Expression and Sequence Evolution of Aromatase cyp19a1 and Other Sexual Development Genes in East African Cichlid Fishes

Astrid Böhne,*1 Corina Heule,1 Nicolas Boileau,1 and Walter Salzburger1

1Zoological Institute, University of Basel, Basel, Switzerland
*Corresponding author: E-mail: astrid.boehne@unibas.ch.
Associate editor: John Parsch

Sequences newly reported in this article have been deposited in GenBank (accession nos. KC684559–KC684586).

Abstract

Sex determination mechanisms are highly variable across teleost fishes and sexual development is often plastic. Nevertheless, downstream factors establishing the two sexes are presumably conserved. Here, we study sequence evolution and gene expression of core genes of sexual development in a prime model system in evolutionary biology, the East African cichlid fishes. Using the available five cichlid genomes, we test for signs of positive selection in 28 genes including duplicates from the teleost whole-genome duplication, and examine the expression of these candidate genes in three cichlid species. We then focus on a particularly striking case, the A- and B-copies of the aromatase cyp19a1, and detect different evolutionary trajectories: cyp19a1A evolved under strong positive selection, whereas cyp19a1B remained conserved at the protein level, yet is subject to regulatory changes at its transcription start sites. Importantly, we find shifts in gene expression in both copies. Cyp19a1 is considered the most conserved ovary-factor in vertebrates, and in all teleosts investigated so far, cyp19a1A and cyp19a1B are expressed in ovaries and the brain, respectively. This is not the case in cichlids, where we find new expression patterns in two derived lineages: the A-copy gained a novel testis-function in the Ectodine lineage, whereas the B-copy is overexpressed in the testis of the speciest-richest cichlid group, the Haplochromini. This suggests that even key factors of sexual development, including the sex steroid pathway, are not conserved in fish, supporting the idea that flexibility in sexual determination and differentiation may be a driving force of speciation.

Key words: sex determination, adaptive evolution, cichlid fishes, aromatase, developmental system drift, gene duplication.

Introduction

Sexual development in animals involves the steps of sex determination, the initial decision to develop either ovaries or testes, and sex differentiation, the subsequent developmental program establishing the male and female phenotype through the action of steroid hormones. Interestingly, the master regulatory triggers of sex determination are not conserved throughout the animal kingdom, and sex differentiation can be rather plastic (Cutting et al. 2013). This is particularly exemplified by teleost fishes, in which all sorts of natural hermaphroditism occur and sex changes can be induced through a variety of external factors (Godwin 2010, and references therein).

Contrary to the observed variation in the initial triggers of sex determination, the downstream genetic factors of sexual development are thought to be conserved between species because they are part of complex genetic networks with often pleiotropic effects. Their modifications presumably have deleterious effects, whereas changes “at the top of the hierarchies” are apparently better tolerated (Marín and Baker 1998). Previous studies have thus mainly focused on sex differentiation and the ubiquitous downstream genes (Valenzuela et al. 2013). Of particular interest here are the presumably conserved genetic and enzymatic network regulating the production of sex steroid hormones in vertebrates, where estrogens mediate ovarian development and androgens testis differentiation. For example, the action of cyp19a1, which aromatizes androgens into estrogens, is considered a general feature of ovarian development in fish, amphibians, reptiles, and mammals (for a review see Nakamura 2010), whereas sox9 appears to be its major counterpart in testis development (Valenzuela et al. 2013).

Sex-determining systems in teleosts, the taxonomically largest group of vertebrates, changed frequently during evolution and vary even between closely related species (Mank et al. 2006). So far, five male master sex-determining genes (i.e., equivalents to the mammalian sry) have been identified: DMY/dmrt1bY in medaka (Oryzias latipes) and its sister species O. curvinotus (Matsuda et al. 2002, 2003; Nanda et al. 2002); GsdY, a member of the transforming growth factor-beta superfamily, in O. luzonensis (Myosho et al. 2012); amhr2Y in the rainbow trout (Yano et al. 2012); amhy, (Y-chromosomal copy of the anti-Müllerian hormone) in the Patagonian pejerrey (Odontesthes hatcheri) (Hattori et al. 2012); and in three Takifugu species (T. rubripes, T. pardalis, and T. poecilonotus) an association between male sex and a (proto-)Y-chromosomal single nucleotide polymorphism in the amh receptor type II (amhr2) was found (Kamiya et al. 2012). Four of these genes, DMY/dmrt1bY, GsdY, amhy and amhr2, exemplify the scenario that genes already implicated downstream in a network (such as the genetic cascade of sex determination) are more readily uprecruited to the top of the hierarchy than...
nonimplicated—genes—a hypothesis originally proposed by Wilkins (1995). Sex determination and differentiation have also been studied in various other model fish species including zebrafish (Anderson et al. 2012), the poeciliid fishes guppy (Tripathi et al. 2009) and platyfish (Böhne et al. 2009), with a main focus on understanding and characterizing the genetic control over the establishment of the two sexes. Also, the downstream factors or particular genetic pathways of, for example, gonad development have been studied in several species (for review see Nakamura 2010; Angelopoulou et al. 2012). Another field of research that has received much attention is sex change in fish and, particularly, the identification of hormones and chemical components (including pollutants) that influence it (for review see Scholz and Klüver 2009). Thus, a considerable body of literature on sex determination and differentiation in fish is available, including data on many candidate genes.

Cichlid fishes of the East African Great Lakes Malawi (LM), Victoria (LV), and Tanganyika (LT) are a prime model system in evolutionary biology and provide an exceptional opportunity to study organismal diversification (Kocher 2004). Each of the Great Lakes harbors cichlid species assemblages counting hundreds of endemic species, which evolved within a few millions of years (LT; Salzburger et al. 2002), to less than 150,000 years (LV; Verheyen et al. 2003) only. This opens up a unique possibility for comparative studies at different evolutionary time scales, involving repeated patterns of diversification. As for the majority of fish species, the triggers of sex determination in cichlids are largely unknown. Yet, it becomes clear that also in this group various mechanisms exist, including genetic systems and environmental triggers such as water pH and temperature (Baroiller 2009; Reddon and Hurd 2013). The known genetic factors in cichlids include, for example, sex determination via B-chromosomes (small supernumerary/accessory chromosomes, Yoshida et al. 2011) and male and female heterogametic sex chromosome systems, with the possibility of both systems co-existing within a single species (Roberts et al. 2009). The best-studied cichlid species with respect to sexual development is the Nile tilapia (Oreochromis niloticus), a member of a more basal lineage, widely distributed in rivers and lakes of Africa. The Nile tilapia has an XX-XY sex-determining system, which can substantially be influenced by temperature (Baroiller 2009). For this species, expression profiles of key genes of sexual development are available (Ijiri et al. 2008), which is not the case for other cichlid species such as the radiating lineages in East Africa. The only exception to some extent is the LT cichlid Astatotilapia burtoni, in which the relation between hormones and behavior have been examined (Renn et al. 2012). To the best of our knowledge, sex determination mechanisms and sex differentiation have not been studied in a phylogenetically representative set of East African lake cichlids. To this end, it is particularly important to focus on the cichlid assemblage of LT, which is the oldest of the three lakes and, hence, home to the genetically, morphologically and ecologically most diverse cichlid species flock (Salzburger et al. 2002). LT’s 250 mostly endemic species (Snoeks 2001) are currently classified into 12 to 16 lineages, so-called tribes (Muschick et al. 2012). The taxonomic situation is quite different in LM and LV, where the adaptive radiations consist of only one of these tribes, the Haplochromini (Salzburger et al. 2002).

Genes involved in sex determination (including its master regulators), differentiation, and reproduction often evolve under a positive selection regime (Sorhannus and Kosakovský Pond 2006; Hassellmann et al. 2008; Morgan et al. 2010; Sobrinho and de Brito 2010). This has been explained by selection for reproductive isolation (which may ultimately lead to speciation, Vacquier and Swanson 2011), intersexual (genomic) conflict (Rice and Holland 1997), sperm competition (Pizzari and Birkhead 2002), and/or male-female coevolution (Swanson et al. 2003). This general trend seems to hold true also for cichlids, in which previous scans for positive selection uncovered three genes with putative roles in sexual development and reproduction: an aromatase gene (cyp19a1A, Gerrard and Meyer 2007), a gene with a gonad-specific expression profile (SPP120, Gerrard and Meyer 2007), and a meiotic maturation factor (20-beta-hydroxysteroid dehydrogenase, Baldo et al. 2011).

