




ARTICLE

DOI: 10.1038/s41467-018-07648-2

OPEN

# Long-term experimental hybridisation results in the evolution of a new sex chromosome in swordtail fish

Paolo Franchini <sup>1</sup>, Julia C. Jones<sup>1,2</sup>, Peiwen Xiong<sup>1</sup>, Susanne Kneitz<sup>3</sup>, Zachariah Gompert<sup>4</sup>, Wesley C. Warren<sup>5</sup>, Ronald B. Walter<sup>6</sup>, Axel Meyer <sup>1,7</sup> & Manfred Scharl <sup>3,8,9</sup>

The remarkable diversity of sex determination mechanisms known in fish may be fuelled by exceptionally high rates of sex chromosome turnovers or transitions. However, the evolutionary causes and genomic mechanisms underlying this variation and instability are yet to be understood. Here we report on an over 30-year evolutionary experiment in which we tested the genomic consequences of hybridisation and selection between two *Xiphophorus* fish species with different sex chromosome systems. We find that introgression and imposing selection for pigmentation phenotypes results in the retention of an unexpectedly large maternally derived genomic region. During the hybridisation process, the sex-determining region of the X chromosome from one parental species was translocated to an autosome in the hybrids leading to the evolution of a new sex chromosome. Our results highlight the complexity of factors contributing to patterns observed in hybrid genomes, and we experimentally demonstrate that hybridisation can catalyze rapid evolution of a new sex chromosome.

<sup>1</sup>Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany. <sup>2</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 75123, Sweden. <sup>3</sup>Physiological Chemistry, Biozentrum, University of Würzburg, Am Hubland, 97074 Würzburg, Germany. <sup>4</sup>Department of Biology, Utah State University, Logan, UT 84322, USA. <sup>5</sup>McDonnell Genome Institute, Washington University School of Medicine, St. Louis 63108 MO, USA. <sup>6</sup>The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos 78666-4616 TX, USA. <sup>7</sup>Radcliffe Institute for Advanced Study, Harvard University, 9 Garden Street, Cambridge, MA 02139, USA. <sup>8</sup>Comprehensive Cancer Centre, University Clinic Würzburg, Josef Schneider Straße 6, 97074 Würzburg, Germany. <sup>9</sup>Hagler Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, TX 77843, USA. Correspondence and requests for materials should be addressed to A.M. (email: [axel.meyer@uni-konstanz.de](mailto:axel.meyer@uni-konstanz.de)) or to M.S. (email: [phch1@biozentrum.uni-wuerzburg.de](mailto:phch1@biozentrum.uni-wuerzburg.de))

Sex chromosomes evolve from autosomes and are typically extremely conserved, not only among species but entire classes of organisms. The best-studied sex chromosome systems are ZW and XY, female and male heterogametic systems respectively. These categorical sex-determining systems have evolved repeatedly and independently in the animal and plant kingdoms<sup>1</sup>. Yet genomic studies have also revealed several examples of rapidly evolving sex-determining mechanisms in closely related species of fish<sup>2–7</sup>, and also in amphibians<sup>8,9</sup> and reptiles<sup>10,11</sup>. Even between sister taxa of fishes, instances are known where either males or females are heterogametic. However, despite much theoretical work and empirical findings of convergent evolution of sex chromosomes<sup>11–16</sup>, why such variety in something as fundamental as the mechanisms of sex determination exists, and how it evolves, remains elusive.

If two closely related lineages differ in their sex chromosome systems, several scenarios for the emergence of this difference are possible<sup>17</sup>: either there was no chromosomal sex determination in the last common ancestor and both systems developed independently, or one lineage retained the ancestral sex chromosomes while in the other lineage a transition to the alternate system occurred. Theoretically, such transitions can happen on the same pair of sex chromosomes (homologous transition) or involve an autosome, which then becomes a new sex chromosome (heterologous transition). At the molecular level, the gene(s) determining sex might have changed their mode of action, e.g. from female to male determination, or a novel sex-determining (SD) gene might have arisen and taken over (SD turnover) in the evolving lineage.

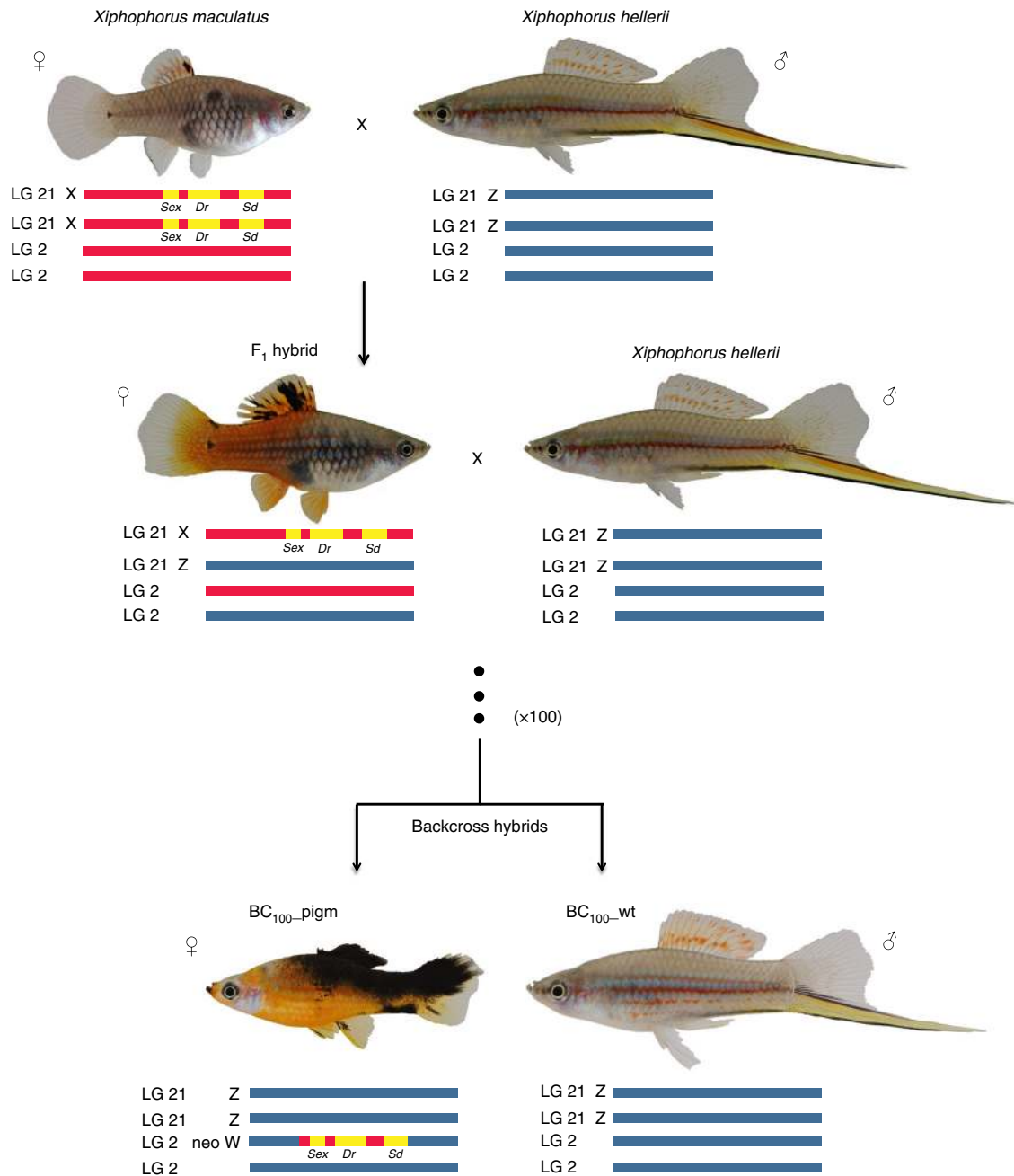
An intriguing scenario in sex chromosome evolution connected to speciation is hybridisation, particularly in crosses of species with different male and female heterogametic sex chromosomes. Based on Haldane's rule, one would expect that the heterogametic sex in hybrids tends to show greater inviability or sterility than the homogametic sex<sup>18</sup> and one proposed explanation for this is dominance theory<sup>19</sup>. Hybrids between different species might be expected to suffer deleterious incompatibilities because alleles from orthologous genes from different species may not interact well in hybrids, and if these genes are on the sex chromosomes and are recessive, the heterogametic sex is likely to suffer the most. This would be expected to result in selection for alternative sex chromosome systems in lineages arising from a hybridisation event. As intriguing as these concepts are, empirical evidence for the evolution of new sex chromosome systems following hybridization is very scant because these events have happened in the evolutionary past and would be expected to take long time periods to complete. Still, in some laboratory animals the loss of a sex-determining locus has been observed (e.g. medaka<sup>20</sup> and zebrafish<sup>21</sup>), but these domesticated lines have not yet developed a new stable mechanism of the same kind (a sex determination turnover) or completed a transition.

