Cell. Mol. Life Sci. (2013) 70:1807–1830
DOI 10.1007/s00018-012-1229-5

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

Structure and evolution of vertebrate aldehyde oxidases: from gene duplication to gene suppression

Mami Kurosaki · Marco Bolis · Maddalena Fratelli · Maria Monica Barzago · Linda Pattini · Gemma Perretta · Mineko Terao · Enrico Garattini

Received: 4 August 2012 / Revised: 29 November 2012 / Accepted: 3 December 2012 / Published online: 21 December 2012

Abstract Aldehyde oxidases (AOXs) and xanthine dehydrogenases (XDHs) belong to the family of molybdo-flavoenzymes. Although AOXs are not identifiable in fungi, these enzymes are represented in certain protists and the majority of plants and vertebrates. The physiological functions and substrates of AOXs are unknown. Nevertheless, AOXs are major drug metabolizing enzymes, oxidizing a wide range of aromatic aldehydes and heterocyclic compounds of medical/toxicological importance. Using genome sequencing data, we predict the structures of *AOX* genes and pseudogenes, reconstructing their evolution. Fishes are the most primitive organisms with an *AOX* gene (*AOXα*), originating from the duplication of an ancestral *XDH*. Further evolution of fishes resulted in the duplication of *AOXα* into *AOXβ* and successive pseudogenization of *AOXα*. *AOXβ* is maintained in amphibians and it is the likely precursors of reptilian, avian, and mammalian *AOX1*. Amphibian *AOXγ* is a duplication of *AOXβ* and the likely ancestor of reptilian and avian *AOX2*, which, in turn, gave rise to mammalian *AOX3L1*. Subsequent gene duplications generated the two mammalian genes, *AOX3* and *AOX4*. The evolution of mammalian *AOX* genes is dominated by pseudogenization and deletion events. Our analysis is relevant from a structural point of view, as it provides information on the residues characterizing the three domains of each mammalian AOX isoenzyme. We cloned the cDNAs encoding the AOX proteins of guinea pig and cynomolgus monkeys, two unique species as to the evolution of this enzyme family. We identify chimeric RNAs from the human *AOX3* and *AOX3L1* pseudogenes with potential to encode a novel microRNA.

Keywords Aldehyde oxidase · Molybdo-flavoenzyme · Drug metabolism · Molybdenum cofactor

M. Kurosaki and M. Bolis contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00018-012-1229-5) contains supplementary material, which is available to authorized users.

M. Kurosaki · M. Bolis · M. Fratelli · M. M. Barzago · M. Terao · E. Garattini (✉)
Laboratory of Molecular Biology, Istituto di Ricerche Farmacologiche "Mario Negri", via La Masa 19, 20156 Milan, Italy
e-mail: enrico.garattini@marionegri.it

L. Pattini
Department of Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

G. Perretta
Istituto di Biologia Cellulare e Neurobiologia, Consiglio Nazionale delle Ricerche, via Anguillarese 301, 00123 Rome, Italy

Introduction

Aldehyde oxidases (AOX, EC 1.2.3.1) are proteins belonging to the family of molybdo-flavoenzymes along with the structurally related xanthine oxidoreductase/xanthine dehydrogenase (XDH). Molybdo-flavoenzymes require a molybdo-pterin cofactor (MoCo) and FAD for their catalytic activity [1–4]. Catalytically active AOXs and XDHs are homodimeric proteins whose subunits are characterized by the same tripartite structure consisting of a 25-kDa N-terminal domain containing two 2Fe/2S centers, a 45-kDa FAD-binding central domain, and an 85-kDa C-terminal domain containing the MoCo-binding site and the substrate pocket. The primary structures of the AOX and XDH subunits share a high degree of amino acid identity. The recent crystallization of the first mammalian AOX demonstrated that the

secondary and tertiary structures of AOXs and XDHs are also very similar [5, 6]. Despite their structural similarity, AOXs and XDHs differ in their substrate specificity and function. Indeed, XDHs act rather specifically on hypoxanthine and xanthine, representing key enzymes in the catabolism of purines. In contrast, AOXs are characterized by broad substrate specificity, catalyzing the oxidation of various types of compounds [4, 7]. The physiological substrates and the homeostatic functions of AOXs are still unknown and the object of active investigation.

While most organisms contain a single XDH, different isoenzymatic forms of AOX have been described in plants and animals [1–3, 8–11]. In plants, four AOXs play a role in the metabolism of abscisic acid, a regulator of growth in stressful conditions [8, 12, 13]. Animal AOXs metabolize a wide range of xenobiotics [7, 14], acting on a broad range of substrates that are not limited to organic molecules containing an aldehyde functionality [1, 4, 15]. In mammals, the number of AOX proteins varies from four in mice and rats (AOX1, AOX3, AOX4 and AOX3L1) to one in humans (AOX1) [1]. The four rodent isozymes are the product of an equivalent number of distinct genes, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1*, clustering on a short region of chromosome 1c1, while the human AOX1 locus maps to chromosome 2q. Unlike XDH whose expression is rather ubiquitous [16], rodent AOXs are expressed in a tissue-specific fashion, suggesting different physiological functions [17–20]. AOX1 and AOX3 have overlapping tissue specificity, with liver representing the richest source for the two enzymes [18, 19]. Liver AOXs are considered to represent an important xenobiotic metabolizing system in animals and humans [21]. AOX4 expression is limited to intra-orbital Harderian glands [22], sebaceous glands, epidermis, mouth epithelium, and esophagus [19]. The *Aox4* knock-out mouse [22] is viable and characterized by alterations in the structure of the epidermis. This animal shows deficient biotransformation of retinaldehyde into retinoic acid, the active metabolite of vitamin A, in the Harderian glands, and the skin. AOX3L1 expression is restricted to the nasal mucosa and Bowman's glands, the major exocrine structures of the nasal cavity [18]. Although, the *Aox3l1* knock-out mouse (MT and EG, unpublished observations) is viable and does not present with gross anatomical abnormalities, we hypothesize that AOX3L1 has a role in the perception of certain odorants and pheromones, as these compounds are often aromatic aldehydes. The human genome is characterized by the presence of *AOX1*, and the two pseudogenes, *AOX3* and *AOX3L1* (currently annotated as a single pseudogene, i.e., *AOX2P*). *AOX1* is expressed predominantly in the liver, although detectable amounts of the encoded enzyme are present in other tissues including lung and kidney.

The evolution of AOXs has been the object of only two studies [1, 23]. Two distinct AOX evolutionary branches have been hypothesized on the basis of the few plant and animal species for which molybdo-flavoenzyme sequences were available [24]. The first one resulted in the development of insect and plant AOXs via a series of gene duplications from a common *XDH* ancestral precursor. The second branch, resulting from a distinct ancestral *XDH* duplication, led to the appearance of vertebrate *AOX*s [1].

In this report, we take advantage of the large number of vertebrate species for which genome sequencing data are available, to predict the structures of numerous *AOX* genes and relative protein products. The evolutionary history of vertebrate *AOX* genes is reconstructed. Our results also provide initial information on the structural details characterizing each mammalian AOX isoenzyme. We also clone and sequence the AOX cDNAs expressed in guinea pig and cynomolgus monkeys, two species of major interest for the evolution of this family of enzymes. Finally, we demonstrate the existence of chimeric RNAs transcribed from the human *AOX3* and *AOX3L1* pseudogenes, which are predicted to be the precursor of a novel microRNA.

Materials and methods

Animals

Mice (*C57BL/6J*) and Dunkin Hartley guinea pigs (stock, HsdDhl:DH) were purchased from Harlan Laboratories, Inc. (Indianapolis, IN, USA). Animals were maintained on standard chow and colonies expanded in standard pathogen-free animal house facilities. All the procedures involving animals were performed according to the relevant national and international legislations and authorized by the internal Ethical Committee of the Animal Care Unit. The *Macaca fascicularis* (cynomolgus monkey) subjects were colony-born individuals from a closed colony established in 1981 and were maintained at the Consiglio Nazionale delle Ricerche, Primate Center in Rome, Italy.

Analysis of the genomes and reconstruction of the *AOX* gene structures

Local *tBlastn* searches were performed to identify exons coding for homologues of the protein products of murine *Aox1*, *Aox3*, *Aox4*, and *Aox3l1* genes in 65 different genomes. Sequences from GenBank and Ensembl genomes were used. Exons were assembled (Online Resource 1 and Online Resource 2) and translated into the corresponding protein products (Online Resource 3). When necessary, we repeated *tBlastn* searches using sequences obtained from genomes phylogenetically less distant and validated with sequences present in the EST databases. When we did not identify the complete set of exons due to sequencing

gaps in the genome, we added stretches of N residues of the expected length. For the generation of the phylogenetic tree, we discarded predictions with N residues accounting for more than 20 % of the total ORF (open reading frame) length. The 71 complete sequences predicted from vertebrate genomes were submitted to GenBank as TPA inferential entries (Online resource 4, Suppl. Table 1) and were given the accession No. BK008648–BK008718. We sorted active *AOX* genes from pseudogenes according to at least one of the following criteria. *AOX* were classified as pseudogenes when we observed: (a) the presence of more than one STOP codon located in distinct exons; (b) the absence of one or more identifiable exons resulting in disruption of the expected downstream reading frame; (c) lack of a recognizable intron/exon junction maintaining the reading frame.

Phylogenetic analysis of the AOX proteins

To determine the evolution of the *AOX* genes, we decided to perform a phylogenetic analysis on the corresponding protein products, as this provides a more accurate and reliable method when a comparison of distant species is necessary. Only proteins whose sequences accounted for at least 80 % of the predicted length were included, as some of the assembled genomes contained gaps which did not allow complete reconstruction of the protein sequences. Phylogenetic analysis pipelines available at the Phylogeny.fr server (http://www.phylogeny.fr/) were used. Multiple sequence alignments were prepared using Muscle (http://www.ebi.ac.uk/Tools/msa/muscle/) and visualized with Ugene (http://ugene.unipro.ru/). Default parameters were used for Muscle analysis (http://www.drive5.com/muscle). These included the following numerical parameters: anchor spacing = 32; diaglength = 24; diagmargin = 5; hydro = 5; maxiters = 16. For the phylogenetic tree containing all the available AOX and XDH protein structures, we performed an alignment of 119 taxa and 232 sequences. Protein phylogeny in vertebrates was estimated from an alignment of 65 taxa and 148 sequences through the maximum likelihood approach implemented in PhyML [25]. A WAG (Whelan and Goldman) model of evolution with four-substitution rate categories was selected; the gamma shape parameter for among-site variation and the proportion of invariable sites were estimated by PhyML. One hundred bootstrap replicates were performed using the parameters estimated from the original data. A comprehensive phylogenetic tree spanning from bacteria to vertebrates was generated through the distance based method BioNJ (http://www.atgc-montpellier.fr/bionj/) [26] using the JTT (Jones Taylor Thornton) substitution matrix. The faster algorithm allowed to compute the bootstrap support on 1,000 replicates. The vertebrate and the comprehensive phylogenetic trees were drawn using FigTree v1.0 (http://evolve.zoo.ox.ac.uk/software.html?id=figtree). The genomes of *Alligator mississippiensis* and *Crocodylus porosus* were downloaded from the crocodilian genome project website [27].

Secondary structure prediction

The secondary structures of the mammalian AOX1, AOX3, AOX4, and AOX3L1 25-kDa N-terminal domains were determined with the algorithm used by the CFSSP prediction server (http://www.biogem.org).

cDNA cloning, nucleotide sequencing, and determination of the intron/exon structure of the guinea pig and cynomolgus monkey AOXs

Total RNA was extracted from *Cavia porcellus* (guinea pig) liver, Harderian gland, and nasal mucosa as well as *Macaca fascicularis* (cynomolgus monkey) liver and nasal mucosa, using RNAspin mini RNA Isolation kit (GE Healthcare). The poly(A^+) fraction was selected according to standard procedures [18].

Cavia porcellus AOX1 cDNA was isolated as three overlapping fragments (corresponding to exons 1–14, 14–25, and 24–35 of the relative gene) by RT-PCR from liver RNA. *Cavia porcellus* AOX4 cDNA was isolated as four overlapping fragments (corresponding to exons 1–10, 9–24, 22–32, and 31–35) by RT-PCR from Harderian gland RNA. *Cavia porcellus* AOX3L1 cDNA was isolated as four overlapping fragments (corresponding to exons 1–14, 13–21, 17–28, and 27–35) by RT-PCR from nasal mucosa RNA. *Macaca fascicularis* AOX1 cDNA was isolated as two overlapping fragments (corresponding to exons 1–18 and 17–35 of the relative gene) by RT-PCR from liver RNA. *Macaca fascicularis* AOX3L1 cDNA was isolated as two overlapping fragments (corresponding to exons 5–21 and 21–34) by RT-PCR from nasal mucosa RNA.

The appropriate DNA fragments were subcloned into the *pCR2.1* plasmid using the TA cloning kit (Invitrogen) and sequenced in both directions by Primm (Milan, Italy). The sequencing technology is based on the use of the fluorescent dye terminator protocol and was performed on an ABI PRISM 3730 DNA Analyzer. Oligodeoxynucleotide primers were custom synthesized by Sigma. The nucleotide and protein sequences of the full-length *Cavia porcellus* and *Macaca fascicularis* AOX cDNAs were compared with the corresponding genomic sequences present in the ENSEMBL public databases. This allowed determination of the exon/intron structure of the corresponding genes.

The cDNA fragments corresponding to the 5′- and 3′-ends of *Cavia porcellus* (AOX1, AOX4 and AOX3L1) and *Macaca fascicularis* (AOX1 and AOX3L1) transcripts were determined by 5′-rapid amplification of cDNA ends (RACE) using the Marathon cDNA amplification kit (Clontech) and 5′-specific primers (5′SP1 and 5′SP2). 3′-RACE experiments were conducted as above with the specific primers

3′SP1 and 3′SP2. PCR products were subcloned in pCR2.1 and multiple clones were sequenced. The sequences were deposited in GenBank with the following accession numbers: *Cavia porcellus* AOX1 = JQ280309; *Cavia porcellus* AOX4 = JQ280310; *Cavia porcellus* AOX3L1 = JQ280311; *Macaca fascicularis* AOX1 = FJ751935; *Macaca fascicularis* AOX3L1 = FJ746636. The structures of all the oligonucleotides used for the cloning studies are listed in (Online Resource 4, Suppl. Table 2).

Tissue distribution of guinea pig and cynomolgus monkey AOX1, AOX3, AOX4, and AOX3L1

Tissue distribution of *Macaca fascicularis* AOX1/AOX3L1 as well as guinea pig and mouse AOX1/AOX4/AOX3L1 were determined by RT/PCR using the oligonucleotide amplimers reported in Online Resource 4, Suppl. Table 2. The PCR-amplified DNA fragments were analyzed using 1 % agarose gel electrophoresis.

Cloning and sequencing of the full-length cDNAs coding for the human AOX3/AOX3L1 chimeric transcripts in thymomas

Human thymoma specimens were kindly provided by Pia Bernasconi (Istituto Neurologico "Carlo Besta", Milan, Italy). Total RNA was prepared by miRNeasy Mini Kit (Qiagen) and used as the template for RT/PCR based cloning of the cDNAs. The primers used for the amplification of the AOX3/AOX3L1 chimeric cDNAs are detailed in Online Resource 4, Suppl. Table 2. The amplified DNA fragments were subcloned in *pCR2.1* as above and sequenced. The sequences were deposited with the following accession numbers: JX431285 and JX431286.