Here, we investigate 28 candidate genes in detail in cichlids. The genes were recovered from a thorough literature survey focusing on genes involved in sex determination and differentiation in vertebrates, and, especially in teleosts (see supplementary table S1 for details, Supplementary Material online). These candidates include gene copies that have emerged in the course of the fish-specific whole-genome duplication (FSGD; Meyer and Schartl 1999). To our knowledge, this has not been done before for genes involved in sexual development, although it seems crucial as gene duplications open the routes to partitioning gene functions and to neo-functionalization. We first investigate the candidate genes for sequence evolution using the available cichlid genomes (including species from all three lakes) and study their expression in ovary, testis and brain tissue in three species from LT (A. burtoni, Ophthalmotilapia ventralis, and Neolamprologus pulcher), which were chosen with respect to available genetic and genomic resources (Baldo et al. 2011 and BROAD; www.broadinstitute.org/models/tilapia, last accessed July 23, 2013). Sequence data analysis of key genes are then extended to a phylogenetically representative set of 32 species belonging to 13 different tribes and the expression analysis is extended to nine East African cichlid species, representing four tribes, including species from all three lakes.

**Results**

**Sequence Evolution of Candidate Genes: Signs of Positive Selection in East African Cichlids**

To test for signs of positive selection in our set of 28 candidate genes, we used the annotated sequences of the Nile tilapia genome as reference to recover full-length coding sequences from all available cichlid genomes (A. burtoni and N. brichardi from LT, Metriaclima zebra from LM, and Pundamilia nyererei from LV; BROAD institute, see supplementary table S2 for details, Supplementary Material online). Using tilapia as outgroup, we tested for signs of positive selection specific to the
East African cichlid radiations. We calculated mean gene wide Ka/Ks (nonsynonymous to synonymous substitution rates; see supplementary table S3 [Supplementary Material online] and Material and Methods for details) for all pairwise comparisons. The two genes with the highest gene-wide Ka/Ks values were cyp19a1A, as also detected by Gherard and Meyer in an expressed sequence tag (EST) wide screen (Gherard and Meyer 2007; discussed earlier), and nanos1A, which has not been studied in this context in cichlids so far. Note that a role for a nanos family member in germ cell differentiation has been proposed in tilapia (Kobayashi 2010).

One drawback with gene-wide Ka/Ks analyses is that they might miss signals of adaptive sequence evolution that are restricted to certain parts of a gene only; it is generally unlikely that all positions of a given gene experience positive selection at the same time, and many cases are known where positive selection is restricted to particular domains of a gene (Salzburger et al. 2007). To account for this, we analyzed our gene set using Selecton (Doron-Faigenboim et al. 2005; Stern et al. 2007) to test for site-wise selection, again using Nile tilapia as reference. We detected positively selected sites in the following genes: cyp19a1A (in 60 amino acid positions out of 511), foxI2B (9 of 268), sox9A (1 of 500), sox9B (7 of 484), tdrd1 (96 of 1,164), and wnt4A (2 of 352). The results for cyp19a1A and tdrd1 remained significant after comparison with the alternative hypotheses (i.e., no positive selection, see supplementary fig. S1 [Supplementary Material online] for details).

Sequence Evolution of Cyp19a1 in Cichlids

As the initial analyses based on the five cichlid genomes pointed out new patterns of sequence evolution for cyp19a1A, whereas we could not detect a single site under positive selection in the B-copy (cyp19a1B; supplementary fig. S1, Supplementary Material online), we decided to further study these two genes in East African cichlids. Overall, cyp19a1B is more conserved among the five cichlid genomes: Cyp19a1A differs in 12.2% (average over all pairwise comparisons to tilapia, done with KaKs_calculator, Zhang et al. 2006) of all amino acid sites (3% synonymous changes and 9.2% nonsynonymous changes), whereas Cyp19a1B varies in 9.8% of all amino acid sites (4.8% synonymous changes and 5% nonsynonymous changes). Especially, nonsynonymous substitutions occur more often in Cyp19a1A compared to Cyp19a1B.

We then extended our analyses to a more representative set of East African cichlids. To this end, we used Ion Torrent next-generation amplicon sequencing (LifeTechnologies) to re-sequence a 3.6 kb genomic fragment of cyp19a1A in 28 cichlid species belonging to 12 different tribes (see Materials and Methods and supplementary table S4 [Supplementary Material online]), GENBANK accession numbers KC684559–KC684586. The cyp19a1A coding sequences of all 28 newly sequenced species plus the sequences obtained from the cichlid genomes (O. niloticus as reference, N. pulcher, A. burtoni, and P. nyererei) were subjected to tests for positive selection as described earlier. The overall gene-wide Ka/Ks of 0.91 resulting from this analysis was similar to the value obtained with the limited data set. However, when checking in the comprehensive data set for site-wise selection with Selecton, we detected an additional 50 amino acid sites (adding up to a total of 108 out of 511 sites) under positive selection. All but three of the sites suggested to have evolved under positive selection in the initial analysis were confirmed by this second, more extensive analysis.

Placing the sites under a positive selection regime into the structural context of the well-studied aromatase protein (fig. 1) revealed that they do not cluster in a specific region of the coding sequence but are rather distributed over the entire protein and also occur in major functional domains. Additionally, the observed amino acid changes (marked with arrows above the strongest selected sites, fig. 1) are mostly not conservative. This suggests that cyp19a1A has experienced important changes in its functionality in East African cichlids.

To obtain a more detailed understanding of cyp19a1A sequence evolution in East African cichlids, we investigated positive selection in the phylogenetic context using HyPhy (Pond et al. 2005) and Codeml (Yang 2007) (fig. 2 and table 1).

These analyses uncovered several branches in the phylogeny with particularly strong adaptive sequence evolution. One of these branches represents the ancestor of the adaptive radiation of cichlids in the East African Great lakes (branch after split from the outgroup O. niloticus; light gray box, fig. 2). The signal of positive selection gets even stronger when the basal Tanganyikan tribes Bathybatini, Trematocarini, and Boulengerochromini are excluded (dark gray box, fig. 2). A strong signal of positive selection was also found in the lineage leading to the taxonomically largest cichlid tribe, the Haplochromini (dark green box, fig. 2, remember that all cichlids of LV and LM belong to this group). The so detected amino acid positions are consistent with the ones shown in figure 1 and fall into the two first categories of strongest positive selection of Selecton (fig. 2 and table 1).

Comparative Sequence Analysis of Cyp19a1A and B: Transcription Start Site Variation and Promoter Evolution

Using cichlid transcriptome databases (access over BROAD cichlid genome consortium, assembled RNA-seq data for the species with sequenced genomes), we investigated the transcription start sites (TSSs) of both cyp19a1A gene copies. Cyp19a1A transcripts were found in transcriptome databases derived from brain, gonad and mixed embryonic tissue. Interestingly, there are substantial differences between the two genes: In cyp19a1A, the TSSs of the three presumably functional transcripts (i.e., transcripts comprising the start codon marked with an orange box in supplementary fig. S2, Supplementary Material online) found in O. niloticus, A. burtoni, and M. zebra are 20 nucleotides apart of each other. Quite to the contrary, the TSSs of cyp19a1B are located more than 1,200 bp apart from each other in the transcripts found for A. burtoni, P. nyererei, and M. zebra. A closer inspection revealed two alternative splice forms of cyp19a1B:...
transcription of cyp19a1b in cichlids can start inside intron 1 or mark the beginning of the untranslated exon 1, in which case intron 1 is correctly spliced out (see supplementary fig. S2 [Supplementary Material online] for the intron location). For the two basal species O. niloticus and N. brichardi, we only detected the splice variant with the TSS inside exon 1.