In vertebrates, teleost fish show an enormous diversity of sex chromosome systems, hence turnovers or transitions must have occurred more often than in most other classes of vertebrates (e.g. ref. 7). Therefore, the study of closely related species of fishes that exhibit different sex chromosome systems would be particularly informative for the investigation of the evolution of mechanisms of sex determination (e.g. ref. 7). Among the 26 species of the genus *Xiphophorus* (swordtails and platyfish) are species with both XY and ZW sex chromosome systems<sup>22,23</sup>. In this genus, these simple heterogametic systems are both present together with more complex situations that can include multiple loci and chromosomes<sup>23,24</sup>. *Xiphophorus maculatus*, for example, one of the most well studied and geographically widespread species in this genus, has three different genetically well-defined sex chromosomes, X, Y and W. Possible female genotypes are XX, XW and YW, and in males XY or YY. Several different models have

been proposed to explain sex determination with three sex chromosomes in this species<sup>25,26</sup>. While in the wild in most populations all these sex chromosome combinations co-exist, laboratory lines have been established which are stable for XY/XX or WY/YY sex determination. WW females can be generated in the laboratory after experimental manipulation and are viable, but such individuals have not been reported in natural populations<sup>23</sup>. In several species, autosomal modifiers, which occur at low frequency in natural populations, have also been reported to act, and may explain instances of atypical sex determination<sup>24,25</sup>. As another example of sex determination variation in this genus, in *Xiphophorus hellerii*, a polyfactorial sex-determining system has been reported to be acting. Additionally, it has been suggested that this species has a main genetic system affected by numerous autosomal modifiers, while more recent studies report that some strains of *X. hellerii* have a XW–YY female heterogametic system<sup>22,23</sup>. To our knowledge, environmental sex determination has not been reported from observations in nature or from laboratory studies in the genus *Xiphophorus*.

In general, this genus of fish comprises well known and widely studied models for organismic and molecular evolution. In particular, *Xiphophorus* species have been used as a model system for studying the evolutionary genetics of hybridisation for more than 50 years<sup>27–32</sup> with a recent resurgence of investigations using modern molecular and genomic data<sup>27,33–38</sup>. Ancient hybridisation events between many species in this genus has been inferred<sup>27,36,37,39</sup> and may be facilitated by apparently weak postzygotic isolation in *Xiphophorus*<sup>39–41</sup>. We previously utilized SNP data from RAD sequencing to estimate the phylogenetic relationships among all 26 *Xiphophorus* species, and uncovered incongruence between the RAD-nuclear vs. mtDNA phylogenies<sup>33,36,42</sup>. This phylogenetic incongruence likely reflects the contribution of hybridisation to the evolution of two *Xiphophorus* species (*Xiphophorus clemenciae* and *Xiphophorus monticolus*)<sup>36</sup>. Such phylogenetic incongruence, in addition to asymmetrical behavioural preferences of first generation hybrids and the morphology of the putative hybrid taxon being much closer to one of the possible parental species<sup>33</sup>, provides support for hybridisation followed by repeated backcrossing having contributed to the evolution of these species. The inferred ancient hybridisation event likely involved a female Southern platyfish, *X. maculatus*, and a male green swordtail, *X. hellerii*.

Evidence for reticulate evolution is typically inferred from analyses of extant populations or species. However, the actual processes contributing to the introduction of an allospecific genome and the fate of the extra-specific genetic material are difficult to identify and document. To better understand the genomic consequences of interspecific hybridisation for sex chromosome evolution and speciation, we conducted a long-term crossing experiment (conducted for >30 years and spanning >100 generations) in fish of the genus *Xiphophorus* (Fig. 1). We experimentally mimic the evolutionary scenario, hybridisation with repeated backcrossing, which is thought to have contributed to the evolution of at least two of the 26 species in the genus *Xiphophorus*<sup>33,36,37</sup>. Of particular interest is that the parental species used in the experimental evolutionary crosses differ in their sex chromosome systems. The maternal lineage used here, the platyfish *X. maculatus* Jp163A strain, has an XY sex chromosome system<sup>43</sup>, whereas the backcross paternal lineage, the swordtail *X. hellerii* strain originating in the Rio Lancetilla, has a ZW sex chromosome system. This experimental set up enabled us to investigate the genomic consequences of known hybrid ancestry and timing, and specifically the consequences of sex chromosome evolution. We find that during the process of continuous introgressive breeding, the sex-determining region of the X chromosome derived from the maternal ancestor was



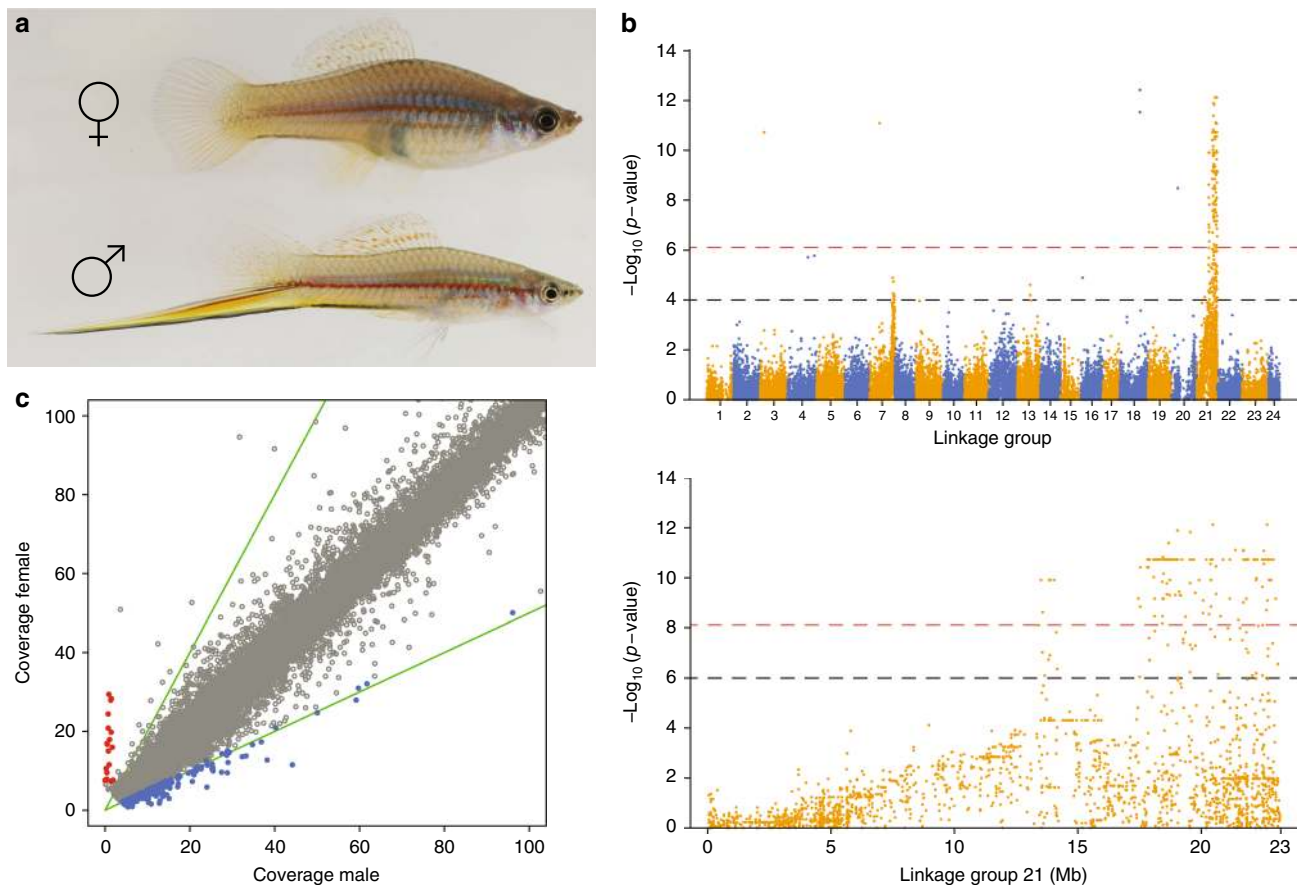
**Fig. 1** Schematic outline of the laboratory cross. The parental species *X. maculatus* and *X. hellerii*, a female specimen of the first generation hybrid and the resulting offspring obtained after approximately 100 generation of backcrosses (see Methods for details) are shown. The dorsal red (*Dr*), the spotted dorsal (*Sd*) and the oncogene *xmrk* loci, closely linked on the X-chromosome of the platyfish (LG21), are also highlighted. A translocation event of a genomic region containing these loci accompanied by recombination suppression has likely contributed to the formation of a new sex chromosome, a neo W chromosome (LG2), during the backcross experiment. Fish images by M. Scharf

translocated to an autosome of the receiving genome, thus providing empirical experimental evidence of a heterologous transition. We also note that at its new chromosomal position the former X-chromosome-specific region is located in a large region of retained maternal sequences, indicating that the loss of recombination accompanied the establishment of the novel W chromosome.

