bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **A dense linkage map of Lake Victoria cichlids improved the *Pundamilia* genome**
2 **assembly and revealed a major QTL for sex-determination**
3
4 Philine G.D. Feulner*[1,2], Julia Schwarzer*[1,2,3], Marcel P. Haesler[1,2], Joana I. Meier[1,2],
5 Ole Seehausen[1,2]
6
7
8
9
10 [1] Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and
11 Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and
12 Technology, Switzerland
13 [2] Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution,
14 University of Bern, Switzerland
15 [3]Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany
16
17 * These authors contributed equally.
18
19 Correspondence: philine.feulner@eawag.ch
20
21
22
23

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 **Abstract**

25 Genetic linkage maps are essential for comparative genomics, high quality genome
26 sequence assembly and fine scale quantitative trait locus (QTL) mapping. In the
27 present study we identified and genotyped markers via restriction-site associated
28 DNA (RAD) sequencing and constructed a genetic linkage map based on 1,597 SNP
29 markers of an interspecific F2 cross of two closely related Lake Victoria cichlids
30 (*Pundamilia pundamilia* and *P.* sp. "red head"). The SNP markers were distributed on
31 22 linkage groups and the total map size was 1,594 cM with an average marker
32 distance of 1.01 cM. This high-resolution genetic linkage map was used to anchor the
33 scaffolds of the *Pundamilia* genome and estimate recombination rates along the
34 genome. Via QTL mapping we identified a major QTL for sex in a ~1.9 Mb region on
35 Pun-LG10, which is homologous to *Oreochromis niloticus* LG 23 (Ore-LG23) and
36 includes a well-known vertebrate sex-determination gene (*amh*).

37

38

39 **Introduction**

40 The haplochromine cichlid lineage of the East African Great Lakes is famous for
41 forming large adaptive radiations often in exceptionally short time, resulting in
42 several hundred species each in Lakes Malawi and Victoria, and dozens of species
43 each in several smaller East African Lakes (Kocher 2004; Seehausen 2015). The Lake
44 Victoria haplochromine cichlid radiation stands out in being the youngest (~15,000
45 years) showing a high degree of diversity in morphology, behavior and ecology
46 (Greenwood 1974; Seehausen 1996). An abundance of studies have been published on
47 the evolution of Lake Victoria cichlids, providing insight to colonization history (Nagl
48 *et al.* 2000; Seehausen *et al.* 2003; Verheyen *et al.* 2003; Meier *et al.* 2017b), species
49 formation (Seehausen *et al.* 1997; Seehausen and van Alphen 1999; Seehausen *et al.*
50 1999; Selz *et al.* 2014a), the interaction of sexual and natural selection (Seehausen
51 2000, Seehausen *et al.* 2008, Maan and Seehausen 2011), and the role of
52 hybridization between distant relatives (Seehausen *et al.* 2003, Keller *et al.* 2013, Selz
53 *et al.* 2014b, Meier *et al.* 2017a, Meier *et al.* 2017b). Recently, several cichlid
54 genomes were published (Brawand *et al*. 2014), among them one from Lake Victoria.
55 This genome has been started to be used to investigate the genomic landscape of
56 speciation (Meier *et al.* accepted). Detailed genetic linkage maps offer a powerful tool
57 to improve the quality of genome assemblies (Fierst 2015) and set the framework for

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 quantitative trait loci (QTL) localization. In the past decade, a number of genetic
59 linkage maps have been published for haplochromine cichlids using various molecular
60 genetic markers (Streelman *et al.* 2003; Sanetra *et al.* 2009; O´Quin *et al.* 2013;
61 Henning *et al.* 2014; 2017). For Lake Victoria cichlids three linkage maps based on
62 two interspecific F2 hybrid crosses were published. The first was based on an F2
63 cross between *Paralabidochromis chilotes* and *Paralabidochromis sauvagei* and
64 contained 184 microsatellites and two SNP markers with a mean marker spacing of
65 6.09 cM on 25 linkage groups (Kudo *et al.* 2015). The two others were based on F2
66 crosses between *Paralabidochromis sauvagei* and *Pundamilia* cf. *nyererei* (Henning
67 *et al*. 2014) and *Paralabidochromis chilotes* and *Pundamilia* cf. *nyererei* (Henning *et*
68 *al*. 2017). Linkage maps were constructed with 867 and 752 single-nucleotide
69 polymorphism (SNP) markers resulting in a mean marker spacing of 1.30 and 1.09
70 cM, respectively on 22 linkage groups (Henning *et al.* 2014; 2017). These linkage
71 maps were then used to identify QTL, such as for lateral stripes, lip size, and head
72 morphology (Henning *et al*. 2014; 2017) and sex determination (Kudo *et al*. 2015).
73 None of the linkage maps has been used to improve the Lake Victoria haplochromine
74 genome assembly.
75
76 In haplochromine cichlids, some polymorphic color patterns are genetically linked to
77 sex determination and are associated with segregating polymorphisms in sex
78 determination (Holzberg 1978; Seehausen *et al*. 1999; Lande *et al*. 2001; Streelman *et*
79 *al*. 2003; Kocher 2004). These observations supported the hypothesis that the rapid
80 evolution of sex determination systems might play a role in the very rapid speciation
81 of haplochromine cichlids (Seehausen *et al*. 1999; Lande *et al*. 2001; Kocher 2004;
82 Ser *et al*. 2010). A high diversity of sex determination systems and high sex
83 chromosome turnover rates are known in fish, including cichlids, with a variety of
84 environmental and genomic factors resulting in male or female phenotypes (reviewed
85 e.g. in Heule *et al*. 2014a). In cichlids, very closely related species, populations within
86 the same species, and even individuals within a population, can have different sex
87 determination mechanisms or non-homologous sex chromosomes. This is evidenced
88 by the presence of both XX-XY and ZZ-ZW sex determination systems within
89 haplochromines of Lakes Victoria and Malawi and in oreochromine cichlids
90 (Seehausen *et al*. 1999; Lande *et al*. 2001; Cnaani *et al*. 2008; Roberts *et al*. 2009; Ser
91 *et al*. 2010). Some candidates for genetic sex determination in cichlids exist and could

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

92 be associated with respective chromosomes. Among different species of
93 *Oreochromis*, sex determination loci have been repeatedly mapped on linkage group
94 (LG) 1 (XY), LG 3 (ZW) and LG 23 (XY) (Cnaani *et al*. 2008) and in haplochromine
95 cichlids, sex determination loci mainly mapped to LG 5 (ZW and XY) and LG 7 (XY)
96 (Ser *et al*. 2010; Kudo *et al*. 2015; Roberts *et al*. 2016; Böhne *et al*. 2016; Peterson *et*
97 *al*. 2017). Some genes that have repeatedly evolved as master sex determination genes
98 in teleost fishes (Kikuchi and Hamaguchi 2013; Heule *et al*. 2014a) seem to play a
99 role in sex determination in cichlids as well. Recent results published on *Astatotilapia*
100 *calliptera*, a haplochromine cichlid from Lake Malawi, and *Oreochromis niloticus*, a
101 distant relative of the East African adaptive radiations, indicate that two of these
102 candidate genes, the gonadal soma-derived factor (*gsdf*) and the anti Müllerian
103 hormone (*amh*) might have been re-used as sex determination loci (Eshel *et al*. 2014;
104 Peterson *et al*. 2017). Those genes are often derived by duplication or allelic
105 diversification from genes with a known function in sex differentiation or gonad
106 development (Heule *et al*. 2014a).
107
108 In the present study we construct a linkage map of an interspecific F2 cross between
109 two very closely related Lake Victoria cichlid species (*Pundamilia pundamilia* and *P.*
110 sp. "red head). The map was build using 1,597 SNPs identified and genotyped via
111 restriction-site associated DNA (RAD) sequencing with an average marker distance of
112 1.01 cM. We then used the linkage map to anchor the scaffolds of the *P. nyererei*
113 reference genome to the 22 linkage groups of the map and to perform a QTL analysis
114 for putative sex determination loci in *Pundamilia*. We identify the LG determining
115 sex in a Lake Victoria cichlid cross, as well as potential candidate genes for sex
116 determination and put these findings into the context of sex determination evolution
117 within a rapidly radiating clade of fish.
118
119
120 **Materials and Methods**
121 **Mapping family and RAD sequencing**
122 The genetic cross was started with a lab bred *Pundamilia* sp. "red head" male from
123 Zue Island in Lake Victoria (lab strain established from wild caught fishes by OS in
124 1993, 4th or 5th lab generation) and a wild *P. pundamilia* female caught by OS at
125 Makobe Island in Lake Victoria in 2003. Eggs were removed from the female's