Regulation of Cyp19a1A and B Expression in (Cichlid) Fish

The cyp19a1A and B promoter regions have previously been studied in several teleost species (Callard et al. 2001; Diotel et al. 2010) just as the promoter of the single copy cyp19a1 gene in tetrapods (Hinselwood et al. 2000; Nakagawa and Iwabuchi 2012). In a first step, to characterize general patterns in the molecular evolution of the cyp19a1 upstream regions in teleosts, we compared the existing cichlid sequences with all other available teleost aromatase promoters (fig. 3 and table 2). Note that cyp19a1B is absent from the available M. zebra (LM) genome assembly version 1, so that we excluded this species in further analysis. Using Vista plots of nucleotide similarity (Mayor et al. 2000; Frazer et al. 2004), we identified three conserved regions in the upstream sequences of both gene copies marked green, yellow, and brown in figure 3. Here, we define the upstream region as 5′-sequence between the start codon of the cyp19a1A and B genes and their respective adjacent gene. For cyp19a1A, we analyzed the 3,394 bp until its neighboring gene gliomedinA (gldnA); for cyp19a1B, we analyzed 2,571 bp to gldnB. We inspected the three conserved regions with regard to common transcription factor binding sites focusing on factors with a known function/response role in brain/central nervous system (CNS) and gonads/germ cells, as these are the tissues in which the aromatase genes are described to be expressed. Although remarkable differences in the expression patterns of cyp19a1A and B exist between teleosts, and especially between cichlids (discussed later), the cyp19a1 promoters still show conserved regions between all teleosts, underlining their general functional importance.
Putative transcription factor binding sites inside the conserved regions and shared between teleosts (underlined text in Table 2) comprise factors belonging to the two investigated categories (brain/CNS and gonads/germ cells). None of these categories seems to dominate, however, neither in \textit{cyp19a1A} nor in \textit{cyp19a1B}.

Using MatInspector (Genomatix Software GmbH), we identified putative estrogen response elements (EREs) in the upstream regions of both \textit{cyp19a1} duplicates, which is consistent with previous studies in fish (Diotel et al. 2010). In the

**Fig. 2.** Ka/Ks rates on a phylogenetic tree of aromatase \textit{cyp19a1A} in cichlids. Phylogenetic tree of aromatase \textit{cyp19a1A} of representatives of the major East African cichlid lineages showing the Ka/Ks rates reconstructed with HyPhy and branch-site selection as calculated with codeml. Nodes are labeled with Bayesian posterior probabilities as obtained with MrBayes. Colored boxes denote clades that were tested against all other branches with codeml (see Materials and Methods for details). Significant obtained Ka/Ks values ($\omega$) are shown next to the boxes in the respective colors. $\omega$ values denoted in blue were calculated with Branch-site REL in HyPhy. Table 1 presents the sites that are under positive selection as detected by codeml in the different clades that show a significant $\omega$ with a probability (BEB) of more than 50%. The values obtained with Selecton are given in addition.
first teleost-conserved block upstream of cyp19a1A (green box, fig. 3), we also detected putative Nr5a2 (Pezzi et al. 2004) and Fox-family binding sites (Wang et al. 2010) shared between teleosts. Here, we also detected partially conserved sites belonging to the category “brain/CNS transcription factors,” such as neuronal factors of the Oct family including Oct-6 and the transcriptional repressor DREAM. Oct-6 is known to be expressed in brain and testis, to play a role in myelination and to act as an RNA polymerase II distal enhancer (Hofmann et al. 2010). DREAM plays a role in noci
ception in the brain and is broadly expressed in testis (Rivas et al. 2004).

Tilapia, platyfish, and medaka contain, in the first block, a potential site for Nr1f1, which is important for the development of the cerebellum, regulation of (lipid) metabolism, lymphocyte development, inflammatory responses, and possibly myogenesis or muscle function (Jetten 2009). Finally, we detected putative CREB (cAMP response element-binding protein)-binding sites in a subset of species. CREB can act as transcriptional activator or repressor and acts in neurogenesis, neuronal survival and plasticity (Barco and Marie 2011; Gruart et al. 2012). Note that CREB binding sites have been described in the cyp19a1A promoter before (Chang et al. 2005). Finally, we detected shared putative Sox-family binding sites in the second conserved block (Callard et al. 2001).

Interestingly, all three conserved blocks of the cyp19a1B teleost promoter contain putative homeobox domain transcription factor binding sites, including sites for members of the HOX family. Additionally, we detected possible binding sites for the Fox-gene-family in block one, which is also the case in the cyp19a1A promoter (blocks 1 and 2). In three teleost species (platyfish, medaka, and stickleback), we found conserved putative sites for Hmx3 (involved in specification of neuronal cell types, required in the hypothalamic-pituitary axis), RFX1 (transcriptional activator of MHC class II genes), and PCE-1 (an element usually found in promoter regions of photoreceptor genes). However, these sites were not detected in tilapia or pufferfishes.

A closer inspection of the cichlid sequences only (supplementary fig. S2, Supplementary Material online) revealed that these contain, in the conserved upstream regions of both gene copies, putative binding sites for SF-1, several Sox-family members and several androgen response elements in addition to the ERE and estrogen receptor (ER)-related sites. The promoter region of cyp19a1A in cichlids contains also putative binding sites for Nr5a2 and Dmrt1. Furthermore, we confirmed the existence of putative binding sites for Nkx6-1 in both cyp19a1 promoters in cichlids, just as previously described for cyp19a1B in tilapia, zebrafish, and medaka (Chang et al. 2005; Diotel et al. 2010).

Expression Profiling of Cyp19a1A and B and Other Sexual Development Genes in Cichlids

In a first step, we tested by quantitative real-time polymerase chain reaction (qRT-PCR), the expression of our set of candidate genes in three different species from LT (A. burtoni, O. ventralis, and N. pulcher, for species choice see Materials and Methods) belonging to three different cichlid lineages. We obtained expression profiles in both gonad and brain tissue for 23 of these candidate genes (see Materials and Methods for details on our working procedure; the remaining five genes, amh, arA, arB, dax1b, and tdrd1 could not successfully be amplified on cDNA from gonadal tissue). Brain tissue was included in these experiments, because several of the

---

**Table 1.** Cyp19a1A Amino Acid Sites under Positive Selection in Different Phylogenetic Clades of East African Cichlids.

<table>
<thead>
<tr>
<th>Cyp19a1A Amino Acid Position in O. niloticus Reference Protein</th>
<th>Category of Positive Selection on Selecton Scale</th>
<th>BEB of Positive Selection in Clades with Significant ω^a^ as Tested with Codeml</th>
<th>1^a^ (Light Gray)</th>
<th>2 (Dark Gray)</th>
<th>3 (Dark Green)</th>
<th>4 (Light Green)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0.701</td>
<td>0.712</td>
<td>0.941</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>0.5</td>
<td>0.81</td>
<td>0.852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>0.624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>1</td>
<td>0.563</td>
<td>0.569</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>1</td>
<td>0.646</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>1</td>
<td>0.998</td>
<td>0.998</td>
<td>0.885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>1</td>
<td>0.978</td>
<td>0.98</td>
<td>0.535</td>
<td>0.628</td>
<td></td>
</tr>
<tr>
<td>257</td>
<td>1</td>
<td>0.858</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>314</td>
<td>1</td>
<td>0.772</td>
<td>0.831</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>362</td>
<td>1</td>
<td>0.816</td>
<td>0.827</td>
<td>0.621</td>
<td>0.696</td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>1</td>
<td>0.777</td>
<td>0.813</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>461</td>
<td>1</td>
<td>0.859</td>
<td>0.659</td>
<td>0.721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>471</td>
<td>1</td>
<td>0.979</td>
<td>0.98</td>
<td>0.944</td>
<td>0.956</td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>1</td>
<td>0.628</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.**—BEB, Bayes Empirical Bayes.

^a^1 and 2 represent the two categories of strongest positive selection as detected with Selecton corresponding to Ka/Ks values over 2.7 for category 1 and 1–2.7 for category 2 (fig. 1).

^a^ω = Ka/Ks.

^a^The phylogenetic categories (1, light gray; 2, dark gray; 3, dark green; 4, light green) correspond to the ones shown in figure 2.
Comparison of the homologous regions surrounding cyp19a1A and cyp19a1B in teleosts. Shuffle-LAGAN Vista plots for cyp19a1A and cyp19a1B. Blue peaks indicate conserved coding regions, pink peaks represent conserved noncoding regions. Light blue regions correspond to UTRs. Green, yellow, and brown boxes surround conserved possible promoter regions for both genes (in the text referred to as conserved boxes). Putative transcription factor binding sites for factors known to be active in gonads/germ cells and brain/CNS identified in these regions are indicated in table 2.
candidate genes are putative members of the hypothalamic–pituitary–gonadal axis, especially those involved in sex steroid synthesis, and hence, are expected to be expressed in the brain. This expectation can also be made for the adult brain because the fish brain produces sex steroid hormones necessary for sexual differentiation, plasticity, and reproduction throughout the entire lifespan of a fish unlike the mammalian brain, which is irreversibly sexualized during early development.