Prediction of the microRNA structure encoded by the human AOX3/AOX3L1 chimeric transcripts

The potential to encode a microRNA precursor was assessed through CID-miRNA (http://mirna.jnu.ac.in/cidmirna/index.html) [28], an algorithm, available as a web application, trained on human miRNAs secondary structures. The predicted hairpin structure was confirmed by the RNAfold software of the Vienna RNA package (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Results and discussion

Evidence of two ancestral AOX lineages from genomic sequences

The focus of our study is vertebrate *AOX* genes and corresponding protein products, whose structures were reconstructed from the available genomic sequencing data in a total of 65 species from fishes to humans using ENSEMBL, NCBI, and specific genome databases (Online Resources 1–3). Using the *tBlastn* algorithm, we searched for exonic sequences coding for polypeptides with similarity to the mouse AOX1, AOX3, AOX4, and AOX3L1 proteins or an evolutionary closer AOX, when necessary. Reconstruction of the *AOX* genes was performed, taking into account conservation of the coding exons in terms of sequence, length, and phase of exon/intron junctions [1, 3]. A major problem was sorting of genes coding for an active AOX protein (active genes) from pseudogenes, which may be transcribed, but do not encode a complete protein [1]. To this purpose, we applied a number of conservative criteria (see "Materials and methods"). Our data indicate that the vast majority of the *AOX* genes predicted to be active consists of 35 exons. Exceptions are represented by the mammalian *AOX3L1* genes most of which contain an additional non-coding 36th exon. *AOX3L1* genes are also characterized by a peculiarity in the exon 2/3 junction, which invariably presents with an anomalous GG or GC splicing donor site rather than the canonical GT dinucleotide [18]. Except for exon 25, which is split by an extra intron in *XDH* [16], conservation of the exon/intron junctions between *AOXs* and *XDHs* is also entirely confirmed [1].

Prediction of the active AOX and XDH gene structures in multiple species resulted in the determination of the corresponding protein products, which were used to reconstruct the evolutionary history of these enzymes. After integration of the vertebrate AOX protein sequences deduced in the present study with the AOX and XDH sequences already available for other eukaryotic organisms and a selection of bacteria (Online Resource 4, Suppl. Table 3), we generated an unrooted phylogenetic tree (Fig. 1). The tree confirms that XDH is represented throughout evolution from bacteria to man. It should be noticed that XDH is represented also in archeal bacteria, such as *Thermogladius cellulolyticus* or *Sulfolobus solfataricus*, although archeal enzymes were not included in our analysis. The prokaryotic and eukaryotic XDH clusters separate vertebrate AOX sequences from the protist, algae, plant (AOX1–AOX4), and insect (AOX1–AOX4) counterparts [1, 23]. The tree contains proteins identified in certain bacteria (bacterial AOXs in Fig. 1) that show remarkable amino acid identity with protist, algae, insect and plant AOXs. The structure of the tree is in line with the idea that independent gene duplication events from a common or distinct *XDH* ancestor produced two separate lineages of *AOX* genes [1, 23]. The first lineage gave rise to the extant complement of *AOXs* present in bacteria, protists, plants, algae, and insects. The second one is at the basis of the *AOX* evolution in vertebrates. An expansion of the vertebrate branch of the protein phylogenetic tree is shown in Fig. 2 and constitutes the basis of most of the following

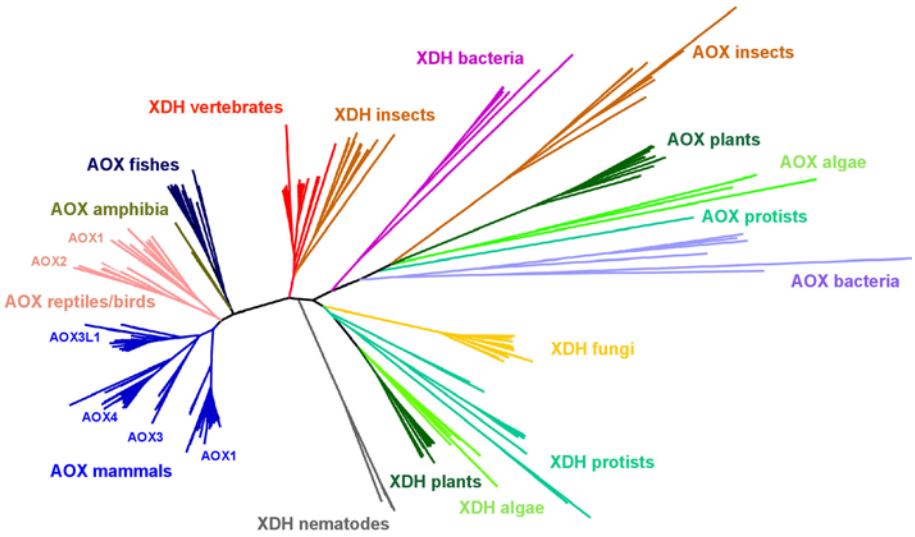


Fig. 1 Phylogenesis of AOX and XDH proteins in prokaryotes and eukaryotes. The unrooted phylogenetic tree was generated from the available prokaryotic as well as eukaryotic AOX and XDH protein sequences. The phylogenetic tree consists of all the AOX and XDH proteins whose structure could be predicted from the cloning of the corresponding cDNAs or deduced from genome sequencing data. Only proteins whose sequences account for at least 80 % of the predicted length are included, as some of the assembled genomes contain gaps, which do not allow complete reconstruction of the protein sequences. The macrogroups corresponding to reptile, bird, and mammal AOX1, AOX2, AOX3, AOX4, and AOX3L1 are indicated

evolutionary considerations. A summary of all the vertebrate species considered and their taxonomy is present in Online Resource 4, Suppl. Table 4.

Fish AOXα and AOXβ: relationship to AOX1

The most ancient vertebrate with a complete assembly of the genome is the jawless fish, *Petromyzon marinum* (lamprey). *P. marinum* belongs to Hyperoartia, the oldest class of fishes to appear 360 million years (Myr) ago [29]. These animals present with a typical *XDH* gene (Fig. 3), but they are devoid of sequences showing similarity with any *AOX*. A similar situation is observed in *Callorhinchus milii* (Australian ghost shark) and *Leucoraja erinacea* (little skate) that are members of the more recent class of Chondrichthyes.

Actinopterygii (ray-finned fishes) are the oldest vertebrates endowed with *AOX* genes. *Danio rerio* (zebrafish) and *Lepisosteus oculatus* (spotted gar) contain two genes, which we named *AOXα* and *AOXβ* (Fig. 3). The nomenclature based on Greek letters adopted for the *AOXs* in fishes, amphibians, and reptiles reflects the present uncertainty as to the evolutionary relationship between some of these genes and the mammalian counterparts, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1*. *AOXα* and *AOXβ* map to the same chromosome in close proximity to each other, consist of the canonical 35 coding exons and code for all the domains present in the catalytically active AOX enzymes. The two genes are transcribed on opposite DNA strands, showing a head-to-head configuration. The more recent *Actinopterygii*, *Takifugu rubripes* (fugu), *Tetraodon nigroviridis* (pufferfish), *Gasterosteus aculeatus* (stickleback), *Xiphophorus maculatus* (platy), *Oryzias latipes* (medaka), *Oreochromis niloticus* (Nile tilapia) (Fig. 3) and *Poecilia reticulata* (guppy; not shown) are all characterized by a single *AOXβ* gene predicted to be active. *Salmo salar* (Atlantic salmon), an intermediate evolutionary species between *Danio rerio* and all the other Actinopterygii considered, possesses the entire *AOXβ* gene and an incomplete set of exons with similarity to *AOXα*, which is consistent with a pseudogenization event. A similar situation occurs in the only representative species of Sarcopterygii (lobe-finned fishes), *Latimeria chalumnae* (coelacanth).

The protein phylogenetic sub-tree constructed for vertebrate AOXs (Fig. 2) indicates that AOXα is basal with respect to AOXβ. Thus, we propose that the *AOXα* gene is the primary duplication product of an ancestor fish *XDH* counterpart, whereas *AOXβ* originated from a further duplication of *AOXα*. Thus, fish *AOXα* must be the product of a duplication event distinct from the one at the origin of plant and insect *AOXs*. *AOXα* was lost during further evolution of Actinopterygii and Sarcopterygii via pseudogenization (*S. Salar* and *L. chalumnae*) followed by complete deletion of the locus. Fish AOXβ proteins are most closely related to amphibian and terrestrial vertebrate AOX1 than to AOX3, AOX4, and AOX3L1. This strongly suggests that *AOXβ* is the ancestor of *AOX1*.

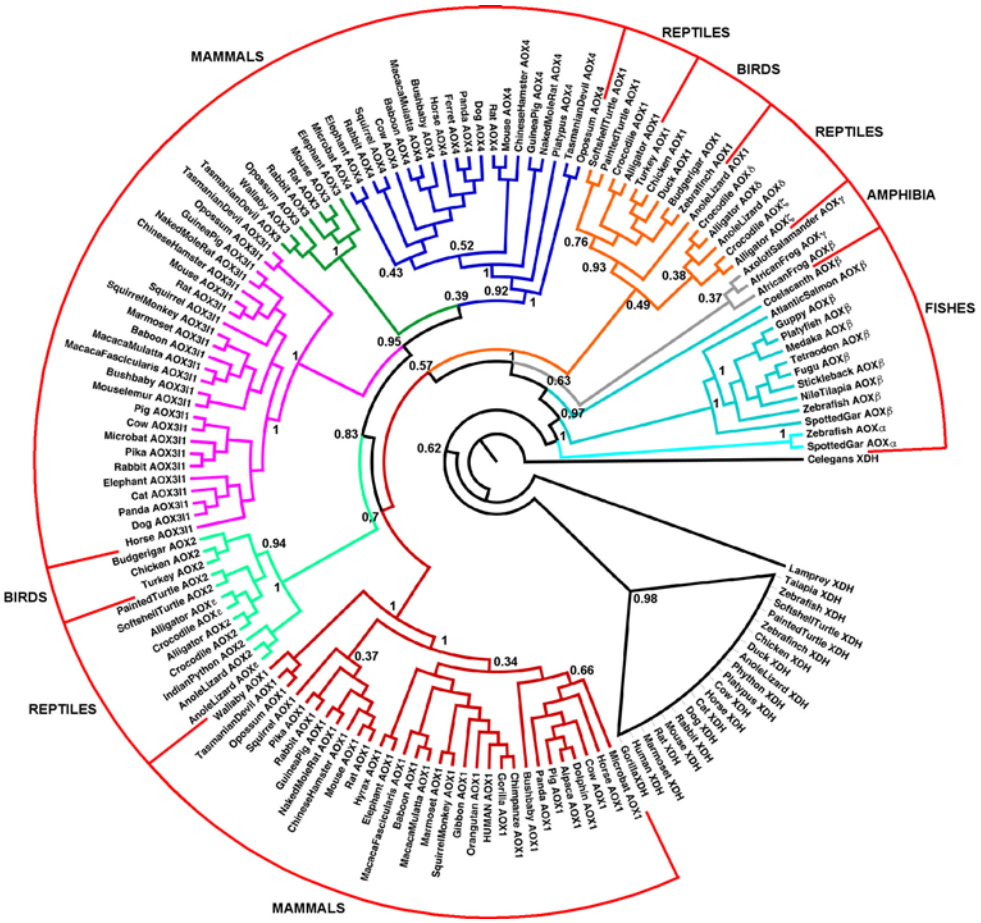


Fig. 2 Phylogenesis of AOX proteins in vertebrates. The rooted circular phylogenetic tree was generated from the available vertebrate AOX protein sequences. The phylogenetic tree consists of all the vertebrate AOX proteins whose structure could be predicted from the cloning of the corresponding cDNAs or deduced from genome sequencing data. The bootstrap values for the most primitive branches of the tree are shown. The position of the XDH sequences in the tree is shown as a reference without an illustration of the corresponding dendrogram. The tree consists of 148 sequences covering at least 80 % of the expected AOX polypeptide lengths to account for sequencing gaps in incomplete genomes

Maintenance of *AOXβ* and occurrence of a second gene duplication producing *AOXγ* in amphibians

Amphibians evolved from Sarcopterygii 315 Myr ago [30] and are divided in the orders of Anura, Caudata, and Gymnophiona. The genome of *Xenopus tropicalis* (African frog) contains the orthologue of fish *AOXβ*, the distinct *AOXγ* gene and an *AOXμ* pseudogene in between. It is unclear whether a similar situation applies also to *Ambystoma mexicanum* (mole salamander) (data not shown), in which only a single *AOX* gene (*AOXγ*) could be deduced from transcriptomic data (http://www.ambystoma.org). In *X. tropicalis*, *AOXβ*, *AOXμ*, and *AOXγ* are contained in the same scaffold (Fig. 3). *AOXβ* and *AOXγ* encode proteins of the expected size and complement of functional domains. On the basis of the sequence of the encoded protein, amphibian *AOXγ* is a gene distinct from both fish *AOXα* and *AOXβ*. The origin of *AOXμ* is currently unclear, while *AOXγ* may have originated from *AOXβ* as the result of a lineage-specific duplication event. Because of a low bootstrapping score (Fig. 2), it is currently difficult to establish the relationship between *AOXγ* and reptilian *AOX2* or mammalian *AOX3*, *AOX3l1*, and *AOX4*. However, it is likely that *AOXγ* is the direct precursor of *AOX2*.

Reptile and avian AOX2 the precursor of mammalian AOX3, AOX4, and AOX3L1

Sauropsida are today's reptiles and birds. Sauropsida became distinct from Synapsida, which gave rise to mammals, around 320 Myr ago. The oldest reptiles for which genomic sequences are available are *Chrysemys picta bellii* (western painted turtle) and *Pelodiscus sinensis* (Chinese softshell turtle). The genomes of *Anolis carolinensis* (anole lizard), *Python molurus* (Indian python), *Alligator mississippiensis* (American alligator) and *Crocodylus porosus* (saltwater crocodile) have also been sequenced.

