



# Aryl Coenzyme A Ligases, a Subfamily of the Adenylate-Forming Enzyme Superfamily

M. E. Arnold,<sup>a</sup> I. Kaplieva-Dudek,<sup>a</sup> I. Heker,<sup>a</sup> R. U. Meckenstock<sup>a</sup>

<sup>a</sup>Environmental Microbiology and Biotechnology (EMB), University of Duisburg-Essen, Essen, Germany

M. E. Arnold and I. Kaplieva-Dudek contributed equally to this work and are ranked by alphabetical order.

**ABSTRACT** Aryl coenzyme A (CoA) ligases belong to class I of the adenylate-forming enzyme superfamily (ANL superfamily). They catalyze the formation of thioester bonds between aromatic compounds and CoA and occur in nearly all forms of life. These ligases are involved in various metabolic pathways degrading benzene, toluene, ethylbenzene, and xylene (BTEX) or polycyclic aromatic hydrocarbons (PAHs). They are often necessary to produce the central intermediate benzoyl-CoA that occurs in various anaerobic pathways. The substrate specificity is very diverse between enzymes within the same class, while the dependency on Mg<sup>2+</sup>, ATP, and CoA as well as oxygen insensitivity are characteristics shared by the whole enzyme class. Some organisms employ the same aryl-CoA ligase when growing aerobically and anaerobically, while others induce different enzymes depending on the environmental conditions. Aryl-CoA ligases can be divided into two major groups, benzoate:CoA ligase-like enzymes and phenylacetate:CoA ligase-like enzymes. They are widely distributed between the phylogenetic clades of the ANL superfamily and show closer relationships within the subfamilies than to other aryl-CoA ligases. This, together with residual CoA ligase activity in various other enzymes of the ANL superfamily, leads to the conclusion that CoA ligases might be the ancestral proteins from which all other ANL superfamily enzymes developed.

**KEYWORDS** 2-naphthoate-CoA ligase, 2-phenanthrene-CoA ligase, ANL superfamily, BTEX degradation, PAH degradation, adenylate-forming enzymes, aryl-CoA ligases, benzoate-CoA ligase, hydrocarbon degradation, phenylacetate-CoA ligase

Thioester, amide, and ester bonds are very common in nature and are often part of typical chemical building blocks. Enzymes catalyzing the formation of these bonds occur in nearly all forms of life. Many of these enzymes belong to the superfamily of adenylate-forming enzymes, which were structured into a novel order of classes and subclasses by S. Schmelz and J. H. Naismith in 2009 (1). Class I of this superfamily is composed of three subclasses: subclass Ia comprises nonribosomal peptide synthetase (NRPS) adenylation domains, subclass Ib includes acyl and aryl coenzyme A (CoA) synthetases/ligases, and subclass Ic includes oxidoreductases (1).

Class I enzymes can serve various functions. For example, they are involved in fatty acid metabolism and transport (2, 3), cell signaling (4), biofilm formation (e.g., in *Candida albicans*) (5), the synthesis of antibiotic compounds (e.g., in *Streptomyces coelicolor*) (6), protein transport (2, 7, 8), and many others.

This review focuses on aryl-CoA ligases (ACLs). ACLs catalyze the formation of a thioester bond between aromatic compounds and CoA using ATP. Their general mechanism is equivalent to that of acyl-CoA ligases, and the first reaction is shared among most adenylate-forming enzymes. In this step, the negatively charged oxygen of the carboxylate substrate performs a nucleophilic attack on the more positive  $\alpha$ -phosphorous to form an aryl-adenylate intermediate and release pyrophosphate (PP<sub>i</sub>) (Fig. 1)

**Citation** Arnold ME, Kaplieva-Dudek I, Heker I, Meckenstock RU. 2021. Aryl coenzyme A ligases, a subfamily of the adenylate-forming enzyme superfamily. *Appl Environ Microbiol* 87:e00690-21. <https://doi.org/10.1128/AEM.00690-21>.

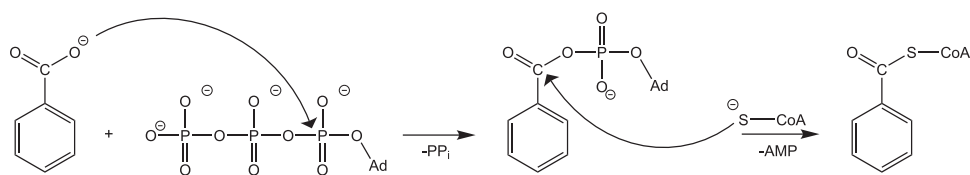
**Editor** Shuang-Jiang Liu, Chinese Academy of Sciences

**Copyright** © 2021 Arnold et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to R. U. Meckenstock, [rainer.meckenstock@uni-due.de](mailto:rainer.meckenstock@uni-due.de).

**Accepted manuscript posted online**  
14 July 2021

**Published** 26 August 2021



**FIG 1** Mechanism of benzoyl-CoA formation by benzoate-CoA ligase. In the first reaction, adenylation, the carboxylic acid of benzoate attacks the  $\alpha$ -phosphate of ATP to form a reactive benzoate-AMP intermediate and  $PP_i$ . In the second reaction, thioesterification, the benzoate-AMP intermediate reacts with the thiol group of the CoA and releases AMP by forming the CoA thioester. Ad, adenosyl group.

(for a comprehensive and detailed review of the mechanism, see reference 9; for a review about adenylate-forming enzymes of all classes, see reference 1).

In the second reaction, the intermediate is attacked by a nucleophile, in the case of the aryl-CoA ligases the thiol group of CoA, with AMP as the leaving group (Fig. 1) (1, 10, 11). For 4-chlorobenzoate:CoA ligase, it is proposed that AMP becomes activated as a leaving group by the interaction of its oxygen atoms with hydroxyl groups of two surrounding threonine residues of the enzyme (11).

CoA thioesters are formed by many organisms and serve various purposes. In many bacteria degrading aromatic or polycyclic aromatic compounds, such as, for example, *Rhodospseudomonas palustris*, the formation of a CoA thioester helps to accumulate substrates inside the cell (12). This accumulation comes to pass for two reasons. First, aromatic substrates entering the cell by diffusion through the membrane are converted to CoA thioesters, maintaining a downhill concentration gradient between the cytoplasm and the outside (12). Second, the CoA group prevents back-diffusion out of the cell due to its bulky size and polar structure (13).

Furthermore, the formation of the CoA thioester activates the substrate for further reactions or degradation steps because the CoA thioester group potentially draws electrons from the aromatic ring, facilitating, for example, the reduction of the ring system (14–16).

In some enterobacteria as well as eucaryotes, membrane-bound acyl-CoA ligases mediate long-chain fatty acid transport into the cell as well as activating the substrate by CoA ligation (3, 17, 18; for a comprehensive review on the role of class I adenylate-forming enzymes in the *trans*-membrane movement of long-chain fatty acids, see reference 2).

### ARYL-CoA LIGASES FROM ANAEROBICALLY AND AEROBICALLY GROWN BACTERIA

Aryl-CoA ligases have been found in a variety of bacteria that grow not only anaerobically but also aerobically with aromatic substrates. All these CoA ligases catalyze a comparable thioesterification reaction that depends on  $Mg^{2+}$ , ATP, and free CoA. All aryl-CoA ligases are oxygen insensitive and have an alkaline pH optimum of between 7.0 and 9.3 (Table 1).

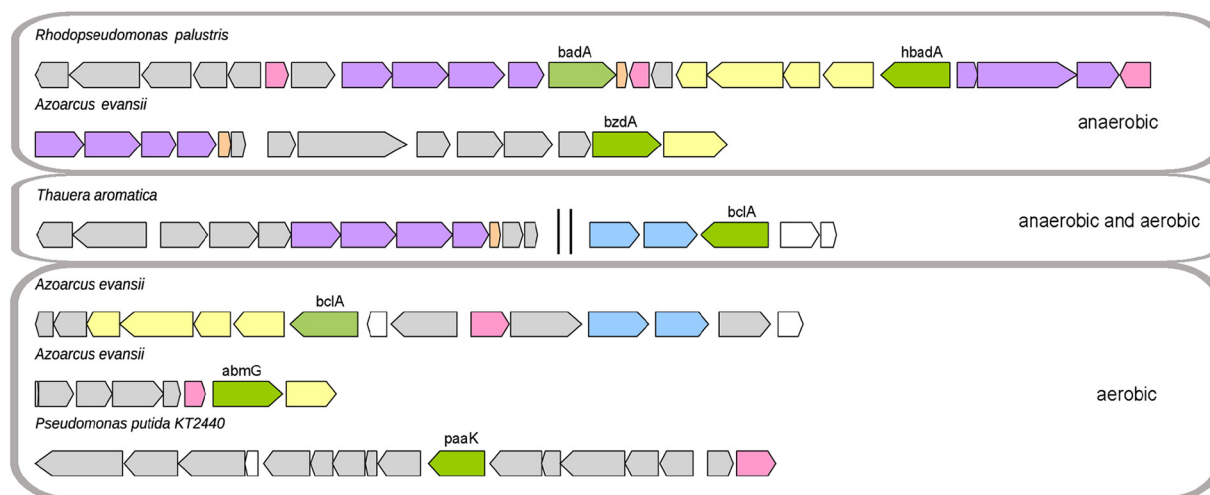
Anaerobic degradation of aromatic acids always involves the formation of the corresponding CoA thioester by an aryl-CoA ligase. CoA thioester formation facilitates the uptake of aromatic hydrocarbons (12, 13, 19) and is certainly crucial for further degradation (15, 16, 20, 21). Genes encoding these aryl-CoA ligases are mostly located adjacent to genes for enzymes involved in the anaerobic degradation pathway, such as reductases (Fig. 2) (22).

Bacteria grown aerobically with aromatic substrates usually use oxygenases to activate the aromatic compound through the formation of a hydroxylated intermediate that is oxygenolytically cleaved (23). In several bacteria, however, the aerobic metabolism of aromatic acids occurs via a novel type of aerobic degradation pathway (*box* pathway). These aromatic acids are activated via an aryl-CoA ligase, similar to anaerobic degradation, and are further hydrolytically cleaved (23). It is proposed that the produced CoA thioester acts as an inducer for the expression of the downstream enzymes in aerobic and anaerobic degradation (24).