Table 3 shows a summary of the expression data for the three tested cichlid species (for box plots and detailed values of all expression tests and statistics see supplementary material S2, Supplementary Material online). Most of the tested gene candidates show—at least to some extent—expression patterns as described for other species (e.g., dmrt1 is also a testis gene in cichlids). The tissue of highest expression level is most often conserved between the examined species. However, we also found deviations from described expression patterns.

We cannot confirm a role for rspo1 as an ovary factor in adult cichlids, where it is mainly expressed in the brain. Rspo1 has been suggested to act as an activator of the wnt/beta-catenin pathway in the developing ovary (Smith et al. 2008) and is overexpressed in zebrafish ovaries (Zhang et al. 2011). Our data do, however, support a role for other members of this pathway, ctnnb1a and B and wnt4a, in the adult ovary, whereas wnt4b, just like rspo1, is predominantly expressed in the brain and testis; furthermore, ctnnb1a seems to also have experienced a shift in expression toward the brain in adult cichlids. Therefore, the role of the wnt-pathway in ovary development seems to be less conserved than previously thought, and, in addition, to differ between paralogous gene copies.
Table 3. Expression Patterns of Candidate Genes in Cichlids (Based on qRT-PCR, Calculated after Simon [2003]. See Supplementary Material S2, Supplementary Material online).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Expression in Astatotilapia burtoni</th>
<th>Expression in Ophthalmotilapia ventralis</th>
<th>Expression in Neolamprologus pulcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>cbh1A*, catenin beta A, intracellular signal transducer, part of the wnt-pathway in ovaries</td>
<td>Brain &gt; gonads</td>
<td>Ovary and brain &gt; testis</td>
<td>Ovary and brain &gt; testis</td>
</tr>
<tr>
<td>cbh1B, catenin beta B</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary &gt; female brain &gt; male brain &gt; testis</td>
</tr>
<tr>
<td>cyp11b2, cytochrome P450 family 11 subfamily B polypeptide 2, androgen synthesis, testis factor</td>
<td>Testis &gt; female brain &gt; male brain &gt; ovary</td>
<td>Testis &gt; brain and ovary</td>
<td>Testis &gt; brain &gt; ovary</td>
</tr>
<tr>
<td>cyp19a1A, aromatase cytochrome P450 family 19 subfamily A polypeptide 1 A, estrogen synthesis, ovarian aromatase</td>
<td>Ovary &gt; brain and testis</td>
<td>Gonads &gt; brain</td>
<td>Ovary &gt; testis &gt; brain</td>
</tr>
<tr>
<td>cyp19a1B, aromatase cytochrome P450 family 19 subfamily A polypeptide 1, brain aromatase</td>
<td>Testis &gt; brain and ovary</td>
<td>Male brain &gt; female brain &gt; gonads</td>
<td>Brain &gt; gonads</td>
</tr>
<tr>
<td>dax1A, dosage-sensitive sex reversal adrenal hypoplasia critical region on chromosome X gene 1 A, orphan receptor, regulates steroidogenesis, interacts with sf-1</td>
<td>Ovary &gt; testis &gt; brain</td>
<td>Ovary and brain &gt; testis</td>
<td>Ovary and brain &gt; testis</td>
</tr>
<tr>
<td>dmt1, doublesex and mab-3 related transcription factor 1, male sex/testis differentiation</td>
<td>Testis &gt; ovary</td>
<td>Testis &gt; ovary</td>
<td>Testis &gt; ovary &gt; female brain &gt; male brain</td>
</tr>
<tr>
<td>figla, factor in the germline alpha, germline transcription factor, acts in folliculogenesis</td>
<td>Ovary &gt; testis</td>
<td>Ovary &gt; testis</td>
<td>Ovary &gt; testis</td>
</tr>
<tr>
<td>foxa2, forkhead transcription factor L2 A, transcription factor, regulates cyp19a1A, ovary development and function</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary &gt; brain &gt; testis</td>
</tr>
<tr>
<td>foxl2B, forkhead transcription factor L2 B, discussed earlier</td>
<td>Gonads</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>gata4, GATA binding protein 4, transcription factor, gonad development, and function</td>
<td>Ovary &gt; testis</td>
<td>Gonads &gt; brain</td>
<td>Gonads &gt; brain</td>
</tr>
<tr>
<td>nanos1A, brain and germ cell development</td>
<td>Brain &gt; ovary &gt; testis</td>
<td>Brain &gt; ovary &gt; testis</td>
<td>Brain &gt; ovary &gt; testis</td>
</tr>
<tr>
<td>nanos1B, discussed earlier</td>
<td>Brain &gt; ovary &gt; testis</td>
<td>Brain &gt; ovary &gt; testis</td>
<td>Brain &gt; ovary &gt; testis</td>
</tr>
<tr>
<td>nrs2A (lrh-1), nuclear receptor subfamily S, group A, member 2, acts in steroidogenesis</td>
<td>Brain &gt; gonads</td>
<td>Brain &gt; gonads</td>
<td>Brain &gt; gonads</td>
</tr>
<tr>
<td>nrs5A, nuclear receptor subfamily S, group A, member 5, related to sf-1, gene of the nrs5a family, regulation of steroid/steroid metabolism</td>
<td>Not expressed</td>
<td>Brain and ovary</td>
<td>Brain &gt; ovary &gt; testis</td>
</tr>
<tr>
<td>nsp1, R-spondin-1, activates wnt/beta-catenin pathway in ovary differentiation</td>
<td>Brain &gt; gonads</td>
<td>Brain &gt; testis &gt; ovary</td>
<td>Brain &gt; gonads</td>
</tr>
<tr>
<td>sf-1 (nrs5a1) steroidogenic factor 1, orphan nuclear receptor, gonadal differentiation, regulates steroidogenesis</td>
<td>Gonads &gt; brain</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>sox9A, SRY (sex determining region Y)-box 9 A, transcription factor, regulates amh expression together with sf-1, gonad development, Sox9 is a conserved vertebrate testis factor</td>
<td>Brain &gt; gonads</td>
<td>Brain &gt; ovary &gt; testis</td>
<td>Brain &gt; gonads</td>
</tr>
<tr>
<td>sox9B, SRY (sex determining region Y)-box 9 B</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary and brain &gt; testis</td>
<td>Brain &gt; ovary &gt; testis</td>
</tr>
<tr>
<td>wnt4A, wingless-type MMTV integration site family member 4A, female reproductive development</td>
<td>Ovary &gt; brain &gt; testis</td>
<td>Ovary and brain &gt; testis</td>
<td>N.A.</td>
</tr>
<tr>
<td>wnt4B, wingless-type MMTV integration site family member 4B</td>
<td>Brain and testis</td>
<td>Brain and testis</td>
<td>Brain &gt; testis &gt; ovary</td>
</tr>
<tr>
<td>wtf1A, Wilm's tumor 1 A, gonad formation, testis differentiation, interacts with sf-1</td>
<td>Gonads &gt; brain</td>
<td>Gonads &gt; brain</td>
<td>Gonad &gt; testis</td>
</tr>
<tr>
<td>wtf1B, Wilm's tumor 1 B</td>
<td>N.A.</td>
<td>Gonads &gt; brain</td>
<td>Gonad &gt; testis</td>
</tr>
</tbody>
</table>

**Note:** N.A., not applicable.

* A and B denote gene copies derived from the FSGD.

bPrimer do not amplify successfully.
In mammals, Sox9 is indispensable for testis differentiation (Kanai et al. 2005) and sox9 is the first gene to show male sex-specific expression in many species (for a review see Morais da Silva et al. 1996). In fish, both sox9 gene copies were found to be (highly) expressed in developing and adult testis (Zhou et al. 2003; Johnsen et al. 2013). In our set of cichlids, sox9A seems to be important in brain tissue and sox9B in ovaries and to some extent in the brain.

