Plant and Animal Glycolate Oxidases Have a Common Eukaryotic Ancestor and Convergently Duplicated to Evolve Long-Chain 2-Hydroxy Acid Oxidases

Christian Esser,[1] Anke Kuhn,[2] Georg Groth,[3,4] Martin J. Lercher,[1,4] and Veronica G. Maurino*[,2,4]

[1]Institute for Computer Science, Heinrich-Heine-University, Düsseldorf, Germany

[2]Institut of Developmental and Molecular Biology of Plants, Plant Molecular Physiology and Biotechnology Group, Heinrich-Heine-University, Düsseldorf, Germany

[3]Institute of Biochemical Plant Physiology, Heinrich-Heine-University, Düsseldorf, Germany

[4]Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-University, Düsseldorf, Germany

***Corresponding author:** E-mail: veronica.maurino@uni-duesseldorf.de.

Associate editor: John True

Abstract

Glycolate oxidase (GOX) is a crucial enzyme of plant photorespiration. The encoding gene is thought to have originated from endosymbiotic gene transfer between the eukaryotic host and the cyanobacterial endosymbiont at the base of plantae. However, animals also possess GOX activities. Plant and animal GOX belong to the gene family of (L)-2-hydroxyacid-oxidases ((L)-2-HAOX). We find that all (L)-2-HAOX proteins in animals and archaeplastida go back to one ancestral eukaryotic sequence; the sole exceptions are green algae of the chlorophyta lineage. Chlorophyta replaced the ancestral eukaryotic (L)-2-HAOX with a bacterial ortholog, a lactate oxidase that may have been obtained through the primary endosymbiosis at the base of plantae; independent losses of this gene may explain its absence in other algal lineages (glaucophyta, rhodophyta, and charophyta). We also show that in addition to GOX, plants possess (L)-2-HAOX proteins with different specificities for medium- and long-chain hydroxyacids (lHAOX), likely involved in fatty acid and protein catabolism. Vertebrates possess lHAOX proteins acting on similar substrates as plant lHAOX; however, the existence of GOX and lHAOX subfamilies in both plants and animals is not due to shared ancestry but is the result of convergent evolution in the two most complex eukaryotic lineages. On the basis of targeting sequences and predicted substrate specificities, we conclude that the biological role of plantae (L)-2-HAOX in photorespiration evolved by co-opting an existing peroxisomal protein.

Key words: **glycolate oxidase, (L)-2-hydroxyacid oxidases, convergent evolution, lactate oxidase.**

Introduction

All photosynthetic organisms fix carbon from CO_2 through a reaction catalyzed by ribulose-1,5-bisphosphate-carboxylase/oxygenase (RuBisCO). RuBisCO also has a biologically relevant affinity to O_2, leading to the production of 2-phosphoglycolate (2PG) (Ogren 2003). 2PG is recycled through the photorespiratory cycle, an indispensable metabolic process for all organisms performing oxygenic photosynthesis (Zelitch et al. 2009; Maurino and Peterhansel 2010). In a first step, 2PG is converted to glycolate, a toxic 2-hydroxy acid. Enzymatic removal of glycolate is an essential part of photorespiration and is catalyzed by two phylogenetically unrelated enzymes. In cyanobacteria and chlorophyta, glycolate oxidation is catalyzed by glycolate dehydrogenase (GlcDH) (Nelson and Tolbert 1970; Beezley et al. 1976; Nakamura et al. 2005; Stabenau and Winkler 2005; Eisenhut et al. 2008). In contrast, charophyceae—which share a common ancestor with land plants to the exclusion of other algal lineages (Karol et al. 2001; Timme et al. 2012)—do not possess GlcDH activity; similar to land plants, charophyceae (e.g., *Mougeotia scalaris, Spirogira* sp., and *Chara* sp.) have glycolate oxidase (GOX) activity, which produces H_2O_2 in leaf-type peroxisomes (peroxisomes in photosynthetically active cells that catalyze essential reactions of photorespiration) (Stabenau 1976; Stabenau and Winkler 2005).

Cyanobacteria conferred parts of their photosynthetic machinery to a eukaryote ancestral to plantae sensu lato (archaeplastida, comprising glaucophytes, red and green algae, and plants) via endosymbiosis (Bauwe et al. 2012). Cyanobacteria encode not only GlcDH but also a homolog of plant GOX. This has led to the hypothesis that plant photorespiratory GOX was derived from a cyanobacterial gene together with other genes involved in photosynthesis (Eisenhut et al. 2008; Hackenberg et al. 2011; Kern et al. 2013).

However, plantae are not the only eukaryotes that possess members of the gene family (L)-2-hydroxyacid-oxidases ((L)-2-HAOXs), to which GOX and its cyanobacterial homolog belong. Mammals (human, mouse, and rat) possess two genes encoding (L)-2-HAOXs, also located to peroxisomes. Human short-chain (L)-2-HAOX (HAOX1 or GOX) is highly expressed in liver, kidney, and pancreas (Jones et al. 2000). Similar to plant GOX, human GOX shows the highest activity for the oxidation of glycolate to glyoxylate, which is then used for peroxisomal synthesis of glycine (Williams et al. 2000).

Article

Human GOX can also use glyoxylate, 2-hydroxyoctanoate, and 2-hydroxydodecanoate as substrates, albeit with much lower catalytic efficiency (Jones et al. 2000; Williams et al. 2000; Vignaud et al. 2007). The oxidation of glyoxylate to oxalate by GOX likely contributes to the physiopathology of hyperoxaluria type I and to the formation of gall and kidney calcium oxalate stones (Danpure and Purdue 1995; Murray et al. 2008). Mouse GOX exhibits similar enzymatic properties (Kohler et al. 1999).

The second human (L)-2-HAOX protein, HAOX2 or lHAOX, uses long-chain 2-hydroxy acids as substrates; it is also highly expressed in liver and kidney (Jones et al. 2000). Enzymatic activity has been shown only with the 2-hydroxy fatty acids 2-hydroxyoctanoate and 2-hydroxydodecanoate (Jones et al. 2000). Rat lHAOX shows activity with L-mandelate and other nonphysiological long-chain 2-hydroxy acids (Belmouden and Lederer 1996). In humans, lHAOX was suggested to be involved in the general pathway of fatty acid alpha-oxidation (e.g., alpha-oxidation of phytanic acid [Verhoeven et al. 1997, 1998]). In rat, lHAOX may play a role in the degradation of xenobiotics (Belmouden and Lederer 1996). Similarly, the model plant *Arabidopsis thaliana* encodes not only three paralogous GOX proteins but also two putative lHAOX proteins of uncharacterized function (Reumann et al. 2004).

Bacterial (L)-2-HAOX proteins exhibit L-lactate oxidase (LOX) activity (Duncan et al. 1989; Dong et al. 1993; Gibello et al. 1999; Umena et al. 2006; Gao et al. 2012). LOX is involved in the usage of L-lactate as an energy source, and it may also remove excess L-lactate to reduce toxic or inhibitory effects. The (L)-2-HAOX enzymes of the cyanobacterium *Nostoc* sp. and the chlorophyte *Chlamydomonas reinhartii* also show high LOX activity (Hackenberg et al. 2011), although these enzymes can also use glyoxylate, glycerate, and glycolate as substrates with low catalytic rates. In N_2 fixing cyanobacteria, LOX may serve as an O_2-scavenging enzyme to protect nitrogenase (Hackenberg et al. 2011).

Thus, although all characterized bacterial and some chlorophyte (L)-2-HAOX proteins display LOX activity, their homologs in land plants and mammals are specialized for either GOX or lHAOX activity. Because of the endosymbiotic origin of eukaryotes, it is likely that LOX activity represents the ancestral state of (L)-2-HAOX proteins. The presence of proteins encoding putative GOX and lHAOX activities in both plants and animals thus indicates a possible duplication and neofunctionalization of a LOX protein into GOX and lHAOX proteins in a common eukaryotic ancestor of animals and plants. However, this hypothesis does not easily account for the presence of LOX activity in some green algae, which prompted previous speculation of an endosymbiotic origin of plant GOX from the primary cyanobacterial endosymbiont at the base of plantae (Eisenhut et al. 2008; Hackenberg et al. 2011; Bauwe et al. 2012; Kern et al. 2013).

Here, we examine the evolution of the (L)-2-HAOX gene family central to plant photorespiration. We first show that both *A. thaliana* lHAOX indeed have substrate specificities comparable to those found in mammalian homologs. Contrary to the parsimonious assumption of a duplication and diversification of an ancestral (L)-2-HAOX in a common ancestor of plants and animals, phylogenetic analysis shows that duplication and diversification occurred independently at the base of deuterostomia and at the base of vascular plants. Although our data clearly indicates that plants did not acquire the ancestor of photorespiratory GOX from cyanobacteria via endosymbiosis, the homologous LOX in chlorophyta is indeed likely a remnant from the cyanobacterial endosymbiont that replaced the ancestral eukaryotic (L)-2-HAOX in this lineage.