**Results**

**Identification of the sex-determining system and sex linkage group in *X. hellerii*.** The sex determination system of the

platyfish strain used for our study is firmly established as XY, not only from a plethora of crossing data but also from molecular genetic evidence, and there is no evidence for an influence of autosomal or environmental modifiers (e.g. the platyfish genome was based on the same strain<sup>43</sup>). However, for *X. hellerii* the situation is less clear and there are conflicting reports ranging from pure ZW heterogamety to a polyfactorial system with the absence of sex chromosomes<sup>44,45</sup>. To identify the sex-determining (SD) system in *X. hellerii*, deep coverage RAD sequencing of 60 *X. hellerii* individuals (30 males and 30 females) of the Rio Lancetilla strain was done (sequencing statistics for each individual is provided in Supplementary Data 1).



**Fig. 2** Sex-determination system in *Xiphophorus hellerii*. **a** A male and female specimen of *X. hellerii*, males are clearly distinguished by the elongated caudal fin, 'the sword' (Fish images by M. Scharfl). **b** Genome-wide association analysis, where sex is set as a binary case/control variable. The 65,417 SNPs that passed quality control were used in this analysis. The Manhattan plot shows the  $-\log_{10} p$ -value (Fisher's exact test) of each SNP across the 24 *X. maculatus* linkage groups (LGs). The dashed red line indicates the genome-wide significance level ( $p$ -value =  $7.79 \times 10^{-7}$ ), while a suggestive significance ( $p$ -value = 0.001) is denoted by the dashed black line. Orange and blue colours are used to distinguish between LGs. The peak showing the strongest association is located on LG21, and is enlarged in the bottom-right section of the panel. **c** Sequence coverage of RAD loci in female and male *X. hellerii* individuals. Each dot in the scatter plot represents the average coverage across all males (x-axis) against the average coverage across all females (y-axis) for each of the selected RAD loci. Red dots indicate potential W-linked loci (female-specific loci), while blue dots indicate loci potentially linked to the Z chromosome (twice the coverage in males than in females). The upper and lower green lines show the expected coverage of X-linked and Z-linked loci in a XY and ZW sex determination system, respectively. The coverage plot showing the total number of selected loci (average coverage across all individuals  $>3$  and  $<400$ ) is reported in Supplementary Fig. 2

To identify genomic regions associated with sex in *X. hellerii*, we employed a genome-wide association study (GWAS) and an analysis of genetic differentiation ( $F_{ST}$ ). For these analyses, we used the *X. maculatus* v4.4.2 as the reference genome<sup>46</sup>, where the sex chromosome pair was clearly detected in linkage group 21 (LG21). Given the high level of synteny conservation between the two species' genomes<sup>47</sup> this approach allowed us to directly compare the sex chromosome regions of *X. maculatus* and *X. hellerii* by plotting the GWAS  $p$ -values and the  $F_{ST}$  estimates on the *X. maculatus* linkage groups. The GWAS approach identified a large genomic region strongly associated with sex distal on LG21, spanning more than 1/3 of the LG (a genomic region of approximately 10 Mb, from ~13 to ~23 Mb, at the end of LG21) (Fig. 2b). This region includes 243 SNPs that exceeded the Bonferroni-corrected genome-wide threshold ( $p$ -value =  $7.64 \times 10^{-7}$ ) and showed an excess of heterozygosity in females vs. males, conforming to expectations under a ZW sex determination system. A weaker association was identified in LG7, where 15 SNPs exceeded a threshold suggestive of significance ( $p < 0.0001$ )

(the location of the SNPs highly associated with sex that exceed the thresholds are shown in Supplementary Data 2). A second region of increased female sex linked heterozygosity besides LG21 representing the W/Z pair was unexpected. A reasonable explanation is that this region may play a role as a minor sex modifier. Such autosomal modifiers of a major sex chromosomal system have frequently been noted in the genus *Xiphophorus*<sup>24</sup>. Further, in our laboratory cross experiments (see below) we observed exceptional cases of males with a BC<sub>100</sub>-pigm phenotype and females with a BC<sub>100</sub>-wt phenotype (Supplementary Table 1). Such cases may be caused by autosomal modifiers.

$F_{ST}$  analyses confirmed our findings using the GWAS approach. Several genomic regions exhibited a high degree of genetic differentiation between males and females on LG21 ( $F_{ST} > 0.5$ ) on a background of low genome-wide differentiation (average  $F_{ST}$  across all loci = 0.004). This pattern was consistent across the three different sliding windows used (1, 10 and 100 Kb).  $F_{ST}$  analysis confirmed the weaker association found on LG7 using the GWAS approach. Across the three windows used,

several regions consistently exceeded the set threshold, showing an average  $F_{ST}$  value in the upper 1% of the distribution of the per window  $F_{ST}$  values. However, the majority of the windows exceeding the threshold were located on LG21 (see Supplementary Data 3 and Supplementary Fig. 1).

To provide further confirmation of the genomic regions associated with sex in *X. hellerii*, we analysed coverage information from males and females in order to identify any non-recombinant regions along chromosomes. Here, we used the *X. hellerii* genome v3.0.1<sup>47</sup> for aligning the RAD reads, derived from a single female, the heterogametic sex in a W/Z sex chromosome system. By investigating copy number variation, we detected 18 candidate female-specific loci (mean coverage in females >7; mean coverage in males <2) (Fig. 2c, Supplementary Fig. 2 and Supplementary Data 4) and no male-specific loci. Again, such a pattern is compatible with female heterogamety and a W/Z sex chromosome system. We then aligned these loci to the *X. maculatus* genome v4.4.2 to compare the location of the sex-determination system in the two *Xiphophorus* species. We found that out of the total 18 female-specific loci, 16 aligned to the *X. maculatus* genome, 14 of which were on LG21, which is the sex linkage group (Supplementary Data 4). This indicates that the sex-chromosome pairs of *X. hellerii* and *X. maculatus* are homologous. We then identified potential Z-linked loci (mean coverage ratio in male/female >1.9) (Fig. 2c, Supplementary Fig. 2 and Supplementary Data 4). Of these, we found that 161 loci aligned to the *X. maculatus* genome, and the highest number of loci aligned to LG21 (54: 33.5%), followed by LG13 (27: 16.8%) and LG10 (17: 10.5%) (Supplementary Fig. 3 and Supplementary Data 4). Both the W-linked female-specific loci and the Z-linked male loci were found to be over-represented on LG21 (two-tailed Fisher exact test:  $p$ -value < 0.001). The occurrence of several Z-linked loci that did not align to *X. maculatus* LG21 could suggest the presence of autosomal modifiers. However, given the potentially high false positive rate when using such an approach, caution is needed in interpreting these findings.

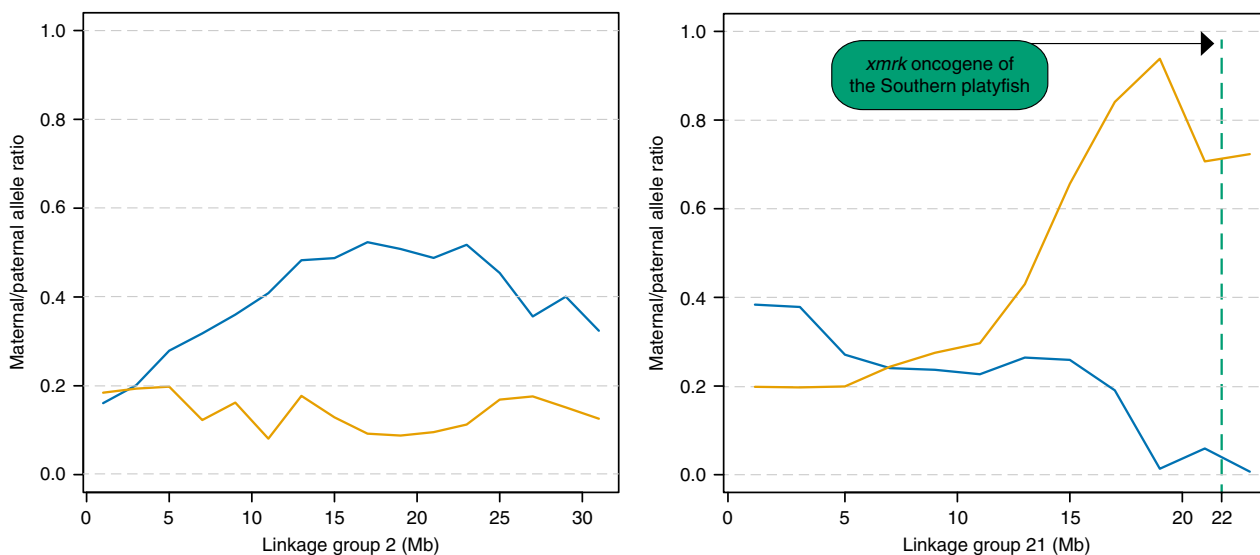
Taken together the GWAS,  $F_{ST}$  and coverage analysis results show that the Rio Lancetilla population of *X. hellerii*, which is the paternal species in our long-term crossing experiment, has a ZW chromosome system. LG21 is the sex chromosome pair and includes a terminal non-recombining region (spanning from approximately 13 to 23 Mb) and a large (0–13 Mb)

pseudoautosomal region (PAR). The boundary between the non-recombining region and the PARs was identified by a sudden drop of SNPs significantly associated with sex, as revealed by the GWAS (Fig. 2b and Supplementary Data 2), a pattern confirmed by the  $F_{ST}$  (Supplementary Fig. 1 and Supplementary Data 3) and the coverage analysis (Supplementary Data 4). Despite *X. maculatus* having a different type of genetic sex determination, the sex chromosomes of both parental species are homologous.