C. picta bellii and *P. sinensis* are endowed with two active genes whose products show the highest similarity to avian/mammalian AOX1 and AOX2 (Fig. 2). Clustering with mammalian AOX3, AOX4, and AOX3L1 proteins indicates that AOX2 is more recent than AOX1. *A. carolinensis* is characterized by the same two genes. In this species, two further genes, *AOXδ* and *AOXε*, showing the canonical and conserved 35-exon structure, can be identified. The four loci map to chromosome 1 on the same strand and in the order, *AOXδ*, *AOX1*, *AOX2*, and *AOXε*. The AOXδ AOXε proteins cluster with reptilian AOX1s and AOX2s, respectively (Fig. 2). Interestingly, the *AOXδ*

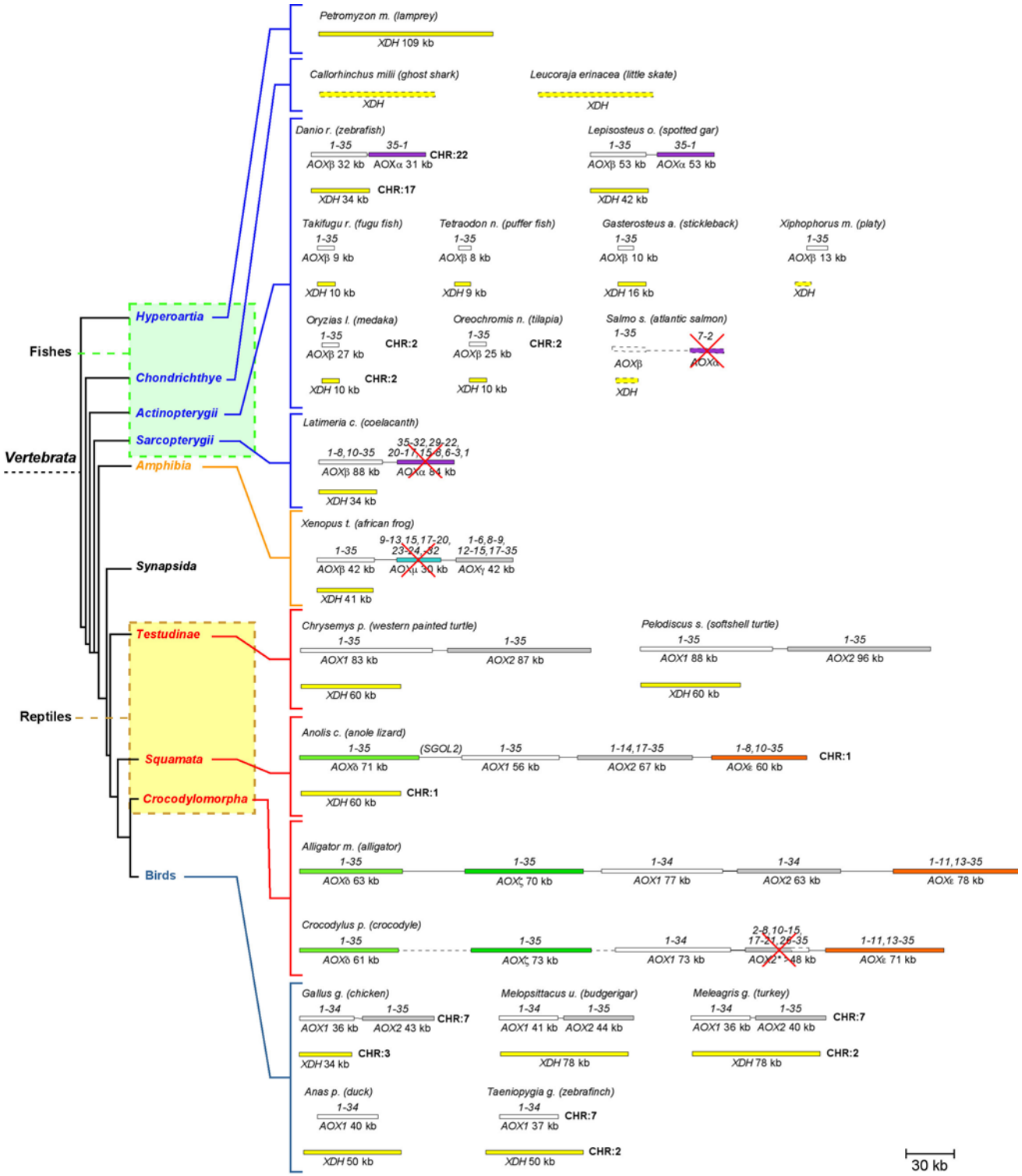


Fig. 3 *AOX* and *XDH* genes in fishes, amphibians, reptiles, and birds. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated vertebrates for which complete or almost complete genomic sequence data are available. Orthologous and predicted precursor genes are indicated *with the same color*. Predicted *AOX* pseudogenes are *crossed through* and *asterisked*. The number of exons in the active *AOX* genes are indicated along with the exons identified in the *AOX* pseudogenes. *Dashed boxes* are used when the length of the genes is unknown. *Dashed lines* are used when the intragenic distances are undetermined. When available, the chromosomal location is shown on the right. The simplified rooted phylogenetic diagram on the left illustrates the relative evolutionary distance between the lineages the animal species analyzed belong to. The Latin names of each species are followed by the corresponding common names in *parenthesis*

gene localizes upstream of *SGOL2* (*shugoshin-like 2*), which generally marks the upstream limit of all vertebrate *AOX* clusters. In contrast, *AOXε* lays downstream of *AOX2*. Chromosomal localization and protein similarity indicate that *AOXδ* and *AOXε* are duplication products of *AOX1* and *AOX2*, respectively. The genomes of *P. molurus*, *A. mississippiensis*, and *C. porosus* are still largely unassembled. In *P. molurus*, we could only reconstruct the exon structure of *AOX2* (data not shown). In addition to the four genes present in *A. carolinensis*, *A. mississippiensis* and *C. porosus* show a further *AOX* duplication, i.e., *AOXζ*, which is located in between *AOXδ* and *AOX1*. The 35-exon *AOXζ* gene is predicted to code for a polypeptide which clusters with AOXδ and all AOX1 proteins (Fig. 2). Noticeably, *C. porosus AOX2* is a pseudogene.

As indicated by the results obtained in *Gallus gallus* (chicken), *Melospittacus undulatus* (budgerigar), and *Meleagris gallopavo* (turkey), the *AOX1* and *AOX2* genes were conserved during the evolution from reptiles to birds. In birds, the two genes lay in close proximity on the same chromosomes (Fig. 3). In chicken, enzymatically active products of *AOX1* and *AOX2* (originally denominated *AOH*) orthologues were demonstrated, although this species is devoid of AOX activity in liver [24]. In *Taeniopigia guttata* (zebrafinch) and *Anas platyrhynchos* (mallard duck), only the *AOX1* gene is present, indicating deletion of *AOX2*. The bird AOX1 protein lays at the root of the mammalian AOX1 branch (Figs. 1, 2), while bird AOX2 lays in close proximity to the three branches corresponding to mammalian AOX3l1, AOX4, and AOX3.

AOX1, *AOX3*, *AOX4*, and *AOX3L1* genes in marsupials

Mammals are the only surviving Synapsida and are divided in Monotremata, Marsupialia, and Eutheria (placental mammals). Extant mammals are extremely diverse [31, 32], being characterized by remarkable divergence between lineages of similar age, from the more than 2,200 species of rodents to the single one of aardvarks [33, 34].

The two earliest splits in the evolution of mammals involved Monotremata approximately 166 Myr ago as well as Marsupialia and Eutheria approximately 148 Myr ago. The monotreme *Ornithorhyncus anatinus* (platypus) is the most primitive mammal for which a genome sequence is available. In *O. anatinus*, exon sequences with similarity to mouse AOX4 are recognized. The gene is likely to code for an active AOX, as the structures of only three exons (1, 29,

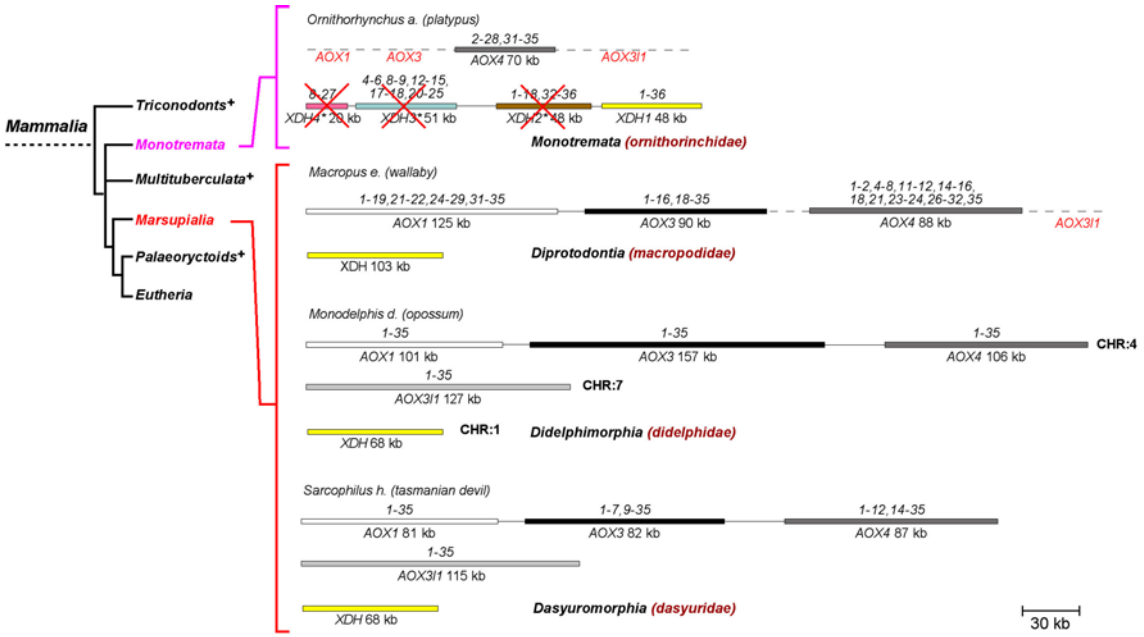


Fig. 4 *AOX* and *XDH* genes in monotremes and marsupials. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated monotremes and marsupials. Orthologous genes are indicated with the *same color*. Predicted pseudogenes are *crossed through* and *asterisked*. The number of exons in the active *AOX* as well as *Ornythorynchus a. XDH* genes are indicated along with the exons identified in the corresponding pseudogenes. *Dashed lines* are used when the intragenic distances are undetermined. Genes whose presence is only hypothesized are marked in *red*. When available, the chromosomal location is shown on the *right*. The simplified rooted phylogenetic diagram on the *left* illustrates the relative evolutionary distance between the lineages the animal species analyzed belong to. Order names and family names are indicated in *black*. The family names are shown in *red* and in *parenthesis*. The Latin names of each species are followed by the corresponding common names in *parenthesis*

and 30) could not be determined (Fig. 4). Sequences with similarity to mouse AOX1, AOX3 and AOX3L1 could not be identified, although this is probably due to the fact that the assembly of the genome is still preliminary. A unique feature of the animal is the predicted presence of an active *XDH* gene (*XDH1*) and three derived pseudogenes (*XDH2*, *XDH3*, and *XDH4*).

Marsupials are classified in seven orders, Didelphimorphia, Paucitubercolata, Peramelemorphia, Notoryctemorphia, Dasyuromorphia, Microbiotheria, and Diprotodontia. *Monodelphis domestica* (opossum) and *Sarcophilus harrisii* (Tasmanian devil) are endowed with the four active genes, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1* (Fig. 4). In *Macropus eugenii* (wallaby), the structure of *AOX1*, *AOX3*, and *AOX4* was reconstructed almost completely. Despite the absence of sequences with similarity to *AOX3L1*, we cannot exclude the existence of this gene either, given the preliminary nature of the genome assembly.

As Didelphimorphia are the oldest marsupial order, we conclude that the process of gene duplication resulting in the full complement of four active *AOX* genes present in rodents, among other Eutheria, must have been completed early during the evolution of Marsupialia. It is even possible that this occurred before the split of Monotremata from Marsupialia. To prove this last point, however, it will be necessary to establish the number of *AOX* genes in platypus or another Monotremata, such as *Tachyglossus aculeatus* (echidna), whose genome sequencing is in progress (http://genome.wustl.edu).

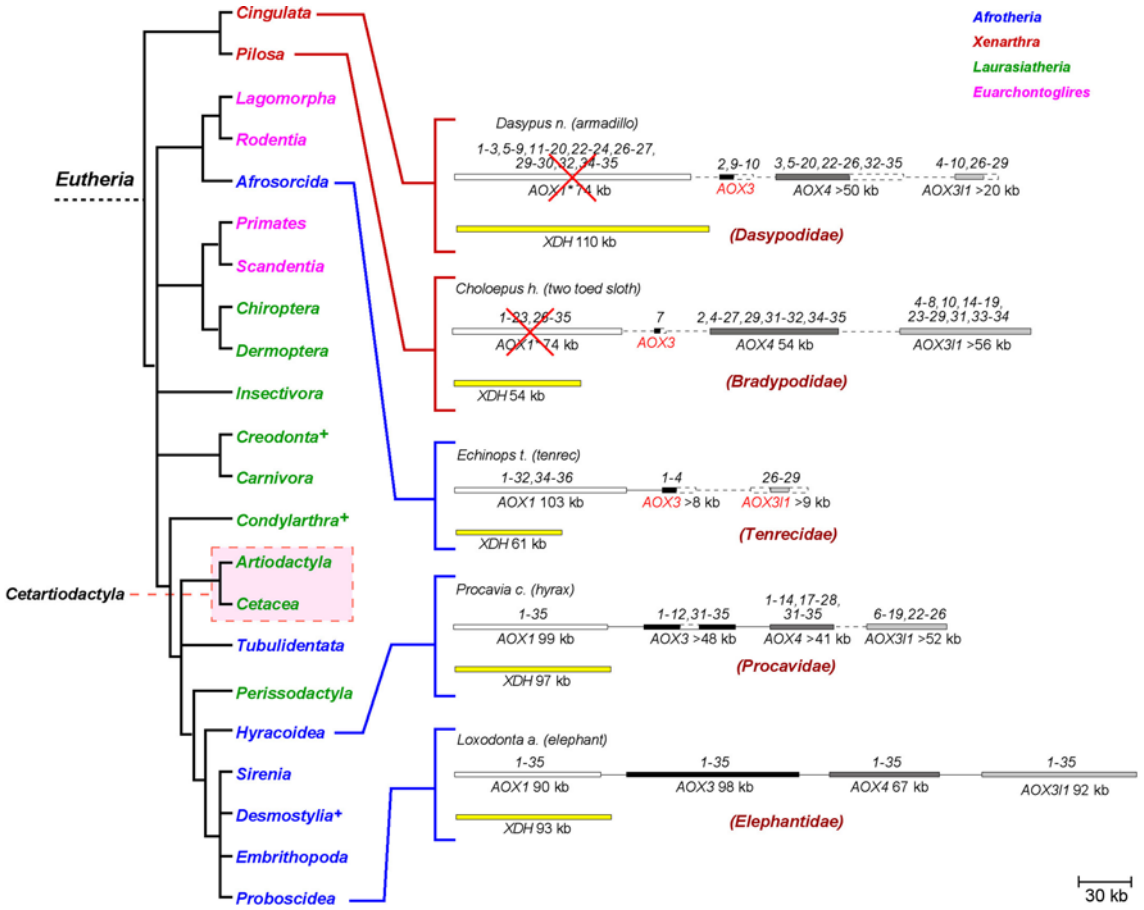


Fig. 5 *AOX* and *XDH* genes in Afrotheria and Xenarthra. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated mammalian species belonging to the Afrotheria and Xenarthra lineages for which complete or almost complete genomic sequences are available. Orthologous genes are indicated with the *same color*. Predicted pseudogenes are *crossed through* and *asterisked*. The number of exons in the *AOX* active genes are indicated along with the exons identified in the AOX pseudogenes. *Dashed boxes* are used when the length of the genes is unknown. *Dashed lines* are used when the intragenic distances are undetermined. Genes whose presence is only hypothesized are marked in *red*. When available, the chromosomal location is shown on the *right*. The simplified rooted phylogenetic diagram on the *left* illustrates the relative evolutionary distance between the orders the animal species analyzed belong to. The family names are shown in *red* and in *parenthesis*. The Latin names of each species are followed by the corresponding common names in *parenthesis*

AOX1 pseudogenization in Xenarthra

Four major lineages are represented in Eutheria, i.e., Afrotheria, Xenarthra, Laurasiatheria, and Euarchontoglires [31]. Afrotheria is the most ancient lineage (time of origin, 101 Myr ago) [31] and it groups the orders of Afrosorcida, Tubulidentata, Proboscidea, Hyracoidea, and Sirenia. The number of *AOXs* present in *Echinops telfairi* (lesser hedgehog tenrec, time of origin, 91 Myr ago) is incompletely defined (Fig. 5). *AOX1* is correctly predicted and few exons corresponding to *AOX3* have been identified in the same incomplete contiguous sequence. This species is also endowed with an *AOX3L1* gene, although the number of exons identified (exons 26–29) is very limited. No prediction can be made for *AOX4*, as the gene is likely to fall within an unsequenced stretch of DNA. In *Loxodonta africana* (African elephant, time of origin, 78 Myr ago), four active *AOX* genes, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1* are predicted. Finally, *Procavia capiensis* (rock hyrax, time of origin, 76 Myr ago) [35] is likely to contain four active *AOX* genes, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1*.

