which codes an alcohol dehydrogenase. Iurescia \textit{et al.} (11)
isolated the N22 mutant as a β-myrcene-negative strain, in which a myrcene alcohol,
myrcen-8-ol, was accumulated, instead of proceeding to mineralization. The initial β-
myrcene-supported growth of M1-N22 cells was significantly compromised, in comparison
with M1 \textit{wt} strain, due to the disruption of the predicted sequential oxidation of the β-
myrcene molecule (β-myrcene >> myrcene alcohol >> myrcene aldehyde >> myrcene}
 Nonetheless, in this work, N22 strain was able to resume growth in liquid cultures supplemented with β-myrcene as sole carbon-source and reached optical densities similar to the *wt* strain, within 48 hours of growth. In contrast, the M1-C19 and M1-C38 mutants were not able to grow in β-myrcene as sole carbon source within the sampled time (growth was extended to five days without increase in the optical density or decrease in pH values).

**The profile of promoter activity in M1-derived strains evidences different β-myrcene-induced enzymatic repertoires**

Further molecular analysis was carried out to describe the effect of mutations in the overall expression of the 28-kb GI and link specific activity of its functional modules to key enzymatic steps of β-myrcene catabolic pathway.

To understand the effect of the three mutations in the activity of the different transcriptional units within the 28-kb island, during β-myrcene induction, the pSEVA637 probes harboring the validated promoters were mobilized to M1-N22, M1-C19 and M1-C38 cells. The fluorescence of the *wt* and mutant strains was measured after five hours of β-myrcene pulse (Figure S1) in lactate-growing cultures.

Although M1-C19 and M1-C38 strains were not able to use β-myrcene as sole carbon source in the sampled time-points from growth kinetics experiments, promoter activity profiles within the GI were slightly different from each other. The disruption of the putative LuxR transcriptional regulator PM1_0216370 in M1-C19 switched-off all the promoters assayed (hence supporting the inability to mineralize β-myrcene, as the ratio of fluorescence levels in the presence and absence of β-myrcene was approximately 1. Therefore, *PM1_0216370* was shown to code for a primary activator of the β-myrcene-induced catabolic machinery. The disruption of the putative PM1_0216370 in M1-C38 only
resulted in the inactivation of promoters P2, P3 and P6, whereas promoters P1, P5 and P7 remained active, but downregulated approximately 2-fold, 3-fold and 2-fold, respectively.

The loss of the myr+ phenotype associated to the inactivation of the GI promoter activity in M1-C38 strongly indicated that PM1_0216365-PM1_0216370 (TU5) is essential for β-myrcene catabolism. In agreement, complementation of M1-C19 and M1-C38 strains with PM10216375 and TU5 sequences, respectively, under the control of the inducible xylS-Pm system, restored the ability of both mutants to grow in MM supplemented with β-myrcene, in the presence of the inducer 3-methylbenzoate (data not shown).

In M1-N22 strain, interruption of myrcene alcohol oxidation into myrcene aldehyde led to an overall reduction of the reporter levels. The activity of promoters P1, P2, P5, P6 and P7 was reduced approximately 3-fold, 6-fold, 6-fold, 5-fold and 2-fold, respectively, whereas the promoter P3 remained inactive.

**Biotransformation of β-myrcene correlates with the expression of promoter reporters**

To validate the role of the functional modules from the 28-kb GI, the metabolites from β-myrcene biotransformation experiments were identified and the cell membrane response to the hydrophobicity of β-myrcene characterized in the different M1 strains.

Metabolites of β-myrcene biotransformation experiments with M1 wt and derived strains were identified by gas chromatography coupled to mass spectrometry. M1 cells were grown in mineral medium supplemented with 48 mM of lactate, to which β-myrcene was added at medium-exponential phase of growth. The supernatants were collected at early- and late-stage of the monoterpene catabolism (30 min and 3.5 hours) and the β-myrcene-derivatives from culture supernatants were identified by gas chromatography coupled to mass spectrometry (Table 4, detailed in Table S2 and Figure S2).
The Table 4 lists the metabolites identified from β-myrcene biotransformation in M1 wt strain and compares it with the profile of mutant strains. The relative quantification was calculated based on their peak area, normalized as the percentage of the total spectral area of the chromatogram (41, 42), which allows to infer differences in metabolite abundances between peaks of the same chromatogram and, to some extent, between peaks of different samples (43).

In all wt cultures assayed after 30 minutes of β-myrcene pulse, the major metabolite detected (compound 2, retention index of 1920) was identified as myrcen-8-ol. Although this compound was not available in databases used for data mining, the obtained mass spectrum matched the mass spectrum of the detected alcohol and synthetic myrcen-8-ol standard published previously (11), described as substrate for the alcohol dehydrogenase MyrB. As evidenced by searching the mass spectrum of myrcen-8-ol, metabolite databases do not comprise a large number of myrcene-derived compounds for MS search, reflecting the lack of knowledge available regarding β-myrcene catabolic pathways, which in turn reduced the rate of metabolite identification.

A putative C10 aldehyde (compound 1, retention index of 1708) and carboxylic acid (compound 5, retention index of 2466) were also detected 30 minutes after addition, which most likely resulted from the oxidation of the terminal hydroxyl group of myrcen-8-ol, mediated by the alcohol dehydrogenase MyrB and the aldehyde dehydrogenase MyrA. A unequivocal identification of compound 1 and 5 was also not possible using available databases. Nevertheless, the comparison of the mass spectrum of compound 5 to the mass spectrum of the 2-methyl-6-methylen-2,7-octadienoic acid methyl ester, reported by Narushima et al. (44) from supernatants of Pseudomonas sp. S4-2 strain, supported its identification as 2-methyl-6-methylen-2,7-octadienoic acid (myrcenoic acid). Moreover, the
detection of the mass peaks 53, 67 and 93 in the m/z spectra of compounds 1 and 5, easily inferred by predicting the fragmentation pattern of the β-myrcene-like backbone and common to myrcene-8-ol, supported the identification of both derivatives as the myrcene aldehyde and myrcenoic acid, respectively. Although not present in all assayed supernatants, two putative derivatives of the predicted myrcenoic acid were detected in M1 wt supernatants at 3.5 hours of biotransformation: the 4-methyl-3-hexenoic acid (compound 4, retention index of 2045) and the 4-methylhexanoic acid (compound 3, retention index of 1925). The former was suggested by Narushima et al. (44) as a product of a β-oxidation-like catabolic step targeting the myrcenoic acid, with subsequent elimination of the carboxyl-containing C3-unit as propionyl-CoA, whereas the latter might result from the dehydrogenation of the C3,C4 double bond of 4-methyl-3-hexenoic acid.