Results and Discussion

Similar Substrate Specificities of Plant and Mammalian lHAOX

To analyze whether the predicted plant lHAOX display similar properties as their mammalian counterparts, lHAOX1 and lHAOX2 of the model plant *A. thaliana* (At3g14130 and At3g14150, respectively) were heterologously expressed in *Escherichia coli* and purified to homogeneity (fig. 1). The isolated recombinant proteins have expected molecular masses of approximately 43 kDa (lHAOX1: 39.9 kDa; lHAOX2: 40.1 kDa; plus an ~3 kDa fragment encoded by the expression vector; fig. 1*A*).

A substrate screening indicated that both recombinant proteins have broad substrate specificities (fig. 1*B*). lHAOX1 displays the highest activity with the long-chain fatty acid 2-hydroxydodecanoic acid (2-hydroxypalmitic acid; 0.27 $\mu mol\ min^{-1}\ mg^{-1}$) and has intermediate activity with 2-hydroxyhexanoic acid (2-hydroxycaproic acid), 2-hydroxyoctanoic acid (2-hydroxycaprylic acid), and the short-chain hydroxyacid L-lactate (fig. 1*B*). With much lower activity, it can also use glycolate, leucic acid, valic acid, and isoleucic acid as substrates. The second lHAOX protein of *A. thaliana*, lHAOX2, exhibits the highest activity with leucic acid (0.51 $\mu mol\ min^{-1}\ mg^{-1}$); it shows intermediate activity with 2-hydroxyhexanoic acid and 2-hydroxyoctanoic acid. lHAOX2 displays lower activity with 2-hydroxydodecanoic acid, valic acid, and isoleucic acid and extremely low activity with glycolate and L-lactate. Neither lHAOX1 nor lHAOX2 use 2-hydroxyhexadecanoic acid or D-lactate as substrates (fig. 1*B*).

Similar to *A. thaliana* lHAOX1, mammalian lHAOX were previously shown to mainly use the long-chain 2-hydroxy fatty acids 2-hydroxyoctanoate and 2-hydroxydodecanoate (Jones et al. 2000). As already suggested for human lHAOX, plant lHAOX1 may be involved in a general medium- and long-chain fatty acid catabolic pathway such as alpha-oxidation (Hutton and Stumpf 1971; Verhoeven et al. 1997, 1998). The presence of a second lHAOX with different substrate specificity in some plants indicates additional gene duplications and neofunctionalizations, with lHAOX2 possibly assuming new metabolic roles.

GOX and lHAOX in Plants and Animals Arose from an Ancestral Eukaryotic (L)-2-HAOX Gene

Based on the analysis of some algal, plant, and bacterial sequences, it was previously suggested that plant

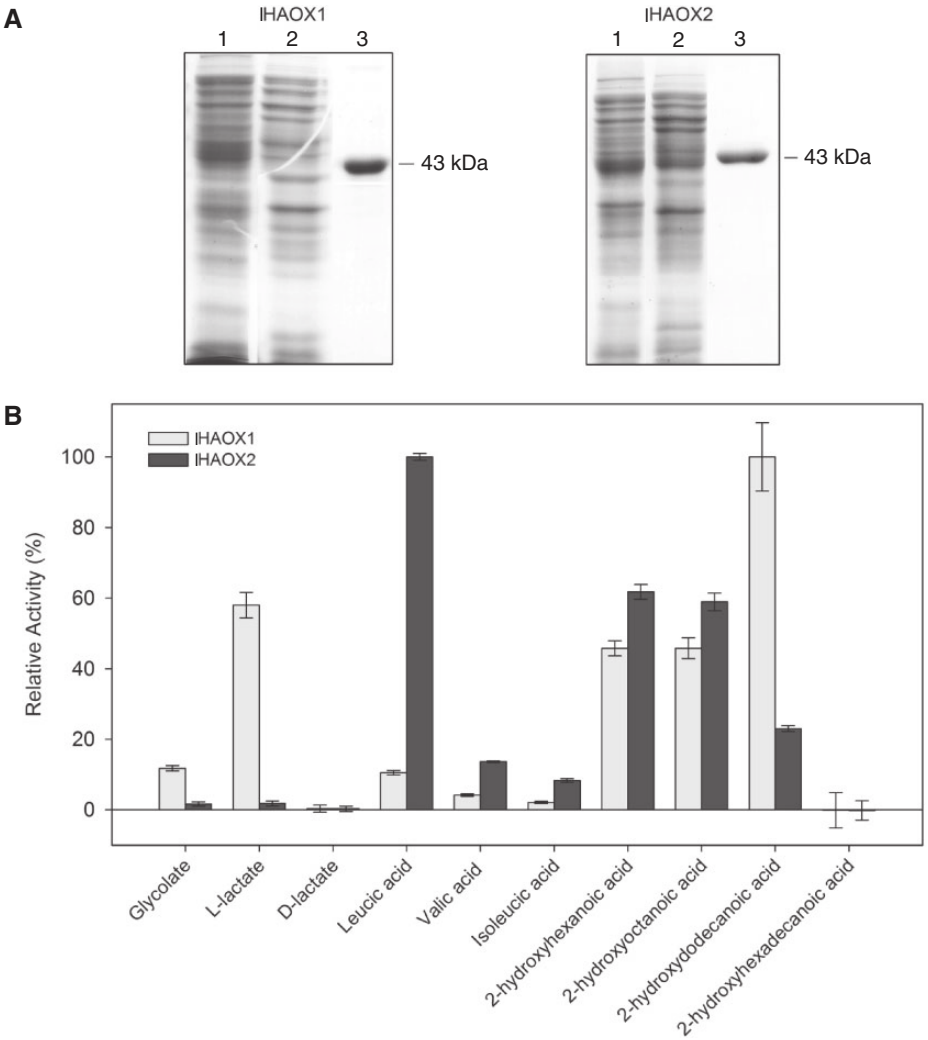


Fig. 1. Isolation and substrate screening of recombinant *Arabidopsis thaliana* lHAOX1 and lHAOX2. (*A*) Coomassie-stained sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of lHAOX1 und lHAOX2. Lane 1, 20 μg of *Escherichia coli* crude extract expressing lHAOX1 or lHAOX2 after IPTG induction; lane 2, wash of the Ni-NTA column; lanes 3, 3 μg of purified recombinant lHAOX1 or lHAOX2. (*B*) Substrate specificity of lHAOX1 und lHAOX2. lHAOX1 exhibits the highest activity with 2-hydroxydodecanoic acid (2-hydroxypalmitic acid; 0.27 μmol min^{-1} mg^{-1}) and has intermediate activity with 2-hydroxyhexanoic acid (2-hydroxycaproic acid), 2-hydroxyoctanoic acid (2-hydroxycaprylic acid), and the short-chain hydroxy acid L-lactate. lHAOX2 has the highest activity with leucic acid (0.51 μmol min^{-1} mg^{-1}) and has intermediate activity with 2-hydroxyhexanoic acid and 2-hydroxyoctanoic acid. Shown are mean values ± SEM of the activities relative to the best substrate of at least two independent experiments.

(L)-2-HAOX genes are derived by endosymbiotic gene transfer from the genome of the cyanobacterial endosymbiont that gave rise to photosynthetic eukaryotes (Eisenhut et al. 2008; Hackenberg et al. 2011). Given that (L)-2-HAOX proteins with GOX and lHAOX activity are found in both plants and animals, it appears more parsimonious to postulate an ancestral ortholog in the common ancestor of these eukaryotic lineages. To test this hypothesis, we built a comprehensive data set of (L)-2-HAOX sequences, using 12 experimentally confirmed (L)-2-HAOX to construct a hidden Markov model (HMM, see Materials and Methods) (Eddy 2011), and then selecting matching sequences (at most one per species) from RefSeq (Pruitt et al. 2012) as well as a plant-specific database (Proost et al. 2009) (Materials and Methods, supplementary table S1, Supplementary Material online). On the basis of the HMM profile, we aligned the protein sequences (supplementary fig. S1*A*, Supplementary Material online) and reconstructed a gene tree using Bayesian and maximum-likelihood (ML) methods (supplementary fig. S2, Supplementary Material online). To reduce the risk of long branch attraction artifacts, we repeated our analysis after excluding sequences that failed Tajima's relative rate test (Tajima 1993) (fig. 2; alignment in supplementary fig. S1*B*, Supplementary Material online). This removed all fungal sequences from our data set, indicating accelerated evolution of (L)-2-HAOX proteins in fungi. Accelerated evolution in fungi is consistent with previous work, which showed that the only functionally characterized fungal gene homologous to *(L)-2-HAOX* encodes a mitochondrial protein that differs substantially in molecular function from the (L)-2-HAOX proteins analyzed in this study (Guiard 1985).