Xenarthra groups the orders of Cingulata and Pilosa and diverged from Afrotheria 101 Myr ago. Preliminary genome sequencing data are available for a member of Cingulata, *Dasypus novemcinctus* (armadillo), and a member of Pilosa, *Choloepus hoffmanni* (two-toed sloth) (Fig. 5). In *D. novemcinctus* and *C. hoffmanni*, we observed exon sequences with similarity to *AOX1*, *AOX3*, *AOX4*, and *AOX3L1*. However, in both species, the only prediction that can be made with confidence is pseudogenization of *AOX1*.

On the basis of our data, it can be concluded that AOX1 pseudogenization must have occurred following the split of Xenarthra from Afrotheria.

Pseudogenization of AOX1, AOX3, and AOX4 in Laurasiatheria

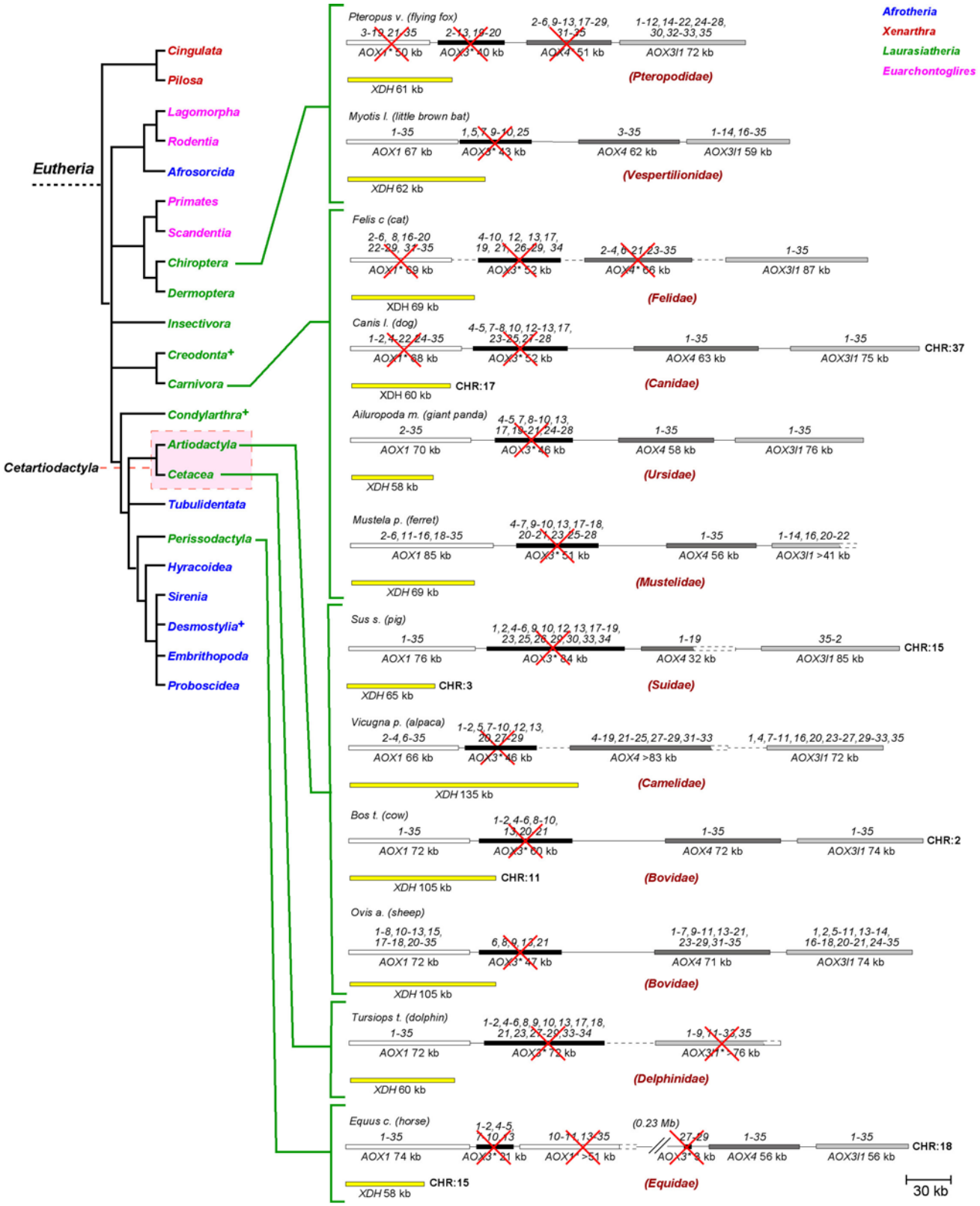
Laurasiatheria groups the orders of Chiroptera, Perissodactyla, Cetartiodactyla, and Carnivora [31]. The calculated time of origin for Chiroptera, the only order of flying Eutheria, is approximately 89 Myr ago [31]. The order is divided into Megachiroptera and Microchiroptera for which the genomes of two representative species, *Pteropus vampyrus* (flying fox) and *Myotis lucifugus* (little brown bat), are available. Megachiroptera originated earlier than Microchiroptera [36] and are estimated to be endowed with one active gene (*AOX3L1*) and three pseudogenes (*AOX1*, *AOX3*, and *AOX4*) (Fig. 6). An ORF of the expected size is predicted from *P. vampyrus AOX3L1* gene only if a likely sequencing error or polymorphism (an extra T, TTTT instead of TTT) on exon 26 is taken into account. Microbats possess a single pseudogene (*AOX3*) and three active genes (*AOX1*, *AOX4*, and *AOX3L1*).

Fig. 6 *AOX* and *XDH* genes in Laurasiatheria. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated mammalian species belonging to the Laurasiatheria lineage for which complete or almost complete genomic sequence data are available. Orthologous genes are indicated with the *same color*. Predicted pseudogenes are *crossed through* and *asterisked*. The number of exons in the active *AOX* genes are indicated along with the exons identified in the corresponding pseudogenes. *Dashed boxes* are used when the length of the genes is unknown. *Dashed lines* are used when the intragenic distances are undetermined. When available, the chromosomal location is shown on the *right*. The simplified rooted phylogenetic diagram on the *left* illustrates the relative evolutionary distance between the orders the animal species analyzed belong to. The family names are shown in *red* and in *parenthesis*. The Latin names of each species are followed by the corresponding common names in *parenthesis*

The time of origin for Perissodactyla and Cetartiodactyla is the same, i.e., approximately 87 Myr ago [31]. Cetartiodactyla groups Artiodactyla with Cetacea according to a recent classification scheme [37]. The complement of *AOX* loci in Perissodactyla (*Equus caballus*, horse) and Artiodactyla (*Sus scrofa*, pig; *Vicugna pacos*, alpaca; *Bos taurus*, cow and *Ovis aries*, sheep) is the same and consists of the *AOX1*, *AOX4*, and *AOX3L1* genes as well as the pseudogene, *AOX3* (Fig. 6). Within these species, the only gene for which limited data are available is *Sus s. AOX4*, because of a large sequencing gap, which prevents determination of putative exons 20–35. Cetacea is the order grouping all the marine eutherian species. *Tursiops truncatus* (dolphin) is predicted to contain an active *AOX1* gene, although assembly gaps prevent identification of exon 34, as well as the pseudogenes, *AOX3* and *AOX3L1*. The current genome assembly of this animal does not contain sequences with similarity to *AOX4*, consistent with a possible deletion event.

Carnivora are divided into many evolutionary lineages including the suborders Caniformia and Feliformia [38]. The base of the carnivore radiation occurred approximately 67 Myr ago [38–40]. Caniformia have traditionally comprised four families (Canidae, Mustelidae, Procyonidae, and Ursidae), while Feliformia consists of Felidae, Herpestidae, Hyaenidae and Viverridae. *Canis lupus* (dog) is characterized by the *AOX4* and *AOX3L1* genes, as well as the *AOX1* and *AOX3* pseudogenes (Fig. 6). *Felix catus*, (cat) contains an active *AOX3L1* gene and three pseudogenes, *AOX1*, *AOX3*, and *AOX4*. A different pattern is observed in *Ailuropoda melanoleuca* (giant panda) and *Mustela putorius furo* (ferret) for which three active *AOX* genes and the same inactive pseudogene, *AOX3*, are predicted.

All this supports the idea that *AOX3* underwent pseudogenization at an early stage during the evolution of Laurasiatheria. Conversely, pseudogenization of *AOX4* in *F. catus*



and *P. vampyrus* are likely to represent distinct events. We propose that *AOX1* underwent three independent pseudogenizations during the evolution of Pteropodidae, Felidae, and Canidae. In fact, branching of Felidae from Canidae predates branching of Canidae from Ursidae [40]. Yet, an active *AOX1* orthologue is present in the Ursidae, giant panda.

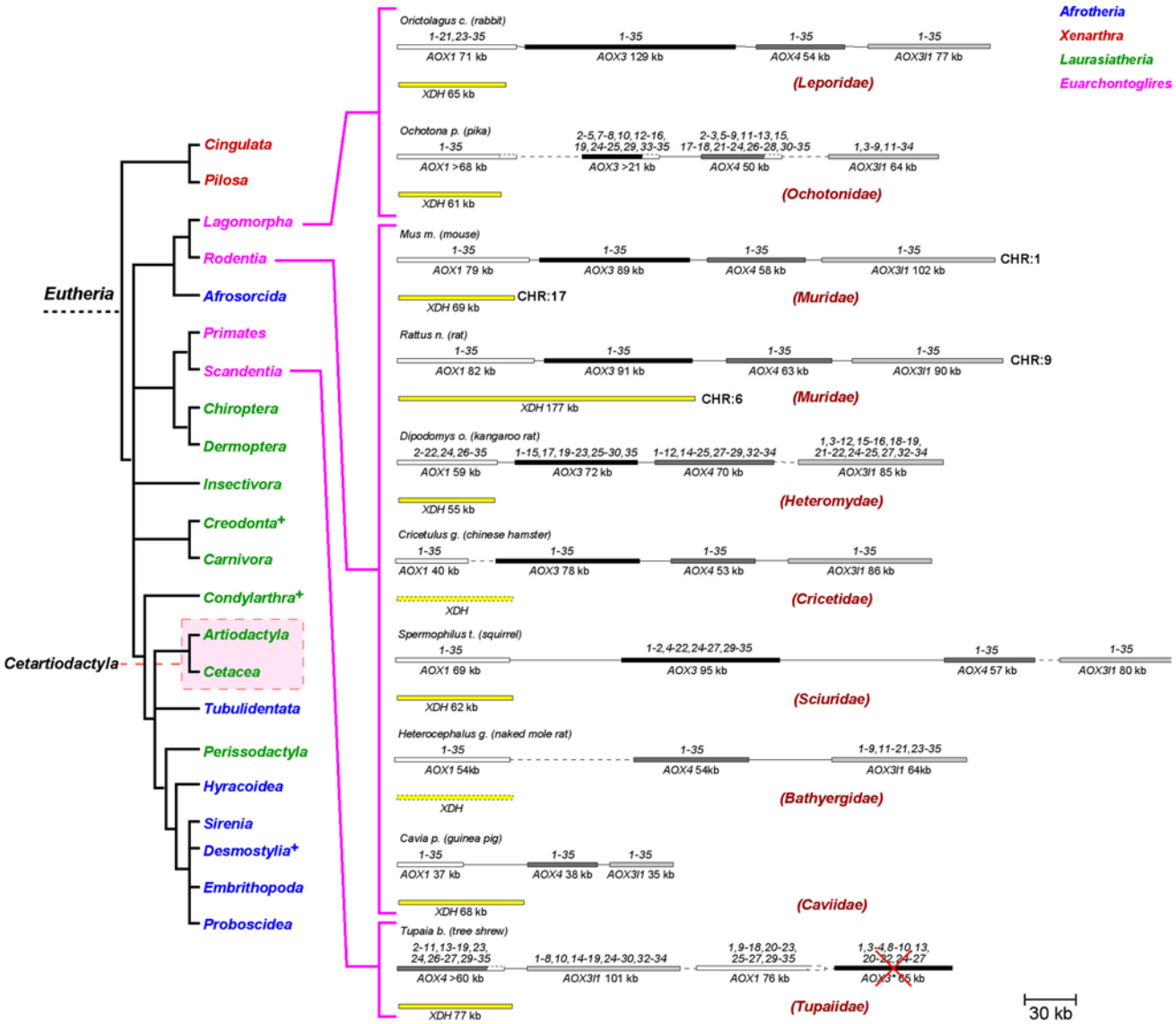


Fig. 7 *AOX* and *XDH* genes in Euarchontoglires. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated mammalian species belonging to the Euarchontoglires lineage for which complete or almost complete genomic sequencing data are available. Orthologous genes are indicated with the *same color*. Predicted pseudogenes are *crossed through* and *asterisked*. The number of exons in the active *AOX* genes are indicated along with the exons identified in the corresponding pseudogenes. *Dashed boxes* are used when the length of the genes is unknown. *Dashed lines* are used when the intragenic distances are undetermined. When available, the chromosomal location is shown on the *right*. The simplified rooted phylogenetic diagram on the *left* illustrates the relative evolutionary distance between the orders the animal species analyzed belong to. The family names are shown in *red* and in *parenthesis*. The Latin names of each species are followed by the corresponding common names in *parenthesis*

The deletion of AOX3 in Euarchontoglires

The diversified lineage of Euarchontoglires gave rise to the orders of Scandentia, Dermoptera, Lagomorpha, Rodentia and Primates, which will be discussed separately. Scandentia are the most ancient eutherian order (time of origin 94 Myr ago) [31]. The Scandentia, *Tupaia belangeri* (northern tree shrew), is predicted to contain three active genes, *AOX1*, *AOX4*, and *AOX3L1* as well as the *AOX3* pseudogene (Fig. 7). The relative distance and location of the four loci on the same scaffold is uncertain, although the current assembly indicates an unusual arrangement, with *AOX1* and *AOX3* mapping downstream of *AOX4* and *AOX3L1* on the same DNA strand. Pseudogenization of *AOX3* in Scandentia is not the consequence of the same event observed in Laurasiatheria, as the gene is active in members of the more recent suborders of Lagomorpha and Rodentia (see below). Similarly, the pseudogenizations of *AOX1* in Felidae and Canidae are unrelated.

