



# Sequence-Based Screening for Rare Enzymes: New Insights into the World of AMDases Reveal a Conserved Motif and 58 Novel Enzymes Clustering in Eight Distinct Families

Janine Maimanakos<sup>1</sup>, Jennifer Chow<sup>1</sup>, Sarah K. Gaßmeyer<sup>2</sup>, Simon Güllert<sup>1</sup>, Florian Busch<sup>2</sup>, Robert Kourist<sup>2</sup> and Wolfgang R. Streit<sup>1\*</sup>

<sup>1</sup> Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Hamburg, Germany, <sup>2</sup> Junior Research Group for Microbial Biotechnology, Ruhr-University Bochum, Bochum, Germany

## OPEN ACCESS

### Edited by:

Peter Neubauer,  
Technical University of Berlin,  
Germany

### Reviewed by:

Shawn Chen,  
Revive Genomics Inc., USA  
Seung Gu Shin,  
Pohang University of Science  
and Technology, South Korea

### \*Correspondence:

Wolfgang R. Streit  
wolfgang.streit@uni-hamburg.de

### Specialty section:

This article was submitted to  
Microbiotechnology, Ecotoxicology  
and Bioremediation,  
a section of the journal  
Frontiers in Microbiology

**Received:** 04 May 2016

**Accepted:** 11 August 2016

**Published:** 25 August 2016

### Citation:

Maimanakos J, Chow J,  
Gaßmeyer SK, Güllert S, Busch F,  
Kourist R and Streit WR (2016)  
Sequence-Based Screening for Rare  
Enzymes: New Insights into the World  
of AMDases Reveal a Conserved  
Motif and 58 Novel Enzymes  
Clustering in Eight Distinct Families.  
Front. Microbiol. 7:1332.  
doi: 10.3389/fmicb.2016.01332

Arylmalonate Decarboxylases (AMDases, EC 4.1.1.76) are very rare and mostly underexplored enzymes. Currently only four known and biochemically characterized representatives exist. However, their ability to decarboxylate  $\alpha$ -disubstituted malonic acid derivatives to optically pure products without cofactors makes them attractive and promising candidates for the use as biocatalysts in industrial processes. Until now, AMDases could not be separated from other members of the aspartate/glutamate racemase superfamily based on their gene sequences. Within this work, a search algorithm was developed that enables a reliable prediction of AMDase activity for potential candidates. Based on specific sequence patterns and screening methods 58 novel AMDase candidate genes could be identified in this work. Thereby, AMDases with the conserved sequence pattern of *Bordetella bronchiseptica*'s prototype appeared to be limited to the classes of *Alpha*-, *Beta*-, and *Gamma*-*proteobacteria*. Amino acid homologies and comparison of gene surrounding sequences enabled the classification of eight enzyme clusters. Particularly striking is the accumulation of genes coding for different transporters of the tripartite tricarboxylate transporters family, TRAP transporters and ABC transporters as well as genes coding for mandelate racemases/muconate lactonizing enzymes that might be involved in substrate uptake or degradation of AMDase products. Further, three novel AMDases were characterized which showed a high enantiomeric excess (>99%) of the (*R*)-enantiomer of flurbiprofen. These are the recombinant AmdA and AmdV from *Variovorax* sp. strains HH01 and HH02, originated from soil, and AmdP from *Polymorphum gilvum* found by a data base search. Altogether our findings give new insights into the class of AMDases and reveal many previously unknown enzyme candidates with high potential for bioindustrial processes.

**Keywords:** arylmalonate decarboxylase (AMDase), racemase, *Variovorax* sp., *Polymorphum gilvum*, decarboxylation, flurbiprofen, sequence-based screening, Hidden Markov Models

## INTRODUCTION

Arylmalonate decarboxylases (AMDase; EC 4.1.1.76) belong to the family of carboxy-lyases (EC 4.1) and are very rare and underexplored enzymes. AMDases are capable of decarboxylating a range of  $\alpha$ -disubstituted malonic acid derivatives to enantiopure products without the need for any cofactor. This cofactor independent and enantioselective decarboxylation could open the door to highly valuable, (enantiopure) fine chemicals and important building blocks for pharmaceutical products with environmentally friendlier biotechnological methods rather than chemical processes. One example for these compounds are 2-arylpropionic acids which belong to the important class of non-steroidal, anti-inflammatory drugs (NSAID) like ibuprofen, flurbiprofen or naproxen (Kourist et al., 2011a). In the early nineties the patent for naproxen production expired and resulted in numerous investigations to optimize the production of the homochiral (*S*)-(+)-isomer (Harrington and Lodewijk, 1997), which is the almost exclusively pharmacologically active enantiomer (Harrison et al., 1970). The great interest in aryl alkanolic acids led to the discovery of the first and best characterized AMDase from *Bordetella bronchiseptica* (Miyamoto and Ohta, 1992a) which was object of a large number of investigations. The crystal structure of the enzyme was solved and enabled a prediction of a putative reaction mechanism (Okrasa et al., 2008) that was recently confirmed by semi-empirical calculations (Lind and Himo, 2014). Furthermore, there were constant efforts to optimize this AMDase for biotechnological use. It started with the inversion of enantioselectivity from (*R*)-products to (*S*)-isomers (Ijima et al., 2005; Obata and Nakasako, 2010) and proceeded with enhancement of AMDase activity of mutant enzymes (Terao et al., 2006; Kourist et al., 2011a; Miyauchi et al., 2011; Yoshida et al., 2015; Gaßmeyer et al., 2016). Engineered variants of the AMDase are able to produce several chiral building blocks from readily accessible alkenyl malonates and a range of enantioenriched, heteroaromatic  $\alpha$ -hydroxycarboxylic acids including pharmaceutical precursors (Okrasa et al., 2009; Lewin et al., 2015). The extensive knowledge about their structure and reaction mechanism makes them suitable candidates for further protein engineering approaches. But besides the intensive efforts of optimizing the enzymes' catalytic capabilities for industrial application only little is known about their occurrence and putative metabolic function. AMDases are members of the widespread Asp/Glu racemase superfamily together with aspartate (EC 5.1.1.13) and glutamate racemases (EC 5.1.1.3), hydantoin racemases (EC 5.1.99.5) and maleate isomerases (EC 5.2.1.1). Modified variants of the AMDase from *B. bronchiseptica* showed racemase activity (Kourist et al., 2011b; Gaßmeyer et al., 2015). The mechanism of this promiscuous reaction is currently an active field of research (Busch et al., 2016). Nevertheless, there are only four enzymes with proven AMDase activity known until today (Miyamoto and Ohta, 1992a; Miyamoto et al., 2007, 2008; Okrasa et al., 2008) followed by a larger number of annotated ORFs which might encode further AMDases.

The aim of this work was to establish a sequence-based screening algorithm for a reliable prediction of enzymes with

AMDase activity of *Bordetella*'s prototype, to classify newly discovered enzymes and to gain insights into the occurrence, spreading and genetic context of these biotechnologically promising enzymes.

## MATERIALS AND METHODS

### Databases Used in This Study

Nucleotide and amino acid sequences of putative AMDases were acquired from the NCBI<sup>1</sup> and IMG database (Integrated Microbial Genomes)<sup>2</sup>. Sequences were compared to others deposited in the NCBI database using BLAST alignment tools<sup>3</sup> (Altschul et al., 1990). Structural information on the enzymes was retrieved from the RCSB-PDB database<sup>4</sup>.

### Bioinformatic Analyses

Sequence data were processed using BioEdit<sup>5</sup>, Clone Manager Suite 9 (SciCentral software, licensed), Staden Package containing Pregap<sup>6</sup> and Gap4 and SeaView<sup>7</sup> 4.5.1/4.3.4 (Gouy et al., 2010). Maximum-likelihood phylogenetic trees based on amino acid sequences were constructed using SeaView<sup>7</sup> 4.5.1 and 4.3.4 (Gouy et al., 2010)].

The representative amino acid sequence of the *B. bronchiseptica* KU1201 AMDase (IMG Locus Tag Ga0077226\_10648), which included all known AMDase-specific motifs (Okrasa et al., 2008), was used for a protein blast search against the IMG and NCBI non-redundant database. Resulting hits with an *e*-value  $\leq 1e-44$  and an identity score  $\geq 43\%$  were analyzed and evaluated manually. In total, 63 candidate AMDases were identified, including 58 new enzymes.

Amino acid sequences of AMDases (**Supplementary Table S1**) were aligned using T-coffee 8.99 in sensitive mode. Subsequently, hmmbuild of the HMMER package 3.1b2 was used to construct a profile HMM (Hidden Markov Model) from the multiple alignment file. The HMM logo was visualized using the skylign webtool<sup>8</sup>. The new domain will be retrievable from the Pfam database collection<sup>9</sup>.

### Cloning and Heterologous Expression of *amdA*, *amdV*, and *amdP* in *Escherichia coli* BL21

For strains and plasmids used in this study see **Tables 1–3**. Cloning of *amdA* (IMG locus tag Ga0099533\_11097; 708 bp) and *amdV* (IMG locus tag Ga0099534\_11097; 711 bp) into the expression vector pET21a (Novagen/Merck, Darmstadt, Germany) was accomplished after amplification from genomic

<sup>1</sup><http://www.ncbi.nlm.nih.gov/>

<sup>2</sup><https://img.jgi.doe.gov/cgi-bin/er/main.cgi>

<sup>3</sup><http://blast.ncbi.nlm.nih.gov/Blast.cgi>

<sup>4</sup>[www.rcsb.org](http://www.rcsb.org)

<sup>5</sup><http://www.mbio.ncsu.edu/bioedit/bioedit.html>

<sup>6</sup><http://staden.sourceforge.net/>

<sup>7</sup><http://doua.prabi.fr/software/seaview>

<sup>8</sup>[www.skylign.org](http://www.skylign.org)

<sup>9</sup><http://pfam.xfam.org/>

**TABLE 1 | Bacterial strains used in this work.**

Strain	Properties <sup>1</sup>	Source
<i>Escherichia coli</i> DH5 $\alpha$	<i>supE44</i> $\Delta$ <i>lacU169</i> ( $\Phi$ 80 <i>lacZ</i> $\Delta$ M15) <i>hsdR17</i> <i>recA1</i> <i>endA1</i> <i>gyrA96</i> <i>thi-1</i> <i>relA1</i>	Invitrogen [Karlsruhe, Germany; (Hanahan, 1983; Bethesda Research Laboratories, 1986)]
<i>Escherichia coli</i> EPI300	Host strain for fosmid libraries F <sup>-</sup> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>d</i> <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74</i> <i>recA1</i> <i>endA1</i> <i>araD139</i> $\Delta$ ( <i>ara, leu</i> ) 7697 <i>galU</i> <i>galK</i> $\lambda$ - <i>rpsL</i> <i>nupG</i>	Epicentre® (Madison, WI, USA)
<i>Escherichia coli</i> BL21 (DE3)	F <sup>-</sup> , <i>ompT</i> , <i>hsdS</i> B (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) <i>gal</i> , <i>dcm</i> , $\lambda$ DE3	Novagen/Merck (Darmstadt, Germany)
<i>Achromobacter</i> sp. HH01	<i>amd</i>	Soil sample, Botanical Garden (Duisburg, Germany), enrichment culture with phenylmalonic acid as sole carbon source
<i>Variovorax paradoxus</i>	Type strain, DSM-30034	DSMZ (Braunschweig, Germany)
<i>Variovorax</i> sp. HH01	<i>amdA</i>	Soil sample, Botanical Garden (Duisburg, Germany), enrichment culture with phenylmalonic acid as sole carbon source
<i>Variovorax</i> sp. HH02	<i>amdV</i>	Soil sample, Botanical Garden (Duisburg, Germany), enrichment culture with phenylmalonic acid as sole carbon source

<sup>1</sup>Geno- and phenotypes according to (Bachmann, 1983).

