Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes

Athimed El Taher[1], Fabrizia Ronco[1], Michael Matschiner[1,2,3], Walter Salzburger[1], Astrid Böhne[1,4*]

[1]Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland

[2]Department of Palaeontology and Museum, University of Zurich, Zurich, Switzerland.

[3]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway.

[4]Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany

*e-mail: a.boehne@leibniz-zfmk.de

Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes

Abstract

Sex is a fundamental trait that is determined, depending on the species, by different environmental and/or genetic factors, including various types of sex chromosomes. However, while the functioning and evolution of sex chromosomes have been explored in species scattered across the eukaryotic tree of life, little is known about tempo and mode of sex chromosome evolution in closely related species. Here, we examined the dynamics of sex chromosome evolution in an archetypical example of adaptive radiation, the cichlid fishes of African Lake Tanganyika. Through inspection of male and female genomes of 244 cichlid taxa and the analysis of transcriptomes from 66 taxa, we identified the sex chromosomes in 79 taxa, involving 12 different linkage groups. We estimated that Tanganyikan cichlids have the highest rates of sex chromosome turnover and heterogamety transitions known to date. That the recruitments of chromosomes as sex chromosomes is not at random and that some chromosomes have convergently emerged as sex chromosomes in cichlids, provides empirical support for the limited options hypothesis of sex chromosome evolution.

Introduction

Sex chromosomes – referred to as Z and W in female and X and Y in male heterogametic sex determination (SD) systems – define, through their properties and combinations, the sex of an individual[1]. The evolutionary trajectories of sex chromosomes differ from those of autosomes: Due to the restriction of one of the two gametologs to one sex (W to females in ZW, Y to males in XY SD systems), their sex-specific inheritance (e.g., XY-fathers pass on their X exclusively to daughters and their Y to sons), and their reduced levels of recombination, sex chromosomes accumulate mutations more rapidly than autosomes, potentially leading to accelerated functional evolution[2,3].

The functioning of a chromosome as sex chromosome is often short-lived in evolutionary time scales. This relative instability of sex chromosomes is due to turnovers (i.e., changes of the actual chromosome pair in use as sex chromosome) caused by a new sex-determining mutation on a previously autosomal locus[4] or the translocation of the ancestral SD gene to another chromosome (e.g.,[5]). Sex chromosome turnovers may be accompanied by a transition in heterogamety[6]. Heterogamety can also change without transition to another chromosome, which in this case likely involves a turnover of, or a mutation within, the actual SD gene[7].

The assumed major driving forces underlying turnovers of sex chromosomes are deleterious mutational load[8] [9], sexually antagonistic loci linked to a newly invading SD gene[10,11], selection on restoring sex-ratios[12] and genetic drift[6,13,14]. These drivers are predicted to differ in their respective outcome: turnovers induced by mutational load tend to preserve heterogamety[8,9], while sexually antagonistic selection driven turnovers more readily induce a change of heterogamety[10].

Finally, the gene repertoire on previously existing sex chromosomes can also be extended by chromosomal fusion with an autosome, which then becomes sex-linked itself, leading to the formation of a neo-sex chromosome[15].

The frequency of occurrence of these different paths of sex chromosome evolution varies substantially across animal clades[16]. For example, in some vertebrates (mammals and birds) the same sex chromosomes are shared across the entire class[1] (but see[12]). Models[13] as well as empirical observations[15] suggest that sex chromosomes such as those of mammals and (most) birds have differentiated to a degree that makes turnovers unlikely; these sex chromosomes are in an "evolutionary trap"[17]. This is because a sex chromosome turnover requires the fixation of one of the previous sex chromosomes as an autosome, which becomes more deleterious and thus unlikely the more specialised and/or degenerated the sex chromosomes are[18]. In other vertebrate lineages (amphibians, reptiles, and fish), frequent turnover events and continued recombination led to many different and mostly non-degenerated (homomorphic) sex chromosomes (e.g.[19,20]).

As of to date, empirical studies on the dynamics of sex chromosome evolution are limited and scattered across different taxa. In an amphibian system with rapid rates of sex chromosome turnover, the true frogs Ranidae, mutational load seems to be the major driving force of sex chromosome turnover[20]. In geckos, a high rate of sex chromosome changes with heterogametic transitions potentially supports sexual antagonism as a key mechanism of these changes[10,11,21]. However, an in-depth analysis of sex chromosome turnovers over short evolutionary timescales and with a broad taxon sampling is currently lacking (reviewed in[16]).

Here, we examined sex chromosome evolution in an archetypical example of rapid organismal diversification, the adaptive radiation of cichlid fishes in African Lake Tanganyika[22] (LT). Teleost fishes are generally known for their species richness[23], but cichlids stand out in this clade on the basis of the "explosive" character of several of their

adaptive radiations, giving rise to a total estimated number of over 3000 species[23]. Rapid speciation in adaptive radiations is usually attributed to ecological specialisation and thus diversification in eco-morphological traits[24]. Here, we were interested if sex determination is keeping pace with other traits in cichlids by determining the diversity of SD systems and by investigating the dynamics of sex chromosome turnover across the entire LT cichlid radiation. The available data from about 30 African cichlid species (reviewed in [25,26]) suggest that sex chromosomes are not conserved in this group with both simple and polygenic sex determination systems being known from the different species investigated. An emerging picture is that certain chromosomes have recurrently been recruited as sex chromosomes in cichlids. However, available studies supporting the convergent recruitment of sex chromosomes have been based on cichlid species belonging to different lineages and the observed patterns have rarely been assessed in a phylogenetic framework, which makes inferences about rates of evolution as well as of convergence *versus* common ancestry difficult (but see[25]). Importantly, as of yet, no inclusive analysis of sex chromosome evolution exists for a cichlid adaptive radiation (nor for radiations in other fish families).

In this study, we inspected genomic and transcriptomic information from 229 Lake Tanganyika cichlid taxa as well as 18 cichlid species belonging to the Haplochromini and Lamprologini lineages, phylogenetically nested within the LT radiation[22,24,27] for signatures of sex chromosomes. Based on this nearly complete taxon sampling of the LT radiation and an available genome-wide phylogenetic hypothesis[24], we estimated the amount and direction of sex chromosome turnovers in a young species flock. This allowed us to test for a possible contribution of sexual antagonism in the evolution of sex chromosomes in LT cichlids. Sexual antagonism has been suggested as a driving force of sex chromosome turnovers in sexually dimorphic cichlids of the Lake Malawi radiation[28,29]. However, unlike the cichlid adaptive radiation in Lake Malawi, which is composed solely of cichlids of the

Haplochromini lineage, the endemic LT cichlid assemblage consists of 16 cichlid lineages (corresponding to the taxonomic assignment into tribes)[22], some of which are sexually dimorphic while others are not.

To assess the dynamics of sex chromosome turnover in fish on a larger scale, we expanded our comparative analyses to other fish systems as well. In particular, we investigated sex chromosome turnovers in ricefishes, another model system for the evolution of sex chromosomes[30].

Sex differences in the recombination rate could contribute to the differentiation of sex chromosomes[31]. Unlike in the extremely heterochiasmic frogs of the family Ranidae[20] and some fish model organisms[32,33], recombination rate along chromosomes do not systematically nor drastically differ between the sexes in cichlids[33,34]. In ricefishes, reduced rates of recombination have been linked to maleness in some species[35] but this does not seem to be a general pattern in this group of fishes[36]. Ricefishes and cichlids may hence have differing, probably lower, rates of sex chromosome degeneration than the heterochiasmic frogs, in which mutational load resulting from sex chromosome degeneration caused by suppressed recombination mainly drives turnover. In general, we expect fewer – if any – cichlid species to be in the "evolutionary trap" of degenerated sex chromosomes and more sex chromosome turnovers caused by sexual antagonism than mutational load[16]. Finally, with the identification of sex chromosomes in genetically very closely related species we pave the way for the subsequent identification of sex-determining genes and/or the causal mutations leading to sex chromosome turnover.

Results

Sex chromosomes in LT cichlids

To identify sex chromosomes in LT cichlids, we screened male and female genomes of 244 taxa[24] as well as six transcriptomes of 66 taxa[27] for sex-linked regions, applying three complementary approaches: genome-wide association study (GWAS on the genomic data, approach 1, see Methods), identification of sex-specific SNPs in the genomic data (approach 2) and allele frequency difference tests on the transcriptome data (approach 3). Genomic locations of sex chromosomes refer to linkage groups (LGs) of the used reference genome of a phylogenetically equidistant outgroup to the cichlid species of the LT radiation, the Nile tilapia (*Oreochromis niloticus*). To estimate sex chromosome turnovers (see below), we used two different datasets, a "permissive dataset" including all sex chromosomes identified with approaches 1-3 and a "stringent dataset" excluding sex chromosomes that had support only in approach 2, i.e. lacking transcriptome data or support for small sex-linked and potentially non-expressed regions in the transcriptome data and occurring in tribes too small to be investigated with approach 1.

By combing the result of approaches 1-3, we detected sex chromosomes in 78 endemic LT cichlid taxa as well as in the riverine Haplochromini *Orthochromis indermauri* (Figs. 1 and 2; Supplementary Tables 1-4, Supplementary Figs. 1-4).

Approach 1 (GWAS), which was applied to the larger (that is, more species-rich) cichlid tribes from LT only, revealed the presence of sex chromosomes shared among several species of their respective tribes. We thus identified an XY SD system on LG19 in Haplochromini/Tropheini, (thereby confirming an XY system previously known from one species in this clade, *Tropheus* sp. "black"[37]), an XY and ZW system on LG05 in Cyprichromini (confirming a ZW system previously described in *Cyprichromis leptosoma*[37]), and an XY system on LG15 and LG20 in Lamprologini. Surprisingly, with this approach, we

did not detect a shared sex chromosome within the second-most species-rich cichlid tribe of LT, Ectodini.

Approach 2, the inspection of the genomes within tribes for an accumulation of sex-specific SNPs (i.e., XY or ZW SNPs) and outlier regions thereof revealed two XY systems on LG19 within Haplochromini/Tropheini, one covering the first ~22 Mb of LG19 (in the genus *Tropheus* and in *O. indermauri*) and a second one located at the end of LG19 co-occurring with XY SNPs at the beginning of LG05 (in the second Tropheini clade grouping all genera but *Tropheus*, that is, 15 species belonging to the genera *Pseudosimochromis*, *Petrochromis,* and *Interochromis,* Fig. 1, Supplementary Fig. 4). We also recovered the narrow sex-linked region on LG20 detected with GWAS in Lamprologini, corroborating the effectiveness of this approach.

As in approach 1, we did not detect a sex-differentiated region shared across species in Ectodini with approach 2.

When applied to the smaller tribes, approach 2 revealed rather narrow but clear outlier regions that were shared between subsets of species in Benthochromini (XY LG10, two species), Trematocarini (ZW LG04, two species) and Cyphotilapiini (XY LG16, two species) as well as in all members of the Eretmodini (XY, LG07 and LG10). We also detected a less pronounced and smaller ZW-outlier region on LG09 in the same two Cyphotilapiini species, a pattern potentially explained by variation in X-linked markers across the different species while simultaneously lacking sites on the Y (hemizygosity in males). Due to this uncertainty and the stronger signal on LG16, we excluded the ZW LG09 system of Cyphotilapiini in the subsequent analysis (in the permissive and stringent datasets). Within Bathybatini and Perissodini, we identified a chromosome-wide increase of ZW SNPs on LG07 and of XY SNPs on LG19, respectively, which however failed our thresholds for the permissive dataset (see Methods). Upon inspection of XY-ZW differences per species within these tribes

(Supplementary Fig. 5), this pattern turned out to be caused by only one species in each tribe (*Hemibates stenosoma* and *Plecodus paradoxus*), which both showed signs of a differentiated sex chromosome across the entire length of the respective LG. Note that a ZW system on LG07 has previously been described in *H. stenosoma*[37] and that we could also confirm the XY system spanning the full length of LG19 with approach 3 (see below) in *P. paradoxus* as well as in another Perissodini species, *Plecodus straeleni*, for which we did not have whole genome data of both sexes. We hence included the sex chromosomes of these species in all the down-stream analyses.

