

Research article

Open Access

## The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event?

Véronique Douard<sup>†1</sup>, Frédéric Brunet<sup>†2</sup>, Bastien Boussau<sup>3</sup>, Isabelle Ahrens-Fath<sup>4</sup>, Virginie Vlaeminck-Guillem<sup>2</sup>, Bernard Haendler<sup>4</sup>, Vincent Laudet<sup>\*2</sup> and Yann Guiguen<sup>1</sup>

Address: <sup>1</sup>INRA-SCRIBE IFR 140, Campus de Beaulieu, 35042 Rennes Cedex, France, <sup>2</sup>Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, UMR 5242 du CNRS, INRA, IFR128 BioSciences Lyon-Gerland, Ecole Normale Supérieure de Lyon, 46, Allée d'Italie, 69364 Lyon Cedex 07, France, <sup>3</sup>Biométrie et Biologie Évolutive UMR CNRS 5558 Université Claude Bernard-Lyon 1, 43, Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France and <sup>4</sup>Bayer Schering Pharma AG, 13342 Berlin, Germany

Email: Véronique Douard - douardve@umdnj.edu; Frédéric Brunet - Frederic.Brunet@ens-lyon.fr; Bastien Boussau - boussau@biomserv.univ-lyon1.fr; Isabelle Ahrens-Fath - isabelle.ahrens-fath@parexel.com; Virginie Vlaeminck-Guillem - Virginie.Vlaeminck.Guillem@ens-lyon.fr; Bernard Haendler - bernard.haendler@bayerhealthcare.com; Vincent Laudet\* - Vincent.Laudet@ens-lyon.fr; Yann Guiguen - yann.guiguen@rennes.inra.fr

\* Corresponding author †Equal contributors

Published: 18 December 2008

Received: 21 June 2008

*BMC Evolutionary Biology* 2008, **8**:336 doi:10.1186/1471-2148-8-336

Accepted: 18 December 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/336>

© 2008 Douard et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Based on the observation of an increased number of paralogous genes in teleost fishes compared with other vertebrates and on the conserved synteny between duplicated copies, it has been shown that a whole genome duplication (WGD) occurred during the evolution of Actinopterygian fish. Comparative phylogenetic dating of this duplication event suggests that it occurred early on, specifically in teleosts. It has been proposed that this event might have facilitated the evolutionary radiation and the phenotypic diversification of the teleost fish, notably by allowing the sub- or neo-functionalization of many duplicated genes.

**Results:** In this paper, we studied in a wide range of Actinopterygians the duplication and fate of the androgen receptor (AR, NR3C4), a nuclear receptor known to play a key role in sex-determination in vertebrates. The pattern of AR gene duplication is consistent with an early WGD event: it has been duplicated into two genes AR-A and AR-B after the split of the Acipenseriformes from the lineage leading to teleost fish but before the divergence of Osteoglossiformes. Genomic and syntenic analyses in addition to lack of PCR amplification show that one of the duplicated copies, AR-B, was lost in several basal Clupeocephala such as Cypriniformes (including the model species zebrafish), Siluriformes, Characiformes and Salmoniformes. Interestingly, we also found that, in basal teleost fish (Osteoglossiformes and Anguilliformes), the two copies remain very similar, whereas, specifically in Percomorphs, one of the copies, AR-B, has accumulated substitutions in both the ligand binding domain (LBD) and the DNA binding domain (DBD).

**Conclusion:** The comparison of the mutations present in these divergent AR-B with those known in human to be implicated in complete, partial or mild androgen insensitivity syndrome suggests that the existence of two distinct AR duplicates may be correlated to specific functional differences that may be connected to the well-known plasticity of sex determination in fish. This suggests that three specific events have shaped the present diversity of ARs in Actinopterygians: (i) early WGD, (ii) parallel loss of one duplicate in several lineages and (iii) putative neofunctionalization of the same duplicate in percomorphs, which occurred a long time after the WGD.

## Background

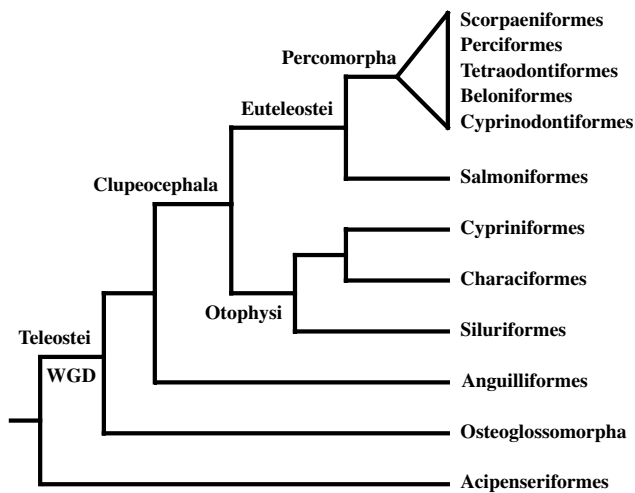
Actinopterygian fishes have provided the first clear demonstration of an ancient whole genome duplication (WGD) in vertebrate evolution [1]. This event was originally suggested based on the finding that zebrafish and medaka possess seven Hox clusters [2-4], compared to four in mammals and one in most invertebrates. It was confirmed later on by comparative mapping [5] and through the analysis of genome sequences of two pufferfishes [1,6]. Indeed, many short duplicated groups of linked genes were identified in the *Takifugu rubripes* and *Tetraodon nigroviridis* genomes [1,7]. The duplication event leading to these duplicates was dated by molecular clock to a window between divergence of Actinopterygians from Tetrapods, and diversification of teleost fish [8,9]. In addition, all chromosomes of *Tetraodon nigroviridis* were assigned to syntenic groups of duplicated genes, demonstrating the genomic scale of the duplication. It was further shown that each pair of duplicated genes was homologous to one non-duplicated human chromosomal region [1].

Direct dating of fish gene duplications based on molecular clock was hampered by saturation of synonymous changes at the time scales considered, as well as by differences in evolutionary rates between mammals and fishes [1,10-13]. Less sensitive to these problems, phylogenies of a few tens of gene families have shown a high frequency of gene duplications to be a common feature among sampled teleosts or euteleosts, but not among other fishes [11,14]. Comparative mapping has recently shown that paralogons are homologous between pufferfishes (which belong to Percomorphs) and zebrafish (a Cypriniform), implying that the whole genome duplication event occurred before the divergence of these two lineages of Euteleosts [7]. In addition, using a small number of genes, Hoegg *et al.* [14] have scrutinized the existence of duplications in basal Actinopterygians and have found that the WGD event took place after the split of the Acipenseriformes from the lineage leading to teleosts but before the divergence of Osteoglossiformes, making it specific to the teleostean fish (Figure 1). Because this event separates the species-poor basal lineages from the species-rich teleost lineages, the same authors have suggested that the additional number of genes resulting from this event might have facilitated the evolutionary radiation and the phenotypic diversification of teleosts [15,16].

One basic question regarding gene duplication is the fate of duplicated genes. According to the Duplication – Degeneration – Complementation (DDC) model proposed by Force *et al.* [17], duplicated genes may have three main fates: the majority of duplicated copies are lost, some duplicated genes are subfunctionalized (*i.e.* they share the ancestral function of their non duplicated ances-

tor) and some others undergo neofunctionalization (*i.e.* they change their function when compared to their ancestor). In most cases the sub- or neofunctionalization events are classically considered to have occurred relatively soon after the duplication event. We recently suggested that a biased subset of genes was retained as duplicates after the genome duplication and that gene retention was biased with regard to biological processes [18]. Most notably, we observed an enrichment of fish genomes in new paralogs implicated in development, supporting the link between genome duplication and fish morphological diversity [15,18]. In addition, several studies have shown that sub- or neofunctionalization events can be observed at the expression level when specific pairs of duplicated genes are studied (see for example [19-21]). Of note, if there are some examples of neofunctionalization events affecting the biochemical function of genes after the vertebrate WGD, much less numerous specific examples of that sort were shown for fish duplicated genes [22-24].

Androgens play essential roles in sex differentiation, sex maturation and behavior in vertebrates, including Actinopterygian fishes [25,26], and their actions are mediated through a specific receptor, AR (NR3C4) which belongs to the nuclear receptor super-family [27]. In Actinopterygians, the mediation of androgen action is far more complicated than in other vertebrates as there is a duality in the active androgens involved in reproduction, *i.e.* regular androgens (DHT) versus 11-oxygenated androgens (11KT) [26]. Like every nuclear receptor, ARs are composed of three main domains [27]: a hypervariable N-terminal domain involved in transcriptional activation, a DNA-binding domain (also referred to as the C Domain) which permits the binding of receptor on target genes and a ligand-binding domain (also referred to as the E Domain). The amino acid sequences of these last two domains (DBD and LBD) are highly conserved from actinopterygians to mammals with 90% and 70% identity with mammalian ARs for the DBD and the LBD respectively [28]. In Actinopterygians, several duplicates of AR were characterized in the rainbow trout, *Oncorhynchus mykiss* [29], the mosquitofish, *Gambusia affinis* [30], *Astatotilapia burtoni* (see Table 1 for accession numbers), the Nile tilapia, *Oreochromis niloticus*, the Japanese eel, *Anguilla japonica* [31,32] and in *Gasterosteus aculeatus* [33]. Interestingly, in addition to these molecular clones corresponding to two duplicated genes, different forms of ARs, termed AR-A and AR-B, have also been biochemically characterized based on their binding affinities for different androgen ligands in the Atlantic croaker, *Micropogonias undulatus* [34] and the kelp bass, *Paralabrax clathratus* [35]. These two receptors display different tissue distributions with AR-A present only in the brain and AR-B found both in the brain and the gonads [34]. In addition, the different binding affinities of these two ARs suggest that the receptors



**Figure 1**  
**Classification of the fish species used in this survey as in [51,52].** The Whole Genome Duplication (also named 3R) event is indicated.

mediate the actions of different androgens, 11KT and DHT, in different tissues of teleost fish [36].

The case of AR in fish is particularly interesting to study as sex determination mechanisms are known to be particularly plastic in Actinopterygians [37,38]. For example, sequential hermaphroditism is common among marine fishes, particularly in tropical and subtropical seas, and can involve females becoming males (protogyny) or males becoming females (protandry), and also bidirectional (repetitive) sex changes [39-42]. Sex changes among species with well organized social and mating systems are controlled by social cues [41,43-45] and involve complete alterations in gonadal anatomy and function, as well as changes in color and behavior. It is known that sex steroid hormones play important roles in sex change and behavior in many fish species, and androgens have been shown to be crucial for completion of this process in many protogynous hermaphrodites [46-48].

Interestingly, in human, mutations of the AR gene represent the molecular basis of androgen insensitivity syndrome (AIS) [49]. AIS is characterized by defective virilization in 46, XY individuals. The phenotypic spectrum of AIS is extremely large: Complete AIS (CAIS) is characterized by completely female external genitalia. In Partial AIS (PAIS) the phenotype ranges from almost female external genitalia through ambiguous forms to predominantly male external genitalia with hypospadias. Minimal (or Mild) forms of AIS exist which are characterized by impaired spermatogenesis with or without a slight virilization deficit. In addition, the androgen receptor is

also implicated in prostate cancer and a specific set of mutations often occurred in patients whose cancer became androgen-independent, an evolution of poor clinical prognosis [50].