The most intriguing expression patterns observed in our data set involved once more the aromatase gene cyp19a1. Although the brain expression patterns of cyp19a1B in two of our study species (A. pulcher and O. ventralis) are consistent with the general trend in teleosts, that is, expression of the A-copy in ovaries and the B-copy in the brain (Callard et al. 2001; Cao et al. 2012), the member of the derived and exceptionally species-rich haplochromine cichlids (A. burtoni) shows a different pattern. In A. burtoni, cyp19a1B is overexpressed in testis and not in the brain, a pattern so far not described. Also cyp19a1A shows a deviation from the general trend in teleosts, in this case in the Ectodini O. ventralis, where we detected a similar expression level in testis and in ovaries (the other two cichlids show overexpression in ovaries, like other teleosts).

Because of the observed new expression patterns and the strong signal for positive selected sites in cyp19a1A, we decided to study both gene copies in more detail and more species focusing on a phylogenetically representative set of the East African cichlids. To this end, we performed qRT-PCR expression analysis in adults of six additional cichlid species: a representative of an extra cichlid tribe (Perissodiscus microlepis, tribe Perissodini) as well as more representatives of the Ectodini (Enantiopus melanogenys), the Lamprologini (Altolamprologus fasciatus), and, particularly, the Haplochromini (Ctenochromis horei, Pseudotropheus pulpican, and Haplochromis obliquidens). Thus, we now also included members of the two major haplochromine adaptive radiations, that of lakes Victoria (H. obliquidens) and Malawi (P. pulpican). The results of these experiments are shown in figure 4 (for further details on values and statistics see supplementary material S3, Supplementary Material online).

First, we confirmed that the newly detected overexpression of cyp19a1B in testis of A. burtoni is representative for haplochromine cichlids in general, indicating that this shift in gene expression has been acquired in the ancestor of the most species-rich group of cichlids.

Second, the expression of cyp19a1A in testis in O. ventralis also seems to be a more general pattern in Ectodini, as evidenced by the analysis of a second representative of this tribe, E. melanogenys. The expression patterns observed in Lamprologini and Perissodini, however, resemble the presumably ancestral patterns known from other teleosts.

**Discussion**

In this study, we investigated supposedly conserved candidate genes implicated in sex determination, sex differentiation, and reproduction in East African cichlid fishes. We found substantial variation in adult expression patterns for several of these genes, suggesting that not only the master initial regulators but also the downstream genetic factors of sex determination are less conserved than previously thought. We could show that this variation depends on the species investigated, and that expression can vary even between closely related species. Furthermore, variation in sequence evolution and expression can act on the level of additional gene copies such as the ones that emerged from the FSGD (e.g., see our data on sox9A and B and wnt4A and B), opening the path for sub- and neofunctionalization. This suggests that not only the master regulatory triggers of sex determination are labile in the animal kingdom but also that the downstream genetic factors can experience substantial changes in coding sequence and function. This adds a further level of plasticity to the processes of sexual development, at least in teleost fish.

**Aromatase Gene Expression in Cichlids**

The aromatase cyp19a1 is considered one of the most important, if not the key gene in vertebrate sexual development (Diotel et al. 2010; Valenzuela et al. 2013). Cyp19a1 is the sensitive gene that reacts to temperature changes and thus induces sexual development in reptiles (Merchant-Larios and Díaz-Hernández 2013), it has been implicated with sex change in fish (Liu et al. 2009; Nozu et al. 2009) and it is one of the first genes to show sex differences in expression in developing tilapia (Ijiri et al. 2008). Cyp19a1 is a fundamental component of the estrogen pathway and it has been suggested that hormonal actions, such as the ones of estrogen, are so important that sex determination relies on them exclusively, regardless of their initial activation (Angelopoulou et al. 2012; Merchant-Larios and Díaz-Hernández 2013).

In adult vertebrates, cyp19a1 typically comes in two flavors, the brain and the ovary aromatase. In tetrapods, this target specificity is achieved through the generation of alternative 5'-UTR transcripts derived from a single gene, mediated by tissue-specific promoters (Toda and Shizuta 1993; Golovine et al. 2003); an exception is the pig, which has three copies of cyp19a1 (Graddy et al. 2000). Teleosts, on the other hand, have two copies of the gene generated by the FSGD and generally express the A-copy in ovaries and the B-copy in the brain (an exception are the European and the Japanese eels that only retained one copy after the FSGD, which is expressed in both, the brain and gonads, Tzchori et al. 2004; Jeng et al. 2012). The high expression levels in the fish brain, especially of cyp19a1B, are explained by its regulation through a feedback loop of estrogen on EREs in its promoter (Callard et al. 2001; Diotel et al. 2010, and references therein). In general, the functions of cyp19a1A and cyp19a1B have been suggested to be highly conserved among teleosts (Callard et al. 2001).

Here, we show that East African cichlid fishes deviate from this trend. Cyp19a1A, which shows strong signs of positive selection in cichlids, has experienced a shift in expression from ovary-specific to testis and ovaries in the Ectodini of LT. This pattern clearly contradicts the general assumption that an upregulation of cyp19a1A is needed to initiate and maintain ovary function in fish, whereas its downregulation induces testis development (Guiguem et al. 2010). Furthermore, the
protein domains of Cyp19a1a could be substantially influenced by the changes occurring in its coding sequence in East African cichlids and hence its functionality in the estrogen pathway remains to be determined in these fish.

Cyp19a1B, on the other hand, shows high sequence conservation in the investigated cichlids suggesting that it preserved its function in testosterone aromatization to estrogen. Still, this gene has experienced a shift in expression from brain to testis in the Haplochromini. It thus appears that in this most species-rich group of cichlids the main site of adult estrogen production is the testis and not the brain. The haplochromines are the cichlid lineage displaying the most pronounced sexual color dimorphism, which has often been implicated with sexual selection via female choice, and they show the most derived female mouthbrooding behavior (Salzburger et al. 2005). Whether the unprecedented recruitment of the otherwise brain-specific gene cyp19a1B in testis is causally linked to some of these features of haplochromine cichlids would need to be investigated in the future.

Adult ectodine cichlids mainly express cyp19a1A in the gonads and haplochromine cichlids overexpress cyp19a1B in the testis. This should lead, at least in the haplochromines, to high estrogen levels in the testis, whereas the functionality of Cyp19a1A in estrogen aromatization to testosterone can be questioned.

High estrogen levels in the haplochromine cichlid A. burtoni and in many other vertebrates are linked to aggressive behavior (Trainor et al. 2006); however, the steroids effecting...
aggressive (but not reproductive) behavior are apparently not of gonadal origin in *A. burtoni* (Soma et al. 1996; Huffman et al. 2013), leaving the function of high aromatase activity in the testis of this species unexplained. Yet, none of the studies on estrogen levels in *A. burtoni* is specific for one of the two *cyp19a1* copies (Huffman et al. 2013). Pharmacologically blocking aromatase (presumably both forms), decreases aggressive behavior, whereas reproductive behavior is not influenced (Huffman et al. 2013, and references therein). We suggest further critical investigation of the usage of such inhibiting substances with respect to the sequence changes that we observed in *cyp19a1A*.

Estrogen production via aromatase in the testis of fish has been described to be involved in spermatogenesis, where estrogen is necessary for the renewal of spermatagonial stem cells and thought to be necessary for germ cell proliferation and possibly differentiation (Schulz et al. 2010). Yet, overexpression of aromatase in testis compared with ovary or brain has not been described in fish before (see table 2 in Pferrer and Blázquez 2005). Furthermore, an increase of estrogen levels has been shown to impair testis function in fish and substantially influence reproduction (Kobayashi et al. 2011), leaving the high expression levels of aromatase, and especially *cyp19a1B* in haplochromine cichlids, hard to explain with the currently available data on estrogen actions on testis/spermiogenesis.

Considering our analysis of the promoter regions of *cyp19a1A* and *B* in cichlids, we think that the presence of SF-1 and Dmrt-1 binding sites in the *cyp19a1A* promoter and the expression of both factors in the testis of cichlids (fig. 3 and table 3) is consistent with a repressive action of these two transcription factors on *cyp19a1A* expression in testis. Conversely, *cyp19a1B* is overexpressed in testis in haplocromines, suggesting that it may not be suppressed by SF-1 in testis, though the cichlid *cyp19a1B* promoter contains a putative binding site for this factor and *sf-1* is overexpressed in testis. The main expression of *nr5a2* in the brain of *A. burtoni* contradicts its activating action on *cyp19a1A* in gonads, although there is a conserved potential binding site for it in the A promoter. Still, *Nr5a2* could drive *cyp19a1A* expression in the brain or earlier during development in the gonads.