**TABLE 2 | Vectors used in this work.**

Vector	Properties <sup>1</sup>	Size (kb)	Source
pDrive	TA-cloning vector, <i>oriEc</i> , P <sub>lac</sub> / <i>lacZ</i> , Amp <sup>R</sup> , Kan <sup>R</sup> , T7 promoter	3.85	QIAGEN (Hilden, Germany)
pET-21a	Expression vector, <i>lacI</i> , Amp <sup>R</sup> , T7- <i>lac</i> - promoter, C-terminal His <sub>6</sub> -tag coding sequence	5.44	Novagen/Merck (Darmstadt, Germany)
pCC1fos <sup>TM</sup>	Fosmid vector, Chl <sup>R</sup> , <i>redF</i> , <i>oriV</i> , <i>ori2</i> , <i>repE</i> , <i>parA</i> , <i>parB</i> , <i>parC</i> , <i>cos</i> , <i>loxP</i> , <i>lacZ<math>\alpha</math></i> , T7 Promoter	8.139	Epicentre® (Madison, WI, USA)
pEXA	Cloning vector, Amp <sup>R</sup> , P <sub>lac</sub> / <i>lacZ</i> , pUC <i>ori</i>	2.450	Eurofins MWG Operon (Ebersberg, Germany)

<sup>1</sup>Geno- and phenotypes according to the manufacturer's specifications.

DNA using gene specific primer pairs with underlined restriction sites (*amdA*\_for: 5'-AGGGGATCCATGACGAAACCCCATCTGGG-3'/*amdA*\_rev: 5'-ATGAAGCTTGGCGAAAAGACGGCCGTGGC-3' and *amdV*\_NdeI\_for: 5'-CATATGGTGACACAGCAGCCCCATC-3'/*amdV*\_XhoI\_rev: 5'-CTCGAGGGCGA-ACAGCAGTCCC-3'). The gene sequence of *amdP* (744 bp) was obtained from IMG (IMG locus tag SL003B\_3169) and synthesized with codon usage adapted to *E. coli* (MWG Eurofins, Germany; sequence deposited in Supplementary Information). The constructs were verified by sequencing (MWG Eurofins, Germany) and were subsequently transformed into *E. coli* BL21. Cultures were grown aerobically in LB medium containing 100  $\mu$ g/ml ampicillin at 37°C until they reached an OD<sub>600</sub> of 0.5. After induction with 0.1 mM IPTG, the corresponding proteins were expressed at 22°C for 20 h harboring a C-terminal sixfold histidine tag. The enzymes AmdA, AmdP, and AmdV were purified from crude cell extract with nickel-ion affinity chromatography using Ni-NTA agarose (Qiagen, Hilden, Germany) and checked for the correct sizes by SDS-PAGE. The enzymes were dialyzed over night at 4°C against 10 mM EDTA

(pH 7.5) in a SERVAPOR tube (SERVA, Heidelberg, Germany;  $\emptyset$  16 mm, MWCO 12000-14000).

## Enrichment Culture of Microorganisms from Soil Samples

Soil samples originated from the Botanical Garden in Duisburg, Germany, (latitude: N51°29'15.8", longitude: E6°46'3.5") and were incubated aerobically at 30°C in 100 mL flasks for 3 days with a selective enrichment medium [1 L contained 5.0 g NaNO<sub>3</sub>  $\times$  3 H<sub>2</sub>O; 1.01 g K<sub>2</sub>HPO<sub>4</sub>  $\times$  3 H<sub>2</sub>O; 0.25 g MgSO<sub>4</sub>  $\times$  7 H<sub>2</sub>O; 0.015 g CaCl<sub>2</sub>  $\times$  7 H<sub>2</sub>O; 0.14 g FeSO<sub>4</sub>  $\times$  7 H<sub>2</sub>O; 0.02 g EDTA. All components were prepared as 100-fold concentrated solutions and autoclaved separately. The medium was supplemented with 10 mL of a trace element solution [100-fold containing 15 mg Co(NO<sub>3</sub>)<sub>2</sub>  $\times$  6 H<sub>2</sub>O; 61.8 mg H<sub>3</sub>BO<sub>4</sub>; 198 mg MnCl<sub>2</sub>  $\times$  4 H<sub>2</sub>O; 57.5 mg ZnSO<sub>4</sub>  $\times$  7 H<sub>2</sub>O; 47.5 mg NiCl<sub>2</sub>  $\times$  H<sub>2</sub>O; 2.5 mg CuSO<sub>4</sub>  $\times$  5 H<sub>2</sub>O; 2.4 mg Na<sub>2</sub>MoO<sub>4</sub>; 4 mg CaCl<sub>2</sub>  $\times$  6 H<sub>2</sub>O) and 10 mM phenylmalonic acid (Sigma-Aldrich)]. For streak dilution, agar plates were prepared with the same medium and incubated at the same temperature to get single colonies, which were then transferred onto indicator medium agar plates [500 mL contained 0.5 g KH<sub>2</sub>PO<sub>4</sub>; 0.5 g yeast extract; 1 g NaCl; 1 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; pH was set to pH 6.0 with KOH/HCl. After autoclaving and cooling to 60°C, sterile filtered 0.025 g kresol red (solved in 70% ethanol) and 1.8 g phenylmalonic acid (solved in H<sub>2</sub>O, pH set to 6 with KOH) was added]. The low amount of yeast extract enabled a fast growth of cell mass. When yeast extract was depleted, phenylmalonic acid was the sole carbon source and its decarboxylation caused a pH increase. Colonies which showed a color shift after 2–3 days provided initial indication of using phenylmalonic acid as a carbon source and were further investigated.

## Biochemical characterization of AmdA, AmdP, and AmdV

For general activity tests, the three different AMDases were assayed using purified recombinant protein. Unless otherwise indicated, the enzymes were added to a substrate solution [10 mM phenylmalonic acid (Sigma-Aldrich) dissolved in 50 mM sodium

**TABLE 3 | Constructs created, resp. used in this work.**

Construct	Vector	Insert size (kb)	Properties
pET-21a::amdA	pET-21a	0.708	<i>amdA</i> derived from <i>Variovorax</i> sp. HH01, inserted at <i>Bam</i> HI and <i>Hind</i> III restriction sites, with C-terminal His <sub>6</sub> -tag coding sequence, Amp <sup>R</sup>
pET-21a::amdV	pET-21a	0.712	<i>amdV</i> derived from <i>Variovorax</i> sp. HH02, inserted at <i>Xho</i> I and <i>Nde</i> I restriction sites, with C-terminal His <sub>6</sub> -tag coding sequence, Amp <sup>R</sup>
pET-21a::amdP	pET-21a	0.745	<i>amdP</i> derived from <i>Polymorphum gilvum</i> SL003B-26A1, inserted at <i>Xho</i> I and <i>Nde</i> I restriction sites, with C-terminal His <sub>6</sub> -tag coding sequence, Amp <sup>R</sup>

phosphate buffer, pH 7.0]. The reaction was stopped by the addition of 2 M HCl and by putting the reaction tubes on ice. The decrease of substrate concentration was measured by isocratic, reverse-phase HPLC (LaChrom Elite®, VWR International GmbH, Germany) with RP-18 columns (LiChroCART 125-3, Purophor RP-18E und LiChroCART 125-4, Purophor RP-18, VWR International GmbH, Germany). The elution solvent was composed of 80% sodium phosphate buffer (50 mM, pH 7) and 20% methanol with a flow rate of 0.5 ml/min. Analysis of measurements was performed by EZChrom Elite (version 3.2.1, licensed, VWR International GmbH, Germany). Retention times of phenylmalonic acid and phenylacetic acid were 1.5–2.5 min resp. 4.5–6.5 min, dependent on column diameter at a wavelength of 214 nm.

All samples were measured in triplicate with three biological replicates. For determination of the optimal temperature, samples were incubated between 4 and 50°C for 45 min. Temperature stabilities were assayed at 22 and 4°C for up to 30 h and at –20°C for up to 70 days. The impact of pH conditions on enzyme activity was measured using phenylmalonic acid substrate solutions prepared in 50 mM citric acid buffer (pH 4.0, 5.0 and 6.0), 50 mM sodium phosphate buffer (pH 6.0 and 7.0) or 50 mM potassium phosphate buffer (pH 6.0, 6.5, and 7.0) and 50 mM Tris buffer (pH 7.0, 8.0, and 9.0). The pH dependent enzyme activity of AmdA was measured using 20 mM phenyl malonic acid substrate solution solved in 100 mM citric acid buffer (pH 4.3, 5.2, 5.4, and 5.8), 200 mM potassium phosphate buffer (pH 5.8, 6.5, and 7.3) and 30 mM Tris buffer (pH 7.0 and 8.0).

The influence of salts and possible cofactors on AmdP, AmdV, and AmdA was assayed with the following compounds in 1 mM final concentration: sodium phosphate, pyridoxal phosphate, NAD, ATP, thiamine diphosphate, zinc chloride, sodium chloride, and magnesium chloride. The reaction using 10 mM phenylmalonic acid was stopped with 2 M HCl after 45 min.

To determine the solvent stability of the AMDases, ethanol, methanol, isopropanol and DMSO were added to the reaction

mixture (10 mM phenylmalonic acid in 50 mM Tris-buffer, pH 7.0) to a final volume of 10% (v/v). The reaction was stopped after 45 min and the activity compared to solvent-free controls.