In this paper, we reconstructed the evolutionary history of AR in Actinopterygians. We observed a complex history shaped by three successive events well separated in time: (i) an ancestral duplication event specific to teleost fishes corresponding to the WGD; (ii) a parallel loss of one duplicated copy (AR-B) in basal Clupeocephala and (iii) a major sequence divergence indicative of a change in functional constraints in the AR-B duplicate of Percomorphs. This evolutionary history together with the striking mutation patterns is indicative of a putative neofunctionalization event that took place late during AR-B evolution. It is tempting to link this neofunctionalization event to the plasticity of sex determination in Percomorphs.

## Results and discussion

### *The Androgen receptor is duplicated in teleost fishes*

Using a combination of RT-PCR with degenerate primers designed in the conserved C and E domains and *in silico* search against various databases, we were able to characterize 26 different new AR cDNA fragments from 20 different fish species. Along with these AR cDNA fragments, we also identified other steroid receptors (NR3C group) in *A. baerii* and *E. stoutii* (Genbank accession numbers [ABF50787](#) and [ABF50785](#)). This is probably due to a combination of the high conservation of the C and E domains used to design primers among all steroid nuclear receptors and the low stringency of touchdown PCR procedure that we used. In this study, five new sequences of AR were also identified using database searches and 21 new sequences of AR were isolated by RT-PCR with degenerated primers, the majority of them being teleosteans (20), one being chondrichthyan, one being Dipnoi and one being chondrosteian (Table 1).

Of note, one expressed sequence tag in *Oryzias latipes* was found to match with the 5' end region of the divergent AR in *Haplochromis burtoni* (AF121257). The corresponding clone was further sequenced and was confirmed to be a medaka AR. No more than one gene was identified in zebrafish by searching both EST databases and the whole genome sequence (see below). Additional file 1 provides an amino acid alignment of a representative choice of these sequences, focusing on complete DBD and LBD sequences. A complete alignment is available upon request to F.B.

Using the 30 full length sequences identified in our screens, we reconstructed the phylogeny of the Actinopterygian ARs using Neighbor-Joining, Maximum Parsimony, Maximum Likelihood and Bayesian methods

**Table 1: Androgen receptors, species names and their accession numbers.**

Genus_AR	Scientific name	Common name	Class_Infraclass	Order	SWP_TrEMBL_ID	Ensembl_Prot_ID	GenBank_Acc_Nb
Homo_AR	Homo sapiens	Human	Mammalia	Primates	ANDR_HUMAN	ENSP00000363822	<a href="#">PI0275</a>
Mus_AR	Mus musculus	Mouse	Mammalia	Rodentia	ANDR_MOUSE	ENSMUSP00000052648	<a href="#">NP_038504</a>
Gallus_AR	Gallus gallus	Chicken	Aves	Galliformes	Q2ACE0_CHICK	ENSGALP00000007301	BAE80463
Xenopus_t_AR	Xenopus tropicalis	European clawed frog	Amphibia	Anura	n.a.	ENSXETP00000011091	n.a.
Xenopus_l_AR	Xenopus laevis	African clawed frog	Amphibia	Anura	P70048_XENLA	_	<a href="#">AAC97386</a>
Protopterus_AR	Protopterus annectens	Protopterus annectens	Sarcopterygii/dipnoi	Lepidosireniformes	A3QQ74_PROAN	_	<a href="#">ABF50783*</a>
Myoxocephalus_AR-B	Myoxocephalus scorpius	Shorthorn sculpin	Actinopterygii/Teleostei	Scorpaeniformes	A3QQ70_MYOSC	_	<a href="#">ABF50779*</a>
Takifugu_AR-A	Takifugu rubripes	Fugu	Actinopterygii/Teleostei	Tetraodontiformes	n.a.	SINFRUP00000156318*	n.a.
Takifugu_AR-B					n.a.	SINFRUP00000135072*	n.a.
Tetraodon_AR-A	Tetraodon nigroviridis	Spotted green pufferfish	Actinopterygii/Teleostei	Tetraodontiformes	Q4S8Q7_TETNG	GSTENP00022234001	<a href="#">CAG02975</a>
Tetraodon_AR-B					Q4RT97_TETNG	GSTENP00029350001	<a href="#">CAG08385</a>
Tetraodon_f_AR-A	Tetraodon fluviatilis	Green pufferfish	Actinopterygii/Teleostei	Tetraodontiformes	A3QQ72_TETFL	_	<a href="#">ABF50781*</a>
Porichthys_AR-A	Porichthys notatus	Plainfin midshipman	Actinopterygii/Teleostei	Batrachoidiformes	Q4F6Z1_PORNO	_	<a href="#">AAZ14095</a>
Oryzias_AR-A	Oryzias latipes	Medaka	Actinopterygii/Teleostei	Beloniformes	Q76LM5_ORYLA	ENSORLP00000011941*	<a href="#">BAC98301</a>
Oryzias_AR-B					A8CMD5_ORYLA	ENSORLP00000010323*	<a href="#">ABV55993</a>
Rivulus_AR-A	Rivulus marmoratus	Mangrove rivulus	Actinopterygii/Teleostei	Cyprinodontiformes	Q15HT9_RIVMA	_	<a href="#">ABC68612</a>
Gambusia_AR-A	Gambusia affinis	Western mosquitofish	Actinopterygii/Teleostei	Cyprinodontiformes	Q5NU07_GAMAF	_	<a href="#">BAD81046</a>
Gambusia_AR-B					Q5NU08_GAMAF	_	<a href="#">BAD81045</a>
Gasterosteus_AR-A	Gasterosteus aculeatus	Three-spined stickleback	Actinopterygii/Teleostei	Gasterosteiformes	Q801Z2_GASAC	ENSGACP000000026869	<a href="#">AAO83572</a>
Gasterosteus_AR-B					n.a.	ENSGACP000000024489*	n.a.
Micropogonias_AR-A	Micropogonias undulatus	Atlantic croaker	Actinopterygii/Teleostei	Perciformes	Q66VR6_MICUN	_	<a href="#">AAU09477</a>
Acanthopagrus_AR-A	Acanthopagrus schlegeli	Black porgy	Actinopterygii/Teleostei	Perciformes	Q800S7_ACASC	_	<a href="#">AAO61694</a>

**Table 1: Androgen receptors, species names and their accession numbers.** (Continued)

Haplochromis_AR-A	Haplochromis burtoni	Burton's mouthbrooder	Actinopterygii/Teleostei	Perciformes	Q8QFV7_HAPBU	_	<a href="#">AAL92878</a>
Haplochromis_AR-B					Q9W6F4_HAPBU	_	<a href="#">AAD25074</a>
Dicentrarchus_AR-A	Dicentrarchus labrax	European sea bass	Actinopterygii/Teleostei	Perciformes	Q4G497_DICLA	_	<a href="#">AAT76433</a>
Dicentrarchus_AR-B					A3QQ67_DICLA	_	<a href="#">ABF50776*</a>
Lepomis_AR-A	Lepomis gibbosus	Lepomis gibbosus	Actinopterygii/Teleostei	Perciformes	A3QQ57_9PERO	_	<a href="#">ABF50766*</a>
Oreochromis_AR-A	Oreochromis niloticus	Nile tilapia	Actinopterygii/Teleostei	Perciformes	Q8UWB7_ORENI	_	<a href="#">BAB20082</a>
Oreochromis_AR-B					Q8UWB8_ORENI	_	<a href="#">BAB20081</a>
Perca_AR-A	Perca fluviatilis	Perch	Actinopterygii/Teleostei	Perciformes	A3QQ55_PERFL	_	<a href="#">ABF50764*</a>
Pagrus_AR-A	Pagrus major	Red sea bream	Actinopterygii/Teleostei	Perciformes	O93497_PAGMA	_	<a href="#">BAA33451</a>
Pomatoschistus_AR-B	Pomatoschistus minutus	Sand goby	Actinopterygii/Teleostei	Perciformes	A3QQ69_POMMI	_	<a href="#">ABF50778*</a>
Halichoeres_AR-A	Halichoeres trimaculatus	Three-spot wrasse	Actinopterygii/Teleostei	Perciformes	Q9DDJ4_HALTR	_	<a href="#">AAG48340</a>
Salmo_AR-A	Salmo salar	Atlantic salmon	Actinopterygii/Teleostei	Salmoniformes	Q8UWF7_SALSA	_	<a href="#">AAL29928</a>
Oncorhynchus_AR-A1	Oncorhynchus mykiss	Rainbow trout	Actinopterygii/Teleostei	Salmoniformes	O93244_ONCMY	_	<a href="#">BAA32784</a>
Oncorhynchus_AR-A2					O93245_ONCMY	_	<a href="#">BAA32785</a>
Cyprinus_AR-A	Cyprinus carpio	Common carp	Actinopterygii/Teleostei	Cypriniformes	A3QQ59_CYPCA	_	<a href="#">ABF50768*</a>
Carassius_AR-A	Carassius auratus	Goldfish	Actinopterygii/Teleostei	Cypriniformes	Q8QFV2_CARAU	_	<a href="#">AAM09278</a>
Ctenopharyngodon_AR-A	Ctenopharyngodon idella	Grass carp	Actinopterygii/Teleostei	Cypriniformes	A3QQ60_CTEID	_	<a href="#">ABF50769*</a>
Labeo_AR-A	Labeo rohita	Indian major carp	Actinopterygii/Teleostei	Cypriniformes	A3QQ58_LABRO	_	<a href="#">ABF50767*</a>
Pimephales_AR-A	Pimephales promelas	Fathead minnow	Actinopterygii/Teleostei	Cypriniformes	Q918F5_9TELE	_	<a href="#">AAF88138</a>
Danio_AR-A	Danio rerio	Zebrafish	Actinopterygii/Teleostei	Cypriniformes	A4GVF3_DANRE	ENSDARP0000016299	<a href="#">ABO21344</a>
Gymnocorymbus_AR-A	Gymnocorymbus ternetzi	Black Widow tetra	Actinopterygii/Teleostei	Characiformes	A3QQ71_9TELE	_	<a href="#">ABF50780*</a>
Clarias_AR-A	Clarias gariepinus	Sharptooth catfish	Actinopterygii/Teleostei	Siluriformes	A3QQ63_CLAGA	_	<a href="#">ABF50772*</a>
Heterotis_AR-A	Heterotis niloticus	Heterotis	Actinopterygii/Teleostei	Osteoglossiformes	A3QQ64_9TELE	_	<a href="#">ABF50773*</a>

**Table 1: Androgen receptors, species names and their accession numbers.** (Continued)

Heterotis_AR-B					A3QQ65_9TELE	–	<a href="#">ABF50774*</a>
Anguilla_AR-A	Anguilla japonica	Japanese eel	Actinopterygii/Teleostei	Anguilliforme	Q9PWG5_ANGJA	–	<a href="#">BAA83805</a>
Anguilla_AR-B					Q9YGV9_ANGJA	–	<a href="#">BAA75464</a>
Acipenser_AR	Acipenser baerii	Siberian sturgeon	Actinopterygii	Acipenseriformes	A3QQ77_ACIBE	–	<a href="#">ABF50786*</a>
Squalus_AR	Squalus acanthias	Spiny dogfish	Chondrichthyes	Squaliformes	Q56VU2_SQUAC	–	<a href="#">AAP55843</a>
Ginglymostoma_AR	Ginglymostoma cirratum	Nurse shark	Chondrichthyes	Orectolobiformes	A3QQ75_GINCI	–	<a href="#">ABF50784*</a>
Leucoraja_AR	Leucoraja erinacea	Little skate	Chondrichthyes	Rajiformes	Q1KXY2_RAJER	–	<a href="#">ABD46746</a>

Names of the species used in this analysis and their accession numbers retrieved from SwissProt, TrEMBL, Ensembl and Genbank. Asterisks show the genes sequenced in this survey. Non available: n.a.