### Introgression of parental DNA in a controlled backcrossing experiment.

We performed a laboratory cross, hybridisation with backcrossing, and sequenced the first and approximately 100th generation backcross individuals (sequencing statistics for each individual are provided in Supplementary Data 5). This crossing experiment mimicked the evolutionary scenario that has potentially given rise to two *Xiphophorus* fish species, *X. clemenciae* and *X. monticolus*, and in which the parental species differ in their sex chromosome systems (see refs. 26,33,36,37,43). Accordingly, we crossed a female Southern platyfish (*X. maculatus*, genotype  $2N = 48$ ; XX) with a male swordtail (*X. hellerii*, genotype  $2N = 48$ ; ZZ). A female  $F_1$  fish was used for producing the first backcross with *X. hellerii*. Further on, female fish were selected in each backcross generation for two pigmentation phenotypes that are encoded on *X. maculatus* X-chromosome (LG21) close to a molecular marker, the *xmrk* oncogene, and mated to *X. hellerii* males.

Not unexpectedly, the crossing of two species with opposing sex chromosomal systems produced an  $F_1$  generation that showed severe signs of hybrid dysgenesis. In such crossing, the number of offspring that developed to adulthood was very low (usually 1–3, compared to 20–50 in the purebred parental species) and a considerable number of these were sterile (Supplementary Table 1). This was particularly the case for females, which carry the *X. maculatus* X chromosome, which was our target for selection in the long-term crossing experiment. Conversely, those offspring from backcrosses to *X. hellerii*, including  $BC_{100}$ , displayed full fertility and a clear-cut sex-linked inheritance of the pigmentation loci (Supplementary Table 2). All fish (99.7%) with these markers were females, while the fish without the selected markers were preponderantly males (91.7%).



**Fig. 3** Parental allele ratios in the first generation backcross. Maternal and paternal allele ratios in  $BC_{1\_wt}$  (blue line) and  $BC_{1\_pigm}$  (orange line) on LG2 and LG21. The location of the melanoma gene (*Xmrk*), linked to the pigmentation genes, in LG21 is shown

**Table 1 Segregation patterns of the parental alleles**

| Group                   | Class                 | Paternal alleles ( <i>X. hellerii</i> ) | Maternal alleles ( <i>X. maculatus</i> ) |
|-------------------------|-----------------------|---|--|
| F <sub>1</sub>          | Observed <sup>a</sup> | 130,379 (0.500)                         | 130,379 (0.500)                          |
| BC <sub>1</sub> _wt     | Observed              | 263,923 (0.743)                         | 91,329 (0.257)                           |
|                         | Expected              | 266,439 (0.750)                         | 88,813 (0.250)                           |
| BC <sub>1</sub> _pigm   | Observed              | 336,355 (0.771)                         | 100,059 (0.229)                          |
|                         | Expected              | 327,310 (0.750)                         | 109,103 (0.250)                          |
| BC <sub>100</sub> _wt   | Observed              | 525,244 (-1.000)                        | 258 (158 sites)                          |
|                         | Expected              | 525,502 (-1.000)                        | -0 (-0.000)                              |
| BC <sub>100</sub> _pigm | Observed              | 513,445 (-1.000)                        | 1073 (476 sites)                         |
|                         | Expected              | 514,518 (-1.000)                        | -0 (-0.000)                              |

For each group, the observed and expected number of alleles and their proportions (in parentheses) are shown

<sup>a</sup>Observed and expected number of alleles have the same values for the F<sub>1</sub> class as a result of the quality filter we applied to the loci selection

A clear enrichment of maternal alleles was observed in the BC<sub>1</sub>\_pigm group (Fig. 3) as expected at LG21 (maternal/paternal allele ratio >0.9) in a genomic region close to the *xmrk* oncogene of *X. maculatus* (Fig. 3). In the >100 backcross generation, maternal polymorphisms were found at 476 SNPs in the BC<sub>100</sub>\_pigm group, and at 158 SNPs in the BC<sub>100</sub>\_wt group (Table 1 and Fig. 4; see the ancestry frequency calculated at each SNP in Supplementary Fig. 4). Notably, in the BC<sub>100</sub>\_pigm a large block of retained maternal polymorphisms (204 SNPs) spanning approximately 10 Mb was found in the central part of LG2 (Fig. 4). Specifically, several loci with *X. maculatus* alleles were retained in the BC<sub>100</sub>\_pigm group, which in the genome of the parental *X. maculatus* exclusively identify scaffolds from LG21 that are located in the region containing *xmrk* and the pigmentation loci. No such enrichment was seen on LG21 of the BC<sub>100</sub>\_pigm group. Hence, this genomic region in the backcross hybrids is derived from the *X. hellerii* genome.

To identify genomic regions in hybrid individuals that harbour *X. maculatus* polymorphisms, parental alleles were sorted according to a set of key criteria, sites carrying a maternal variant in at least one individual of the BC<sub>100</sub> generation were identified and annotated, and the ancestry frequency of each SNP was calculated. For this analysis, the region in the BC<sub>100</sub>\_pigm showing high maternal allele density in LG2 was masked because of the hypothesized different mechanisms by which it was produced (translocation of the *X. maculatus* sex-determining region to an autosome of the receiving genome). After removing these sites, a total of 204 loci with maternal alleles were identified in the BC<sub>100</sub>. Our simulations show that the number of retained *X. maculatus* polymorphisms is higher than expected under a neutral model of hybridisation with backcrossing ( $p < 0.001$ ; Fig. 4). This suggests that selection has shaped the hybrid genomes beyond LG2.