Approach 3, the species-specific investigations of sex-specific SNPs based on replicate transcriptome data, confirmed all sex-differentiated regions that spanned larger chromosomal regions (i.e., the XY systems on LG19 and LG05/LG19 in Haplochromini/Tropheini, the XY and ZW systems on LG05 in Cyprichromini, the XY system in Eretmodini). With this approach we also detected a ZW system on LG15 in two Ectodini species (*Xenotilapia boulengeri* and *Enantiopus melanogenys*). Approach 3 further permitted us to identify sex-linked LGs unique to eight additional species and not shared with their respective sister species. For example, we detected an XY system on LG23 in the Ectodini *Callochromis pleurospilus,* and a ZW system on LG20 in the Benthochromini *Benthochromis horii* (Fig. 1, Supplementary Table 2). In another four species, the RNA data showed a significant overrepresentation of either XY or ZW-SNP windows that could not be placed unambiguously on reference LGs (Fig. 1, Supplementary Table 2).

Overall, in nine of the 13 investigated tribes of the cichlid radiation in LT, several species shared the same SD system (chromosomal region and heterogametic type); however, we did not find a shared sex chromosome across members of different tribes.

We detected sex linkage on 12 out of the 23 reference LGs (Fig. 2). Eight of these reference LGs were sex-linked in species belonging to different tribes (Fig. 2a). Two

196 reference LGs (LG14 and LG18) that we did not identify as sex chromosomes within any of
197 the LT cichlid species, have respectively been identified as sex chromosomes in labstrains
198 and one natural population of the haplochromine cichlid *Astatotilapia burtoni* (occurring in
199 LT and affluent rivers)[38,39]. In addition to the published data for *A. burtoni*, we also included
200 the XY LG07 sex chromosome of *Pseudocrenilabrus philander* (Lake Chila)[25] in our
201 subsequent analyses, a haplochromine species integrated in the phylogenetic reconstruction
202 used here (Fig. 1) but represented with a single individual in the WGS dataset.

203 In 62 of the LT cichlids (79.5% of the LT species with a sex chromosomal signal), the
204 sex linkage was compatible with an XY system (Fig. 2b).

205

206 **Sex chromosome evolution in LT cichlids**

207 Next, to determine when particular sex chromosomes emerged and to trace heterogamety
208 transitions in the course of the cichlid adaptive radiation in LT, we performed ancestral state
209 reconstructions along a time-calibrated species tree[24]. We performed these analyses on the
210 permissive as well as on the stringent dataset.

211 We reconstructed 30 sex chromosome turnovers in the radiation and LG04 as the
212 likely sex chromosome at its root (permissive dataset; 27 turnover events with the stringent
213 dataset), translating into an estimated rate of 0.186 turnovers per Myr (Fig. 1, Supplementary
214 Fig. 7, permissive dataset; turnover rate stringent dataset 0.187). On average, we therefore
215 expect one sex chromosome turnover event between two species that diverged ~2.7 Ma. This
216 rate-estimate was ten times higher than the one that we calculated for ricefishes
217 (Adrianichthyidae; 0.02 transitions per Myr; Supplementary Fig. 9; 19 species investigated,
218 see Methods).

219 The distribution of sex chromosomes in LT cichlids differed from random
220 expectations (Fig. 2d). There was no correlation between the size of a reference LG, the

number of genes on a reference LG, or the number of known sex-candidate genes on a reference LG and the frequency at which these LGs appeared as sex chromosome in LT cichlids (Fig. 2d). Our findings thus corroborate that SD is a rapidly and non-randomly evolving trait in cichlids. We further found that the number of turnovers in a tribe is correlated with its species richness (Fig. 2c, pGLS: P=0.0043, coeff=0.039), suggesting that the turnover rate has been relatively constant throughout the radiation.

Our heterogamety reconstructions further suggested that XY is the most likely ancestral state in the cichlid adaptive radiation in LT (Supplementary Fig. 8). Subsequently, 11 transitions occurred from XY to ZW (permissive dataset; 11 towards ZW and one towards XY in the stringent dataset). Heterogamety transitions are predicted to have a directional bias towards new dominant sex chromosomes[13], suggesting that in cichlids from LT – just like in cichlids from Lake Malawi[28,29] – the new W chromosomes are dominant over ancestral Ys.

When integrating over the reconstructed transitions in heterogamety and sex chromosomes, we found heterogamety changes that were uncoupled from turnovers in LGs and that were hence not captured in our rate estimate of sex chromosome turnover: A transitions from XY to ZW was detected on LG05 in Cyprichromini and on LG04 in Trematocarini and in Bathybatini (*H. stenosoma*) (Fig. 1; Supplementary Figs. 7-8).

The overlap of heterogametic and sex chromosome turnovers also showed that the majority (23 *versus* seven) of the observed sex chromosome turnovers in LT cichlids preserved the heterogametic state, suggesting that mutational load, predicted to preserve heterogametic state[20], might be a major driver of sex chromosome turnover in cichlids as well. The transitions with a change in heterogamety offer the possibility to investigate the actual potential of sexual antagonistic selection between very young species (the divergence time between e.g. *Paracyprichromis* and *Cyprichromis* between which a turnover has occurred is ~3.8 Ma). The heterogametic status of the four species for which we could not

identify the sex-linked LG (see above) lead to additional heterogamety transitions not captured in the sex chromosome turnover rate.

Overall, the heterogamety transition rate in LT cichlids (0.028 transitions per Myr permissive dataset; 0.031 per Myr stringent dataset) was about four times higher than in ricefishes (0.007 transitions per Myr; ancestral state: ZW). To explore heterogamety changes on a greater taxonomic scale, we also calculated heterogamety transition rates for all ray-finned fishes available in the Tree of Sex database (http://www.treeofsex.org/) that were included in a recent comprehensive phylogeny[40] (543 species analysed in total). Our analysis revealed a rate of 0.009 transitions per Myr for ray-finned fishes as a whole and identified XY as the ancestral state (Supplementary Table 6; Supplementary Fig. 9). Across the ray-finned fish phylogeny, transitions from XY to ZW were significantly younger than those from ZW to XY (Supplementary Fig. 9b, P=0.01428).

Chromosome fusions and novel sex chromosomes

Novel sex chromosomes can be created by chromosome fusions[41], which can contribute to reproductive isolation and eventually drive speciation over miss-segregation at meiosis, changes in recombination rates, novel physical combinations of loci, and changes in gene expression[42,43,44]. The here identified signatures of sex linkage suggest that several sex-chromosome/autosome fusions have occurred in the course of the cichlid radiation in LT or that autosome/autosome fusions occurred prior to the recruitment of these fused autosomes as sex chromosome (Fig. 1). The distribution of sex-differentiated genomic regions indicated a fusion (or large chromosomal translocations) between LG05 and LG19 in Haplochromini/Tropheini and between LG15 and LG20 in Lamprologini (Fig. 1, Supplementary Fig. 1). There was also some support for the previously described genome rearrangements in the tribe Eretmodini[45], which showed an increase of XY SNPs on several

LGs (Fig. 1, Supplementary Fig. 3). Additional sex-differentiated regions point to species-specific fusion events (e.g. LG11 and LG15 in *Gnathochromis pfefferi,* LG05 and LG13 in *C. leptosoma*). Our analyses also confirmed the reported sex linkage of LG04 as well as of LG07 in *H. stenosoma*[37] (Supplementary Fig. 5). Chromosome fusions have previously been implicated with the evolution of novel sex chromosomes in other taxa, as well as in the riverine haplochromine cichlid *A. burtoni*[38,39].

Interestingly, the so far only karyotypically investigated member of the tribe Tropheini, *Ctenochromis horei*, has a reduced number of chromosomes in a male and an unsexed individual (2n=40) compared to other Haplochromini, which usually feature 2n=42[45]. We did not detect the LG05/LG19 XY system found in many other Tropheini in *C. horei*. Hence, while the karyotype of this species indeed supports chromosomal fusions in the Haplochromini/Tropheini, this data cannot help to resolve when and how these events occurred. The data at hand are sparse but it might be that several large chromosomal rearrangements occurred before the novel chromosomes were recruited as sex chromosomes, making inferences of the driving forces of these fusions difficult.

Convergent evolution of sex chromosomes

On some LGs, the regions that showed sex linkage largely overlapped between members of different tribes (Supplementary Fig. 10), which can either be explained by common ancestry or by the independent (convergent) recruitment of those LGs as sex chromosome. In particular on LG19, several closely related species including six *Tropheus* species (Haplochromini/Tropheini), the riverine haplochromine *Orthochromis indermauri*, and *P. paradoxus,* and *P. straeleni* (Perissodini) feature an XY system in the same chromosomal region (Supplementary Fig. 10). Our ancestral state reconstruction suggested an independent origin of the LG19 SD system in Perissodini and *Tropheus*, in each case early in their tribe's

296 evolutionary history, and another independent origin in the terminal branch leading to *O.*
297 *indermauri* (Supplementary Fig. 7). Phylogenetic inference from Y- and X-haplotypes indeed
298 supported the independent evolution of LG19 as XY sex chromosome in Perissodini (Fig. 3),
299 while grouping together the Y-haplotypes of the *Tropheus* species and *O. indermauri*. This
300 suggests common ancestry of the XY system in the two haplochromine clades with an origin
301 either early on in haplochromines (implying several losses later in the evolution of this tribe;
302 likely because of this, such a scenario was not supported by ancestral state reconstruction) or
303 a later origin and inheritance of the sex chromosomal system in *Tropheus* and *O. indermauri*
304 from an extinct or unsampled taxon.

305 The XY sex chromosome system on LG05/LG19 found in the second clade of
306 Haplochromini/Tropheini (grouping all genera except *Tropheus*) must be derived from
307 another independent evolutionary event, since the regions on LG19 that show XY alleles in
308 the two Haplochromini/Tropheini clades are not overlapping (Supplementary Fig. 10) and
309 also do not group together in the phylogenetic tree of LG19 (Fig. 3). Other convergent cases
310 of sex chromosome recruitment supported by our ancestral state reconstruction involved
311 LG05 (in Cyprichromini and the haplochromine *A. burtoni*[38,39]) and LG07. LG07 has
312 independently been recruited as a sex chromosome in *H. stenosoma* (Bathybatini)[37], in
313 Eretmodini, in the lamprologine *Neolamprologus cylindricus* (Fig. 1 , Supplementary Fig. 7),
314 in several Lake Malawi cichlids[28,29] (Haplochromini), as well as in *P. philander*[25]
315 (Haplochromini), making it the most widespread sex-linked LG known in cichlids to date.