(Figure 2 and see Additional file 2). The shorter PCR fragments were used for sequence signature analysis to assess their orthology relationships. Note that given the overall strong conservation of the AR sequences and the relatively short size of the alignable conserved regions, we obtained many branches supported by relatively weak bootstrap values. The AR gene is thus certainly not an adequate marker to decipher fish phylogeny and the topologies we obtained are often not in accordance with accepted concepts in fish phylogeny [51,52]. Nevertheless, the main branches discussed in this paper are well supported and allow the drawing of clear conclusions.

In all cases, only one AR sequence was identified in each tetrapod. Based on partial sequence analysis, the AR from the west African lungfish branched at the root of this tetrapod cluster (not shown). We obtained only one AR sequence in the sturgeon, *Acipenser baerii* and this sequence is, as expected, clearly located at the base of the Actinopterygian sequences, suggesting that it corresponds to a non-duplicated version of the AR gene (Figure 3). In contrast in *Heterodontis niloticus*, a member of the Osteoglossiformes, which is located at the base of the teleosts according to recent phylogenies based on mitochondrial and nuclear genes [51,53], we found two AR sequences, as in many other teleost fishes such as the eel, the medaka or the cichlids (Table 1 and Figure 2). This suggests that the duplication giving rise to the AR-A and AR-B genes occurred specifically at the base of the teleost tree after the split of the Acipenseriformes from the lineage leading to teleosts but before the divergence of Osteoglossiformes (see Figure 1). This phylogenetic dating obtained with AR is in accordance with the data recently obtained with other genes such as Sox11 and tyrosinase by Hoegg *et al.*, [15]. In contrast to the data from these authors, our data set does not contain sequences from Semionotiformes (gars) or Amiiiformes (the bowfins, *Amia calva*) that would allow us to confirm more accurately that the WGD occurred specifically in teleosts. Nevertheless, our data are fully consistent with this likely scenario.

That the two sequences AR-A and AR-B are indeed the product of the ancestral WGD specific of teleosts is further supported by the chromosomal location of the two tetraodon AR sequences. Indeed, these two genes are located on chromosomes T1 and T7 in the tetraodon genome (Figure 4). From a global synteny analysis, we have previously shown that these two chromosomal regions have a common origin [1,18]. This is further supported by the observation that in the medaka genome, the two AR genes are located on chromosomes M10 and M14 that also share many duplicated genes ([16] and data not shown). Taken together, these data unambiguously show that the AR

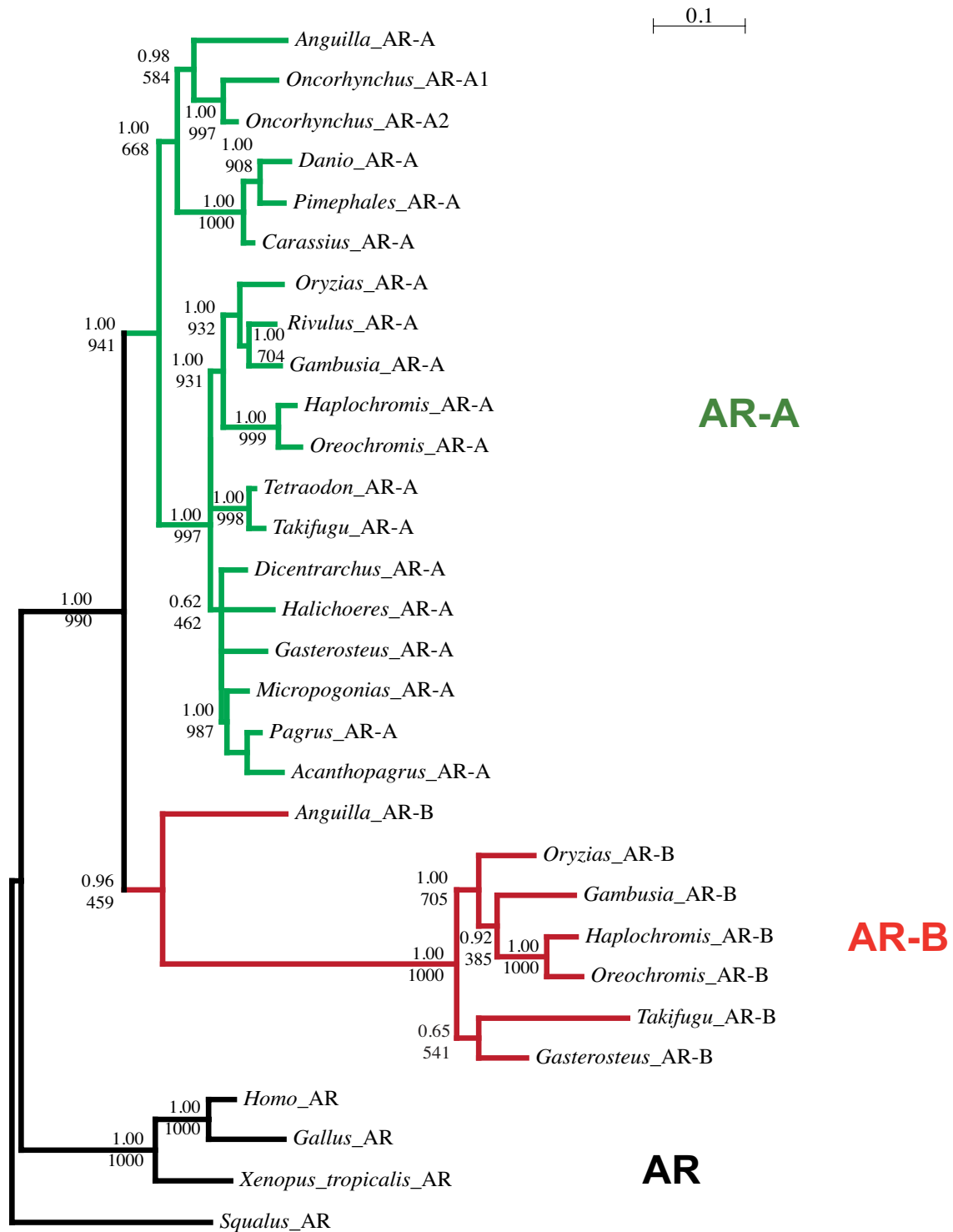
gene was duplicated to give rise to two paralogs AR-A and AR-B during the teleost specific WGD.

#### **AR-B was secondarily lost in basal Clupeocephala, including zebrafish**

When analyzing the tree presented in Figure 1, we were puzzled to observe that the AR-B sequence can be found in basal teleosts (*Heterotis*, *Anguilla*) as well as in many Percomorphs but is missing in many basal Clupeocephala lineages.

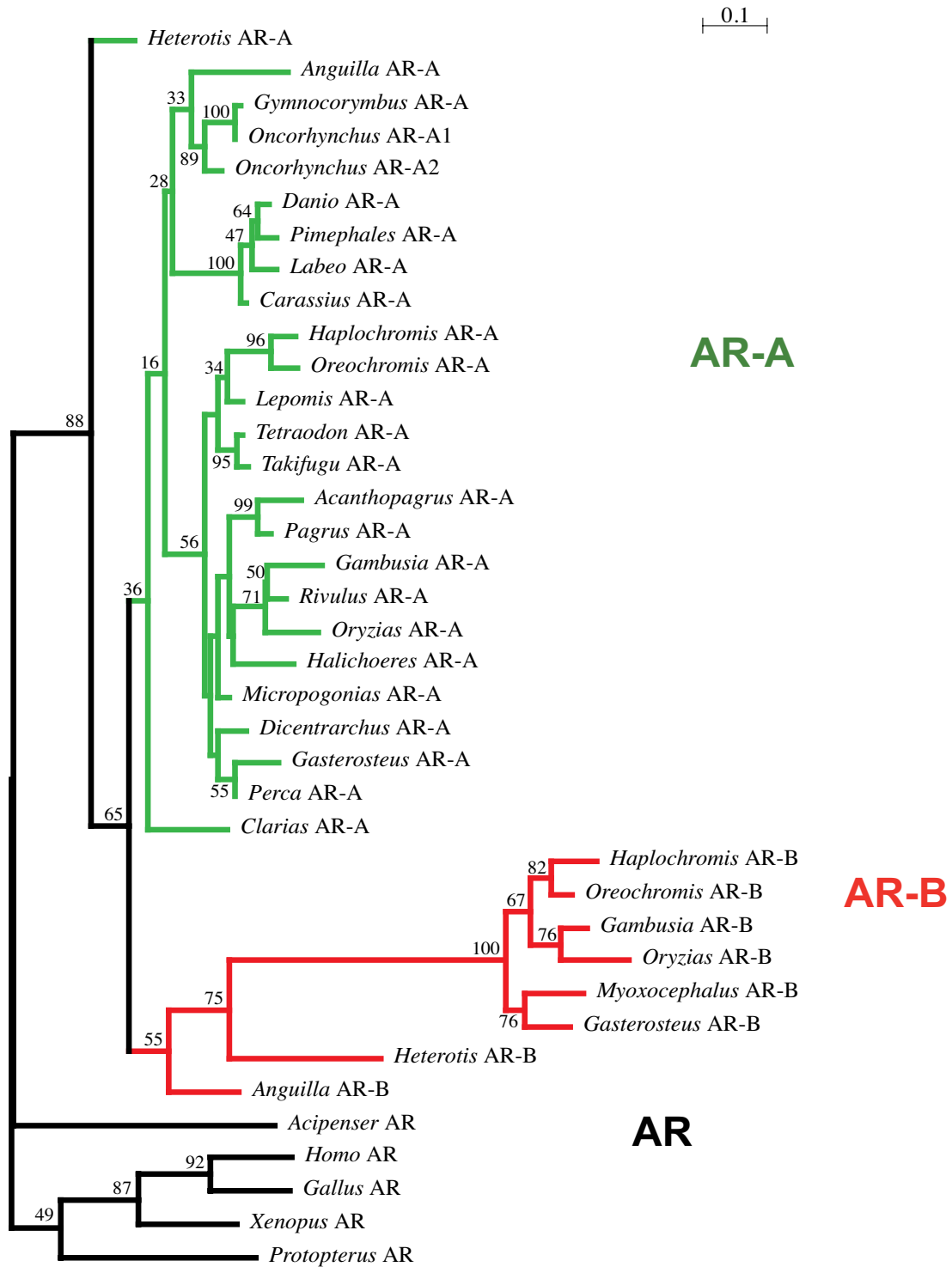
As this observation could be due to an experimental bias linked to a failure to amplify a divergent gene by PCR, we then first checked that this observation was not due to an artifactual lack of detection of the AR-B gene. To this end, we focused our analysis on zebrafish for which a large number of data (complete genome, ESTs, *etc.*) are available. First, we carried out RT-PCR experiments using different batches of primers and several RNA extracts of zebrafish embryos at various developmental stages, as well as adult organs. In all cases, we detected only one AR sequence whereas our primer batches were able to detect divergent NR3C steroid receptors such as GR, MR or PR. PCR experiments based on DNA amplification of short fragments contained in only one exon also failed. Finally, we intensively screened the release Zv7 (13 July 2007) of the zebrafish genome using various fragments of the AR gene as baits without any significant hit. Of note, no sequence reminiscent of a pseudogene was detected.