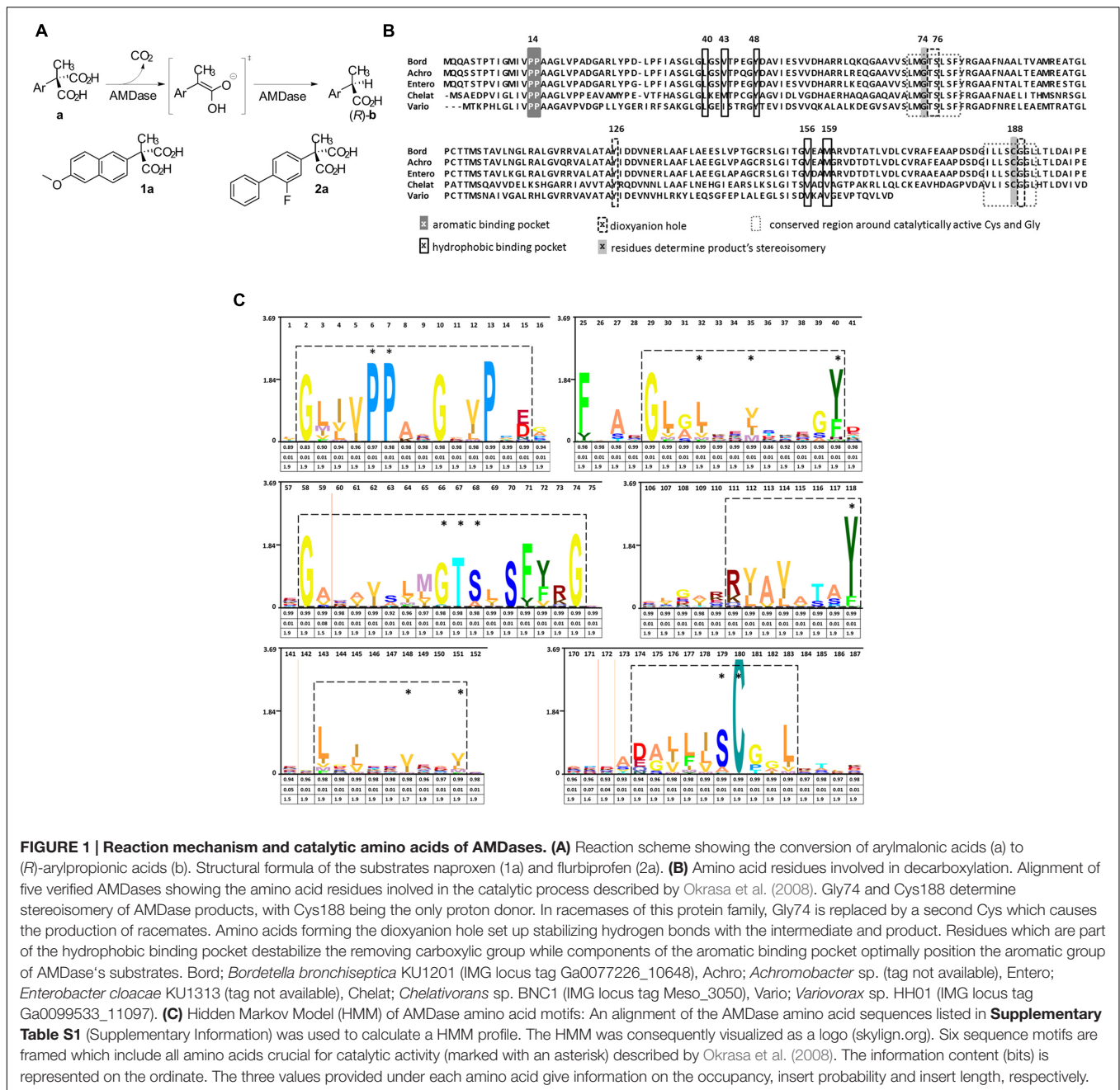
## Selectivity of AmdA, AmdV, and AmdP on Derivatives of 2-Arylpropionic Acids

The selectivity of the three different AMDases in the synthesis of optically pure 2-arylpropionic acids was analyzed by using purified recombinant protein. The enzymes were added to a substrate solution [10 mM **1a** or **1b** dissolved in 50 mM tris buffer, pH 7.5]. **1a** or **1b** (see **Figure 1A**) were prepared as described in (Gaßmeyer et al., 2016). The reaction was stopped by the addition of 2 M HCl and 2-arylpropionic acids were extracted twice with methyl *tert*-butyl ether (500 µL). Methanol (100 µL) and trimethylsilyl-diazomethane (25 µL) were added and the mixture incubated at room temperature for 30 min. Acetic acid (5 µL) was added, the solvents evaporated to dryness and the residue was re-dissolved in ethyl acetate (200 µL). Conversion and optical purity were then analyzed by chiral gas chromatography on a Shimadzu GC Plus 2010 device (Gaßmeyer et al., 2016); using a FS-Hydrodex-β-6TBDM column (Macherey-Nagel), with a method operating isothermic at 160°C (for naproxen) and 170°C (for flurbiprofen) with an injection split of 1/20. The elution order was for both 2-arylpropionic acids (*S*) and (*R*) with baseline separation and a peak difference of 0.5 min. *ee*-values were determined from the relative peak areas according to standard procedures. *ee* stands for enantiomeric excess and represents the excess of one enantiomer in a racemic mixture. Errors of *ee*-value determination were extrapolated from the area difference obtained from the respective racemic authentic standard. From this, the precision of *ee*-values was found to be ±0.3%.

## RESULTS

### Conserved AMDase-Specific Amino Acid Motif as the Basis for a Search Algorithm

For the determination of AMDase-specific searching criteria an amino acid sequence alignment of known AMDases was established initially (**Figure 1B**). We then verified, that the catalytically crucial amino acids described in the postulated reaction mechanism (Okrasa et al., 2008; Lind and Himoto, 2014) are conserved within all representatives. For the alignment only enzymes were chosen which (i) have a proven AMDase activity of the order of *Bordetella*'s prototype and (ii) are able to decarboxylate phenylmalonic acid and more complex prochiral substrates. Five enzymes met these requirements (**Figure 1B**): Four of them are the characterized AMDases mentioned above and additionally one enzyme from *Variovorax* sp. HH01. The strain is an isolate of an enrichment culture derived from soil with phenylmalonic acid as sole carbon and energy source (see Material and Methods) and the respective enzyme was characterized in the course of this work. Based on the published AMDases and our *Variovorax* enzymes, we were able to establish a total of 12 conserved motifs that could



**FIGURE 1 | Reaction mechanism and catalytic amino acids of AMDases. (A)** Reaction scheme showing the conversion of arylmalonic acids (a) to (*R*)-arylpropionic acids (b). Structural formula of the substrates naproxen (1a) and flurbiprofen (2a). **(B)** Amino acid residues involved in decarboxylation. Alignment of five verified AMDases showing the amino acid residues involved in the catalytic process described by Okrasa et al. (2008). Gly74 and Cys188 determine stereoisomery of AMDase products, with Cys188 being the only proton donor. In racemases of this protein family, Gly74 is replaced by a second Cys which causes the production of racemates. Amino acids forming the dioxanion hole set up stabilizing hydrogen bonds with the intermediate and product. Residues which are part of the hydrophobic binding pocket destabilize the removing carboxy group while components of the aromatic binding pocket optimally position the aromatic group of AMDase's substrates. Bord; *Bordetella bronchiseptica* KU1201 (IMG locus tag Ga0077226\_10648), Achro; *Achromobacter* sp. (tag not available), Entero; *Enterobacter cloacae* KU1313 (tag not available), Chelat; *Chelativorans* sp. BNC1 (IMG locus tag Meso\_3050), Vario; *Variovorax* sp. HH01 (IMG locus tag Ga0099533\_11097). **(C)** Hidden Markov Model (HMM) of AMDase amino acid motifs: An alignment of the AMDase amino acid sequences listed in **Supplementary Table S1** (Supplementary Information) was used to calculate a HMM profile. The HMM was consequently visualized as a logo (skylign.org). Six sequence motifs are framed which include all amino acids crucial for catalytic activity (marked with an asterisk) described by Okrasa et al. (2008). The information content (bits) is represented on the ordinate. The three values provided under each amino acid give information on the occupancy, insert probability and insert length, respectively.

be used to identify AMDase sequence homologs (Table 4). The conserved amino acids described in Okrasa et al. (2008) and listed in Table 4 include all AMDase-specific protein structural motifs known today, like the hydrophobic pocket, the dioxanion hole and the aromatic binding pocket and additionally two sequence motifs surrounding the residues, which determine the product's stereoisomery. These conserved sequence motifs were used to screen data bases for further enzymes with AMDase activity. In these analyses only proteins, which met all 12 criteria with identical amino acids or amino acid residues with similar physicochemical properties were chosen. With the established search algorithm, 58 further ORFs

which most likely encoded enzymes with AMDase activity were identified from the IMG and NCBI databases. The amino acid sequence alignment of these enzymes was used to create a Hidden Markov Model (HMM) shown in Figure 1C and enabled the deduction of six conserved sequence motifs (Table 5) which seemed specific for the prototype AMDase from *B. bronchiseptica* and include all known catalytically active residues. The HMM was implemented into the search algorithm and used for a reliable prediction of AMDase-active enzymes. AMDase are members of the protein superfamily aspartate/glutamate racemases, including aspartate racemases (EC 5.1.1.13), glutamate racemases (EC 5.1.1.3), hydantoin

**TABLE 4 | Detailed list of 12 determined searching criteria for the identification of AMDases in data bases (numbering of amino acid according to the AMDase from *Bordetella bronchiseptica* KU1201).**

Number	Criteria	Function
(1)	P14 P15	Part of large binding pocket
(2)	T75 S76	Part of dioxyanion hole
(3)	Y126	
(4)	G189	
(5)	L40	Building blocks of hydrophobic binding pocket
(6)	V43	
(7)	Y48	
(8)	V156	
(9)	M159	
(10)	G74	Distinctive for AMDases instead of second active site Cys characteristic for the related racemases.
	C188	Protonation of postulated endiolat intermediate
(11)	LMGTSLSF	Conserved sequence motifs surrounding residues that determine product's stereoisomery (positions 72–79 and 184–191).
(12)	ILLSCGGL	

racemases (EC 5.1.99.5) and maleate *cis-trans* isomerases (5.2.1.1). AMDases of *B. bronchiseptica*'s prototype with their conserved sequence patterns are clearly defined within the superfamily, as can be seen in **Figure 2**. Characteristic catalytic elements are the dioxyanion hole, the hydrophobic pocket, the large binding pocket for aromatic residues and a thiolate hole for the accommodation of the deprotonated Cys residue. The search algorithm and deduced sequence patterns allow a well-defined distinction between “true” AMDases and other members of the aspartate/glutamate superfamily for the very first time.