316

317 **Sex chromosome differentiation**

318 A comparison of the proportion of sex-specific sites on the different sex-linked LGs revealed
319 a continuum of sex chromosome differentiation in the cichlid adaptive radiation in LT (Fig.
320 4, Supplementary Fig. 10), ranging from a few kb (LG20 in Lamprologini) to almost full

chromosomal length (LG05 in Cyprichromini, LG19 in *Tropheus* and Perissodini). Varying lengths of sex-differentiated regions were even detected within the same LG when being used as sex chromosome by different lineages (e.g., the sex-differentiated region on LG05 spans only 8 Mb in Tropheini versus the entire LG in Cyprichromini).

The canonical model of sex chromosome evolution predicts progressing differentiation of sex chromosomes with time[2]. Contrastingly, we found no correlation between the estimated age of origin of a sex chromosome and its degree of differentiation (Fig. 4, pGLS: P=0.9049, coeff=0.0011). Some very young sex chromosomes showed signs of differentiation, i.e., sex-specific sites, along almost the full length of an LG, suggesting wide-spread suppression of recombination along their sex chromosomes.

The amount of sex-specific sequences in cichlids of the LT radiation inferred from a "subtraction pipeline" (approach 4) of expressed male and female sequences[46] was somewhat higher in XY than in ZW systems (P=0.058), but not when accounting for phylogenetic signal (phylogenetic ANOVA: P=0.2, Supplementary Fig. 11a). The observed pattern could suggest that Y-chromosomal genes are more highly expressed in adult (gonadal) tissues of closely related species than W-chromosomal genes. The difference in the amount of sex-specific contigs between XY and ZW systems does not reflect a difference in the degree of differentiation between the two heterogamety types (Supplementary Fig. 11b, phylogenetic ANOVA: P=1).

Candidate genes of sex determination in LT cichlids

Our inspection of known genes implicated in SD revealed that such genes were located on all LGs, including those, for which no sex linkage was detected, with no particular overrepresentation on certain LGs (Supplementary Fig. 12). The regions with the strongest signal for being sex-differentiated did not contain any of these genes (Supplementary Table

346 2). However, through the inspections of the regions with the strongest signs of sex linkage we
347 identified promising new candidate genes for SD in these regions, such as *tox2* in
348 Lamprologini, a HMG-box transcription factor involved in the hypothalamo-pituitary-
349 gonadal system. *Tox2* resembles the mammalian master SD gene *Sry*[47], which also codes for
350 an HMG-box protein.

351 In cichlids from lakes Malawi[28] and Victoria[48], sexually antagonistic colour genes
352 underlying a characteristic orange-blotched colour pattern are linked to SD genes, creating
353 the potential for speciation by sexual selection. In LT cichlids, which in general do not
354 feature the orange-blotched phenotypes, we did not find any obvious pattern in the
355 localisation of colour genes on sex-linked LGs (Supplementary Fig. 12).

Discussion

Here we report the identification of sex chromosomes for 79 taxa of cichlid fishes, most of which belonging to the cichlid adaptive radiation of LT, based on the analysis of whole-genome data from virtually all cichlid species of the radiation[22,24] and transcriptome data from a representative set of 66 taxa[27].

Models[13] and empirical observations[17] suggest that, beyond a certain degree of differentiation, sex chromosome turnover becomes unlikely. On the other hand, frequent turnovers, sex reversal, and continued recombination can contribute to counteract sex chromosome differentiation[9,49]. Our analyses revealed that, in the cichlid adaptive radiation of LT, sex chromosome turnovers seem to have occurred very frequently (Fig. 1), indicating that the cichlids' sex chromosomes have not (yet) reached a threshold preventing turnover, but that their sex chromosomes remain dynamic instead.

Sex chromosome recruitment in LT cichlids is non-random with respect to the recruited chromosome (Fig. 2), which is compatible with the hypothesis that certain chromosomes are more likely to become sex chromosomes than others[20,50]. This pattern becomes even more apparent when the LT cichlids are compared to other African cichlid species (Fig. 5), revealing that some LGs (in particular LG05, LG07, and LG19) emerged multiple times as sex chromosomes whereas others never appeared as such. This corroborates the hypothesis that particular chromosomes are preferentially[50] or even cyclically[9,49] recruited as sex chromosomes. Within LT cichlids, sex chromosome turnovers have likely been driven by a combination of mutational load and sexual antagonism. However, we detected a prevailing persistence of male heterogamety in LT cichlids, which is a common pattern in fish[51], suggesting a smaller role for sexual antagonism than previously postulated. The observed prevalence of XY systems is compatible with models of speciation driven by sexual

selection and sex-ratio distortion in cichlids that predict higher probabilities for the maintenance of male heterogamety[52].

The evolution of a novel sex determiner driven by linkage to a sexually antagonistic colour locus has previously been documented in Haplochromini from Lake Malawi[28], which are characterised by pronounced levels of sexual dimorphism. In our set of mostly riverine Haplochromini and Tropheini (the LT representatives of this clade) species, we found that a single sex chromosome system prevails, XY on LG05/LG19, which was probably established after a turnover from the rather strongly differentiated XY LG19 system present in the genus *Tropheus*. It thus appears that in the Tropheini, in which sexual dimorphism is much less pronounced (and even absent in some species) compared to the radiations of Haplochromini in lakes Malawi and Victoria, sexual antagonism does not play a prominent role as a driving force for sex chromosome turnover. Still, several Tropheini species seem to have lost the XY LG05/LG19 SD system and we were unable to detect most of the a new system that replaced it based on the available data, probably because the sex-linked chromosomal regions are rather small. These species will be particularly interesting to investigate further for the presence and the drivers of very young, novel sex chromosomes with a potential role for sexual antagonism impacting sex chromosome turnover. In addition, the observed cases of young homologous sex chromosome turnovers between closely related species (e.g. in the genus *Cyprichromis* on LG05 or in Trematocarini on LG04), which are indeed compatible with a role for sexual antagonism as driving force in cichlid sex chromosome evolution[10], open the route for further analysis of the causal mutations driving sex chromosome turnovers. Especially the presence of several ZW as well as XY species in Cyprichromini, potentially caused by a single transition event on the same chromosome, will allow in the future to trace which alleles have been affected by a heterogamety change. Such analyses may eventually

reveal the causal mutation(s) (supposedly within the SD gene) of the heterogamety turnover and the dominance relationships between XY and ZW systems.

We failed to detect sex-linked LGs in several of the LT cichlids, which certainly can, to some extent, be explained by our limited sample size per species and the limited power to detect small sex-specific regions, especially when using transcriptome data, as well as complex polygenic SD systems. While the limited per species sample size in the currently available data certainly leaves some SD regions undetected, we found it particularly intriguing that we could identify sex-linked regions only in three species across the second-most species-rich tribe, Ectodini. Some species of this tribe display an impressive level of sexual dimorphism, suggesting similar or even more pronounced sexual antagonistic selection compared to tribes such as the Haplochromini/Tropheini, which show relatively strongly differentiated sex chromosomes. It will thus be interesting to examine if Ectodini (and also members of other LT cichlid tribes) have very small if any sex-specific genome regions that our approaches failed to detect. This would further reveal if selective forces on SD differ within the radiation and if our sex chromosome turnover rate is indeed underestimating the true dynamics of sex chromosome change in LT cichlids.

Although the sex chromosomal status of many species remains to be identified in LT cichlids, our ancestral state reconstructions estimated a sex chromosome turnover rate ten times higher than the one in ricefishes, another group of fishes with an astonishing diversity of sex chromosomes, as well as the one published for true frogs, which was previously considered the fastest sex chromosome turnover rate known in vertebrates[20]. Note that extremely high numbers of SD system turnovers have also been described in geckos[21], but these have so far not been quantified using a comparable rate estimate.

Chromosome fusions could drive speciation through incompatibilities in genome structure[42,43,44] and cytogenetic analyses have indeed provided evidence for chromosome

fusion and fissions in some cichlid species[45]; however, their impact on cichlid diversification has not yet been assessed. Sex-chromosome/autosome fusions generating an odd number of chromosomes in one sex and leading to the formation of neo-sex chromosomes can be driven by altering expression of genes on the translocated chromosome[53,54], sexually antagonistic selection resolving conflict by restricting an antagonistic allele to a sex chromosome[55], or meiotic drive[56]. Until now, differences in chromosome number between male and female cichlids have not been reported, with the notable exception of copy number variations in female-determining B chromosomes in Lake Victoria and Lake Malawi cichlids[57,58]. For the limited number of cytogenetically investigated cichlid species, males and females have the same number of regular chromosomes and across cichlids in general, chromosome numbers differ little[45]. Overall, our analyses provide support for several large chromosomal rearrangements between the identified sex-linked LGs, suggesting that structural changes in the genome and the emergence of sex chromosomes are coupled in cichlids. The causality of this relationship remains to be investigated, just as the impact of genome rearrangements on reproductive isolation and eventually diversification. The available data on rearrangements are sparse, but it might be that several large chromosomal rearrangements occurred before the novel chromosomes were recruited as sex chromosomes, making inferences of the driving forces of these fusions difficult.

A next necessary step will be the identification of sex-determining genes and mutations causing sex chromosome turnover, facilitated by the close relatedness of the LT cichlids allowing the generation of interspecies hybrids and the opportunity to study multiple sex chromosome turnover events and directions, including the repetitive occurrence of heterogamety transitions without sex chromosome change. While the repeated recruitment of the same LG as sex chromosome indicates a particularly powerful core set of SD genes on the one hand, several transitions to novel LGs on the other hand question their dominance.

454 Although this could represent recycling of sex chromosomes to some extend, we lack the
455 molecular and most importantly functional evidence for any master SD gene in cichlids of the
456 Great Lakes radiations.

457 In conclusion, the estimated rapidity of sex chromosome turnover within (LT)
458 cichlids supports the hypothesis that SD mechanisms, albeit serving the unifying function of
459 sex determination, can be extremely labile. It remains to be tested if sex chromosome
460 turnovers simply keep pace with other fast evolving traits in radiating cichlid fishes or if they
461 are particularly frequent, potentially contributing to speciation.

Methods

Sequencing data

We used whole genome sequencing (WGS) data in the form of mapped reads in BAM files as well as variant call format from Ronco *et al.*[24] and raw transcriptome data from El Taher *et al.*[27] (see Supplementary Table 1 for details on species included and per species sample sizes). Based on a recent compilation of LT cichlid species[22], the WGS data included 225 taxa (174 described species with 4 of those represented with two local variants/populations each, and 47 undescribed species). The data further included 16 non-LT radiation haplochromine cichlid taxa (13 described species one of which represented with two local variants, and two undescribed species) and three riverine non-LT Lamprologini taxa (two described and one undescribed species) summing to a total of 244 taxa and 469 individual genomes (Supplementary Table 1). The transcriptome data that we used, were comprised of 66 taxa of LT cichlids (4 undescribed species, 61 described species one of which represented with two regional variants), with typically three males and three females per species (details are provided in Supplementary Table 1; 7 out of the 66 species had differing replicate numbers) and three tissues per individual (brain, gonad, gills; details on read numbers provided in Supplementary Table 3). In total, the dataset comprised 248 cichlid taxa.

Variant calling for WGS data

Mapped reads in coordinate-sorted BAM format were derived from Ronco *et al.*[24] (for mapping coverage statistics see Extended Data Table 1 in Ronco *et al.*[24]), which are based on mapping against the Nile tilapia (*Oreochromis niloticus*) genome (NBI RefSeq GCF_001858045.1_ASM185804v2). Unplaced scaffolds of the reference genome were concatenated lexicographically into an "UNPLACED" super chromosome.