Due to the availability of a complete and assembled zebrafish genome sequence, we tried to better understand the fate of the AR-B gene in zebrafish. In *Tetraodon*, we found two AR genes, AR-A and AR-B (Table 1 and Figure 2). Since we and others previously showed that an extensive synteny persists between Tetraodontiformes and zebrafish genomes [1,7,54], we precisely mapped in *Tetraodon* and zebrafish the syntenic regions containing AR-A and AR-B. Figure 4 clearly shows that AR genes map in a large duplicated region corresponding to chromosomes T1 and T7 in *Tetraodon*. Chromosome T7 in *Tetraodon* is syntenic to chromosomes Z5, Z10 and Z21 in zebrafish. Interestingly, the zebrafish AR-A ortholog is present in chromosome Z5, as predicted based on conserved synteny. A detailed map shows that the organization of this region is conserved between tetraodon and zebrafish (data not shown). The *Tetraodon* AR-B gene map to chromosome T1 and the region encompassing the gene corresponds mainly to the zebrafish chromosome Z14. The mapping of the region containing the *Tetraodon* AR-B sequence on the zebrafish genome shows that this region has been scrambled during evolution. Many gene orders are not conserved and large fragments are missing or were exchanged (data not shown). The same consideration is

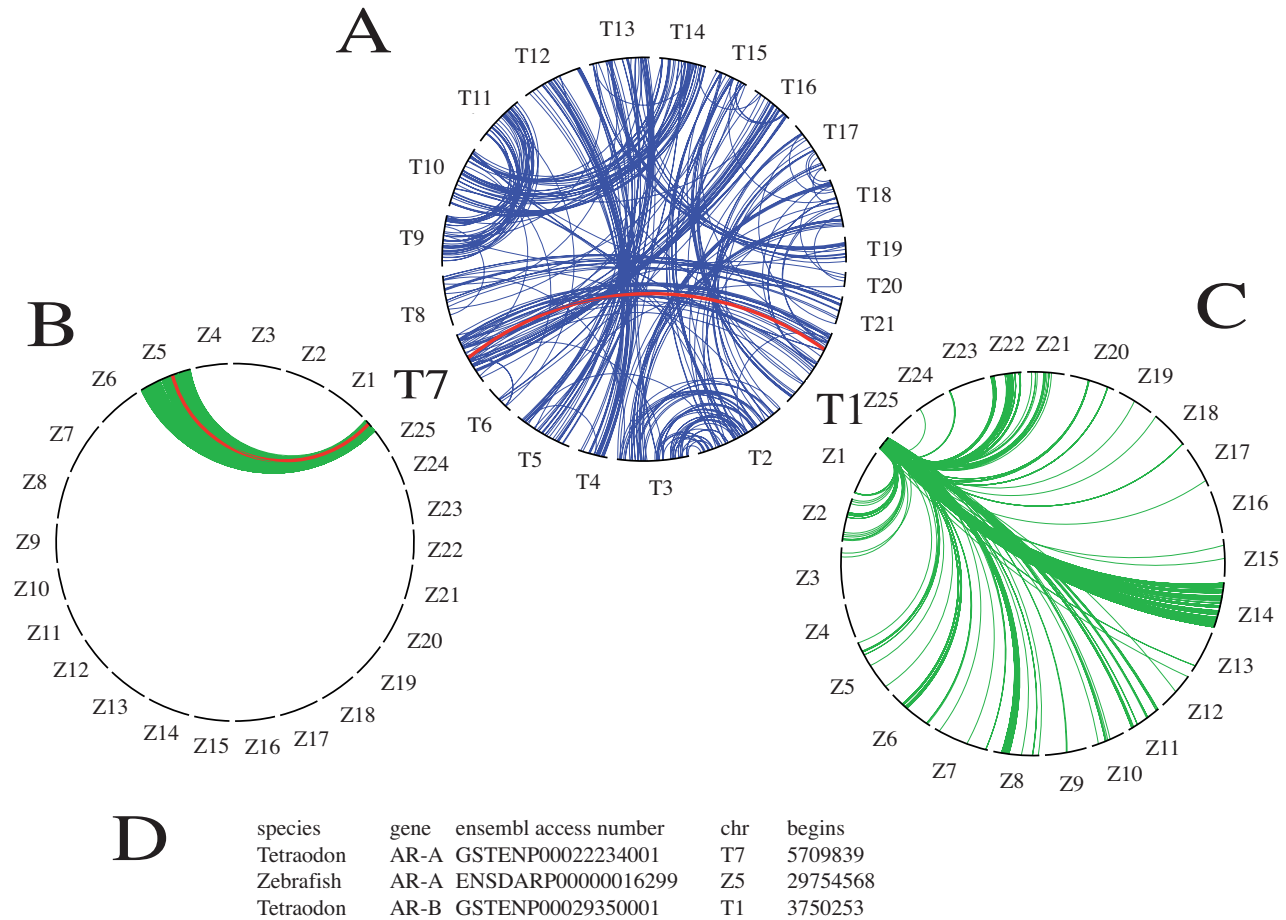


**Figure 2**  
**Phylogenetic analysis of the AR in fish.** Top values are posterior probabilities given with MrBayes. Bottom values are maximum likelihood for 1000 iterations of bootstrap values. Default parameters of PhyML were used. AR-A branches are in green, AR-B branches are in red.





**Figure 3**  
**Maximum likelihood analysis of all species using PhyML.** Default parameters of PhyML were used. 100 bootstrap replicates were used and only values above 30 are indicated. Since short sequences are included here, the bootstrap support is lower than on the tree based on complete sequences as in Figure 2. AR-A branches are in green, AR-B branches are in red.



**Figure 4**

**Syntenic analysis of the AR in fish.** We used the rose window method as described in Jaillon *et al.* [1] and Brunet *et al.* [18] (see also Additional file 3). (A) Rose window showing the 21 Tetraodon chromosomes, illustrating the two-by-two relation between duplicates originating from the teleost specific WGD event. AR-A and AR-B in Tetraodon are located respectively on chromosome T1 and T7, chromosomes that have been described as originating from the teleosts WGD. The red line shows the relationship between these two genes. (B and C) relationship between the genes present in Tetraodon chromosomes T7 (B) and T1 (C) with the genes present in the 25 zebrafish chromosomes. As shown in panel B, the AR A gene is found in the zebrafish on chromosome Z5 (red line) and a very strong synteny exists in general between T7 and Z5 chromosomes. As shown in panel C, a clear synteny is found between T1 and the Z14 chromosomes, although less univocal than in the previous case (orthology based from Ensembl v48, in agreement with reciprocal best-hit analyses we performed, data not shown). Importantly, the AR-B ortholog in the zebrafish could not be detected neither on chromosome Z14 nor on other locations in the whole genome. (D) Table indicating the positions and Ensembl accession numbers of the relevant AR genes.

reached when the medaka genome is considered (see Additional file 3).

These data indicate clearly that a secondary loss of AR-B occurred in zebrafish. Interestingly, in related Cypriniformes (5 species), Characiformes (1 species) and Siluriformes (1 species) that altogether form the well-supported clade Otophysi [55], we also found only one AR-A sequence and no AR-B one. We recently screened EST data available for all these species and we could not

find any sequence reminiscent of AR-B. Of course, although complete genome sequences are not available and RT-PCR results can artifactually miss a divergent sequence, these data collectively suggest that a unique event of loss of AR-B occurred early on in the Otophysi lineage.

Strikingly, the exact same situation was observed in another independent lineage of Clupeomorph: the Salmoniformes. In the 2 species analyzed one more time, we

found only AR-A and not AR-B. Of note, in Salmoniformes two AR-A sequences are observed (corresponding to the two subtypes named *ar-alpha* and *ar-beta* known in rainbow trout [29]). These correspond to the tetraploidization event that occurred 25 to 100 MYr ago specifically in the salmonid lineage [56]). It is important to note that even in the divergent AR-B sequences that we analyzed in Percomorphs (see below), the regions targeted by the various PCR primers that we used are well-conserved. In addition, once again, the screening of the ESTs available in salmon and trout (the most widely used species of Salmoniformes in aquaculture and genomic research) has not delivered any AR-B type sequence. Thus, although this conclusion is only tentative in the absence of a complete genome, the most likely scenario, considering all existing data, is that AR-B gene was lost in Salmoniformes as in Otophysi (Figure 5).

It is clear that loss of duplicated genes is a very common fate after a genome duplication event but the present analysis nicely illustrates a late case of neofunctionalization. Our data suggest that basal teleosts and percomorphs kept two functional copies of AR whereas "intermediate" lineages such as Otophysi and Salmoniformes lost it secondarily (see below). According to the topology of teleost fish phylogeny presented in Figure 1, our results imply two independent losses of AR-B, one at the base of Otophysi and one at the base of Salmoniformes (see also Figure 5). This is based on the assumption that the current topology based essentially on complete mitochondrial DNA analysis is correct in the respective placement of Salmoniformes and Otophysi [51,55,57]. If, as suggested by some authors, these two groups form a monophyletic clade, it may be possible that in fact only one ancestral event of loss occurred [58]. In that case, we can predict that AR-B should not be found in Esociformes. In any case, our present data plead for the search of AR-A and AR-B in orders of Actinopterygians located at key positions in the evolutionary tree: it would be interesting for example to see if AR-B is present in other Ostariophysi lineages such as Gonorhynchiformes, or Clupeomorphs [55] as well as other Protacanthopterygii such as Esociformes, Argentinioidea and Osmeroidea. This will allow a more precise determination of when the events of loss occurred [51].

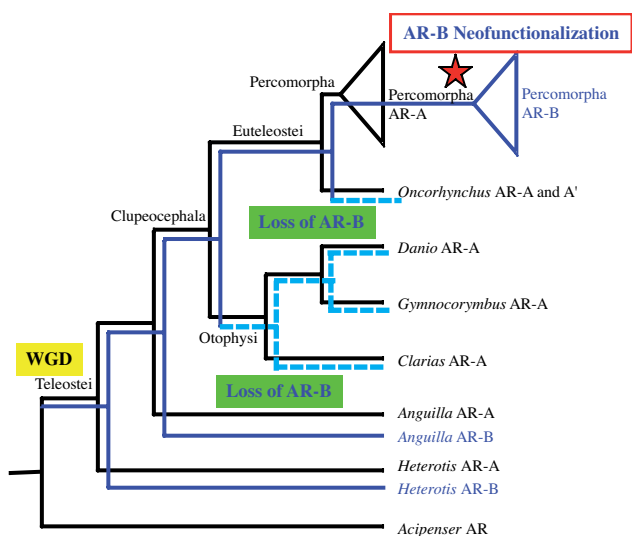
It is difficult to speculate with the data available why the Otophysi and the salmonids apparently do not need a second AR-B gene. Given the major function of AR in sex determination and sex organ differentiation, it is tempting to link these events with these processes but given that these data on Otophysi and salmonids are limited to some specific models such as zebrafish, salmon and trout, it is up to now difficult to find an obvious connection. It is striking that zebrafish and salmonids are extremely different regarding sex determination and sex differentiation.