Following terminology for unrooted trees suggested by Wilkinson et al. (2007), we use "clan" to refer to a group of

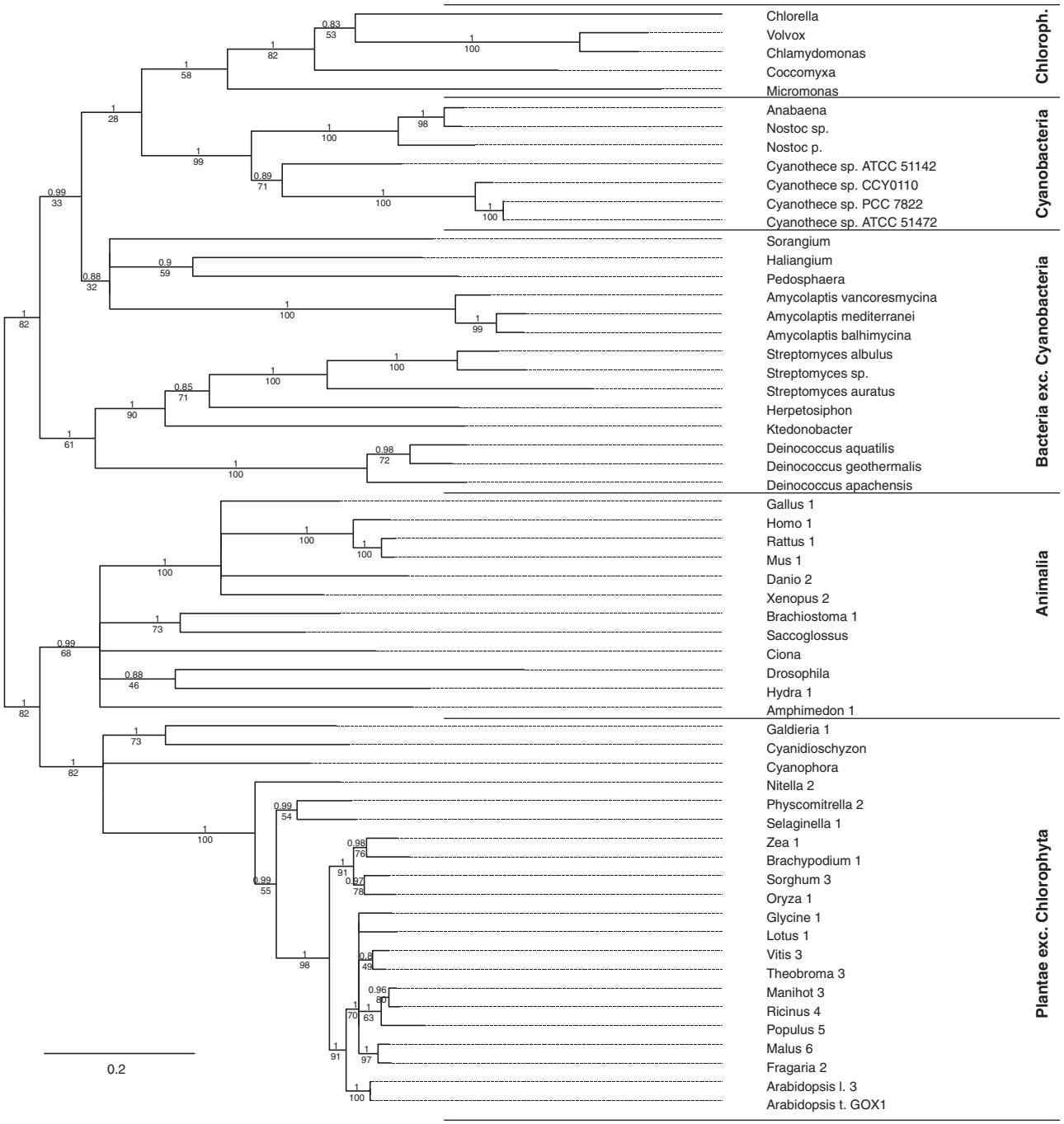


Fig. 2. (L)-2-HAOX protein family tree based on Bayesian phylogenetic inference, including the best matching sequence from each species. Eukaryotic (L)-2-HAOX are monophyletic with the exception of chlorophyta, which cluster with cyanobacteria and other bacterial sequences. Numbers above branches show posterior probabilities; numbers below branches report bootstrap support from an ML analysis of the same alignment. Species names and gene identifiers are listed in supplementary table S1, Supplementary Material online. Branches with posterior probabilities < 0.8 were deemed unreliable and were collapsed (see supplementary fig. S3, Supplementary Material online, for the uncollapsed tree). The tree shown was calculated excluding fast-evolving proteins; see supplementary table S1, Supplementary Material online, for the results of the relative rate tests, and supplementary figure S2, Supplementary Material online, for the tree including all sequences. Species names and gene identifiers are listed in supplementary table S1, Supplementary Material online.

sequences that form a connected subtree after removal of one branch (i.e., a clan is one side of a split and is a monophyletic group in some rootings of the tree). Although we cannot specify the exact position of the root, we assume that a clan of bacteria (or at least one bacterium) forms an outgroup to the other sequences. We only use the term "monophyletic group" for a clan if the corresponding group of sequences is monophyletic across all such rootings.

As expected from the broad phylogenetic distribution of (L)-2-HAOX across eukaryotic kingdoms, we found that eukaryotic sequences—except chlorophyta—form a

monophyletic group (Bayesian posterior probability PP = 1.0, ML bootstrap support BS = 82% after exclusion of fast-evolving sequences, fig. 2; PP_{all} = 0.99 including all sequences, supplementary fig. S2, Supplementary Material online). In agreement with established phylogenetic relationships, this eukaryotic clade is further split into animals and fungi (forming the clade of opisthokonts) and plantae *sensu lato* (supplementary fig. S2, Supplementary Material online). Known substrate specificities of (L)-2-HAOX proteins are consistent with the phylogenetic grouping of chlorophyta with cyanobacteria and other bacteria (LOX activity) and of animals with plants (GOX/lHAOX activities). We conclude that plant, animal, and fungal (L)-2-HAOX proteins are derived from an (L)-2-HAOX encoding gene in a common eukaryotic ancestor.

This ancestral (L)-2-HAOX was most likely already located to peroxisomes, as we find peroxisomal targeting signal type 1 (PTS1) motifs (Reumann et al. 2004) in almost all plant, animal, and fungi (L)-2-HAOX (supplementary fig. S4, Supplementary Material online).

Convergent Evolution in Vascular Plants and Deuterostomia

Does the diversification into GOX and lHAOX functions in plants and animals go back to a gene duplication in a common ancestor of these kingdoms? To examine this question, we assembled an extended data set of genes encoding (L)-2-HAOX proteins from eukaryotes, cyanobacteria, and other bacteria (Materials and Methods, supplementary table S1 and fig. S4, Supplementary Material online). Figure 3 shows the phylogenetic tree built from the resulting protein sequence alignment after the removal of fast-evolving proteins (Tajima 1993), whereas supplementary figure S5, Supplementary Material online, displays the tree including all sequences.

We find monophyly of animal GOX and lHAOX (PP = 1.0, BS = 89%, fig. 3; PP_{all} = 0.76 for the monophyly of animal (L)-2-HAOX, and PP_{all} = 1.0 for the monophyly of opisthokonts (animal + fungi) (L)-2-HAOX, supplementary fig. S5, Supplementary Material online). We also find monophyly of streptophyta GOX and lHAOX (PP = 1.0, BS = 99%, fig. 3; PP_{all} = 1.0, supplementary fig. S5, Supplementary Material online). This separation of plant and animal (L)-2-HAOX proteins clearly contradicts the hypothesis of an early duplication and diversification in the common ancestor of plants and animals. Instead, figure 3 indicates a striking case of convergent evolution, with independent duplications and convergent diversifications of (L)-2-HAOX in both eukaryotic kingdoms.

The duplication leading to animal GOX and lHAOX occurred before the common ancestor of all vertebrates represented in figure 3 (for the two groups of sequences that cluster with human GOX and lHAOX, respectively: PP = 1.0, BS = 100%, and PP_{all} = 1.0). The two basal deuterostomia included in our study (*Branchiostoma floridae* and *Saccoglossus kowalevskii*) each also contain more than one (L)-2-HAOX sequence, forming two clusters, one of which is monophyletic with the group containing vertebrate lHAOX (PP = 0.97, BS = 64%, PP_{all} = 0.99). In contrast, all more basal animals (Srivastava et al. 2010) in figure 3 possess only one (L)-2-HAOX gene (the sponge *Amphimedon queenslandica*, the cnidarian *Hydra magnipapillata*, and the fly *Drosophila melanogaster*). Furthermore, all fungal species in supplementary figure S5, Supplementary material online, encode only one (L)-2-HAOX protein. This indicates that the gene duplication leading to the subfamilies containing human GOX and lHAOX, respectively, occurred early in the history of deuterostomia. That the duplication occurred after the divergence of deuterostomia from the common ancestor with *D. melanogaster* is supported by the observation that all arthropod species in RefSeq contain at most one (L)-2-HAOX protein that matches our selection criteria (data not shown).