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

126 mouth five days after spawning and reared in isolation from the adults. After reaching
127 maturity, four F1 individuals were crossed, resulting in two F2 families with together
128 more than 300 individuals. When F2 individuals were adult and sexually mature sex
129 was determined based on coloration, then fish were euthanized with MS222, and a fin
130 clip was taken and stored in 98% ethanol for genetic analyses. Genomic DNA of 218
131 F2 progeny, the four F1 parents, and the two F0 grandparents was extracted using
132 phenol-chloroform. Restriction-site associated DNA (RAD) sequencing libraries were
133 prepared following Marques *et al.* (2016) using a protocol slightly modified from
134 Baird *et al.* (2008). In brief, genomic DNA was digested with *Sbf*I followed by
135 shearing and size selection of 300 to 500 bp. Equimolar proportions of DNA from 11
136 to 48 individuals carrying different barcode sequences were pooled into one library.
137 Each library was amplified in four reactions of 50 µl aliquots. A total of nine libraries
138 were single-end sequenced (100 bp) each on a single lane of an Illumina HighSeq
139 2500 platform either at the Next Generation Sequencing Platform of the University of
140 Bern or at the Genomic Technologies Facility of the University of Lausanne. Some
141 individuals and all F0 grandparents were sequenced in up to three libraries to increase
142 coverage. Together with each library, we sequenced about 10% reads of
143 bacteriophage PhiX genomic DNA (Illumina Inc.) to increase complexity at the first
144 10 sequenced base pairs. During read processing, PhiX reads were further utilized to
145 recalibrate libraries to equalize base quality scores across Illumina lanes utilizing
146 GATK version 3.2 (McKenna *et al.* 2010).
147

148 **Sequence processing and genotyping**

149 Before recalibration, read qualities were inspected using fastQC
150 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and filtered using
151 FASTX Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) requiring
152 a minimum quality of 10 at all bases and of 30 in at least 95% of the read. After PhiX
153 removal, reads were demultiplexed, cleaned, and trimmed to 92 bp with
154 process_radtags implemented in Stacks v1.26 (Catchen *et al.* 2013). Reads were
155 mapped against the *P. nyererei* reference genome (Brawand *et al.* 2014) using
156 bowtie2 version 2.2.6 (Langmead and Salzberg 2012). Mapped reads of individuals
157 run in multiple libraries were merged using Picard tools version 1.97 and filtered for a
158 mapping quality of at least 30. After the filtering pipeline we were left with a total of
159 719,720,265 sequences across the nine RAD libraries (on average 79,970,000 reads

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

160 per library). For the female and male parental samples, 1.364,225 and 6,459,242 reads
161 respectively, were mapped and remained after filtering. For the 222 progeny
162 individuals (including the F1) we obtained on average 2,008,826 reads per individual.
163 All 224 individuals (218 F2, the two grandparents and four F1) were genotyped using
164 freebayes version 1.0.0 (Garrison and Marth 2012). As a first filter, sites were kept if
165 bi-allelic, had less then 50% missing data, a quality of more than 2, a minor allele
166 frequency of more than 5%, and a minimal depth of 3. Utilising a script established to
167 filter freebayes genotype calls based on RAD sequencing
168 (https://github.com/jpuritz/dDocent/blob/master/scripts/dDocent_filters), genotypes
169 were further excluded (thresholds given in brackets) on criteria related to allelic
170 balance at heterozygote sites (< 0.28 allele balance between reads), quality versus
171 depth (ratio <0.5), strand presentation (overlapping forward and reverse reads), and
172 site depth (one standard deviation from mean and a quality score lower than twice the
173 depth first, followed by an additional maximum mean depth cutoff of 67). Multi-
174 allelic variants and indels were removed, resulting in 7,401 SNPs. Lastly, the 2,052
175 SNPs that were differentially fixed homozygote genotypes in the grandparents were
176 used for creating the linkage map.
177

178 **Linkage map**

179 A linkage map was constructed with JoinMap 4.0 (Van Ooijen 2006) using 212 F2
180 progeny derived from two F1 families. Out of the 224 genotyped individuals
181 (including the 2 F0 and 4 F1), 2 F1 and 6 F2 were removed due to missing data (>
182 25%). Out of the 2,052 loci, homozygous for alternative alleles in the grandparents,
183 we placed 1,597 in the final linkage map. Loci were excluded if positioned identical
184 with another locus. Markers showing segregation distortion ($\chi 2$ test, $P < 0.001$) were
185 excluded for linkage map reconstruction. Linkage groups were identified based on an
186 independent logarithm of odds (LOD) threshold of 12. Unlinked markers were
187 excluded. The strongest cross-link (SCL) in the final map is 5.4. The linkage map was
188 built using the regression mapping algorithm, a recombination frequency smaller than
189 0.40, and a LOD larger than 3. Up to three rounds of marker positioning were
190 conducted with a jump threshold of 5. A ripple was performed after the addition of
191 each new marker. Map distances were calculated using the Kosambi mapping
192 function. All markers resolved onto 22 linkage groups were matched to positions in
193 the *Oreochromis niloticus* genome using a chain file (Brawand *et al.* 2014) with

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

194 liftover (UCSC Genome Browser LiftOver tool; Hinrichs *et al*. 2006) to examine
195 synteny of chromosomal locations and allow comparisons with other published
196 studies.
197
198 **Anchoring of reference scaffolds**
199 In order to reconstruct a chromosomal reference genome for *Pundamilia*, we used the
200 linkage map to anchor the scaffolds of the *Pundamilia* genome from Brawand *et al*.
201 (2014) onto the 22 *Pundamilia* linkage groups (Pun-LGs) identified during mapping
202 (see paragraph above). We ordered and oriented the scaffolds with ALLMAPS (Tang
203 *et al*. 2015). Gaps between the scaffolds were then estimated using interpolated
204 recombination rate estimates based on the conversion between map distances (cM)
205 and physical distances (bp) as implemented in the ALLMAPS function
206 "estimategaps" (Tang *et al*. 2015). In addition to an improved reference version,
207 resolving linkage groups, we compiled a chain file for converting positions on the
208 original *Pundamilia nyererei* reference (Brawand *et al*. 2014) to our new reference
209 (*Pundamilia* reference version 2.0) with ALLMAPS and in the opposite direction
210 using chainSwap from kentUtils (https://github.com/ENCODE-DCC/kentUtils). We
211 could then use the chain file to liftover the position of all 7,401 genotyped loci, using
212 Picard liftoverVcf (http://broadinstitute.github.io/picard/index.html). In addition, we
213 generated a new version of the NCBI *Pundamilia nyererei* RefSeq annotation file
214 with the positions for reference version 2.0 by lifting over the positions from the
215 NCBI PunNye1.0 annotation release 101
216 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Pundamilia_nyererei/101/#Bu
217 ildInfo) using the UCSC liftOver tool (Hinrichs *et al*. 2006) and custom-made chain
218 files (see Table 2). By comparing physical (bp) and recombination distances (cM), we
219 estimated recombination rates along the different linkage groups. First, we pruned the
220 linkage map for markers generating negative recombination rates and markers that
221 were less than 20 kb apart. Then we fitted a cubic smoothing spline to the physical
222 (bp) and recombination (cM) distances using the R function "smooth.spline" setting
223 the smoothing parameter (spar) to 0.7 and inferred the recombination positions in cM
224 for the genomic positions as the first derivative of the "predict.smooth.spline"
225 function.
226
227