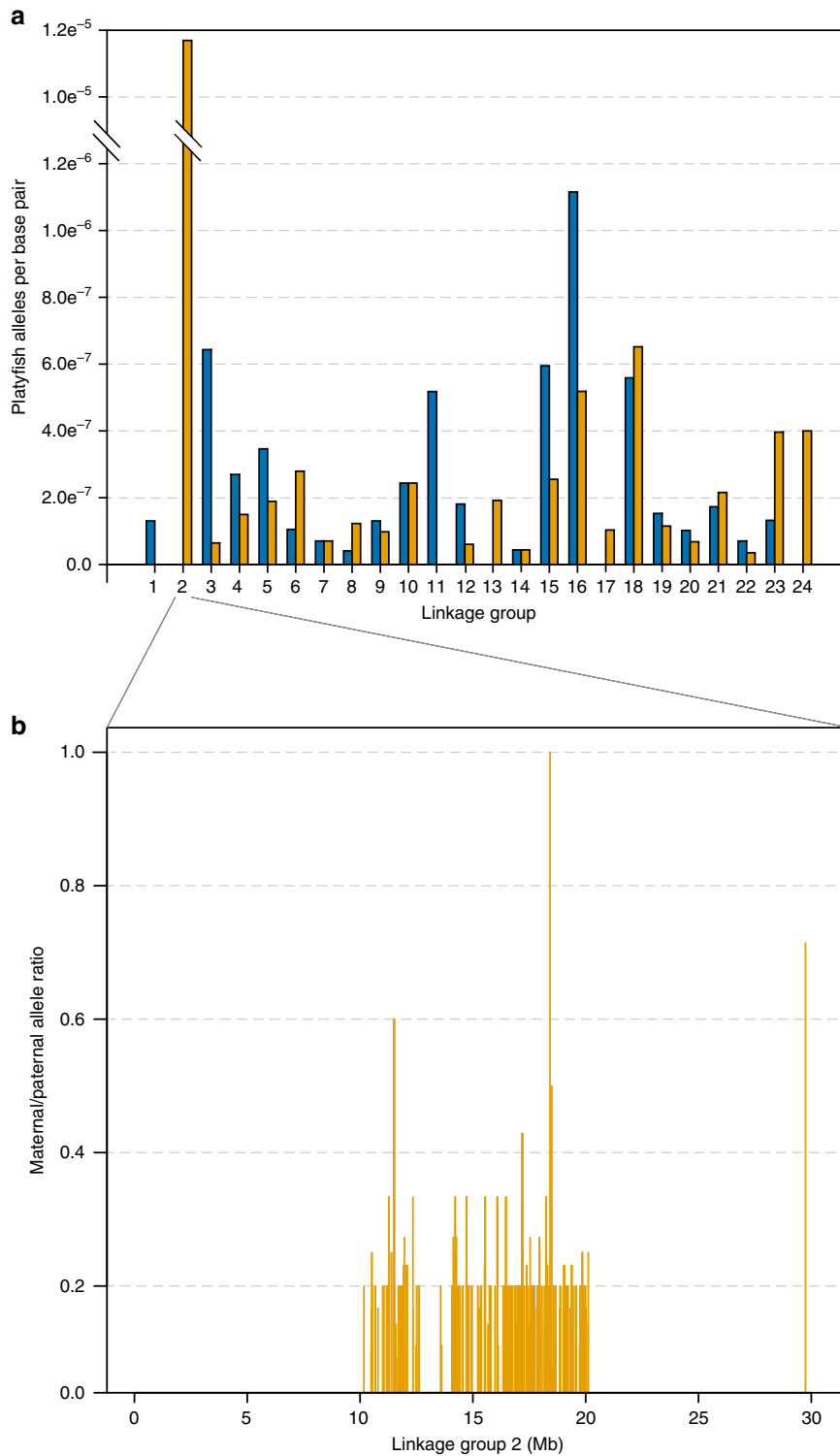
## Discussion

By conducting a long-term evolutionary crossing experiment, we demonstrate the genomic consequences, and particularly the consequences of sex chromosome evolution, of hybridisation with backcrossing to the paternal species. This same mechanism has potentially contributed to the evolutionary history of two *Xiphophorus* fish species. Importantly, the parental species in this cross differ in their sex chromosome systems. We find that artificial selection for pigmentation phenotypes in backcrossed individuals results in the retention of a large, maternally derived genomic block after 100 generations (>30 years). Interestingly, this experimental design has unexpectedly promoted the formation of a new sex chromosome, and although no other known selection is imposed under our experimental conditions, there is also a higher proportion of maternal SNPs retained elsewhere in the genome than expected under a neutral model. The retention

of certain *X. maculatus* alleles in synthetic swordtail–platyfish hybrids thus appears to be a phenomenon that occurs across a very short timescale and may be driven by a range of non-mutually exclusive forces.

In the synthetic hybrid lineage, maternal (*X. maculatus*) alleles are concentrated within a region of approximately 10 Mb on LG2 after about 100 backcross generations. Such a concentration is not seen elsewhere in the genome, or in the wild-type non-selected hybrid lineage. We suggest that a combination of two proximate mechanisms may have contributed to the retention of this unexpectedly large maternal genomic block. Firstly, a mechanism likely contributing to the generation and/or maintenance of this genomic pattern is a translocation event accompanied by recombination suppression. Specifically, potentially sometime during the establishment of the selected backcross line the region of LG21 known to encompass the black pigmentation locus *Sd* including the *xmrk* oncogene, the female sex-determining locus, and the red colour pattern locus *Dr* was translocated into the region on LG2. By imposing selection for these genes (a small region of former LG21<sup>46</sup>), the genes and the surrounding non-recombining regions on LG2 are transmitted as a large block. Second, theoretical predictions suggest that blocks of an introgressing genome are expected to be longer around selected loci<sup>48</sup>. Therefore, due to artificial selection of two genes, sometime during the early backcross generations a large maternally derived block encompassing the selected genes may have been maintained. We note that the higher proportion of maternal SNPs retained elsewhere in the genome may be due to the selection of fertile individuals in each backcross generation, and therefore the purging of incompatible combinations of alleles while an excess of maternal alleles are left.

Given that the core of the retained genomic block originates from the X chromosome and includes the sex-determining locus (from *X. maculatus*), the genomic consequences of this form of hybridisation followed by repeated backcrossing and selection appears to include the formation of a new sex chromosome. Our results strongly suggest that an X chromosome from the platyfish, *X. maculatus*, was introgressed into the genome of the swordtail, *X. hellerii*, which has a ZW sex chromosome system. After approximately 100 generations of backcrossing, the platyfish X-chromosome derived sex-determining region (X-SD) has moved to what was previously an autosomal location. While in the purebred platyfish homozygosity at the *Sd* region of X is required for female development, in the BC<sub>100</sub> heterozygosity is sufficient. Thus the recombinant neo-sex chromosome behaves like a W with a dominant effect on SD. In the backcrosses, only male *X. hellerii* were used, and the almost exclusive female sex of the fish expressing the selected colour loci on LG2 in the highest backcross generations show that the Z-chromosome (LG21) in *X. hellerii* has no male determining power. Our observation that the



**Fig. 4** Segregation patterns of parental alleles. **a** Frequency of loci carrying platyfish alleles in the BC<sub>100</sub>\_wt (blue bars) and BC<sub>100</sub>\_pigm (orange bars) in each of the 24 linkage groups of the *X. maculatus* genome; **b** frequency of maternal alleles across linkage group 2 (LG2)

sex-determining region of the X was translocated to an autosome provides the first empirical evidence of a heterologous transition. Additionally, at its new chromosomal position the former X-chromosome specific region is located within a large region of retained maternal sequences, indicating that the loss of recombination accompanied the establishment of the novel W chromosome. This formation of a new sex chromosome under a hybridisation with backcrossing scenario highlights the possibility

that the variation in genetic sex-determining mechanisms found in the genus *Xiphophorus*, which include simple male and female heterogametic systems together with more complicated situations<sup>24</sup>, may be contributed to by the many hybridisation events inferred among species in this genus. More specifically, this formation of a new sex chromosome provides a key target for future functional studies, and for further investigations of the *Xiphophorus* species thought to have arisen through hybridisation with

backcrossing in the wild. For example, could this new sex chromosome play a role in reproductive isolation? And do the sex chromosomes of wild hybrids play similar roles?

The X chromosome has been found to play a special role in reproductive isolation in many species, most likely due to Haldane's rule, meiotic drive or faster evolution ('faster X')<sup>49</sup>. However, to date a definitive role for the X chromosome in reproductive isolation and speciation remains elusive in *Xiphophorus*. In another *Xiphophorus* species, *Xiphophorus nezhualcoyotl*, that exhibits significant hybrid ancestry, it has been indicated that the putative X chromosome has lower levels of coding and non-coding introgression than average, although reportedly not a clear outlier from other chromosomes<sup>27</sup>. Sex chromosome systems in closely related species of fish are generally highly variable, suggesting changes occur rapidly in these systems, and neo-sex chromosomes or novel SD regions have been reported in natural populations<sup>50–55</sup>.

More broadly, further studies of hybridisation in this genus are imperative for fully understanding the genomic and adaptive consequences of the process, including at different time scales. Importantly, recent advances have been made in the development and application of genomic methods for detecting hybridisation and inferring individual ancestries<sup>56–58</sup>. Indeed, fine-scale signals of hybridisation can be detected using genomic analyses that allow the identification of introgressed genes and gene combinations which may be important in adaptation (e.g. refs. 59–61). Also, long-read de novo assemblies of various species allow more refined measures of genomic introgression events, a necessary measure to develop evolutionary theories of molecular mechanisms. Broadly, research to date suggests that hybridisation can be a mechanism for co-opting functional elements from another genome<sup>61,62</sup>. For instance, modern humans carry genomic regions derived from Neanderthal ancestors including loci involved in skin and hair phenotypes suggested to be adaptive for humans migrating to non-African environments<sup>61,62</sup>. Importantly, previous studies have suggested that the rate of hybrid genome stabilisation varies across taxa and is dependent on demographic differences, selection or strong genetic drift<sup>57</sup>. In rare and extreme cases, whole genome hybridisations can alter the course of major traits such as meiosis<sup>38</sup>.

In the current study, by using an experimental backcross for over 100 generations we could directly observe the genomic consequences of hybridisation with backcrossing. Importantly this was done with two parental species with different sex (determining) chromosome systems, and where we strongly selected for distinct phenotypes. It is likely that a combination of different processes drove the observed genomic outcomes, including exogenous selection (i.e. artificial selection for the pigmentation phenotype and female sex), and endogenous selection such as gene–gene interactions. Further, the genomic patterns observed here could have come about through a number of different scenarios that require future investigation. For instance, the patterns may reflect the potential for a connection between chromosomal organisation and adaptive evolution in this system. Alternatively, and potentially equally parsimoniously, the translocation event may have also been neutral. Recombination suppression associated with chromosomal rearrangement can promote local adaptation and the accumulation of genetic incompatibilities between species (reviewed in refs. 63,64). In *Helianthus* sunflowers, where large linkage blocks appearing to resist recombination were observed in both experimental and ancient hybrids, gene interactions were invoked as the most likely explanation for the observed concordance in genomic pattern<sup>65</sup>. In another more recent study where introgression and local adaptation in two poplar species were investigated using fine-scale genomic techniques, admixed individuals of one species

harbour a telomeric region on one chromosome which had introgressed from the other species, and this region was found to contain several candidate genes for local adaptation<sup>59</sup>. Interestingly, a paralogous block of genes on another chromosome showed no signs of introgression or signatures of selection<sup>59</sup>. Natural hybridisation between several species in the genus *Xiphophorus* has been inferred and has reportedly occurred across different timescales, including contemporary and more ancient events<sup>27,34,36,37,39</sup>, suggesting hybridisation, such as the scenario investigated in our study, may be a key contributor to the evolutionary history of this group of fishes.