## Classification of AMDases

To date, 63 ORFs encoding putative AMDases were identified (status January 2016). All of the 25 *amd*-carrying bacterial genera belong exclusively to the class of *Alpha*-, *Beta*- and *Gamma*-*proteobacteria* (**Figure 3**). *Alpha*-*proteobacteria* include 25 *amd* genes, distributed over 16 genera, *Beta*-*proteobacteria* contain 34 AMDase genes distributed over six genera and within *Gamma*-*proteobacteria* four ORFs encoding AMDases were found, spread over three genera (**Figures 4A,B**). In strains with complete genome sequences, AMDase genes are encoded on the chromosomes. A singularity represents the strain *Nitratireductor pacificus* pht-3B which is the only isolate that encodes two different *amd* genes. *N. pacificus* pht-3B is part of a pyrene-degrading consortium of enriched sediment from the Pacific Ocean (Nie et al., 2012).

It is intriguing that for most of the genera in which AMDases were identified; only one or two different strains with the *amd* genes could be detected. However, within the genus *Variovorax* 23 strains harbored an *amd* gene and within the genus *Alcaligenes* seven strains that carried *amd* genes were detected (**Figure 5B**). This observation may indicate that the *amd* gene was originally evolved within the genus *Variovorax* or a closely related group of

bacteria. The findings also suggest that the *amd* gene is not part of an essential core genome of any of these bacteria but rather part of the accessory pan-genome.

Bacterial strains with *amd* genes derived from six isolation sources (**Figure 4C**): soil, plants (plant associated or endophytes), wastewater, oceans, hosts, and air. For 16% of the strains no distinct isolation source was mentioned. The majority (32%) is composed of strains that have been isolated from soil. All strains and gene locus tags are listed in **Supplementary Table S1** (Supplementary Information).

Comparison of *amd* flanking sequences enabled a more precise classification and characterization of the AMDases (**Figure 5**). One of the most widely spread ORFs was a gene encoding a mandelate racemase/muconate lactonizing enzyme (MR/MLE, e.g., IMG gene ID of *Variovorax* sp. HH01: 2639906160) which was present in 61% of the available sequences. Mandelate racemases (EC 5.1.2.2) and muconate cycloisomerases (EC 5.5.1.1) belong to the enolase superfamily and are involved in aromatic metabolism (Hasson et al., 1998). Moreover, genes commonly occurring were encoding tripartite tricarboxylate transporters (TTT). These sodium ion-dependent transporters translocate various substances, including citrate and its derivatives and are frequently found in *Beta*-*proteobacteria* (Antoine et al., 2003). Further genes cumulating in the sequence surrounding *amd* are ORFs coding for TRAP transporters and ABC transporters (Schneider and Hunke, 1998). TRAP transporters are tripartite ATP-independent periplasmic transporters which are common in prokaryotes (Mulligan et al., 2011). Like TTTs and ABC transporters, they consist of cytoplasmic membrane proteins and a periplasmic solute-binding protein.

Comparison of *amd* neighboring sequences combined with the amino acid sequence similarities allowed a first classification of all known AMDases into eight enzyme clusters (**Figure 5B**). Cluster TTT-I to TTT-V comprise isolates with genes encoding a TTT next to the *amd* genes. With exception of *Azoarcus* sp. all *Beta*-*proteobacteria* belong to the clusters TTT-I to TTT-III. Cluster TTT-IV and TTT-V are composed of *Alpha*-*proteobacteria*. By now there is only one cluster with *amd* genes flanked by ORFs encoding ABC transporters. This cluster includes the marine isolates *Cucumibacter marinus* DSM 18995 and *N. pacificus* pht-3B as well as the isolates of *Devosia* sp. There are two clusters of AMDases with ORFs encoding TRAP transporters nearby. Thereby the cluster TRAP-I is the only enzyme cluster with representatives of all three proteobacterial classes carrying AMDase coding genes.

## Biochemical Characterization of the Three Novel AMDases AmdA, AmdV, and AmdP

For further investigations the enzymes AmdA, AmdV, and AmdP were chosen. AmdA and AmdV originated from *Variovorax* sp. HH01 resp. *Variovorax* sp. HH02. The *Variovorax* strains are isolates from one soil sample and showed the highest activity on indicator medium plates. Furthermore, special interest lies in them as no other enzymes from this genus have been characterized so far. AmdP has been found by data base search

**TABLE 5 | Conserved sequence patterns and catalytically active amino acids of AMDases and their function.**

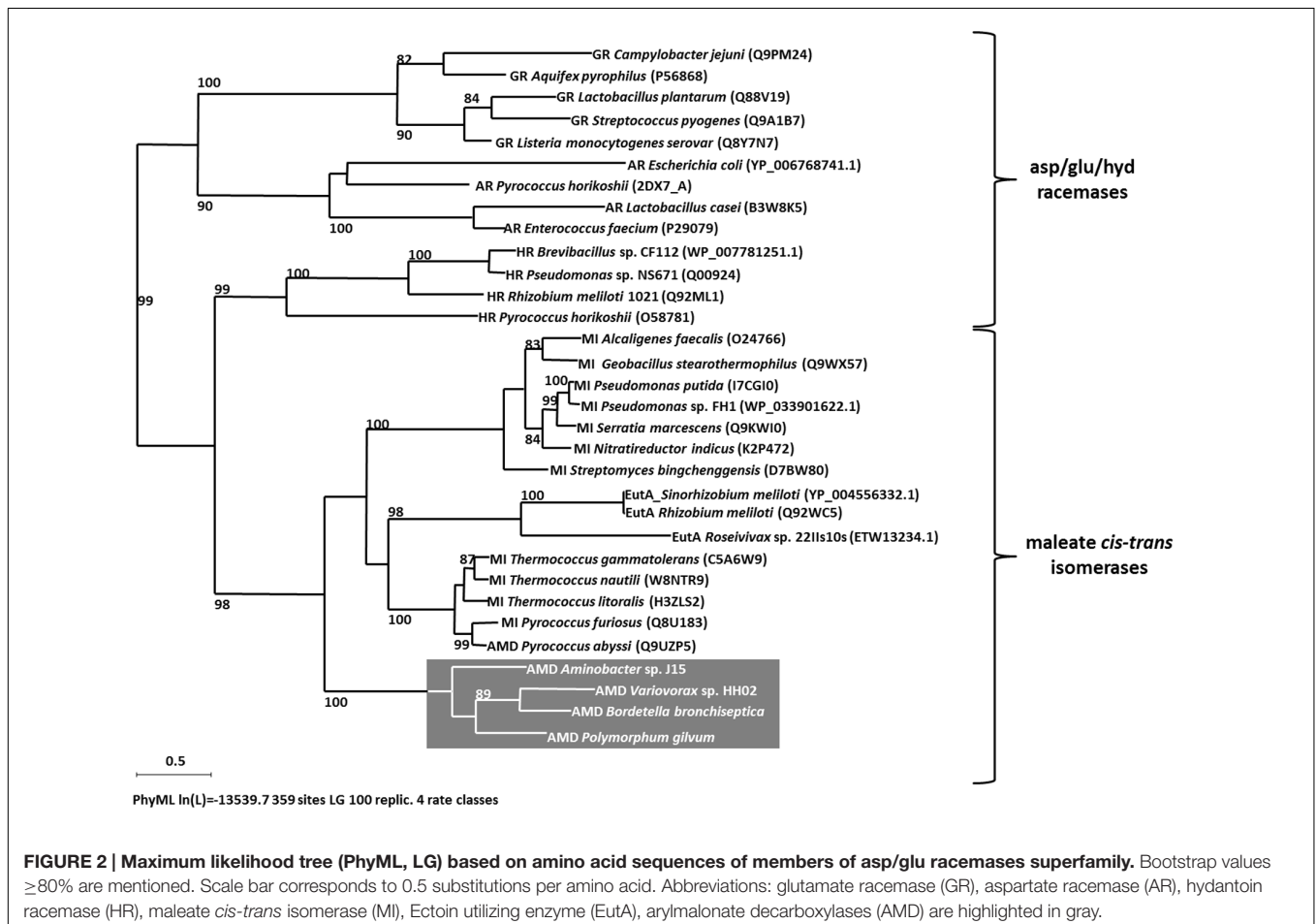
Number	Sequence pattern	Function of conserved amino acids
(1)	GLIV <b>PPA</b> xGxVPxE <sub>(10–23)</sub>	Binding of aromatic residues
(2)	GLGLxxV <b>xxxGY</b> <sub>(37–48)</sub>	Destabilization of leaving carboxylic group
(3)	G <b>AxxVxL</b> MG <b>TSL</b> SFYRG <sub>(66–82)</sub>	G74 determines product's stereoisomery; hydrogen bridge bond
(4)	RVAVx <b>TAY</b> <sub>(119–126)</sub>	Hydrogen bridge bond
(5)	Lx <b>lxxVxxV</b> <sub>(151–159)</sub>	Destabilization of leaving carboxylic group
(6)	DALLIS <b>CG</b> xL <sub>(182–191)</sub>	Active site Cys determines product's stereoisomery, hydrogen bridge bond