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

228 **QTL mapping of sex**

229 QTL mapping of the sex-determining region was performed with Rqtl (Broman *et al.*
230 2003) based on 209 individuals (3 F2 were discarded prior to analysis as they were
231 juveniles) and 1,597 SNP markers. 137 males and 72 females were included. Sex was
232 mapped by standard interval mapping as a binary trait and significance was
233 determined by permutation (n = 1000). Bayesian confidence intervals were estimated
234 as implemented in Rqtl and the highest LOD score was used to calculate the percent
235 variance explained following $1 - 10^{-2\ \mathrm{LOD}\ /\ n}$ (Broman and Sen 2009). Plotting
236 phenotypic sex against the genotypes for the marker most strongly associated with
237 sex, revealed two individuals labeled as females, but carrying a male genotype. Those
238 individuals were dissected and their gonads were inspected, showing immature or
239 undeveloped gonads indicating an error in phenotyping. The same plot also revealed
240 both males (n = 74) and females (n= 32) that were heterozygous at the locus strongly
241 associated with sex. To investigate if sex in those individuals was explained by
242 another locus, we extracted the genotypes of these individuals and repeated the
243 interval mapping. Further, we made use of 366 markers positioned on the linkage
244 group containing the sex QTL and investigated segregation patterns at those loci in
245 more detail in the larger of our mapping families (n = 122 F2 offspring). Based on the
246 improved, annotated reference (v.2.0) we determined the number of annotated genes
247 in the QTL interval and screened for candidate genes in sex determination.

248

249 **Data availability**

250 All genomic resources (see Table 2) will be made available upon publication. Raw
251 read sequencing files will be deposited on short read archive (fastq files for all 224
252 individuals). Genotype (vcf format) and phenotype file will also be made available.

253

254 **Results and Discussion**

255 **Linkage map**

256 The linkage map comprises 22 linkage groups containing 1,597 markers with an
257 average marker distance of 1.01 cM adding up to a total map length of 1593.72 cM
258 (Figure 1, Table 1). It is slightly longer than other maps published on Lake Victoria
259 cichlids (1130.63 cM in Henning *et al.* 2014, 1133.2 cM in Kudo *et al.* 2015 and
260 1225.68 cM in Henning *et al.* 2017), but contains more markers with a lower average
261 marker distance (1.30 cM (Henning *et al.* 2014) 1.09 cM (Henning *et al.* 2017) and

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

262 6.09 cM (Kudo *et al*. 2015)). The detection of 22 linkage groups is consistent with the
263 expected number of chromosomes in haplo-tilapiine cichlids (Guyon *et al*. 2012). Out
264 of 1,597 markers used to build the *Pundamilia* linkage map, 1,182 markers could be
265 positioned onto *Oreochromis niloticus* linkage groups (Ore-LG). Figure 2 reveals
266 extensive synteny between the chromosomes of these distantly related cichlid species.
267 The linkage map presented here will facilitate comparative genomics and will enable
268 comparisons of previous QTL results with newly established results (for an example
269 see paragraph below on QTL for sex-determination) using Ore-LGs as a reference
270 point.
271
272 **Improvement of the genomic resources for Lake Victoria cichlids (*Pundamilia*)**
273 The *Pundamilia* linkage map provides a new chromosome framework for whole
274 genome sequence assembly and map integration with more anchoring points then
275 previous published maps. The anchored genome encompasses 78.7% of the total
276 bases (653,642,680 bp) of the original *P. nyererei* reference genome based on 383
277 anchored scaffolds, of which 233 are now oriented. This is a slightly higher fraction
278 than in the Lake Malawi cichlid *Metriaclima zebra*, where 564,259,264 bp (66.5%) of
279 the genome sequence could be anchored to linkage groups (O'Quin *et al*. 2013). The
280 mean marker density is 2.4 per megabase (Mb). The 6,853 remaining scaffolds could
281 not be anchored due to lack of informative markers. This improved resolution of the
282 new reference assembly (v2.0) will greatly facilitate genome scan approaches in Lake
283 Victoria cichlids. Such approaches rely on the information from neighboring genomic
284 positions to identify signatures of selection due to genetic hitchhiking. Any
285 approaches evaluating or making use of linkage information, like linkage
286 disequilibrium (LD) based genome scans, association studies or evaluations of the
287 genomic landscape of divergence will now become feasible or more powerful.
288 Together with the improved reference, we provide chain files to liftover positions
289 from the previous version (v1.0) to the new chromosome level resolved reference
290 version (v2.0). We further provide a matching annotation file based on the NCBI
291 annotation (see Table 2 for a complete list of all genomic resources). Finally, we
292 estimated recombination rates and show that those are highly variable across the
293 genome ranging from 0 to 9.4 cM/Mb (Table 2), with a mean recombination rate of
294 2.3 cM/Mb. Knowledge of fine-scale patterns of recombination rate variation (see

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

295 Figure 4C) will be useful for future studies of adaption and speciation (Stapley *et al*.
296 2017) in the exceptional species radiation of Lake Victoria cichlids.
297
298 **Characterization of sex-determination in *Pundamilia***
299 Our knowledge of sex determination in *Pundamilia*, a prime model system of
300 sympatric speciation in Lake Victoria, had been limited. Here, we mapped sex to Pun-
301 LG10, which is homologous with Ore-LG23 (Figure 3; $p << 0.001$, LOD = 26.5). We
302 did not find any further associations on any of the other LGs. Ore-LG23 has been
303 previously identified as one potential sex-determining LG in *Oreochromis* (Cnaani *et*
304 *al*. 2008; Palaiokostas *et al*. 2013) and in four cichlid tribes from Lake Tanganyika
305 overexpression of male specific genes accumulate on Ore-LG23 (Böhne et al. 2014).
306 Early work on sex determination in Lake Victoria cichlids had suggested
307 polymorphisms at several unlinked genomic regions to be associated with sex, and
308 invoked a major effect locus and some modifiers (Seehausen et al. 1999). Recent QTL
309 mapping identified genomic regions involved in sex determination in Lake Victoria
310 cichlids on Ore-LG5 and Ore-LG2 (Kudo *et al*. 2015) or on derived, female specific B
311 chromosomes (Yoshida *et al*. 2011). Ore-LG5 was repeatedly found to be involved in
312 sex-determination in other cichlids, e.g. in the riverine haplochromine cichlids
313 *Astatotilapia burtoni* and *Astatotilapia calliptera* from Lakes Tanganyika and Malawi
314 and associated rivers (Roberts *et al*. 2016; Böhne *et al*. 2016; Peterson *et al*. 2017), in
315 *Cyprichromis leptosoma* from Lake Tanganyika (Gammerdinger *et al*. 2018) and in
316 *Labeotropheus trewavasae* and across some *Metriaclima* species from Lake Malawi
317 (Ser *et al*. 2010; Parnell and Streelman 2013).
318
319 The mapping interval (Bayesian confidence interval of 5.7 cM, 21.7 to 27.4 cM;
320 Figure 4B) in total covers four markers and spans a region of ~1.9 Mb (Figure 4C).
321 The marker showing the strongest association with sex in our study (Figure 4A)
322 explains 44% of the phenotypic variance in sex. Sex is not entirely explained by this
323 marker as we had misidentified two likely sub-adults (gonads appear not developed at
324 hindsight inspection) as females, and due to 106 individuals, both males and females,
325 which are heterozygous at this position (Figure 4A). Repeating the mapping
326 procedure for those individuals again identified a region on Pun-LG10 (Ore-LG23) as
327 weakly associated with sex ($p = 0.177$, LOD = 3.33, position right to previous interval
328 at 28.8 cM). This suggest that none of the markers used to build the linkage map is