Further studies should unravel the functionality of the two aromatase copies in cichlids, their function throughout brain and gonad development and their action on adult steroid production involved in aggression and reproduction. Also the regulation of their expression by neuronal/gonadal transcription factors as well as through estrogens and androgens (as suggested by the presence of response elements for both steroid types in the cichlid promoter regions of *cyp19a1A* and *B*) need further investigation.

**Contrasting Sequence Evolution Trajectories of the Two Aromatase Gene Copies in Cichlids**

The relative contribution to phenotypic divergence of changes in the *cis*-regulatory elements of a gene versus changes in the coding sequence is still nebulous and a matter of debate (Hoekstra and Coyne 2007), although it has been argued that rapid evolution in closely related taxa such as cichlids is more likely to be mediated by regulatory changes (Baldo et al. 2011). Here, we show that even two copies of the same ancestral gene emerging from a duplication event may pursue these distinct evolutionary trajectories, both likely leading to functional divergence. *Cyp19a1A* has undergone various, presumably adaptive changes in its amino acid sequence during cichlid evolution that might substantially influence its function. *Cyp19a1B*, on the other hand, seems to be under strong purifying selection in cichlids, indicating that functional constraints act on at least one of the duplicate gene copies, though the effects of aromatase action remain illusive. Still, also the B-copy shows new evolutionary paths in cichlids: in addition to expression shift in Haplochromines (overexpression in testis) we could show the usage of different TSSs in Ectodini and Haplochromini, a feature already described for the single copy *cyp19a1A* gene in tetrapods (mentioned earlier). This shift seems to be caused by *cis* or *trans* regulatory changes. The other, more basal cichlids investigated here show conserved expression patterns in both genes resembling other teleosts. This illustrates that paralogs can be subjects to differential evolutionary fates also at late, that is, derived stages of an evolutionary radiation.

**More Flexibility in the Sex-Determining Cascade: Sox9 Expression in Cichlids—More than a Testis Factor**

In mammals, Sox9 drives testis differentiation (Kanai et al. 2005) and a similar role has been suggested for this gene in other vertebrate groups, after the observation that it is often the first gene to show male sex-specific expression (Morais da Silva et al. 1996). In teleost fish, both sox9 gene copies were described to be (highly) expressed in developing and adult testis (Zhou et al. 2003; Johnsen et al. 2013). Apparently, however, the expression pattern is more flexible in adult fish, it varies by species and depends on the gene copy investigated. In line with this, we show that sox9A seems to rather play a role in the brain and sox9B in ovaries and also to some extent in the brain. This illustrates that substantial differences exist in the expression pattern of sox9 between teleost fishes and underlines once more the importance of studying both duplicates of a single copy tetrapod gene, when present in fish, and not to conclude a function from one copy only. This is also illustrated by our data on *wnt4A* and *B*. Sox9A and B are obviously another example of a sub- or neofunctionalization of gene copies after duplication (Ohno 1970).

**Sexual Development Is Plastic in Teleost Fishes**

It has been proposed that a set of genes forms a common regulatory network of sexual differentiation in vertebrates (Angelopoulou et al. 2012; Valenzuela et al. 2013). This study shows that some of the genes implicated in sexual development and reproduction are indeed conserved between lineages. The function of these genes, however, as reflected by their varying expression patterns in different
species (this study and see Cutting et al. 2013 for a review) and changing coding sequence as we show here for the aromatase gene copies, seems to be less conserved than previously thought.

Plasticity in sexual development has been studied in detail in sex-changing fish and species that possess temperature dependent sex-determining mechanisms, where gene expression changes in response to temperature have been assessed. Most other available studies on sexual development have focused on one gene or a particular pathway of a few interacting genes. Our work was based on a broad screen of many genes and with respect to both, gene expression and sequence evolution, in a prime model system in evolutionary biology, the East African cichlids. We detected substantial differences in expression within different cichlid species and between cichlids, or certain lineages thereof, and other teleosts. The most intriguing outcome of our study is the lineage-specific shifts in gene expression in the two copies of the aromatase cyp19a1, a major core gene of sexual development, behavior, and reproduction in chordates. That the changes in gene expression in the two cyp19a1 copies occurred in derived cichlid lineages, but not in the ancestral tribes, suggests that functional divergence between two gene copies derived from the FSGD adds another level of flexibility to the sex-determining cascade of teleosts.

Taken together, most of the studied genes are part of the cascade of sex determination, differentiation, and reproduction across teleosts but their usage may differ substantially at different taxonomic levels. This can be interpreted as a developmental system drift (DSD), a change in the molecular pathway with preservation of the morphological/phenotypic outcome (True and Haag 2001), in this case the establishment, maintenance, and functioning of separate sexes. Indeed, changes of master control genes in the sex-determining cascade between closely related taxa have been called “the most striking examples of diversity in a conserved regulatory process” (True and Haag 2001), exemplifying DSD. Our data and also those of others (Wang and Sommer 2011; Valenzuela et al. 2013) indicate that this diversity not only affects the top of the sex-determining cascade but also includes sexual differentiation genes, hormone pathways and the molecular networks necessary for the development of reproductive organs and behavior. The occurring modifications can thus affect all levels of a developmental process and may depend on cis or trans changes, novel wiring, new patterns of coding sequence evolution, different gene usage and de novo recruitment and evolution of genes.

Especially the observed sequence and expression changes of the aromatase gene copies could have implications on the evolution of sex chromosomes, which might be driven by the action of hormones (Howard 2002). Our study once more demonstrates that understanding evolutionary shifts in sex-determining systems requires the characterization of downstream gene variability (Uller and Helanterä 2011). Cichlid fishes are, in this respect, an ideal model system since many closely related, hence genetically very similar species, with different sex-determining mechanisms and all sorts of reproductive strategies exist.

Materials and Methods

Tissue Sampling and Nucleic Acid Extraction

The presented data set comprises DNA sequence data for 33 species of East African cichlid fishes (see supplementary table S4, Supplementary Material online) and expression data for nine species (A. burtoni, C. horei, O. ventralis, E. melanogenys, N. pulcher, Alt. fasciatus, P. microlepis, H. obliquidens, and Pse. pulpican). The investigated species belong to 13 different tribes and thus cover a great portion of the taxonomic and phylogenetic diversity of East African cichlids (Salzburger et al. 2002, 2005). The species that were tested for candidate gene expression were chosen with regards to available genetic resources and their respective phylogenetic positions. All candidate genes were tested in A. burtoni (a member of the Haplochromini), O. ventralis (Ectodini), and N. pulcher (Lamprologini). These three cichlid tribes differ remarkably in many life history aspects and with regards to sexual development and, especially, their reproductive strategies. A. burtoni and O. ventralis are maternal mouthbreeders, whereas N. pulcher is a bi-parental substrate spawner within a cooperative helper group. Additionally, A. burtoni shows a rather strong sexual color dimorphism.

Samples of the species from LT were collected during field expeditions to the lake and surrounding rivers between 2007 and 2012 under a research permit issued by the Lake Tanganyika Research Unit, Department of Fisheries, Mpulungu, Zambia. Tissue samples of N. callipterus were kindly provided by M. Taborsky. Adult and mature specimens for two additional species (H. obliquidens from Lake Victoria and P. pulpican from Lake Malawi) were derived from the aquarium trade. Life fish were kept under standard conditions (12 h light, 12 h dark, 25 °C) at the animal facility of the Salzburger group, University of Basel, Switzerland. Research was performed under the cantonal veterinary permit no. 2317.