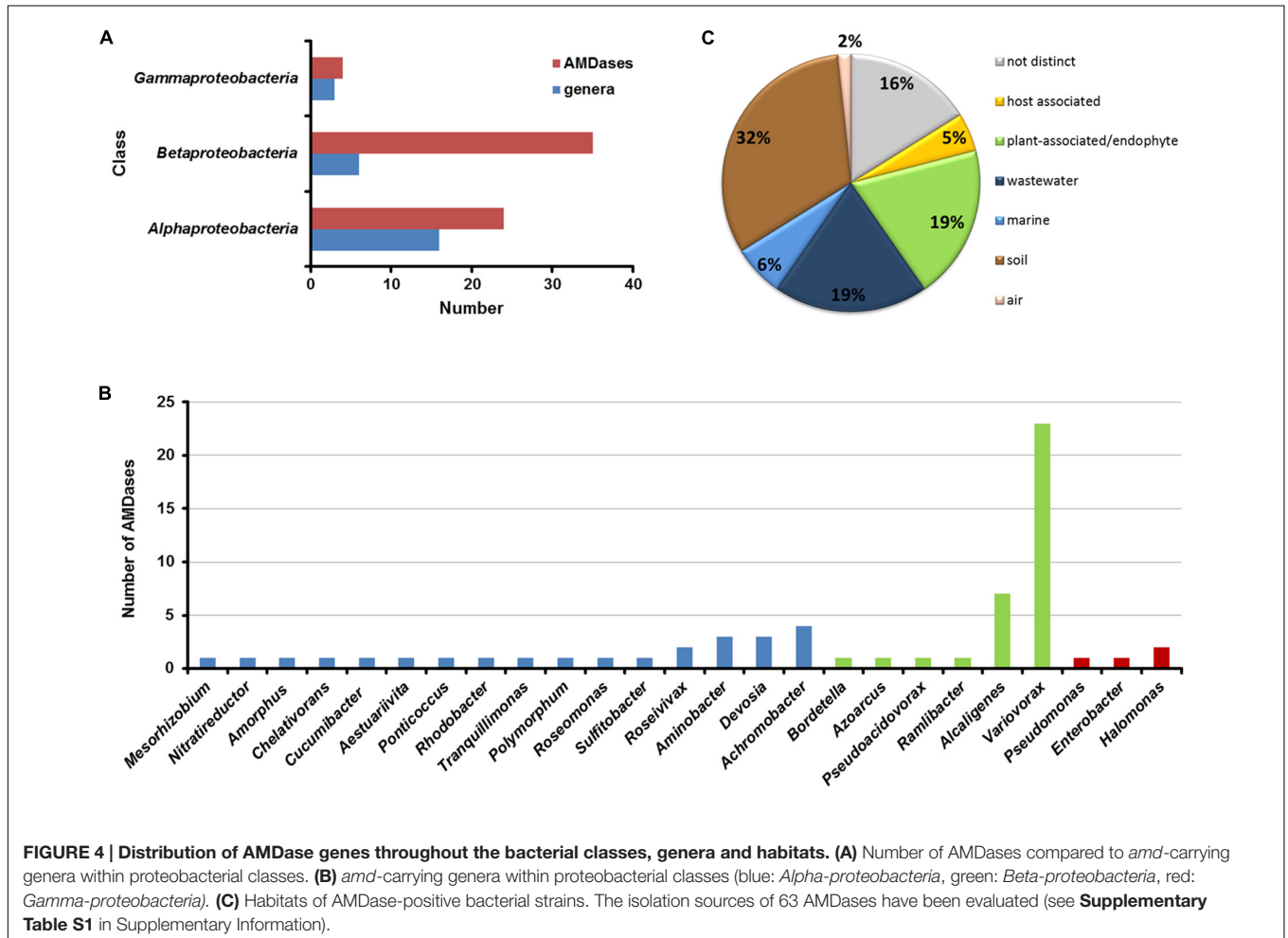
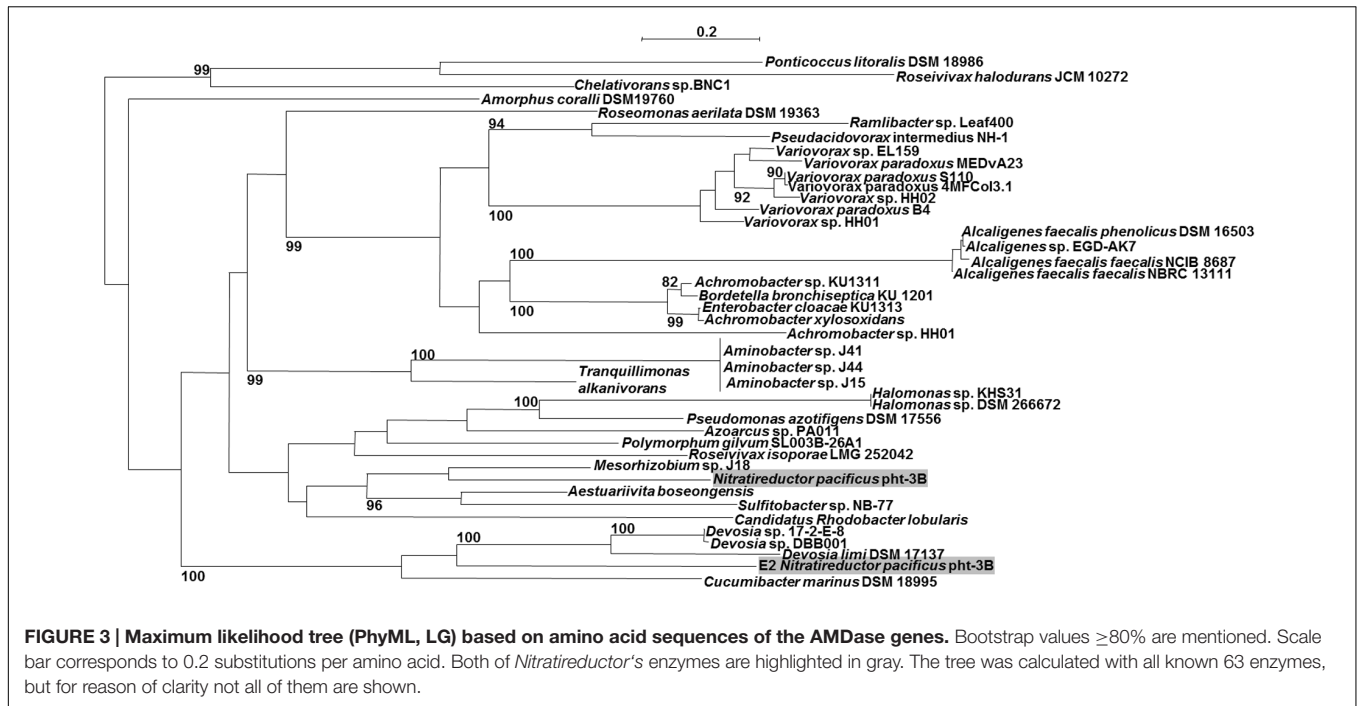
The letter *x* indicates an unconserved position within the sequence pattern. Numbering of amino acids is according to the AMDase from *B. bronchiseptica* KU1201. Catalytically active amino acids are printed in bold.

as a candidate to prove the applicability and reliability of the searching criteria. The corresponding gene originated from *Polymorphum gilvum* SL003B-26A1, a strain which was originally isolated from a crude oil-polluted saline soil and was able to

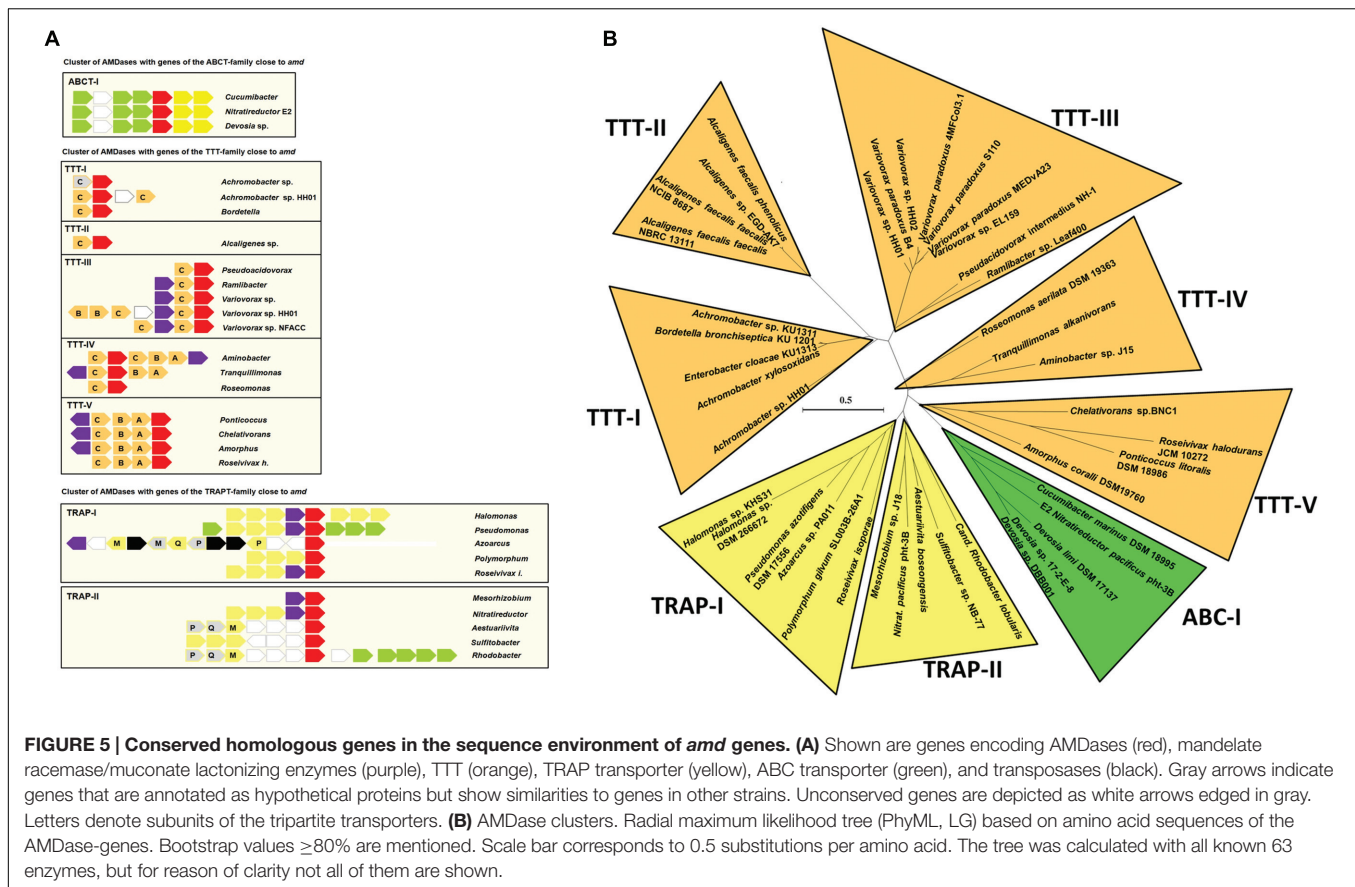
grow at temperatures up to 50°C (Nie et al., 2012). This makes AmdP a promising candidate for putative industrial applications. The enzymatic properties of the newly identified AmdA, AmdP, and AmdV were assayed using recombinant purified protein. A turnover of 10 mM of the substrate phenylmalonic acid was measured by HPLC. Concerning the temperature optimum, all three enzymes had their optimal activity in a comparable mesophilic range between 30 and 37°C (**Figure 6A**). AmdV showed maximal activity at 30°C and possessed  $44 \pm 3\%$  activity at 10°C. Above 30°C, however, activity rapidly decreased with  $6 \pm 2\%$  remaining at 50°C. The highest activity of AmdA from *Variovorax* sp. HH01 could be detected at 34°C. At 10°C, the activity decreased to 33% and above 40°C, activity decreased rapidly until almost no enzyme activity could be measured above 50°C. The highest conversion of the substrate by AmdP could be measured at 37°C. Below this temperature, the activity of this enzyme was lower compared to AmdA and AmdV, at 10°C even 2.6-fold lower than the activity of AmdV. Above 37°C, AmdP showed a better activity compared to the other two AMDases with 4.8-fold more substrate conversion at 45°C than AmdV.