In addition, as discussed above, the exact phylogenetic range of this loss of AR-B is still unclear.

#### Functional shift of AR-B in Euteleosts

In the tree presented in Figure 2, we noticed the presence of a highly divergent terminal group of AR-B sequences. This is confirmed when a larger dataset including partial sequences is used to construct a phylogeny with any of the 4 methods used (Figure 3 and data not shown). In all cases, we found a long terminal branch uniting divergent AR-B sequences. This divergent AR-B subtype unambiguously (bootstrap value: 1000 out of 1000; posterior probabilities: 1.00) clusters AR sequences of fish belonging to the percomorphs, *i.e.*, the seabass *D. labrax*, the sand goby, *P. minutus*, the Nile tilapia, *O. niloticus* and *A. burtoni*, a scorpaeniforme with *M. scorpius*, the shorthorn sculpin, a beloniforme with the medaka, *O. latipes*, a cyprinodontiforme with the mosquitofish, *G. affinis*, and 2 tetraodontiformes with the tetraodon, *T. nigroviridis* and the fugu, *T. rubripes*. Indeed, when we considered the sequence alignment (see Additional file 1), we observed a series of mutations that are present only in the percomorph AR-B sequences (highlighted in green). The divergence of these sequences corresponds to a transient episode of sequence divergence as the AR-B sequences clustered inside this group are not particularly variable. Thus, all these data suggest that the percomorphs AR-B are connected to the basal teleosts AR-B through a long branch and exhibit some striking sequence divergence at key positions. From the phylogenetic range of species in which these divergent AR-B sequences are found, it is likely that this acceleration occurred specifically in percomorphs, although this remains to be fully established by a broader taxonomic sampling including other Neoteleostei lineages such as basal Acanthomorphs (e.g. Gadiformes; [59]) as well as Beryciformes [52]. To really assess if this event is found in all Percomorphs, some basal lineages (e.g. Ophidiiformes) of this extremely vast group of fishes should also be studied [52]. In the mean time, given our observation that divergent AR-Bs are found only in percomorphs from our dataset, we will refer to these divergent sequences as "percomorph AR-B".

It is important to insist on the fact that basal teleosts (*Anguilla* and *Heterotis*) clearly contained AR-A and AR-B paralogs. For AR-A, this is not difficult to establish given that this gene is present in a wide phylogenetic range of species. For AR-B, the assignment is less obvious since, as discussed above, this gene has been lost in basal Clupeocephala. The fact that the AR-Bs from *Anguilla* and *Heterotis* are indeed orthologs of the percomorph AR-B is indicated by several features: (i) these sequences exhibit a few key sequence signatures that represent synapomorphies of AR-B sequences (highlighted in yellow and orange in Additional file 1), this is for example the case of



**Figure 5**  
**Model of evolution of the AR in fish based on the same phylogenetic background as Figure 1.** A WGD occurred at the base of the Teleostei, and two successive losses occurred in the lineages leading to the Otophysi and the *Oncorhynchus*. Last, in the percomorphs, all the AR-B genes harbor an event of acceleration, while the AR-A ones show no sign of such rapid evolution at the molecular level.

Gly633, Ser861 and Ser928 in the LBD; (ii) the topology of the phylogenetic tree supports this assumption albeit with a moderate support (posterior probability of 0.96, bootstrap value of 459%; Figure 2). Of note, we constructed trees based on Bayesian analysis which confirm that the topology presented in Figures 2 and 3 is robust (see Additional file 2).

The most likely scenario accounting for the data available concerning Actinopterygian AR evolution is therefore a three step model (Figure 5): (i) ancestral duplication of a unique AR gene during the WGD event specific of teleost fishes. This explains why *Anguilla* and *Heterotis* have two AR sequences, AR-A and AR-B; (ii) secondary loss of AR-B in basal Clupeocephala (Otophysi and Salmoniformes) explaining the restricted phylogenetic occurrence of AR-B when compared to AR-A; (iii) a late specific divergence of AR-B. The long branch connecting percomorph AR-B to the basal AR-B sequences is indicative of the accumulation of numerous mutations and we thus proposed that it corresponds to a functional shift that has affected the AR-B protein.

We therefore wanted to test whether the two groups of paralogous genes AR-A and percomorph AR-B were under different selective pressures. We reasoned that if selective pressures differed between the two groups, there should be sites undergoing substitutions in the AR-A subtree and

constrained in the percomorph AR-B subtree, and symmetrically sites constrained in the AR-A subtree undergoing substitutions in the AR-B subtree. Patterns of evolutionary rates in one subtree versus the other were compared to answer this question: are they significantly more different than they would be if branches were picked at random among the two subtrees? Expected numbers of substitutions were estimated for all branches of the tree and all sites of the alignment [60]. A non-symmetric correspondence analysis was applied on these numbers of substitutions, and the percentage of variance between branches explained when branches are clustered according to the two subtrees was computed. The significance of this percentage was assessed by a permutation test based on 500 000 replicates, where branches are picked randomly from the two subtrees. Among the 500 000 random clusterings, only 0.36% explained a higher percentage of variance among branches than the clustering according to the paralogous subtrees (See Additional file 4). As branches have been normalized with respect to their lengths by the correspondence analysis, this variance comes from differences in patterns of substitutions, not branch lengths. Therefore, patterns of substitutions are significantly more different between the two groups of paralogous genes than between two random groups of branches. This suggests that selective pressures differ between AR-A and AR-B genes, which is in favor of a possible neofunctionalization. The same conclusion is reached with the use of the PAML software [76]. A significant change in the selective pressure onto the branch specific to the AR-B in percomorphs (p-value = 1.688343e-08) is unequivocally detected, although the test is not sensitive enough to tell whether it is a relaxation of the selective pressure or positive selection that drove this change. It should be made clear that, in the absence of a functional characterization, including a comparison of a basal non-duplicated AR (e.g. sturgeon), duplicated AR with a non-divergent AR-B (e.g. eel) and duplicated AR with a divergent AR-B (e.g. medaka) this neofunctionalization cannot yet be formally proved and should be regarded as only putative.

This pattern of a late spectacular divergence of a duplicated gene in a precise taxonomic group is an interesting case in which the duplication and the functional shifts are clearly two recognizable events that were decoupled in time. The AR-B gene will thus be a very interesting model to study the precise functional and biological impact of these two events since we have sequences of non-duplicated fish AR (sturgeon), duplicated AR-A (in eel and medaka for example), duplicated and non divergent AR-B (eel) and duplicated and divergent AR-B (medaka). In addition, we have other interesting cases for comparison such as a unique zebrafish AR-A gene with secondary loss of AR-B. The fact that AR is a gene encoding a nuclear hormone receptor with a known ligand, a clear biological role

and for which several functional tests are available renders this gene particularly suitable for a precise integrated study of the consequences and respective roles of duplication and evolutionary sequence divergence. For example, it may be very interesting to study if, as proposed recently at a broader scale for nuclear receptors, sequence divergence is correlated to expression divergence [19].

#### **Analysis of the substitution pattern in relation to human Androgen Insensitivity Syndrome**

As a first step to analyze the possible consequences of the duplication and divergence of AR-B in Percomorphs, we scrutinized the mutations found in divergent AR-B *versus* AR-A. We first detected how the various mutations observed in Actinopterygian ARs are located in function of the complex evolutionary history described above. We thus categorized the mutations in four classes (Figure 6 and see Additional file 5): (i) mutations found only in AR-A and not in AR-B or ARs from Amniotes (yellow in Figure 6); (ii) mutations observed specifically in divergent AR-B (red star clade in Figure 5; shown in blue in Figure 6); (iii) positions found to be identical in AR-A and AR-B but different in Amniotes AR (red in Figure 6); and finally (iv) positions that are different in Amniotes AR-A and AR-B (green in Figure 6). From this analysis, it is obvious that the divergent AR-B effectively accumulated mutations. Out of 58 amino acid mutations, we found that AR-As exhibit 14 (= 5 + 9; see Figure 6) specific mutations whereas divergent AR-B have 38 (= 29 + 9) specific mutations. This difference is highly significant (Chi-square test  $\geq 15.34^{***}$ ).

We thus studied in more detail the 38 mutations found in the divergent AR-B sequences in order to see if some of them could have obvious functional consequences. Of note, and not surprisingly, none of these mutations affect the positions known to directly interact with the ligand or implicated in coactivator binding as determined in the 3D structure of the AR LBD complexed with various ligands [61-63]. Few specific changes are observed such as L744V, M749L, Q783H and M895I (numbering according to the consensual human AR mutations database). Some substitutions are also observed in the AF-2 region: A898 is substituted to a S in most AR-Bs and to a G in most AR-As, as well as I899 is substituted to a V in most AR-Bs. Of note, they are observed in both AR-As and divergent AR-Bs and overall, they are unlikely to account for significant functional consequences.

Then, we scrutinized the positions in the DBD and LBD divergent in AR-B and we checked whether these mutations are affecting amino acids found mutated in human pathologic conditions (Figure 6 and see Additional file 6). We were particularly interested by mutations occurring in Androgen Insensitivity Syndrome (AIS) or prostate cancer since these pathologies affect the ability of the receptor to

regulate transcription of target genes in response to ligand binding. AIS is a pathologic condition in humans defined by the eventual occurrence of female differentiation despite the male XY genome and results from germinal mutations in the human AR gene. As discussed above, AIS can be complete (CAIS), partial (PAIS) or mild (MAIS) [49]. We noted that effectively, some divergent AR-B specific mutations are localized in close proximity to functionally relevant residues and may thus impact, in a subtle manner, the function of the receptor. For example the substitution Y739L is observed in all divergent AR-Bs (see Additional file 1). Located close to M742, another amino acid involved in ligand binding, Y739 could influence ligand binding by itself since its substitution to aspartic acid has been described in a CAIS patient [64]. The same question addresses several other amino acids specifically different in divergent AR-Bs as compared to AR-As, such as the F856L substitution which has been observed in patients with CAIS [65] (see Additional file 6).

We thus compared the mutation pattern observed in the divergent percomorphs AR-B gene (that is 38 mutations found in the LBD as well as 22 mutations in the DBD) which are linked to a neofunctionalization event to the mutation pattern occurring in the pathological conditions (Figure 6 and Table 2; see also Additional file 7). When these various types of mutations are mapped on the structure of the human receptor (see Figure 6 for the LBD and Additional file 5 for the DBD), we found that most of the mutations found in AR-B are distinct from pathological mutations. Nevertheless, as discussed above, we also observed some positions mutated both in AR-B and in human patients. Globally, 6 of the 38 LBD mutations found in the divergent AR-B involve amino acids known to be implicated in PAIS + MAIS, whereas only 3 are found in CAIS (see Androgen Receptor Gene Database [75] and references herein; see Additional file 6). Although this difference is not statistically significant, this is in accordance with the notion that AR-B is a functional gene the function of which is only modified (and not drastically reduced) by the specific mutations arising during the putative neofunctionalization event. The fact that we found 7 mutations common between fish AR-B and prostate cancer mutants is more difficult to interpret since mutated ARs in prostate cancer are functionally diverse (from totally inactive receptors to receptors activated by antagonists).