TABLE 1 Characteristics of different aryl-CoA ligases

Aryl-CoA ligase	Microorganism	Growth condition	Native protein form	Subunit size (kDa) <sup>a</sup>	pH optimum	K <sub>m</sub> (μM) <sup>b</sup>	Substrate affinity	Reference
Benzoate:CoA ligase	<i>Rhodospseudomonas palustris</i>	Anaerobic	Monomer	60	8.4–8.9	0.6–2	Benzoate and 2- and 4-fluorobenzoate	44
	<i>Azoarcus evansii</i>	Anaerobic Aerobic	Homodimer Homodimer	120 56 (130)	9.3 9	11 11	Benzoate, 2-, 3-, and 4-fluorobenzoate, and 2-aminobenzoate	25, 26
	<i>Thauera aromatica</i>	Aerobic/anaerobic	Monomer	58	8.5	16	Benzoate, 2-, 3-, and 4-fluorobenzoate, and 2-aminobenzoate	34
	<i>Magnetospirillum</i> sp. strain TS-6	Aerobic/anaerobic	Homodimer	60 (120)	9	30	Benzoate and 2- and 4-fluorobenzoate	35
4-Chlorobenzoate:CoA ligase	<i>Pseudomonas</i> sp. strain CBS3	Aerobic	Homodimer	57	7.5	8.5 ± 0.09	4-Chlorobenzoate, benzoate, 4-bromobenzoate, 4-iodobenzoate, and 4-methylbenzoate	45
2-Aminobenzoate:CoA ligase	<i>Azoarcus evansii</i>	Aerobic	Monomer	65	8.5	10	2-Aminobenzoate and benzoate	29
	<i>Rhodospseudomonas palustris</i>	Anaerobic	Monomer	60	8.5–9.2	13	2-, 3-, and 4-fluorobenzoate	25
4-Hydroxybenzoate:CoA ligase	<i>Thauera aromatica</i>	Anaerobic	Dimer	61 (117)	>9	22	4-Hydroxybenzoate, benzoate, and cyclohex-1,4-dienecarboxylate	79
	<i>Pseudomonas putida</i>	Aerobic	Monomer	48	8.5	37	4-Hydroxybenzoate and 4-aminobenzoate	36
Phenylacetate:CoA ligase	<i>Azoarcus evansii</i>	Aerobic Anaerobic Aerobic	Monomer Monomer Monomer	48 ± 1 52 ± 2 52	7.0–8.5 8.5 8–8.5	16.5 60 14	Phenylacetate	39 31 32
	<i>Burkholderia cenocepacia</i> Paak1/Paak2	Aerobic	Homodimer	48 (82)	7.5	62 ± 4/150 ± 7		72
3-Hydroxybenzoate:CoA ligase	<i>Thauera aromatica</i>	Anaerobic	Monomer	60	≥9	60 ± 5	3- and 4-hydroxybenzoate	37

<sup>a</sup>Subunit sizes in parentheses denote the size of the full dimer.<sup>b</sup>K<sub>m</sub> values given here refer to the main substrate.



**FIG 2** Organization of gene clusters involved in the anaerobic or/and aerobic catabolism of benzoate, 4-hydroxybenzoate, 2-aminobenzoate, or phenylacetate in *Rhodopseudomonas palustris*, *Azoarcus evansii*, *Thauera aromatica*, or *Pseudomonas putida* KT2440, created using Gene Graphics (80). Genes are represented by tags: green, genes encoding CoA ligases (*badA*, *bzdA*, and *bclA*, benzoate:CoA ligase; *hbadA*, 4-hydroxybenzoate:CoA ligase; *abmG*, 2-aminobenzoate:CoA ligase; *paaK*, phenylacetate:CoA ligase); violet, genes encoding the subunits of the corresponding reductases for anaerobic degradation; orange, genes encoding ferredoxin; pink, regulatory genes; yellow, putative transport genes; gray, genes encoding enzymes for the downstream degradation pathway; white, genes of unknown function. Two vertical lines indicate that the genes are not adjacent in the genome.

There are a few bacteria known that are able to degrade certain aromatic acids such as benzoate, 2-aminobenzoate, and also phenylacetate aerobically and anaerobically using aryl-CoA ligases. This is considered an adaptation of these bacteria to changes between anoxic and oxic conditions (16). The denitrifying bacterium *Azoarcus evansii* is a great example of a bacterium that induces two distinct aryl-CoA ligases for several aromatic acids, depending on the availability of oxygen during growth. Benzoate:CoA ligases expressed in anaerobically or aerobically grown *Azoarcus evansii* cells are homodimers with different molecular masses (25, 26) and gene locations on the chromosome. One benzoate:CoA ligase, BzdA, is encoded in the cluster coding for enzymes involved in anaerobic degradation (*bzd* cluster), whereas the other one, BclA, is located within the cluster for enzymes required for aerobic degradation (*box* cluster) (Fig. 2) (27, 28). Two other monomeric aryl-CoA ligases were found in *A. evansii* growing with 2-aminobenzoate: one is induced under oxic conditions (29), and the other is found in anaerobically grown cultures (25). Two very similar clusters encoding enzymes for the aerobic degradation of 2-aminobenzoate, including the 2-aminobenzoate:CoA ligase, are present in the genome (30), while genes involved in anaerobic degradation are barely known. *A. evansii* grown anaerobically on 2-aminobenzoate also expresses a small amount of the aerobically induced 2-aminobenzoate:CoA ligase (25). The expression of both 2-aminobenzoate:CoA ligases during anaerobic growth indicates that *A. evansii* is able to switch rapidly from anaerobic to aerobic degradation of 2-aminobenzoate. Furthermore, this implies that the expression of both ligases is not exclusively under the control of the availability of oxygen and is thus regulated by other factors.

*A. evansii* also expresses two distinct aryl-CoA ligases in aerobic or anaerobic cultures grown with phenylacetate. Both phenylacetate:CoA ligases are monomers, have similar molecular masses and nearly the same pH optima, and are active only with phenylacetate. Furthermore, both ligases require glycerol for stability. Despite this similarity, both phenylacetate:CoA ligases have totally different N-terminal amino acid sequences and are encoded in the vicinity of enzymes required for aerobic or anaerobic degradation pathways, respectively (22, 31–33).

Based on the position of these aryl-CoA ligase genes in the genome of *A. evansii*, it might be that all these aryl-CoA ligases are expressed together with the enzymes required for the corresponding degradation pathway, as has been shown for the aerobically induced 2-aminobenzoate:CoA ligase (30).

In contrast to *A. evansii*, the denitrifying bacteria *Thauera aromatica* and *Magnetospirillum* sp. strain TS-6 expressed the same benzoate:CoA ligase (BclA) when the cultures were grown aerobically or anaerobically with benzoate (34, 35). The gene for benzoate:CoA ligase of *Thauera aromatica* was found as a part of the gene cluster for the novel aerobic benzoate oxidation pathway (*box* pathway) and not in the gene cluster for anaerobic benzoyl-CoA metabolism (Fig. 2) (34). In contrast, the *bclA* gene of *Magnetospirillum* sp. is located in neither the *bzd* cluster nor the *box* cluster (35). This indicates that the benzoate:CoA ligase is expressed independently from enzymes involved in the corresponding degradation pathway. Thus, the presence or absence of oxygen induces the expression of the respective downstream enzymes (34, 35). This opportunity may enable these bacteria to survive in environments where the oxygen conditions are changing rapidly and they need to switch between anaerobic and aerobic degradation pathways.

### SUBSTRATE SPECIFICITY

Even though all aryl-CoA ligases utilize the same reaction mechanism, substrate specificities differ widely (Table 1). Eight different aryl-CoA ligases have been found in the denitrifying bacterium *A. evansii* alone: two benzoate:CoA ligases, two 2-aminobenzoate:CoA ligases, two phenylacetate:CoA ligases, a 3-hydroxybenzoate:CoA ligase, and a 4-hydroxybenzoate:CoA ligase (25, 26, 29, 31, 32). Each of these aryl-CoA ligases is encoded in the vicinity of its corresponding downstream genes. *A. evansii* grown anaerobically in a batch culture with 2-aminobenzoate showed the presence of 2-aminobenzoate:CoA ligase, benzoate:CoA ligase, and a small amount of the 2-aminobenzoate:CoA ligase usually produced during aerobic growth. All three aryl-CoA ligases prefer the same substrates (25). Even benzoate:CoA ligase induced under oxic conditions with benzoate as the substrate showed similar substrate affinities (26). This suggests a regulatory function of these ligases in the downstream degradation pathway and, due to the similar substrate specificities, may also allow switching easily from one degradation pathway to another. Interestingly, *A. evansii* expresses four enzymes with nearly the same substrate specificity and not only one as in the denitrifying bacterium *T. aromatica*. The benzoate:CoA ligase induced during aerobic and anaerobic growth on benzoate or 2-aminobenzoate in *T. aromatica* has a broad substrate specificity (34). This may enable *T. aromatica* to alternate rapidly between different degradation pathways depending on the occurrence of the substrates. Two further aryl-CoA ligases were identified in *T. aromatica*, the distinct monomeric aryl-CoA ligases for the isomers 3-hydroxybenzoate and 4-hydroxybenzoate (36, 37). These two aryl-CoA ligases have a very narrow substrate specificity compared to the benzoate:CoA ligase. This suggests that 3-hydroxybenzoate:CoA ligase and 4-hydroxybenzoate:CoA ligase, in contrast to the benzoate:CoA ligase, are involved in the regulation of the downstream genes and are able to activate only their corresponding degradation pathway.

The phototrophic bacterium *R. palustris* grown with 4-hydroxybenzoate in the absence of oxygen showed 4-hydroxybenzoate:CoA ligase activity. In contrast to the high substrate specificity of the 4-hydroxybenzoate:CoA ligase of *T. aromatica*, this ligase has a broad substrate affinity. The gene for 4-hydroxybenzoate:CoA ligase (*hbadA*) is located adjacent to the gene clusters coding for enzymes involved in 4-hydroxybenzoate degradation as well as enzymes required for benzoate degradation (Fig. 2) (38). This high substrate diversity may enable *R. palustris* to also regulate the expression of enzymes required for the downstream degradation of other aromatic acids such as benzoate.

A highly substrate-specific aryl-CoA ligase is the phenylacetate:CoA ligase induced in *A. evansii* and *Pseudomonas putida* when grown with phenylacetate. The aerobically and anaerobically induced phenylacetate:CoA ligases from *A. evansii* and the aerobically induced one from *P. putida* could convert only phenylacetate to phenylacetyl-CoA. No other substrates were converted to their corresponding CoA thioesters (31, 32, 39), indicating a potential regulatory function in the expression of enzymes exclusively involved in the downstream degradation of phenylacetate.

This shows that the substrate specificity of monoaromatic aryl-CoA ligases differs from ligase to ligase, even if they are expressed during growth with the same substrate; this may indicate different regulatory functions.

The substrate specificity of polycyclic aromatic aryl-CoA ligases is under investigation. The polycyclic aromatic aryl-CoA ligases expressed in the naphthalene-degrading, sulfate-reducing culture N47 (40) or NaphS2 (41) showed broad substrate specificity. They can convert fluorinated naphthoates to their corresponding CoA thioester at the same rate as naphthoate, similar to what is known for all benzoate:CoA ligases. Several hydroxylated naphthoates were also used by the naphthoate:CoA ligase but with lower conversion rates (M. E. Arnold, F. Kaschani, and R. U. Meckenstock, unpublished results). In the sulfate-reducing, phenanthrene-degrading enrichment culture TRIP1 (42, 43), the phenanthroate:CoA ligase is very specific for 2-phenanthroate. Only 3-phenanthroate was utilized by the phenanthroate:CoA ligase as well but with a much lower conversion rate (I. Kaplieva-Dudek, F. Kaschani, and R. U. Meckenstock, unpublished results).