The optimal pH conditions of the enzymes were assayed using phenylmalonic acid substrate solutions. All three enzymes are pH-dependent and showed the highest activity at a neutral to









**FIGURE 5 | Conserved homologous genes in the sequence environment of *amd* genes. (A)** Shown are genes encoding AMDases (red), mandelate racemase/muonate lactonizing enzymes (purple), TTT (orange), TRAP transporter (yellow), ABC transporter (green), and transposases (black). Gray arrows indicate genes that are annotated as hypothetical proteins but show similarities to genes in other strains. Unconserved genes are depicted as white arrows edged in gray. Letters denote subunits of the tripartite transporters. **(B)** AMDase clusters. Radial maximum likelihood tree (PhyML, LG) based on amino acid sequences of the AMDase-genes. Bootstrap values  $\geq 80\%$  are mentioned. Scale bar corresponds to 0.5 substitutions per amino acid. The tree was calculated with all known 63 enzymes, but for reason of clarity not all of them are shown.

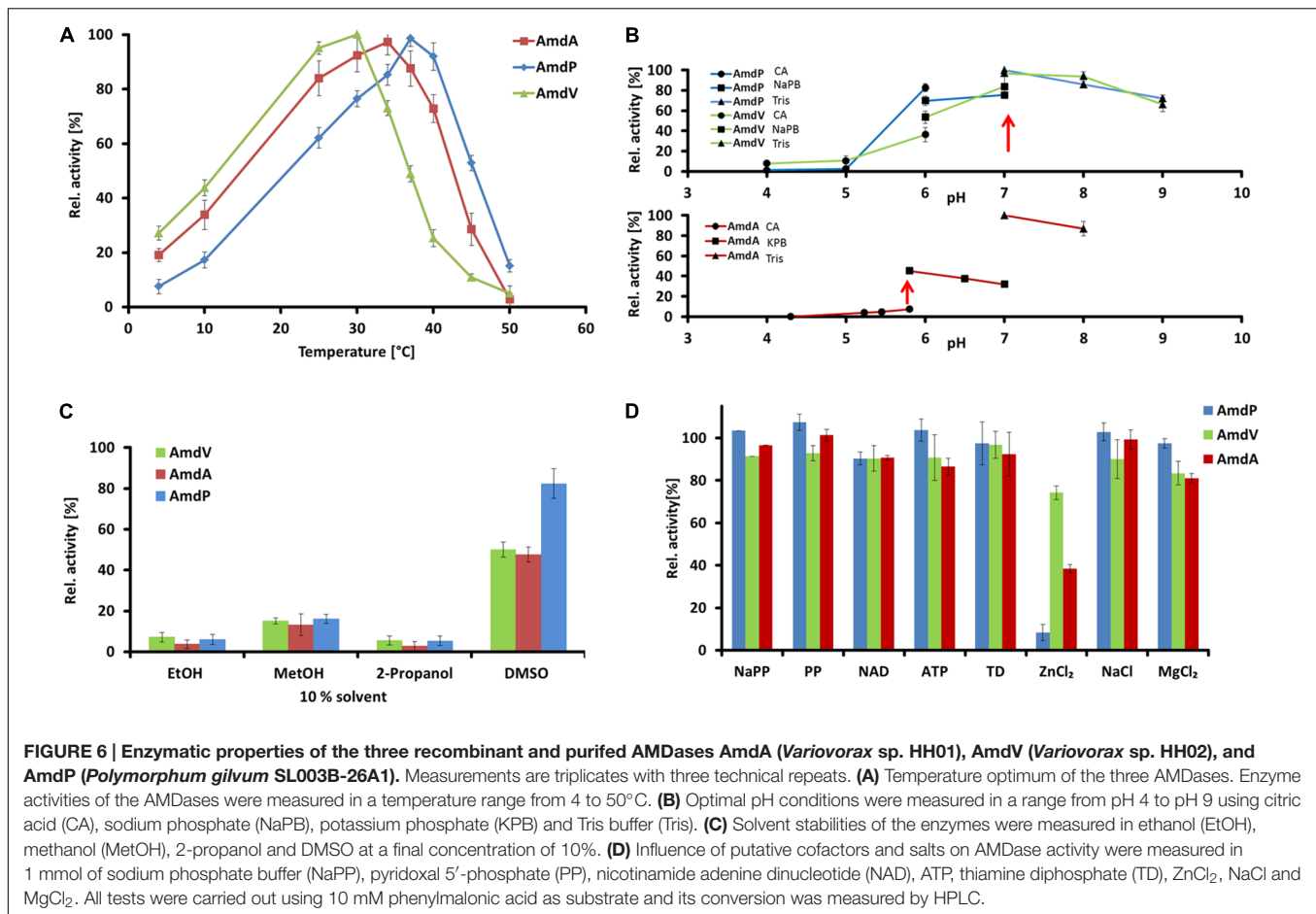
slightly acidic pH, mostly with Tris and sodium phosphate buffer (Figure 6B). AmdA showed its highest activity at pH 6.0, and AmdV as well as AmdP at a pH of 7.0 (Tris buffer). Below pH 4.0, no enzymatic activity could be detected, whereas at a pH of 8.0 and 9.0, 85% residual activity could be measured for AmdA and at pH 9.0, more than 65% activity remained in case of AmdV and AmdP. The preference for neutral or slightly basic conditions agrees well with that of the well-characterized AMDase from *B. bronchiseptica* (Miyamoto and Ohta, 1992a).

The solvent stabilities of the three AMDases were measured in 10% solutions (v/v) of ethanol, methanol, isopropanol, and DMSO diluted in 10 mM Tris buffer pH 7.0. AmdA, AmdV, and AmdP were similarly susceptible to the negative influences of the solvents (Figure 6C). Ethanol and isopropanol decreased the enzyme activities of the AMDases to less than 10% of the controls without solvent. With methanol, the substrate turnover was also strongly reduced to approx. 15% (AmdA  $13 \pm 5\%$ , AmdV  $15 \pm 2\%$ , and AmdP  $16 \pm 2\%$ ). DMSO reduced the enzymatic activity of AmdA and AmdV only to approx. half of the control ( $48 \pm 4$  and  $50 \pm 4\%$ , respectively) and AmdP still possessed  $82 \pm 7\%$  activity in this solvent.

Concerning the influence of salts and possible cofactors, none of the tested effectors sodium phosphate, pyridoxal phosphate, NAD, ATP, thiamine diphosphate, zinc chloride, sodium chloride, and magnesium chloride (all 1 mM in final concentration)

showed a beneficial effect on the enzymatic activity of AmdA, AmdV, and AmdP, proving that the enzymes function cofactor-independently (Figure 6D). Magnesium and especially zinc chloride even reduced enzyme activity significantly ( $\text{MgCl}_2$ :  $81 \pm 1\%$  residual activity for AmdA and  $83 \pm 5\%$  for AmdV.  $\text{ZnCl}_2$ :  $74 \pm 9\%$  for AmdV,  $38 \pm 5\%$  for AmdA and  $8 \pm 4\%$  for AmdP compared to the effector-less control).

The thermal stabilities and optimal storage conditions of the three AMDases were assayed at 22, 4, and  $-20^\circ\text{C}$  in 50 mM Tris buffer pH 7.0 and 50 mM sodium phosphate buffer pH 7.0, respectively, by measuring the enzymes' residual activities. In case of all three enzymes, storage in sodium phosphate buffer seemed to be beneficial compared to Tris buffer (Figure 7A). The residual activity of AmdA (*Variovorax* sp. HH01) after 7 days at  $22^\circ\text{C}$  in Tris buffer comprised only  $6 \pm 4\%$  while in sodium phosphate buffer,  $64 \pm 10\%$  activity were still measurable after 19 days at the same temperature. The best storage conditions for AmdA were at  $4^\circ\text{C}$  in sodium phosphate buffer with  $61 \pm 12\%$  residual activity after 24 days. AmdP from *P. gilvum* SL003B-26A1 was stable for more than 7 days with  $40 \pm 6\%$  residual activity after incubation at  $22^\circ\text{C}$  and with  $69 \pm 7\%$  at  $4^\circ\text{C}$  (Figure 7C). Although AmdV (*Variovorax* sp. HH02) lost most of its activity after storage at  $22^\circ\text{C}$  for 7 days ( $11 \pm 5\%$  residual activity), it was highly stable at  $4^\circ\text{C}$  and retained  $96 \pm 5\%$  of its initial activity after 11 days (Figure 7B). As shown in Figure 7D, for a long-term storage of the three enzymes, it is recommendable to choose



sodium phosphate buffer (50 mM and pH 7.0) and  $-20^{\circ}\text{C}$  as the AMDases keep more than 85% of their initial activity after at least 70 days (AmdA  $89 \pm 13\%$ ; AmdV  $85 \pm 9\%$ , and AmdP  $89 \pm 7\%$ ).

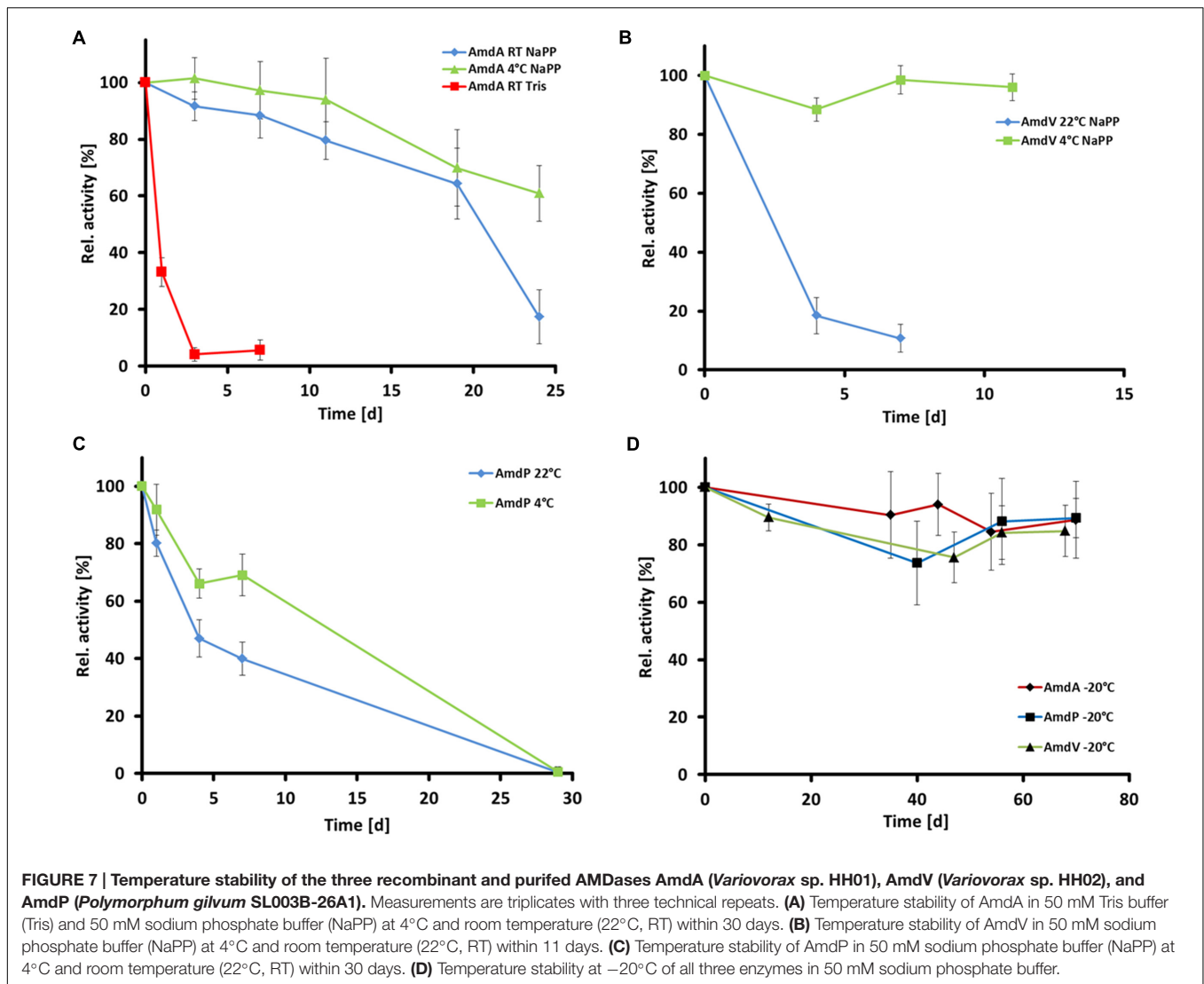
### Selectivity of AmdA, AmdV, and AmdP on Derivatives of 2-Arylpropionic Acids