All samples for the expression studies were collected during the above mentioned field expeditions in 2011/2012. Tissues (brain, gonad, Gill, anal fin, and liver) were collected from wild caught adult fish within one hour to reduce gene expression shifts caused by stress. Tissues were directly stored in RNA later (Life Technologies). DNA samples (muscle and fin tissue) were taken at the same time and stored in 100% ethanol. Fish were sexed based on external morphological traits and visual gonad inspection. Samples were only included in the study when ovary or testis structure could be determined unambiguously. For the mouth-breeding species only non-breeding females were dissected. Given the sampling conditions, reliable determination of GSI (gonadosomatic index) or plasma steroid levels to infer reproductive status of the invested specimens was not feasible. Note however, that previous studies on stable laboratory populations of male A. burtoni have shown that changes in gonad mass are not necessarily reflecting reproductive maturity but are rather signs of interstitial cell division (Huffman et al. 2012, and references therein). Additionally, testis of both subordinate and dominant males contains all spermatogenic stages (Maruska and Fernald 2011) and sperm proliferation does not differ
between them (Kustan et al. 2011). A reliable assessment of reproductive status and dominance over gene expression analysis is, however, possible (Huffman et al. 2012) and sex-specific expression in our data set is consistent between samples of one sex.

Genomic DNA was extracted from ethanol preserved muscle or fin tissue applying manually a Proteinase K digestion followed by sodium chloride extraction and ethanol precipitation (Laird et al. 1991) or using the robotic workstation BioSprint 96 (Qiagen) following the manufacturer’s instructions.

Samples stored in RNAlater were incubated overnight in Trizol (LifeTechnologies) at 4 °C (this initial incubation step increases RNA yield after long-term storage in RNAlater) prior to extraction and subsequently homogenized in Trizol using a FastPrep24 (MP Biomedicals Europe). Total RNA was extracted following the Trizol protocol. RNA concentration and quality was measured using a NanoDrop1000 Spectrophotometer (ThermoScientific). RNA samples were treated with DNA-free Kit (LifeTechnologies) as recommended by the manufacturer.

Reverse Transcription and Quantitative Real-Time PCR
DNase-treated RNA was reverse transcribed using the High capacity RNA-to-cDNA kit (LifeTechnologies) following the manufacturer’s protocol. Real-time PCR experiments were run on a StepOnePlus Real-Time system (LifeTechnologies) on a final cDNA concentration of 0.5 ng/μl for the experiment shown in table 3 and 1 ng/μl for the experiment shown in figure 4 with 200 mM final primer concentration and the SYBR Green master (RoX) dye (Roche) in 20 μl final volume with the following cycling conditions: 95 °C 10 min, 40 cycles 95 °C 10 s, and 58 °C 30 s for experiments shown in column 3 of table 3 and 95 °C 10 min, 40 cycles 95 °C 15 s, and 58 °C 60 s for all other experiments shown). All qRT-PCR amplifications included a melt curve step after cycling. Primers were constructed in conserved regions chosen based on the A. burtoni and N. brichardi (N. brichardi forms together with the expression investigated for N. pulcher, the N. brichardi/pulcher species complex, Duftner et al. 2007) genome sequences and sequenced amplicons of the corresponding regions for O. ventralis to guarantee equal binding efficiencies between species.

All used primers (see supplementary table S5 Supplementary Material online for primer sequences) were initially validated on serial dilutions of factor 5 of A. burtoni juvenile whole body mixed cDNA (34 individuals of 7.8–24 mm standard length, 3 weeks to 2 months age, laboratory strain, equimolar pooled RNA) and for N. brichardi/pulcher specific primers on serial dilutions of factor 5 of N. brichardi juvenile whole body mixed cDNA (six individuals pooled, 18.4–24.6 mm standard length, 3 months age, kindly provided by H. Gante, University of Basel) starting with 10 ng/μl DNA final concentration. qRT-PCR efficiencies (E) were calculated for each reaction from the slope of the standard curve using the equation \( E = 10^{1-1/slope} \) as implemented in the StepOnePlus software (LifeTechnologies) with an efficiency of 2 being equal to 100% (\( E = [10^{1-1/slope}] - 1 \) \times 100) and an indicator of a robust assay.

Expression was measured as comparative C\(_T\) experiments using rpl7 (ribosomal protein L7) as control gene and gills as reference tissue. Rpl7 was chosen after an initial comparison with two other presumably house keeping genes (df1a and beta-actin). rpl7 showed the most consistent expression between samples (i.e., smallest C\(_T\) range) and has been used previously in fish qRT-PCR experiments (Böhne et al. 2010). Gill tissue was chosen as reference for the delta-delta-C\(_T\) method, because it did not show differences between sexes and expressed the genes of interest. Furthermore, expression in gills was most consistent between species and in comparison to liver and anal fin tissue. For quantification, data were first analyzed using the delta-delta-C\(_T\) method (Livak and Schmittgen 2001) and custom R-scripts. As lowest threshold for expression we applied 37 cycles. As a second, and this time reference tissue independent, analysis method, mean normalized gene expression was calculated as described in (Simon 2003) again using rpl7 as control gene.

All experiments were carried out with three technical replicates. The details of all experiments including used samples, raw data, graphs, and statistics are given in supplementary materials S2 and S3, Supplementary Material online.

Cyp19a1A Gene Sequencing
A fragment including the entire coding region was amplified by PCR on genomic DNA of 28 cichlid species from LT. The species cover 12 tribes of the cichlid assemblage of this lake; the number of species sequenced per tribe was chosen according to species-richness of the tribe. When available, a male and a female specimen were sequenced (supplementary table S4, Supplementary Material online).

Long-range PCR reactions were carried out using Phusion Master Mix (New England Biolabs) in 40 μl reaction volume on 50 ng genomic DNA with the primers for 5'-TGAACTAGGGCCTGTAAACCCAGAG-3' and rev 5'-AAGACTTTTGCACACACAGCTAGG-3' (cycling conditions: 98 °C 30 s followed by 35 cycles of 98 °C for 10 s, 64 °C for 30 s, 72 °C for 4 min and with a final elongation of 10 min at 72 °C, primers ordered from Microsynth).

PCR products were visualized using SYBRSafe DNA Gel Stain (LifeTechnologies) on a 1.5% agarose gel. Thirty microliters of PCR product was gel purified using QIAquick Gel Extraction Kit (Qiagen). 100 ng of purified PCR product was used for barcoded library preparations using NEBNextFast DNA fragmentation kit (fragmentation time 28 min) and Ion Xpress Plus Fragment Library kit and Ion Xpress Barcode Adapters (LifeTechnologies) according to manufacturer’s recommendation. Libraries were pooled after barcoding for subsequent analysis. Quantification prior to sequencing was done with the Ion Library Quantitation Kit (LifeTechnologies) on a StepOnePlus Real-Time system (LifeTechnologies).

Libraries were sequenced on three 314 chips on an Ion PGM Sequencer (LifeTechnologies). Demultiplexed sequences
were loaded into Geneious (Biomatters Ltd.) and mapped to the reference amplicon reconstructed from the *A. burtoni* genome (BROAD) as closest available relative for all self-sequenced species. Alignments were corrected manually. The subsequent sequence analyses were performed in Geneious using MUSCLE (Edgar 2004) as alignment algorithm. Alignments per species were combined to construct one consensus sequence per species with 75% identical threshold for ambiguous sites. Sequences for *O. niloticus*, *P. nyererei*, and *N. brichardi* were retrieved from the cichlid genomes (BROAD) and included in the alignment.

The final alignment based on the grouped nucleotide sequences was trimmed and used for the reconstruction of a phylogenetic tree with MrBayes (5 million generations, default settings, version 3.2.1, Ronquist et al. 2012) under the GTR + I + gamma model as determined with jModelTest2 (Darriba et al. 2012). The obtained tree topology was in accordance with a maximum likelihood tree (1,000 bootstraps) obtained with PhyML (Guindon et al. 2010) (data not shown). An alignment covering only the coding region from start to stop codon according to the Nile tilapia sequence available in Ensembl (ENSONIT00000000198) was used for selection analysis.

**In Silico Sequence Analysis**

Sequences of the candidate genes were identified using the reciprocal best hit BLAST method with annotated sequences from the medaka (*O. latipes*) genome assembly (http://www.ensembl.org/Oryzias_latipes, last accessed July 23, 2013) as queries against the Nile tilapia (*Oreochromis niloticus*) EST data available in Genbank (www.ncbi.nlm.nih.gov, last accessed July 23, 2013) and, upon availability, against the Nile tilapia genome (http://wwwensembl.org/Oreochromis_niloticus, last accessed July 23, 2013). The retrieved tilapia sequences were used as queries in a local BLASTn search against the cichlid fish genomes (BROAD) as references. If the reciprocal best hit method was not conclusive, especially to distinguish cichlid/teleost copies of the same original tetrapod gene, phylogenetic analysis were carried out (data not shown) including more teleost sequences to characterize the A- and B-copies according to the annotation used for medaka in Ensembl (wwwensembl.org, last accessed July 23, 2013).