### INFLUENCE OF CATIONS AND THIOL GROUP-MODIFYING AGENTS ON ARYL-CoA LIGASE ACTIVITY

All known adenylate-forming enzymes require magnesium ions for their catalytic reaction. The absence or replacement of magnesium by other cations leads to a loss of enzyme activity. For example, the addition of the known inhibitor of magnesium-dependent enzymes NaF to phenylacetate:CoA ligase of *A. Evansii* decreased the ligase activity by up to 20% (31). The 4-chlorobenzoate:CoA ligase of *Pseudomonas* sp. strain CBS3 has 12-fold-lower activity without  $Mg^{2+}$  (11), indicating that magnesium is necessary for optimal enzyme activity. Magnesium ions neutralize the charge of ATP and later the charge of the leaving group pyrophosphate. Furthermore, magnesium ions can interact with and stabilize the acyl-AMP intermediate. In nearly all aryl-CoA ligases,  $Mg^{2+}$  can be replaced only by  $Mn^{2+}$  without loss of activity (25, 29, 31, 39, 44). For the 4-chlorobenzoate:CoA ligase from *Pseudomonas* sp. strain CBS-31, it was also shown that  $Mg^{2+}$  can be replaced by  $Co^{2+}$  as a cofactor (45). In most cases, other cations such as  $Zn^{2+}$ ,  $Cu^{2+}$ ,  $Ni^{2+}$ ,  $Hg^{2+}$ , and  $Mo^{5+}$  have strong inhibitory effects on ligase activity. Copper ions can probably interact with the thiol group and thus prevent the CoA from binding to the substrate. Monovalent cations like  $K^+$ ,  $Na^+$ ,  $Li^+$ , and  $Rb^+$  have no effect on aryl-CoA ligase activity (25, 29, 31, 39, 44). In contrast to monocyclic aromatic aryl-CoA ligases, polycyclic aromatic aryl-CoA ligases were also active with either  $Na^+$  or  $K^+$  instead of  $Mg^{2+}$  when using cell extracts (Arnold et al., unpublished; Kaplieva-Dudek et al., unpublished). However, it cannot be excluded that a sufficient amount of  $Mg^{2+}$  is present in the ligase assays since a cell extract was used and not purified enzymes. So far, similar findings have also been reported for xenobiotic/medium-chain fatty acid:CoA ligases of bovine liver mitochondria and cinnamate:CoA ligase of *Hypericum calycinum* cell cultures (46–48). The phenanthroate:CoA ligase from the sulfate-reducing, phenanthrene-degrading culture TRIP1 showed higher ligase activity when  $Na^+$  or  $K^+$  was added to cell extracts instead of  $Mg^{2+}$ . With a combination of  $Mg^{2+}$  and  $K^+$ , the greatest conversion of 2-phenanthroate was measured. Furthermore, phenanthroate:CoA ligase was inhibited with increasing  $Mg^{2+}$  concentrations (Kaplieva-Dudek et al., unpublished). This indicates a difference between aryl-CoA ligases converting monocyclic aromatic and polycyclic aromatic hydrocarbon (PAH)-carboxylic acids regarding the influence of cations on ligase activity. Other strong inhibitory effects on monocyclic aromatic aryl-CoA ligases were detected by thiol group-modifying agents such as *N*-ethylmaleimide, iodoacetamide, *p*-hydroxymercuribenzoate, 5,5'-dithiobis-(2-nitrobenzoic acid), and *p*-chloromercuribenzoic acid, which indicates that a free thiol group of the CoA is necessary for enzyme activity (25, 32, 36).

### ARYL-CoA LIGASES FORM TWO MONOPHYLETIC CLADES

A neighbor-joining phylogenetic tree of amino acid sequences of the adenylate-forming enzyme superfamily (ANL superfamily) was recently constructed by L. Clark et al.

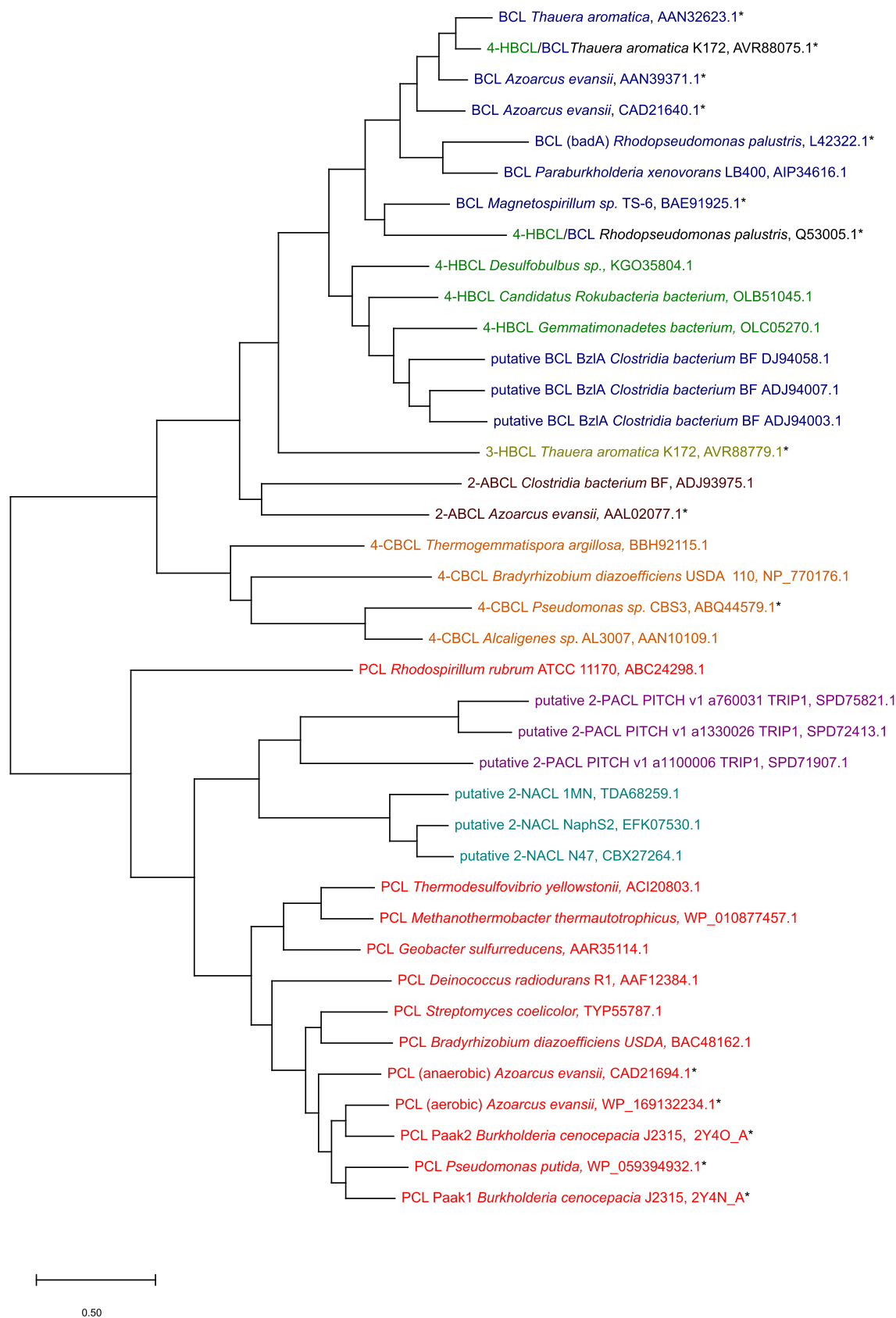
(49), aligning in total 374 protein sequences of the ANL superfamily, of which 49 belonged to aryl-CoA ligases. The unrooted tree classified the proteins into nine distinct groups: luciferases, three groups of fatty acyl-CoA synthetases, nonribosomal peptide synthetases (NRPSs), fatty acid-AMP ligases, methylmalonyl-CoA synthetases, mycobacterial FadD10 long-chain fatty acyl-CoA ligases, and aryl-CoA ligases. The alignment of the sequences yielded 5 amino acids conserved in all groups: Glu328, Gly384, Asp418, Arg433, and Lys524 (residue positions correspond to the *Thermus thermophilus* sequence as noted in this paper). Conserved residues were mostly found surrounding the active site of the enzymes containing the AMP- and CoA-binding site rather than the protein surface. Group-specific residues were identified using the GEnt program (50), calculating the conservation of specific residues inside the aryl-CoA group of proteins versus the whole alignment. Two of the eight identified residues interacted directly with the substrate: Asn411 of the 4-chlorobenzoate:CoA ligase from *Alcaligenes* forms a hydrogen bond to the  $\alpha$ -phosphate of AMP in the thioesterification conformation, and His207 interacts with the substrate during the adenylation reaction.

Recently, Robinson et al. (51) used a machine-learning approach to identify enzymes of the ANL superfamily and construct a maximum likelihood phylogenetic tree of the characterized proteins. This tree shows that  $\beta$ -lactone synthetases, NRPSs, as well as luciferases form very distinct monophyletic clades. In contrast to the study by L. Clark et al. (49), especially the aryl/acyl-CoA ligases are distributed widely throughout the phylogenetic clades and show a closer relationship to the enzyme subfamilies than to each other. Hence, aryl/acyl-CoA ligases seem to be clearly polyphyletic, and those authors conclude that the ANL superfamily developed from enzymes with an active site comparable to that of modern aryl/acyl-CoA ligases using CoA-SH as a cofactor. In this case, divergent evolution from ancestral CoA ligase-like scaffolds led to the diverse specialized functions of the ANL superfamily. A similar radial topology and common ancestor have been shown previously for the nitroreductase superfamily, where evidence suggests divergent evolution from an ancestral minimal flavin-binding scaffold (52). This line of reasoning is supported by experiments showing low-level CoA-ligase activity in members of the ANL superfamily, specialized primarily toward other functions. Five NRPS A domains were bifunctional, showing CoA ligase activity (53), similar to luciferases (54). Moreover, the fact that single point mutations were able to restore CoA ligase activity in ANL superfamily members strengthens the hypothesis of an ancestral CoA ligase-like scaffold (51). Nonetheless, the alternative of CoA ligase activity developing independently at several times throughout ANL superfamily evolution, as shown with other superfamilies, cannot be conclusively excluded at this point.

We constructed a maximum likelihood phylogenetic tree with Mega-X (55, 56) using 39 amino acid sequences of the aryl-CoA enzyme family (Fig. 3). Here, the group of the aryl-CoA ligases was divided into two phylogenetically distinct subgroups: benzoate:CoA ligase-like enzymes and phenylacetate:CoA ligase-like enzymes. In this tree, not only do benzoate:CoA ligases and phenylacetate:CoA ligases cluster into separate clades, but putative polycyclic aromatic compound ligases for naphthoate (57, 58) and phenanthroate (43) are also part of the phenylacetate:CoA ligase-like monophyletic clade. The only relatively conserved residues of the aryl-CoA ligases, leading to clustering according to their substrate spectrum inside these two major phylogenetic clades, can be found surrounding the substrate-binding pocket (49).

### STRUCTURAL FEATURES AND CONSERVED RESIDUES OF ARYL-CoA LIGASES

Overall, benzoate:CoA ligase-like and phenylacetate:CoA ligase-like enzymes conform to the general two-domain structure of ANL enzymes, with a larger N-terminal domain with parallel  $\beta$ -sheets forming the core surrounded by  $\alpha$ -helices and the active site situated in the linker region between the N- and C-terminal domains. A 140° rotation of the C-terminal domain enables the enzyme to catalyze the two distinct reactions of adenylation and thioesterification (9, 59). During the first reaction, the enzyme



**FIG 3** Maximum likelihood phylogenetic tree of aryl-CoA ligases of the ANL superfamily. The tree was constructed in Mega-X using 39 amino acid sequences of the aryl-CoA enzyme family (see GenBank accession numbers at the right). Enzymes (Continued on next page)



adopts the adenylation conformation when ATP and the carboxylate bind to the active site in the N-terminal region. This configuration places the carboxylate in the ideal position to attack the  $\alpha$ -phosphate, which leads to the release of PP<sub>i</sub> and the formation of the adenylate intermediate (60). For the second reaction, a rotation of the C-terminal domain is necessary because the pantetheine tunnel, which is part of the CoA-binding site, is blocked in the adenylation conformation. The pantetheine or thiol group of CoA needs to be close to the intermediate to form the thioester bond and displace the AMP. During the C-terminal domain rotation and the subsequent formation of the thiolation conformation, the pantetheine tunnel is open, and CoA can be placed correctly into the binding pocket, leading to the formation of the aryl-CoA product (61).