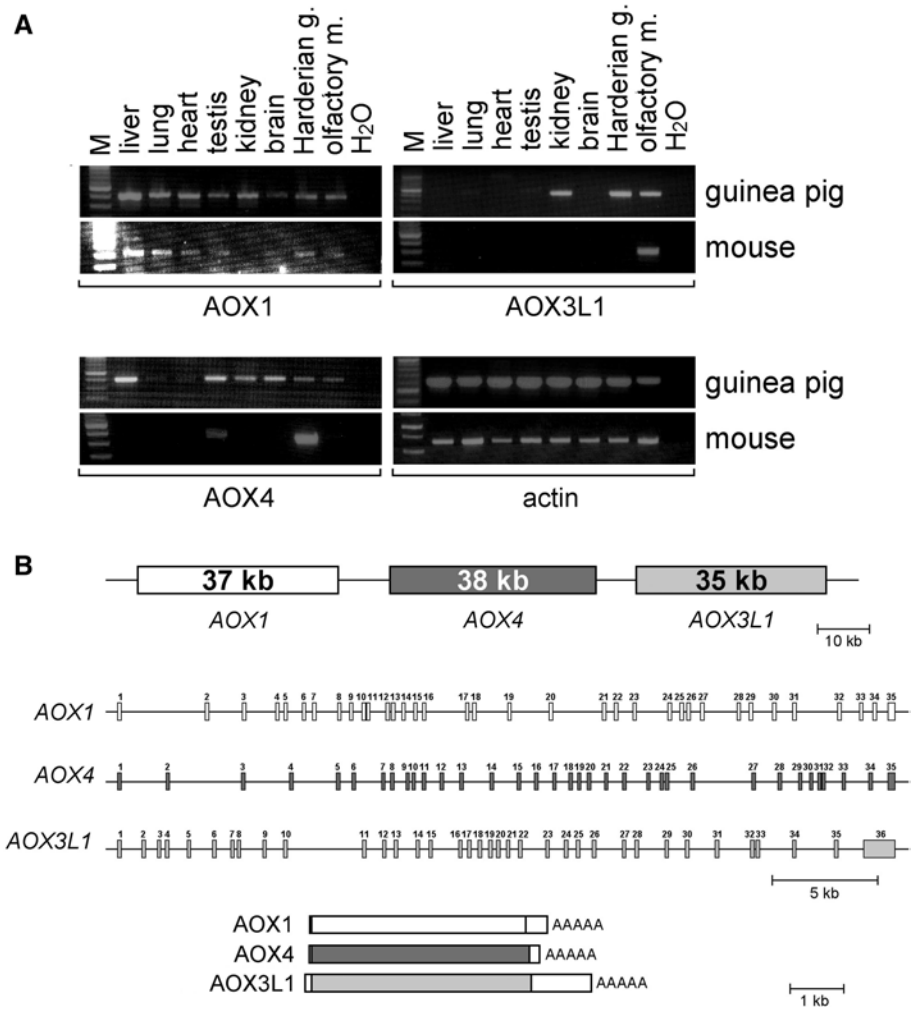


Fig. 8 Tissue-specific expression and exon–intron structures of the guinea pig *AOX1*, *AOX4*, and *AOX3L1* transcripts and genes. **a** A comparison of the tissue distribution of the guinea pig and mouse AOX1, AOX4 and AOX3L1 mRNAs is illustrated. Total RNA was extracted from the indicated tissue and amplified by RT/PCR using specific couples of amplimers for the various transcripts. Actin is used as an internal control of the experiment. Harderian g., Harderian gland; olfactory m., olfactory mucosa; M, molecular weight markers. **b** *Top* Schematic representation of the guinea pig cluster of *AOX* genes. *Middle* The exon–intron structures of the guinea pig *AOX1*, *AOX4*, and *AOX3L1* genes were deduced by comparison of the corresponding cDNA and genomic sequences. A schematic representation of the three genes is illustrated and exons are *numbered*. *Bottom* A schematic representation of the AOX1, AOX4, and AOX3L1 transcripts is illustrated. The *central box* corresponding to the coding region lays in between the boxes corresponding to the 5′ and 3′ ends of the transcript. *AAAAA* polyadenylated tail

Lagomorpha is the order closest to Rodentia and includes *Oryctolagus cuniculus* (rabbit, family Leporidae) and *Ochotona princeps* (pika, family Ochotonidae). In both species, *AOX1*, *AOX3*, *AOX4*, and *AOX3L1* are predicted to be active genes (Fig. 7). Rodentia is the order containing the largest number of species for which genome sequences are available and it consists of five suborders, Myomorpha, Sciuromorpha, Hystricomorpha, Anomaluromorpha, and Castorimorpha. The split of Myomorpha from Sciurimorpha predates branching of Hystricomorpha [31]. *Dipodomys ordii* (kangaroo rat), *Cricetulus griseus* (Chinese hamster), *Mus musculus* (mouse) and *Rattus norvegicus* (rat) belong to Myomorpha. As experimentally proven in *M. musculus* and *R. norvegicus* [18], *D. ordii* and *C. griseus* are also predicted to contain active *AOX1*, *AOX3*, *AOX4*, and *AOX3L1* genes. *Spermophilus tridecemlineatus* (squirrel), the only member of Sciuromorpha, shows the same complement of four *AOX* genes as Myomorpha. The genomes of the Hystricomorpha, *Cavia porcellus* (guinea pig) and *Heterocephalus glabrus* (naked mole rat) are exceptions, containing only *AOX1*, *AOX4*, and *AOX3L1* (Fig. 7). No sequences with similarity to *AOX3* are evident in either species, suggesting a unique gene deletion event.

Cloning and sequencing of guinea pig AOX cDNAs

The predicted absence of *AOX3* in naked mole rat and particularly in guinea pig has far-reaching implications. From an evolutionary point of view, it demonstrates that loss of *AOX3* is observed in rodents, which are generally characterized by four active *AOX* genes. Loss is not due to pseudogenization as observed in other eutherian species, but rather the consequence of a rare gene deletion event. From an applied perspective, *AOX3* deletion has important ramifications for the use of guinea pig as a human proxy in drug metabolism studies [7]. For these reasons and to validate our predictions on the *AOX* genes in guinea pig, we cloned the corresponding cDNAs, using total RNA extracted from the liver, Harderian gland and

nasal mucosa, the organs expected to contain the largest amounts of AOX1, AOX4, and AOX3L1, respectively. The full-length forms of the AOX1, AOX4, and AOX3L1 cDNAs were obtained after further experiments of 5′- and 3′-anchor PCR. Despite the use of degenerate amplimers designed against conserved sequences of mouse AOX3, we did not amplify an AOX3 cDNA from liver or any of the tissues considered.

The tissue distribution of the AOX1, AOX4, and AOX3L1 mRNAs in guinea pig and mouse was compared (Fig. 8a). Expression of AOX1 is relatively ubiquitous and similar in the two species, with liver containing the largest amounts of the transcript. In contrast, the expression profiles of guinea pig and mouse AOX4 are different. In guinea pig, significant amounts of AOX4 mRNA are present in many tissues, including Harderian glands, liver, testis, kidney, brain, and nasal mucosa. This is at variance with what is observed in the mouse, where expression of AOX4 is restricted to the Harderian gland, the epidermis, the mouth, and the esophagus [22]. Significantly, the relative levels of the AOX4 mRNA observed in guinea pig Harderian glands are much lower than those observed in the mouse counterpart. Guinea pig AOX3L1 mRNA is expressed in the kidney, the Harderian glands and the olfactory mucosa. This is also at variance with what is observed in the mouse where expression is restricted to the nasal mucosa [18]. Taken together, the results indicate that two distant rodent species are characterized by different profiles of AOX4 and AOX3L1 tissue-specific expression. Of particular interest is the presence of AOX1 and AOX4 mRNAs in guinea pig liver. Was this confirmed at the protein level, guinea pigs, like mice and rats, would represent inadequate proxies for the human situation in drug metabolism studies.

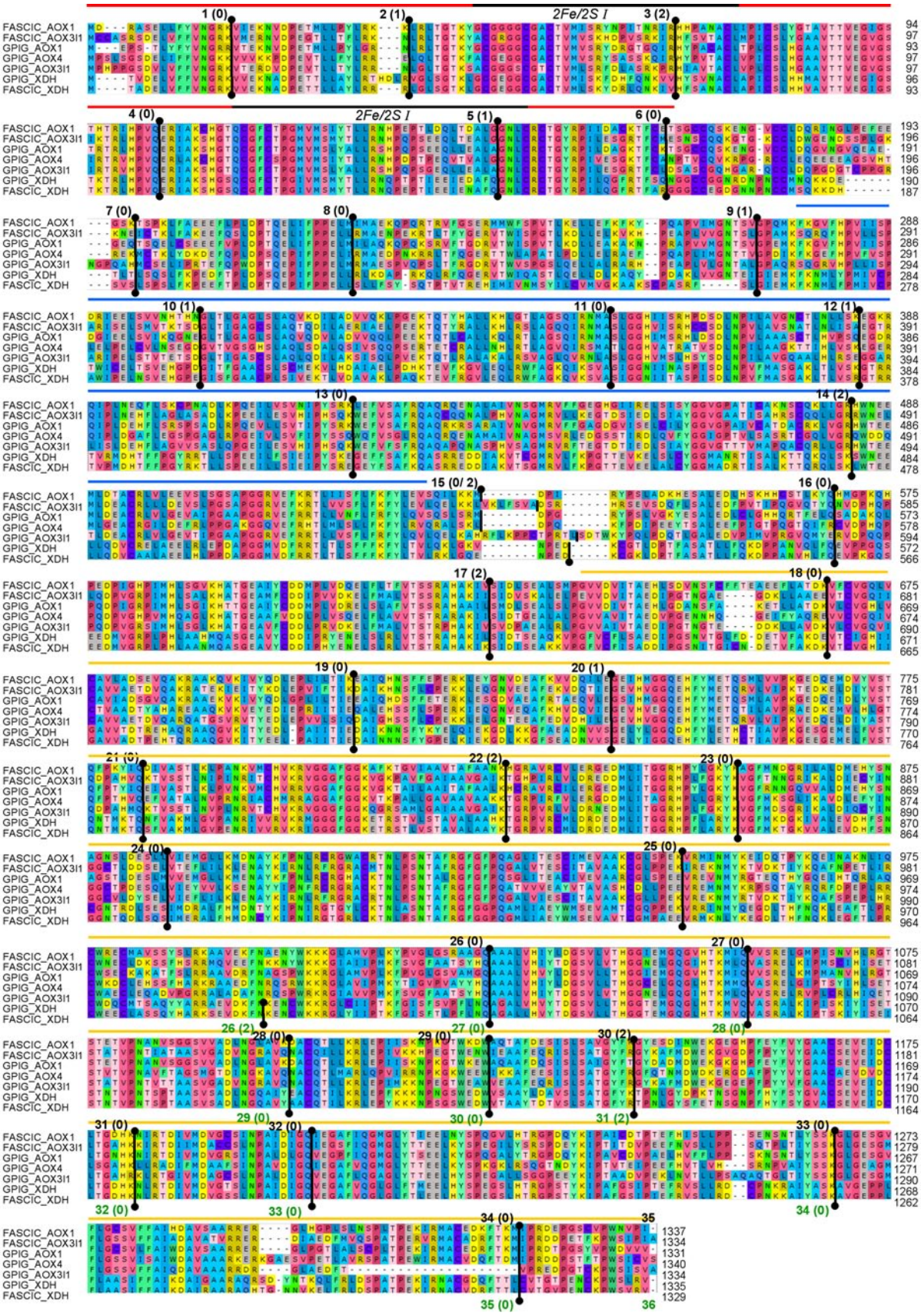
Sequencing of the full-length cDNAs confirmed the predicted structures of guinea pig *AOX1*, *AOX4* and *AOX3l1* genes almost entirely (Fig. 8b and Online Resource 4, Suppl. Tables 5–7). The exon length and exon–intron boundaries of the 35 coding exons of the three genes are strictly conserved, as demonstrated by the position of the exon junctions along the sequence of the corresponding protein products (Fig. 9). The only exception is represented by exon 34 of *AOX3L1*, which is much shorter than predicted from the sequence alignment on the genome. As expected, there is remarkable conservation also with the exon structure of the predicted guinea pig *XDH* gene. This last gene is part of a different contiguous and is likely to reside on a distinct chromosome relative to the *AOX* cluster. This confirms the absence of genomic sequences with similarity to AOX3 in guinea pig. The primary structure of guinea pig AOX1, AOX4, and AOX3L1 proteins (Fig. 9) show the same level of amino acid identities observed in other mammalian species [1] and have the typical structure of all molybdo-flavoenzymes consisting of three conserved domains (25, 40, and 85 kDa from the N- to the C-terminus) separated by non-conserved hinge regions.

Fig. 9 Alignment of the guinea pig and *Macaca fascicularis* AOX1, AOX4, AOX3L1, and XDH proteins. The indicated amino acid sequences were aligned with the CLUSTAL-W algorithm and the regions corresponding to the two 2Fe/2S redox centers are indicated with a *black line* above the sequences. The *red*, *blue*, and *yellow lines* indicate the 25-kDa amino-terminal domain, the 45-kDa NAD-binding domain and the 85-kDa MoCo and substrate-binding domain, respectively. The *vertical lines* connecting the *dots* indicate the position of the exon–intron junctions. The *AOX* exon number and intron junction phase (*in parenthesis*) are indicated in *black* above the sequences, while the XDH exon numbers from 26 on are indicated underneath the sequences in *green*. FASCIC, *Macaca fascicularis*; GPIG, guinea pig. The color code of the amino acids is defined by the Ugene (http://ugene.unipro.ru/) software used for the alignment ▶

Early pseudogenization of AOX3 and late pseudogenization/deletion of AOX4 and AOX3L1 in Strepsirrhini and Haplorrhini

The first branching of Primates occurred approximately 68 Myr ago with the appearance of Strepsirrhini (wet-nose primates) and Haplorrhini (dry-nose primates). The Strepsirrhini suborder subsequently branched into the Lorisiformes and Lemuriformes infraorders between 62 and 65 Myr ago. In the Lorisiformes, *Otolemur garnettii* (bushbaby) three active genes, *AOX1*, *AOX4*, and *AOX3L1*, and one pseudogene, *AOX3*, are predicted (Fig. 10). The same complement of genes and pseudogenes is present in the Lemuriformes, *Microcebus murinus* (mouse lemur). The *AOX3* pseudogenization may be the result of the same event that gave rise to the AOX3 pseudogenes observed in Chiroptera and Scandentia. Thus, we propose that pseudogenization of AOX3 preceded the evolution of Primates.

Haplorrhini split in the two distinct infraorders of Tarsiiformes and Simiiformes, which consists of the Platyrrhini (New World monkeys) and Catarrhini (Old World monkeys) parvorders, about 40 Myr ago [41, 42]. The full complement of *AOX* genes and pseudogenes in the Tarsiiformes, *Tarsius syrichta* (tarsier) is unknown. Nevertheless, the available data are consistent with the presence of at least three active genes, i.e., *AOX1*, *AOX4*, and *AOX3L1*. Currently, it is unclear whether *AOX3* is a gene or a pseudogene, although we favor the second hypothesis on the basis of what observed above. The infraorder of Platyrrhini branched into the superfamilies of Cebidae, Pitheciidae, and Atelidae approximately 25 Myr ago [43]. In Cebidae, both *Saimiri boliviensis*, (squirrel monkey) and *Callitrix jacca* (marmoset) are characterized by two active genes (*AOX1*, *AOX3L1*) and two pseudogenes, (*AOX3*, *AOX4*) (Fig. 10). Thus, *AOX4* is a pseudogene in Platyrrhini, whereas it is an active gene in Lorisiformes and Lemuriformes.