Similarly to what was report by Iurescia et al. (11) the disruption of myrB in M1-N22 strain led to the early accumulation of the identified myrcen-8-ol, with concomitant loss of the ability to produce the putative myrcene aldehyde (compound 1) or the 4-methyl-3-hexenoic acid (compound 4), not detected in supernatants of this mutant. Both C19 and C38 strains did not produce any of the metabolites proposed for the catabolic pathway.

The different GI genotypes were also translated into different physiological traits at the membrane level. Since hydrophobic molecules are expected to be membrane active compounds (45, 46), changes in the membrane lipid profile and cell viability were analyzed when M1 and derived strains were exposed to the saturating amounts of β-myrcene. Cells were sampled and FAMEs were analyzed by gas chromatography (Figure 3) in co-metabolism experiments, by using lactate as primary carbon-source. Control cells grown solely with lactate were collected at the same time-points of FAME analysis, before β-myrcene supplementation and corresponding to 3 hours after the monoterpenes exposure.
The major changes observed in all M1 strains after β-myrcene supplementation were associated with the cis-to-trans isomerization of C16:1 and C18:1 unsaturated fatty acids (Figure 3A). Whereas the trans/cis ratio in both controls was approximately 0.02, these values increased up to 8-fold higher after 30 minutes of β-myrcene pulse in the assayed strains (from a ratio of 0.06 in M1-C19 and M1-C38 to a ratio > 0.1 in M1 wt and M1-N22 strains). This change in the content of membrane lipids coincided with the production of myrcen-8-ol in the wt strain (and accumulation in M1-N22) as depicted in Table 4. In contrast, the myr- M1-C19 and M1-C38 strains showed the lowest increase of the trans/cis ratios, most probably caused by the presence of β-myrcene alone, since the extensive downregulation of the 28-kb GI promoters/TUs impaired the successful biotransformation and utilization of β-myrcene to support cell growth.

M1 cells were capable to rapidly adapt to the hydrophobic properties of β-myrcene and to the over-accumulation of monoterpenoid derivatives as a proficient biocatalyst, in which significant inhibitory effects were not detected in cell viability (Figure 3B).

The LuxR-family transcriptional regulator PM1_0216375 is a key activator of the enzymatic machinery coded by the 28-kb GI

As indicated by previous results, PM1_0216375 is an essential activator of the β-myrcene inducible 28-kb GI. Moreover, the role of the two other putative transcriptional regulators, PM1_0216360 and PM1_0216405 coded in this locus remained undetermined. As depicted in the Figure 4, both PM1_0216360 and PM1_0216375 belong to the LuxR family, included in the MalT-subfamily, albeit the percentage of amino acid identity with the closest homologs, from the NCBI non-redundant protein database, is only 46 % (from Burkholderia sp. MSh2; Genbank accession: WP_063934625.1) and 28 % (from...
Polycyclovorans algicola; Genbank accession: WP_029891500.1), respectively. The PM1_0216405 is a member of the LysR family (Figure 4) with 83% of amino acids shared with a putative regulator from *P. aeruginosa* (Genbank: WP_023104498.1).

Similarly to the MalT-like regulators, PM1_0216375 is a large protein (872 a.a.), harboring 3 main functional domains (Figure 4). The protein domains depicted in Figure 4 are well described in MalT from *Escherichia coli* K-12 and confer a one-component switch-like function to the protein. The ATPase domain comprises a conserved nucleotide oligomerization domain, which alternates between the resting (ADP-bound) and the active (ATP-bound) conformation, being required for the oligomerization of the regulator into a protein complex. The sensor domain is responsible for the recognition of the inducer metabolite in MalT (47). The interaction with the inducer and ATP results in conformational changes in the MalT regulator (48), which allow the effector domain (winged helix-turn-helix – WHT-H) to bind the recognition site and activate transcription (48). Although not so extensively characterized, the MalT-like AcoK regulator is hypothesized to use a slightly different mechanism of regulation. Contrary to MalT, in AcoK mutants the ATPase domain appeared to not be essential for efficient recognition of the DNA-binding site. AcoK was hypothesized to bind the regulated DNA prior to interacting with the ATP or inducer, although both molecules were still essential for a conformational switch between an inactive and active form (49).

The a.a. identity between the functional domains of PM1_0216375 and other characterized MalT-like regulators is shown in Table S3 from Supplemental Material, whose respective sequence alignments are shown in Figures S3, S4 and S5. Despite the low identity levels observed between the predicted domains of PM1_0216375 and the MalT, AcoK and AlkS regulators, the conservation of key motifs, such as the Walker A and B motifs involved...
in the ATPase activity, suggests a similar mechanism of transcriptional activation in the presence of β-myrcene.

PM1_0216370 is more closely related to hydroxylases than fatty acid desaturases and probably catalyzes the terminal hydroxylation of the β-myrcene backbone.