In addition to the variant file containing all species derived from Ronco *et al.*[24], variants were called for each tribe separately with GATK's[59] (v.3.7) HaplotypeCaller (per individual and per chromosome) and GenotypeGVCFs (per 1 Mb window), and merged with GATK's CatVariants. Variants were further filtered with BCFtools (https://github.com/samtools/bcftools, v.1.6), applying the settings ReadPosRankSum<-0.5, MQRankSum<-0.5, FS<20.0, QD>2.0, MQ>20.0 and placing tribe-specific thresholds on minimum and maximum read depths to account for varying sample sizes (Bathybatini: 50-300; Benthochromini: 25-100; Cyphotilapiini: 50-200, Cyprichromini: 100-400; Ectodini: 250-1500; Eretmodini: 50-200; Tropheini/Haplochromini: 375-1375; Lamprologini: 700-3000; Limnochromini: 50-300; Trematocarini: 50-300). For the tribes Lamprologini, Tropheini/Haplochromini, Ectodini, and Limnochromini we further applied InbreedingCoeff>-0.8.

Indels were normalised with BCFtools's norm function, monomorphic sites were excluded, and SNPs around indels were masked depending on the size of the indel: for indels with a size of 1 bp, 2 bp were masked on both sides, and 3, 5, and 10 bp were masked for indels with sizes of 3 bp, 4-5 bp, and >5 bp, respectively. Individual genotypes were then masked with VCFtools[60] (v.0.1.14) if they had low quality (--minGQ 20) or depth (--minDP 4). Filtered variants were phased and missing genotypes were imputed with Beagle[61] (v.4.1). We then retained only biallelic sites that had no more than 50% missing data prior to phasing. For sites that were polymorphic but no individual had the reference genome allele, we set the first alternative allele as reference allele.

Approach 1 Tribe-wise association tests for sex on WGS data using GWAS

In total, we applied four approaches to identify sex-linked genomic regions (approach 1-4). Approaches 1 and 2 were applied on the tribe-level, the taxonomic rank above genus, but

511 below family and which, in the case of the LT radiation, groups species with divergence
512 times of 9.7-6.2My.

513 For approach 1 (Supplementary Figs. 1 and 2), the phased sets of variants for tribes
514 with at least 10 species (Lamprologini: sample size of 196 individuals representing 100
515 species; Ectodini: sample size of 81 individuals representing 40 species; Haplochromini
516 including the LT-endemic Tropheini: sample size of 99 individuals of 55 species; and
517 Cyprichromini: sample size of 21 individuals of 11 species) were each transformed into bim
518 and bed format with PLINK[62] (v.1.90b). Next, we ran association tests (GWAS) for sex on
519 these tribe-specific variant files using the univariate linear mixed model integrated in
520 GEMMA[63] (v.0.97) accounting for population stratification. After visual inspection of
521 GWAS results for potentially sex-associated regions (i.e., peaks or shifts of increased
522 significance), genotypes of the 100 most significantly with sex associated sites for
523 Haplochromini and Cyprichromini (broad signal for sex-association along the entire length of
524 LG19 and LG05, respectively) and of outlier SNPs (narrow peak regions on LG15, LG20 and
525 unplaced contigs comprising 51 SNPs with a -log10(*P*-value) > 3 ; extraction of the top 100
526 most significantly sex-associated SNPs revealed same clustering and no further sex-
527 associated region since those SNPs were scattered across the genome) were clustered and
528 visualised with the R package Pheatmap (v.1.0.12, https://cran.r-
529 project.org/web/packages/pheatmap/index.html) in R (v.3.5.2) and inspected for grouping by
530 sex.

531

532 **Approach 2 Tribe-wise tests for an accumulation of sex-specific SNPs**

533 Approach 2 was applied to tribes that contain more than a single species (Supplementary
534 Table 1 for all sample sizes): We here tested for an accumulation of sites with sex-specific
535 alleles, referred to as XY and ZW sites depending on the heterogametic sex (Supplementary

Figs. 3 and 4), under the assumption that a sex chromosomal region will show an accumulation of sex-specific alleles due to linkage caused by suppressed recombination. To this end, we subsetted the unphased, filtered sets of variants per tribe and included only species for which we had individuals of both sexes (Supplementary Table 1). We then removed indels and sites with more than 20% missing data and more than two alleles with VCFtools (v.0.1.14). The resulting files were loaded into R (v.3.5.0) with VCFR[64] (v.1.8.0.9). Sex-specific sites were classified as follows: Each variant site was recoded per species as a "nosex" site if the male and the female individual had the same genotype; as "noinfo" if one or both individuals had no genotype call; as "XY" if the male was heterozygous and the female homozygous; and as "ZW" if the female was heterozygous and the male homozygous. Next, we calculated within each tribe the sum of nosex, ZW, and XY sites in windows of 10 kb with a slide of 2 kb as well as the difference between XY and ZW sites per window. Next, we calculated the mean genomewide percentage of nosex, ZW, and XY sites over all windows and multiplied these values with the number of called sites per window to obtain expected values for XY, ZW, and nosex under the assumption that most variant sites across the genome show no particular sex difference. The expected values per window were compared to the observed values using a Fisher's Exact test with the exception of the tribe Lamprologini in which the large counts of sites per window rendered exact calculations with a Fisher's Exact test impossible so that we applied a Pearson's Chi-squared test. These tests will indicate windows that significantly differ from the genome-wide mean. We next inspected how windows differed from the genome-wide mean and designated and plotted a window with its corresponding *P*-value as (i) XY if the observed XY value was greater than the expected one, and the observed ZW value smaller than the expected one and, as (ii) ZW if the observed ZW value was greater than the expected and the observed XY value smaller than the expected one. If both, the observed XY and observed ZW values were larger than the

561 expected value, a window was declared ambiguous and not further considered. If both,
562 observed XY and observed ZW values were equal or smaller than the expected values, a
563 window was declared nosex and not considered further. Fisher's Exact test and Pearson's Chi-
564 squared *P*-values of XY and ZW windows were plotted together (on -log10 and log10 scale
565 respectively) and with an overlay of the calculated XY-ZW difference for each window
566 normalised by dividing the obtained value through the number of species analysed. The
567 obtained plots were inspected for the presence of LG-wide or regional shifts in XY-ZW
568 difference and outliers from the expected XY or ZW sites. We also calculated and visualised
569 the XY-ZW difference in each window at the species level. In order to assess a false-
570 discovery threshold, we permutated the observed data within each tribe 100 times by
571 randomly assigning the SNPs to the observed genomic positions. We recalculated the XY-
572 ZW difference per window as well as the expected values. We assessed from each
573 permutation the highest absolute XY-ZW difference of a window and the smallest *P*-value
574 for XY/ZW sites. The largest absolute XY-ZW difference normalised by species number
575 across all permutations within each tribe was then used as minimal threshold to define the
576 sex-linked regions in the observed data. The lowest *P*-value of all XY/ZW windows across
577 all permutations was -log10(*P*-value)=5.04 in Haplochromini/Tropheini, which corresponds
578 to a FDR ~4 after Bonferroni correction. To minimise the possibility of false positives after a
579 comparison of all observed data across all tribes, we finally retained only drastic XY/ZW-
580 outlier regions that in addition of exceeding the tribe-wise threshold of XY-ZW difference
581 derived from each tribes permutation, also had a -log10(*P*-value)>20 (corresponding to a
582 FDR of 2.30×10^{-26} after Bonferroni correction).

583 Upon a first inspection of sequence content of sex-linked regions, we noticed in the
584 empirical RNA and DNA data XY and ZW peaks in different tribes/species within the same
585 region on LG02 of the reference genome. This region (21.36 Mb - 21.93 Mb) is annotated

586 with 26 protocadherin tandem gene copies. We suspect that this array of similar genes
587 impacts mapping and hence masked this region from our sex chromosomal call. Furthermore,
588 it has previously been shown that LG03 is a sex chromosome in *Oreochromini* and that the
589 assembly quality of this region is poor due to presence of repetitive elements leading to
590 difficulties in the identification of sex-linked regions on this LG[65]. This is also reflected in
591 our data by an increase of missing data on this LG and hence less reliable SNP data. We
592 therefore also excluded outlier regions on LG03 as potential sex chromosome.

593 Since approach 2 was run on the tribal level we next needed to identify how many and
594 which species are responsible for the sex chromosome signals detected within a tribe, i.e.
595 identify the sex chromosomes on the species level from this approach. To this aim, we
596 visualised, per window, species level XY-ZW differences in the outlier regions and clustered
597 individual genotypes (with the possible values "homozygous reference", "homozygous
598 alternative", "heterozygous") with divisive hierarchical clustering based on a pairwise
599 dissimilarities matrix of Gower's distances calculated with the R package FSA (v.0.8.30)
600 (https://github.com/droglenc/FSA). Resulting dendrograms were inspected for grouping by
601 sex rather than species and boxplots of species-specific XY-ZW difference for support by
602 increased absolute XY-ZW difference. Due to the reduced sample size, final calls based on
603 the outlier regions were only made if at least two species within a tribe shared the same sex
604 linkage (Supplementary Table 2).
605
606 **Approach 3 Species-specific association tests for sex on transcriptome data**
607 For approach 3, species-specific association tests (Supplementary Fig. 6), we pooled tissue-
608 specific transcriptomes of brain, gonad and gills into one transcriptome per individual and
609 quality filtered and trimmed these with Trimmomatic[66] (v.0.33) with a 4 bp window size, a
610 required window quality of 15 and a minimum read length of 30 bp resulting in typically six

multi-tissue transcriptomes per species (Supplementary Table 3 for read numbers). The following analysis was then run per species: We performed reference-free *de novo* variant calling with KisSplice[67] (v.2.4.0) with settings "-s 1 -t 4 -u" and "--experimental". The identified SNPs were placed on the Nile tilapia genome assembly with STAR[68] (v.2.5.2a) (settings "--outFilterMultimapxNmax 1 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4"). The genome index used for this mapping was generated with the corresponding STAR parameters: --runMode genomeGenerate, --sjdbOverhang 124, --sjdbGTFfeatureExon exon and the genome annotation file (RefSeq GCF_001858045.1_ASM185804v2). Kiss2Reference[67] was used to classify KisSplice variants aligned to the Nile tilapia reference genome, and kissDE[67] (v.1.4.0) was applied to determine variants that differed between the two sexes. The resulting files were loaded into R. The KisSplice events were filtered with the following attributes: Only SNPs were kept; SNPs placed on mitochondrial DNA or on unplaced scaffolds of the reference genome were removed; only SNPs with significant *P*-values for an allele difference between the sexes ($P \leq 0.05$ after adjustment for multiple testing following the Benjamini and Hochberg method[69]) were retained. Significant SNPs were classified as (i) "XY" if they had zero read counts in all females and a minimum of one count in at least two males and as (ii) "ZW" if they had zero counts in all males and a minimum of one count in at least two females. Next, the density of these XY and ZW SNPs was assessed in 10 kb non-overlapping windows (Supplementary Fig. 6 first plot).