1 (0)
2 (1)
2Fe/2S I
3 (2)
4 (0)
2Fe/2S I
5 (1)
6 (0)
7 (0)
8 (0)
9 (1)
10 (1)
11 (0)
12 (1)
13 (0)
14 (2)
15 (0/ 2)
16 (0)
17 (2)
18 (0)
19 (0)
20 (1)
21 (0)
22 (2)
23 (0)
24 (0)
25 (0)
26 (0)
27 (0)
26 (2)
28 (0)
27 (0)
29 (0)
28 (0)
30 (2)
28 (0)
31 (0)
32 (0)
30 (0)
31 (2)
33 (0)
32 (0)
33 (0)
34 (0)
35
34 (0)
35 (0)
36
FASCIC_AOX1
FASCIC_AOX3I1
GPIG_AOX1
GPIG_AOX4
GPIG_AOX3I1
GPIG_XDH
FASCIC_XDH

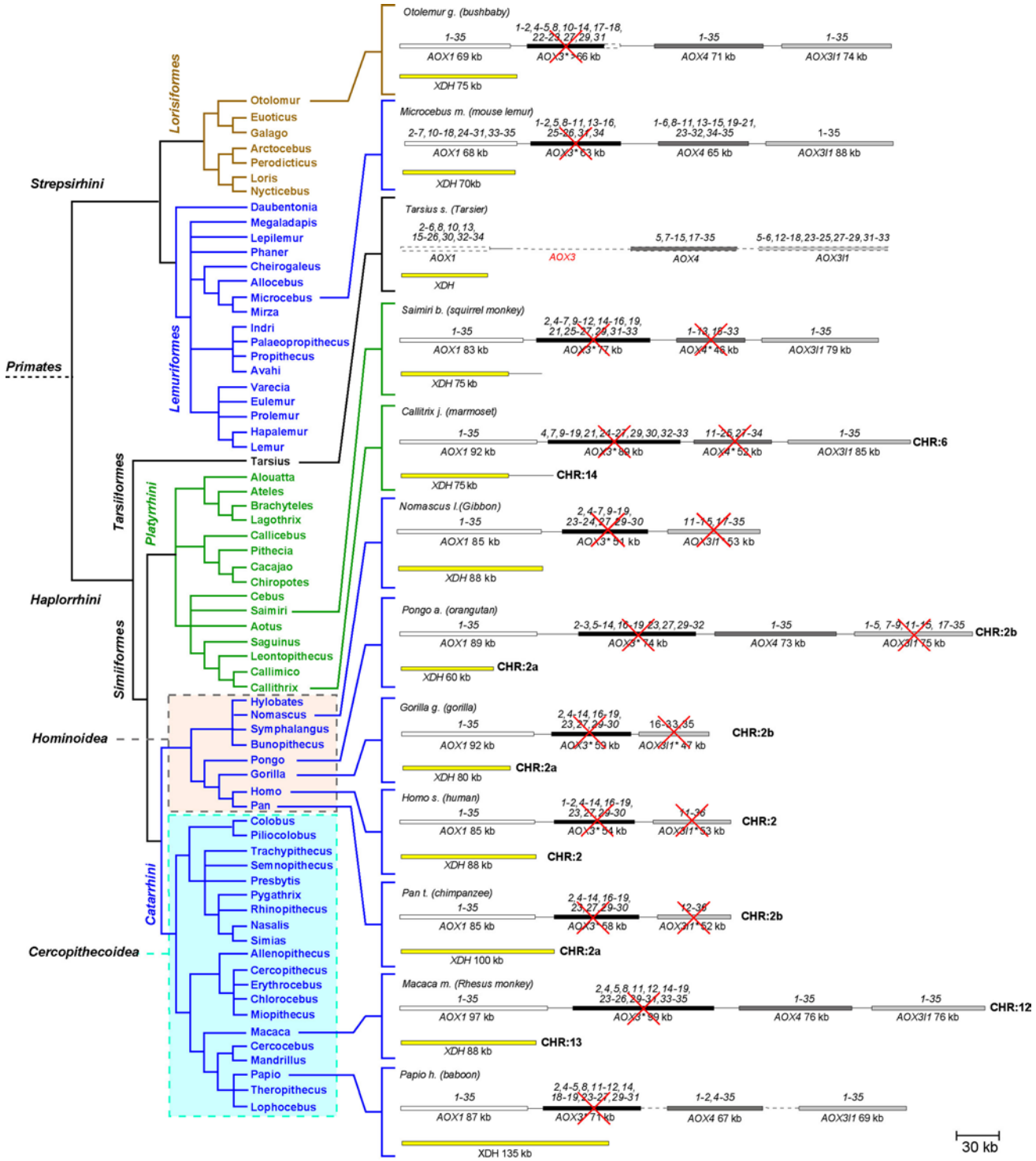


Fig. 10 *AOX* and *XDH* genes in Primates. The figure shows a schematic representation of the *AOX* and *XDH* genes in the indicated mammalian species belonging to the order of Primates for which complete or almost complete genomic sequence data are available. Orthologous genes are indicated with the *same color*. Predicted pseudogenes are *crossed through* and *asterisked*. The number of exons in the active *AOX* genes are indicated along with the exons identified in the corresponding pseudogenes. *Dashed boxes* are used when the length of the genes is unknown. *Dashed lines* are used when the intragenic distances are undetermined. When available, the chromosomal location is shown on the *right*. Genes whose presence is only hypothesized are marked in *red*. When available, the chromosomal location is shown on the *right*. The simplified rooted phylogenetic diagram on the *left* illustrates the relative evolutionary distance between the families the animal species analyzed belong to. The Latin names of each species are followed by the corresponding common names in *parenthesis*

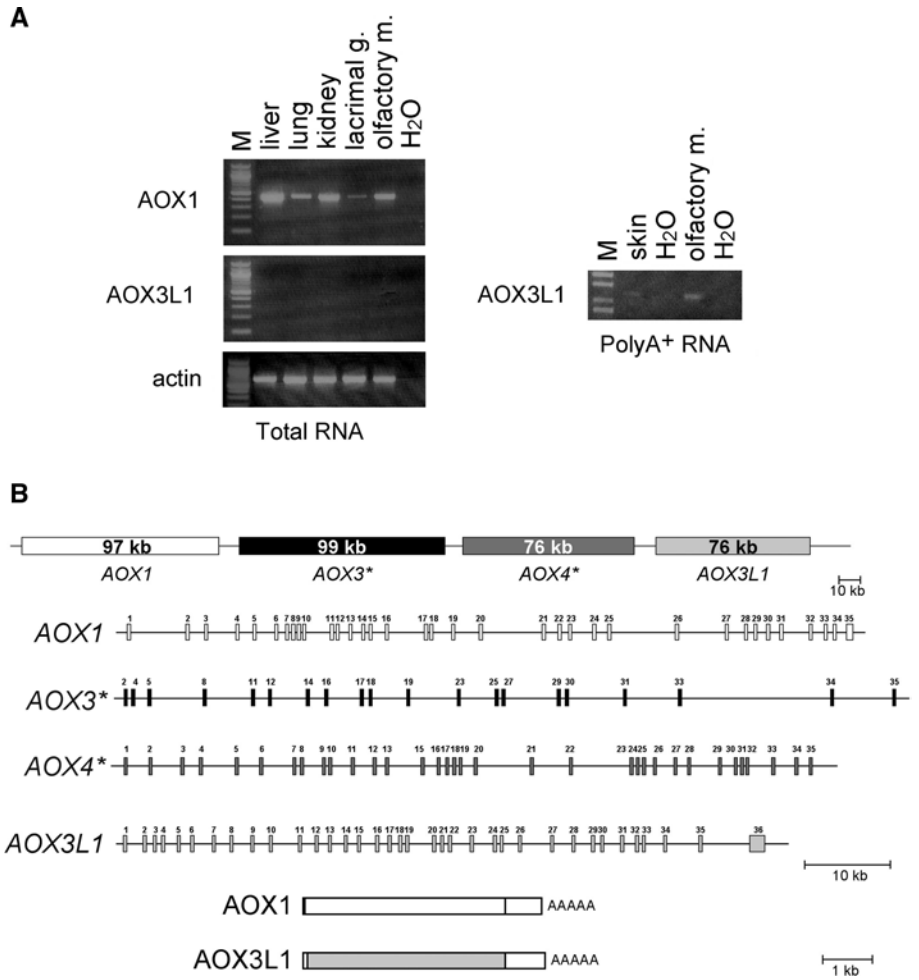


Fig. 11 Tissue-specific expression and exon–intron structures of the *Macaca fascicularis AOX1* and *AOX3L1* genes and transcripts. **a** A comparison of the tissue distribution of the *Macaca fascicularis* AOX1 and AOX3L1 mRNAs is illustrated. Total or polyadenylated (PolyA$^+$) RNA was extracted from the indicated tissue and amplified by RT/PCR using specific couples of amplimers for the two transcripts. Actin is used as an internal control of the experiment. Lacrimal g., lacrimal glands; olfactory m., olfactory mucosa; M, molecular weight markers. **b** *Upper* Schematic representation of the *Macaca fascicularis* cluster of *AOX* genes. *Middle* The exon–intron structures of the *Macaca fascicularis AOX1*, and *AOX3L1* genes, as well as the *AOX3** and *AOX4** pseudogenes were deduced by comparison of the corresponding cDNA and genomic sequences. A schematic representation of the four loci is illustrated and exons are *numbered*. *Lower* A schematic representation of the AOX1 and AOX3L1 transcripts is illustrated. The *central box* corresponding to the coding region lays in between the boxes corresponding to the 5′ and 3′ ends of the transcript. *AAAAA* polyadenylated tail

The AOX profile in Catharrini, cloning of the AOX cDNAs in *Macaca fascicularis*

Approximately 25 Myr ago, Catharrini branched into the superfamilies of Cercopithecoidea and Hominoidea. In the primitive Cercopithecoidea, *Macaca mulatta* (rhesus monkey), the active *AOX1*, *AOX4*, and *AOX3L1* genes as well as the *AOX3* pseudogene are predicted. An identical situation is observed in the more recent Cercopithecoidea, *Papio hamadryas* (hamadryas baboon).

We evaluated the situation in *M. fascicularis* (cynomolgus monkey) directly, by cloning the corresponding cDNAs. The Cercopithecoidea, *M. fascicularis* was selected because it is closely related to *M. mulatta* and it is a popular experimental model [44, 45], providing easy access to tissue samples. First, the best sources for the isolation of the AOX1, AOX3L1, and AOX4 cDNAs were identified screening a selected number of tissues by RT/PCR. AOX1 was easily amplified in liver and other tissues (Fig. 11a). In contrast, AOX3L1 could be amplified only after enrichment for the polyadenylated RNA fraction extracted from the nasal mucosa, indicating that the corresponding protein is present in this organ, albeit in much smaller amounts relative to rodents (Fig. 11a). As observed in mice [22], AOX4 was detectable in the skin and, surprisingly, also in the hepatic tissue (data not shown). The structures of the full-length AOX1 and AOX3L1 cDNAs were elucidated by anchor-PCR

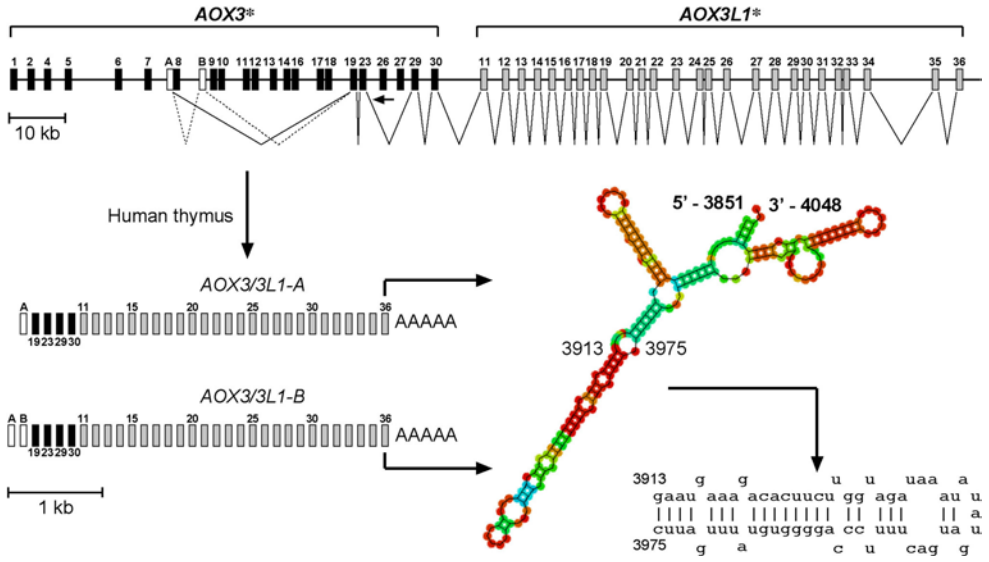


Fig. 12 Structure of the human *AOX3* and *AOX3L1* pseudogene-derived chimeric transcripts representing potential pri-microRNAs. The exon–intron structures of the human *AOX3** and *AOX3L1** pseudogenes was deduced from the sequence of the human genome and is shown on the *top*. Exons derived from AOX3* are shown as *black boxes*, exons derived from *AOX3L1** as *grey boxes* and *numbered*. The two exons A and B originating from *AOX3** introns 7 and 8 are shown as *white boxes*. The two alternative splicing events giving rise to the distinct chimeric and polyadenylated mRNAs (AOX3/3L1-A and AOX3/3L1-B) shown on the lower left are indicated with *solid lines* and *dashed lines* underneath the exons. A diagrammatic representation of the stem-loop structure of the predicted pre-microRNA along with the sequence of the predicted mature microRNA are shown on *lower right*. Residues are indicated in *different colors* according to their pairing probability; from *blue* to *red* = low to high probability

experiments. This allowed determination of the encoded protein sequences (Fig. 9). Except for a few amino acids, AOX1 and AOX3L1 sequences in *Macaca fascicularis* and *M. mulatta* are almost identical. The exon structure of *M. fascicularis AOX1* and *AOX3L1* genes reconstructed from the corresponding cDNAs, using the genome of *M. mulatta* as a template (Fig. 9 and Online Resource 4, Suppl. Tables 8–9), indicates that the former is characterized by 35 exons, while the latter consists of the expected 36 exons [1].

Regardless of its origin (liver or skin), assembly of the full-length AOX4 cDNA never resulted in an ORF of the expected size because of various in-frame stop codons. In addition, some of the cDNA clones isolated were characterized by intron-retention events, which also resulted in short ORFs. Altogether our results indicate that *M. fascicularis AOX4* is transcribed and the corresponding mRNA contains all the expected exons. However, the transcript does not code for an active AOX. This indicates that *M. fascicularis* and *M. mulatta* differ only for *AOX4*. Despite incomplete assembly of the genome, the structure of the *AOX4* pseudogene in *Macaca fascicularis* consists of the canonical 35 exons observed also in *M. mulatta*. Hence, *M. fascicularis AOX4* is the first example of pseudogenization without loss of recognizable exons. These results along with the data obtained in *S. boliviensis* and *C. jacca* indicate that the evolution of Haplorrhini is associated with two distinct *AOX4* pseudogenizations, which occurred separately in Catharrini and Platyrrhini.

Evolution of the AOX loci in Hominoidea: potential microRNA precursors from chimeric AOX3/AOX3L1 pseudogene-derived transcripts in humans

In all the species of Hominoidea analyzed, i.e., *Nomascus leucogenys* (northern white-cheeked gibbon), *Pongo abelii* (sumatran orangutan), *Gorilla gorilla* (western gorilla), *Pan troglodytes* (common chimpanzee) and *Homo sapiens*, *AOX1* is always predicted to be an active gene, while *AOX3* and *AOX3L1* are invariably pseudogenes. In *Pongo a.*, *AOX4* is an active gene consisting of a complete set of exons. No traces of *AOX4* are present in *N. leucogenys*, *G. gorilla*, *P. troglodytes*, or *H. sapiens*, indicating deletion of the gene. The deletion involves also part of the *AOX3L1* from exon 1 to exon 10. Despite being considered a more ancient primate than *Pongo a.*, it is surprising that *N. leucogenys* presents with the same *AOX4* deletion observed in *G. gorilla*, *P. troglodytes*, and *H. sapiens*.

The two human pseudogenes *AOX3* and *AOX3L1* are annotated as a single non-coding locus (*AOX2P*) in the NCBI database. The two human pseudogenes are transcribed and are present as expressed sequence tags (ESTs) in the same database. Most of these ESTs were isolated from the thymus. We noticed that some of the ESTs

identified contained exon sequences deriving from both *AOX3* and *AOX3L1*, consistent with the existence of chimeric transcripts. The corresponding full-length cDNAs were amplified, cloned, and sequenced using RNA from human thymus and derived thymomas. Two different types of cDNAs varying only at the 5′end were identified (Online Resource 4, Suppl. Fig. 2). The first transcript (*AOX3/3L1-A*) is made of 31 exons (Fig. 12). The 5′-region consists of a first exon contained in intron 7 of the *AOX3* pseudogene and it is spliced to *AOX3* exons 19, 23, 29, and 30. The 3′-region consists of exons 11 to 36 of *AOX3L1*. The second transcript (*AOX3/3L1-B*) differs just for an extra exon, which is contained in intron 8 of *AOX3* and is located downstream of the first *AOX3/3L1-A* exon. Both transcripts are polyadenylated and are non-coding.