The predicted TU5 codes for an oxidoreductase system, composed by 2 genes annotated as a putative fatty acid desaturase (PM1_0216370) and a rubredoxin (PM1_0216365). The loss of the ability to use β-myrcene as sole carbon source in M1-C38 strain (Figure 2 and Figure S1) indicated that the TU5, namely the PM1_0216370 gene, plays an essential role in the catabolic pathway. The PM1_0216370 protein shared only 52% of a.a. identity with its closest homolog from the NCBI non-redundant protein database (fatty acid desaturase from Novosphingobium sp. AAP93, Genbank accession; WP_054123145.1). Moreover, a detailed inspection of the protein sequence of PM1_0216370 predicted that the enzyme might fold into several transmembrane helix domains (5 to 6 domains predicted in silico by several tools in Figure S6) and contains three histidine-rich motifs, highly conserved in proteins of the membrane non-heme di-iron family: HX₃₋₄H, HX₂₋₃HH and HXXHH. The conservation of the histidine-rich motif is shown in Figure 5A (protein alignment in Figure S7), which are described to be required for iron binding and for the oxidative enzyme activity (50). The non-heme di-iron family of proteins comprises not only membrane fatty acid desaturases but also membrane hydroxylases/monooxygenases (e.g. alkane 1-­monooxygenase AlkB, xylene monooxygenase XylM, p-cymene monooxygenase CymAa) (50), as shown in Figure 5A.

Furthermore, both the protein alignment (Figure 5A and Figure S7) and the homology-based clustering in Figure 5B showed that PM1_0216370 shared higher sequence homology with the β-carotene hydroxylases (light blue in Figure 5B), whose
amino acid identity considering their complete protein sequences ranged from 41% to 22%
. The alignment of PM1_0216370 with protein sequences of the other tree clusters
registered identity values below 15%. Therefore, PM1_0216370 might code for the β-
myrcene monooxygenase, not identified in previous studies, which, putatively, catalyzes
the first enzymatic step of the β-myrcene catabolic pathway and conversion to myrcen-8-ol.

**M1 cells are able to biotransform different cyclic and acyclic monoterpenes**

Previous reports regarding biotransformation of plant-derived volatiles by
*Pseudomonas* sp. M1 have described solely the utilization of β-myrcene, β-citronellol and
citral as carbon sources (13, 14). To further evaluate the biotechnological potential of M1
enzymatic repertoire to other plant volatiles, biotransformation experiments were
performed with *wt* cells by supplementing lactate-growing cultures with other acyclic
(linalyl acetate) and cyclic (R-(+)-limonene and (-)-β-pinene) monoterpenes. After
monoterpene supplementation, the metabolite profile of culture supernatants was identified
by GC-MS (Table 5, detailed in Table S4 and Figure S8).

Supplementation with linalyl acetate resulted in the identification of linalool
(retention index of 1552), geraniol (retention index of 1852) and geranic acid (retention
index of 2342) by mass spectra and retention indices correspondence. M1 strain was also
able to hydroxylate the cyclic structure of R-(+)-limonene into limonene-1,2-diol (retention
index of 2287) and (-)-β-pinene biotransformation experiments resulted in the identification
of α-terpineol (retention index of 1703) and citronellic acid (retention index of 2249) in
culture supernatants.
In order to determine the potential involvement of the 28-kb GI from M1 strain in the enzymatic oxidation of other monoterpene backbones, the reporter activity of the GFP-based promoter-probes was evaluated in the presence of different plant-derived volatiles.

**The 28-kb GI promoters are responsive to other plant-derived volatiles**

Besides β-myrcene, the acyclic linalyl acetate, (±)-β-citronellol and geraniol and the cyclic R-(+)-limonene, (-)-carveol, (-)-α-terpineol and (-)-β-pinene were used as potential inducers. The promoter activity is shown in Figure 6, measured as fold induction between the GFP levels detected in lactate-grown cells challenged with a monoterpene and in cells grown with lactate as sole carbon source (control cultures).

Under the tested conditions, all promoter regions assayed showed higher induction levels in the presence of β-myrcene, confirming this substrate as the more effective enhancer of the 28-kb genomic island (Figure 6A).

The 28-kb GI was responsive to other monoterpene backbones, albeit an induction pattern related to the structure of the molecule was not apparent (Figure 6B). From the acyclic monoterpenes tested, geraniol activated promoters P6 by 4-fold, P2 by approximately 3-fold, P1 and P7 by 2-fold. Supplementation with linalyl acetate increased the GFP levels of P2 reporter by 3-fold and of P6 and P7 by approximately 2-fold. (±)-β-Citronellol, a substrate for the acyclic terpene utilization pathway, was only able to induce promoter P2, by 2.4-fold (data not shown). The monocyclic R-(+)-limonene induced promoter P2 by approximately 4-fold and P6 by 3-fold, while the cyclic alcohol (-)-carveol only induced P2, by 2.4-fold (data not shown). The addition of (-)-α-terpineol did not result in substantial fluorescence levels of the reporter (below 1.5-fold) for any of the promoter regions tested. The bicyclic (-)-β-pinene mainly induced the activity of promoter P2, by 3-fold, as well as promoters P1 and P3 by approximately 2-fold.
Discussion

This report describes the first functional characterization of a 28-kb GI associated with β-myrcene metabolism in *Pseudomonas* sp. M1. By monitoring the β-myrcene-induced expression of promoter reporter systems in different GI genotypes, the overall genetic organization of the 28-kb island was validated and the promoter activity correlated with the biotransformation of β-myrcene, according to the predicted pathway (11, 13). In particular, the comparison performed between M1 *wt* and derived strains, based on promoter reporter activity and GC-MS identification of β-myrcene derivatives from biotransformation experiments, provided evidences about the involvement of genes coded by TU5 and TU6 in the initial oxidation steps of β-myrcene and their biotechnological relevance as targets for engineering approaches and development of β-myrcene catalysts. Moreover, the molecular and physiological approaches described in this work allowed the identification of two key gene products, essential to activate the β-myrcene core-code in the 28-kb GI and confer the ability to use this monoterpane as sole carbon source in *Pseudomonas* sp. M1 cells: the novel putative β-myrcene hydroxylase (PM1_0216370), involved in the initial oxidation of β-myrcene backbone (mainly based on experiments with M1-C38 strain), and the one-component switch-like regulator (PM1_0216375), capable of readily sensing β-myrcene-derived stimulus and of activating the 28-kb GI (mainly based on experiments with M1-C19 strain). Considering the homology of PM1_0216375 protein sequence with regulators from the MalT subfamily, this protein may alternate between a stable inactive monomeric conformation (which might be constitutively expressed at low concentration) and an oligomeric active conformation when the inducer molecule is recognized. The induction of the active conformation could
be performed by β-myrcene, or a β-myrcene-dependent stimulus, derived from the alterations occurring in the membrane (e.g. a physical stimulus derived from membrane stress or from the hydrophobic environment in the interface between the aqueous medium and β-myrcene droplets; interaction with a signaling/sensor protein located in the membrane). The active PM1_0216375 might activate the transcription of the membrane-bound β-myrcene hydroxylation system coded by the TU5, comprised by the PM1_0216370 and PM1_0216365 gene products, for instance, by recruiting the RNA polymerase or altering the promoter region topology which leads to transcription initiation.