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

329 determining sex directly, but that the causal locus can be found close by and indicates
330 that there are no further major genetic determines of sex segregating in this cross.
331 Investigating the segregation patterns in the larger of the F2 mapping-families (n =
332 122) more in detail, the loci selected to built the map, reciprocal homozygous in F0
333 female (AA) and male (BB) and heterozygous in both F1 (AB), segregate as expected
334 in a 50:50 ratio of AA:AB in F2 females and AB:BB in F2 males (Figure 5A).
335 However, evaluating segregation patterns of the additional markers genotyped but not
336 used for the construction of the linkage map, indicate that the sex determination
337 system on Pun-LG10 is male heterogametic (XY, Figure 5B). We identified 57 loci
338 between 0 and 35 Mb that were homozygous in the F0 and F1 females and
339 heterozygous in the F0 and F1 males; these markers are similarly homozygous in all
340 F2 females and heterozygous in all F2 males, consistent with females being XX and
341 males being XY (Figure 5B, the plot also shows 13 loci > 35 Mb). Additional
342 evidence comes from markers heterozygous in the F0 female (AB) and homozygous
343 in the F0 male (BB), for which we find all 35 loci for positions < 33 Mb heterozygous
344 (AB) for both F1 individuals. The heterozygous loci in both F1 are a segregation
345 pattern only consistent with male heterogametic (XY) segregation. If females would
346 be heterogametic (ZW) those loci would need to be homozygous (BB) in one of the
347 F1 and not heterozygous (AB) in both as observed in that 33 Mb region. Sex-averaged
348 recombination rates around the QTL are low and even close to zero within 20 Mb
349 proximity to the mapping interval (Figure 4C). Such a pattern, potentially due to
350 suppressed recombination in the heterogametic sex (males), might indicate initial
351 steps toward the evolution of a heteromorphic (degenerated) sex (Y) chromosome
352 (Charlesworth, 1991).
353
354 Within our mapping interval of ~1.9 Mb, 65 genes, based on the NCBI annotation for
355 the new *Pundamilia* reference assembly, can be found (Table S1). Among them is the
356 anti-müllerian hormone (*amh)*, a master gene for sex determination in other fish. *Amh*
357 is part of the transforming growth factor beta pathway, responsible for the regression
358 of Müllerian ducts in tetrapods (Josso *et al*. 2001). Even though teleost fish do not
359 have Müllerian ducts, the *amh* pathway has a prominent role in sex determination for
360 several distantly related fish species. In the Japanese pufferfish (*Takifugu rubripes*), a
361 mutation in the receptor of the *amh* (amhrII) determines sex (Kamiya *et al*. 2012). The
362 *amhy* (Y chromosome-specific anti-müllerian hormone) gene has been inserted

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

363 upstream of *amh* in the cascade of male development in the neotropical silverside
364 *Odonthestes hatcheri* (Hattori *et al*. 2012). Similarly, in *Oreochromis niloticus*, a Y-
365 linked duplicate of *amh* acts as a major sex determination locus (Eshel *et al*. 2012; Li
366 *et al*. 2015). In *Oryzias luzonensis*, a mutation of an *amh* related ligand gsdf^{Y} is
367 responsible for sex determination (Myosho *et al*. 2012). The same ligand is suggested
368 to be involved in sex determination in the haplochromine cichlid *Astatotilapia*
369 *calliptera* (Peterson *et al*. 2017). Beside the two master sex determination genes in
370 *Oreochromis niloticus* on LG23 (*amh*) and in *Astatotilapia calliptera* on LG7 (*gsdf*)
371 (Peterson *et al*. 2017), candidates for sex determination in cichlids have not been
372 shown to be directly involved in sex determination in other species (Heule *et al.*
373 2014b, Böhne *et al.* 2016, Gammerdinger *et al.* 2018, but see Böhne *et al*. 2014).
374 They might in this matter act as so-called "newcomers" (Herpin and Schartl 2015).
375 Our results indicate that in the Lake Victoria cichlid *Pundamilia* Pun-LG10 (Ore-LG
376 23) acts as an (evolving) sex chromosome, even though it might not be the only
377 region controlling sex in *Pundamilia*. The anti-müllerian hormone *amh* (or a derived
378 copy) appears to be a very good candidate influencing sexual development in
379 *Pundamilia*, but further work is warranted to characterize the genomic candidate
380 region and the impact of this candidate gene on sex determination.
381
382 A recent meta-analysis showed that transitions between sex determination systems are
383 frequent across various fish species, including transitions to and between
384 heteromorphic sex chromosomes (Pennell *et al.* 2018). In cichlids a high turnover of
385 sex determination systems was described in Lake Malawi (Ser *et al*. 2010), Lake
386 Tanganyika (Böhne *et al*. 2014; Gammerdinger *et al*. 2018), and oreochromine
387 cichlids (Cnaani *et al*. 2008). The circumstance that *amh*, Pun-LG10 or a homologous
388 region was not invoked in sex determination in other Lake Victoria cichlids that have
389 previously been used for mapping sex (Kudo *et al*. 2015; Yoshida *et al.* 2011) implies
390 that multiple sex determining systems segregate among the species of Lake Victoria
391 cichlid fish as well. This is consistent with early work on sex determination in this
392 group (Seehausen *et al*. 1999). Given the extreme youth of the Lake Victoria species
393 radiation (~15,000 years; Seehausen 2006), this may be surprising at first. Recent
394 work, however, has shown that much of the genetic variation in the radiation is much
395 older than the species radiation and took its origin in a hybridization event between
396 two anciently divergent cichlid lineages from which all 500+ species of the radiation

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

397 evolved (Meier *et al*. 2017a). It is tempting to speculate that the variation in sex
398 determination systems between and within species of this radiation traces its roots to
399 these ancient lineages too, something that can now be tested.
400
401
402

