Evolution and conservation of *Characidium* sex chromosomes

Utsunomia Ricardo[1]*, Scacchetti Priscilla C.[1]*, Hermida Miguel[2], Fernández-Cebrián Raquel[3], Taboada Xoana[4], Fernández Carlos[2], Bekaert Michaël[5], Mendes Natália J.[1], Robledo Diego[4,6], Mank Judith E.[7], Taggart John B.[5], Oliveira Claudio[1], Foresti Fausto[1†], Martínez Paulino[2†]

[1]Departamento de Morfologia, Instituto de Biociências, UNESP, 18618-689 Botucatu, São Paulo, Brazil
[2]Departamento de Zooloxía, Xenética e Antropoloxía Física, Facultade de Veterinaria, Universidade de Santiago de Compostela, Campus Terra, 27002 Lugo, Spain
[3]GENEAQUA S.L., Rúa Cántigas e Flores, 6 Entresuelo B, 27002 Lugo, Spain
[4]Departamento de Zooloxía, Xenética e Antropoloxía Física, Facultade de Bioloxía, Universidade de Santiago de Compostela, Campus Vida, 15782 Santiago de Compostela, Spain
[5]Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, UK
[6]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG Scotland, UK
[7]Department of Genetics, Evolution and Environment, The Darwin Building, London, WC1E 6BT, UK

[†]Corresponding author:
Paulino Martínez, Departamento de Zooloxía, Xenética e Antropoloxía Física, Facultade de Veterinaria, Universidade de Santiago de Compostela, Campus Terra, 27002 Lugo, Spain
Telephone and fax: +34 982822428
e-mail: paulino.martinez@usc.es
Fausto Foresti, Departamento de Morfologia, Instituto de Biociências, UNESP, 18618-689 Botucatu, São Paulo, Brazil
Telephone :+55 14 3880 0467
e-mail: fforesti@ibb.unesp.br

Running title: *Characidium* sex chromosomes evolution

Word count for main text: 6180 words

* These authors contributed equally to this work

Abstract

Fish species exhibit substantial variation in the degree of genetic differentiation between sex chromosome pairs, and therefore offer the opportunity to study the full range of sex chromosome evolution. We used Restriction site Associated DNA sequencing (RAD-seq) to study the sex chromosomes of *Characidium gomesi*, a species with conspicuous heteromorphic ZW/ZZ sex chromosomes. We screened 9,863 SNPs, corresponding to ~1 marker/100 kb distributed across the genome for sex-linked variation. With this dataset, we identified 26 female-specific RAD loci, putatively located on the W chromosome, as well as 148 sex-associated SNPs showing significant differentiation (average F_{ST} = 0.144) between males and females, and therefore in regions of more recent divergence between the Z and W chromosomes. Additionally, we detected 25 RAD loci showing extreme heterozygote deficiency in females but which were in Hardy-Weinberg equilibrium in males, consistent with degeneration of the W chromosome and therefore female hemizygosity. We validated seven female-specific and two sex-associated markers in a larger sample of *C. gomesi*, of which three localised to the W chromosome, thereby providing useful markers for sexing wild samples. Validated markers were evaluated in other populations and species of the genus *Characidium*, this exploration suggesting a rapid turnover of W-specific repetitive elements. Together, our analyses point to a complex origin for the sex chromosome of *C. gomesi* and highlight the utility of RAD-seq for studying the composition and evolution of sex chromosomes systems in wild populations.

Key words: *Characidium*, sex chromosome evolution, female-heterogamety, RAD sequencing, repetitive elements, comparative mapping

Introduction

Sex chromosomes are usually defined as a pair of chromosomes which carry the sex determination (SD) locus, and they typically occur as either male (XX/XY) or female (ZW/ZZ) heterogamety in diploid organisms (Martínez *et al.,* 2014). Sex chromosome divergence is often assumed to result from selection against recombination between the X and Y or Z and W chromosomes in order to maintain the association between a sex determining allele and a sexually antagonistic allele at a locus in close proximity (Bull, 1983). The loss of recombination on the sex-limited W or Y can result in degenerative processes, such as the accumulation of recessive deleterious mutations and repetitive DNA elements (Ellegren, 2011). If allowed to progress for a sufficient period of time, this process can produce major differences in size and gene content, referred to as sex chromosome heteromorphy (Ohno, 1967; Bull, 1983).

In contrast to mammals and birds, which show highly heteromorphic sex chromosomes that are conserved within each clade, fish exhibit a wide variety of non-orthologous sex chromosomes that have emerged independently throughout their evolution. This is reflected both by the rapid rate of sex chromosome turnover as well as significantly reduced heteromorphism compared to mammals and birds (Mank and Avise, 2009; Bachtrog *et al*., 2014). This rapid turnover has produced many examples of different sex chromosome systems between congeneric species, and even different systems within species (Martínez *et al*., 2014).