The study of the adenylation domain of NRPSs yielded 10 conserved motifs, 9 of which can be found in the wider ANL superfamily (A1 to A8 and A10) (9, 62). Three of these motifs have been previously described (63). A recent publication by L. Clark et al. (49) compared 374 amino acid sequences of the ANL family and found 10 conserved motifs, corresponding partially to the previously established motifs. A comparison of these different motifs and their function can be found in Table 2.

Crystal structures of the 4-chlorobenzoate:CoA ligase elucidated the role of conserved amino acids in these motifs (64, 65). One of the most conserved regions found in all enzymes of this superfamily is the ATP-binding site. Motif A7, containing a conserved aspartic acid residue, as well as motif A8, containing an arginine residue, interact with ribose hydroxyls. The ATP molecule is further stabilized via the A5 motif. In the 4-chlorobenzoate:CoA ligase, an aromatic tyrosine is the key residue for this interaction. The key catalytic residue lysine in motif A10 interacts in the adenylation conformation of the enzyme with the  $\alpha$ -phosphate during the nucleophilic attack on ATP. This lysine is highly conserved in all aryl-CoA ligases and throughout the acyl-CoA and fatty acid acyl-CoA ligases. The lysine residue is the target of acetylation as a posttranslational regulation mechanism (66, 67). Reversible lysine acylation in prokaryotes is a relatively new field of study. The earliest mention of this regulatory mechanism in bacteria was in 2002 (68). The benzoate:CoA ligase BadA from *Rhodospseudomonas palustris* is one example of regulation through the interaction of acetyltransferases and deacetylases (66). Another, possibly related regulatory mechanism was found in *Azoarcus* sp. strain CIB. Here, the genes for benzoate degradation are located in one operon. Gene expression is regulated by the repressor BzdR, which is dependent on the benzoyl-CoA concentration in the cell (69). Posttranslational lysine acetylation might be a possible way to regulate not just ligase activity but also the expression of aryl-CoA ligase genes through the benzoyl-CoA concentration. This might provide a diverse toolset for bacteria to regulate an energy-consuming reaction.

The position of the CoA-binding site is not well conserved within the group of aryl-CoA ligases, but it consists of a nucleotide-binding site on the protein surface and a pantetheine tunnel interacting with motif A8. The substrate-binding site within this group shows almost no conservation and is highly diverse even between enzymes with the same main substrate (9). The alignment created for the phylogenetic tree (Fig. 3) has been searched for the motifs described here (Table 2). The sequence alignments containing at least 50% sequence identity with these motifs are marked in Fig. 4.

C. K. Thornburg et al. (59) elucidated the structure of the benzoate:CoA ligase BadA from *R. palustris* (Fig. 5) and determined the reaction mechanism and wild-type substrate specificity through mutations of the active center. BadA shows a preference for *ortho*-substituted substrates in contrast to *meta* or *para* isomers. Steric effects influence kinetics more than electrostatic interactions. The enzyme shows a two-domain structure with the N-terminal domain containing the benzoate-binding site, typical for the ANL superfamily. The substrate is centered in the active site with the help of charged

### FIG 3 Legend (Continued)

marked with an asterisk are listed in Table 1. BCL, benzoate:CoA ligase; 4-HBCL, 4-hydroxybenzoate:CoA ligase; 3/4-CBCL, 3/4-chlorobenzoate:CoA ligase; 2-ABCL, 2-aminobenzoate:CoA ligase; 2-PACL, 2-phenanthroate:CoA ligase; 2-NACL, 2-naphthoate:CoA ligase; PCL, phenylacetate:CoA ligase.

**TABLE 2** Conserved sequence motifs in the ANL superfamily<sup>a</sup>

Motif described by Clark et al. <sup>b</sup>	Motif described by Marahiel et al., <sup>c</sup> A1–A10, and Chang et al., <sup>d</sup> I–III	Amino acid sequence <sup>e</sup>		A1–A10 (I–III)	Function(s) and result(s) of mutation <sup>f</sup>	Motif found in aryl-CoA ligases <sup>g</sup>
		1–10				
6	A1	(R/K)LANAixxxLG(V/I/L)(K/P)GD(R/V)(A/G)(L/V/I/L)		Ψ(S/T)Ωx(E/Q)Ψ	Structural	
10	A2	(A/I)GAV(I)VVVP(L/I)NPRx(Y/L)TPK(E/D)(I/L)xYR(L/I)N		(R/K/F)ΨGΨ	Structural	
3	A3 (I)	(T/S)SG(T/S)ITGLPKGV(M/L)(L/H)(T/S)H		ΨΨx(S/T)(S/T/G) G(S/T)TGxPK	P-loop Thr161, 2,000-fold-lower activity Gly163, 1,000-fold-lower activity Gly166, 14-fold-lower activity Pro168, no activity Lys169, 4-fold-lower activity	X (Gly189)
	A4			Ω	Aromatic residue, part of the acyl-binding pocket in the active center His207, 500-fold-lower activity	
9		(I/L)(E/Q)(K/Y/E)(K/R)(V/I)Tx(L/F)xG(V/A)PTIYR(F/A)L(L/A)(K/Q)				X (Pro278)
7	A5 (II)	DLSSL(R/K)KLV(S/G)(A/G)(A/E)(P/A)LN(P/K)EV(L)xE		Ω(G/W)x(A/T/E)	Positioning of ATP and binding of Mg <sup>2+</sup> Tyr304, no change in activity Thr307, 100-fold-lower activity Glu306, 50-fold-lower activity	X X
8		ExKPGSVG(K/R)(P/V)VP(G/N)(E/D)(K/R)(I/V/L)(V/I)DP				
5	A6	GE(I/L)C(V/I)(R/K)x <sub>5</sub> GPG(V/I/L)(M/F/A)KGY(W/Y/L)N		GE <sub>X<sub>10-14</sub></sub> G (S/T)GD	Structural	
1	A7 (III)	(Y/L/F)H(T/S)GD(L/I)(G/A)(Y/R)xDEDGY(F/L)(W/F)(W/T) (G/D)Rx(K/D)(L/M)(K/I)(S(G/K/S)(Y/E/F)(R/N/Q)(I/W) GPAE(I/V/L)ESAL		Rx(D/K)x <sub>6</sub> G	Positioning of ATP, binding/interacting with the ribose hydroxyls of ATP Asp385, 500-fold-lower activity Arg400, 100-fold-lower activity	X (Arg424)
4		HPA(V/I)A(E/D)AAV(V/I)G(V/I)P(D/H)(P/E)x(W/A/R)G(E/Q)(V/P/V)				
2	A10	P(R/D)x(V/I)(E/V)FUDE(L/I)PK(T/N)(P/A)(S/T)GK(L/D)(R/K) (R/K)ELR		Px <sub>3</sub> GKΨx(R/K)		X (Lys518)

<sup>a</sup>A1–A10, conserved motifs of the adenylation domain of NRPS (62); I–III, conserved sequences of the acyl ligase family (63); 1–10, conserved motifs of the ANL superfamily (49); P-loop, phosphate-binding loop.

<sup>b</sup>See reference 49.

<sup>c</sup>See reference 62.

<sup>d</sup>See reference 63.

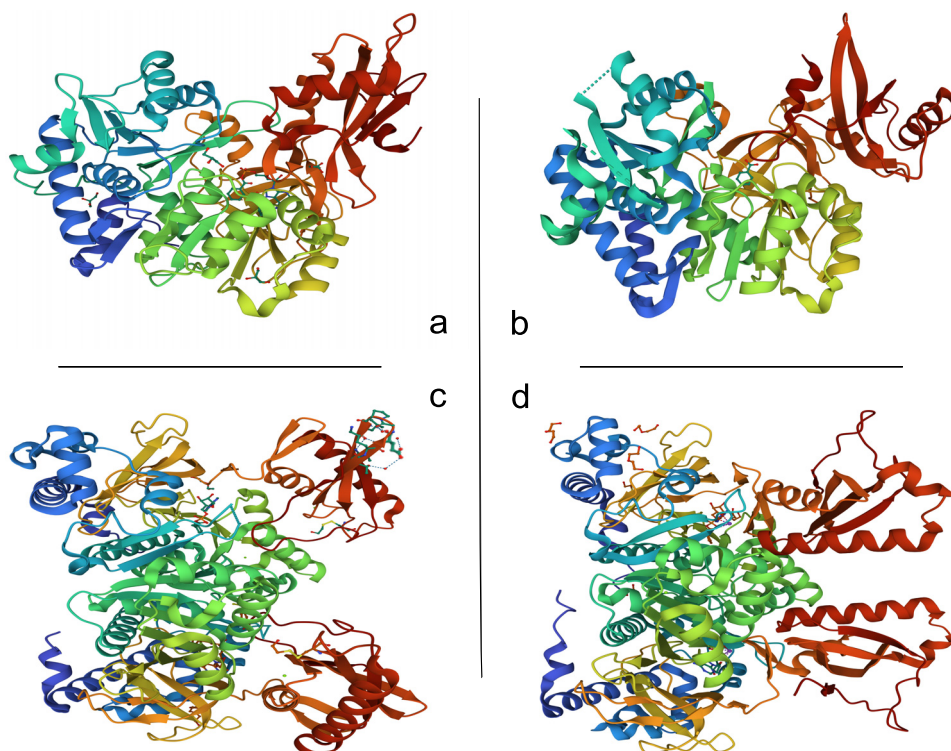
<sup>e</sup>Ω, aromatic amino acid (F, Y, H, or W); Ψ, aliphatic amino acid (A, V, L, I, or M); x, any amino acid.

<sup>f</sup>See references 9, 11, and 63.

<sup>g</sup>Amino acids used in the phylogenetic analysis of aryl-CoA ligases (Fig. 3) were aligned using the MUSCLE algorithm (Mega-X) and examined for conserved sequences. The 100% conserved amino acids and their positions in the benzoate:CoA ligase of *Thauera aromatica* (GenBank accession number AAN32623.1) are written in three-letter code. A motif was considered present if 50% of the amino acid code was concurrent with the described motifs.