The selectivity of all three AMDases was analyzed with aryl malonic acid derivatives of the arylpropionic acids naproxen and flurbiprofen. All enzymes showed a very high (*R*)-selectivity for both substrates as the obtained *ee*-values were greater than 99% indicating a huge excess of the (*R*)-enantiomer (Figures 1A and 8). In the case of flurbiprofen, spontaneous decarboxylation during the preparation of the aryl malonic starting material limited the *ee*. In the synthesis of naproxen, all three AMDases showed (*R*)-preference and a very high enantioselectivity ( $>99\%$  *ee*). In the case of flurbiprofen, the malonic acid is of limited stability. During the synthesis of the starting material, a small amount of racemic flurbiprofen ( $<4\%$ ) is formed that cannot be separated from the flurbiprofen malonate. As a consequence of this contamination, the enantiomeric excess of flurbiprofen is somewhat decreased. This effect is known and can be circumvented by the optimization of the reaction conditions. Nevertheless, the selectivity of

the three AMDases in the synthesis of flurbiprofen can be assumed to be very high. Terao et al. (2003) showed an *ee* of only 92% for (*R*)-flurbiprofen using the AMDase from *B. bronchiseptica*. This effect is known and can be circumvented by the optimization of the reaction conditions (Gaßmeyer et al., 2016). In the synthesis of naproxen, all three AMDases showed (*R*)-preference and a very high enantioselectivity ( $>99\%$  *ee*).

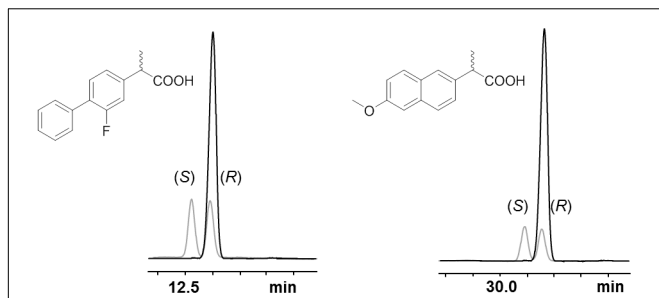
## DISCUSSION

Within this work, we have developed a novel search algorithm, allowing the identification of AMDase gene candidates from genomes and metagenomes. Altogether we were able to identify 58 novel possible enzyme candidates. A first classification of AMDases due to their protein sequence similarities and occurrence of conserved homologous genes in the sequence environment of *amd* enabled the formation of eight enzyme clusters within this work (Figure 5). Moreover, it should be noted that a large number of proteins have been automatically annotated as AMDases. However, as these proteins lack essential catalytic features like the hydrophobic binding pocket, an activity toward arylmalonates is thus highly unlikely. We confirmed this by cloning and testing the gene WP\_013776825 from

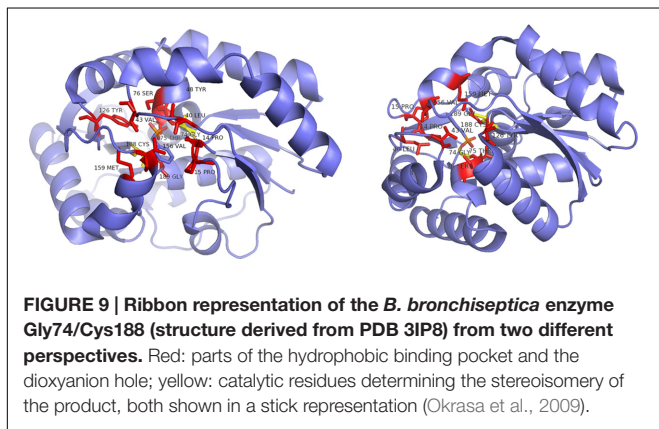


*Acidianus hospitalis*. Common features in the *amd* flanking sequences are the occurrence of genes encoding TRAP, ABC, and TTTs. Furthermore striking was the prevalence of a gene coding for MR/MLE, which was present in 61% of the available sequences. Five clusters comprise strains with genes coding for TTTs next to their *amd* gene (Figure 5B). Alpha-proteobacterial strains of the TTT clusters contain genes encoding all three subunits of TTTs. In contrast, all of the beta-proteobacterial representatives just carry “orphan” genes encoding periplasmic receptor proteins (Figure 5A). Genes coding for TctC are known to be overrepresented in *Beta-proteobacteria* (Antoine et al., 2003). Only a small amount is part of polycistronic operons. The majority of these so-called *bug* (*Bordetella* uptake gene) genes are “orphan” genes whose products most probably operate as extracytoplasmic substrate receptors (Antoine et al., 2003). There are suggestions that different receptor proteins interact with the same TctAB subunits (Antoine et al., 2003). Alternatively, the receptor protein BctC from *B. pertussis* appears to be involved in signal transduction

by interacting with the periplasmic domain of the sensor protein BctE (Antoine et al., 2005). The high number of TTT cluster’s representatives might indicate an evolutionary origin of AMDases within these strains or at least beneficial conditions that have been favored by these clusters. TRAP transporters are tripartite, ATP independent periplasmic transporters that have been found in bacteria and archae and enable the uptake of C<sub>4</sub>-carboxylates and further organic anions, e.g., lactate, glutamate, gluconate, ectoine, or benzoate derivatives by using an electrochemical gradient (Forward et al., 1997; Kelly and Thomas, 2001; Fischer et al., 2010; Mulligan et al., 2011). The uptake of 4-chlorobenzoate in *Comamonas* sp. DJ-12 is carried out by the TRAP transporter FcbT1-T3 (Chae and Zylstra, 2006). The corresponding genes are part of an inducible operon together with a gene encoding an enzyme that catalyzes the conversion of 4-chlorobenzoate to 4-hydroxybenzoate. An analogous correlation between genes of TRAP transporters and AMDases is conceivable. ABC transporter genes adjacent to *amd* genes are annotated as putative peptide/nickel transporters



**FIGURE 8 | Selectivity of AmdA, AmdV, and AmdP on derivatives of 2-arylpropionic acids.** Selected GC-chromatograms of (*R*)-flurbiprofen (**left**) and (*R*)-naproxen (**right**) after AMDase-catalyzed decarboxylation (here using AmdA) of the corresponding aryl malonic acids. Racemic arylpropionates are shown in gray. The measurement was done using a FS-Hydrodex- $\beta$ -6-TBDM column from Macherey-Nagel with an isothermic temperature of 160°C for naproxen and 170°C for flurbiprofen.



**FIGURE 9 | Ribbon representation of the *B. bronchiseptica* enzyme Gly74/Cys188 (structure derived from PDB 3IP8) from two different perspectives.** Red: parts of the hydrophobic binding pocket and the dioxanion hole; yellow: catalytic residues determining the stereoisomerism of the product, both shown in a stick representation (Okrasa et al., 2009).

(Navarro et al., 1993). Genes downstream of *amd* encode two ATP binding domains. Genes upstream of *amd* encoding a permease subunit and a substrate binding protein, are separated by an ORF annotated as putative oxidoreductase. The insertion of these two genes makes the functionality of the ABC transporters questionable although single genes are conserved. Because of the frequent occurrence of genes coding for various transporters next to the *amd* genes, it is likely that they are involved in substrate uptake.

The enzyme properties of the newly characterized AMDases from *Variovorax* sp. and *P. gilvum* fit well into the overall picture of known AMDases with an optimum pH from 5.5 to 8.5 and the highest substrate conversion at temperatures from 30 to 45°C (Miyamoto and Ohta, 1992a; Miyamoto et al., 2007, 2008; Okrasa et al., 2008). So far, all known wildtype AMDases have a strict selectivity for the formation of the (*R*)-enantiomer. This (*R*)-selectivity is based in the position of the catalytic Cys at position 188 in the active site (**Figure 9**). By moving this residue to the opposite side to position 74, the enantioselectivity of AMDase could be switched to an (*S*)-preference (Terao et al., 2007). None of the identified 58 putative AMDases, however, has a Cys in position 74, indicating that the natural enantioselectivity of AMDase is strictly for the (*R*)-enantiomer.

The selectivity seems to be highly conserved through all known AMDases. As the natural substrate of AMDase remains to be identified, and it is not even known if it is prochiral, the highly conserved enantioselectivity remains a puzzling feature of the decarboxylases.

All *amd* genes have a high GC content that is mostly higher than the average GC content of the corresponding genome (**Supplementary Figure S1**). Genomes with high GC content are characteristic for soil organisms with complex metabolic abilities (Foerster et al., 2005). The restriction to *Proteobacteria* could indicate a recent evolution within this group. The isolated occurrence of AMDases in most of the known genera with only one representative per species or strain implies that AMDases provide a slight competitive advantage for their carriers and might be lost in most bacteria.