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

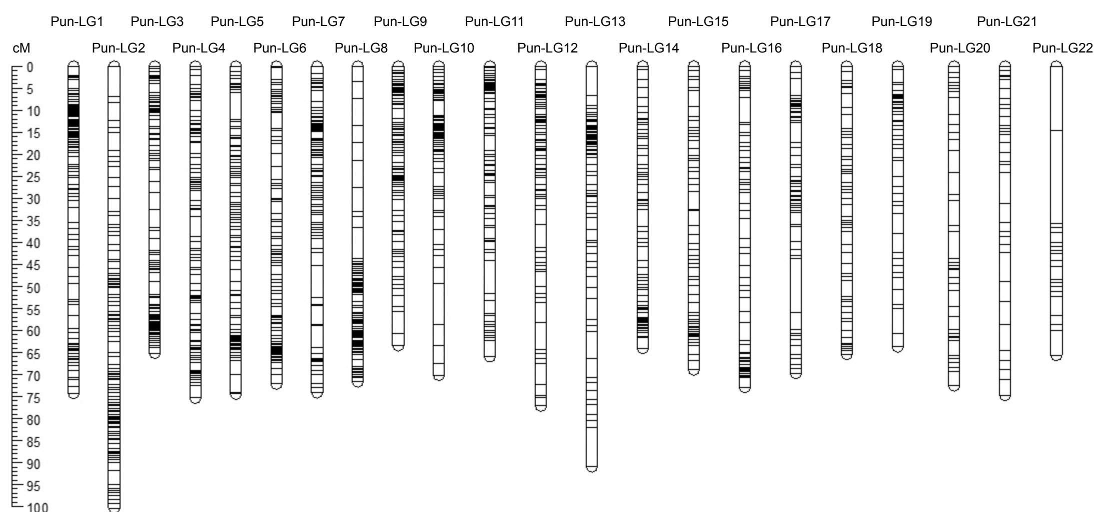
403 **Tables and Figures**

404

405 Figure 1: Linkage map indicating the positioning of 1,597 markers and Kosambi

406 mapping length (cM) of 22 linkage groups.

407



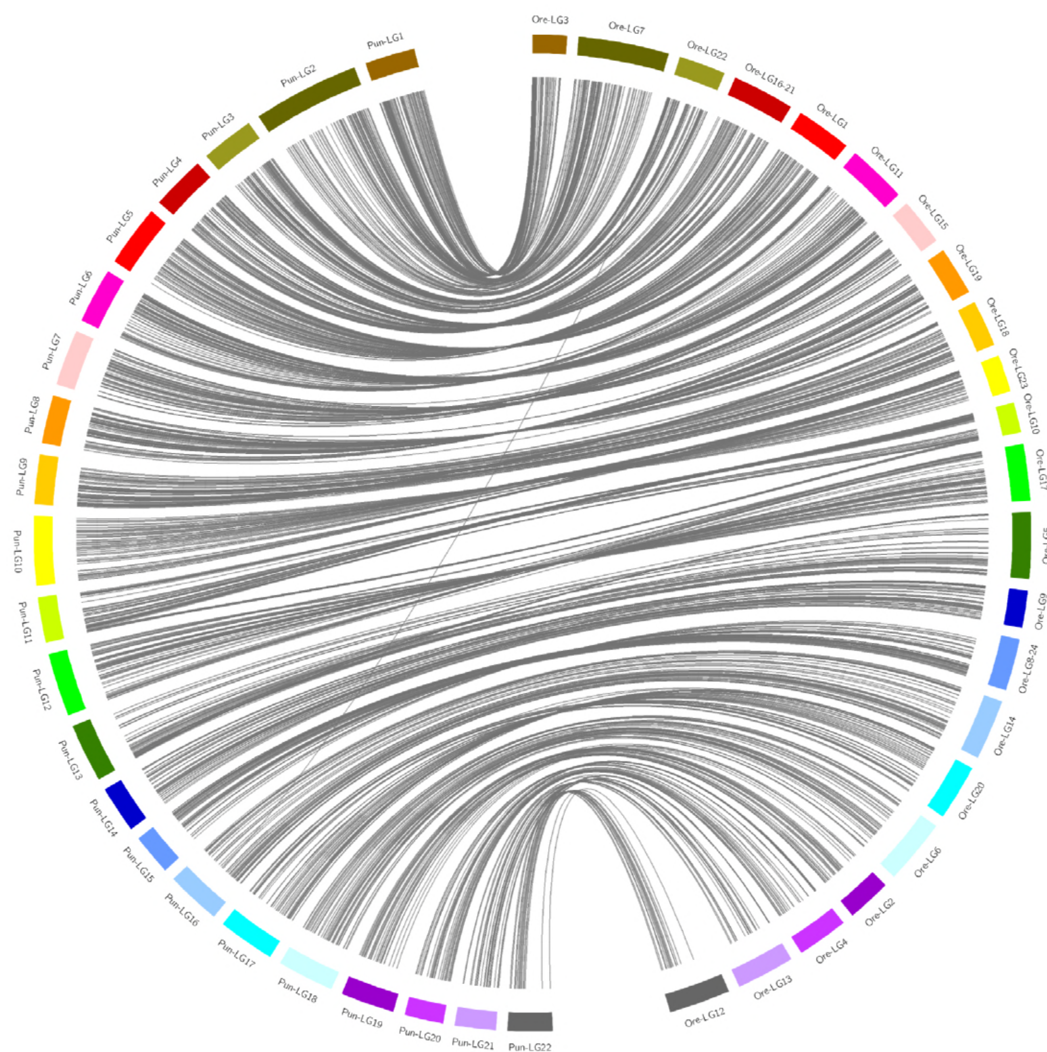
408

409

410

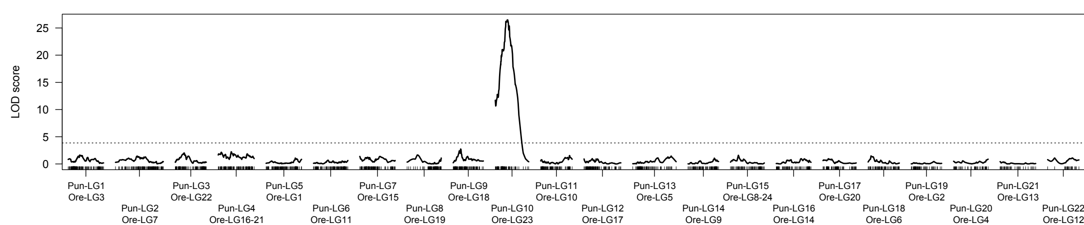
bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2: Synteny plot showing the correspondence of *Pundamilia* linkage groups (Pun-LG) with *Oreochromis niloticus* linkage groups (Ore-LG). Lines indicated markers used in linkage map construction, which could be positioned in the *Pundamilia* reference (v2.0) and lifted over to the *Oreochromis* reference.



bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

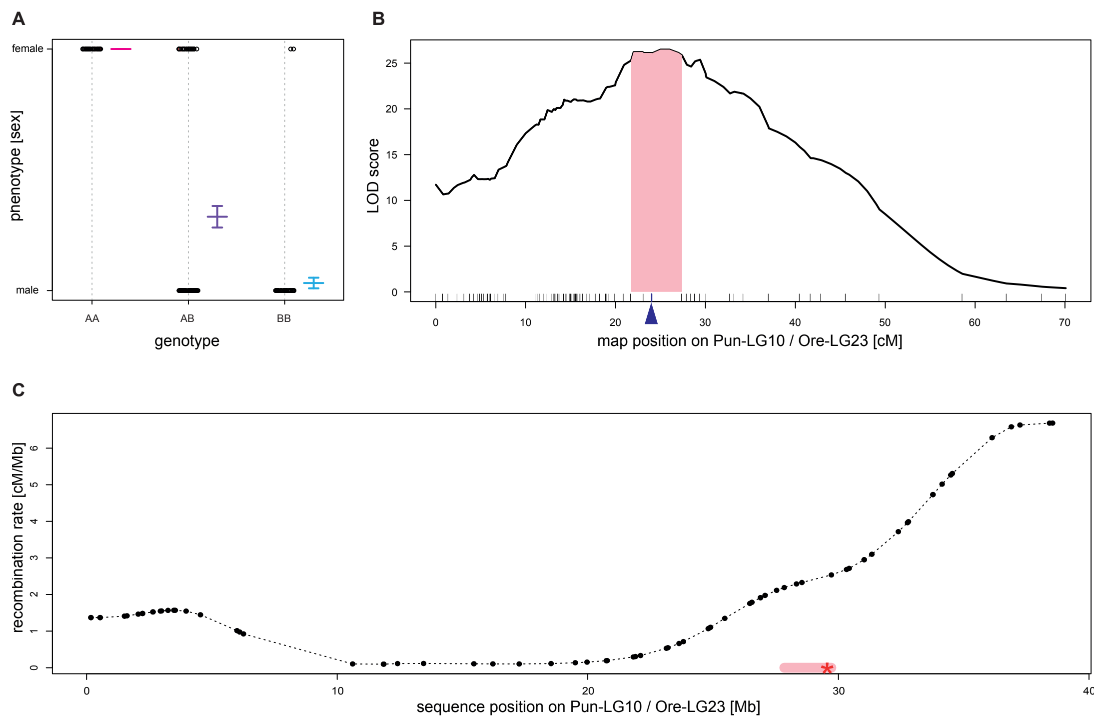
419 Figure 3: QTL mapping of sex. LOD scores across the 22 linkage groups are shown.
420 Genome-wide significance levels are indicated by horizontal lines (alpha = 0.05
421 dotted line). Marker loci are indicated along the x-axis.