Additionally, the profiles of promoter reporter expression suggested the sequential activation of transcriptional units TU5 >> TU6 >> TU3 during β-myrcene catabolism, in which the initial hydroxylation of β-myrcene into myrcen-8-ol seemed to be an important trigger for initial activation of the 28-kb GI. In particular, the reporter analysis (Figure 2) showed a dependence of the successful activation of TU6 (the dehydrogenase-coding cluster myrABC) on the expression of TU5 gene products, whereas the induction of TU3, which was not active in M1-C19, M1-C38 or in M1-N22, is in fact dependent on the successful recognition of β-myrcene stimulus and activation of TU5 and TU6 to carry the initial oxidations. This interplay between transcriptional units hints a regulatory coordination, in which a modulation by β-myrcene derivatives might be involved. Nevertheless, a putative activation mechanism mediated by metabolic intermediates does not fully explain the molecular outcomes registered. Both TU5 and TU6 (and other transcriptional units) might require additional regulatory elements to confer full metabolic potential to the 28-kb GI, as a downregulation of the reporter GFP was observed in both M1-N22 and M1-C38 strains for promoters P5 and P6 (Figure 2). Furthermore, a polar effect resulting from transposon insertion, altering transcription of downstream genes coded
in the polycistronic messenger RNA molecule, could also affect the transposon-derived phenotypes. However and particularly in the mutation of PM1_0216375 and PM1_0216370, the respective transcriptional units are not predicted to code more than one key protein (the regulator or the β-myrcene hydroxylase), thus, the transposon insertion was unlikely to cause downstream transcription inhibition, which is supported by the differences observed in the profile of promoter activity between M1-C19 and M1-C38 strains and mutant complementation.

Additionally, further molecular characterization will also be focus on describing the role of the PM1_0216360 and PM1_0216405 regulators. As described for MalT in *E. coli* (51) and other regulators of the aromatic hydrocarbon catabolic pathways (e.g. XylS and XylR from *P. putida* TOL plasmid (52), the regulatory protein PM1_0216375 might be synthesized constitutively in residual concentration, only becoming active in the presence of the inducer. A similar mechanism could also occur in the P4-controlled PM1_0216360 regulator, which shows similar one-component sensor-like structure as PM1_0216375. Thus, the absence of a significant increase in fold change of the reporter level of promoters P4 and P8 could be explained by a post-translational modulation of the regulator activity, without a significant increase in the expression of the respective transcriptional unit upon β-myrcene induction.

Although the 28-kb GI from M1 strain might have evolved towards the specialized mineralization of β-myrcene as shown by the reporter levels in Figure 6A, this locus was also responsive to other monoterpenes, with concomitant biotransformation of the inducer compounds (Figure 6B and Table S4 in Supplemental Material). In M1 *wt* strain, the promoter regions P2 and P6 (and to a lesser extent P1 and P3) were activated by acyclic (linalyl acetate) and cyclic (R-(+)-limonene and (-)-β-pinene) monoterpene-backbones (or their derivatives): P2 putatively involved in sensing a wide range of monoterpenic-related molecules, while P6-
controlled enzymes would be involved in the oxidation of different terpene alcohols. The putative gene products coded by TU3, namely the monooxygenase PM1_0216335 and epoxide hydrolase PM1_0216350, resemble the enzymatic machinery reported for the biotransformation of some cyclic monoterpenes (53–55) and could be involved in the oxidation or ring-opening of cyclic backbones. Nevertheless, comparison between monoterpenes was performed by using β-myrcene as reference compound and based on theoretical values regarding their chemical characterization (Table S1 in Supplemental Material). Monoterpenes used showed different physico-chemical properties (structure, volatility, solubility in aqueous medium, membrane partition) which influence substrate availability, thus, the outcome of biological experiments. Following work will focus on the detailed characterization of M1 response to β-myrcene and related monoterpenes, regarding i) the expression kinetics of the 28-kb GI TUs complemented with substrate titration, ii) effect of carbon sources/energetic balance on the performance of monoterpene biotransformation in co-metabolism experiments, iii) putative catabolic repression systems and cell adaptation mechanisms, as well as iv) detail possible regulatory cross-talks between secondary/redundant enzymatic systems, which might explain the broad range of monoterpane substrate of the 28-kb catabolic island and the ability of several M1 mutants (e.g. M1-N22) to overcome the disruption of the predicted central catabolic pathway for β-myrcene.

The versatility of M1 enzymatic repertoire and the ability of M1 cells to maintain membrane integrity when different catabolic steps are impaired, and harmful metabolites are accumulated (e.g. myrcen-8-ol) (46, 56, 57), emphasizes the environment-driven adaptation of this potential biochasssis towards particular carbon sources derived from plant material and/or with high hydrophobic character. The metabolic tuning of M1 strain, and in particular of the 28-kb genomic island, may prompt the development of optimized
biochassis and bioblocks for tailored biotransformation of natural and non-natural compounds, targeting industrial wastes and by-products as carbon feedstock for sustainable biotechnological processes. Furthermore, the functional elements characterized in this work might present potential targets for engineering strategies and incorporation as functional parts in plant-volatile-responsive circuits.

Acknowledgments

Vectors from the Standard European Vector Architecture (SEVA) library and pBAM1 used in this work were kindly provided by Prof. Victor de Lorenzo (CNB-CSIC, Madrid).