Neotropical fish in particular show a remarkable diversity of sex determination mechanisms, and this diversity offers a natural laboratory to explore and test evolutionary hypotheses about the origin and evolution of sex chromosomes. Different sex chromosome systems have been described by classical and molecular cytogenetics, including conspicuous sex chromosome heteromorphisms (Oliveira *et al.,* 2009), however, aside from recent work in *Poecilia reticulata* (Wright *et al*., 2017), sequencing-based studies on sex chromosome evolution have not yet been conducted in Neotropical species.

Characidium, a genus within the order Characiformes, shows a wide Neotropical distribution. All cytogenetic studies on this genus have thus far revealed a conserved diploid chromosome number (2n=50), and most analyses have revealed visibly heteromorphic female heterogametic sex chromosomes (Scacchetti *et al.,* 2015), suggesting that the ZW/ZZ chromosome system of *Characidium* originated once in the ancestor of the genus (Pansonato-Alves *et al.,* 2014). Following this putative single origin, the sex chromosomes in *Characidium* diversified among different species and populations, as there is significant inter- and intra-specific cytogenetic variation in size and heteromorphism (Scacchetti *et al.,* 2015).

In order to characterize the degree of divergence of the sex chromosomes, as well as variation across populations and related species, we used restriction-site associated DNA sequencing (RAD-seq) (Baird *et al.,* 2008), which permits the simultaneous discovery and robust scoring of large numbers of single nucleotide polymorphisms (SNPs) across many individuals. This approach now provides marker densities appropriate for detailed studies of sex chromosomes characterization and its evolution in fish (Martínez *et al.,* 2014; Pan *et al.,* 2016)

and RAD-seq has been applied for the identification of SD regions through linkage analysis in model and aquaculture (Robledo *et al*., 2017) as well as wild (Wilson *et al.*, 2014; Böhne *et al.*, 2016) species. Results presented here provide new information into the differentiation of the W chromosome and its variation across populations and species within the genus *Characidium*.

Material and Methods

Biological material

Sexually mature adults of *C. gomesi* were collected in the Água da Madalena tributary, Paranapanema River Basin Brazil (PR onwards; Figure 1). Gonad inspection under light microscope and cytogenetic analyses (Figure 2) were used to sex individuals and to check for convergence between histological and cytogenetic data. In order to minimize false positives, a total of 21 females and 18 males of *C. gomesi* from the PR population were used to construct RAD-seq libraries.

DNA from additional sexed specimens from the PR population (14 females and 7 males) and from another distant population (10 males and 10 females, Alambari River, Tietê River Basin, Brazil; TR onwards; Figure 1) were analysed to validate the identified sex-associated markers and to check for their intraspecific conservation. Furthermore, specimens from two other *Characidium* species, *C. zebra* (5 males and 5 females), a basal *Characidium* species which lacks sex chromosome heteromorphism, and *C. pterostictum* (5 males and 5 females), a distant relative of *C. gomesi* with significant sex chromosome heteromorphism (Pansonato-Alves *et al.*, 2014), were used for testing trans-specific conservation of sex-associated markers identified in *C. gomesi*.

All samples were collected in accordance with Brazilian environmental protection legislation (collection permission MMA/IBAMA/SISBIO—number 3245), and the procedures for sampling, maintenance and analysis of the specimens were performed in compliance with the Brazilian College of Animal Experimentation (COBEA) procedures and was approved (protocol 595) by the Bioscience Institute/UNESP Ethics Committee on use of animals (CEUA).

RAD sequencing and SNP genotyping

Genomic DNA from 21 females and 18 males of *C. gomesi* was extracted using the NucleoSpin Tissue Kit (Macherey-Nagel) and treated with RNAse to remove residual RNA from the samples. DNA quantity and quality were evaluated by fluorescence (Qubit) and agarose gels prior to library construction.

In order to help achieve an even representation of sequenced individuals, two RAD libraries were prepared, one with 20 individuals, the second with 19 individuals, both with approximately equal numbers of each sex. The RAD library preparation protocol, including the design of RAD-specific P1 and P2 paired-end adapters and library amplification PCR primer sequences, followed Houston *et al*. (2012). Briefly, each sample (n = 39; 200 ng DNA) was individually digested with 1.6 U SbfI high fidelity restriction enzyme (New England Biolabs; NEB) in 1× Reaction Buffer 4 (NEB) at 37 °C for 45 min. The reactions (10 µL final volumes) were then heat inactivated at 65°C for 20 min. Individual specific P1 adapters, each with a unique 5 or 7 base barcode (Table S1) were ligated to the SbfI digested DNA, at 22 °C for 90 min, by adding 0.5 µL 100 nM P1 adapter, 0.12 µL 100 mM rATP (Promega), 0.2 µL 10× Reaction Buffer 2 (NEB), 0.1 µL T4 ligase (NEB, 2 M U/mL) and reaction volumes made

up to 12 µL volume with nuclease free water. Following ligation, the samples were heat inactivated at 65 °C for 20 min, cooled to room temperature, then combined into one or other of two library pools. Shearing (Covaris S2 sonication) and initial size selection (c. 200–500 bp) by agarose gel separation of both library pools was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation and library amplification. An equimolar combination of two P2 adapters with 5 and 6 base barcodes (1 µL of 10 µM P2 adapter mix per library) was used to identify each library. A total of 150 µL of each amplified library (16 PCR cycles) was prepared and size selected (c. 320–650 bp) by gel electrophoresis. Following a final gel elution step into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were quantified by fluorimetry.