The recent findings on AMDases described in this publication will significantly extend the knowledge of these enzymes and provide promising candidates for the production of enantiopure flurbiprofen. The eight novel enzyme clusters give new insights to possible reactions preceding decarboxylation (e.g., substrate uptake) and enzymatic reactions taking place subsequently to this step. Besides, the clustering shows the restriction of this conserved *Bordetella* prototype to some classes of *Proteobacteria*. Together with the newly established AMDase-specific search algorithm the screening for novel AMDases in data bases will be improved and will enable an effective use of the continuously growing sequence data for a reliable prediction of AMDase activity for the very first time.

The screening method is particularly useful for enzymes with a known structure and reaction mechanism. The generated HMM enables a specific search for AMDases of the *Bordetella* prototype and is beneficial for time-saving screenings of massive sequence data like metagenomic samples. Accordingly, the chance of finding biocatalysts with suitable properties can be increased dramatically. Furthermore, it is an appropriate method to find organisms as well as habitats where the desired enzymes occur and are more or less abundant. This allows a first insight into possible physiological and ecological functions and consequently also their natural substrate.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: JM, JC, RK, and WS. Performed the experiments: JM, SaG, SiG, and FB. Analyzed the data: JM, JC, SaG, RK, and WS. Contributed reagents/materials/analysis tools: RK and WS. Contributed to the writing of the manuscript: JM, JC, SaG, SiG, RK, and WS.

## ACKNOWLEDGMENTS

Research was in part funded by the Landesexzellenzinitiative Hamburg (C1-Chemistry in Resource and Energy Management) and the Deutsche Bundesstiftung Umwelt (DBU).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01332>

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Loch, C., et al. (2005). The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. *J. Mol. Biol.* 351, 799–809. doi: 10.1016/j.jmb.2005.05.071
- Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S., and Loch, C. (2003). Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *J. Bacteriol.* 185, 1470–1474. doi: 10.1128/JB.185.4.1470-1474.2003
- Bachmann, B. J. (1983). Linkage map of *Escherichia coli* K-12, edition 7. *Microbiol. Rev.* 47, 180–230.
- Bethesda Research Laboratories (1986). BRL pUC host: *E. coli* DH5 $\alpha$ <sup>TM</sup> competent cells. *Bethesda Res. Lab. Focus* 8, 9–12.
- Busch, F., Enoki, J., Hülsemann, N., Miyamoto, K., Bocola, M., and Kourist, R. (2016). Semiempirical QM/MM calculations reveal a step-wise proton transfer and an unusual thiolate pocket in the mechanism of the unique arylpropionate racemase AMDase G74C. *Catal. Sci. Technol.* 6, 4937–4944. doi: 10.1039/c5cy01964h
- Chae, J.-C., and Zylstra, G. J. (2006). 4-Chlorobenzoate uptake in *Comamonas* sp. strain DJ-12 is mediated by a tripartite ATP-Independent periplasmic transporter. *J. Bacteriol.* 188, 8407–8412.
- Fischer, M., Zhang, Q. Y., Hubbard, R. E., and Thomas, G. H. (2010). Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends Microbiol.* 18, 471–478. doi: 10.1016/j.tim.2010.06.009
- Foerster, K. U., Mering, C. V., Hooper, S. D., and Bork, P. (2005). Environments shape the nucleotide composition of genomes. *EMBO Rep.* 6, 1208–1213. doi: 10.1038/sj.embor.7400538
- Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R., and Kelly, D. J. (1997). TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J. Bacteriol.* 179, 5482–5493.
- Gaßmeyer, S. K., Wetzig, J., Mügge, C., Assmann, M., Enoki, J., Hilterhaus, L., et al. (2016). Arylmalonate decarboxylase-catalyzed asymmetric synthesis of both enantiomers of optically pure Flurbiprofen. *ChemCatChem* 8, 916–921. doi: 10.1002/cctc.201501205
- Gaßmeyer, S. K., Yoshikawa, H., Enoki, J., Hülsemann, N., Stoll, R., Miyamoto, K., et al. (2015). STD-NMR-based protein engineering of the unique Arylpropionate-Racemase AMDase G74C. *ChemBioChem* 16, 1943–1949. doi: 10.1002/cbic.201500253
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform Graphical User Interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580.
- Harrington, P. J., and Lodewijk, E. (1997). Twenty years of naproxen technology. *Organ. Process Res. Dev.* 1, 72–76. doi: 10.1021/op960009e
- Harrison, I. T., Lewis, B., Nelson, P., Rooks, W., Roszkowski, A., Tomolonis, A., et al. (1970). Nonsteroidal antiinflammatory agents. I. 6-substituted 2-naphthylacetic acids. *J. Med. Chem.* 13, 203–205.
- Hasson, M. S., Schlichting, I., Moulai, J., Taylor, K., Barrett, W., Kenyon, G. L., et al. (1998). Evolution of an enzyme active site: the structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10396–10401. doi: 10.1073/pnas.95.18.10396
- Ijima, Y., Matoishi, K., Terao, Y., Doi, N., Yanagawa, H., and Ohta, H. (2005). Inversion of enantioselectivity of asymmetric biocatalytic decarboxylation by site-directed mutagenesis based on the reaction mechanism. *Chem. Commun.* 21, 877–879. doi: 10.1039/b416398b
- Kelly, D. J., and Thomas, G. H. (2001). The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol. Rev.* 25, 405–424. doi: 10.1111/j.1574-6976.2001.tb00584.x
- Kourist, R., Maria, P. D. D., and Miyamoto, K. (2011a). Biocatalytic strategies for the asymmetric synthesis of profens-recent trends and developments. *Green Chem.* 13, 2607–2618. doi: 10.1039/c1gc15162b
- Kourist, R., Miyauchi, Y., Uemura, D., and Miyamoto, K. (2011b). Engineering the promiscuous racemase activity of an arylmalonate decarboxylase. *Chem. A Eur. J.* 17, 557–563. doi: 10.1002/chem.201001924
- Lewin, R., Goodall, M., Thompson, M. L., Leigh, J., Breuer, M., Baldeus, K., et al. (2015). Enzymatic enantioselective decarboxylative protonation of heteroaryl malonates. *Chem. A Eur. J.* 21, 6557–6563. doi: 10.1002/chem.201406014
- Lind, M. E. S., and Himo, F. (2014). Theoretical study of reaction mechanism and stereoselectivity of arylmalonate decarboxylase. *ACS Catal.* 4, 4153–4160. doi: 10.1021/cs5009738
- Miyamoto, K., and Ohta, H. (1992a). Purification and properties of a novel arylmalonate decarboxylase from *Alcaligenes bronchisepticus* KU 1201. *Eur. J. Biochem.* 210, 475–481. doi: 10.1111/j.1432-1033.1992.tb17445.x
- Miyamoto, K., Yatake, Y., and Ohta, H. (2008). Screening, cloning, expression, and purification of an acidic arylmalonate decarboxylase from *Enterobacter cloacae* KU1313. *Appl. Microbiol. Biotechnol.* 78, 793–799. doi: 10.1007/s00253-008-1375-8
- Miyamoto, K., Yatake, Y., Tamura, K., Terao, Y., and Ohta, H. (2007). Purification and characterization of arylmalonate decarboxylase from *Achromobacter* sp. KU1311. *J. Biosci. Bioeng.* 104, 263–267. doi: 10.1263/jbb.104.263
- Miyauchi, Y., Kourist, R., Uemura, D., and Miyamoto, K. (2011). Dramatically improved catalytic activity of an artificial (S)-selective arylmalonate decarboxylase by structure-guided evolution. *Chem. Commun.* 47, 7503–7505. doi: 10.1039/c1cc11953b
- Mulligan, C., Fischer, M., and Thomas, G. H. (2011). Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol. Rev.* 35, 68–86. doi: 10.1111/j.1574-6976.2010.00236.x
- Navarro, C., Wu, L.-F., and Mandrand-Berthelot, M.-A. (1993). The nik operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transporter system for Nickel. *Mol. Microbiol.* 9, 1181–1191. doi: 10.1111/j.1365-2958.1993.tb01247.x
- Nie, Y., Tang, Y.-Q., Li, Y., Chi, C.-Q., Cai, M., and Wu, X.-L. (2012). The genome sequence of *Polymorphum gilvum* SL003B-26A1T reveals its genetic basis for crude oil degradation and adaptation to the saline soil. *PLoS ONE* 7:e31261. doi: 10.1371/journal.pone.0031261
- Obata, R., and Nakasako, M. (2010). Structural basis for inverting the enantioselectivity of arylmalonate decarboxylase revealed by the structural analysis of the Gly74Cys/Cys188Ser mutant in the liganded form. *Biochemistry* 49, 1963–1969. doi: 10.1021/bi9015605
- Okrasa, K., Levy, C., Hauer, B., Baudendistel, N., Leys, D., and Micklefield, J. (2008). Structure and mechanism of an unusual malonate decarboxylase and related racemases. *Chem. A Eur. J.* 14, 6609–6613. doi: 10.1002/chem.200800918
- Okrasa, K., Levy, C., Wilding, M., Goodall, M., Baudendistel, N., Hauer, B., et al. (2009). Structure-guided directed evolution of alkenyl and arylmalonate decarboxylases. *Angew. Chem. Int. Ed.* 48, 7691–7694. doi: 10.1002/anie.200904112
- Schneider, E., and Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.* 22, 1–20. doi: 10.1111/j.1574-6976.1998.tb00358.x
- Terao, Y., Ijima, Y., Kakidani, H., and Ohta, H. (2003). Enzymatic synthesis of (R)-flurbiprofen. *Bull. Chem. Soc. Jpn.* 76, 2395–2397. doi: 10.1246/bcsj.76.2395

**FIGURE S1 | GC content of *amd* genes in comparison to the average genomic GC content.** The information on the GC contents was retrieved from the genomes listed in Supplementary Table S1 that have been deposited in the IMG database.

**TABLE S1 | List of putative *amd* genes found in this work.**

- Terao, Y., Ijima, Y., Miyamoto, K., and Ohta, H. (2007). Inversion of enantioselectivity of arylmalonate decarboxylase via site-directed mutation based on the proposed reaction mechanism. *J. Mol. Catal. BEnzym.* 45, 15–20. doi: 10.1016/j.molcatb.2006.11.002
- Terao, Y., Miyamoto, K., and Ohta, H. (2006). Improvement of the activity of arylmalonate decarboxylase by random mutagenesis. *Appl. Microbiol. Biotechnol.* 73, 647–653. doi: 10.1007/s00253-006-0518-z
- Yoshida, S., Enoki, J., Kourist, R., and Miyamoto, K. (2015). Engineered hydrophobic pocket of (S)-selective arylmalonate decarboxylase variant by simultaneous saturation mutagenesis to improve catalytic performance. *Biosci. Biotechnol. Biochem.* 79, 1965–1971. doi: 10.1080/09168451.2015.1060844

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Maimanakos, Chow, Gaßmeyer, Güllert, Busch, Kourist and Streit. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.