1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	-----MS SVKFLFTEEN VGRKHGVYDP EEQ--DRTK PYSEKFMKSV --ETLPIERI REIQMER-FR NIVQFAYARS	[ 80]
2. putative_2-NACL_N47_CBX27264.1	-----MRL---VTQK ATESPYMNVE M-ETMPREKL DQLHLRR-LQ RLKIYAYENI	[ 80]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	-----MPVKT PPSGDL-EP I-EKASQDEL RALQLER-LK WSVRHAYENV	[ 80]
4. PCL Pseudomonas putida WP_059394932.1*	-----MNMV HADRALLDP M-ETASVDAL RQHLER-LR WSLKHAYDNN	[ 80]
5. BCL Thauera aromatica AAN32623.1*	-----MYTSL VADSNHTPPA IKI--PERY NAADDLGRN L-LAGRGGKT VYDDAG-S YTYDELALRV	[ 80]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	-----M TSPNASSEHA LDI--PRDY NAATVFDLRH L-KDGRADKV AFIDADG--S HTYQGLADKV	[ 80]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	-----MPTLS AADHTASPEE IRI--PRHY NAADDLGRN L-DAGRGPKI AYIDDEG--R YSFDELARLV	[ 80]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	-----MP--LSDY NAAVDFVDRN V-AEGRGKI AFIDPQR--S LSYGELRADV	[ 80]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	-----M QTVHEMLRRA V--SRVPHRW AIVDAAR-ST FDICRTGETS	[ 80]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	-----MSEQ LQP--QQSM NAAEIIIGRP L-AQGLGEQT AMLCAER--S ITYRELDAAE	[ 80]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	-----MTSHVDTF ADRLRPPPEQ QPEFLFELPS LQF--PPLM NCAVELLDRR V-LGEGGERV CLRAPGG-LR WTYRDLGLHA	[ 80]
<b>A3/3</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	IRHPVITVTK DFG----- --DDQKENPPY GTAFTSPNPT QLKYM----- --QTSGTTA KPRLWETKE	[240]
2. putative_2-NACL_N47_CBX27264.1	WEKPSIDPK DIL----- --RYQAANPPF GGSIVRDSDE YLTF----- --FQTSGTTG TPMKEZGYVR	[240]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	LAKFPPTAGK DL----- --RDNYPF G-MFAVPRE KVARV----- --HASSGTTG KPTVVGTYLK	[240]
4. PCL Pseudomonas putida WP_059394932.1*	LAKFPPTGKN DL----- --RDNYPF G-MFAVQPE EVVRL----- --HASSGTTG KPTVVGTYLN	[240]
5. BCL Thauera aromatica AAN32623.1*	LGVKPTLEHL VVA-G-GA--GEDSLA ALLATGSEQE E--AAATPRD DFCF----- --LYSSGSGT APKGTVHHS	[240]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	LDPMLPEHV VIS-GKNAH--GHTLSD DLLAKAEPLK K--TAETTRD DVAHF----- --LYSSGSGT APKGVAHQR	[240]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	LGVKPTLERI IVA-GDAAA--GADSLA ALLAQGREAF P--AAPTAVD DACF----- --LYSSGSGT APKGTVHHS	[240]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	AADLPHLRTI IAV-GDAAA--PTLQLA NLLATEQEGE A--PAATCAD DIAYH----- --QYSSGSGT MPKGVHHS	[240]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	RTRVPSVCLD ALD----- --DLVSGS RVPEVAGKSL P--PPCPPE QAGFV----- --FYTSGTTG LPKGVAIQR	[240]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	AGELEQPQV VVR-GDEAP--APATIAFK HFLDQQAAL E--SVQVAPD DVAHY----- --LYSSGTTG RPKVAHHR	[240]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	VATRPSVAHV VSF-GDAGG--AGLE AAMARQSGEF D--NWTASD DTCIL----- --AFTSGTTG QPKATFHHR	[240]
<b>A3/3</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	-----DNLGIFLY SRGLYAHGIR PG--HRGFGF FSYPPFIAFH LCHSACESHG CQIVYKGP-L STKANHLG-L --IKNLSGTG	[320]
2. putative_2-NACL_N47_CBX27264.1	-----DMSTGIAFV -KWAHAGIR PA-DIPIY FPFGTMAFH CAYDVAAG AQVITSGG-M TTEQVRG-- --ILEKPTG	[320]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	-----DIDHTATVV ARSIRASGGR AG--DMWHTA YGYGLFTGGL GAHYGAELGK CTVVPMSSG-G QTEKQIQL-- --IQDFKPD	[320]
4. PCL Pseudomonas putida WP_059394932.1*	-----DINTWAVV ARSIRASGGR KG--DKVHVA YGYGLFTGGL GAHYGAELGK CTVVPMSSG-G QTEKQIQL-- --IRDQFDP	[320]
5. BCL Thauera aromatica AAN32623.1*	-----DLHTAELY ARPIL--GIR ED--DVVFA AKLFFAYVGL NGLIFPLAVG ATAVLMAERP TPAAVFER-- --LRRHQPT	[320]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	-----DLPATAVHY GQVVL--GIR ED--DVVFA AKLFFAYVGL NGMTFPLAVG ATSVLKDPR TPAAVFKL-- --LKDQPT	[320]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	-----SLHTAELY ARPVL--GIR ED--DVVFA AKLFFAYVGL NGLTFPLAVG ATAVLMAERP TPAAVFER-- --LRHQPT	[320]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	-----SPRVMAENA GRRV--GIR ED--DVVFA AKLFFAYVGL NAMFCPMGIG ATSVLVPERP TADSVFDT-- --LRLHQPT	[320]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	-----AAESRVLFM ATQA--GIR ED--DVVFA AKLFFAYVGL NAMFCPMGIG ATSVLVPERP TADSVFDT-- --LRLHQPT	[320]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	-----SVLIADRL EREYF--GIR PG--DRVFTF SKMFFGNSLG HSLMGLQCG ATVIVAPGMP DAERVMAT-- --AARHPRT	[320]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	-----DVIIACRCM PPHVL--RQP PD--DVIFGS PPLAFTFGLG GMLFFPSVGE ASTVLEQ-A SPPKLLDA-- --IGEGFAT	[320]
<b>A5</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	VDSFLAATPT FTMHRVEMAE ELGINLKLNL -IKVLTMAGE PGACVPSTK YLENAAAKA -HDQLGSVET SGP--VWYSC	[400]
2. putative_2-NACL_N47_CBX27264.1	-----VLVATPT YAMRIAEAMR EMGVDDIAHS -IKYITSAE AGVYLVPTRD AVEKANGAKA -LDLYGISDL WGS--TSMHC	[400]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	-----IIMVTPS YMLVLDEME RFGIDPHQTS -LKVGFIGAE PMT--QAMRA AMEARAGIDA -VDIYGLSEV MGP-GVAMEC	[400]
4. PCL Pseudomonas putida WP_059394932.1*	-----IIMVTPS YMLNLDAEIE RFGIDPHDLK -LRLGIFGAE PMT--DELRR SIEQRALGINA -LDIYGLSET MGP-GVAMEC	[400]
5. BCL Thauera aromatica AAN32623.1*	-----IFYGVPT LYASMLANDP CPKE--GELR -LRACTSAGE ALP--EDVGR RWAQRFVGTI -LDIGSSTEM LHI-FLSMRA	[400]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	-----IFYGVPT LYGIILADQP YRRET-ASTR -LRACTSAGE ALP--EDVGR RWEREFGAAI -LDIGSSTEM LHI-FLSMRH	[400]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	-----IFYGVPT LYASMLASDP CPAR--GELN -LRACTSAGE ALP--EIEGR RWTERVGLDI -LDIGSSTEM LHI-FLSMRP	[400]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	-----LLFAVPT LYAAHLADPR SRTET-LPDR -LRLCVSAGE PLP--AQVGL WRRNRFGHDI -VMNGSSTEM GHL-FTMLP	[400]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	-----AMFATPT HDLALTAIVE QAGL-RLES -LEHTVFAQA TWP--DTYLE RVNRFPSEK -NVIYGTTEA MSLYMRAR	[400]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	-----ILEFSTP MYRNLREGA GESA--AMRD -IRHVSAGE KLP--EMIQO QMLDFTGPII -TEGIGASET VFL-FLCARP	[400]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	-----ILFATPT SYRAMAEFAR GERLAPLGG PLVKCVSAGE VLP--AATRA LKWDATGIEI -IDGIGATEM FHI-FISADE	[400]
<b>A5</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	AEQEEENMS DNLHLS-D LVELVDVPT -LAKPVGDGE PGATVV--TA LTRFGMPTIR F-----LLGDW MITSYKECRC	[480]
2. putative_2-NACL_N47_CBX27264.1	P--VHP--- DRLHLETIAI YPLVDKEG -KLVDPDGG IGEWLT--TN Y-STVMPLIK Y-----RTHDT VENHKTECEC	[480]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	I--EAK--- DGPVIEDHF YPEIDPHT -GEVLPDGS EGELVF--TN LTKAEMPVIR Y-----RTRDL TRLLPPTARS	[480]
4. PCL Pseudomonas putida WP_059394932.1*	I--ETK--- DGPVIEDHF YPEIDPHT -GEVLPDGG EGELVF--TS LSKEALPMVR Y-----RTRDL TRLLPPTARS	[480]
5. BCL Thauera aromatica AAN32623.1*	G--DVHYG-T SGKVPVG--Y RRLRIDEGD -AEITTAGV AGELQI--SG PSSAVMYWNN -----PEKTA ATFMGETRPS	[480]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	G--EVRYG-T SGKVPVG--Y RRLRIDEGD -HEV-PQGE MGEVLV--RG PSSATAYWNN -----REKSL KTRFGVETHT	[480]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	G--QVRYG-T SGMPVPG--Y RLVKIDAGG -NEITAPGE PGEQLI--SG PPSAVMYWNN -----REKSR STFLGVEHTS	[480]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	H--AVEYGT-S SGVPVGD--Y RRLVGDGRG -QDV-ADDE IGEVLL--SG GSSAAGYWNN -----RDKTR TTFVGEVTRT	[480]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	I-----A GVMRPFQFYS EFRIVRVGG -DVEDDGCPTVK RASMRW-----RRNRPFQA -----TLTNL RLLQKSFRA	[480]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	D--AYRIG-S CGKRVPP--A EVRLDELG -NEITTPDT PGLIAI--RM ASQFVYMKL -----PETTE KALRDSGNYT	[480]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	E--HARPG-A TGTVPVG--Y RARIVDDEG -REV-PAGT VGRVLA--KG PTGCRYLD-----DSRQR NVYGDGWNYP	[480]
<b>A7/A8/1</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	GR---TLPLA KGG---IKAR SDDLIIKGT NIYPSLIENS VRSIEGLSPE YRIRVKRT-- --NAIVMVE AK-PGIK---	[560]
2. putative_2-NACL_N47_CBX27264.1	GR---TWTWL RGG---VLGR TQWMTIKGT NVFPTAIQGI IGGDGLTEN LEIHLITG-E GGDIEDIKVE AA-PDVP---	[560]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	MR---RMAK-----ITGR SDDMLIIRGV NLFPQTVEEL ICKNPKLAP YLLEVDKD-G HMDTLTKVKE IN-PEANVR---	[560]
4. PCL Pseudomonas putida WP_059394932.1*	MR---RIGK-----ITGR SDDMLIIRGV NVFPTQIEEL IWKIKLSEM YEITHLYRN-G NLDSEVHVEH LR-AECQ-HL---	[560]
5. BCL Thauera aromatica AAN32623.1*	GD---KYLVM DEGYVYVYAGR SDDMLKVSIGI YVSPFIEVSA LIAEAVLEA AVGWEDD-E HLIXPKAFIV LK-PGVY---	[560]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	GD---KYYVD DGYVYVYAGR GDDMLKVSIGI WSPFVEVSA LISHDKVLEA AVWGEDA-E GLVKPKAFVY LA-PGET---	[560]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	GD---KYLVD QDGYVYVYAGR SDDMLKVSIGI YVSPFIEVSA LIAEAVLEA AVWGEDA-E HLIXPKAFIV LK-PGLE---	[560]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	GD---KYHRR ADGVTYTCGR TDDIFKVSIGI WSPFIEVSA LMSHAKVLEA AVIPAEDT-D GLIXPKAFIV LASRGI---	[560]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	GTGRACIVRD GSGNIVLVRG VDDMIISGGE NHPSEVERT LAAPGVAVL VVIGVQDE-R WQSVVACV LQ-PGAS---	[560]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	GD---MFSFD ADGVTYVYVYAGR ADDMLKVSIGI WSPFIEVSA ASAVPGTAEV VVAVVPM-D GLTRTLTFIV PE-OPSA---	[560]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	GD---AYYMD ADGVTYVYVYAGR LDDMIVSAGY NIGAPEVEDA LMQHAPVAEC AVIGVPE-E RGQIVAKFVY PR-PGHW---	[560]
<b>A10</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	-----KTDYQKLSK -----LLEEDIR -DKTSVRLMI EVNPP-GTLP R-EIDVQTKRI IRE-----	[640]
2. putative_2-NACL_N47_CBX27264.1	-----AEKYEVLLK -----HLRGLQ -FRIGVMKV DLLP-KTLP R-YEVKAKRV FDRHKKR-----	[640]
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	-----PEQKALAK -----ELQHDIK -TFVGSYAKV HCEP-FAIE RTVIGAKRV VDRRPEK-----	[640]
4. PCL Pseudomonas putida WP_059394932.1*	-----DEGQRKLVI -----ELSKQIK -NYLIGISTQV HLQAC-GTLK R-SEKACHV YDKRLAS-----	[640]
5. BCL Thauera aromatica AAN32623.1*	-----AGEALRT -----DLKAHKV -HLLAPYKVP RWIEFVDDLK KTATGKIQRF KLRSA-----	[640]
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	-----GSEVLKE -----ELQAVK -SKLAPYKVP RWIEFVDEAL KTATGKIQRF KLRGTLGSK-----	[640]
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	-----ADEALRE -----MLKNHKV -AMLAPYKVP RWIEFVDEAL KTATGKIQRF KLRGMSAR-----	[640]
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	-----DPGALFD -----ELKEHKV -SAIGPKYKVP RWIEFVDEAL KTATGKIQRF LKRGMSAR-----	[640]
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	-----ASAE -----RLDFACR ASALADFKRP RRYVFLDELK KSAMNKLVR LQHQVSTAS SAUVVAPAV KQRTYVPSGN	[640]
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	-----SQQLKSE -----AVMTTLR -GTLSTYKCP RTIQFLEELP RTATGKIQRF RLRDMLQATL SAUVVAPAV KQRTYVPSGN	[640]
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	-----AGELLVR -----ELQDFVK -RTIAPYKVP RAIEFRDLSL RTETGKIQRF RLREGKP-----	[640]
<b>A11</b>		
1. putative_2-PACL_PITCH_v1_a1100006_TRIP_SPD71907.1	---- [644]	
2. putative_2-NACL_N47_CBX27264.1	---- [644]	
3. PCL (aerobic) Azoarcus Evansii WP_169132234.1*	---- [644]	
4. PCL Pseudomonas putida WP_059394932.1*	---- [644]	
5. BCL Thauera aromatica AAN32623.1*	---- [644]	
6. BCL Magnetospirillum sp. TS-6_BAE91925.1*	---- [644]	
7. 4-HBCL/Thauera aromatica_K172_AVR88075.1*	---- [644]	
8. 4-HBCL/BCL Rhodopseudomonas palustris_Q53005.1*	GR-- [644]	
9. 4-CBCL Pseudomonas sp. CBS3_ABQ44579.1*	AIAR [644]	
10. 3-HBCL Thauera aromatica_K172_AVR88779.1*	---- [644]	
11. 2-ABCL Azoarcus Evansii_AAL02077.1*	---- [644]	