## Methods

**Sex determination in *Xiphophorus hellerii*.** A total of 30 female and 30 male *X. hellerii* individuals from the same strain (origin Rio Lancetilla), which were also used for the backcrossing experiment, were genotyped to identify sex-related markers, and thus characterize the sex determination system of this swordtail species strain.

**Sequencing and data filtering.** Genomic DNA was extracted from fin clips using standard phenol–chloroform extraction<sup>66</sup>, followed by RNase treatment. DNA quality of each sample was assessed by agarose gel electrophoresis and quantified using a Qubit v2.0 fluorometer (Life Technologies, Darmstadt, Germany). Approximately 100 ng of DNA template of each sample was used to construct a single quaddRAD<sup>67</sup> library (PstI and MspI were used as rare and frequent restriction enzymes, respectively), size selected from 450 to 550 bp using a Pippin Prep platform (Sage Sciences, Beverly, USA) and paired-end sequenced in an Illumina HiSeq2500 (rapid run mode, 2 × 151 cycles).

The raw fastq files obtained from one Illumina lane (NCBI's SRA database, accession no. PRJNA493969) were first processed using the *clone\_filter* module implemented in the Stacks v1.46<sup>68,69</sup> package to identify and remove duplicate reads. The retained sequence data, stripped of the four random bases at the 5' end of each paired read, was then separated by the dual index inner barcodes using the Stacks' *process\_radtags* script (option: *-inline\_inline*) with no quality filters applied. Next, the sequences were processed by the dDocent v2.2.25 pipeline<sup>70</sup>, with default parameters, in order to get the individuals' genotypes. For this RAD data set, the dDocent pipeline was preferred to Stacks because of its superior ability to deal with paired-end reads when calling genotypes. Briefly, dDocent calls Trimmomatic v0.36<sup>71</sup>, bwa v0.7.15<sup>72</sup> and Freebayes v1.1.0<sup>73</sup> to respectively quality control the raw reads, align the filtered sequences to the reference genome and infer individuals' genotypes at polymorphic loci. Trimmomatic was used to remove bases with a quality score <20 from the beginning and end of reads, and additionally trim bases with an average quality score <10 in a sliding 5 bp window. The 'mem' algorithm implemented in bwa was applied setting the match score value (-A) to 1, the penalty for a mismatch (-B) to 4 and the gap penalty (-O) to 6. Freebayes was used to call genotypes setting a minimum mapping and base quality score to 10.

**Sequence variation analysis: filtering and GWAS of sex in *X. hellerii*.** The raw VCF variant file was obtained by setting the *X. maculatus* v4.4.2 as the reference genome<sup>46</sup>. In this version of the *X. maculatus* genome, the original scaffolds<sup>43</sup> were tiled along a dense genetic map to create chromosome-length genome assemblies. The VCF file was filtered using VCFtools v0.1.15<sup>74</sup> to retain loci present in a minimum of 80% of the individuals and having mapping quality (MQ) >30. Further stringent filtering was applied using the *dDocent\_filters* script (<http://ddocent.com/filtering/>) that relies on both VCFtools and the *vcfilter* module in the *vcflib* package (<https://github.com/vcflib/>). A total of 65,417 SNPs passed the quality control applied and were used for the GWAS and the  $F_{ST}$  analyses (the full set of filters, and the number of sites removed at each step, are reported in Supplementary Table 3; the individuals' genotypes were deposited into the Dryad Data Repository: <https://doi.org/10.5061/dryad.7h54h66>).

The final VCF file was used to perform a standard genome-wide association analysis using PLINK v1.90b4.9<sup>75</sup> setting sex as the binary case/control variable. The  $-\log_{10} p$ -value (obtained with a Fisher's exact test) of each SNP was shown as a function of genomic position on the 24 *X. maculatus* LGs<sup>46</sup> in a Manhattan plot constructed with the R package qqman v0.1.4<sup>76</sup>. The threshold for genome-wide significance was set at a  $p$ -value =  $7.64e^{-7}$  after Bonferroni correction for multiple comparisons (0.05/65,417 variants). To identify highly divergent, potential candidate sex-chromosome regions using an additional line of investigation, the program VCFtools was used to calculate  $F_{ST}$ -values between males and females across different non-overlapping sliding windows (1, 10 and 100 Kb). As a suggestive threshold for significance, we used the upper 1% percentile distribution of the per window  $F_{ST}$  values (1 Kb window:  $F_{ST} = 0.184$ ; 1 Kb window:  $F_{ST} = 0.181$ ; 1 Kb window:  $F_{ST} = 0.157$ ).

**Coverage analysis to identify sex association in *X. hellerii*.** As an alternative and independent approach to identify and verify potential sex-associated SNPs, we



analysed mapped sequence coverage information from *X. hellerii* males and females to identify loci located in the non-recombining and recombination restricted regions of the sex chromosomes. We first created a set of reference loci using the *ustacks* module of the Stacks package in default mode, but setting the minimum depth of coverage required to create a stack (-m) to 5. As a reference for aligning the RAD loci, we used the *X. hellerii* genome v3.0.1 (NCBI accession number: GCA\_001443345.1)<sup>47</sup> that is derived from a single female, strain Sarabia, from the Xiphophorus Genetic Stock Center (TX, USA). Existing evidence suggested that the female is the heterogametic sex in this species; therefore, because the genome of *X. hellerii* was generated using a single female individual, using the *X. hellerii* genome would allow us to recover a higher number of W-linked loci than using the *X. maculatus* genome. We used this approach to confirm the findings of the statistically more robust GWAS approach, and possibly to identify specific *X. hellerii* sex-determining regions (e.g. autosomal modifiers). The total 70,763 loci, extracted from the Stacks' catalogue formatted in fasta format (deposited into the Dryad Data Repository: <https://doi.org/10.5061/dryad.7h54h66>), were used as a reference for the alignment of the sequence set of each sample carried out with Bowtie2 v2.3.0<sup>77</sup> in '-very-sensitive' mode. To avoid redundant information, only one of the paired loci was retained to serve as a reference. Next, raw coverage values—i.e. the number of reads mapping to each locus—were extracted from each individual mapping file using SAMtools v1.2.1<sup>78</sup> *idxstats* and normalized using the Median Ratio Normalization (MRN) function implemented in DESeq2 v1.8.1<sup>79</sup>.

We then calculated the mean coverage for each locus (including males and females), but first removed loci with coverage below 3 and above 400 because they most likely indicate mapping errors and repetitive elements, respectively (the number of loci was reduced to 29,444 after the coverage filters were applied). We assumed a ZW sex-determination system is operating in this system, given the results from our sequence variation analysis (see the previous paragraph), and from the filtered coverage data set we selected (1) potentially W-linked female-specific loci and (2) Z-linked loci. According to expectations, W-linked loci should be present only in females, the heterogametic sex, so we applied coverage filters to satisfy this condition (mean coverage in females >7; mean coverage in males <2). On the other hand, the Z-linked loci are expected to be present in both sexes, but they should have twice the coverage in males than in females, as they have two copies of the Z chromosome (coverage filter to identify these loci: mean coverage ratio in male/female >1.9). We note that this type of filtering criteria is dependent on the sequencing depth of coverage used in the experiment, and the use of these arbitrary filters means the results should be interpreted with caution. However, the main purpose of the coverage analysis was to confirm the results of the GWAS analysis. Finally, we used a two-tailed Fisher's exact test to test for over-representation of these candidate sex-chromosome loci in the *X. maculatus* LGs that were associated with sex (SNP analysis).

**Synthetic hybrids: laboratory cross experiment.** Laboratory crosses were conducted which mimic the hybridisation with backcrossing evolutionary scenario that has potentially given rise to at least two *Xiphophorus* fish species, and may have contributed to the ancestry of other taxa. We sampled both first generation and approximately 100th generation backcrossed individuals that had been artificially selected for two pigmentation phenotypes. We note that all sampled backcross generations were raised under the same standard laboratory conditions, therefore excluding any environmental effects.

In these laboratory crosses, one parental species is *X. maculatus*, where the strain Jp163A (origin Rio Jamapa) was used, which is also the same strain from which the reference genome was produced<sup>43</sup>. This strain has been kept as a brother-sister mating line for more than 100 generations. The other parental species used in the laboratory crosses is *X. hellerii*. The stock is derived from fish collected from the Rio Lancetilla and has been kept for over 50 years in closed colony breeding. The *X. hellerii* stock has been through several bottlenecks (due to the breakdown of the stock) where only a handful of fish or even a single pair was used to regenerate the stock. Both parental species are representatives of the two clades that were earlier shown to have been involved in the initial hybridisation event generating *X. clemenciae* and *X. monticolus*<sup>33,36,37</sup>.