The corresponding gene duplication in plants occurred before the diversification of higher plants (spermatophyta), indicated by two monophyletic groups of spermatophyta containing *A. thaliana* GOX and lHAOX proteins, respectively (*A. thaliana* GOX clade: PP = 1.0, BS = 78%, PP_{all} = 1.0; *A. thaliana* lHAOX clade: PP = 1.0, BS = 100%, PP_{all} = 1.0; fig. 3 and supplementary fig. S5, Supplementary Material online). The spikemoss *Selaginella moellendorffii* encodes three (L)-2-HAOX proteins, two of which cluster basal to the group containing *A. thaliana* lHAOX (PP = 1.0, BS = 93%, PP_{all} = 1.0). In contrast, all more basal plantae in figure 3 possess only one *(L)-2-HAOX* gene or paralogous gene cluster (the glaucophyte *Cyanophora paradoxa*, the rhodophyta *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, the charophyte *Nitella hyaline*, and the moss *Physcomitrella patens*). This indicates that the gene duplication leading to the two subfamilies containing *A. thaliana* GOX and lHAOX likely occurred early in the history of tracheophyta.

Chlorophyta Replaced Eukaryotic GOX/lHAOX with Cyanobacterial LOX

The relative position of (L)-2-HAOXs in plantae is largely consistent with the accepted phylogeny (Martin et al. 1998), with glaucophyta and rhodophyta (red algae) basal to charophyta, mosses, and vascular plants (figs. 2 and 3). Surprisingly, the chlorophyta (L)-2-HAOX are not a sister group to charophyta sequences; instead, they form a well-supported clan together with bacterial sequences to the exclusion of all other plantae (fig. 2: PP = 1.0, BS = 82%, PP_{all} = 0.99; fig. 3: PP = 1.0, BS = 73%, PP_{all} = 1.0).

The grouping of (L)-2-HAOX sequences from chlorophyta and bacteria is consistent with the shared LOX activity of these evolutionarily distant clades (Hackenberg et al. 2011). This pattern is unlikely to be an artifact of convergent evolution (or shared conservation) of the substrate specificity, as chlorophyta and cyanobacteria form a clan also in a tree based on an alignment that excludes regions involved in the substrate specificity (i.e., the 11 alignment positions identified as directly involved in substrate binding, plus all positions within 6 Å in the protein structure; supplementary table S2, Supplementary Material online).

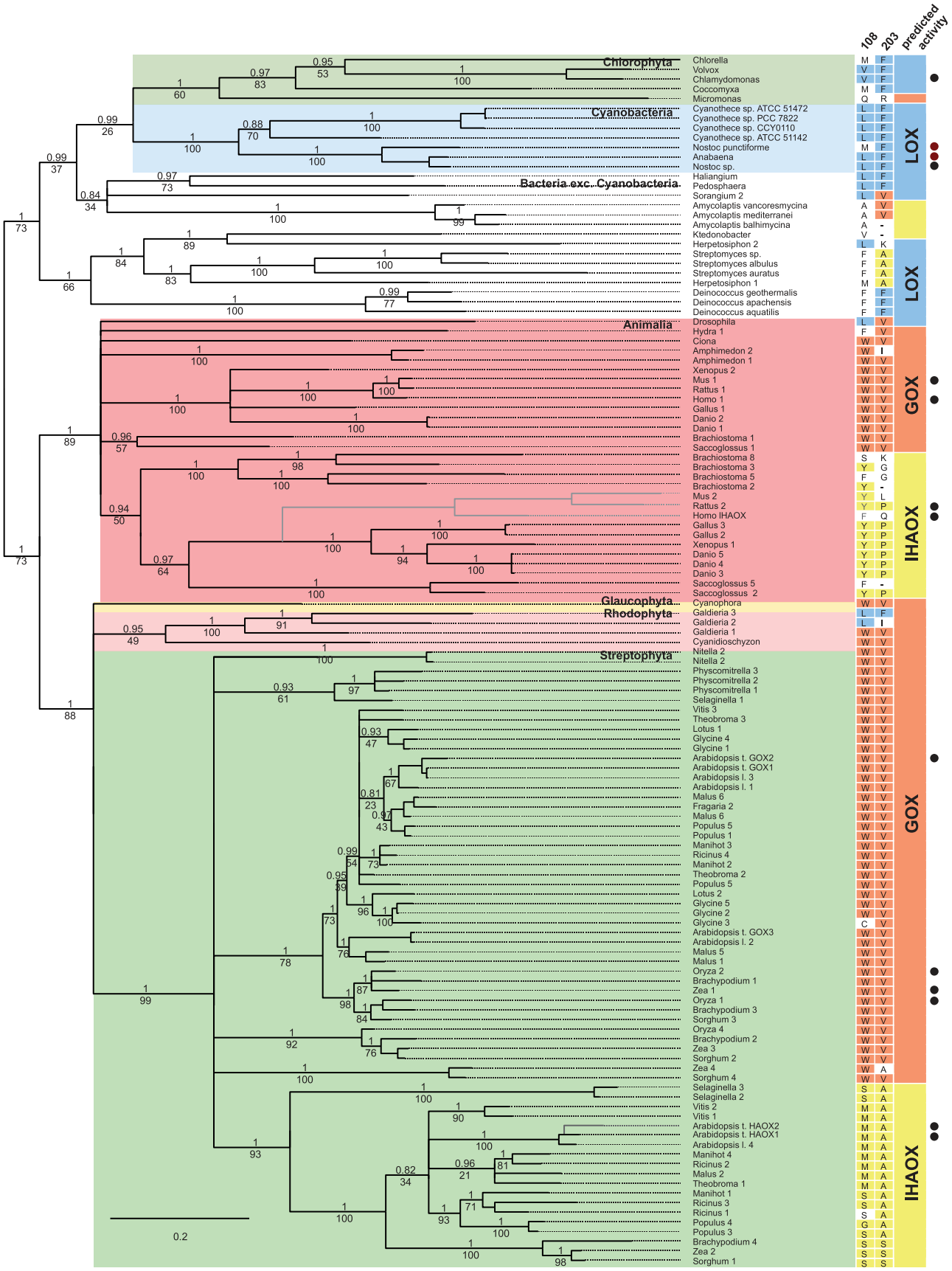


FIG. 3. Extended (L)-2-HAOX protein family tree based on Bayesian phylogenetic inference, together with measured and predicted substrate specificities. Diversification into GOX and lHAOX subfamilies occurred prior to the divergence of deuterostomia and independently prior to the divergence of vascular plants. Chlorophyta form a clan together with cyanobacterial and other bacterial LOX, to the exclusion of all other eukaryotic sequences. Numbers above branches show posterior probabilities; numbers below branches report bootstrap support from an ML analysis of the same alignment.

(continued)

Instead, differential gene loss in the chlorophyta versus the glaucophyta, rhodophyta, and charophyta lineages appears to be the most parsimonious explanation for this grouping (fig. 4). The origin of algae and plants can be traced back to the endosymbiosis of a cyanobacterium in a eukaryotic host, with the cyanobacterium later evolving into plastids (Dagan et al. 2013). It is thus noteworthy that the trees in figures 2 and 3 indicate cyanobacteria as the bacterial clade that forms the closest clan with chlorophyta. We thus hypothesize that the LOX in chlorophyta was originally encoded on the endosymbiotic plastid genome and was later transferred to the host genome, a process termed endosymbiotic gene transfer (Martin et al. 1998). According to our differential gene loss hypothesis, this LOX was retained in the chlorophyta, whereas it was lost independently in glaucophyta, rhodophyta, and charophyta (fig. 4). Conversely, chlorophyta lost the ancestral eukaryotic *(L)-2-HAOX* gene (fig. 4). Such differential gene loss is frequently found among genes "duplicated" through endosymbiosis (Martin et al. 1998). However, horizontal gene transfer between free-living bacteria and algae has been observed (Schönknecht et al. 2013), and we thus cannot rule out a bacterial origin of LOX in chlorophyta independent from the acquisition of plastids.

The activity of a GOX protein is essential for photorespiration in all algal and plant lineages except chlorophyta, which, as cyanobacteria, use GlcDH for this biological function (Nelson and Tolbert 1970; Beezley et al. 1976; Nakamura et al. 2005; Stabenau and Winkler 2005; Eisenhut et al. 2008). The mitochondrial glycolate pathway (using GlcDH) is energetically favorable, as it produces ATP, but it also has a low capacity and thus glycolate is excreted under conditions that favor high rates of glycolate synthesis (Stabenau and Winkler 2005). Because of the presence of a carbon concentrating mechanism, cyanobacteria and chlorophyta have lower rates of photorespiratory glycolate synthesis (Giordano et al. 2005; Stabenau and Winkler 2005; Huege et al. 2011). Plants, as well as algae other than chlorophyta, cope with large rates of glycolate synthesis by the use of GOX, as this enzyme has a higher catalytic activity than GlcDH (Stabenau and Säftel 1982; Macheroux et al. 1992); the high turnover rate of GOX may have been crucial for the evolution of land plants, as these had to grow in the presence of substantial atmospheric oxygen concentrations.