#### **Conclusion**

The above analyses suggest that these sequence differences between AR-A and AR-B will affect the functionality of these receptors and are linked to a putative neofunctionalization event. This remains of course to be directly addressed through *in vitro* and *in vivo* analysis of the role of AR-A and AR-B in suitable fish models. How this functionality is precisely affected remains therefore an open question. The teleost duality in terms of active androgens

involved in reproduction [26] is of great interest in that context. Of note, AR-A and AR-B have been shown to both bind regular androgens and the fish specific 11-oxygenated androgens (11KT) although no direct comparison between AR-A and AR-B has been carried out in the same species until now. We thus have no clear comparison of the respective affinities and potencies of AR-A and AR-B for DHT and 11KT. Of special interest are the *in vivo* binding studies carried out in two perciform species, *i.e.* the Atlantic croaker, *Micropogonias undulates* and the kelp bass, *Paralabrax clathratus*, that demonstrated the existence of two different nuclear androgen receptors that may mediate the physiological actions of different androgens [34-36]. The mutation pattern we observed in AR-B is indicative of a neofunctionalization event at the functional level, but it is likely that this may also be coupled to differences at the expression level. Indeed, neofunctionalization of the expression pattern has been suggested in the cichlid fish *A. burtoni* in which it has been shown that AR-A and AR-B have distinct expression patterns in the brain [66], with a differential implication of these receptors in the maintenance of social dominance status of male fish [67]. Taking into account the functional shift that we specifically observed in the Percomorph lineage, it is tempting to link this functional shift and the sexual lability that is observed in this lineage as the Percomorphs contain nearly 90% of all the hermaphrodite species known to date [37]. One may thus hypothesize that the existence of two functionally divergent AR genes play a role in the plasticity of sex determination often observed in these fishes. In this sense, the presence of the divergent AR-B gene could be viewed as a permissive factor allowing the evolvability of divergent sex determination in these fishes.

## Methods

### Fish and RNA extraction

Common and scientific names of all fish species used in this study are given in Table 1. Tissues samples (whole fish, ovaries, testis and brain) were obtained from fish specimens collected in the wild (*Perca fluviatilis*, *Lepomis gibbosus*, *Pomatoschistus minutus*, *Myoxocephalus scorpius*, *Protopterus annectens*, *Labeo rohita* and *Eptatretus stoutii*), bred in captivity in experimental or aquaculture facilities (*Dicentrarchus labrax*, *Ctenopharyngodon idella*, *Gambusia fluvalis*, *Tetraodon fluvalis*, *Danio rerio*, *Carassius auratus*, *Clarias gariepinus*, and *Heterotis niloticus*) or from local aquarium fish retailers (*Gymnocorymbus ternetzi* and *Acipenser baerii*). For *Ginglymostoma cirratum*, we used the Epigonal Nurse Shark cDNA Library provided by Dr. M.F. Flajnik at the University of Miami School of Medicine [68]. All animals were anesthetized with 2-phenoxyethanol and then sacrificed by decapitation before the sampling of tissues. Total RNA was prepared after homogenization in TRIZOL® reagent (Invitrogen, Cergy Pontoise, France) following the manufacturer's instruc-

tions and mRNA was further purified from 500 µg of total RNA using the OligoTex mRNA kit (Qiagen).

### Reverse transcription

For cDNA synthesis, 2 µg of mRNA were denatured in the presence of oligo (dT) (0.5 µg) for 5 min at 70°C, and then chilled on ice. Reverse transcription (RT) was performed at 37°C for 1 hour using M-MLV reverse transcriptase (Promega, Madison, WI) as described by the manufacturer. Namely, 2 µg of mRNA were reverse transcribed with 200 units of M-MLV reverse transcriptase in the presence of 1.25 µl of each dNTP at 10 mM, 5 µl of M-MLV 5× reaction buffer and 25 units of RNasin® (Promega, Madison, WI, USA), in a total volume of 25 µl.

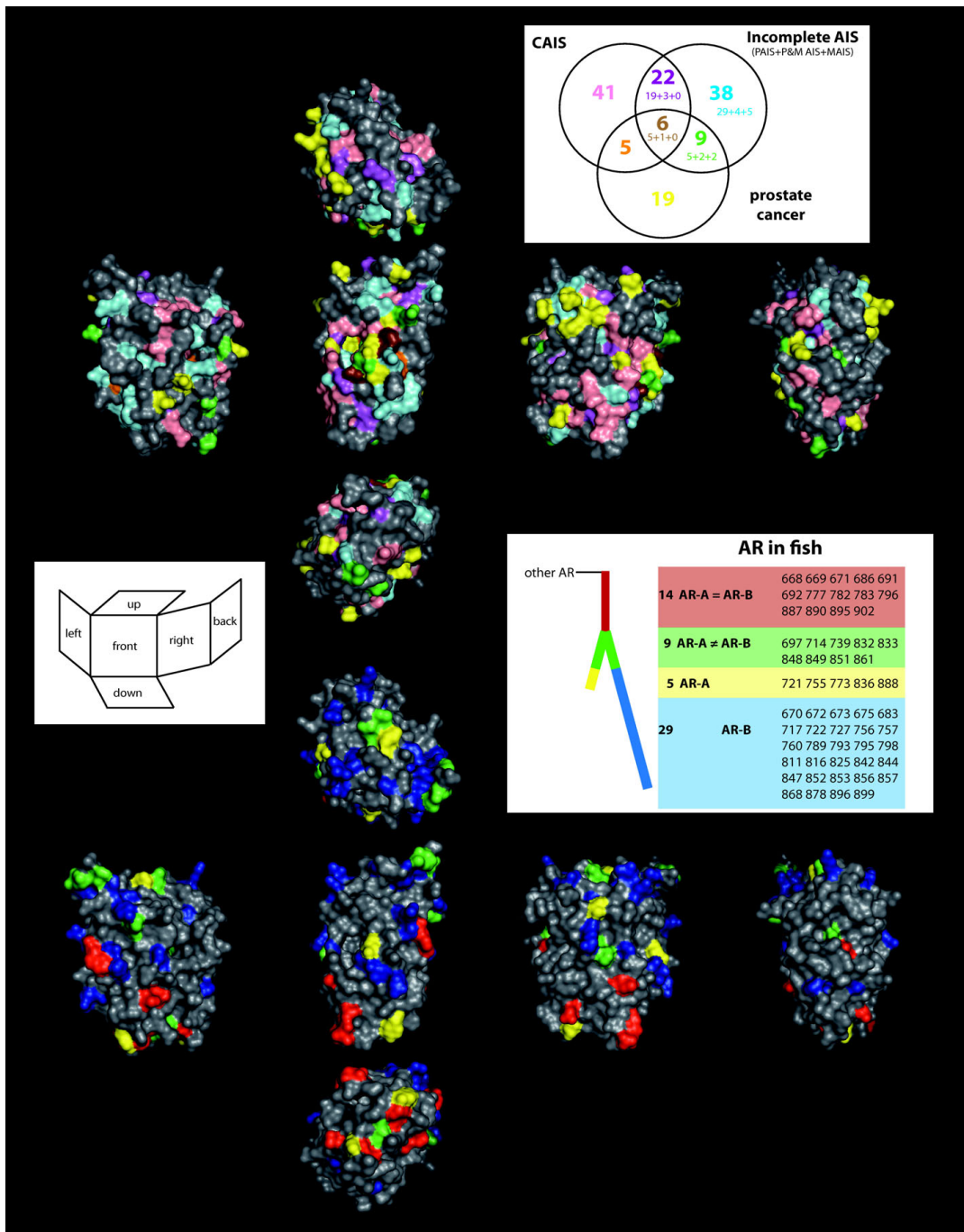
### Primer design

Degenerate oligonucleotide primers were designed after alignment of various fish, reptile, bird and mammalian AR amino acids sequences. Two degenerate primers, AR.AS: TGY TAY GAR GCI GGI ATG AC [CYEAGM] and AR.AAS: AAI ACC ATI ACI BYC CTC CA [WMGVMF], were selected respectively in the highly conserved DBD (C domain) and steroid LBD (E domain) regions. The alignment of the divergent *H. burtoni* AR-B (see Table 1) sequence was used to design another specific degenerate primer, AR.BS: TGC TTY ATG KCG GGN ATG [CFMSGM] in the same region as AR.AS. AR.AS and AR.BS were both used in conjunction with AR.AAS [AR.AS × AR.AAS and AR.BS × AR.AAS]. A second set of degenerate primers outside the PCR fragments produced by AR.AS × AR.AAS or AR.BS × AR.AAS was designed: AR3S: GTI TTY TTY AAR AGR GCI GC [VFFKRAA] and AR3AS: CCA ICC CAT IGC RAA IAA IAC CAT [MVFAMGW] and were used in a nested PCR strategy.

### Touchdown RT-PCR and cloning of AR sequences

The RT-PCR strategy was used according to Escriva *et al.* [69]. PCR reactions were set up using 2 µl of cDNA or 2 µl of Epigonal Nurse Shark cDNA library and 0.5 units of *Taq* Polymerase (Sigma), 200 µM of dNTPs, 30 µmol of each degenerated primer and 2.5 µl of *Taq* buffer 10× (Sigma). The total volume of the reaction was 25 µl and the cycling Touchdown PCR conditions were: 94°C for 1 min, 20 cycles of regularly decreasing annealing temperature from 50°C to 40°C for 30 sec and 72°C for 30 sec, and 30 cycles at the annealing temperature of 40°C, ending at 72°C for 30 sec. PCR products were analyzed on agarose gel (1%) and amplified DNA fragments of the anticipated length (average between 370 and 450 bp) were subsequently subcloned into pCR 2.1 plasmid. Bacteria (INVαF', *E. coli* TOP10 cells, Invitrogen) were transformed by electroporation, spread on LB-ampicillin agar plates and incubated overnight at 37°C. From 10 to 20 randomly selected recombinant colonies were then screened either using PCR with primers amplifying the inserts (T7 and M13 reverse primers) or by hybridization of nitrocel-





**Figure 6**  
**Location of mutations in the 3D structure of the androgen receptors.** Cubic view of localization of the substitutions onto the surface representation of the LBD of the human AR (1XNN.pdb). (Top): number of substitutions characterizing CAIS, PAIS + MAIS, and the prostate cancer. The mutations are those listed in Additional file 6. Amino acid changes (regardless of the kind of substitutions) can lead to different phenotypes: e.g. 5 is the number of aa found in CAIS in a patient, and prostate cancer in another; 9 is the sum of 5 aa reported as found in both PAIS and prostate cancer, etc... See details in Additional file 7. Color codes are reported in the 3D structure. (Bottom): Localization of the substitutions characterizing AR-A and AR-B in fish compared to all other AR. There are 14 identical substitutions in the 2 ARs (that are found in fish AR but not in other vertebrates). We found respectively 14 and 38 substitutions in AR-A and AR-B, 9 of them being common with different substitutions in these two ARs (in green). Color codes are reported onto the 3D representation.