Funding information

This work was supported by the strategic programme UID/BIA/04050/2013 (POCI-01-0145-FEDER-007569) funded by national funds through the FCT I.P. and by the ERDF through the COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI), and through a Ph.D grant (grant number SFRH/BD/76894/2011) to P.S-C.
References


32. **Tropel D, van der Meer JR.** 2004. Bacterial transcriptional regulators for degradation


**Figure legends**

**Figure 1:** Operonic organization of the newly identified 28-kb GI harboring the β-myrcene core-code, predicted *in silico* by Soares-Castro and Santos (12) (A) and validated by promoter reporters (B). Genes are organized in transcriptional units (TU), promoter regions (P) are illustrated by blue/green arrows, and terminator sites depicted as red circles. Microscopy images correspond to the total amplification of 400x. Controls of promoter induction are shown for P2 and are representative of the absence of fluorescence in all reporters. Cultures of M1*wt* cells harboring pSEVA637 constructs were supplemented with 48 mM lactate or 47.8 μmol of β-myrcene in solution.

**Figure 2:** Overview of the promoter activity of the 28-kb genomic island in M1 *wt* and mutant strains, obtained from whole-genome mutagenesis. M1 strains harbored the reporter pSEVA637, fused to each one of the 6 validated promoter regions (P1, P2, P3, P5, P6, P7). Cells were grown in MM with 48 mM lactate and GFP fluorescence was measured after 5 hours of the supplementation with 9.6 μmol of β-myrcene. The heatmap (colored squares aligned with predicted transcriptional units – TU) indicates the variation in fold induction of the GFP reporter for promoters, between presence and absence of β-myrcene in the culture medium, as myrcene/lactate ratio of relative light units normalized with optical density of the culture. The knockout of genes or cluster of genes is shown for each mutant by the gray coloring and dotted line: disruption of *PM1_0216390* (coding the alcohol dehydrogenase MyrB) in M1-N22, by transposon insertion; disruption of *PM1_0216375* (coding for a regulatory protein from the LuxR family) in M1-C19, by transposon insertion;
disruption of \textit{PM1\textsubscript{0216370}} (coding a fatty acid desaturase) in M1-C38, by transposon insertion.

**Figure 3:** Effect of β-myrcene exposure in the membrane fatty acid content (A) and cell viability (B) of M1 and derived strains. (A) Membrane adaptive response, regarding the \textit{trans/cis} ratio of unsaturated acyl-chains (UFA) was analyzed by GC of fatty acid methyl esters, at 0.5 hour, 1.5 hour and 3 hours after supplementation with saturating amount of β-myrcene, in three independent experiments. Kinetics of M1 is shown in black, whereas M1 mutants are shown in gray: M1-N22 derived from the interruption of the alcohol dehydrogenase MyrB (♦), M1-C19 derived from the interruption of the transcriptional regulator \textit{PM1\textsubscript{0216375}} (▲), M1-C38 derived from the interruption of the putative myrcene hydroxylase \textit{PM1\textsubscript{0216370}} (●).(B) The number of cells as colony forming units was quantified during the same time range to assess cell viability. The specific growth rate for each strain growing in lactate (lac) or lactate with β-myrcene pulse (myr), during the 3 hours assayed, was calculated (μ) and expressed as ratio of $\mu_{\text{myr}}/\mu_{\text{lac}}$.

**Figure 4:** Homology of the 3 putative regulators coded by the β-myrcene-inducible 28-kb GI of M1 strain. The eight major families of transcriptional regulators involved in the catabolism of acyclic and cyclic/aromatic hydrocarbons are depicted in different colored branches. The regulators \textit{PM1\textsubscript{0216360}}, \textit{PM1\textsubscript{0216375}} and \textit{PM1\textsubscript{0216405}} are highlighted by a black box. The alignment of protein sequences was performed with MAFFT to generate the unrooted tree. A schematic representation of the protein domains predicted by InterPro for the \textit{PM1\textsubscript{0216360}}, \textit{PM1\textsubscript{0216375}} and \textit{PM1\textsubscript{0216405}} protein sequences, as well as for other characterized members of the respective families of regulators, is shown. ATPase, nucleotide-binding oligomerization...
domain; WHTH/HTH, Helix-Turn-Helix domain; a.a., amino acid residues. The uniprot accession numbers of the sequences used are the following: AcoK, Q48411; AlkS, P17051; AtuR, Q9HZW2; BadR, O07458; BenR, Q4K9X4; CatM, P07774; CatR, P20667; ClcR, Q05840; CymR, O33453; DmpR, Q06573; FadR, P0A8V6; HcaR, Q47141; IclR, P16528; LuxR, P35327; MalT, P06993; MarR, P27245; NahR, P10183; NtrC, P10577; PcaQ, P0A4T6; PcaR, I7C334; PobR, Q43992; StyR, O30989; TcbR, P27102; TfIR, Q46M57; TodT, I7CA98; XylR, P06519; XylS, P07859.

Figure 5: Conservation of the histidine-rich motifs HX3-4H, HX2-3HH and HXXHH in the β-myrcene-induced PM1_0216370 (A), common to all membrane non-heme hydroxylases, monooxygenases and fatty acid desaturases, described by . The “*” indicates positions which have a single, fully conserved residue (black). Residues conserved between the query sequences and PM1_0216370 are highlighted in grey. The numbering of each sequence represents the position of each residue in the original protein sequence. The homology-based analysis of the PM1_0216370 gene product (highlighted by a black box in panel A), annotated as a fatty acid desaturase, is showed in panel B. Protein sequences of PM1_0216370 and other reported membrane non-heme di-iron enzymes were aligned with MAFFT and the protein region comprising the 3 characteristic histidine-rich motifs was used to generate the unrooted tree. The accession numbers of the protein sequences retrieved from Uniprot are indicated in figure.