The difference between XY and ZW SNPs per 10 kb window was then calculated and only outlier windows were kept (Supplementary Fig. 6 second and third plot). These outlier windows were defined as windows with a difference of XY-ZW SNPs greater than the 75th percentile value + 1.5 times the inter quartile range. We then compared the distribution of XY and ZW SNPs in all outlier windows with a paired two-sided Mann-Whitney test

(Supplementary Fig. 6 fourth plot). If the two distributions were significantly different from each other (*P*-value<0.05), the heterogametic system was defined as the distribution (XY or ZW) with the higher total amount of SNPs. As a last step, we quantified XY or ZW SNPs of outlier windows (depending on the previously defined heterogametic system) per reference LG (corrected by chromosome size) and defined as potential sex chromosome the LG(s) with an amount of SNPs higher than the 75th percentile value + 3 times the inter-quartile range. In order to keep only the most extreme outliers and to avoid false positives, only the LGs with a number of SNPs higher than the standard deviation were kept for this final call. In species for which a heterogametic system was identified, we further visualized all SNPs of the outlier windows of that system along the genome (Supplementary Fig. 6 fifth plot).

Approach 4 Inference of heterogamety from sex-specific "sequence subtraction"

We applied approach 4 to all taxa of the LT radiation, which had both transcriptome and assembled WGS data of a female and a male available. For this, we adapted a "subtraction pipeline" from[46] to infer (expressed) sex-specific contigs. *De novo* assembled draft genomes from[24] were used as species-specific references. For each species, we pooled all male and female transcriptome data of all three tissues and quality filtered them with Trimmomatic[66] (v.0.33) with a 4 bp window size, a required minimum window Phred score quality of 15, and a minimum read length of 80 bp. Next, we modified the pipeline as follows[46]: In step 1, we used STAR (v.2.5.2a) to map RNA reads of one sex to the DNA *de novo* assembly of the opposite sex[24] (--outFilterMultimapNmax 10 --outFilterMatchNminOverLread 0.4 --outFilterScoreMinOverLread 0.4 --outFilterMismatchNmax 100 --seedSearchStartLmax 20 --seedPerReadNmax 100000 --seedPerWindowNmax 1000 --alignTranscriptsPerReadNmax 100000 --alignTranscriptsPerWindowNmax 10000). In step 2, we assembled all transcriptome reads that did not map to the genome of the opposite sex as previously

661 described[46]. In step 3, we used GMAP-GSNAP[70] (v.2017-08-15) with a minimum trimmed
662 coverage of 0.9 and a minimum identity of 0.98, to map the sex-specific *de novo* assembled
663 transcripts (generated in step 2) to the genome of the opposite sex. In step 4, we used STAR
664 and BEDTools[71] (v.2.26.0) to remove presumed sex-specific contigs that had more than 50%
665 of their length covered with RNA-reads from the opposite sex. In step 5, we used the CD-
666 HIT-EST function of Cd-hit[72] (v.4.6.4) to merge and extend sex-specific contigs. In step 6,
667 we mapped male and female genomic reads to sex-specific contigs as described in the
668 original pipeline[46]. In step 7, we used RSEM[73] (v.1.2.31) to calculate RPKM (reads per kb of
669 transcript per million mapped reads) values. We did not apply steps 8 (a repeat filter) and 9 (a
670 transcript length filter). We tested for a correlation between the type of heterogametic system
671 and the difference in the number of sex-specific contigs with a phylogenetic ANOVA using
672 phytools[74] (v.0.6-67).
673
674 **Final sex chromosome systems definition**
675 Sex-linked chromosomes, sex-differentiated regions and heterogametic state (XY/ZW) per
676 species were inferred from sex-association in GWAS (approach 1), the sex-specific allele test
677 (approach 2) and species-specific sex-differentiated site accumulations identified by allele
678 differences test based on transcriptomes (approach 3). For approach 1 and 2, in which we first
679 report the result per tribe, we defined sex chromosomes on the species level as follows: We
680 required the same sex-linked region to be present in at least two species to base a sex
681 chromosomal call on WGS data only. This might underestimate the presence of sex
682 chromosomes in our dataset but further reduces the number of false positives. Based on
683 approach 3, which includes more individuals per species and was run on the species-level, we
684 could confirm the sex chromosomes with larger sex-differentiated regions identified by
685 approach 1 and 2. We failed to detect some of the rather small regions with approach 3, such

as the narrow ~5 kb region in Lamprologini, which we think is due to a combination of the low number of genes present in these regions and probably low levels of expression of these genes in adults. However, approach 3 allowed us to identify and confirm species-specific sex chromosomes that we would not call/identify otherwise (note that similar sample sizes and transcriptome approaches have previously been used to identify sex chromosomes e.g.[46,75]). The effectiveness of our method is evidenced by our ability to recall all three previously identified sex chromosomes of LT cichlids[37].

Still, given the particularly reduced sample sizes for the small tribes in approach 2, we further decided to generate two sex chromosome call sets, a permissive dataset retaining all sex chromosomes identified by either approaches 1, 2 and 3 or combinations thereof and a stringent dataset, excluding all sex chromosomes that were exclusively identified in approach 2. We performed all subsequent analyses with both sets and report the results.

Approach 4 had previously been established for flies using a comparable sequencing set-up[46]. Within LT cichlids, we noted that this method seemed to have different levels of sensitivity depending on the type of heterogamety. We thus cautiously decided to not base our final sex chromosome call on this approach, but still report the outcome.

Reconstruction of sex chromosome turnovers in cichlids

In order to reconstruct sex chromosome evolution across the LT radiation, we coded the final sex chromosome set as a probability matrix which included 14 different LGs identified in at least one species as sex-linked, incorporating the published data for two labstrains and a cross derived from a natural population of *A. burtoni*[38,39] and a population of *P. philander*[25] (permissive dataset; stringent dataset 13 LGs). Note that *P. philander* was present in the current dataset with a single individual only and the *A. burtoni* samples included here derive from two different populations not allowing the confirmation of previously published data.

711 Species, for which we could not identify a sex-linked LG and none was published to the best
712 of our knowledge, were included in the analysis and attributed equal probability for all 14
713 LGs (permissive dataset; 13 LGs in the stringent dataset). We placed these sex chromosome
714 identities onto a time-calibrated phylogeny of LT cichlids[24], which we pruned to include only
715 the species studied here, using phytools. We followed the approach described in Jeffries *et*
716 *al.*[20] and inferred ancestral sex chromosome states using a stochastic mapping approach
717 implemented in phytools. We compared the likelihood scores (based on the Akaike
718 Information Criterion (AIC)) for three different transition rate models, equal rates (ER),
719 symmetrical (SYM), and all rates different (ARD), which identified ARD as the best model
720 for transition rates between states. We simulated 1000 stochastic character maps along the
721 phylogeny. In addition, we ran stochastic mapping for each chromosome separately, coding
722 the use of the chromosome as a sex chromosome in a given species as a binary (yes/no) trait
723 to account for the fact that some tips of the phylogeny are in two or more states (i.e., two or
724 more reference LGs showed sex-linkage likely due to chromosomal rearrangements/fusions)
725 rather than having the equal probability of being in one out of two states. Note, that for *A.*
726 *burtoni,* even four different LGs have been reported as sex chromosome[38,39]. We then
727 combined the 14 separate reconstructions (permissive dataset; 13 in the stringent dataset) into
728 one phylogenetic representation. The results obtained with the two approaches were very
729 similar and we hence continued calculations with the binary reconstructions.

730 We determined the timepoints of sex chromosome turnover events as points on
731 branches where the inferred probability of using a given chromosome as a sex chromosome
732 dropped below 0.5 for the first time starting from the tips of the phylogeny with the function
733 densityMap of phytools. Based on Jeffries *et al.*[20] we did not consider species that had no
734 detectable sex chromosome as having losses but only considered transition events that led to
735 the emergence of a new sex chromosome, i.e., we only considered gains.

Likewise, we ran a second independent analysis with 1000 stochastic mappings to reconstruct ancestral states for the type of heterogamety (XY/ZW). In addition to the reconstructed turnover points, we here added a turnover on the terminal branch leading to *A. burtoni*, since for this species, both XY and ZW sex chromosomes have been described[39].

To test if gene content or chromosome size drives the observed pattern of sex chromosome recruitment in LT cichlids, we randomly picked 30 times (the number of sex chromosome recruitments derived from ancestral state reconstruction) a window of 10 kb of the reference genome and attributed the LG containing this window as sex chromosome to a species. We simulated this operation 10000 times and counted how many times each LG was recruited in each simulation. We than counted in how many simulations nine or more LGs were not recruited, as this was the observed pattern.

We then tested for a correlation of the number of sex chromosome turnovers leading to the tips of each tribe with the number of species investigated in each tribe with a phylogenetic generalized linear model (pGLS) using the R package ape[76] (v.5.2).

Reconstruction of sex chromosome turnovers in other teleosts

We then ran the same two analyses for ricefishes (Adrianichthyidae), which, to the best of our knowledge, are the only fish family with detailed data on sex chromosomes with synteny inference based on a comparison to a common reference genome (*Oryzias latipes*). Information on sex chromosomes was taken from Hilgers and Schwarzer[30] and placed on a time-calibrated phylogeny of the family Adrianichthyidae (19 species, Supplementary Table 5), extracted from a recent comprehensive ray-finned fish phylogeny[40]. We could not include sex chromosome data of three species (*Oryzias sakaizumii*, *Oryzias wolasi*, and *Oryzias woworae*), as these were not included in the phylogeny and no other comprehensive time-calibrated tree comprising these fishes was available to us.

To compare our data on a larger scale, we calculated transition rates for ray-finned fishes of the Tree of Sex database (http://www.treeofsex.org/). We used all Tree of Sex species that were also included in the recent comprehensive ray-finned fish phylogeny[40] (Supplementary Table 6). As several species names were not initially included in the phylogeny[40], we inspected species names of Tree of Sex for typos, older versions of species names and synonyms in FishBase (www.fishbase.org) and Eschmeyer's Catalog of Fishes Online Database (https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes), and corrected the names accordingly. This allowed us to map SD data for 472 species from the Tree of Sex database onto the phylogeny. We further added published data for cichlids[25,29,37,38,39] and this study, resulting in an additional 72 species. Sex determination data from the Tree of Sex database were simplified and coded as a probability matrix with three states, namely "XY" (including species classified by Tree of Sex as "XY heteromorphic", "XY homomorphic", "XO", "XY polygenic"), "ZW" (including species classified by Tree of Sex as "ZW heteromorphic" and "ZW homomorphic", "ZO", "ZW polygenic") and "NonGSD" (including species classified by Tree of Sex as "apomictic", "hermaphrodite", "ESD_other", "pH", "size", "density", "TSD", "other"). The final matrix is provided in Supplementary Table 6. Similar to our approach described above, all other species with no information on sex determination were included with an equal probability for all three states.