**Table 2: Common substitutions in fish AR and AR diseases in human.**

	CAIS		PAIS+MAIS		prostate cancer			
	DBD	LBD	DBD	LBD	DBD	LBD		
Length/nb	86	257	14	74	17	75	7	39
AR-A all	2	14	0	1	0	1	0	2
AR-B all	22	38	0	3	0	6	1	7
AR-A = B	2	14	0	1	0	3	0	4

Number of substitutions common to the various human phenotypes of AR diseases and AR-A and AR-B in fish (see Additional file 6 for details)

lulose membrane lifts with a rainbow trout AR radiolabelled (dCTP<sup>32</sup>) probe. Positive clones were sequenced using a dideoxy cycle-sequencing method with the Dye Terminator Cycle Sequencing Kit (Applied Biosystems) and reaction sequences were read on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). A secondary nested PCR was carried out for RNA samples from *G. ternaezi*, *D. rerio*, *P. annectens*, and *A. baerii* species (see Table 1) using as template a first PCR reaction, obtained using the primers [AR3S × AR3AS], at a 1/100 dilution and a second set of nested degenerate primers [AR.AS × AR.AAS] or [AR.Bs × AR.AAS]. PCR conditions, subsequent cloning, clone selection, and sequencing were as described above.

#### Searching AR in sequence databases

Homologous DNA and protein fish ARs were searched on available public databases (non redundant, Expressed Sequence Tags) using the various BLAST programs available through the network servers at the National Center of Biotechnology Information <http://www.ncbi.nlm.nih.gov/BLAST/>. We also retrieved AR sequences from the whole genome databases at the Ensembl Genome browser <http://www.ensembl.org/index.html>. From Ensembl v48 (Aug. 2007), we retrieved AR sequences belonging to the Ensembl family ENSF00000000291.

#### Sequence and structural analysis

Multiple alignments of the deduced amino acid sequences were generated with Muscle using the default parameters [70]. Phylogenetic trees were realized by multiple alignments of deduced amino acid sequences using the neighbor-joining and parcimony methods implemented in PhyloWin [71]. PhyML [72] was used to generate maximum likelihood phylogenetic trees. Bayesian trees were generated using MrBayes v3 <http://mrbayes.csit.fsu.edu/index.php>.

To test whether the two groups of paralogous genes AR-A and AR-B were under different selective pressures, patterns

of substitutions estimated in the AR-A subtree *versus* the AR-B one in percomorphs were compared. Expected numbers of substitutions per site and per branch were estimated with the CoMap program [60] based on the Bio++ library [73]. This produced a matrix containing branches of the tree as rows, and sites of the alignment as columns. Branches belonging to the AR-A and AR-B subtrees were selected, discarding the two eel sequences, so that the number of branches was the same in the two subtrees. A non-symmetric correspondence analysis was applied on the resulting submatrix, and the percentage of variance between branches explained when branches are clustered according to the two subtrees computed. The significance of this percentage was assessed by a permutation test based on 500,000 replicates, where branches were clustered randomly. All these analyses were conducted with the ade4 package [74] in the R environment (R development core team).

In order to determine any change in the selective pressure along the branch leading to the percomorphs AR-Bs beside of this previous test, we also used PAML version 4 <http://abacus.gene.ucl.ac.uk/software/paml.html> [76].

Location of all the substitutions found in the LBD of the human AR was retrieved from Bruce Gottlieb's database at his web site <http://androgendb.mcgill.ca/> [75]. Amino acid substitutions in this human AR database and those specific to the two main lineages were thus positioned onto the 3D structure using PyMOL (by Warren L. Delano, version 2004, <http://pymol.sourceforge.net/>).

#### Authors' contributions

VD performed most of the experimental work (cloning, sequencing of fish ARs), FB performed the bioinformatic analysis (phylogeny, synteny and structural data), BB studied the selective pressures acting on the ARs, IA and BH contributed to the cloning/sequencing steps, VVG and FB compared the pattern of mutations present in fish AR



with those found in human patients, VL and YG designed the study and wrote the paper.

## Additional material

### Additional file 1

**ARs sequences alignment.** Alignment of the DBD and LBD sequences of ARs compared to that of human (see Table 1 for accession numbers). In this alignment, a dot refers to the same aa as in the first sequence.

Sequences not known in 5' and 3' are shown by a hyphen ("-") sign, as well as gaps. AR-B sequences are visualized by the grey background. Alignment with other non AR sequences are available upon request. In red are the gaps or insertions characterizing the ARs in fish. Amino acids shown in green are those conserved in divergent AR-B sequences whereas those in yellow are the substitutions characterizing all AR-Bs. Positions in blue are those specific to AR-A and those in violet are common to the actinopterygian ARs.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S1.doc>]

### Additional file 2

**Phylogenetic analyses of complete AR-A and AR-B.** Phylogenetic analyses of complete AR-A and AR-B with sequences encompassing the DBD and the LBD. Four methods were used: bayesian with MrBayes, maximum likelihood (ML) with PhyML with 1000 bootstrap replicates; maximum parsimony (MP) and neighbor joining (NJ) as implemented in Seaview.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S2.pdf>]

### Additional file 3

**Orthology relationships between medaka, tetraodon and zebrafish chromosomes.** As referenced in legend Figure 4, these rose windows show the orthology relationship between chromosomes on which AR A and AR-B are located in the medaka, Tetraodon and the zebrafish. The excellent synteny observed between the chromosomes are strong remnants of the WGD that occurred specifically in the Teleost lineage. The red lines show the orthology link of the ARs among all the other orthologs (orthology based from Ensembl v48, in agreement with reciprocal best-hit analyses we performed, data not shown) shown here in green. (A) AR-A is found on chromosome T7 in the Tetraodon and on chromosome M14 in the medaka. A strong synteny is observed between these two chromosomes. (B) AR-B is found on chromosome T1 in the Tetraodon and on chromosome M10 in the medaka. A strong synteny is observed is also observed between these two chromosomes. (C) AR-A is found on chromosome Z5 in the zebrafish and on chromosome M14 in the medaka. An unequivocal synteny is observed between these two chromosomes, as shown in Figure 4-B between this zebrafish chromosome and that of the Tetraodon. (D) Although AR-B is found on chromosome M10 in the medaka and that a good synteny is observed with the chromosome Z14 in zebrafish, as observed in Figure 4-C for this species with the chromosome T1 of Tetraodon, the zebrafish lacks AR-B to the point we could not detect its pseudo-gene. (E) Table indicating the positions and Ensembl accession numbers of the relevant AR genes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S3.pdf>]

### Additional file 4

**Statistical distributions of the AR-A and AR-B substitutions in fish.**

Statistical representation of the distributions of the AR-A and AR-B substitutions in fish. The bell distribution is a random distribution of the substitutions, the diamond shows that the specificity of AR-A and AR-B versus the other AR is statistically clearly not random.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S4.pdf>]

### Additional file 5

**Human and fish substitutions along the DBD of the human AR.** Representation of the DBD of the human AR modified from [77]. The aa in green are the ones specific to AR-B; in blue, the ones specific to AR-A; in orange and red, when respectively the aa is hit by a common substitution or a different one when compared to other vertebrate ARs. Arrowheads refer to mutations found in CAIS, PAIS, MAIS and prostate cancer respectively in color pink, orange, blue and black. Different aa substitutions are shown by arrowheads side by side.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S5.pdf>]

### Additional file 6

**List of human substitutions in AR leading to AIS phenotypes and prostate cancer.** List of substitutions in human AR leading to CAIS, PAIS, MAIS and prostate cancer phenotype, as referenced into Bruce Gottlieb's database ([androgendb.mcgill.ca/AR23C.pdf](http://androgendb.mcgill.ca/AR23C.pdf)) [75], and list of the AR-A and AR-B specific substitutions detailed in Additional file 1.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S6.xls>]

### Additional file 7

**Detailed analysis of Table 2.**

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-336-S7.xls>]

## Acknowledgements

We are grateful to Marc Robinson-Rechavi, Marie Sémon, Gérard Benoit, Michael Schubert and Hector Escriva for critical reading of the manuscript. This work has been funded by MENRT, CNRS for financial support. The research leading to these results has received funding from the EC Sixth Framework Program under the project X-OMICS, grant agreement n° LSHG-CT-2004-512065.

## References

1. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al: **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431(7011):946-957.**
2. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, et al: **Zebrafish *hox* clusters and vertebrate genome evolution.** *Science* 1998, **282(5394):1711-1714.**
3. Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, et al: **A detailed linkage map of medaka, *Oryzias latipes*: comparative**