422
423
424

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

425 Figure 4: A) Phenotypic effect of genotypes at the locus most strongly associated with
426 sex. The plot identifies two females as likely phenotypic errors and 106 individuals
427 heterozygote at that locus. B) Plot of LOD scores indicating the region of strongest
428 association with sex on Pun-LG10 (Ore-LG23). The Bayesian confidence interval is
429 highlighted in light pink. Marker loci are indicated along the x-axis. The locus shown
430 in panel A is indicated by a blue arrow. C) Variation in recombination rates (sex-
431 averaged) along Pun-LG10 (Ore-LG23). The Bayesian confidence interval (pink
432 highlight) is situated next to a region of low recombination. The red star indicates the
433 position of *amh* (candidate gene for sex determination).
434

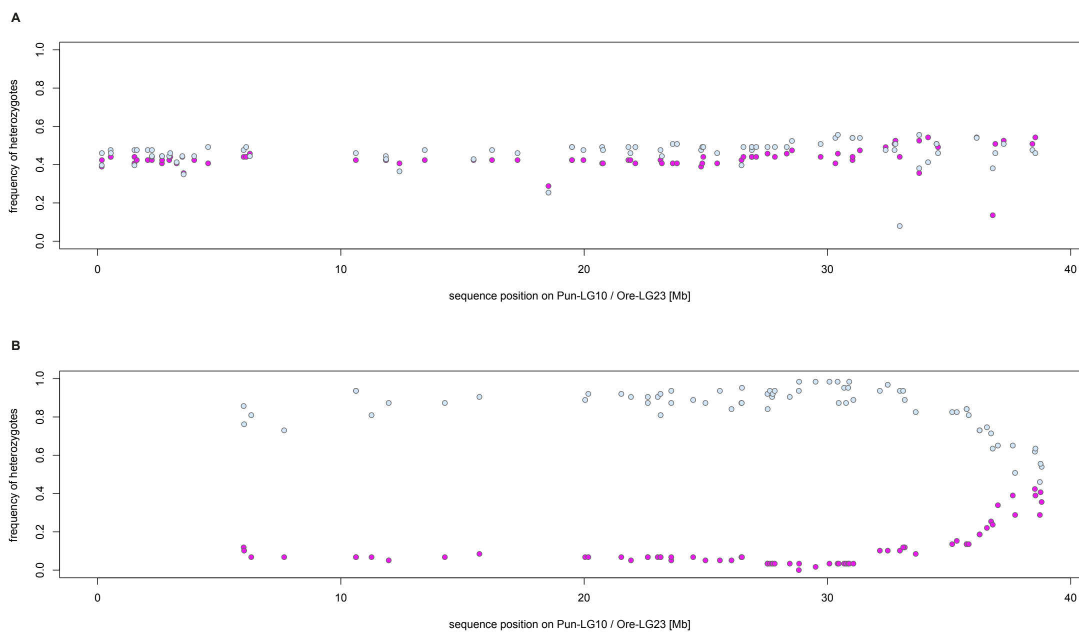


435
436
437

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438 Figure 5: Frequency of heterozygote individuals (n = 122), separated by sex (63
439 males: light blue, 59 females: light pink) for markers selected by their segregation
440 pattern in the larger mapping family and their position on Pun-LG10 (Ore-LG23). A)
441 78 markers, selected as reciprocally homozygous (AA/BB) in the F0 and heterozygote
442 in both F1 (AB/AB), segregate as expected in a 50:50 ratio of AA:AB in F2 females
443 and AB:BB in F2 males, resulting in frequency of heterozygous F2 individuals around
444 0.5 for both sexes. B) 70 markers, selected as homozygote in the F0 and F1 females
445 and heterozygote in the F0 and F1 males, segregate similarly in the F2; i.e. the
446 frequency of heterozygous individuals is approaching 0 in females and 1 in males for
447 positions < 35 Mb.

448



449
450
451

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

452 Table 1: Summary of length and number of markers for each linkage group. Synteny
453 between this study (Pun-LG) and the *Oreochromis niloticus* reference (Ore-LG) is
454 indicated.

***Pundamilia* LG**	***Oreochromis* LG**	**length [cM]**	**# SNPs**
Pun-LG1	Ore-LG3	74.182	120
Pun-LG2	Ore-LG7	100.271	103
Pun-LG3	Ore-LG22	65.086	100
Pun-LG4	Ore-LG16-21	75.325	94
Pun-LG5	Ore-LG1	74.372	90
Pun-LG6	Ore-LG11	72.082	90
Pun-LG7	Ore-LG15	74.093	88
Pun-LG8	Ore-LG19	71.579	88
Pun-LG9	Ore-LG18	63.352	82
Pun-LG10	Ore-LG23	70.089	77
Pun-LG11	Ore-LG10	65.94	76
Pun-LG12	Ore-LG17	77.124	75
Pun-LG13	Ore-LG5	90.878	74
Pun-LG14	Ore-LG9	63.956	71
Pun-LG15	Ore-LG8-24	70.914	61
Pun-LG16	Ore-LG14	72.914	60
Pun-LG17	Ore-LG20	69.652	58
Pun-LG18	Ore-LG6	65.41	54
Pun-LG19	Ore-LG2	63.732	44
Pun-LG20	Ore-LG4	72.481	39
Pun-LG21	Ore-LG13	74.665	33
Pun-LG22	Ore-LG12	65.627	20
		1593.724	1597

455

456 Table 2: List of genomic resources provided with this manuscript

Type	**Name**	**Source**
Text file giving the position of 1,597 loci on the *Pundamilia* linkage map and the respective positions on *Pundamilia* and *Oreochromis* references	P_cross.MarkerPositions.txt	**XXX**
Fasta file of the improved *Pundamilia* reference genome (v2.0)	P_nyererei_v2.fasta.gz	**XXX**
Chain files to convert position between original (v1.0) and new reference (v2.0)	P_nyererei_v1.To.P_nyererei_v2.chain, P_nyererei_v2.To.P_nyererei_v1.chain	**XXX**
Annotation file matching reference v2.0 position based on NCBI annotation release 101	P_nyererei_v2.gff.gz	**XXX**
Text file giving the extrapolated recombination	P_nyererei_v2.RecRates.txt	**XXX**

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rates along *Pundamilia* reference genome

457

458 Table S1: List of genes overlapping the mapping interval on Pun-LG10 (Ore-LG23).
459 Gene position as annotated in *Pundamilia* (v2.0) annotation file based on NCBI
460 annotations using Uni-Prot.