Figure 6: Activity of the promoter regions P1, P2, P3, P5, P6 and P7 in the presence of different monoterpenes. Cultures of M1 cells harboring the pSEVA637 constructs were supplemented with β-myrcene (A) and equimolar volumes of other monoterpenes (B), commonly used in industrial processes. Only those monoterpenes that showed significant
activation of the promoter regions are depicted: linalyl acetate and geraniol (aliphatic), R-(+)-limonene (monocyclic) and (-)-β-pinene (bicyclic). GFP expression was measured after 5 hours of incubation with 9.6 μmol of terpene substrate in solution. The promoter activity is shown as fold induction ratio of the relative fluorescence units normalized by OD$_{600\text{nm}}$ between cells challenged with the monoterpene and control cells grown only with 48 mM lactate. Statistical significance was calculated by two-way ANOVA test, using the fold change of lactate control (FC = 1) as reference: *, P-value < 0.05; **, P-value < 0.01; ***, P-value < 0.001.
Table 1: Bacterial strains and plasmids used in this work.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli S17.1λpir</td>
<td>recA thi pro hasD R  M+ RP4-2Tc::Mu-Km::Tn7  Tp R Sm R λpir; recipient strain for conjugation experiments</td>
<td>(15)</td>
</tr>
<tr>
<td>Pseudomonas sp. M1</td>
<td>wild type; phenol+ (o); toluene+ (m); myrcene+; cresol+; citronellol+; citral+</td>
<td>(11)</td>
</tr>
<tr>
<td>Pseudomonas sp. M1-N22</td>
<td>derived from M1 strain, mini-Tn5-Km cassette inserted in myrB (PM1_0216390)</td>
<td>(11)</td>
</tr>
<tr>
<td>Pseudomonas sp. M1-C19</td>
<td>derived from M1 strain, pBAM1-P2::GFP cassette inserted in PM1_021675</td>
<td>This work</td>
</tr>
<tr>
<td>Pseudomonas sp. M1-C38</td>
<td>derived from M1 strain, pBAM1-P2::GFP cassette inserted in PM1_021670</td>
<td>This work</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSEVA637</td>
<td>GFP promoter-probe vector for transcriptional fusions; pBBR1; Gm&lt;sup&gt;+&lt;/sup&gt;; 3.7 kbp</td>
<td>(16)</td>
</tr>
<tr>
<td>pSEVA637-P1</td>
<td>pSEVA637 derivative containing the promoter region P1 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.0 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P2</td>
<td>pSEVA637 derivative containing the promoter region P2 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.2 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P3</td>
<td>pSEVA637 derivative containing the promoter region P3 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.2 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P4</td>
<td>pSEVA637 derivative containing the promoter region P4 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.0 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P5</td>
<td>pSEVA637 derivative containing the promoter region P5 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 3.9 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P6</td>
<td>pSEVA637 derivative containing the promoter region P6 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.0 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P7</td>
<td>pSEVA637 derivative containing the promoter region P7 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.0 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA637-P8</td>
<td>pSEVA637 derivative containing the promoter region P8 from the M1 28-kb GI; Gm&lt;sup&gt;+&lt;/sup&gt;; 4.1 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pBAM1</td>
<td>Tn5-derived vector for random mutagenesis; R6K; Km&lt;sup&gt;+&lt;/sup&gt; Amp&lt;sup&gt;+&lt;/sup&gt;; 4.4 kbp</td>
<td>(17)</td>
</tr>
<tr>
<td>pBAM1-P2::GFP</td>
<td>pBAM1 derivative containing the promoter region P2::GFP; R6K; Km&lt;sup&gt;+&lt;/sup&gt; Amp&lt;sup&gt;+&lt;/sup&gt;; 5.7 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA638</td>
<td>pSEVA238 derivative with Gm&lt;sup&gt;+&lt;/sup&gt; cassette; xyl5-Pm based expression vector; pBBR1; 4.9 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA638-PM1_0216375</td>
<td>pSEVA638 derivative containing the promoterless PM1_0216375 sequence; Gm&lt;sup&gt;+&lt;/sup&gt;; 5.2 kbp</td>
<td>This work</td>
</tr>
<tr>
<td>pSEVA638</td>
<td>pSEVA638 derivative containing the promoterless TUS (PM1_0216365-PM1_0216370) sequence; Gm&lt;sup&gt;+&lt;/sup&gt;; 5.0 kbp</td>
<td>This work</td>
</tr>
</tbody>
</table>
Table 2: Primers used in this work.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’-3’)</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-FWD-gfp-BamHI</td>
<td>aataggatccctcatgctcgccagtggacg</td>
<td>pSEVA637-P1</td>
</tr>
<tr>
<td>P1-REV-gfp-HindIII</td>
<td>agagaagctttacacagttccagctgcg</td>
<td>pSEVA637-P1</td>
</tr>
<tr>
<td>P2-FWD-gfp-HindIII</td>
<td>aacaaagcttcacccccgagcatggg</td>
<td>pSEVA637-P2</td>
</tr>
<tr>
<td>P2-REV-gfp-BamHI</td>
<td>aagtggatccttggttccttcatcc</td>
<td>pSEVA637-P2</td>
</tr>
<tr>
<td>P3-FWD-gfp-BamHI</td>
<td>aactggatccacccccgagcatggg</td>
<td>pSEVA637-P3</td>
</tr>
<tr>
<td>P3-REV-gfp-HindIII</td>
<td>gcaaaagcttttcacccccgagcatggg</td>
<td>pSEVA637-P3</td>
</tr>
<tr>
<td>P4-FWD-gfp-HindIII</td>
<td>aactaaagctccattcagttccggggagattg</td>
<td>pSEVA637-P4</td>
</tr>
<tr>
<td>P4-REV-gfp-BamHI</td>
<td>attaggatcctcttgacgaagctttgag</td>
<td>pSEVA637-P4</td>
</tr>
<tr>
<td>P5-FWD-gfp-HindIII</td>
<td>aactaagctgegtcctgccacgcgggttgc</td>
<td>pSEVA637-P5</td>
</tr>
<tr>
<td>P5-REV-gfp-BamHI</td>
<td>aactaagcttacccccgagcatggg</td>
<td>pSEVA637-P5</td>
</tr>
<tr>
<td>P6-FWD-gfp-HindIII</td>
<td>aactaagcttgccgagctttgag</td>
<td>pSEVA637-P6</td>
</tr>
<tr>
<td>P6-REV-gfp-BamHI</td>
<td>attaggatccaaagaggagagtcgctgc</td>
<td>pSEVA637-P6</td>
</tr>
<tr>
<td>P7-FWD-gfp-BamHI</td>
<td>aacagggatccagttccgagccgttggag</td>
<td>pSEVA637-P7</td>
</tr>
<tr>
<td>P7-REV-gfp-HindIII</td>
<td>acttaggttaagagggaggagagtcgc</td>
<td>pSEVA637-P7</td>
</tr>
<tr>
<td>P8-FWD-gfp-BamHI</td>
<td>aacaggtctgtggcccgagcttgcc</td>
<td>pSEVA637-P8</td>
</tr>
<tr>
<td>P8-REV-gfp-HindIII</td>
<td>aactaagcttacccccgagcatggg</td>
<td>pSEVA637-P8</td>
</tr>
</tbody>
</table>