**FIG 4** Alignment of 11 out of 39 aryl-CoA ligase protein sequences, with structural motifs highlighted in blue. The protein sequences were taken out of the alignment of 39 aryl-CoA ligases used for the maximum likelihood phylogenetic (Continued on next page)



**FIG 5** Crystal structures of benzoate:CoA ligases (a and b) and phenylacetate:CoA ligases (c and d). (a) Crystal structure of benzoate:CoA ligase (BadA) from *Rhodospseudomonas palustris* with benzoyl-AMP (PDB accession number 4ZJZ) (46). (b) Crystal structure of benzoate:CoA ligase from *Paraburkholderia xenovorans* LB400 (PDB accession number 2V7B) (47). (c and d) Crystal structures of the phenylacetate:CoA ligases PaaK1 (PDB accession number 2Y4N) and PaaK2 (PDB accession number 2Y4O) from *Burkholderia cenocepacia* J2315 with phenylacetyl adenylate (49). Images were created using Mol\* Viewer (78) from the RCSB PDB database.

interactions between a lysine residue contained in the C-terminal domain and the carboxylate group of the benzoate. Contrary to other benzoate:CoA ligases, which adopt the adenylation conformation during cocrystallization with benzoate (65, 70), BadA takes on the thiolation conformation even in the absence of CoA. This indicates two possible subgroups of benzoate:CoA ligases: enzymes with a resting state in the thiolation conformation and those in the adenylation conformation. Those authors propose a multistep reaction mechanism for BadA, analogous to that of the human medium-chain acyl-CoA synthetase ACSM2A (71). Here, the carboxylate binds to the enzyme in the thiolation conformation, followed by a domain rearrangement to the adenylation conformation in which ATP binds to the substrate and acyl-AMP is formed. The last step is the typical domain movement to the thiolation conformation for the thioesterification reaction with CoA.

Crystallization studies of the phenylacetate:CoA ligases PaaK1 and PaaK2 from *Burkholderia cenocepacia* J2315 (72) (Fig. 5) show the phenylacetate:CoA ligases as a distinct subgroup of the aryl-CoA ligases besides the benzoate:CoA ligases. PaaK1 and PaaK2 show structures similar to those of homologous aryl-CoA ligases apart from one additional feature in the N-terminal domain of these enzymes, a novel microdomain of ~70 residues. This microdomain is built with a leucine-zipper-like organization and forms a continuous three-helical bundle structure. The microdomain is stabilized by the hydrophobic interactions of 6 leucine residues and Tyr30, Phe44, and Phe63 of PaaK1. This alteration in the N-terminal domain leads to a change in the typical conformation of the group of four

#### FIG 4 Legend (Continued)

tree in Fig. 3. The whole alignment can be found in the supplemental material. The structural motifs found in aryl-CoA ligases (Table 2) are highlighted in blue. Amino acids that are 100% conserved in the whole alignment are marked in yellow.

$\beta$ -sheets surrounded by five  $\alpha$ -helices as can be found in the benzoate:CoA ligase from *Burkholderia xenovorans* (70) or the 4-chlorobenzoate:CoA ligase of *Pseudomonas* sp. (63). The monomeric N-terminal domain of PaaK1 is comprised of three  $\beta$ -sheets surrounded by nine  $\alpha$ -helices, which leaves the phosphate-binding loop (P-loop) (motif A3) isolated, leading to the necessary dimerization of this domain. Cocrystallization of PaaK1 with ATP further proved the binding of the phosphates from ATP to the P-loop. Hydrogen bonds formed between the amide backbone of the P-loop and the phosphates from ATP reinforce this coupling, with additional stabilization through hydrogen bonds with Ser94, Thr96, and Thr97. As mentioned above, the lysine residue of the A10 motif is highly conserved in these two enzymes as well as in other aryl-CoA ligases.

### PHENYLACETATE:CoA LIGASES, AN EXPANDING GROUP

Recent studies have elucidated new enzymes, expanding the group of phenylacetate:CoA ligases with substrates outside the already known aryl carboxylates. The 2-hydroxyisobutyric acid:CoA ligase from *Aquicola tertiarycarbonis* L108 (73) shows a closer relationship to the phenylacetate:CoA ligases PaaK1 and PaaK2, with 34% and 33% coverages of homologous sequences, compared to 20% with other short- and medium-chain acid ANL ligases. The enzyme shows a similar tertiary structure typical of phenylacetate:CoA ligases, forming a dimer with a seven-stranded  $\beta$ -sheet connecting the two subunits. The acyl-binding pocket contains the unique active-site residues Tyr164 and Ser239, leading to a smaller and more polar environment.

A unique set of enzymes from the benzoxazole biosynthesis pathway has recently been identified as a subset of the phenylacetate:CoA ligase family (74, 75). The enzymes AjiA1 and the closely related ligase NatL2 show 30% sequence identity with PaaK1 but do not undergo a domain alteration; instead, these newly described enzymes feature a domain-swapping reaction involving a C-terminal loop constricting the enzymes to one conformation.

As described above in this review, phylogenetic analyses of putative 2-phenanthroate:CoA and 2-naphthoate:CoA ligases (43, 76, 77) (Fig. 3) indicate that they belong to the monophyletic clade of phenylacetate:CoA ligases. These ligases show a sequence similarity of  $\sim$ 20 to 25% with the described benzoate:CoA ligases, compared to  $\sim$ 30% sequence similarity with phenylacetate:CoA ligases.

### CONCLUSION

The group of aryl-CoA ligases is a central group of the ANL superfamily. Recent evidence even suggested an aryl-CoA ligase-like enzyme as an ancestral enzyme from which the more specialized functions of the ANL superfamily have evolved (51).

Most aryl-CoA ligases are encoded either in the same operon as or close to the corresponding degradation genes of their respective pathways (22). The regulation of these genes is still not fully understood, as some aryl-CoA ligases are regulated by oxygen availability (24), while others are regulated by posttranslational acetylation of a C-terminal conserved lysine residue (66). This in turn influences benzoyl-CoA levels and might be used in some cases to regulate the gene expression of aryl-CoA ligases through a transcription repressor (69).

Acetylation of catalytic residues in prokaryotes is still a relatively new topic, and as such, further research is necessary. It is still unclear what factors influence the expression of acetyltransferases and deacetylases, as many times, these enzymes are not encoded near their corresponding genes. In the future, more aryl-CoA ligases as well as other ANL superfamily enzymes need to be examined for signs of acetylation, as the lysine residue is highly conserved throughout the family.

Crystallization studies give further insights into and elucidate the key residues responsible for reaction mechanisms while also showing new strategies emerging from newly identified members of this group. The aryl-CoA ligases are divided into two major groups, the benzoate:CoA ligase-like and the phenylacetate:CoA ligase-like

enzymes, with their members clustering according to their substrate specificity, as the active site contains most of the conserved amino acid residues (9).

The newly discovered ligases from anaerobic polycyclic aromatic hydrocarbon degradation pathways fall into the second group of phenylacetate:CoA ligases, which expands our knowledge of the essential requirements of these enzymes. In recent years, the group of aryl-CoA ligases has been steadily growing, with more and more enzymes being identified. These enzymes share little sequence similarity, but newly emerging bioinformatic tools and the expansion of machine-learning approaches demonstrate new ways of identifying possible candidate enzymes (51). Future studies will have to focus on these enzymes and their role in the aryl-CoA ligase family as well as the importance of this group in the context of the wider ANL superfamily.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

I. Kaplieva-Dudek and R. U. Meckenstock acknowledge support by the EU-Horizon 2020 program through the ERC advanced grant EcOILogy number 666952.

We declare no conflict of interest.