The first generation hybrid cross was between a female *X. maculatus* and a male *X. hellerii*. Next, a backcross was made between an F<sub>1</sub> female, carrying the pigmentation phenotypes intended for selection and an *X. hellerii* male. Females were then selected for each successive backcross according to the two colour markers. One is the erythrophore pigmentation pattern *dorsal red*, *Dr*, which causes a dark red colouration of the dorsal fin and in hybrids extends from there over the whole body and the unpaired fins. The second pattern is caused by the closely linked macromelanophore pattern locus *spotted dorsal*, *Sd*. It is expressed as jet-black pigmentation spots at the base of the dorsal fin in purebred *X. maculatus*. In the genetic background of *X. hellerii*, the macromelanophores cover the whole dorsal fin, and from there invade the dorsal and caudal body compartment of the fish. As an effect of the *Sd* locus component, *xmrk* oncogene<sup>74</sup>, the macromelanophore pattern in the genetic background of *X. hellerii* is enhanced to melanotic hyperpigmentation that can progress to malignant melanoma (for an overview of the pigmentation genetics of platyfish/swordtail hybrids, see ref. <sup>40</sup>). *Dr* and *Sd/xmrk* are closely linked on the X-chromosome of the platyfish corresponding to linkage group 21<sup>43,46</sup>. In the serial backcross experiment, each generation consists of 50% fish with and 50% without the chromosome region

where both genes are located, designated as 'pigm' or 'wt' group. Selecting fish with the two linked colour markers for breeding mimics natural selection for a certain phenotype. Therefore, we expect fish with the pigmentation phenotype to carry the parental platyfish genetic region of the X-chromosome. Backcrossing was continued in this fashion for at least 100 generations (more than 30 years of laboratory crosses have been performed, counting was terminated after backcross 50 and on average three backcross generations were produced per year). For the approximate 100th generation backcross offspring, we expect fish carrying the pigmentation genes (further on referred to as BC<sub>100</sub>-pigm) to be enriched for genetic markers of platyfish LG21. By contrast, segregants not carrying the colour genes (wild type pigmented, further on referred to as BC<sub>100</sub>-wt) should not be enriched for platyfish sequences of this region of the genome.

**Sequencing, filtering and genotype calling.** Genomic DNA was extracted from pooled organs (brain, eyes, gills, liver, spleen, kidney) of the two parental species (two males and two female individuals each for *X. maculatus* and *X. hellerii*), four F<sub>1</sub> females, 16 first generation backcross individuals and 18 individuals generated after about 100 backcross generations. DNA extraction was carried out using the standard phenol-chloroform extraction<sup>66</sup>. The DNA quality of each sample was determined by agarose gel electrophoresis and quantified using a Qubit v2.0 fluorometer. Approximately 300 ng of DNA template of each sample was used to construct double-digest restriction site-associated DNA (ddRAD)<sup>80</sup> libraries following the modifications introduced in ref. <sup>81</sup>. DNA digestion was carried out using the restriction enzymes PstI (rare cutter) and MspI (frequent cutter).

A single ddRAD library containing 46 barcoded individuals (see Supplementary Data 3 for details) was constructed, size selected from 360 to 430 bp using a Pippin Prep system and single-end sequenced in an Illumina HiSeq2500 using rapid run mode with 151 cycles (raw reads were deposited in the NCBI's SRA database, accession no. PRJNA493969).

The *proccs\_radtags* script implemented in the Stacks package was used for individual demultiplexing and for filtering erroneous and low-quality reads (options: -c -q). After this process, an average of 2.95 million (*X. maculatus*) and 2.08 million (*X. hellerii*) sequences were obtained for the parents; an average of 3.05, 1.91 and 2.34 million sequences were obtained for the F<sub>1</sub>, BC<sub>1</sub> and BC<sub>100</sub> individuals, respectively (see Supplementary Data 3). The sequence length of each read was 146 bp after removing its 5-bp barcode.

Filtered reads were individually aligned to the anchored version of the *X. maculatus* genome<sup>46</sup> using Bowtie2 with default end-to-end mode. SAMtools was then used to filter the mapping result by retaining only sites with high quality score (MQ ≥ 30). Loci construction and genotype calling were performed using the Stacks' *ref\_map.pl* pipeline requiring a minimum of five reads (-m 5) to form a locus and calling genotypes at a 5% significance level using the bounded error model (upper bound of 0.05). The allelic variants (SNPs) were exported in VCF format using the Stacks' *populations* module.

**Sorting of parental alleles in the laboratory cross dataset.** An in-house Perl script was used to identify the subset of sites fixed for alternative alleles in the two parental species (*X. maculatus* and *X. hellerii*) of the laboratory cross following these criteria: (1) the individuals' genotypes are homozygous within each parental species, but they have different allelic variants; (2) the sites selected previously are heterozygous including both parentals' allelic variants in F<sub>1</sub> individuals; (3) at these sites, all BC<sub>1</sub> and BC<sub>100</sub> individuals have at least one paternal (*X. hellerii*) allele. After these filtering steps, we obtained a total of 34,632 SNPs (the individuals' genotypes were deposited into the Dryad Data Repository: <https://doi.org/10.5061/dryad.7h54h66>).

**Tests for excess ancestry.** As we focused on SNPs fixed between the parental species (see above), each SNP was fully informative about ancestry. Thus, we were able to directly calculate the frequency of platyfish ancestry in the BC<sub>100</sub> lines. We then used a stochastic simulation to develop null expectations for the frequency of platyfish ancestry expected in the absence of selection after 100 generations of backcrossing. We sampled platyfish ancestry frequencies each generation stochastically as  $y_t = \text{binomial}(p_{t-1} * 0.5, 2N)$  and then set  $p_t = y_t/2N$ . Here,  $y_t$  is the number of platyfish alleles in the sample of  $N$  diploid hybrid fish and  $p_t$  and  $p_{t-1}$  are the ancestry frequencies in the current and previous generation, respectively. The expectation for the binomial is reduced by half each generation as the BC line is backcrossed to *X. hellerii*. Using this iterative process and starting from  $p_0 = 0.5$  (F<sub>1</sub>s), we simulated expectations for 100 generations of backcrossing with  $N = 10$  or 15 hybrid fish per generation (conservative estimates of the average number of fish used to maintain the line over time). We repeated this procedure 10,000 times. In all cases, we found expected platyfish ancestry at generation 100 of  $p_{100} = 0$ .

**Ethical statement.** All animals were kept and sampled in accordance with the applicable EU and national German legislation governing animal experimentation. In particular, all experimental protocols were approved through an authorization (568/300-1870/13) of the Veterinary Office of the District Government of Lower Franconia, Germany, in accordance with the German Animal Protection Law (TierSchG).

**Data availability**

Raw Illumina sequences were deposited in the NCBI's Sequence Read Archive (SRA) database with accession no. PRJNA493969. The genotypes of each data set that we described in the manuscript (VCF format) and the Stacks' loci we used in the coverage analysis (FASTA format) have been uploaded to the Dryad Digital Repository (<https://doi.org/10.5061/dryad.7h54h66>). A reporting summary for this article is available as a Supplementary Information file.