The precursors of microRNAs (pri-microRNAs) are often transcription products originating from the introns of coding genes [46]. In addition, there are examples of microRNAs deriving from pseudogenes [47, 48]. For these reasons, we evaluated whether the two *AOX3/3L1* transcripts represent potential pri-microRNA that could be subsequently processed into pre-microRNAs and mature microRNAs [49] possessing a predictable stem-loop structure. Two independent algorithms predict the presence of a potential pre-microRNA with a relatively high degree of confidence (Fig. 12). The predicted microRNA is not a mirtron, i.e., a type of microRNA located in the introns of mRNA-encoding host genes, as it maps to exon 36 of the *AOX3L1* pseudogene. If validated experimentally, the observation would indicate that pseudogenization of human *AOX3* and *AOX3L1* resulted in the acquisition of new regulatory functions. The absence of *AOX3/3L1-A* and *AOX3/3L1-B* in multiple human tissues using both in silico (EST databases) and experimental (RT/PCR amplification, data not shown) approaches suggests thymus-specific expression of these RNAs.

Identification of isoenzyme-specific amino acid residues by comparison of mammalian AOXs

The data collected on AOX proteins in multiple mammalian species are of major relevance not only from an evolutionary, but also from a structural perspective. In fact, comparison of the predicted amino acid sequences permits identification of conserved and common residues as well as amino acids specific to each mammalian AOX isoenzyme. A consensus alignment based on 31 AOX1, 21 AOX4, 7 AOX3 and 25 AOX3L1 from many different lineages is presented in Fig. 13.

The amino terminal 25-kDa domain of AOXs (mouse AOX1 = Met1-Lys165; Fig. 13 = Met1-aa175) contains two non-identical *2Fe/2S* centers (*2Fe/2S[I]* and *2Fe/2S[II]*). In both *2Fe/2S[I]* and *2Fe/2S[II]*, the Cys residues involved in the coordination of the iron atoms are invariably conserved in all mammalian AOXs (see black dots underneath the Cys residues in Fig. 13). The only exception is elephant AOX3L1 whose Cys48 is substituted by an Arg residue (see red dot underneath the relevant Cys residue in Fig. 13). Since this difference is likely to result in an inactive enzyme, the finding should be confirmed to rule out a sequencing error or a SNP. The domain does not contain amino acid residues that are specific to AOX1, AOX3, AOX4, or AOX3L1. This is consistent with a predicted strong similarity in the secondary structure of the AOX1, AOX3, AOX4, or AOX3L1 25-kDa domains (CFSSP prediction server, http://www.biogem.org; Suppl Fig. 3).

The sequence of the hinge region (mouse AOX1 = Ala166-Pro230; Fig. 13 = aa176-aa250) separating the 25-kDa from the 45-kDa FAD-binding domain is variable. Indeed, only nine residues are conserved in all mammalian AOXs. Six of the conserved amino acids lay in adjacent positions (mouse AOX1 = Ile216-Leu221; Fig. 13 = Ile235-Leu240), suggesting definition of a structurally important subdomain. The hinge region contains an amino acid that is specific to AOX3 (mouse AOX3 = Val197; Fig. 13 = Val215) and is predominantly substituted by a Thr in AOX1 or a non-polar residue in AOX4 and AOX3L1. A residue specific to AOX4 is also observed (mouse AOX4 = Ile224; Fig. 13 = Ile241).

The FAD-binding domain is divided into two subdomains (FAD1 and FAD2) corresponding to Lys231-Ser418 and Arg419-Leu533 of mouse AOX1. The first sub-domain contains most of the residues involved in FAD binding [3]. In XDH, this sub-domain is also characterized by a NAD-binding site (Phe390 to Glu402 of bovine XDH), which is absent in AOXs [1, 3]. In all mammalian AOXs, the Tyr residue (bovine XDH = Tyr393; Fig. 13 = X428) critical for NAD binding is invariably substituted. Two Gln residues (mouse AOX3 = Gln388, Gln409; Fig. 13 = Gln415, Gln437) in the first sub-domain are specific to AOX3, since the equivalent positions are occupied by Arg and Glu residues in AOX1, AOX4, and AOX3L1. Conversely, the second sub-domain is characterized by an AOX4-specific Asp residue (mouse AOX4 = Asp489; Fig. 13 = Asp530). This Asp residue is substituted by the equally acidic amino acid, Glu, in the majority of the other AOXs.

Hinge region 2 is the most variable and does not contain amino acids specific to any of the AOX isoenzymes. The 85-kDa domain contains the substrate pocket, which lays close to the MoCo-binding region, and it is divided into four sub-domains. In AOXs, sub-domain I (mouse AOX1 = Pro575-Pro697 and Glu744-Gly842) and sub-domain IV (mouse AOX1 = Phe1009-Val1158) contain not only the determinants necessary for substrate access and positioning but also residues involved in the dimerization of the two monomeric subunits, as observed in the case of XDH [1, 3, 6, 50–52]. Sub-domains II

(mouse AOX1 = Leu698-Gln743 and Arg843-Phe963) and III (mouse AOX1 = Ser964-Lys1008 and Phe1159-Val1333) are equally important for the interactions with the substrate and are exposed to the solvent. Sub-domain I presents with one AOX1- (mouse AOX1 = Leu646; Fig. 13 = Leu706), one AOX4- (mouse AOX4 = Gly772; Fig. 13 = Gly834) and three AOX3L1-specific residues (mouse AOX3L1 = Arg763, Lys792 and Thr793; Fig. 13 = Arg817, Lys843, and Thr844). In sub-domain II, two amino acids are specific to each isoenzyme (mouse AOX1 = Tyr846 and Gly885; Fig. 13 = Tyr911 and Gly950. Mouse AOX3 = Ile713 and Leu949; Fig. 13 = Ile777 and Leu1013. Mouse AOX4 = Gly772 and Ser892; Fig. 13 = Gly834 and Ser954. Mouse AOX3L1 = Leu879 and Ala973; Fig. 13 = Leu950 and Ala1027). AOX4-specific amino acids (mouse AOX4 = Ala1190, Glu1241, and Val1251; Fig. 13 = Ala1274, Glu1327 and Val1337) dominate in sub-domain III, where three such residues compare with a single one for AOX1 (mouse AOX1 = Leu1301; Fig. 13 = Leu1405), AOX3 (mouse AOX3 = Thr1250; Fig. 13 = Thr1338) and AOX3L1 (mouse AOX3L1 = Asp1305; Fig. 13 = Asp1418), respectively. The sub-domain with the highest density of residues specific for any single AOX isoenzyme is sub-domain IV. Here, six residues specific to AOX1 (mouse AOX1 = Pro1010, Ser1015, Ile1040, Arg1068, Gly1069, and Glu1073; Fig. 13 = Pro1075, Ser1080, Ile1136, Arg1150, Gly1151, and Glu1155), three residues specific to AOX4 (mouse AOX4 = Cys1043, Met1088, and Thr1145; Fig. 13 = Cys1116, Met1167, and Thr1224) and two residues specific to AOX3 (mouse AOX3 = Lys1016 and Met1071; Fig. 13 = Lys1080 and Met1152) are observed. No amino acids specific to AOX3L1 are evident in this region. Thus, the 85-kDa domain is characterized by the highest concentration of AOX isoenzyme-specific amino acids. The observation is consistent with the idea that AOX1, AOX3, AOX4, and AOX3L1 have overlapping, but different substrate specificities.

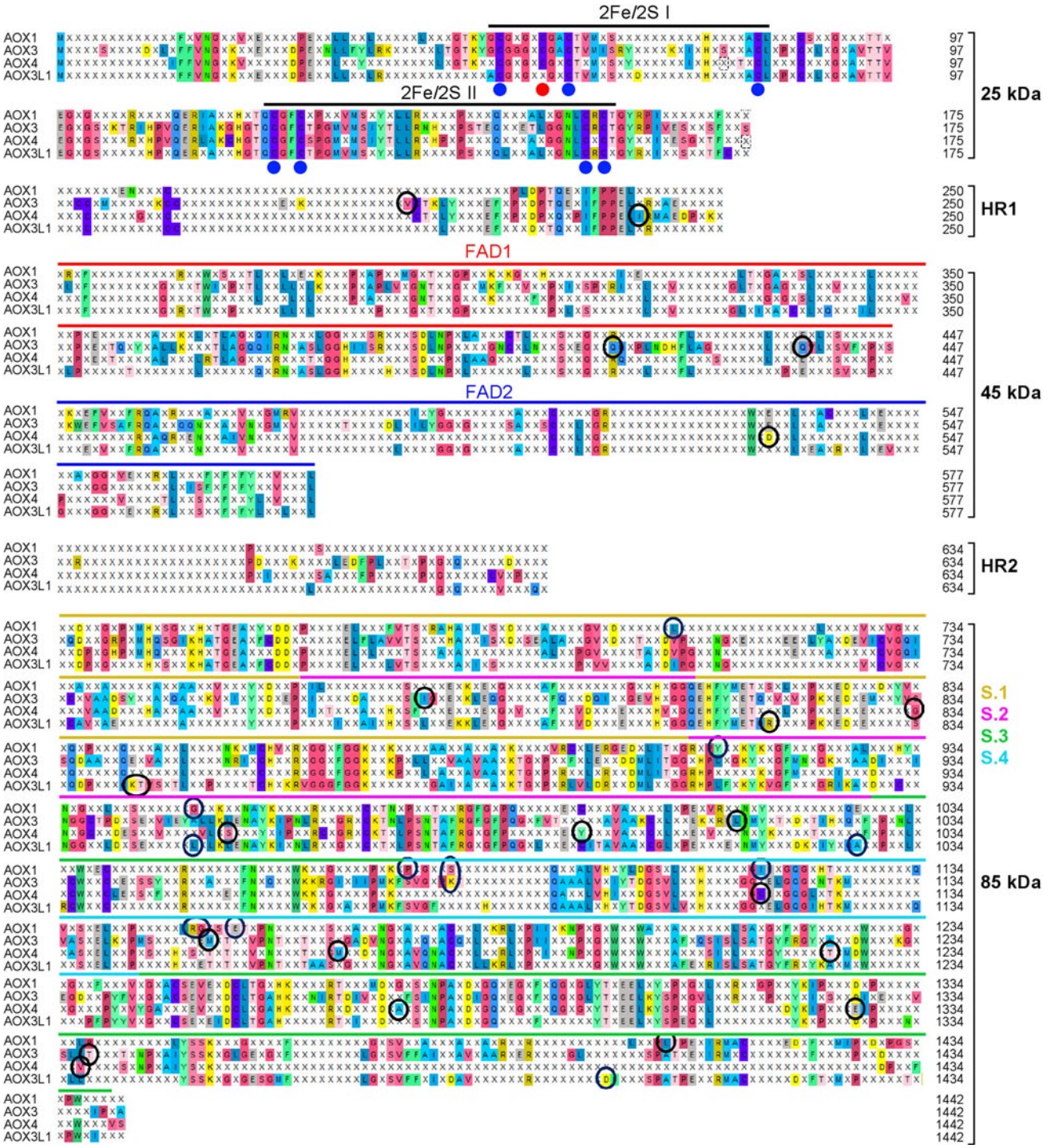
The data obtained will be instrumental in the definition of differences and similarities in the ability of the AOX isoenzymes to accommodate substrates inside the binding pocket, since the coordinates of the first crystallized mammalian AOX have become available [5].

Conclusions

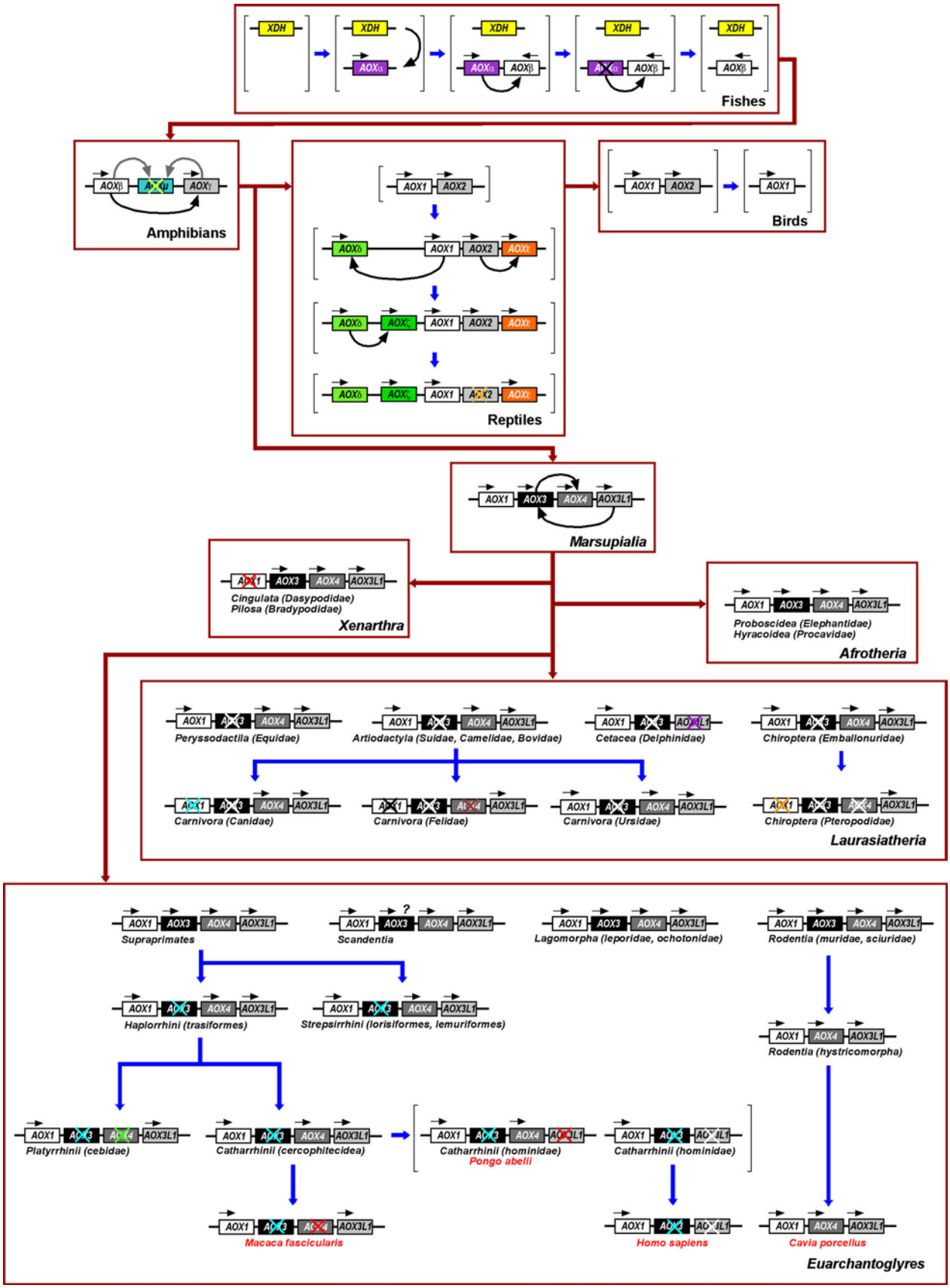
According to our reconstruction (Fig. 14), the most ancient vertebrate AOX enzyme is fish AOXα, which originated by duplication of an ancestral fish *XDH*. AOXα duplicated into *AOXβ*, the direct precursor of reptilian, avian, and mammalian *AOX1*. Further evolution of fishes resulted in the duplication of *AOXα* into *AOXβ* and successive pseudogenization of *AOXα*. *AOXβ* was maintained in amphibians and it is the likely precursors of reptilian, avian, and mammalian *AOX1*. In turn, amphibian *AOXγ* is a duplication of *AOXβ* and the likely ancestor of reptilian and avian *AOX2*, the proposed precursor of mammalian *AOX3L1*. The appearance of AOXγ/AOX2/AOX3L1 proteins may be related to the evolution of the olfactory system in terrestrial and avian vertebrates. In fact, Bowman's glands, which are major nasal structures in reptiles, birds, and mammals [53, 54], are the richest source of AOX3L1 in rodents (Fig. 6b) [18]. The evolution of reptiles is associated with three further and reptile-specific duplications resulting in the appearance of *AOXδ*, *AOXε*, and *AOXζ*. In mammals, the gene duplication processes resulting in the appearance of the four *AOX* genes so far identified was completed during the evolution of Monotreamata or Marsupialia with the acquisition of *AOX3* and *AOX4*. On the basis of the available data, we suggest the following evolutionary sequence from the most ancient to the most recent mammalian gene, i.e., *AOX1*, *AOX3L1*, *AOX3*, and *AOX4*.