Chlorophycean LOX has similar kinetic properties as cyanobacterial LOX (Hackenberg et al. 2011). In cyanobacteria, LOX may serve as an O_2-scavenging enzyme, protecting nitrogenase (Hackenberg et al. 2011). In chlorophyta, LOX is likely located to peroxisomes, as the proteins possess PTS1 motifs (Reumann et al. 2004). The cellular function of the chlorophycean LOX is not known; however, the presence of this enzymatic activity indicates L-lactate metabolism in chlorophycean peroxisomes.

Substrate Specificity of Plant and Animal GOX Is Mediated by Similar Amino Acids

Given that animal and plant (L)-2-HAOX gene families independently diversified into GOX and lHAOX subfamilies, it is interesting to examine whether similar substrate specificities were achieved through the same amino acid replacements. To identify residues crucial for substrate binding, homology models were generated using the crystal structure of spinach GOX1 as a template, superimposed on the structure of human GOX1 complexed with glyoxylate (pdb-code 2RDU). We identified 11 amino acids directly involved in substrate binding: Y24, W108, Y129, R164, F172, L199, A203, A204, K230, R254, and R257 (fig. 5; residue numbers correspond to the *A. thaliana* GOX1 sequence). We used a support vector machine (SVM) to predict substrate specificities of all (L)-2-HAOX sequences in figure 3, assuming that amino acids within a radius of 6 Å around the 11 identified positions are likely to contribute to substrate interactions (Röttig et al. 2010). As a training set for the SVM active site classification, we used 14 sequences with experimentally determined substrate specificities (fig. 3; for details, see Materials and Methods).

The resulting predictions are shown in figure 3. LOX activity is predicted in all cyanobacteria and chlorophyta except *Micromonas*. If the tree is rooted at a bacterial clan (or single bacterial sequence), then the predicted lHAOX activities in deuterostomia and vascular plants correspond to monophyletic groups. (L)-2-HAOX proteins in glaucophyta, rhodophyta, charophyta, and moss are predicted to exhibit GOX activity, consistent with the involvement of GOX in photorespiration. (L)-2-HAOX proteins of the most basal animals (*Amphimedon* and *Hydra*) are also predicted to have GOX activities (fig. 3). Thus, all analyzed (L)-2-HAOX proteins found in animal outgroups of deuterostomia (except *Drosophila*) and plantae outgroups of vascular plants are predicted to exhibit GOX activity. If we assume that all substrate changes were equally likely, then this distribution of predicted activities suggests that the ancestral eukaryotic (L)-2-HAOX already used glycolate as its main substrate (fig. 4). This scenario requires at least four changes in substrate specificity in the evolution of the *(L)-2-HAOX* gene family: in the ancestral

Fig. 3. Continued

The tree was estimated excluding fast-evolving sequences (see supplementary table S1, Supplementary Material online), among them human lHAOX together with close mammalian homologs and *A. thaliana* lHAOX2. Gray branches indicate the position of these lHAOX proteins according to the full analysis (supplementary fig. S5, Supplementary Material online). The two columns to the right of sequence names show the amino acids in the alignment position corresponding to *A. thaliana* GOX1 residues V108 and W203. The colors in the bar to the right indicate predicted main activity of the (L)-2-HAOX proteins (LOX, GOX, and lHAOX); symbols to the right of the colored bar indicate the sequences used for building the HMM model (black + red dots) and for training the SVM (black dots). Branches with posterior probabilities <80% were deemed unreliable and were collapsed (see supplementary fig. S6, Supplementary Material online, for the uncollapsed tree). Species names and gene identifiers are listed in supplementary table S1, Supplementary Material online.

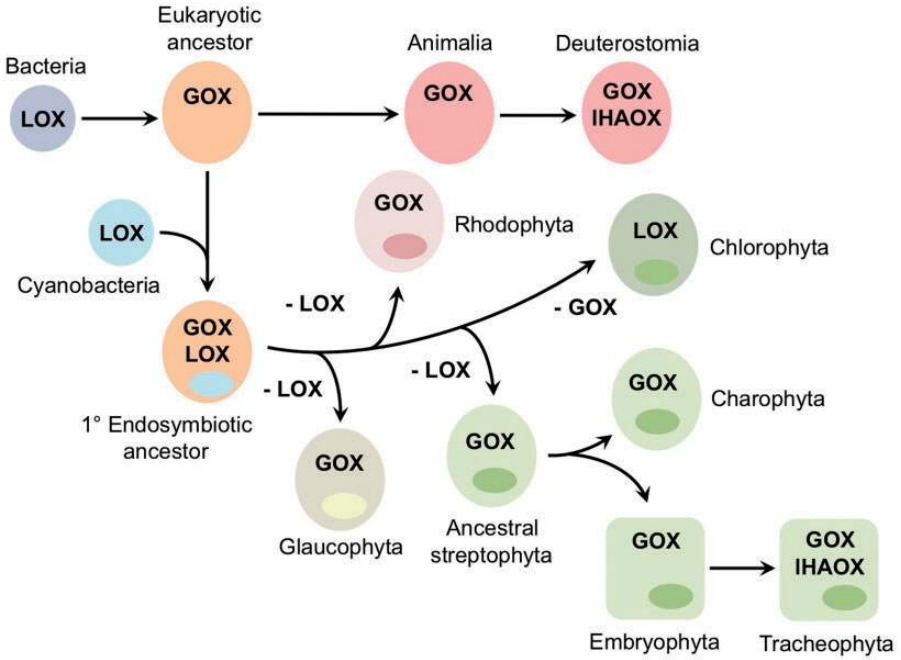


FIG. 4. Hypothesis for the evolution of the *(L)-2-HAOX* gene family. The common eukaryotic ancestor may have already possessed GOX activity. An additional LOX sequence was acquired by the common ancestor of plants and algae from the cyanobacterial endosymbiont. Although most algal lineages (glaucophyta, rhodophyta, and charophyta) independently lost the cyanobacterial *LOX*, chlorophyta replaced eukaryotic *GOX* with cyanobacterial *LOX*. Duplication and diversification of an ancestral *GOX* occurred independently in ancestral deuterostomia and vascular plants (tracheophyta), giving rise to two *lHAOX* subfamilies.

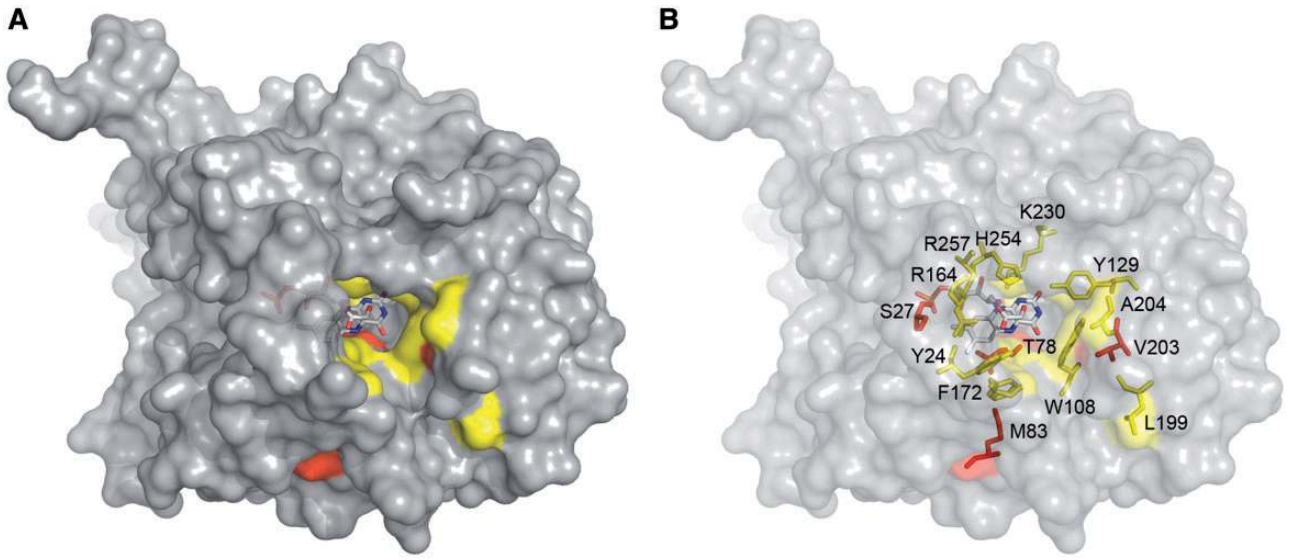
eukaryote to GOX (as it is likely that the bacterial ancestor of eukaryotes encoded a LOX), later to lHAOX in both ancestral deuterostomia and vascular plants, plus a reversion to LOX in an ancestor of *Drosophila*. If we postulated a function other than GOX activity in the ancestral eukaryote, additional substrate specificity changes would be required to explain the distribution of activities in figure 3, making alternative scenarios less parsimonious.