Received: 25 June 2018 Accepted: 13 November 2018

Published online: 03 December 2018

**References**

- Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, CA, USA, 1983).
- Phillips, R. B., Konkol, N. R., Reed, K. M. & Stein, J. D. Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus*, *Salmo*, and *Salvelinus* (Salmonidae). *Genetica* **111**, 119–123 (2001).
- Woram, R. A. et al. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* **13**, 272–280 (2003).
- Mank, J. E., Promislow, D. E. & Avise, J. C. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87**, 83–93 (2006).
- Volff, J. N., Nanda, I., Schmid, M. & Scharl, M. Governing sex determination in fish: regulatory putches and ephemeral dictators. *Sex. Dev.* **1**, 85–99 (2007).
- Mank, J. E. & Avise, J. C. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex. Dev.* **3**, 60–67 (2009).
- Kitano, J. & Peichel, C. L. Turnover of sex chromosomes and speciation in fishes. *Environ. Biol. Fish.* **94**, 549–558 (2012).
- Miura, I. An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. *Sex. Dev.* **1**, 323–331 (2008).
- Stöck, M. et al. A cryptic heterogametic transition revealed by sex-linked DNA markers in Palearctic green toads. *J. Evol. Biol.* **24**, 1064–1070 (2011).
- Gamble, T. et al. Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Mol. Biol. Evol.* **32**, 1296–1309 (2015).
- Graves, J. A. M. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu. Rev. Genet.* **42**, 565–586 (2008).
- Scharl, M. Sex chromosome evolution in non-mammalian vertebrates. *Curr. Opin. Genet. Dev.* **14**, 634–641 (2004).
- Blaser, O., Neuenschwander, S. & Perrin, N. Sex-chromosome turnovers: the hot-potato model. *Am. Nat.* **183**, 140–146 (2014).
- van Doorn, G. S. & Kirkpatrick, M. Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics* **186**, 629–645 (2010).
- Bachtrog, D. et al. Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899 (2014).
- Herpin, A. & Scharl, M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep.* **16**, 1260–1274 (2015).
- Scharl, M. Sex determination by multiple sex chromosomes in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10575–10576 (2015).
- Haldane, J. B. S. Sex ratio and unisexual sterility in animal hybrids. *J. Genet.* **12**, 101–109 (1922).
- Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005).
- Nanda, I., Hornung, U., Kondo, M., Schmid, M. & Scharl, M. Common spontaneous sex-reversed XX males of the medaka *Oryzias latipes*. *Genetics* **163**, 245–251 (2003).
- Wilson, C. A. et al. Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* **198**, 1291–1308 (2014).
- Woolcock, B. et al. Allele-specific marker generation and linkage mapping on the *Xiphophorus* sex chromosomes. *Zebrafish* **3**, 23–37 (2006).
- Schultheis, C., Böhne, A., Scharl, M., Wolff, J. N. & Galiana-Arnoux, D. Sex determination diversity and sex chromosome evolution in Poeciliid fish. *Sex. Dev.* **3**, 68–77 (2009).
- Scharl, M., Galiana-Arnoux, D., Schultheis, C., Böhne, A. & Wolff, J. N. *Ecology and Evolution of Poeciliid Fishes* (eds Evans, J. P. et al.) (University of Chicago Press, Chicago, IL, USA, 2011).
- Kallman, K. D. *Evolutionary Genetics of Fishes* (ed Turner, B. J.) 95–171 (Springer, New York, NY, USA, 1984).
- Volff, J. N. & Scharl, M. Variability of genetic sex determination in poeciliid fishes. *Genetica* **111**, 101–110 (2001).
- Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G. & Andolfatto, P. Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* **25**, 2661–2679 (2016).
- Gordon, M. Heritable color variations in the Mexican swordtail-fish aquarium species as the *Drosophila* of fish genetics. *J. Hered.* **28**, 223–230 (1937).
- Atz, J. W. Effects of hybridization on pigmentation in fishes of the genus *Xiphophorus*. *Zoologica* **47**, 153–181 (1962).
- Häussler, G. Über Melanombildungen bei Bastarden von *Xiphophorus helleri* und *Platypoecilus maculatus* var. *Rubra*. *J. Mol. Med.* **7**, 1561–1562 (1928).
- Gordon, M. The genetics of a viviparous top-minnow *Platypoecilus*; the inheritance of two kinds of melanophores. *Genetics* **12**, 253–283 (1927).
- Kosswig, K. Über Bastarde der Teleostier *Platypoecilus* und *Xiphophorus*. *Z. Indukt. Abstamm. Vererbungslehre* **44**, 150–158 (1928).
- Meyer, A., Salzburger, W. & Scharl, M. Hybrid origin of a swordtail species (Teleostei: *Xiphophorus clemenciae*) driven by sexual selection. *Mol. Ecol.* **15**, 721–730 (2006).
- Culumber, Z. W. et al. Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. *Mol. Ecol.* **20**, 342–356 (2011).
- Jones, J. C., Perez-Sato, J.-A. & Meyer, A. A phylogeographic investigation of the hybrid origin of a species of swordtail fish from Mexico. *Mol. Ecol.* **21**, 2692–2712 (2012).
- Jones, J. C., Fan, S., Franchini, P., Scharl, M. & Meyer, A. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Mol. Ecol.* **22**, 2986–3001 (2013).
- Kang, J. H., Scharl, M. & Meyer, A. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. *BMC Evol. Biol.* **13**, 25 (2013).
- Schumer, M. et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656–660 (2018).
- Cui, R. et al. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**, 2166–2179 (2013).
- Scharl, M. Evolution of Xmrk: an oncogene, but also a speciation gene? *Bioessays* **30**, 822–832 (2008).
- Rosenthal, G. G. & Ryan, M. J. Conflicting preferences within females: sexual selection versus species recognition. *Biol. Lett.* **7**, 525–527 (2011).
- Meyer, A., Morrissey, J. M. & Scharl, M. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature* **368**, 539–542 (1994).
- Scharl, M. et al. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* **45**, 567–572 (2013).
- Peters, V. G. Vergleichende Untersuchungen an drei Subspecies von *Xiphophorus helleri* Heckel (Pisces): (Morphologische Merkmale, geschlechtliche Differenzierung, Wachstum und Geschlechtsverhältnisse). *J. Zool. Syst. Evolut. Res.* **2**, 185–271 (1964).
- Kazianis, S. *Viviparous Fishes* (eds Uribe, M. C. & Grier, H.) 381–400 (New Life Publications, Mexico, 2005).
- Amores, A. et al. A RAD-tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics* **197**, 625–641 (2014).
- Shen, Y. et al. *X. couchianus* and *X. hellerii* genome models provide genomic variation insight among *Xiphophorus* species. *BMC Genomics* **17**, 37 (2016).
- Sedghifar, A., Brandvain, Y. & Ralph, P. Beyond clines: lineages and haplotype blocks in hybrid zones. *Mol. Ecol.* **25**, 2559–2576 (2016).
- Presgraves, D. C. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* **24**, 336–343 (2008).
- Kitano, J. et al. A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079–1083 (2009).
- Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D. & Peichel, C. L. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* **5**, e1000391 (2009).
- Kondo, M. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Res.* **16**, 815–826 (2006).
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S. & Sakaizumi, M. Evidence for different origins of sex chromosomes in closely related oryzias fishes: substitution of the master sex-determining gene. *Genetics* **177**, 2075–2081 (2007).
- Roberts, R. B., Ser, J. R. & Kocher, T. D. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* **326**, 998–1001 (2009).
- Ser, J. R., Roberts, R. B. & Kocher, T. D. Multiple interacting loci control sex determination in Lake Malawi cichlid fish. *Evolution* **64**, 486–501 (2010).
- Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360 (2016).
- Abbott, R. J., Barton, N. H. & Good, J. M. Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* **25**, 2325–2332 (2016).

58. Gompert, Z. & Mandeville, E. G. Analysis of population genomic data from hybrid zones. *Annu. Rev. Ecol.* <https://doi.org/10.1146/annurev-ecolsys-110316-022652> (2017).
59. Suarez-Gonzalez, A. et al. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol. Ecol.* **25**, 2427–2442 (2016).
60. VonHoldt, B. M., Kays, R., Pollinger, J. P. & Wayne, R. K. Admixture mapping identifies introgressed genomic regions in North American canids. *Mol. Ecol.* **25**, 2443–2453 (2016).
61. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
62. Vernot, B. & Akey, J. M. Resurrecting surviving neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014).
63. Faria, R. & Navarro, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669 (2010).
64. Ortiz-Barrientos, D., Engelstädter, J. & Rieseberg, L. H. Recombination rate evolution and the origin of species. *Trends Ecol. Evol.* **31**, 226–236 (2016).
65. Rieseberg, L., Sinervo, B., Linder, C., Ungerer, M. & Arias, D. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* **272**, 741–745 (1996).
66. Blin, N. & Stafford, D. W. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* **3**, 2303–2308 (1976).
67. Franchini, P., Monné Parera, D., Kautt, A. F. & Meyer, A. quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Mol. Ecol.* **26**, 2783–2795 (2017).
68. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
69. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3* **1**, 171–182 (2011).
70. Puritz, J. B., Hollenbeck, C. M. & Gold, J. R. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431 (2014).
71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
72. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
73. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at arXiv:1207.3907v2 [q-bio.GN] (2012).
74. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
75. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
76. Turner, S. D. qqman: an R package for visualizing GWAS results using Q–Q and Manhattan plots. *bioRxiv* 1–2. <https://doi.org/10.1101/005165> (2014).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
80. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135 (2012).
81. Franchini, P. et al. Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Mol. Ecol.* **23**, 1828–1845 (2014).

### Acknowledgements

We thank L. da Conceicao Ferrao Beck and P. Fischer for laboratory assistance. P.F. was supported by the Deutsche Forschungsgemeinschaft (grant number FR 3399/1-1). J.J. was supported by Uppsala University. P.X. was supported by the China Scholarship Council (fellowship number 201306380094). A.M. was supported by the Deutsche Forschungsgemeinschaft (ME 1724/1). M.S. was supported by the Deutsche Forschungsgemeinschaft (grant numbers SCHA 408/12-1, 408/10-1) and the Department of Biology of Texas A&M University. *Xiphophorus* sequencing was supported by grants to R.B.W. (R24-OD-011120) and W.C.W. (2R24OD011198-04A1).

### Author contributions

M.S. carried out the long-term crossing experiment. P.F., J.J., A.M. and M.S. designed the study. P.F., J.J., P.X., S.K., Z.G. and M.S. devised and performed analyses. R.B.W. and C.W. participated in *Xiphophorus* sequencing and data review. P.F., J.J. and M.S. wrote the manuscript with input from all authors.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-07648-2>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018