The picture of further mammalian evolution is dominated by a series of lineage-specific pseudogenizations. Inactivation of *AOX3* is a very frequent event, as at least four lineage specific pseudogenizations are evident in Afrotheria, Xenarthra, Laurasiatheria, and Euarchontoglires. Rodentia are unique in terms of *AOX3* evolution, as this order is generally characterized by the presence of

Fig. 13 Consensus sequences of mammalian AOX1, AOX3, AOX4 and AOX3L1 proteins. An alignment of the consensus sequences determined for mammalian AOX1, AOX3, AOX4 and AOX3L1 proteins deduced after comparison of all the available primary structures is illustrated. This alignment is based on 31 AOX1 (3 Marsupialia, 1 Hyracoydea, 1 Proboscidea, 1 Chiroptera, 2 Carnivora, 3 Artiodactyla, 1 Cetacea, 1 Perossidactyla, 2 Lagomorpha, 6 Rodentia, 10 Primates), 7 AOX3 (3 Marsupialia, 1 Proboscidea, 1 Lagomorpha, 2 Rodentia), 21 AOX4 (1 Monotreamata, 2 Marsupialia, 1 Proboscidea, 1 Chiroptera, 3 Carnivora, 1 Artiodactyla, 1 Perissodactyla, 1 Lagomorpha, 6 Rodentia, 4 Primates) and 25 AOX3L1 (2 Marsupialia, 1 Proboscidea, 1 Chiroptera, 3 Carnivora, 2 Artiodactyla, 1 Perossidactyla, 2 Lagomorpha, 6 Rodentia, 7 Primates). The AOX structural domains are indicated on the *right*. *HR1* and *HR2* = hinge regions 1 and 2. The alignment position is indicated by the *numbers* shown on the *right*. All the *boxed* and indicated residues are conserved across the species in at least one of the AOX isoforms. Amino acid residues specific to only one of the four AOX isoforms are circled in *black*. When the residues are not conserved they are indicated with an "*X*". The two 2Fe/2S subdomains are indicated by a black line above the sequences. The *black dots* underneath the sequences of the two domains indicate the Cys residues involved in the coordination of the iron atoms. The amino acid corresponding to the non-conserved Cys residue observed in elephant AOX3L1 is indicated by a *red dot*. The two FAD-binding sites (FAD1 and FAD2) are indicated with a *red* and a *dark blue line*, respectively. The substrate pocket subdomains (S1–S4) are indicated above the sequences with lines of *different colors*. Color coding of the amino acid residues is the same as in Fig. 9



an active gene, with a few exceptions. Pseudogenization or deletion events involving *AOX4* are the second most frequent occurrence in Eutheria being particularly widespread in Primates. In Primates, pseudogenization or deletion of *AOX4* may be related to the well-known involution of the Harderian glands. Pseudogenization of *AOX1*, the most ancient gene of the family, is limited to a few species, such as dogs and cats, suggesting that the two encoded proteins exert important homeostatic functions in mammals. By the same token, *AOX3L1* is an active gene in all mammals, with the notable exception of dolphins and Hominoidea. Inactivation of *AOX3L1* in Hominoidea, but not other types of primates, may be related to differences in the evolution of the peripheral olfactory system, as the nasal mucosa is the richest source of the corresponding protein.


XDH
AOXα
AOXβ
Fishes
AOXβ
AOXμ
AOXγ
Amphibians
AOX1
AOX2
AOXδ
AOXε
AOXζ
Reptiles
Birds
AOX3
AOX4
AOX3L1
Marsupialia
Cingulata (Dasypodidae)
Pilosa (Bradypodidae)
Xenarthra
Proboscidea (Elephantidae)
Hyracoidea (Procavidae)
Afrotheria
Perissodactila (Equidae)
Artiodactyla (Suidae, Camelidae, Bovidae)
Cetacea (Delphinidae)
Chiroptera (Emballonuridae)
Carnivora (Canidae)
Carnivora (Felidae)
Carnivora (Ursidae)
Chiroptera (Pteropodidae)
Laurasiatheria
Supraprimates
Scandentia
Lagomorpha (leporidae, ochotonidae)
Rodentia (muridae, sciuridae)
Haplorrhini (trasiformes)
Strepsirrhini (lorisiformes, lemuriformes)
Rodentia (hystricomorpha)
Platyrrhinii (cebidae)
Catharrhinii (cercophitecidea)
Catharrhinii (hominidae)
Pongo abelii
Catharrhinii (hominidae)
Macaca fascicularis
Homo sapiens
Cavia porcellus
Euarchantoglyres

◀ **Fig. 14** Evolutionary history of the vertebrate *AOX* genes. A schematic representation of the evolutionary history of the vertebrate *AOX* genes is shown. The various *AOX* genes are represented as *colored boxes*. Predicted *AOX* orthologues or *AOX* genes showing an ancestor/descendant relationship are drawn in the *same color*. Pseudogenes are *crossed*. *Crosses* drawn in the *same color* indicate predicted derivation from the same pseudogenization event. The *black arrows* above the genes indicate the direction of transcription. The *curved black arrows* indicate a duplication event. The *two arrows* referring to reptile *AOXζ* are shown in *grey* to indicate that it is currently unknown whether the gene is a duplication of *AOX1* or *AOXδ*. The *question mark* above *AOX3* in Scandentia indicates that it is currently unclear whether the locus is a gene or a pseudogene

In conclusion, the present data provide a more detailed picture of the evolution of vertebrate AOX genes and proteins. This is a first step along the determination of the biological and functional significance of this highly conserved family of proteins.

Acknowledgments Grants from the Telethon-Italy Foundation, the Fondazione Cariplo, the Fondazione Italo Monzino, and the Negri-Weizmann Foundation were fundamental for the completion of this work. The work was also partially supported with grants from the Associazione Italiana per la Ricerca contro il Cancro (AIRC). We would like to thank the talented student, Valentina Novelli for sequence analysis. We acknowledge the help of Felice Deceglie with the artwork.

References

1. Garattini E, Fratelli M, Terao M (2008) Mammalian aldehyde oxidases: genetics, evolution and biochemistry. Cell Mol Life Sci 65(7–8):1019–1048
2. Garattini E, Fratelli M, Terao M (2009) The mammalian aldehyde oxidase gene family. Hum Genomics 4(2):119–130
3. Garattini E et al (2003) Mammalian molybdo-flavoenzymes, an expanding family of proteins: structure, genetics, regulation, function and pathophysiology. Biochem J 372(Pt 1):15–32
4. Garattini E, Terao M (2011) Increasing recognition of the importance of aldehyde oxidase in drug development and discovery. Drug Metab Rev 43(3):374–386
5. Coelho C et al (2012) The first mammalian aldehyde oxidase crystal structure: insights into substrate specificity. J Biol Chem 287(48):40690–40702
6. Enroth C et al (2000) Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: structure-based mechanism of conversion. Proc Natl Acad Sci USA 97(20):10723–10728
7. Garattini E, Terao M (2012) The role of aldehyde oxidase in drug metabolism. Expert Opin Drug Metab Toxicol 8(4):487–503
8. Sekimoto H et al (1997) Cloning and molecular characterization of plant aldehyde oxidase. J Biol Chem 272(24):15280–15285
9. Seo M et al (2004) Comparative studies on the *Arabidopsis* aldehyde oxidase (AAO) gene family revealed a major role of AAO3 in ABA biosynthesis in seeds. Plant Cell Physiol 45(11):1694–1703
10. Seo M et al (2000) The *Arabidopsis* aldehyde oxidase 3 (AAO3) gene product catalyzes the final step in abscisic acid biosynthesis in leaves. Proc Natl Acad Sci USA 97(23):12908–12913
11. Yang L et al (2009) Cloning and expression analysis of an aldehyde oxidase gene in *Arachis hygogaea* L. J Environ Biol 30(1):93–98
12. Taylor NJ, Cowan AK (2004) Xanthine dehydrogenase and aldehyde oxidase impact plant hormone homeostasis and affect fruit size in 'Hass' avocado. J Plant Res 117(2):121–130
13. Yesbergenova Z et al (2005) The plant Mo-hydroxylases aldehyde oxidase and xanthine dehydrogenase have distinct reactive oxygen species signatures and are induced by drought and abscisic acid. Plant J 42(6):862–876
14. Hartmann T et al (2012) The impact of single nucleotide polymorphisms on human aldehyde oxidase. Drug Metab Dispos 40(5):856–864
15. Pryde DC et al (2010) Aldehyde oxidase: an enzyme of emerging importance in drug discovery. J Med Chem 53(24):8441–8460
16. Cazzaniga G et al (1994) Chromosomal mapping, isolation, and characterization of the mouse xanthine dehydrogenase gene. Genomics 23(2):390–402
17. Kurosaki M et al (1999) Molecular cloning of the cDNA coding for mouse aldehyde oxidase: tissue distribution and regulation in vivo by testosterone. Biochem J 341(Pt 1):71–80
18. Kurosaki M et al (2004) The aldehyde oxidase gene cluster in mice and rats. Aldehyde oxidase homologue 3, a novel member of the molybdo-flavoenzyme family with selective expression in the olfactory mucosa. J Biol Chem 279(48):50482–50498
19. Terao M et al (2001) Purification of the aldehyde oxidase homolog 1 (AOH1) protein and cloning of the AOH1 and aldehyde oxidase homolog 2 (AOH2) genes. Identification of a novel molybdo-flavoprotein gene cluster on mouse chromosome 1. J Biol Chem 276(49):46347–46363
20. Terao M et al (2000) Cloning of the cDNAs coding for two novel molybdo-flavoproteins showing high similarity with aldehyde oxidase and xanthine oxidoreductase. J Biol Chem 275(39):30690–30700
21. Obach RS et al (2004) Human liver aldehyde oxidase: inhibition by 239 drugs. J Clin Pharmacol 44(1):7–19
22. Terao M et al (2009) Role of the molybdoflavoenzyme aldehyde oxidase homolog 2 in the biosynthesis of retinoic acid: generation and characterization of a knockout mouse. Mol Cell Biol 29(2):357–377
23. Rodriguez-Trelles F, Tarrio R, Ayala FJ (2003) Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. Proc Natl Acad Sci USA 100(23):13413–13417
24. Terao M et al (2006) Avian and canine aldehyde oxidases. Novel insights into the biology and evolution of molybdo-flavoenzymes. J Biol Chem 281(28):19748–19761
25. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704
26. Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14(7):685–695
27. St John JA et al (2012) Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. Genome Biol. 13(1):415
28. Tyagi S et al (2008) CID-miRNA: a web server for prediction of novel miRNA precursors in human genome. Biochem Biophys Res Commun 372(4):831–834
29. Gess RW, Coates MI, Rubidge BS (2006) A lamprey from the Devonian period of South Africa. Nature 443(7114):981–984
30. San Mauro D (2010) A multilocus timescale for the origin of extant amphibians. Mol Phylogenet Evol 56(2):554–561
31. Bininda-Emonds OR et al (2007) The delayed rise of present-day mammals. Nature 446(7135):507–512
32. Price SA et al (2012) Tempo of trophic evolution and its impact on mammalian diversification. Proc Natl Acad Sci USA 109(18):7008–7012

33. Davies TJ et al (2008) Colloquium paper: phylogenetic trees and the future of mammalian biodiversity. Proc Natl Acad Sci USA 105(Suppl 1):11556–11563
34. Fritz SA, Bininda-Emonds OR, Purvis A (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecol Lett 12(6):538–549
35. Barrow EC, Seiffert ER, Simons EL (2010) A primitive hyracoid (Mammalia, Paenungulata) from the early Priabonian (Late Eocene) of Egypt. J Syst Palaeontol 8(2):213–244
36. Pumo DE et al (1998) Complete mitochondrial genome of a neotropical fruit bat, *Artibeus jamaicensis*, and a new hypothesis of the relationships of bats to other eutherian mammals. J Mol Evol 47(6):709–717
37. Gingerich PD et al (2001) Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. Science 293(5538):2239–2242
38. Nyakatura K, Bininda-Emonds OR (2012) Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biol 10:12
39. Springer MS et al (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci USA 100(3):1056–1061
40. Eizirik E et al (2010) Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Mol Phylogenet Evol 56(1):49–63
41. Groves C (1998) Primate evolution—in and out of Africa. Curr Biol 8(21):R747 (author reply 747–748)
42. Schrago CG, Russo CA (2003) Timing the origin of New World monkeys. Mol Biol Evol 20(10):1620–1625
43. Wildman DE et al (2009) A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). Mol Phylogenet Evol 53(3):694–702
44. Chono H et al (2011) In vivo safety and persistence of endoribonuclease gene-transduced CD4+ T cells in cynomolgus macaques for HIV-1 gene therapy model. PLoS One 6(8):e23585
45. Van Rompay KK (2012) The use of nonhuman primate models of HIV infection for the evaluation of antiviral strategies. AIDS Res Hum Retroviruses 28(1):16–35
46. Smalheiser NR (2003) EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. Genome Biol 4(7):403
47. Devor EJ (2006) Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. J Hered 97(2):186–190
48. Guo X et al (2009) Small RNAs originated from pseudogenes: *cis*- or *trans*-acting? PLoS Comput Biol 5(7):e1000449
49. Ambros V (2004) The functions of animal microRNAs. Nature 431(7006):350–355
50. Eger BT et al (2000) Purification, crystallization and preliminary X-ray diffraction studies of xanthine dehydrogenase and xanthine oxidase isolated from bovine milk. Acta Crystallogr D Biol Crystallogr 56(Pt 12):1656–1658
51. Kuwabara Y et al (2003) Unique amino acids cluster for switching from the dehydrogenase to oxidase form of xanthine oxidoreductase. Proc Natl Acad Sci USA 100(14):8170–8175
52. Nishino T, Okamoto K (2000) The role of the [2Fe–2s] cluster centers in xanthine oxidoreductase. J Inorg Biochem 82(1–4): 43–49
53. Rehorek SJ, Firth BT, Hutchinson MN (2000) The structure of the nasal chemosensory system in squamate reptiles. 2. Lubricatory capacity of the vomeronasal organ. J Biosci 25(2):181–190
54. Rehorek SJ, Firth BT, Hutchinson MN (2000) The structure of the nasal chemosensory system in squamate reptiles. 1. The olfactory organ, with special reference to olfaction in geckos. J Biosci 25(2):173–179