Whole-genome mutagenesis

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’-3’)</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2:: gfp-FWD-BamHI</td>
<td>attggatcctggtggcctctcatcc</td>
<td>pBAM-P2.GFP</td>
</tr>
<tr>
<td>P2:: gfp-REV-EcoRI</td>
<td>gatggaattcctggattctcaccaaaaaacgc</td>
<td>pBAM-P2.GFP</td>
</tr>
</tbody>
</table>

ARB6

ggcacgctgcagactaactacnnnnnnnnacgcc

ME-1-extR

tctgattccgatcaatagttgctc

gfp-extF

cgttggcctgttcttcccc

ARB2

ggcacgctgcagactaactac

ME-1-intR

cagttttattgttcatgatgatata

gfp-intF

tctcagaacgcctgcggttgcg

Complementation

tatagaactcatgatcgtcagcactacaaggg | pSEVA638-TU5 |

aaaagtgcgtcatactgtgtaaatgac | pSEVA638-TU5 |

atatggattccgtaaagactgtgtgcagcc | pSEVA638-PM1_0216375 |

taagtgcgtcatactgtgtaaatgac | pSEVA638-PM1_0216375 |
Table 3: Correspondence between the genes comprising the β-myrcene-inducible 28-kb GI in *Pseudomonas sp.* M1 with the promoter regions and transcriptional units predicted *in silico* by Soares-Castro and Santos (12).

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Predicted product</th>
<th>Promoter/TU</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM1_0216305</td>
<td>CoA dehydrogenase</td>
<td>P1/TU1</td>
</tr>
<tr>
<td>PM1_0216310</td>
<td>CoA dehydrogenase</td>
<td>P1/TU1</td>
</tr>
<tr>
<td>PM1_0216315</td>
<td>CoA acetyltransferase</td>
<td>P1/TU1</td>
</tr>
<tr>
<td>PM1_0216320</td>
<td>Sensory transducer</td>
<td>P2/TU2</td>
</tr>
<tr>
<td>PM1_0216325</td>
<td>Membrane protein</td>
<td>P2/TU2</td>
</tr>
<tr>
<td>PM1_0216330</td>
<td>CoA acetyltransferase</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216335</td>
<td>Monoxygenase</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216340</td>
<td>Oxidoreductase</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216345</td>
<td>CoA dehydrogenase</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216350</td>
<td>Epoxide hydrolase</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216355</td>
<td>Hypothetical protein</td>
<td>P3/TU3</td>
</tr>
<tr>
<td>PM1_0216360</td>
<td>LuxR-family transcriptional regulator</td>
<td>P4/TU4</td>
</tr>
<tr>
<td>PM1_0216365</td>
<td>Rubredoxin</td>
<td>P5/TU5</td>
</tr>
<tr>
<td>PM1_0216370</td>
<td>Fatty acid desaturase</td>
<td>P5/TU5</td>
</tr>
<tr>
<td>PM1_0216375</td>
<td>LuxR-family transcriptional regulator</td>
<td>P6/TU6</td>
</tr>
<tr>
<td>PM1_0216380</td>
<td>Oxidoreductase</td>
<td>P6/TU6</td>
</tr>
<tr>
<td>PM1_0216385</td>
<td>CoA synthetase MyrC</td>
<td>P6/TU6</td>
</tr>
<tr>
<td>PM1_0216390</td>
<td>Alcohol dehydrogenase MyrB</td>
<td>P6/TU6</td>
</tr>
<tr>
<td>PM1_0216395</td>
<td>Aldehyde dehydrogenase MyrA</td>
<td>P6/TU6</td>
</tr>
<tr>
<td>PM1_0216400</td>
<td>CoA hydratase</td>
<td>P7/TU7</td>
</tr>
<tr>
<td>PM1_0216405</td>
<td>LysR-family transcriptional regulator</td>
<td>P8/TU8</td>
</tr>
<tr>
<td>PM1_0216410</td>
<td>Hypothetical protein</td>
<td>P8/TU8</td>
</tr>
</tbody>
</table>
Table 4: Metabolites identified in β-myrcene (C\textsubscript{10}H\textsubscript{16}) biotransformation experiments with M1 wt. The effect of the different genotypes of the mutant strains is compared to the profile of the wt cells. The detailed characterization of the detected metabolites is shown in Table S2 and Figure S2 in the Supplemental Material.

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>GC-MS</th>
<th>Predicted ID (Formula)</th>
<th>M1 (wt)</th>
<th>N22</th>
<th>C19</th>
<th>C38</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 h</td>
<td>3.5 h</td>
<td>0.5 h</td>
<td>3.5 h</td>
</tr>
<tr>
<td>Myrcenal (C\textsubscript{10}H\textsubscript{14}O) (1)</td>
<td></td>
<td></td>
<td>0.8±0.2</td>
<td>1.8±0.4</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Myrcen-8-ol (C\textsubscript{10}H\textsubscript{16}O) (2)</td>
<td></td>
<td></td>
<td>9.0±3.6</td>
<td>6.0±1.3</td>
<td>14</td>
<td>50±10</td>
</tr>
<tr>
<td>4-Methylhexanoic acid (C\textsubscript{7}H\textsubscript{14}O\textsubscript{2}) (3)</td>
<td></td>
<td></td>
<td>ND</td>
<td>0.4±0.3</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>4-Methyl-3-hexenoic acid (C\textsubscript{7}H\textsubscript{12}O\textsubscript{2}) (4)</td>
<td></td>
<td></td>
<td>ND</td>
<td>1.3±0.2</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Myrcenoic acid (C\textsubscript{10}H\textsubscript{14}O\textsubscript{2}) (5)</td>
<td></td>
<td></td>
<td>0.4±0.2</td>
<td>17±0.6</td>
<td>0.1</td>
<td>5.8±0.3</td>
</tr>
</tbody>
</table>