In-frame alignments of the coding sequences from start codon (included) to stop codon (excluded for Selecton analysis, see below) were performed with CodonCodeAligner (CodonCode Corporation). All alignments were checked and corrected manually with respect to exon–intron boundaries and in-frame indels.

Overall Ka/Ks estimates were calculated using KaKs_calculator (Zhang et al. 2006) with the MA method for each pairwise comparison to the Nile tilapia sequence over the entire length (supplementary tables S2 and S3, Supplementary Material online).

Site-wise Ka/Ks estimates were performed with Selecton (Doron-Faigenboim et al. 2005; Stern et al. 2007) using the tilapia sequence as query. The M8 (M8, beta + w ≥ 1) model enabled for positive selection was used (Yang et al. 2000).

When positively selected sites were detected under this model, it was tested against the null model (no positive selection, M8a, beta + w = 1, Swanson et al. 2003).

The distribution of positively selected sites in *cyp19a1A* along the phylogeny was tested with HyPhy (Pond et al. 2005) and codeml, which is part of the PAML package (Yang 2007). To test for branch-specific signs of positive selection we run the Branch-site REL model (Kosakovsky Pond et al. 2011) as implemented in HyPhy on www.Datamonkey.org (last accessed July 23, 2013) (Pond and Frost 2005; Delport et al. 2010). To test for site- and branch-specific signs of selection, we used codeml with the branch-site model and a test for positive selection (Yang and Nielsen 2002; Yang et al. 2005; Zhang et al. 2005) to test different clades against all other clades of the phylogeny (fig. 2).

**Promoter Analysis**

Promoter analyses were carried out on the flanking regions of the *cyp19a1A* and *cyp19a1B* genes exported from all available teleost genomes on www.ensembl.org (last accessed July 23, 2013) (Release 69, October 2012) and on the sequences available for *A. burtoni*, *P. nyererei*, and *N. brichardi* (BROAD). Annotation files were exported from Ensembl or, whenever necessary, created manually for the nonannotated genomes. The annotation for *glDNB* on the Nile tilapia sequence was added manually, as this gene is not annotated in the current genome release. Annotation was done using the coding sequence present in Genbank (Accession number XM_003443944). Alignments for all teleost promoters were performed using mVISTA (Mayor et al. 2000; Frazer et al. 2004) with Shuffle-LAGAN as alignment algorithm. Alignments including only cichlid sequences were performed with MUSCLE (Edgar 2004) as implemented in Geneious (Biomatters Ltd). Putative transcription factor binding sites were annotated according to (Tong and Chung 2003; Yoshiura et al. 2003; Chang et al. 2005; Ohmuro-Matsuyama et al. 2007; Wang et al. 2007; Dietel et al. 2010) and with the help of MathInspector as implemented in the Genomatix software suite V2.6.2 (Genomatix Software GmbH) with a matrix similarity cutoff of >0.9. We selected only transcription factors that fell in the categories brain and/or CNSs and ovary, testis, and germ cells. Note that putative transcription factor binding sites that are described as “transcription factor –F” (e.g., Hbox-F) include more than three members of the same family predicted to bind at the same site and are thus abbreviated with –F for family.

**Supplementary Material**

Supplementary material S1–S3, tables S1–S5, and figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

A.B., C.H., and W.S. designed the study and wrote the manuscript. A.B. performed sequence analysis. A.B. and C.H. performed qRT-PCR expression analysis. A.B. and N.B. sequenced the *cyp19a1A* locus. The authors thank A. Indermaur, F. Kim,
B. Meyer, B. Egger, and A. Theis for help with sampling; B. Meyer and L. Baldo for helpful discussion on data analysis; F. Pferrer for helpful comments on cyp19a1, the Broad Institute as well as the Cichlid Genome Consortium for providing access to cichlid genomes and transcriptomes, and two anonymous referees and the associate editor for valuable comments on a previous version of this manuscript. This work was supported by the Forschungsfond Universität Basel, the Volkswagenstiftung (Postdoctoral Fellowship Evolutionary Biology, grant number 86 031), and the DAAD (German academic exchange service, grant number D/10/52114) to A.B.; travel grant of the Swiss Academy of Sciences (SZG-Reisestipendium) to C.H.; and the European Research Council (ERC, Starting Grant “INTERGENADAPT”) and the Swiss National Science Foundation to W.S. This research was done at the Zoological Institute, University of Basel, Basel, Switzerland.

References


Mank JE, Promislow DEL, Avise JC. 2006. Evolution of alternative sex-

Ijiri S, Kustan JM, Maruska KP, Fernald RD . 2011. Subordinate male cichlids in development, immunity, circadian rhythm, and cellular metabol-


Liu J-F, Guiguen Y, Liu S-J. 2009. Aromatase (P450arom) and 11β-hydroxylase (P45011β) genes are differentially expressed during the sex change process of the protogynous rice field eel, Monopterus albus. Fish Physiol Biochem. 35:511–518.


Marin I, Baker SB. 1998. The evolutionary dynamics of sex determina-


Matsuda M, Nagahama Y. 2008. Sexual dimorphic expression of \(cyp19a1\) in development, immunity, circadian rhythm, and cellular metab-


Merchant-Larios H, Diaz-Hernández V. 2013. Environmental sex deter-


Toda K, Shizuta Y. 1993. Molecular cloning of a cDNA showing alter-
ting in a color gene involved in the formation of the charac-
teristic egg-dummies of male haplochromine cichlid fishes. BMC Biol. 5:51.

Salzburger W, Braasch I, Meyer A. 2007. Adaptive sequence evolu-

Salzburger W, Mack T, Verheyen S, Meyer A. 2005. Out of Tanganyika:
genesis, explosive speciation, key-innovations and phylogeography
of the haplochromine cichlid fishes. BMC Evol Biol. 5:17.

Phylogeny of the Lake Tanganyika cichlid species flock and its rela-
tion to the Central and East African Haplochromine cichlid fish

Scholz S, Klüver N. 2009. Effects of endocrine disruptors on sexual, gon-
nadal development in fish. Sex Dev. 3:136–151.

Schulz RW, de França LR, Lareyre JJ, LeGac F, Chiarini-Garcia H, Nobrega


Cloning and expression of R-Spondin1 in different vertebrates
suggests a conserved role in ovarian development. BMC Dev Biol.
8:72.


Sobrinho I, de Brito R. 2010. Evidence for positive selection in the gene
fruitless in Anostepha fruit flies. BMC Evol Biol. 10:293.

Soma KK, Francis RC, Wingfield JC, Fernald RD. 1996. Androgen regu-
lation of hypothalamic neurons containing gonadotropin-releasing
hormone in a cichlid fish: integration with social cues. Horm Behav.
30:216–226.

Sorhannus U, Kosakovsky Pond S. 2006. Evidence for positive selection
on a sexual reproduction gene in the diatom genus Thalassiosira

Stern A, Doron-Faigenboim A, Erez E, Mertz E, Bacharach E, Pupko T.
2007. Selecton 2007: advanced models for detecting positive and
purifying selection using a Bayesian inference approach. Nucleic
Acids Res. 35:W506–W511.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in

Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Kobayashi T, Cabau C,
evolved into the master sex-determining gene in rainbow trout,

Yoshihara K, Terai Y, Mizoi S, et al. (11 co-authors). 2011. B chromo-
somes have a functional effect on female sex determination in Lake

Yoshiura Y, Senthilkumar B, Watanabe M, Oba Y, Kobayashi T,
Nagahama Y. 2003. Synergistic expression of Ad4BP/SF-1 and cyto-
chrome P-450 aromatase (ovarian type) in the ovary of Nile tilapia,
Oreochromis niloticus, during vitellogenesis suggests transcriptional

likelihood method for detecting positive selection at the molecular

differential expression of R-spondin1 in zebrafish (Danio rerio)

calculating Ka and Ks through model selection and model averaging.
Genomics Proteomics Bioinformatics. 4:259–263.

Similar gene structure of two Sox9a genes and their expression
patterns during gonadal differentiation in a teleost fish, rice field