Analysis of the alignment positions across all proteins of our extended data set indicates only four amino acid positions that allow a limited discrimination of (L)-2-HAOX with different predicted substrate specificities (fig. 5 and supplementary fig. S4, Supplementary Material online). Amino acid position 27 (numbered according to the *A. thaliana* GOX1 sequence) is generally glycine or alanine in lHAOX sequences, whereas the GOX sequences contain mostly serine, and LOX sequences contain serine, threonine, or glycine at this position. In most cases, amino acid position 78 is methionine in LOX sequences, whereas lHAOX sequences contain threonine, and GOX sequences contain threonine or serine. Amino acid position 83 is methionine in GOX sequences, whereas LOX sequences contain leucine, and lHAOX sequences contain leucine or isoleucine. Finally, amino acid position 203 shows a near-perfect correlation with predicted substrate specificities, with the exception of animal lHAOX sequences. All LOX sequences except *Micromonas* contain phenylalanine at this position. Almost all GOX proteins contain valine; it is interesting to note that exceptions occur only in cases where the same species also encodes another predicted GOX that contains the consensus valine at this position (*Amphimedon*, *Galdieria*, and *Zea*). Plant lHAOX sequences present alanine in all eudicots and serine in all monocots. In contrast, animal lHAOX sequences do not present a specific amino acid at this position (fig. 3).

Amino acid positions 108 and 203 were previously identified as important for substrate preference based on site-directed mutagenesis (Stenberg et al. 1995; Hackenberg et al. 2011). Figure 3 lists the amino acid at the corresponding alignment position for all proteins in our data set. However, amino acid position 108 (fig. 3), as well as all other amino acid positions identified by the SVM (supplementary fig. S7, Supplementary Material online), shows less clear patterns than position 203 when examined in the extended data set. Thus, although the identity of a few amino acids is routinely used to predict the substrate specificities of (L)-2-HAOX sequences (Reumann et al. 2004; Hackenberg et al. 2011), our analysis clearly indicates that this method is not generally reliable. Moreover, although the amino acid at position 203 contains enough information to predict most substrate specificities of evolved (L)-2-HAOX sequences, a single amino acid change may not be sufficient to confer different substrate specificities. The SVM identified 56 amino acid positions that could distinguish between at least two of the three substrates with an accuracy of more than 80% and 20 positions reaching an accuracy of more than 90%. It is thus likely that substrate specificity is conferred by the interaction of tens of amino acids distributed along the sequence of this gene family.

Conclusions

Chlorophyta are the only sequenced species of plantae that appear to have inherited an (L)-2-HAOX (LOX) enzyme from cyanobacteria. All other eukaryotic (L)-2-HAOX sequences



C

Amino acid position*	S27	T78	M83	V203
GOX	S	T, S	M	V
lHAOX	G,A	T	L,I	P,A,S
LOX	S,T	M	L	F

*corresponding to *A. thaliana* GOX1

Fig. 5. Homology model of *Arabidopsis thaliana* GOX1 based on the crystal structure of spinach GOX (PDB code 1GOX). The model is shown in surface representation facing the active site with bound glyoxylate and the FMN cofactor. Conformations and arrangement of substrates at the active sites were derived from superposition with human GOX (PDB code 2RDU). Residues in close proximity of active site that are potentially involved in substrate binding are shown in yellow. Amino acids highlighted in red discriminate (L)-2-HAOX with different predicted substrate specificities; the corresponding amino acids are listed in the associated table. (*A*) Surface representation showing solvent-accessible residues involved in substrate binding (yellow) and substrate specificity (red). (*B*) Transparent surface representation showing side chain orientation and residue number of the amino acids that are potentially involved in substrate binding (yellow) and in substrate specificity (red). (*C*) Table showing the amino acids that allow a limited discrimination of (L)-2-HAOX with different predicted substrate specificities.

can be traced back to a single enzyme present in the last common eukaryotic ancestor. What substrate specificity did this ancestral enzyme have? Given that all (L)-HAOX from animalia and streptophyta in our data set except *Drosophila* have either predicted GOX or lHAOX activity and that the (L)-2-HAOX proteins from species basal to deuterostomia and insects (sponge, *Hydra*) and basal to tracheophyta (glaucophyta, rhodophyta, charophyta) have predicted GOX activities, it is most parsimonious to hypothesize that the common eukaryotic ancestor already used glycolate as its main substrate (figs. 3 and 4). We thus conclude that eukaryotic (L)-2-HAOXs—with the exception of chlorophycean LOX—are likely derived from an ancestral GOX.

It is noteworthy that the duplication of GOX and neofunctionalization into the lHAOX subfamily occurred independently in the two eukaryotic clades that contain the morphologically and physiologically most complex species: deuterostomia and tracheophyta (higher plants). We have shown that *A. thaliana* lHAOX1—similar to animal lHAOX—preferentially oxidizes long-chain 2-HA, although *A. thaliana* lHAOX2 preferentially oxidizes medium-chain 2-HA. It is thus likely that lHAOX in both plant and animal lineages is involved in the conversion or degradation of 2-hydroxy acids produced during the metabolism of fatty acids or amino acids.

Although plants synthesize the corresponding substrates themselves, it is conceivable that animals use lHAOX to degrade long- and medium-chain 2-hydroxy acids ingested from plants or from intermediate members of the food chain (Verhoeven et al. 1998). Involvement of animal lHAOX in the degradation of food components or xenobiotic compounds would also be consistent with the localization of lHAOX in liver and kidney (Belmouden and Lederer 1996; Jones et al. 2000). Alternatively, the independent evolution of lHAOX activity in the peroxisomes of plants and animals may indicate a requirement for additional metabolic pathways in complex multicellular organisms, for example, the oxidative degradation of medium- and long-chain fatty acids and amino acids.

Previous speculations suggested that plants convert photorespiratory glycolate with the aid of a GOX derived from a cyanobacterial LOX, acquired through the primary endosymbiosis at the origin of plantae (Hackenberg et al. 2011; Kern et al. 2013). Moreover, it was suggested that plant GOX with its high reactivity toward glycolate evolved only when land plants appeared (Bauwe et al. 2012). However,

the ancestral eukaryotic (L)-2-HAOX was most likely a GOX, that is, already had a high specificity for glycolate. Furthermore, it was most likely already located to peroxisomes; indeed, GOX activity is expected to coincide with peroxisomal localization, as GOX produces H_2O_2, which is detoxified in this organelle. Recruitment of the ancestral eukaryotic GOX by the ancestors of archaeplastida would have been evolutionarily much faster than the change of both substrate specificity and localization required for the recruitment of the cyanobacterial LOX. This is indeed what we find: the GOX of plantae is an ancient peroxisomal eukaryotic enzyme recruited into photorespiration.

Materials and Methods

Generation of Vectors for the Heterologous Expression of *A. thaliana* lHAOX1 and lHAOX2

To obtain *lHAOX1* and *lHAOX2* full-length coding sequences, total RNA was isolated from 100 mg siliques and leaf, respectively, and converted into first strand cDNA using Pfu-Polymerase (Thermo Scientific, Fermentas). Because lHAOX1 and lHAOX2 are highly homologous, two polymerase chain reaction (PCR) steps were conducted. The first PCR was performed using 5′-UTR-primers in combination with the 3′-cloning primers, followed by a second PCR specific for the coding sequences. The primers used were as follows: 1) *lHAOX1*, H1-5′-UTR-fw (5′-TCTCCGACAGGACCGACTCT GTT-3′) and H1-rv (5′-GGATCCTCAGAGCATAGATTTAATT C-3′) for the first PCR, and H1-fw (5′-GGATCCGGATCAAATC GTTAACGTGG-3′) and H1-rv for the second PCR; 2) *lHAOX2*, H2-5′-UTR-fw (5′-GCCTGACATTCTTGGTTCTGCTTCTGG-3′) and H2-rv (5′-GGATCCTCAGAGCATAGAGTGAAGTC-3′) for the first PCR, and H2-fw (5′-GGATCCGGATCAAATC GTTAATGTGG-3′) and H2-rv for the second PCR. The final PCR products were cloned into pCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA). The inserts were cut out with BamHI and ligated into a BamHI linearized pET-16 b vector (Novagen, Merck KGaA, Darmstadt), which yielded fusion proteins with an N-terminal His-tag.

Expression and Purification of Recombinant Proteins

The expression vectors, pET-16b-lHAOX1 and pET-16b-lHAOX2, were transformed into *E. coli* BLR DE3 pLysS strain (Novagen). The transformed cells were grown overnight in LB medium at 37 °C and agitation at 220 rpm in the presence of 50 μg/ml carbenicillin, 75 μg/ml chloramphenicol, and 5 μg/ml tetracycline. Expression cultures were inoculated with the overnight grown cultures in a 1:50 relationship and grown at 37 °C until the culture reached an absorbance at 600 nm of 0.6. Protein expression was induced by the addition of 0.1 mM isopropyl β-D-thiogalactopyranoside. In the case of lHAOX1, the culture was transferred to room temperature and grown for 20 h before harvesting. In the case of lHAOX2, the culture was grown at 37 °C for 2–3 h before harvesting. Cells were harvested by centrifugation at 4,000 rpm for 15 min at 4 °C. Cells were resuspended in 20 mM Tris-HCl (pH 8.0) containing 0.5 mg/ml phenylmethylsulfonyl fluoride, sonicated, and centrifuged at 10,000 rpm for 15 min at 4 °C to remove debris. The proteins were further purified from the supernatant, using immobilized metal ion chromatography on Ni^{2+}-nitrilotriacetic acid-agarose (Ni^{2+}-NTA, Qiagen). The Ni^{2+}-NTA column was equilibrated and washed with buffer A (consisting of 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 1 mM DTT) containing 40 mM imidazole. After two further washing steps with buffer A containing 80 and 100 mM imidazole, respectively, the proteins were eluted using buffer A containing 300 mM imidazole. The eluted proteins were desalted using a NAP5 Sephadex G-25 DNA Grade column (Amersham Pharmacia Biotech AB, Uppsala, Sweden) eluted with 20 mM Tris-HCl pH 8.0 and 1 mM DTT.