* Peak area from identified metabolites normalized as percentage of the total area of the chromatogram, shown as averaged value ± standard deviation when replicates are available, for 30 min and 3.5 hours of β-myrcene biotransformation. ND, not detected in any replicate.
Table 5: Metabolites identified in biotransformation experiments with M1 wt, grown in MM supplemented with 48 mM lactate and acyclic (linalyl acetate and (+)-β-citronellol) or cyclic (R- (+)-limonene and (-)-β-pinene) monoterpenes. The detailed characterization of the detected metabolites is shown in Table S4 and Figure S8 in the Supplemental Material.

<table>
<thead>
<tr>
<th>GC-MS</th>
<th>Metabolite identification</th>
<th>Predicted ID</th>
<th>Formula</th>
<th>ID method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linalyl acetate (C₁₂H₂₀O₂)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1552</td>
<td>β-Linalool</td>
<td>C₆H₁₀O</td>
<td>MS, RI, Std</td>
<td></td>
</tr>
<tr>
<td>1852</td>
<td>Geraniol</td>
<td>C₆H₁₀O</td>
<td>MS, RI, Std</td>
<td></td>
</tr>
<tr>
<td>2342</td>
<td>Geranic acid</td>
<td>C₆H₁₀O₂</td>
<td>MS, RI, Std</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limonene (C₁₀H₁₆)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2287</td>
<td>Limonene-1,2-diol</td>
<td>C₆H₁₀O₂</td>
<td>MS, RI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-Pinene (C₁₀H₁₆)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1703</td>
<td>α-Terpineol</td>
<td>C₆H₁₀O</td>
<td>MS, RI, Std</td>
<td></td>
</tr>
<tr>
<td>2249</td>
<td>Citronelllic acid</td>
<td>C₆H₁₀O₂</td>
<td>MS, RI</td>
<td></td>
</tr>
</tbody>
</table>

RI, retention index; MS, mass spectrum search; Std, validation with pure standard.
A. Genomic locus-based analysis

B. Promoter probe-based analysis

M1 harboring promoter-probes 5 hours after β-myrcene pulse  
Controls

Fluorescence Bright Field

P1  P2  P3  P5  P6  P7  Lactate  Empty vector
### Table A

<table>
<thead>
<tr>
<th>Protein query</th>
<th>a.a.</th>
<th>HX3-4H</th>
<th>a.a.</th>
<th>HX2-3HH</th>
<th>a.a.</th>
<th>HXXHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM1_0216370</td>
<td>68</td>
<td>HDA-SH</td>
<td>106</td>
<td>H-MC-HH</td>
<td>244</td>
<td>HDM-HH</td>
</tr>
<tr>
<td>O85832</td>
<td>71</td>
<td>HDA-SH</td>
<td>110</td>
<td>H-NK-HH</td>
<td>247</td>
<td>HIHH</td>
</tr>
<tr>
<td>A3WE62</td>
<td>83</td>
<td>HEA-CH</td>
<td>124</td>
<td>H-LD-HH</td>
<td>261</td>
<td>HVHH</td>
</tr>
<tr>
<td>Q01332</td>
<td>23</td>
<td>HRVIMH</td>
<td>36</td>
<td>H-ES-HH</td>
<td>119</td>
<td>HRL-HH</td>
</tr>
<tr>
<td>A6N7G2</td>
<td>155</td>
<td>HRLA-H</td>
<td>193</td>
<td>H-RSHH</td>
<td>330</td>
<td>HNY-HH</td>
</tr>
<tr>
<td>Q12SO0</td>
<td>58</td>
<td>HRLWSH</td>
<td>96</td>
<td>H-RI-HH</td>
<td>234</td>
<td>HNF-HH</td>
</tr>
<tr>
<td>P12691</td>
<td>138</td>
<td>HEL-CH</td>
<td>169</td>
<td>HNGKH</td>
<td>313</td>
<td>HSD-HH</td>
</tr>
<tr>
<td>P21395</td>
<td>113</td>
<td>HEL-MH</td>
<td>144</td>
<td>HVNT-HH</td>
<td>283</td>
<td>HTN-HH</td>
</tr>
<tr>
<td>Consensus</td>
<td>1</td>
<td>*</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

### Diagram B

**Tree clusters of membrane non-heme di-iron enzymes:**

- β-Carotene hydroxylase cluster, which defines the CrtR-like domain from NCBI (cd03514)
- Fatty acid desaturases
- Other β-carotene hydroxylases from the literature
- Alkane and aromatic monoxygenases

<table>
<thead>
<tr>
<th>A9-FA desaturase, A6N7G2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9-FA desaturase, Q12SO0</td>
</tr>
</tbody>
</table>

- β-Carotene hydroxylase CrtZ, P94792
- β-Carotene hydroxylase CrtZ, Q44262
- β-Carotene hydroxylase CrtZ, Q01332
- β-Carotene hydroxylase CrtZ, P21688
- β-Carotene hydroxylase CrtZ, Q88HV7
- β-Carotene hydroxylase, A3WE62
- β-Carotene hydroxylase, O85832
- β-Carotene hydroxylase, Q8YQ27
- β-Carotene hydroxylase CrtR, AAP99312

- Alkane 1-monooxygenase AlkB, P12691
- Alkane 1-monooxygenase AlkB, O31250
- Alkane 1-monooxygenase AlkB, Q9AEN3
- Alkane 1-monooxygenase AlkB, E5G6V9

- Xylose monoxygenase XylM, P21395
- p-Cymene monoxygenase CymAa, O33456