Convergent evolution of sex chromosomes

We detected the same region on LG19 as sex-linked in species belonging to the tribes Haplochromini and Perissodini. Within Haplochromini, this sex-linked region was present in six species of the genus *Tropheus* (tribe Tropheini, the endemic LT Haplochromini) as well as in *O. indermauri* (a distantly related riverine haplochromine from the Lufubu river, which drains into LT). Our ancestral state reconstruction suggested an independent origin of the

786 LG19 SD system in Perissodini, *Tropheus* and *O. indermauri.* To further investigate the
787 hypothesis of convergence, we extracted all SNPs from LG19 which were XY in at least one
788 of the species with a positionally overlapping XY-system from a variant call file containing
789 all species investigated in the present study[24] with VCFtools. In addition to a male and a
790 female of these eight species, we included representatives without this XY-LG19 system
791 from the other tribes (a male and a female of each of *Xenotilapia flavipinnis*, *Plecodus*
792 *elaviae*, *Petrochromis trewavasae*, *Eretmodus cyanostictus*, *Benthochromis tricoti*, *C.*
793 *leptosoma*, *Limnochromis auritus*, and *Cyphotilapia frontosa*) as well as other haplochromine
794 species (a male *P. philander*, a male and a female *A. burtoni*, a male *Ctenochromis polli*). We
795 only kept variants with less than 10% missing data. We next extracted the two haplotype
796 sequences of each individual for all variants in FASTA format. Assuming that the variant
797 phasing with beagle was not error-free across whole chromosomes, we inspected the
798 haplotypes and corrected the phasing for the eight LG19-XY species. This was done such that
799 for sites where an XY male was heterozygous while the corresponding XX female was
800 homozygous, the allele of the male shared with the female was designated as haplotype 1 (the
801 presumed X-allele) and the other allele as haplotype 2 (the presumed Y-allele). We then
802 inferred a phylogenetic tree by maximum likelihood with IQ-TREE[77] (v.1.7-beta12) under
803 the GTR+F+ASC substitution model to account for ascertainment bias and assessing branch
804 support with 1000 ultrafast bootstrap approximations[78]. We rooted the obtained phylogenetic
805 tree in accordance with the species tree (Fig. 1).
806
807 **Defining the degree of sex chromosome differentiation, potential sex-determining**
808 **regions, and candidate genes**
809 On the above-defined sex chromosomes, we characterised species-specific sex-differentiated
810 regions by counting the numbers of XY- and ZW-SNPs in windows of 10 kb. The density of

XY- or ZW-windows is shown in Supplementary Fig. 10. We defined the size of the sex-differentiated region as the proportion of the LG covered by windows that have a density of sex-specific SNPs that is more than twice as high as the genome-wide mean over all windows such that the sum of all sex-differentiated windows defines the cumulative length of the sex-differentiated regions and the minimum and maximum window coordinates define the range of the sex-differentiated region on the LG. We tested for a correlation between sex chromosome differentiation and the estimated age of origin of the sex chromosome derived from the turnover point with a phylogenetic generalized linear model (pGLS) using the R package ape. From the results of approaches 1-3, we identified sex-differentiated regions shared between several species and overlaid these with candidate genes involved in sex determination and pigmentation. Pigmentation genes in the reference genome were defined over gene ontology annotations including the term "pigmentation" and its child terms. We also retrieved orthologous sequences of the Nile tilapia to the medaka pigmentation genes defined by Braasch *et al.*[79] over Biomart, Ensembl release 96 (www.ensembl.org). Since this Nile tilapia genome is a different genome release than the reference genome used by us, we searched the NCBI database for the obtained Ensembl gene IDs and translated them to the assembly version that we used with the NCBI Genome Remapping Service. Candidate genes for SD included genes previously identified through a literature search[80,81] and a gene ontology analysis based on a GO annotation matching the word "sex" (list of gene IDs of candidate genes for SD and pigmentation in Supplementary Table 7). We further investigated all annotated genes that were partially or fully included in the window(s) with the maximum number of sex-specific SNPs on the sex chromosome (Supplementary Table 2).

Data availability

Sequencing data used in this study were published elsewhere[24,27]. All other data are provided in this paper and its Supplementary Information. Data were analysed with publically available software and their included functions as detailed in the methods section.

References

1. Rice WR. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735-742 (1984).
2. Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat* **130**, 113-146 (1987).
3. Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol* **27**, 3831-3851 (2018).
4. Myosho T*, et al.* Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* **191**, 163-170 (2012).
5. Tennessen JA, Wei N, Straub SCK, Govindarajulu R, Liston A, Ashman T-L. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol* **16**, e2006062 (2018).
6. Bull JJ, Charnov EL. Changes in the heterogametic mechanism of sex determination. *Heredity* **39**, 1-14 (1977).
7. Ogata M, Lambert M, Ezaz T, Miura I. Reconstruction of female heterogamety from admixture of XX-XY and ZZ-ZW sex-chromosome systems within a frog species. *Mol Ecol* **27**, 4078-4089 (2018).
8. Blaser O, Grossen C, Neuenschwander S, Perrin N. Sex-chromosome turnovers induced by deleterious mutation load. *Evolution* **67**, 635-645 (2013).

9. Blaser O, Neuenschwander S, Perrin N. Sex-chromosome turnovers: The hot-potato model. *Am Nat* **183**, 140-146 (2014).

10. van Doorn GS, Kirkpatrick M. Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics* **186**, 629-645 (2010).

11. van Doorn GS, Kirkpatrick M. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**, 909-912 (2007).

12. Eshel I. Selection on sex-ratio and the evolution of sex-determination. *Heredity* **34**, 351 (1975).

13. Veller C, Muralidhar P, Constable GWA, Nowak MA. Drift-induced selection between male and female heterogamety. *Genetics* **207**, 711-727 (2017).

14. Saunders PA, Neuenschwander S, Perrin N. Sex chromosome turnovers and genetic drift: a simulation study. *J Evol Biol* **31**, 1413-1419 (2018).

15. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**, 113-124 (2013).

16. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat Ecol Evol* **3**, 1632-1641 (2019).

17. Pokorná M, Kratochvíl L. Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zool J Linn Soc* **156**, 168-183 (2009).

18. Bull JJ. *Evolution of sex determining mechanisms*. Benjamin/Cummings Publishing (1983).

19. Heule C, Salzburger W, Böhne A. Genetics of sexual development – an evolutionary playground for fish. *Genetics* **196**, 579-591 (2014).

20. Jeffries DL*, et al.* A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat Commun* **9**, 4088 (2018).

882 21. Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D. Restriction site-
883 associated DNA sequencing (RAD-seq) reveals an extraordinary number of
884 transitions among gecko sex-determining systems. *Mol Biol Evol* **32**, 1296-1309
885 (2015).
886 22. Ronco F, Indermaur A, Büscher H, Salzburger W. The taxonomic diversity of the
887 cichlid fish fauna of ancient Lake Tanganyika, East Africa. *J Great Lakes Res* **46**,
888 1067-1078 (2020).
889 23. Nelson JS, Grande TC, Wilson MVH. *Fishes of the World. 5th Edition*. John Wiley &
890 Sons, Hoboken (2016).
891 24. Ronco F*, et al.* Drivers and dynamics of a massive adaptive radiation in African
892 cichlid fish. *Nature* **Accepted**, (2020).
893 25. Böhne A*, et al.* Repeated evolution versus common ancestry: Sex chromosome
894 evolution in the haplochromine cichlid *Pseudocrenilabrus philander*. *Genome Biol*
895 *Evol* **11**, 439–458 (2019).
896 26. Gammerdinger WJ, Kocher TD. Unusual diversity of sex chromosomes in African
897 cichlid fishes. *Genes* **9**, 480 (2018).
898 27. El Taher A*, et al.* The evolution of gene expression levels during rapid organismal
899 diversification. *Nat Ecol Evol* **Accepted**, (2020).
900 28. Roberts RB, Ser JR, Kocher TD. Sexual conflict resolved by invasion of a novel sex
901 determiner in Lake Malawi cichlid fishes. *Science* **326**, 998-1001 (2009).
902 29. Ser JR, Roberts RB, Kocher TD. Multiple interacting loci control sex determination in
903 Lake Malawi cichlid fish. *Evolution* **64**, 486-501 (2010).
904 30. Hilgers L, Schwarzer J. The natural history of model organisms: The untapped
905 potential of medaka's wild relatives. *eLIFE* **8**, e46994 (2019).

906 31. Charlesworth D. Evolution of recombination rates between sex chromosomes. *Philos*
907 *Trans R Soc Lond B Biol Sci* **372**, 20160456 (2017).
908 32. Sardell JM*, et al.* Sex differences in recombination in sticklebacks. *G3* **8**, 1971-1983
909 (2018).
910 33. Feulner PGD, Schwarzer J, Haesler MP, Meier JI, Seehausen O. A dense linkage lap
911 of Lake Victoria cichlids improved the *Pundamilia* genome assembly and revealed a
912 major QTL for sex-setermination. *G3* **8**, 2411-2420 (2018).
913 34. Conte MA*, et al.* Chromosome-scale assemblies reveal the structural evolution of
914 African cichlid genomes. *Gigascience* **8**, (2019).
915 35. Matsuda M, Sotoyama S, Hamaguchi S, Sakaizumi M. Male-specific restriction of
916 recombination frequency in the sex chromosomes of the medaka, *Oryzias latipes*.
917 *Genet Res* **73**, 225-231 (1999).
918 36. Tanaka K, Takehana Y, Naruse K, Hamaguchi S, Sakaizumi M. Evidence for
919 different origins of sex chromosomes in closely related *Oryzias* fishes: Substitution of
920 the master sex-determining gene. *Genetics* **177**, 2075-2081 (2007).
921 37. Gammerdinger W, Conte M, Sandkam B, Ziegelbecker A, Koblmüller S, Kocher T.
922 Novel sex chromosomes in 3 cichlid fishes from Lake Tanganyika. *J Hered* **109**, 489-
923 500 (2018).
924 38. Böhne A, Wilson CA, Postlethwait JH, Salzburger W. Variations on a theme:
925 Genomics of sex determination in *Astatotilapia burtoni*. *BMC Genomics* **17**, 883
926 (2016).
927 39. Roberts NB*, et al.* Polygenic sex determination in the cichlid fish *Astatotilapia*
928 *burtoni*. *BMC Genomics* **17**, 835 (2016).
929 40. Rabosky DL*, et al.* An inverse latitudinal gradient in speciation rate for marine fishes.
930 *Nature* **559**, 392-395 (2018).

931 41. Pennell MW*, et al.* Y fuse? Sex chromosome fusions in fishes and reptiles. *PLoS*
932 *Genet* **11**, e1005237 (2015).
933 42. Kitano J*, et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature*
934 **461**, 1079-1083 (2009).
935 43. Lucek K. Evolutionary mechanisms of varying chromosome numbers in the radiation
936 of Erebia butterflies. *Genes* **9**, (2018).
937 44. King M. *Species evolution - the role of chromosome change*. Cambridge University
938 Press (1993).
939 45. Ozouf-Costaz C, Coutanceau JP, Bonillo C, Mercot H, Hermon Y, Guidi-Rontani C.
940 New insights into the chromosomal differentiation patterns among cichlids from
941 Africa and Madagascar. *Cybium* **41**, 35-43 (2017).
942 46. Mahajan S, Bachtrog D. Convergent evolution of Y chromosome gene content in
943 flies. *Nat Commun* **8**, 785 (2017).
944 47. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. Male development
945 of chromosomally female mice transgenic for Sry. *Nature* **351**, 117 - 121 (1991).
946 48. Seehausen O, van Alphen JJM, Lande R. Color polymorphism and sex ratio distortion
947 in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol*
948 *Lett* **2**, 367-378 (1999).
949 49. Rodrigues N, Studer T, Dufresnes C, Perrin N. Sex-chromosome recombination in
950 common frogs brings water to the fountain-of-youth. *Mol Biol Evol* **35**, 942-948
951 (2018).
952 50. Marshall Graves JA, Peichel CL. Are homologies in vertebrate sex determination due
953 to shared ancestry or to limited options? *Genome Biol* **11**, 205 (2010).
954 51. Pennell MW, Mank JE, Peichel CL. Transitions in sex determination and sex
955 chromosomes across vertebrate species. *Mol Ecol* **27**, 3950-3963 (2018).

52. Lande R, Seehausen O, van Alphen JJ. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112-113**, 435-443 (2001).

53. Ohno S. *Sex chromosomes and sex-linked genes.* Springer (1967).

54. Dobigny G, Ozouf-Costaz C, Bonillo C, Volobouev V. Viability of X-autosome translocations in mammals: an epigenomic hypothesis from a rodent case-study. *Chromosoma* **113**, 34-41 (2004).