Polyacrylamide Gel Electrophoresis and Protein Determination

Protein concentration was determined according to the method of Bradford (Bradford 1976), using bovine serum albumin as standard. Denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed using 12.5% (w/v) polyacrylamide gels according to Laemmli (Laemmli 1970). Proteins were visualized with Coomassie Blue.

Enzymatic Activities and Substrate Screening

Enzymatic activities of *A. thaliana* lHAOX1 and lHAOX2 were assayed spectrophotometrically using 100 mM Tris-HCl (pH 7.5) and 50 mM K_2HPO_4 (pH 6.5), respectively. The reaction mixtures also contained 0.5 mM ethylenediaminetetraacetic acid, 0.05 mM flavin mononucleotide (FMN), 5 mM $MgCl_2$, 4 mM phenylhydrazine, and 2.5–5.2 μg HAOX1 or 3.5–3.6 μg HAOX2 in a final volume of 0.2 ml. The substrates tested were glycolate, D- and L-lactate, leucic acid, valic acid, and isoleucic acid (dissolved in water) at a concentration of 5 mM, and 2-hydroxyhexanoic acid, 2-hydroxyoctanoic acid, 2-hydroxydodecanoic acid, and 2-hydroxyhexadecanoic acid (dissolved in ethanol) at a concentration of 0.5 mM. The activities were determined at 25 °C by monitoring the change in absorption at 324 nm due to the production of a phenylhydrazone. The extinction coefficient of glyoxylate-phenylhydrazone ($\varepsilon = 17\ cm^{-1}mM^{-1}$) was used for all tested substrates. A baseline was determined by monitoring the absorption of the reaction mix for 3.5 min before adding the substrate. Activities were determined with at least two different enzyme batches, each containing at least triplicate determinations. The highest activity of each enzyme was set at 100%, and the other activities were calculated relative to this value.

Collection of (L)-2-HAOX Sequences

To assemble a representative data set of (L)-2-HAOX sequences, we first obtained the sequences of 12 (L)-2-HAOX proteins with experimentally verified substrate specificities (marked by asterisks in fig. 3): *Nostoc* sp. (Hackenberg et al. 2011) and *C. reinhartii* (Hackenberg et al. 2011) LOX; human GOX and lHAOX (Jones et al. 2000); mouse GOX (Kohler et al. 1999); rat lHAOX (Belmouden and Lederer 1996); *A. thaliana*

GOX2 (Hackenberg et al. 2011); *A. thaliana* lHAOX1 and lHAOX2 (this work); *Oryza sativa* GOX (GLO1 and GLO3) (Zhang et al. 2012); and *Zea mays* GOX (Zelitch et al. 2009). To expand the bacterial training set, we additionally included two sequences from the *Nostoc–Anabaena* clade closely related to *Nostoc* sp. From these 14 (L)-2-HAOX protein sequences, we constructed an HMM) using the program HMMbuild from the HMMER3 package (Eddy 2011).

We restricted a database of 18 plants (PLAZA 2.5) (Proost et al. 2009) to those sequences with at least 45% amino acid identity to *A. thaliana* GOX1 in a Needleman–Wunsch global alignment and with a length that deviated less than 20% from *A. thaliana* GOX1. We then used the HMM model in conjunction with the program HMMsearch from the HMMER3 package to select sequences from this data set that matched the model with *e* value $< 10^{-130}$, retaining the best matching sequence from each species. Because not all sequences known from the literature were found in this database, we replaced sequences for some species by homologs from other databases. Additionally, we searched for the best homologous sequences in selected completely sequenced algae, animals, fungi, cyanobacteria, and other bacteria obtained from RefSeq (Pruitt et al. 2012) (supplementary table S1, Supplementary Material online). For the charophyte *N. hyaline*, we identified the best matching homologous sequence in the National Center for Biotechnology Information (NCBI) expressed sequence tag (EST) Database using TBlastN (Altschul et al. 1997). We only retained sequences that showed a Blast query coverage more than 70% of the alignment. From this data set, we again selected those sequences that matched the HMM profile with an *e* value $< 10^{-130}$, retaining only the best matching sequence from each species. Previous work described three human (L)-2-HAOX sequences (Jones et al. 2000). However, the gene designated HAOX3 in Jones et al. (2000) is 96% identical to mouse lHAOX2 (Refseq ID NP_062418.3) and has been marked as a potential contamination in the NCBI EST database; we thus excluded this sequence from further analysis. No sequences from archaea or fungi matched the HMM profile below our cutoff. To still include fungi in our analyses, we additionally selected the five best-scoring fungal sequences above the cutoff. The resulting data set consisted of 65 sequences from different bacterial and eukaryotic species (supplementary table S1, Supplementary Material online).

We created a second, expanded data set using the same pipeline as described earlier but now including all sequences from each species that passed our filters. If two identical amino acid sequences were found in the same species, one of them was disregarded. This resulted in a second data set of 137 (L)-2-HAOX sequences.

Phylogenetic Analyses

For the two sets of 65 and 137 (L)-2-HAOX protein sequences, we constructed multiple sequence alignments using the above HMM profile with HMMalign from the HMMER3 package (Eddy 2011) (supplementary figs. S1 and S3, Supplementary Material online).

According to the Akaike information criterion in Prottest3 (Darriba et al. 2011), for both data sets, the most appropriate evolutionary model implemented in MrBayes (Ronquist and Huelsenbeck 2003) is WAG + I + G (i.e., the substitution matrix from Whelan and Goldman [2001], a proportion of invariant sites, and a gamma model of rate heterogeneity); the most appropriate model implemented in RAxML (Stamatakis 2006) is LG + I + G (i.e., the substitution matrix from [Le and Gascuel 2008], a proportion of invariant sites, and a gamma model of rate heterogeneity).

We reconstructed trees using the Bayesian approach implemented in MrBayes (Ronquist and Huelsenbeck 2003), running 1,000,000 generations for each data set, and disregarding the first 25% of samples as burn-in (figs. 2 and 3, and supplementary figs. S2, S3, S5, and S6, Supplementary Material online). Convergence was assessed by inspection of the likelihoods and by comparison of likelihoods and tree topologies between independent runs. Additionally, we reconstructed ML trees with 1,000 bootstrap replicates using RAxML (Stamatakis 2006) and added bootstrap support values to the trees shown in figures 2 and 3.

Supplementary figures S2 and S5, Supplementary Material online, include all 65 and 137 sequences, respectively. To reduce the risk of long branch attraction artifacts, figures 2 and 3 of the main text exclude sequences with significantly accelerated evolution ($P < 0.05$) according to Tajima's relative rate test (Tajima 1993), using a bacterial sequence (*Pedosphaera parvula*) as the outgroup and contrasting substitutions with *A. thaliana* GOX1 (supplementary table S1, Supplementary Material online). In total, six sequences were excluded from figure 2, and 16 sequences were excluded from figure 3 based on this criterion (supplementary table S1, Supplementary Material online). In particular, all fungal sequences are fast evolving and are thus included only in supplementary figures S2 and S5, Supplementary Material online. Mammalian lHAOX sequences, as well as *A. thaliana* lHAOX2, also failed the relative rate test; their position according to the full analysis (supplementary figure S5, Supplementary Material online) is indicated in figure 3.

Prediction of Substrate Specificities

Homology models of all *A. thaliana* (L)-2-HAOX proteins were generated by comparative modeling using MODELLER software version 9v9 (Sali et al. 1995) with default parameters. Based on sequence similarity and structural considerations, the crystal structure of spinach GOX (pdb-code 1GOX) was used as a template. Constructed models were ranked on the basis of an internal scoring function, and those with the least internal scores were superimposed in PyMOL (DeLano 2002) on the structure of human GOX1 complexed with glyoxylate (pdb-code 2RDU) to identify residues potentially involved in substrate binding.

We used the active site classification server (Röttig et al. 2010) to predict the most probable substrates of the proteins in our data set. We trained the SVM at the webserver using the 12 sequences employed for building the HMM profile for which substrate specificities had been determined

experimentally (fig. 3), plus two additional bacterial proteins from *Pseudomonas stutzeri* (Gao et al. 2012) and *E. coli* (Dong et al. 1993), which were tested experimentally but did not pass our stringent filters for the inclusion in the phylogenetic analyses.