461

462

463 **Acknowledgements**

464 We thank Katie Peichel for her critical input during the analysis of the data and
465 manuscript preparation. We thank Andreas Taverna, Alan Smith, and Ola Svensson
466 for fish maintenance and breeding. We thank David Marques for lifting over the
467 annotation file. We thank the team from the Genetic Diversity Center (GDC) at ETH
468 Zürich for bioinformatics support. Genomic analyses were performed using the
469 computing infrastructure of the GDC and the Euler computer cluster at ETH Zurich.
470 JS was supported by the German Research Foundation grant SCHW-1690/1-1.

471

472

473 **Author's contributions**

474 PGDF, JS, MPH, and JIM performed the experiment and the analysis. OS conceived
475 the original idea and supervised the project. PGDF and JS took the lead in writing the
476 manuscript. All authors provided critical feedback and helped shape the research,
477 analysis, and manuscript.

478

479

480 **References**

481

482 Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver *et al.*, 2008 Rapid
483 SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:
484 e3376.

485

486 Böhne, A., T. Sengstag, and W. Salzburger, 2014 Comparative transcriptomics in East
487 African cichlids reveals sex-and species-specific expression and new candidates for
488 sex differentiation in fishes. Genome Biol. Evol. 6: 2567-2585.

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

489

490 Böhne, A., C. A. Wilson, J. H. Postlethwait, and W. Salzburger 2016 Variations on a
491 theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. BMC
492 Genom. 17:883

493

494 Brawand, D., C.E. Wagner, Y.I. Li, M. Malinsky, I. Keller *et al.*, 2014 The genomic
495 substrate for adaptive radiation in African cichlid fish. Nature 513: 375–81.

496

497 Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in
498 experimental crosses. Bioinf. 19: 889-890.

499

500 Broman, K. W., and S. Sen, 2009 A Guide to QTL Mapping with R/qtl (Vol. 46).
501 Springer, New York.

502

503 Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko, 2013
504 Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22: 3124-
505 3140.

506

507 Charlesworth, B., 1991 The evolution of sex chromosomes. Science 251: 1030-1033.

508

509 Cnaani, A., B.-Y. Lee, N. Zilberman, C. Ozouf-Coastaz, G. Hulata *et al.*, 2008
510 Genetics of sex determination in tilapiine species. Sex. Dev. 2: 43-54.

511

512 Eshel, O., A. Shirak, J.I. Weller, G. Hulata, and M. Ron, 2012 Linkage and physical
513 mapping of sex region on LG23 of Nile Tilapia (*Oreochromis niloticus*). G3 2: 35–42.

514

515 Eshel, O., A. Shirak, L. Dor, M. Band, T. Zak *et al.*, 2014 Identification of male-
516 specific amh duplication, sexually differentially expressed genes and microRNAs at
517 early embryonic development of Nile tilapia (*Oreochromis niloticus*). BMC Genom.
518 15: 774.

519

520 Fierst, J. L., 2015 Using linkage maps to correct and scaffold de novo genome
521 assemblies: methods, challenges, and computational tools. Front. Genet. 6: 220.

522

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

523 Gammerdinger, W. J., M. A. Conte, B. A. Sandkam, A. Ziegelbecker, S. Koblmüller,
524 and T.D. Kocher, 2018 Novel sex chromosomes in three cichlid fishes from Lake
525 Tanganyika. J. Hered. 1: 12.
526
527 Garrison E, and G. Marth, 2012 Haplotype-based variant detection from short-read
528 sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN] 2012
529
530 Greenwood, P.H., 1974 Cichlid fishes of Lake Victoria, East Africa: the biology and
531 evolution of a species flock. Bull. Brit. Mus. Zool. Suppl. 6: 1-134.
532
533 Guyon, R., M. Rakotomanga, N. Azzouzi, J.P. Coutanceau, C. Bonillo *et al.*, 2012 A
534 high-resolution map of the Nile tilapia genome: a resource for studying cichlids and
535 other percomorphs. BMC Genom. 13: 222.
536
537 Hattori, R.S., Y. Murai, M. Oura, S. Masuda, S.K. Majhi *et al.*, 2012 A Y-linked anti-
538 Müllerian hormone duplication takes over a critical role in sex determination. Proc.
539 Natl. Acad. Sci. USA 109: 2955–2959.
540
541 Henning, F., H. J. Lee, P. Franchini, and A. Meyer, 2014 Genetic mapping of
542 horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD
543 markers for dense linkage mapping. Mol. Ecol. 23: 5224-5240.
544
545 Henning, F., G. Machado‐Schiaffino, L. Baumgarten, and A. Meyer, 2017 Genetic
546 dissection of adaptive form and function in rapidly speciating cichlid fishes. Evolution
547 71: 1297-1312.
548
549 Herpin, A., and M. Schartl, 2015. Plasticity of gene‐regulatory networks controlling
550 sex determination: of masters, slaves, usual suspects, newcomers, and usurpators.
551 *EMBO Reports* 16: 1260-1274.
552
553 Heule, C., W. Salzburger, and A. Böhne, 2014a Genetics of sexual development – an
554 evolutionary playground for fish. Genetics 196: 579–91.
555

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

556 Heule, C., C. Göppert, W. Salzburger, and A. Böhne, 2014b Genetics and timing of
557 sex determination in the East African cichlid fish Astatotilapia burtoni. BMC Gen. 15:
558 140.
559
560 Hinrichs AS, D. Karolchik, R. Baertsch G. P. Barber, G. Bejerano *et al.*, 2006 The
561 UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 1; 34
562
563 Holzberg, S., 1978 A field and laboratory study of the behaviour and ecology of
564 *Pseudotropheus zebra* (Boulenger), an endemic cichlid of Lake Malawi (Pisces;
565 Cichlidae). J. Zool. Syst. Evol. Res. 16: 171-187.
566
567 Josso, N., N. di Clemente, and L. Gouédard, 2001 Anti-Müllerian hormone and its
568 receptors. Mol. Cell. Endocrinol. 179: 25–32.
569
570 Kamiya, T., W. Kai, S. Tasumi, A. Oka, T. Matsunaga *et al.*, 2012 A trans-species
571 missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish,
572 *Takifugu rubripes* (fugu). PLoS Genet. 8: e1002798.
573
574 Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz *et al.*, 2013 Population
575 genomic signatures of divergent adaptation, gene flow and hybrid speciation in the
576 rapid radiation of Lake Victoria cichlid fishes. Mol. Ecol. 22: 2848-2863.
577
578 Kikuchi, K., and S. Hamaguchi, 2013 Novel sex-determining genes in fish and sex
579 chromosome evolution. Dev. Dyn. 242: 339–53.
580
581 Kocher, T.D., 2004. Adaptive evolution and explosive speciation: the cichlid fish
582 model. Nat. Rev. Gen. 5: 288-298.
583
584 Kudo, Y., M. Nikaido, A. Kondo, H. Suzuki, K. Yoshida *et al.*, 2015 A microsatellite-
585 based genetic linkage map and putative sex-determining genomic regions in Lake
586 Victoria cichlids. Gene. 560: 156–64.
587
588 Lande, R., O. Seehausen, and J. J. Van Alphen, 2001 Mechanisms of rapid sympatric
589 speciation by sex reversal and sexual selection in cichlid fish. Genetica 112: 435-443.