55. Charlesworth D, Charlesworth B. Sex-differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet Res* **35**, 205-214 (1980).

56. de Villena FPM, Sapienza C. Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**, 1179-1189 (2001).

57. Yoshida K, *et al.* B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genet* **7**, e1002203 (2011).

58. Clark FE, Kocher TD. Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Sci Rep* **9**, (2019).

59. McKenna A, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).

60. Danecek P, *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).

61. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-1097 (2007).

62. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

63. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824 (2012).

64. Knaus BJ, Grunwald NJ. VCFR: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* **17**, 44-53 (2017).

65. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* **18**, 341 (2017).

66. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

67. Lopez-Maestre H*, et al.* SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res* **44**, e148-e148 (2016).

68. Dobin A*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

69. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B* **57**, 289-300 (1995).

70. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).

71. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

72. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).

73. Li B, Dewey C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

74. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**, 217-223 (2012).

1005 75. Harrison PW, *et al.* Sexual selection drives evolution and rapid turnover of male gene
1006 expression. *Proc Natl Acad Sci U S A* **112**, 4393-4398 (2015).
1007 76. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and
1008 evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
1009 77. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective
1010 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
1011 **32**, 268-274 (2015).
1012 78. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving
1013 the ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518-522 (2018).
1014 79. Braasch I, Brunet F, Volff J-N, Schartl M. Pigmentation pathway evolution after
1015 whole-genome duplication in fish. *Genome Biol Evol* **1**, 479-493 (2009).
1016 80. Böhne A, Heule C, Boileau N, Salzburger W. Expression and sequence evolution of
1017 aromatase *cyp19a1* and other sexual development genes in East African cichlid fishes.
1018 *Mol Biol Evol* **30**, 2268-2285 (2013).
1019 81. Heule C, Göppert C, Salzburger W, Böhne A. Genetics and timing of sex
1020 determination in the East African cichlid fish *Astatotilapia burtoni*. *BMC Genet* **15**,
1021 140 (2014).
1022

1023 ## Acknowledgements

1024 We thank Daniel Jeffries and Guillaume Lavanchy for code sharing for ancestral state
1025 reconstructions and Milan Malinsky for discussions on sex-specific sites identification.
1026 Calculations were performed at the sciCORE (http://scicore.unibas.ch/) scientific computing
1027 center at the University of Basel, with support by the SIB (Swiss Institute of Bioinformatics),
1028 and at the Abel computer cluster at the University of Oslo. This work was funded by the

Swiss National Science Foundation (SNSF, Ambizione grant PZ00P3_161462) to AB and the European Research Council (ERC, CoG 617585 "CICHLID~X") to WS.

Author Contributions

A.B. designed the study with input from A.E., F.R., and W.S.; A.E. and A.B. analysed all data, M.M. performed variant calling and phylogenetic reconstructions for convergent sex chromosome evolution, and helped with ancestral state reconstructions as well as with statistics, F.R. helped in analysing data for ancestral state reconstructions and sex-specific site identifications, A.B. and A.E. wrote the manuscript with final contributions from all authors. All authors read and approved the final manuscript.

Competing interest declaration

The authors declare no competing interests.

Materials & Correspondence

Correspondence and requests to materials should be addressed to a.boehne@leibniz-zfmk.de

Figures

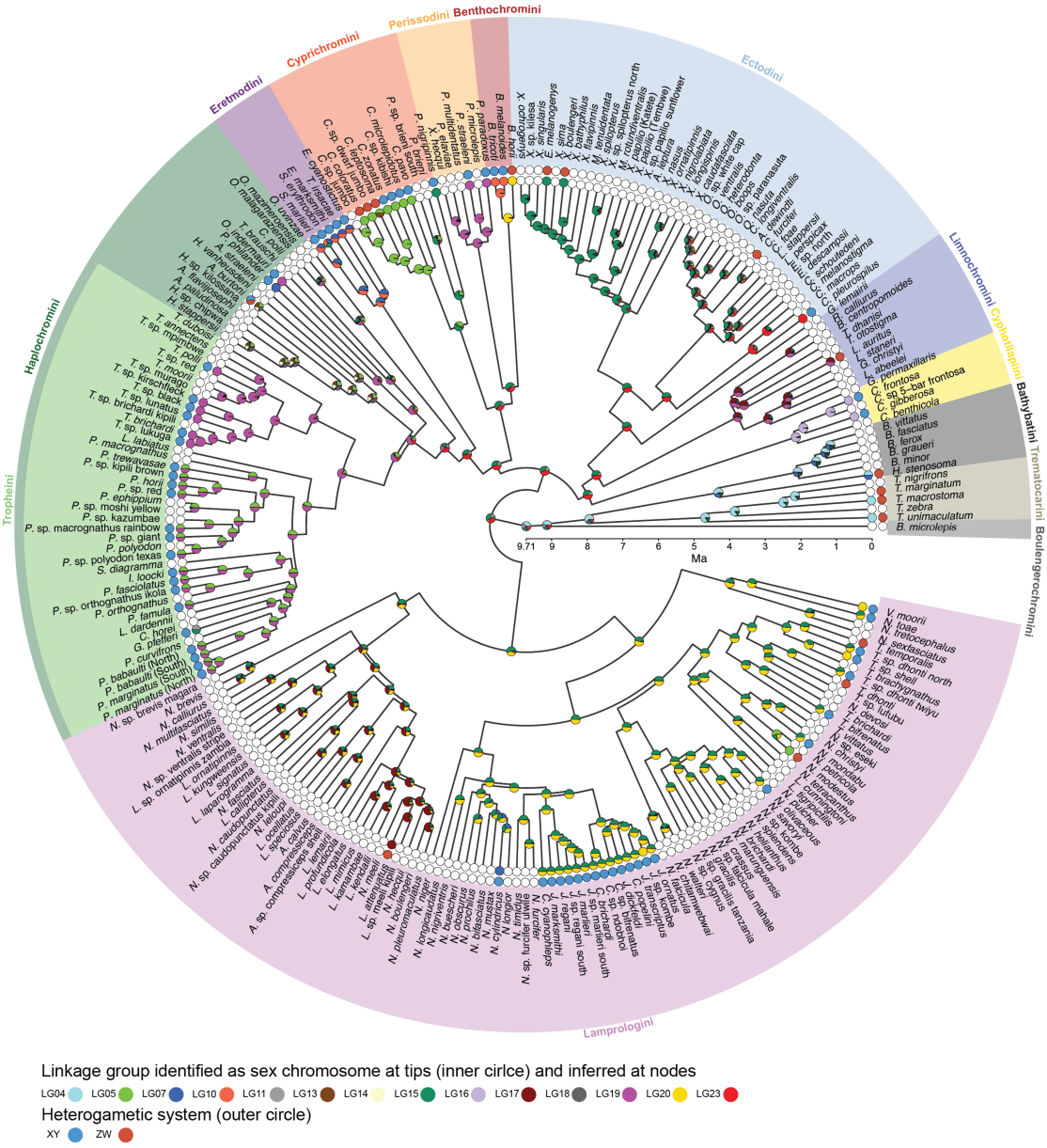

Fig. 1: Sex chromosome evolution in the cichlid adaptive radiation in Lake Tanganyika. Sex chromosome state and ancestral state reconstruction in LT cichlids are placed on a time-calibrated species tree[24] with tribe-grouping indicated by colour shading. The inner circle at -tips shows identified sex-linked LGs. Colours refer to LGs of the reference genome; two- or more coloured symbols at tips indicate sex chromosomal signals that were detected on two or more reference LGs suggesting chromosomal rearrangements. The outer circle indicates the heterogametic status of each species (blue: XY; red: ZW). White circles at tips indicated that

no sex chromosome could be identified. Pie charts at nodes represent the probability for an LG being a sex chromosome at this time derived from ancestral state reconstructions.

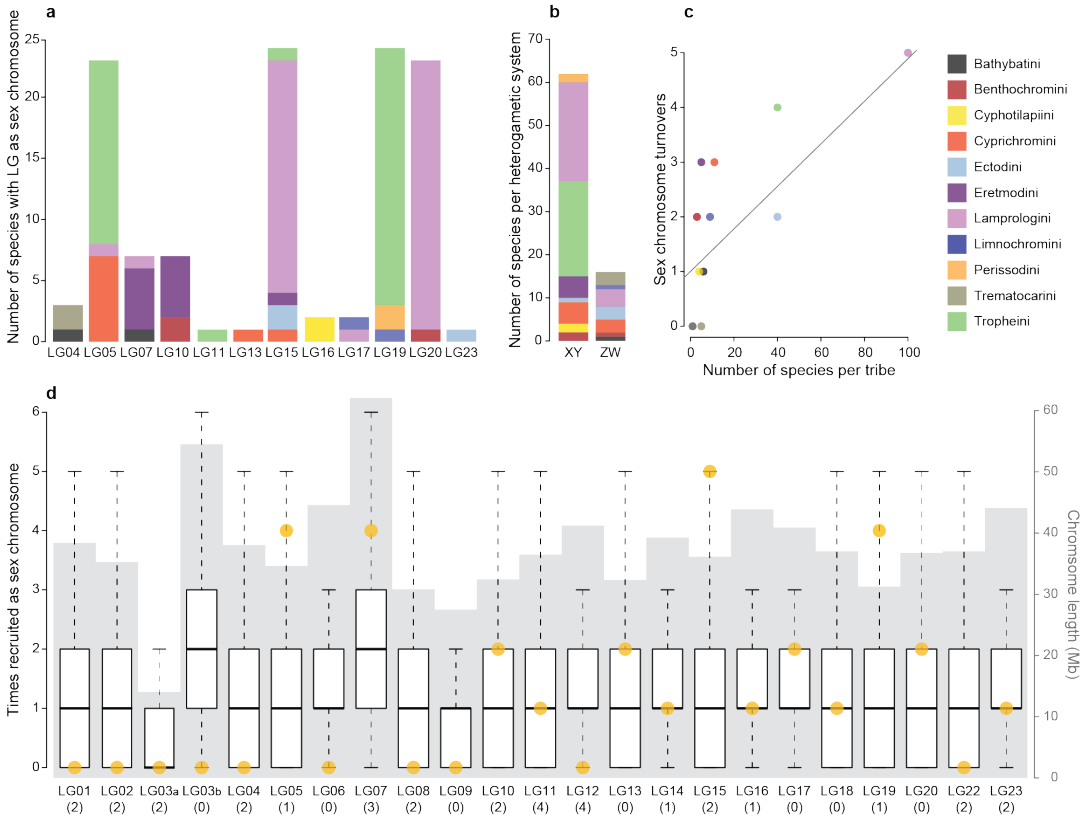

Fig. 2: Non-random sex chromosome distribution in Lake Tanganyika cichlids.

a Use of different LGs as sex chromosomes. Bars represent the number of times an LG has been detected as sex-linked at the species level and are coloured according to tribe. **b** The occurrence of sex determination (SD) systems. Bars represent how often an XY or ZW SD system was identified at the species level and are coloured according to tribe. **c** Correlation between species richness and sex chromosome turnover. The number of sex chromosome turnovers leading to the tips of each tribe is correlated with the number of species investigated in each tribe (pGLS: P=0.0043, coeff=0.039). Dots are coloured according to tribes; the line represents the linear model fitted to the data. **d** Boxplots showing the expected number of sex chromosome recruitments if recruitment was at random (10000 permutations). Boxplot centre lines represent the median, box limits the upper and lower quartiles, and

whiskers the 1.5x interquartile range. Outliers are not shown. Ten reference LGs were never implicated in a turnover event in LT cichlids. Under random recruitment in the simulations this pattern occurred only in 2.01 % of all simulations. Yellow dots indicate the number of observed sex chromosome recruitments per LG derived from ancestral state reconstructions, grey background shading represents chromosome length in Mb and numbers below each boxplot indicate the number of previously described sex-determining genes on these LGs.