We further provided the SVM with the 3D structure of human GOX1 (pdb-code 2RDU), as well as the 11 amino acid positions potentially involved in substrate binding that we identified above. The SVM algorithm then examined all amino acids at positions within a radius of 6 Å around each of these residues. The predictions using the Muscle alignment option of the active site classification server are shown in figure 3. Precision and recall are 0.80 and 1.00 for HAOX; 1.00 and 0.25 for LOX; and 0.75 and 1.00 for GOX. Details of the predictions are listed in supplementary figure S7 and table S2, Supplementary Material online.

Supplementary Material

Supplementary tables S1, S2, and figures S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (FOR 1186 to V.G.M.; EXC 1028 to V.G.M. and M.J.L.). The authors thank Bill Martin for helpful discussions.

References

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

Bauwe H, Hagemann M, Kern R, Timm S. 2012. Photorespiration has a dual origin and manifold links to central metabolism. *Curr Opin Plant Biol.* 15:269–275.

Beezley BB, Gruber PJ, Frederick SE. 1976. Cytochemical localization of glycolate dehydrogenase in mitochondria of *Chlamydomonas*. *Plant Physiol.* 58:315–319.

Belmouden A, Lederer F. 1996. The role of a beta barrel loop 4 extension in modulating the physical and functional properties of long-chain 2-hydroxy-acid oxidase isozymes. *Eur J Biochem.* 238:790–798.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 72:248–254.

Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, Tandeau de Marsac N, et al. 2013. Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol Evol.* 5:31–44.

Danpure CJ, Purdue PE. 1995. Primary hyperoxaluria. In: Scriver CR, Baudet AL, Sly WS, Valle D, editors. The metabolic and molecular bases of inherited disease. 7th ed. New York: McGraw-Hill. p. 2385–2424.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.

DeLano WL. 2002. *PyMOL* [cited 2014 Jan 27]. Available from: http://www.pymol.org.

Dong JM, Taylor JS, Latour DJ, Iuchi S, Lin EC. 1993. Three overlapping lct genes involved in L-lactate utilization by *Escherichia coli*. *J Bacteriol.* 175:6671–6678.

Duncan JD, Wallis JO, Azari MR. 1989. Purification and properties of *Aerococcus viridans* lactate oxidase. *Biochem Biophys Res Commun.* 164:919–926.

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 7: e1002195.

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M. 2008. The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proc Natl Acad Sci U S A.* 105:17199–17204.

Gao C, Jiang T, Dou P, Ma C, Li L, Kong J, Xu P. 2012. NAD-independent L-lactate dehydrogenase is required for L-lactate utilization in *Pseudomonas stutzeri* SDM. *PLoS One* 7:e36519.

Gibello A, Collins MD, Dominguez L, Fernandez-Garayzabal JF, Richardson PT. 1999. Cloning and analysis of the L-lactate utilization genes from *Streptococcus iniae*. *Appl Environ Microbiol.* 65: 4346–4350.

Giordano M, Beardall J, Raven JA. 2005. CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol.* 56:99–131.

Guiard B. 1985. Structure, expression and regulation of a nuclear gene encoding a mitochondrial protein: the yeast L(+)-lactate cytochrome c oxidoreductase (cytochrome b2). *EMBO J.* 4: 3265–3272.

Hackenberg C, Kern R, Huge J, Stal LJ, Tsuji Y, Kopka J, Shiraiwa Y, Bauwe H, Hagemann M. 2011. Cyanobacterial lactate oxidases serve as essential partners in N2 fixation and evolved into photorespiratory glycolate oxidases in plants. *Plant Cell* 23:2978–2990.

Huege J, Goetze J, Schwarz D, Bauwe H, Hagemann M, Kopka J. 2011. Modulation of the major paths of carbon in photorespiratory mutants of synechocystis. *PLoS One* 6:e16278.

Hutton D, Stumpf PK. 1971. Fat metabolism in higher plants. LXII. The pathway of ricinoleic acid catabolism in the germinating castor bean (*Ricinus communis* L.) and pea (*Pisum sativum* L.). *Arch Biochem Biophys.* 142:48–60.

Jones JM, Morrell JC, Gould SJ. 2000. Identification and characterization of HAOX1, HAOX2, and HAOX3, three human peroxisomal 2-hydroxy acid oxidases. *J Biol Chem.* 275:12590–12597.

Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* 294:2351–2353.

Kern R, Eisenhut M, Bauwe H, Weber AP, Hagemann M. 2013. Does the Cyanophora paradoxa genome revise our view on the evolution of photorespiratory enzymes? *Plant Biol (Stuttg).* 15:759–768.

Kohler SA, Menotti E, Kuhn LC. 1999. Molecular cloning of mouse glycolate oxidase. High evolutionary conservation and presence of an iron-responsive element-like sequence in the mRNA. *J Biol Chem.* 274:2401–2407.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol.* 25:1307–1320.

Macheroux P, Mulrooney SB, Williams CH Jr, Massey V. 1992. Direct expression of active spinach glycolate oxidase in *Escherichia coli*. *Biochim Biophys Acta.* 1132:11–16.

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165.

Maurino VG, Peterhansel C. 2010. Photorespiration: current status and approaches for metabolic engineering. *Curr Opin Plant Biol.* 13: 249–256.

Murray MS, Holmes RP, Lowther WT. 2008. Active site and loop 4 movements within human glycolate oxidase: implications for substrate specificity and drug design. *Biochemistry* 47:2439–2449.

Nakamura Y, Kanakagiri S, Van K, He W, Spalding MH. 2005. Disruption of the glycolate dehydrogenase gene in the high-CO2-requiring mutant HCR89 of *Chlamydomonas reinhardtii*. *Can J Bot.* 83: 820–833.

Nelson EB, Tolbert NE. 1970. Glycolate dehydrogenase in green algae. *Arch Biochem Biophys.* 141:102–110.

Ogren WL. 2003. Affixing the O to Rubisco: discovering the source of photorespiratory glycolate and its regulation. *Photosynth Res.* 76: 53–63.
Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K. 2009. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* 21:3718–3731.
Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40:D130–D135.
Reumann S, Ma C, Lemke S, Babujee L. 2004. AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. *Plant Physiol.* 136:2587–2608.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
Röttig M, Rausch C, Kohlbacher O. 2010. Combining structure and sequence information allows automated prediction of substrate specificities within enzyme families. *PLoS Comput Biol.* 6:e1000636.
Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins* 23: 318–326.
Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam A, Baker BJ, Banfield JF, Garavito RM, et al. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339:1207–1210.
Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, et al. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466:720–726.
Stabenau H. 1976. Microbodies from spirogyra: organelles of a filamentous alga similar to leaf peroxisomes. *Plant Physiol.* 58:693–695.
Stabenau H, Säftel W. 1982. A peroxisomal glycolate oxidase in the alga Mougeotia. *Planta* 154:165–167.
Stabenau H, Winkler U. 2005. Glycolate metabolism in green algae. *Physiol Plant.* 123:235–245.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
Stenberg K, Clausen T, Lindqvist Y, Macheroux P. 1995. Involvement of Tyr24 and Trp108 in substrate binding and substrate specificity of glycolate oxidase. *Eur J Biochem.* 228:408–416.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607.
Timme RE, Bachvaroff TR, Delwiche CF. 2012. Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* 7: e29696.
Umena Y, Yorita K, Matsuoka T, Kita A, Fukui K, Morimoto Y. 2006. The crystal structure of L-lactate oxidase from *Aerococcus viridans* at 2.1A resolution reveals the mechanism of strict substrate recognition. *Biochem Biophys Res Commun.* 350:249–256.
Verhoeven NM, Jakobs C, ten Brink HJ, Wanders RJ, Roe CR. 1998. Studies on the oxidation of phytanic acid and pristanic acid in human fibroblasts by acylcarnitine analysis. *J Inherit Metab Dis.* 21: 753–760.
Verhoeven NM, Schor DS, Roe CR, Wanders RJ, Jakobs C. 1997. Phytanic acid alpha-oxidation in peroxisomal disorders: studies in cultured human fibroblasts. *Biochim Biophys Acta.* 1361:281–286.
Vignaud C, Pietrancosta N, Williams EL, Rumsby G, Lederer F. 2007. Purification and characterization of recombinant human liver glycolate oxidase. *Arch Biochem Biophys.* 465:410–416.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 18:691–699.
Wilkinson M, McInerney JO, Hirt RP, Foster PG, Embley TM. 2007. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol Evol.* 22:114–115.
Williams E, Cregeen D, Rumsby G. 2000. Identification and expression of a cDNA for human glycolate oxidase. *Biochim Biophys Acta.* 1493: 246–248.
Zelitch I, Schultes NP, Peterson RB, Brown P, Brutnell TP. 2009. High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiol.* 149:195–204.
Zhang Z, Lu Y, Zhai L, Deng R, Jiang J, Li Y, He Z, Peng X. 2012. Glycolate oxidase isozymes are coordinately controlled by GLO1 and GLO4 in rice. *PLoS One* 7:e39658.