## REFERENCES

- Schmelz S, Naismith JH. 2009. Adenylate-forming enzymes. *Curr Opin Struct Biol* 19:666–671. <https://doi.org/10.1016/j.sbi.2009.09.004>.
- Black PN, DiRusso CC. 2003. Transmembrane movement of exogenous long-chain fatty acids: proteins, enzymes, and vectorial esterification. *Microbiol Mol Biol Rev* 67:454–472. <https://doi.org/10.1128/MMBR.67.3.454-472.2003>.
- Overath P, Pauli G, Schairer H. 1969. Fatty acid degradation in *Escherichia coli*: an inducible acyl-CoA synthetase, the mapping of old-mutations, and the isolation of regulatory mutants. *Eur J Biochem* 7:559–574.
- Murakami K, Ide T, Nakazawa T, Okazaki T, Mochizuki T, Kadowaki T. 2001. Fatty-acyl-CoA thioesters inhibit recruitment of steroid receptor co-activator 1 to alpha and gamma isoforms of peroxisome-proliferator-activated receptors by competing with agonists. *Biochem J* 353:231–238. <https://doi.org/10.1042/0264-6021:3530231>.
- Tejima K, Ishiai M, Murayama SO, Iwatani S, Kajiwara S. 2018. *Candida albicans* fatty acyl-CoA synthetase, CaFaa4p, is involved in the uptake of exogenous long-chain fatty acids and cell activity in the biofilm. *Curr Genet* 64:429–441. <https://doi.org/10.1007/s00294-017-0751-2>.
- Banchio C, Gramajo H. 2002. A stationary-phase acyl-coenzyme A synthetase of *Streptomyces coelicolor* A3(2) is necessary for the normal onset of antibiotic production. *Appl Environ Microbiol* 68:4240–4246. <https://doi.org/10.1128/AEM.68.9.4240-4246.2002>.
- Glick BS, Rothman JE. 1987. Possible role for fatty acyl-coenzyme A in intracellular protein transport. *Nature* 326:309–312. <https://doi.org/10.1038/326309a0>.
- Pfanner N, Glick BS, Arden SR, Rothman JE. 1990. Fatty acylation promotes fusion of transport vesicles with Golgi cisternae. *J Cell Biol* 110:955–961. <https://doi.org/10.1083/jcb.110.4.955>.
- Gulick AM. 2009. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol* 4:811–827. <https://doi.org/10.1021/cb900156h>.
- Chang K-H, Dunaway-Mariano D. 1996. Determination of the chemical pathway for 4-chlorobenzoate:coenzyme A ligase catalysis. *Biochemistry* 35:13478–13484. <https://doi.org/10.1021/bi961284w>.
- Wu R, Cao J, Lu X, Reger AS, Gulick AM, Dunaway-Mariano D. 2008. Mechanism of 4-chlorobenzoate:coenzyme A ligase catalysis. *Biochemistry* 47:8026–8039. <https://doi.org/10.1021/bi800698m>.
- Harwood CS, Gibson J. 1986. Uptake of benzoate by *Rhodospseudomonas palustris* grown anaerobically in light. *J Bacteriol* 165:504–509. <https://doi.org/10.1128/jb.165.2.504-509.1986>.
- Verleur N, Hetteema EH, Van Roermond CWT, Tabak HF, Wanders RJA. 1997. Transport of activated fatty acids by the peroxisomal ATP-binding-cassette transporter Pxa2 in a semi-intact yeast cell system. *Eur J Biochem* 249:657–661. <https://doi.org/10.1111/j.1432-1033.1997.00657.x>.
- Lynen F, Reichert E, Rueff L. 1951. Zum biologischen Abbau der Essigsäure V11). “Aktivierte Essigsäure,” ihre Isolierung aus Hefe und ihre chemische Natur. *Justus Liebigs Ann Chem* 574:1–32. <https://doi.org/10.1002/jlac.19515740102>.
- Buckel W, Keese R. 1995. One-electron redox reactions of CoA-SH esters in anaerobic bacteria—a mechanistic proposal. *Angew Chem Int Ed Engl* 34:1502–1506. <https://doi.org/10.1002/anie.199515021>.
- Fuchs G. 2008. Anaerobic metabolism of aromatic compounds. *Ann N Y Acad Sci* 1125:82–99. <https://doi.org/10.1196/annals.1419.010>.
- Black PN, Kianian SF, DiRusso CC, Nunn WD. 1985. Long-chain fatty acid transport in *Escherichia coli*. Cloning, mapping, and expression of the *fadL* gene. *J Biol Chem* 260:1780–1789. [https://doi.org/10.1016/S0021-9258\(18\)89661-2](https://doi.org/10.1016/S0021-9258(18)89661-2).
- Faergeman NJ, Black PN, Zhao XD, Knudsen J, DiRusso CC. 2001. The acyl-CoA synthetases encoded within *FAA1* and *FAA4* in *Saccharomyces cerevisiae* function as components of the fatty acid transport system linking import, activation, and intracellular utilization. *J Biol Chem* 276:37051–37059. <https://doi.org/10.1074/jbc.M100884200>.
- Merkel SM, Eberhard AE, Gibson J, Harwood CS. 1989. Involvement of coenzyme A thioesters in anaerobic metabolism of 4-hydroxybenzoate by *Rhodospseudomonas palustris*. *J Bacteriol* 171:1–7. <https://doi.org/10.1128/jb.171.1.1-7.1989>.
- Boll M, Fuchs G. 1995. Benzoyl-coenzyme A reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. ATP dependence of the reaction, purification and some properties of the enzyme from *Thauera aromatica* strain K172. *Eur J Biochem* 234:921–933. <https://doi.org/10.1111/j.1432-1033.1995.921.a.x>.
- Thiele B, Rieder O, Jehmlich N, von Bergen M, Müller M, Boll M. 2008. Aromatizing cyclohexa-1,5-diene-1-carbonyl-coenzyme A oxidase—characterization and its role in anaerobic aromatic metabolism. *J Biol Chem* 283:20713–20721. <https://doi.org/10.1074/jbc.M802841200>.
- Carmona M, Zamarro MT, Blázquez B, Durante-Rodríguez G, Juárez JF, Valderrama JA, Barragan MJL, García JL, Díaz E. 2009. Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiol Mol Biol Rev* 73:71–133. <https://doi.org/10.1128/MMBR.00021-08>.
- Díaz E, Jiménez JI, Nogales J. 2013. Aerobic degradation of aromatic compounds. *Curr Opin Biotechnol* 24:431–442. <https://doi.org/10.1016/j.copbio.2012.10.010>.
- Valderrama JA, Durante-Rodríguez G, Blázquez B, García JL, Carmona M, Díaz E. 2012. Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. *J Biol Chem* 287:10494–10508. <https://doi.org/10.1074/jbc.M111.309005>.
- Altenschmidt U, Oswald B, Fuchs G. 1991. Purification and characterization of benzoate-coenzyme A ligase and 2-aminobenzoate-coenzyme A