- genomics and genome evolution. *Genetics* 2000, **154(4)**:1773-1784.
4. Wittbrodt J, Meyer A, Schartl M: **More genes in fishes?** *Bioassay* 1998, **20**:511-515.
  5. Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS: **Zebrafish comparative genomics and the origins of vertebrate chromosomes.** *Genome Res* 2000, **10(12)**:1890-1902.
  6. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al.: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* 2002, **297(5585)**:1301-1310.
  7. Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS: **The zebrafish gene map defines ancestral vertebrate chromosomes.** *Genome Res* 2005, **15(9)**:1307-1314.
  8. Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B: **Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes.** *Mol Biol Evol* 2004, **21(6)**:1146-1151.
  9. Vandepoel K, De Vos W, Taylor JS, Meyer A, Peer Y Van de: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci USA* 2004, **101(6)**:1638-1643.
  10. Jordan IK, Marino-Ramirez L, Wolf YI, Koonin EV: **Conservation and coevolution in the scale-free human gene coexpression network.** *Mol Biol Evol* 2004, **21(11)**:2058-2070.
  11. Robinson-Rechavi M, Laudet V: **Evolutionary rates of duplicate genes in fish and mammals.** *Mol Biol Evol* 2001, **18(4)**:681-683.
  12. Peer Y Van de, Taylor JS, Braasch I, Meyer A: **The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes.** *J Mol Evol* 2001, **53(4-5)**:436-446.
  13. Vankatesh B, Tay A, Dandona N, Patil J, Brenner S: **A compact cartilaginous fish model genome.** *Curr Biol* 2005, **15(3)**:R82-R83.
  14. Hoegg S, Brinkmann H, Taylor JS, Meyer A: **Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish.** *J Mol Evol* 2004, **59(2)**:190-203.
  15. Hoegg S, Meyer A: **Hox clusters as models for vertebrate genome evolution.** *Trends Genet* 2005, **21(8)**:421-424.
  16. Kasahara M: **The 2R hypothesis: an update.** *Curr Opin Immunol* 2007, **19(5)**:547-552.
  17. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151(4)**:1531-1545.
  18. Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: **Gene loss and evolutionary rates following whole-genome duplication in teleost fishes.** *Mol Biol Evol* 2006, **23(9)**:1808-1816.
  19. Bertrand S, Thisse B, Tavares R, Sachs L, Chaumont A, Bardet PL, Escriva H, Duffraisse M, Marchand O, Safi R, et al.: **Unexpected Novel Relational Links Uncovered by Extensive Developmental Profiling of Nuclear Receptor Expression.** *PLoS Genet* 2007, **3(11)**:e188.
  20. Bruce AE, Oates AC, Prince VE, Ho RK: **Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate hoxB5 genes.** *Evol Dev* 2001, **3(3)**:127-144.
  21. Hurley IA, Scemama JL, Prince VE: **Consequences of hoxb1 duplication in teleost fish.** *Evol Dev* 2007, **9(6)**:540-554.
  22. Braasch I, Salzburger W, Meyer A: **Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogs involved in teleost coloration.** *Mol Biol Evol* 2006, **23(6)**:1192-1202.
  23. Escriva H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, Duffraisse M, Holland L, Gronemeyer H, Laudet V: **Neofunctionalization in vertebrates: the example of retinoic acid receptors.** *PLoS Genet* 2006, **2(7)**:e102.
  24. Howarth DL, Law SH, Barnes B, Hall JM, Hinton DE, Moore L, Maglich JM, Moore JT, Kullman SV: **Paralogous Vdr's in Teleosts: Transition of Nuclear Receptor Function.** *Endocrinology* 2008.
  25. Liley NR, Stacey NE: **Hormones, pheromones and reproductive-behavior in fish.** In *Fish Physiology Volume 9*. Edited by: Hoar DJ, RaE MD WS. New York.: Academic Press; 1983:1-63.
  26. Borg B: **Androgens in Teleost fishes.** *Comparative Biochemistry and Physiology C-pharmacology Toxicology and endocrinology* 1994, **109(3)**:219-245.
  27. Laudet V, Gronemeyer H: **The nuclear receptors.** Edited by Press A. London; 2002.
  28. Touhata K, Kinoshita M, Tokuda Y, Toyohara H, Sakaguchi M, Yokoyama Y, Yamashita S: **Sequence and expression of a cDNA encoding the red seabream androgen receptor.** *Biochim Biophys Acta* 1999, **1450(3)**:481-485.
  29. Takeo J, Yamashita S: **Two distinct isoforms of cDNA encoding rainbow trout androgen receptors.** *J Biol Chem* 1999, **274(9)**:5674-5680.
  30. Ogino Y, Katoh H, Yamada G: **Androgen dependent development of a modified anal fin, gonopodium, as a model to understand the mechanism of secondary sexual character expression in vertebrates.** *FEBS Lett* 2004, **575(1-3)**:119-126.
  31. Ikeuchi T, Todo T, Kobayashi T, Nagahama Y: **cDNA cloning of a novel androgen receptor subtype.** *J Biol Chem* 1999, **274(36)**:25205-25209.
  32. Todo T, Ikeuchi T, Kobayashi T, Nagahama Y: **Fish androgen receptor: cDNA cloning, steroid activation of transcription in transfected mammalian cells, and tissue mRNA levels.** *Biochem Biophys Res Commun* 1999, **254(2)**:378-383.
  33. Olsson PE, Berg AH, von Hofsten J, Grahn B, Hellqvist A, Larsson A, Karlsson J, Modig C, Borg B, Thomas P: **Molecular cloning and characterization of a nuclear androgen receptor activated by 11-ketotestosterone.** *Reprod Biol Endocrinol* 2005, **3**:37.
  34. Sperry TS, Thomas P: **Characterization of two nuclear androgen receptors in Atlantic croaker: comparison of their biochemical properties and binding specificities.** *Endocrinology* 1999, **140(4)**:1602-1611.
  35. Sperry TS, Thomas P: **Identification of two nuclear androgen receptors in kelp bass (*Paralabrax clathratus*) and their binding affinities for xenobiotics: comparison with Atlantic croaker (*Micropogonias undulatus*) androgen receptors.** *Biol Reprod* 1999, **61(4)**:1152-1161.
  36. Sperry TS, Thomas P: **Androgen binding profiles of two distinct nuclear androgen receptors in Atlantic croaker (*Micropogonias undulatus*).** *J Steroid Biochem Mol Biol* 2000, **73(3-4)**:93-103.
  37. Devlin RH, Nagahama Y: **Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences.** *Aquaculture* 2002, **208**:191-364.
  38. Volff JN: **Genome evolution and biodiversity in teleost fish.** *Heredity* 2005, **94(3)**:280-294.
  39. Atz JW: **Hermaphroditic Fish.** *Science* 1965, **150(3697)**:789-797.
  40. Kuwamura T, Nakashima Y: **New aspects of sex change among reef fishes: recent studies in Japan.** *Environmental Biology of Fishes* 1998, **52**:125-135.
  41. Munday PL, Caley MJ, Jones GP: **Bi-directional sex change in a coral-dwelling goby.** *Behav Ecol Sociobiol* 1998, **43**:371-377.
  42. Policansky D: **Sex choice and the size advantage model in jack-in-the-pulpit (*Arisaema triphyllum*).** *Proc Natl Acad Sci USA* 1981, **78(2)**:1306-1308.
  43. Fricke H, Fricke S: **Monogamy and sex change by aggressive dominance in coral reef fish.** *Nature* 1977, **266(5605)**:830-832.
  44. Shapryro DY: **Differentiation and evolution of sex-change in fishes I.** *Bioscience* 1987, **37(7)**:490-497.
  45. Warner RR: **Mating-behavior and hermaphroditism in coral-reef fishes.** *American Scientist* 1984, **72(2)**:128-136.
  46. Bhandari RK, Alam MA, Soyano K, Nakamura M: **Induction of female-to-male sex change in the honeycomb grouper (*Epinephelus merra*) by 11-ketotestosterone treatments.** *Zool Sci* 2006, **23(1)**:65-69.
  47. Cardwell JR, Liley NR: **Hormonal control of sex and color change in the stoplight parrotfish, *Sparisoma viride*.** *Gen Comp Endocrinol* 1991, **81(1)**:7-20.
  48. Kroon FJ, Liley NR: **The role of steroid hormones in protogynous sex change in the Blackeye goby, *Coryphopterus nicholsii* (Teleostei: Gobiidae).** *Gen Comp Endocrinol* 2000, **118(2)**:273-283.
  49. McPhaul MJ: **Androgen receptor mutations and androgen insensitivity.** *Mol Cell Endocrinol* 2002, **198(1-2)**:61-67.
  50. Heinlein CA, Chang C: **Androgen receptor in prostate cancer.** *Endocr Rev* 2004, **25(2)**:276-308.
  51. Ishiguro NB, Miya M, Nishida M: **Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality**

- of the "Protacanthopterygii". *Mol Phylogenet Evol* 2003, **27(3)**:476-488.
52. Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, et al: **Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences.** *Mol Phylogenet Evol* 2003, **26(1)**:121-138.
  53. Lavoue S, Sullivan JP: **Simultaneous analysis of five molecular markers provides a well-supported phylogenetic hypothesis for the living bony-tongue fishes (Osteoglossomorpha: Teleostei).** *Mol Phylogenet Evol* 2004, **33(1)**:171-185.
  54. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al: **The medaka draft genome and insights into vertebrate genome evolution.** *Nature* 2007, **447(7145)**:714-719.
  55. Lavoue S, Miya M, Inoue JG, Saitoh K, Ishiguro NB, Nishida M: **Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala.** *Mol Phylogenet Evol* 2005, **37(1)**:165-177.
  56. Allendorf FW, Stahl G, Ryman N: **Silencing of duplicate genes: a null allele polymorphism for lactate dehydrogenase in brown trout (*Salmo trutta*).** *Mol Biol Evol* 1984, **1(3)**:238-248.
  57. Saitoh K, Miya M, Inoue JG, Ishiguro NB, Nishida M: **Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography.** *J Mol Evol* 2003, **56(4)**:464-472.
  58. Steinke D, Salzburger W, Braasch I, Meyer A: **Many genes in fish have species-specific asymmetric rates of molecular evolution.** *BMC Genomics* 2006, **7**:20.
  59. Teletchea F, Laudet V, Hanni C: **Phylogeny of the Gadidae (sensu Svetovidov, 1948) based on their morphology and two mitochondrial genes.** *Mol Phylogenet Evol* 2006, **38(1)**:189-199.
  60. Duthheil J, Pupko T, Jean-Marie A, Galtier N: **A model-based approach for detecting coevolving positions in a molecule.** *Mol Biol Evol* 2005, **22(9)**:1919-1928.
  61. Matias PM, Donner P, Coelho R, Thomaz M, Peixoto C, Macedo S, Otto N, Joschko S, Scholz P, Wegg A, et al: **Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations.** *J Biol Chem* 2000, **275(34)**:26164-26171.
  62. Sack JS, Kish KF, Wang C, Attar RM, Kiefer SE, An Y, Wu GY, Schefler JE, Salvati ME, Krystek SR Jr, et al: **Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone.** *Proc Natl Acad Sci USA* 2001, **98(9)**:4904-4909.
  63. He B, Mingos JT, Lee LW, Wilson EM: **The FXXLF motif mediates androgen receptor-specific interactions with coregulators.** *J Biol Chem* 2002, **277(12)**:10226-10235.
  64. Suzuki K, Fukabori Y, Nakazato H, Hasumi M, Matsui H, Ito K, Kurokawa K, Yamanaka H: **Novel amino acid substitutional mutation, tyrosine-739-aspartic acid, in the androgen receptor gene in complete androgen insensitivity syndrome.** *Int J Androl* 2001, **24(3)**:183-188.
  65. Ahmed SF, Cheng A, Dovey L, Hawkins JR, Martin H, Rowland J, Shimura N, Tait AD, Hughes IA: **Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome.** *J Clin Endocrinol Metab* 2000, **85(2)**:658-665.
  66. Harbott LK, Burmeister SS, White RB, Vagell M, Fernald RD: **Androgen receptors in a cichlid fish, *Astatotilapia burtoni*: structure, localization, and expression levels.** *J Comp Neurol* 2007, **504(1)**:57-73.
  67. Burmeister SS, Kailasanath V, Fernald RD: **Social dominance regulates androgen and estrogen receptor gene expression.** *Horm Behav* 2007, **51(1)**:164-170.
  68. Rumpf LL, Lohr RL, Dooley H, Flajnik MF: **Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark.** *BMC Immunol* 2004, **5(1)**:8.
  69. Escriva H, Robinson M, Laudet V: **Evolutionary biology of the nuclear receptor superfamily.** In *Steroid/Nuclear receptor superfamily: Practical Approach* Edited by: Picard D. Academic Press; 1999:1-28.
  70. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32(5)**:1792-1797.
  71. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny.** *Comput Appl Biosci* 1996, **12(6)**:543-548.
  72. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52(5)**:696-704.
  73. Duthheil J, Gaillard S, Bazin E, Glemin S, Ranwez V, Galtier N, Belkhir K: **Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics.** *BMC Bioinformatics* 2006, **7**:188.
  74. Thioulouse J, Lobry JR: **Co-inertia analysis of amino-acid physico-chemical properties and protein composition with the ADE package.** *Comput Appl Biosci* 1995, **11(3)**:321-329.
  75. Gottlieb B, Lehvaslaiho H, Beitel LK, Lumbroso R, Pinsky L, Trifiro M: **The Androgen Receptor Gene Mutations Database.** *Nucleic Acids Res* 1998, **26(1)**:234-238.
  76. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13(5)**:555-6.
  77. Verrijdt G, Tanner T, Moehren U, Callewaert L, Haelens A, Claessens F: **The androgen receptor DNA-binding domain determines androgen selectivity of transcriptional response.** *Nuclear Receptors: Structure, Mechanisms and Therapeutic Targets. Biochemical Society Transactions* 2006, **34(6)**:1089-1094.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