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

590
591 Langmead, B., and S. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2.
592 Nature Met. 9: 357-359.
593
594 Li, M., Y. Sun, J. Zhao, H. Shi, S. Zeng *et al.*, 2015 A tandem duplicate of Anti-
595 Müllerian hormone with a missense SNP on the Y chromosome is essential for male
596 sex determination in Nile Tilapia, *Oreochromis niloticus*. PLoS Gen. 11: e1005678.
597
598 Maan, M. E., and O. Seehausen, 2011 Ecology, sexual selection and speciation. Ecol.
599 Let. 14: 591-602.
600
601 Marques, D.A., K. Lucek, J. I. Meier, S. Mwaiko, C.E. Wagner *et al.*, 2016 Genomics
602 of rapid incipient speciation in sympatric threespine stickleback. PLoS Genet. 12:
603 e1005887.
604
605 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010. The
606 Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation
607 DNA sequencing data. Genome Res. 20: 1297-1303.
608
609 Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier *et al.*, 2017a
610 Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8:
611 14363.
612
613 Meier, J. I., V. C. Sousa, D. A. Marques, O. M. Selz, C. E. Wagner *et al.*, 2017b
614 Demographic modelling with whole‐genome data reveals parallel origin of similar
615 *Pundamilia* cichlid species after hybridization. Mol. Ecol. 26: 123-141.
616
617 Meier, J. I., D. A. Marques, C. E. Wagner, L. Excoffier, and O. Seehausen, 2018
618 Genomics of parallel ecological speciation in Lake Victoria cichlids. Mol. Biol. Evol.
619 accepted
620

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

621 Myosho, T., H. Otake, H. Masuyama, M. Matsuda, Y. Kuroki *et al*., 2012 Tracing the
622 emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics
623 191: 163–170.
624
625 Nagl, S., H. Tichy, W. E. Mayer, N. Takezaki, N. Takahata *et al.*, 2000 The origin
626 and age of haplochromine fishes in Lake Victoria, East Africa. Proc. R. Soc. B. 267:
627 1049-1061.
628
629 O'Quin, C., A. C. Drilea, M. A. Conte, and T. D. Kocher, 2013 Mapping of
630 pigmentation QTL on an anchored genome assembly of the cichlid fish, *Metriaclima*
631 *zebra*. BMC Genom. 14: 287.
632
633 Palaiokostas, C., M. Bekaert, M. G. Khan, J. B. Taggart, K. Gharbi *et al*, 2013
634 Mapping and validation of the major sex-determining region in Nile tilapia
635 (*Oreochromis niloticus* L.) using RAD sequencing. PLoS One, 8: e68389.
636
637 Parnell, N. F., and J. T. Streelman, 2013 Genetic interactions controlling sex and color
638 establish the potential for sexual conflict in Lake Malawi cichlid fishes. Heredity,
639 110(3), 239.
640
641 Pennell, M.W., J. E. Mank, and C. L. Peichel, 2018 Transitions in sex determination
642 and sex chromosomes across vertebrate species. Molecular Ecology: in press.
643
644 Peterson, E. N., M.E. Cline, E. C. Moore, N.B. Roberts, and R.B Roberts, 2017
645 Genetic sex determination in *Astatotilapia calliptera*, a prototype species for the Lake
646 Malawi cichlid radiation. Sci. Nat. 104: 41.
647
648 Roberts, R. B., J. R. Ser, and T. D. Kocher, 2009 Sexual conflict resolved by invasion
649 of a novel sex determiner in Lake Malawi cichlid fishes. Science 326: 998-1001.
650
651 Roberts, N.B., S. A. Juntti, K. P. Coyle, B. L. Dumont, M. K. Stanley *et al*, 2016.
652 Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics*
653 17: 835.
654

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

655 Sanetra, M., F. Henning, S. Fukamachi, and A. Meyer, 2009 A microsatellite-based
656 genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a
657 comparison of genomic architectures among rapidly speciating cichlids. Genetics 182:
658 387-397.
659
660 Seehausen, O., 1996 Lake Victoria rock cichlids: taxonomy, ecology and distribution.
661 Verduyn Cichlids, Zevenhuizen.
662
663 Seehausen, O., J. J. Van Alphen, and F. Witte, 1997 Cichlid fish diversity threatened
664 by eutrophication that curbs sexual selection. Science 277: 1808-1811.
665
666 Seehausen, O., and J. J. van Alphen, 1999 Can sympatric speciation by disruptive
667 sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? Ecol.
668 Lett. 2: 262-271.
669
670 Seehausen, O., J. J. M. van Alphen, and R. Lande, R., 1999 Color polymorphism and
671 sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by
672 sexual selection. Ecol. Lett. 2: 367-378.
673
674 Seehausen, O., 2000 Explosive speciation rates and unusual species richness in
675 haplochromine cichlid fishes: effects of sexual selection. Advan. Ecol. Res. 31: 237-
676 274.
677
678 Seehausen, O., E. Koetsier, M. V. Schneider, L. J. Chapman, C. A. Chapman *et al.*,
679 2003 Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic
680 origin of the Lake Victoria cichlid species flock. Proc. R. Soc. B 270: 129-137.
681
682 Seehausen, O., 2006 African cichlid fish: a model system in adaptive radiation
683 research. Proc. R. Soc. B 273: 1987–1998.
684
685 Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. Mrosso *et al.*, 2008
686 Speciation through sensory drive in cichlid fish. Nature 455: 620-626.
687

bioRxiv preprint doi: https://doi.org/10.1101/275388; this version posted March 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

688 Seehausen, O., 2015 Process and pattern in cichlid radiations–inferences for
689 understanding unusually high rates of evolutionary diversification. New Phytol. 207:
690 304-312.
691
692 Selz, O. M., M. E. Pierotti, M. E. Maan, C. Schmid, and O. Seehausen, 2014a Female
693 preference for male color is necessary and sufficient for assortative mating in two
694 cichlid sister species. Behav. Ecol. 25: 612-626.
695
696 Selz, O. M., R. Thommen, M. E. Maan, and O. Seehausen, 2014b Behavioural
697 isolation may facilitate homoploid hybrid speciation in cichlid fish. J. Evol. Biol. 27:
698 275-289.
699
700 Ser, J. R., R. B. Roberts, and T. D. Kocher, 2010 Multiple interacting loci control sex
701 determination in lake Malawi cichlid fish. Evol. 64: 486–501.
702
703 Stapley, J., P.G. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja, 2017
704 Variation in recombination frequency and distribution across eukaryotes: patterns and
705 processes. Phil. Trans. R. Soc. B, 372: 20160455.
706
707 Streelman, J. T., R. C. Albertson, and T.D. Kocher, 2003 Genome mapping of the
708 orange blotch colour pattern in cichlid fishes. Mol. Ecol. 12: 2465-2471.
709
710 Tang, H., X. Zhang, C. Miao, J. Zhang, R. Ming *et al.*, 2015 ALLMAPS: robust
711 scaffold ordering based on multiple maps. Gen. Biol. 16: 3.
712
713 Van Ooijen, J. W., 2006 JoinMap 4. Software for the calculation of genetic linkage
714 maps in experimental populations. Kyazma BV, Wageningen, Netherlands, 33.
715
716 Verheyen, E., W. Salzburger, J. Snoeks, and A. Meyer, 2003 Origin of the superflock
717 of cichlid fishes from Lake Victoria, East Africa. Science 300: 325-329.
718
719 Yoshida, K., Y. Terai, S. Mizoiri, M. Aibara, H. Nishihara *et al.*, 2011 B
720 chromosomes have a functional effect on female sex determination in Lake Victoria
721 cichlid fishes. PLoS Genet. 7: e1002203.