- ligases from a denitrifying *Pseudomonas* sp. J Bacteriol 173:5494–5501. <https://doi.org/10.1128/jb.173.17.5494-5501.1991>.
26. Altenschmidt U, Oswald B, Steiner E, Herrmann H, Fuchs G. 1993. New aerobic benzoate oxidation pathway via benzoyl-coenzyme A and 3-hydroxybenzoyl-coenzyme A in a denitrifying *Pseudomonas* sp. J Bacteriol 175:4851–4858. <https://doi.org/10.1128/jb.175.15.4851-4858.1993>.
  27. Harwood CS, Burchhardt G, Herrmann H, Fuchs G. 1998. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. FEMS Microbiol Rev 22:439–458. <https://doi.org/10.1111/j.1574-6976.1998.tb00380.x>.
  28. Gescher J, Zaar A, Mohamed M, Schägger H, Fuchs G. 2002. Genes coding for a new pathway of aerobic benzoate metabolism in *Azoarcus evansii*. J Bacteriol 184:6301–6315. <https://doi.org/10.1128/JB.184.22.6301-6315.2002>.
  29. Altenschmidt U, Fuchs G. 1992. Novel aerobic 2-aminobenzoate metabolism. Eur J Biochem 205:721–727. <https://doi.org/10.1111/j.1432-1033.1992.tb16835.x>.
  30. Schühle K, Jahn M, Ghisla S, Fuchs G. 2001. Two similar gene clusters coding for enzymes of a new type of aerobic 2-aminobenzoate (anthranilate) metabolism in the bacterium *Azoarcus evansii*. J Bacteriol 183:5268–5278. <https://doi.org/10.1128/JB.183.18.5268-5278.2001>.
  31. Mohamed ME, Fuchs G. 1993. Purification and characterization of phenylacetate-coenzyme A ligase from a denitrifying *Pseudomonas* sp., an enzyme involved in the anaerobic degradation of phenylacetate. Arch Microbiol 159:554–562. <https://doi.org/10.1007/BF00249035>.
  32. El-Said Mohamed M. 2000. Biochemical and molecular characterization of phenylacetate-coenzyme A ligase, an enzyme catalyzing the first step in aerobic metabolism of phenylacetic acid in *Azoarcus evansii*. J Bacteriol 182:286–294. <https://doi.org/10.1128/JB.182.2.286-294.2000>.
  33. Rost R, Haas S, Hammer E, Herrmann H, Burchhardt G. 2002. Molecular analysis of aerobic phenylacetate degradation in *Azoarcus evansii*. Mol Genet Genomics 267:656–663. <https://doi.org/10.1007/s00438-002-0699-9>.
  34. Schühle K, Gescher J, Feil U, Paul M, Jahn M, Schägger H, Fuchs G. 2003. Benzoate-coenzyme A ligase from *Thauera aromatica*: an enzyme acting in anaerobic and aerobic pathways. J Bacteriol 185:4920–4929. <https://doi.org/10.1128/JB.185.16.4920-4929.2003>.
  35. Kawaguchi K, Shinoda Y, Yurimoto H, Sakai Y, Kato N. 2006. Purification and characterization of benzoate-CoA ligase from *Magnetospirillum* sp. strain TS-6 capable of aerobic and anaerobic degradation of aromatic compounds. FEMS Microbiol Lett 257:208–213. <https://doi.org/10.1111/j.1574-6968.2006.00165.x>.
  36. Biegert T, Altenschmidt U, Eckerskorn C, Fuchs G. 1993. Enzymes of anaerobic metabolism of phenolic compounds. Eur J Biochem 213:555–561. <https://doi.org/10.1111/j.1432-1033.1993.tb17794.x>.
  37. Laempe D, Jahn M, Breese K, Schagger H, Fuchs G. 2001. Anaerobic metabolism of 3-hydroxybenzoate by the denitrifying bacterium *Thauera aromatica*. J Bacteriol 183:968–979. <https://doi.org/10.1128/JB.183.3.968-979.2001>.
  38. Egland PG, Pelletier DA, Dispensa M, Gibson J, Harwood CS. 1997. A cluster of bacterial genes for anaerobic benzene ring biodegradation. Proc Natl Acad Sci U S A 94:6484–6489. <https://doi.org/10.1073/pnas.94.12.6484>.
  39. Martínez-Blanco H, Reglero A, Rodríguez-Aparicio LB, Luengo JM. 1990. Purification and biochemical characterization of phenylacetyl-CoA ligase from *Pseudomonas putida*. A specific enzyme for the catabolism of phenylacetic acid. J Biol Chem 265:7084–7090. [https://doi.org/10.1016/S0021-9258\(19\)39262-2](https://doi.org/10.1016/S0021-9258(19)39262-2).
  40. Meckenstock RU, Annweiler E, Michaelis W, Richnow HH, Schink B. 2000. Anaerobic naphthalene degradation by a sulfate-reducing enrichment culture. Appl Environ Microbiol 66:2743–2747. <https://doi.org/10.1128/AEM.66.7.2743-2747.2000>.
  41. Galushko A, Minz D, Schink B, Widdel F. 1999. Anaerobic degradation of naphthalene by a pure culture of a novel type of marine sulphate-reducing bacterium. Environ Microbiol 1:415–420. <https://doi.org/10.1046/j.1462-2920.1999.00051.x>.
  42. Himmelberg AM, Bröls T, Farmani Z, Weyrauch P, Barthel G, Schrader W, Meckenstock RU. 2018. Anaerobic degradation of phenanthrene by a sulfate-reducing enrichment culture. Environ Microbiol 20:3589–3600. <https://doi.org/10.1111/1462-2920.14335>.
  43. Kraiselburd I, Bröls T, Heilmann G, Kaschani F, Kaiser M, Meckenstock RU. 2019. Metabolic reconstruction of the genome of candidate *Desulfatiglans* TRIP\_1 and identification of key candidate enzymes for anaerobic phenanthrene degradation. Environ Microbiol 21:1267–1286. <https://doi.org/10.1111/1462-2920.14527>.
  44. Geissler JF, Harwood CS, Gibson J. 1988. Purification and properties of benzoate-coenzyme A ligase, a *Rhodospseudomonas palustris* enzyme involved in the anaerobic degradation of benzoate. J Bacteriol 170:1709–1714. <https://doi.org/10.1128/jb.170.4.1709-1714.1988>.
  45. Chang KH, Liang HG, Beck W, Scholten JD, Dunaway-Mariano D. 1992. Isolation and characterization of the three polypeptide components of 4-chlorobenzoate dehalogenase from *Pseudomonas* sp. strain CBS-3. Biochemistry 31:5605–5610. <https://doi.org/10.1021/bi00139a025>.
  46. Vessey DA, Kelley M, Lau E, Zhang SZ. 2000. Monovalent cation effects on the activity of the xenobiotic/medium-chain fatty acid:CoA ligases of bovine liver mitochondria. Biochim Biophys Acta 1382:243–248. [https://doi.org/10.1016/s0167-4838\(97\)00163-5](https://doi.org/10.1016/s0167-4838(97)00163-5).
  47. Vessey DA, Kelley M, Lau E, Zhang SZ. 2000. Monovalent cation effects on the activity of the xenobiotic/medium-chain fatty acid:CoA ligases are substrate specific. J Biochem Mol Toxicol 14:162–168. [https://doi.org/10.1002/\(SICI\)1099-0461\(2000\)14:3<162::AID-JBT6>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1099-0461(2000)14:3<162::AID-JBT6>3.0.CO;2-8).
  48. Gaid MM, Sircar D, Müller A, Beuerle T, Liu B, Ernst L, Hänsch R, Beerhues L. 2012. Cinnamate:CoA ligase initiates the biosynthesis of a benzoate-derived xanthone phytoalexin in *Hypericum calycinum* cell cultures. Plant Physiol 160:1267–1280. <https://doi.org/10.1104/pp.112.204180>.
  49. Clark L, Leatherby D, Krilich E, Ropelewski AJ, Perozich J. 2018. In silico analysis of class I adenylate-forming enzymes reveals family and group-specific conservations. PLoS One 13:e0203218. <https://doi.org/10.1371/journal.pone.0203218>.
  50. Hempel J, Perozich J, Wymore T, Nicholas HB. 2003. An algorithm for identification and ranking of family-specific residues, applied to the ALDH3 family. Chem Biol Interact 143–144:23–28. [https://doi.org/10.1016/S0009-2797\(02\)00165-5](https://doi.org/10.1016/S0009-2797(02)00165-5).
  51. Robinson SL, Terlow BR, Smith MD, Pidot SJ, Stinear TP, Medema MH, Wackett LP. 2020. Global analysis of adenylate-forming enzymes reveals  $\beta$ -lactone biosynthesis pathway in pathogenic *Nocardia*. J Biol Chem 295:14826–14839. <https://doi.org/10.1074/jbc.RA120.013528>.
  52. Akiva E, Copp JN, Tokuriki N, Babbitt PC. 2017. Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. Proc Natl Acad Sci U S A 114:E9549–E9558. <https://doi.org/10.1073/pnas.1706849114>.
  53. Linne U, Schäfer A, Stubbs MT, Marahiel MA. 2007. Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains. FEBS Lett 581:905–910. <https://doi.org/10.1016/j.febslet.2007.01.066>.
  54. Oba Y, Ojika M, Inouye S. 2003. Firefly luciferase is a bifunctional enzyme: ATP-dependent monooxygenase and a long chain fatty acyl-CoA synthetase. FEBS Lett 540:251–254. [https://doi.org/10.1016/S0014-5793\(03\)00272-2](https://doi.org/10.1016/S0014-5793(03)00272-2).
  55. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. <https://doi.org/10.1093/bioinformatics/8.3.275>.
  56. Kumar S, Stecher G, Li M, Niyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>.
  57. Meckenstock RU, Boll M, Mouttaki H, Koelschbach JS, Cunha Tarouco P, Weyrauch P, Dong X, Himmelberg AM. 2016. Anaerobic degradation of benzene and polycyclic aromatic hydrocarbons. J Mol Microbiol Biotechnol 26:92–118. <https://doi.org/10.1159/000441358>.
  58. Marozava S, Mouttaki H, Muller H, Laban NA, Probst AJ, Meckenstock RU. 2018. Anaerobic degradation of 1-methylnaphthalene by a member of the *Thermoanaerobacteraceae* contained in an iron-reducing enrichment culture. Biodegradation 29:23–39. <https://doi.org/10.1007/s10532-017-9811-z>.
  59. Thornburg CK, Wortas-Strom S, Nosrati M, Geiger JH, Walker KD. 2015. Kinetically and crystallographically guided mutations of a benzoate CoA ligase (BadA) elucidate mechanism and expand substrate permissivity. Biochemistry 54:6230–6242. <https://doi.org/10.1021/acs.biochem.5b00899>.
  60. Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC. 2003. The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine 5'-propylphosphate and coenzyme A. Biochemistry 42:2866–2873. <https://doi.org/10.1021/bi0271603>.
  61. Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D. 2009. The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate:coenzyme A ligase. Biochemistry 48:4115–4125. <https://doi.org/10.1021/bi9002327>.
  62. Marahiel MA, Stachelhaus T, Mootz HD. 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem Rev 97:2651–2674. <https://doi.org/10.1021/cr960029e>.
  63. Chang KH, Xiang H, Dunaway-Mariano D. 1997. Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. Biochemistry 36:15650–15659. <https://doi.org/10.1021/bi971262p>.

64. Gulick AM, Lu X, Dunaway-Mariano D. 2004. Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states. *Biochemistry* 43:8670–8679. <https://doi.org/10.1021/bi049384m>.
65. Reger AS, Wu R, Dunaway-Mariano D, Gulick AM. 2008. Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* 47:8016–8025. <https://doi.org/10.1021/bi800696y>.
66. Crosby HA, Heiniger EK, Harwood CS, Escalante-Semerena JC. 2010. Reversible N epsilon-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in *Rhodopseudomonas palustris*. *Mol Microbiol* 76:874–888. <https://doi.org/10.1111/j.1365-2958.2010.07127.x>.
67. Burckhardt RM, VanDrisse CM, Tucker AC, Escalante-Semerena JC. 2020. New AMP-forming acid:CoA ligases from *Streptomyces lividans*, some of which are posttranslationally regulated by reversible lysine acetylation. *Mol Microbiol* 113:253–269. <https://doi.org/10.1111/mmi.14414>.
68. Starai VJ, Celic I, Cole RN, Boeke JD, Escalante-Semerena JC. 2002. Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science* 298:2390–2392. <https://doi.org/10.1126/science.1077650>.
69. Barragán MJ, Blázquez B, Zamarro MT, Mancheño JM, García JL, Díaz E, Carmona M. 2005. BzdR, a repressor that controls the anaerobic catabolism of benzoate in *Azoarcus* sp. ClB, is the first member of a new subfamily of transcriptional regulators. *J Biol Chem* 280:10683–10694. <https://doi.org/10.1074/jbc.M412259200>.
70. Bains J, Boulanger MJ. 2007. Biochemical and structural characterization of the paralogous benzoate CoA ligases from *Burkholderia xenovorans* LB400: defining the entry point into the novel benzoate oxidation (box) pathway. *J Mol Biol* 373:965–977. <https://doi.org/10.1016/j.jmb.2007.08.008>.
71. Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW. 2009. Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J Mol Biol* 388:997–1008. <https://doi.org/10.1016/j.jmb.2009.03.064>.
72. Law A, Boulanger MJ. 2011. Defining a structural and kinetic rationale for paralogous copies of phenylacetate-CoA ligases from the cystic fibrosis pathogen *Burkholderia cenocepacia* J2315. *J Biol Chem* 286:15577–15585. <https://doi.org/10.1074/jbc.M111.219683>.
73. Zahn M, Kurteva-Yaneva N, Schuster J, Krug U, Georgi T, Müller RH, Rohwerder T, Sträter N. 2019. Structures of 2-hydroxyisobutyric acid-CoA ligase reveal determinants of substrate specificity and describe a multi-conformational catalytic cycle. *J Mol Biol* 431:2747–2761. <https://doi.org/10.1016/j.jmb.2019.05.027>.
74. Song H, Rao C, Deng Z, Yu Y, Naismith JH. 2020. The biosynthesis of the benzoxazole in nataxazole proceeds via an unstable ester and has synthetic utility. *Angew Chem Int Ed Engl* 59:6054–6061. <https://doi.org/10.1002/anie.201915685>.
75. de Paiva FCR, Chan K, Samborsky M, Silber AM, Leadlay PF, Dias MVB. 2020. The crystal structure of AjiA1 reveals a novel structural motion mechanism in the adenylate-forming enzyme family. *Acta Crystallogr D Struct Biol* 76:1201–1210. <https://doi.org/10.1107/S2059798320013431>.
76. DiDonato RJ, Young ND, Butler JE, Chin KJ, Hixson KK, Mouser P, Lipton MS, DeBoy R, Methe BA. 2010. Genome sequence of the deltaproteobacterial strain NaphS2 and analysis of differential gene expression during anaerobic growth on naphthalene. *PLoS One* 5:e14072. <https://doi.org/10.1371/journal.pone.0014072>.
77. Bergmann FD, Selesi D, Meckenstock RU. 2011. Identification of new enzymes potentially involved in anaerobic naphthalene degradation by the sulfate-reducing enrichment culture N47. *Arch Microbiol* 193:241–250. <https://doi.org/10.1007/s00203-010-0667-4>.
78. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. 2021. Mol\* Viewer: modern Web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* 49:W431–W437. <https://doi.org/10.1093/nar/gkab314>.
79. Gibson J, Dispensa M, Fogg GC, Evans DT, Harwood CS. 1994. 4-Hydroxybenzoate-coenzyme A ligase from *Rhodopseudomonas palustris*: purification, gene sequence, and role in anaerobic degradation. *J Bacteriol* 176:634–641. <https://doi.org/10.1128/jb.176.3.634-641.1994>.
80. Harrison KJ, Crecy-Lagard V, Zallot R. 2018. Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 34:1406–1408. <https://doi.org/10.1093/bioinformatics/btx793>.