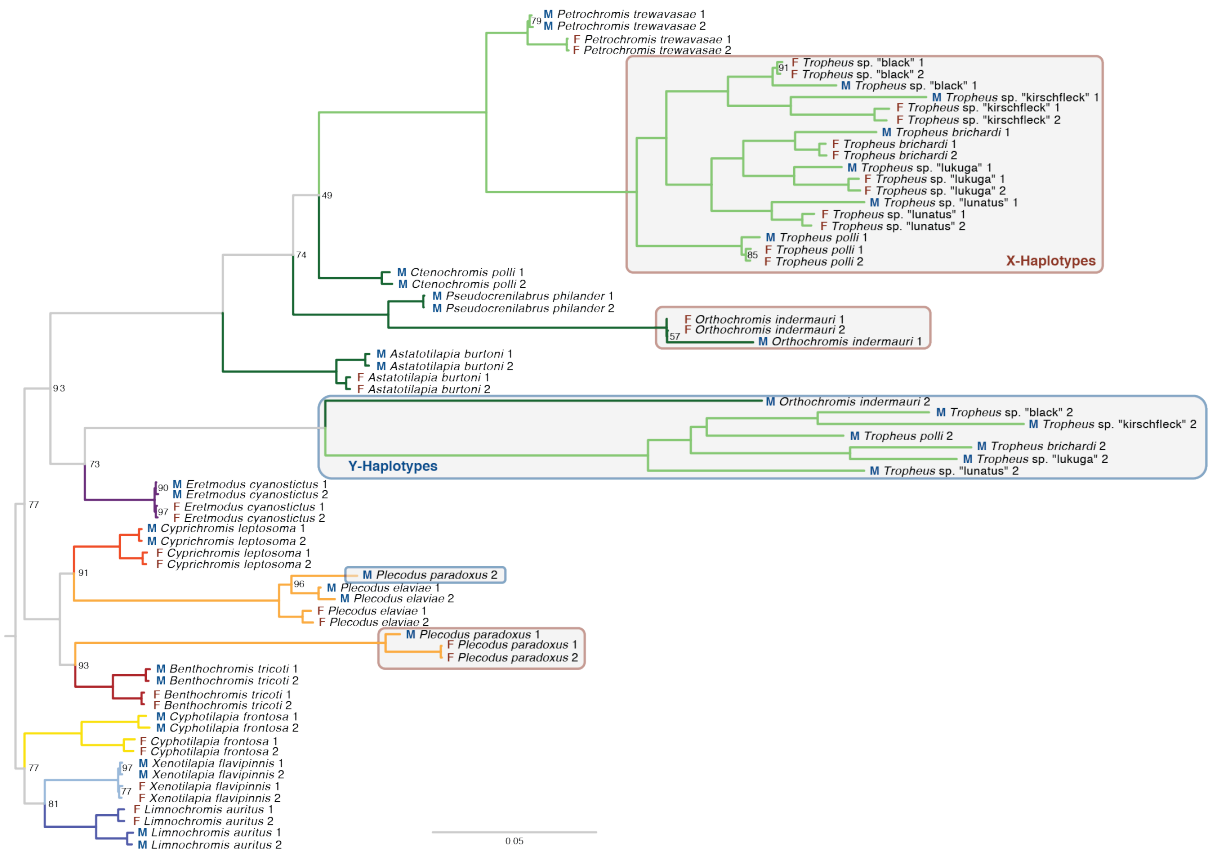

Fig. 3: Convergent evolution of LG19 as XY sex chromosome in two Tanganyikan cichlid tribes

Phylogenetic tree of X- and Y-haplotype sequences does not group the *P. paradoxus* Y-haplotype with the *Tropheus* Y-haplotypes but supports the species tree, suggesting convergent evolution. The Y-haplotype of the non-LT riverine haplochromine *O. indermauri* groups with the Y-haplotypes of the *Tropheus* species, supporting monophyly of this sex

chromosomal system. The scale bar indicates the number of substitutions per site; values at nodes represent bootstrap support (% of 1000 bootstraps, if no value is shown the node support was 100%).

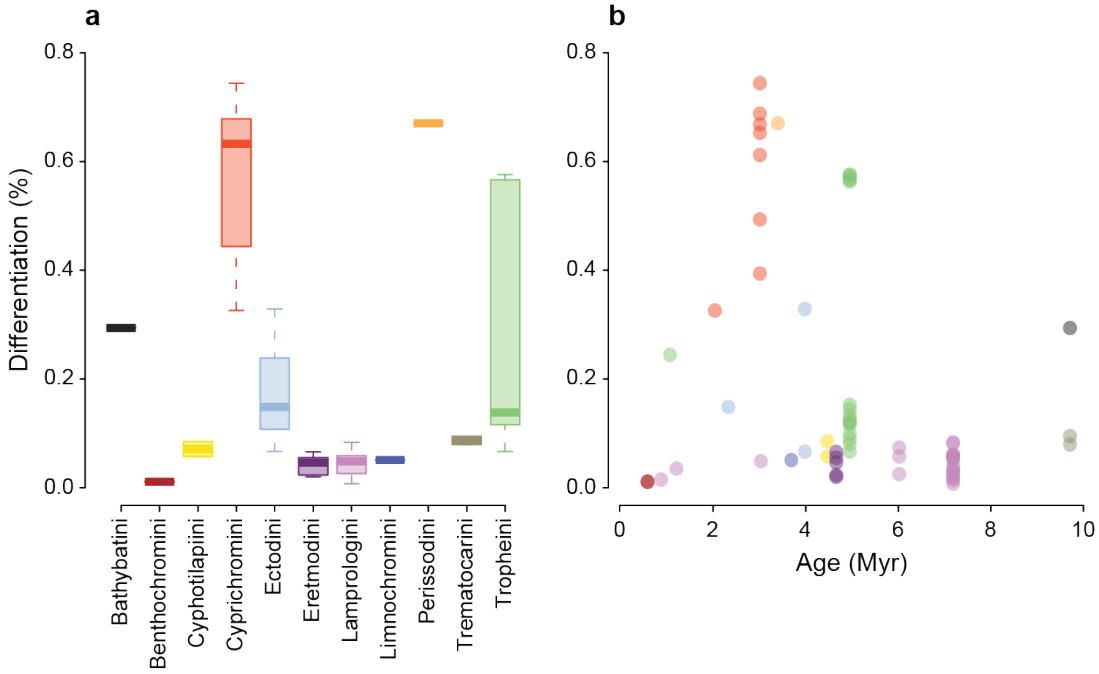

Fig. 4: Sex chromosome differentiation in Lake Tanganyika cichlids.

a Size distribution of sex-differentiated regions. The size of these regions corresponds to the proportion of the LG with windows that have more sex-specific SNPs than two times the mean across all windows. **b** Per-species proportion of the chromosome(s) showing sex differentiation and corresponding estimated ages of the sex chromosomal system based on ancestral state reconstructions on a time-calibrated species tree. The degree of differentiation is not correlated with the estimated age of origin (pGLS: P=0.9049, coeff=0.0011).

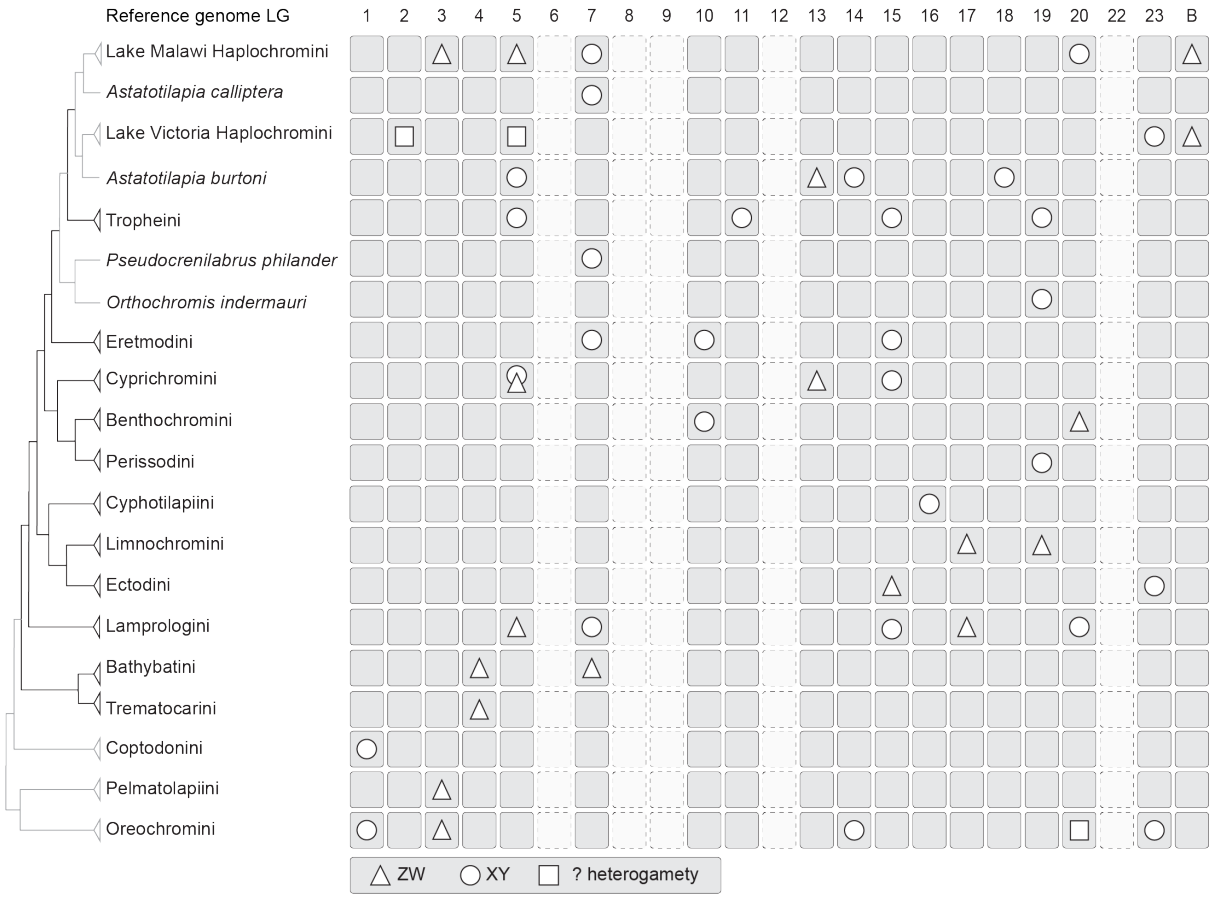

Fig. 5: Sex chromosome evolution in African cichlids. Phylogenetic relationships[24,25] and sex chromosome occurrence with reference to the genome of the Nile tilapia (*O. niloticus*) in African cichlids. Cichlid lineages found in Lake Tanganyika are indicated in black, cichlids from other lakes or rivers in grey. Sex chromosome information is derived from this study, and previously published summaries[25,26].