Equimolar amounts of both libraries were combined and sequenced in one lane of an Illumina Genome Analyzer II (100 base paired-ends (PE) reads) at the Wellcome Trust Centre for Human Genetics Sequencing Platform. Raw reads retrieved from the sequencing platform (NCBI BioProject: PRJNA391395) were processed using Stacks v1.08 (Catchen *et al.*, 2011). First, the *process_radtags* module was used to demultiplex raw reads of each individual, discarding reads with uncalled bases, missing restriction site, ambiguous barcodes or average quality score below 20. Barcodes were also removed and all Read 1 sequences (i.e. those starting at the RE site) were 3’ trimmed to 93bp. Next, *denovo_map.pl* was used to align these processed reads into exactly-matching stacks, and to score SNPs at each locus using a maximum likelihood framework. The main parameters were as follows: minimum stack depth (M = 3), maximum nucleotide mismatches allowed within stacks (m = 2), and

mismatches between sample tags when building the catalogue (n = 1). Third, two datasets were extracted from the SNP data: i) the *populations* module was used to generate an unfiltered set of all RAD loci, from which lists of RAD loci present in all females but not in males and *vice versa* were manually identified; ii) the *export_sql.pl* and *populations* modules were used to select a set of highly consistent, robust SNPs. For the latter, RAD-loci with a minimum depth of 10 reads, containing only one SNP with two variants, and genotyped at least in ~75% of the 28 samples were selected. Finally, the 3' end read sequences (Illumina P2 reads) of the RAD loci selected in i) and ii) were retrieved and collated using the *sort_read_pairs.pl* module. As the 3' end of RAD-tags are generated by random shearing it is possible to assemble multiple reads from the same RAD-tag into longer more informative contigs (Etter *et al*., 2011). CAP3 software was employed for this (Huang and Madan, 1999) using the suggested parameters in the CAP3 manual for short reads assembly.

Identification, annotation and population parameters of sex-linked markers in C. gomesi

Two different types of sex-linked markers were identified in the PR samples used for RAD-seq. First, female-specific RAD loci were defined as sequences only identified in females from the unfiltered outputs dataset. Second, sex-associated markers were defined as SNPs with significant genotypic association with sex, starting from the filtered SNP dataset. In light of the likely occurrence of false positives when analysing very large numbers of markers, two statistical approaches were used to identify confident sex-associated SNPs ($P < 0.05$). First, we calculated exact G tests using default parameters of GENEPOP

(Rousset, 2008). Second, we calculated a logistic regression based genome-wide association (GWAS) strategy using the fast score test for association between a trait and genetic polymorphism implemented in GenABEL (Aulchenko *et al*., 2007), with sex coded as a binary trait. The common set of markers identified with both approaches was considered as consistently associated with sex, while those markers identified by only one the two analytical methods were considered suggestive, but not included in the sex-associated SNP list.

RAD loci identified with both approaches were annotated using BLASTn and BLASTx homology searches (Altschul *et al.*, 1990) against NCBI's nr/nt database [http://www.ncbi.nlm.nih.gov/blast] using both the 5' end sequence and the 3' end contig sequence of each RAD locus. To assign predictive genomic locations for the sex-linked RAD loci of *C. gomesi*, comparisons were made to the genomes of the blind cave fish (*Astynax mexicanus, Characidae*; NCBI BioProject accession PRJNA89115), the closest related species with an assembled genome within Characiformes, and zebrafish (*Danio rerio, Cyprinidae*; Ensembl GRCz10), the fish species with the most extensive genomic resources and best quality genome assembly within the superorder Ostariophysi. Searches for homology (E-value < 10^{-5}) against these reference genomes were carried out using both 5' and 3' end sequences of each RAD locus. Gene mining around the RAD locus in blind cave fish and zebrafish genomes (2 Mb windows: RAD locus position ± 1 Mb) was performed with BioMart [www.ensembl.org], to identify relevant genes related to gonad differentiation. Additionally, all sex-linked RAD loci were screened with Repeatmasker to identify putative interspersed repeats and low complexity regions (www.repeatmasker.org; Smit *et al.*, 2015).

In the case of sex-associated SNPs, the relative coefficient of genetic differentiation (F_{ST}) was used to measure the extent of genetic differentiation between male and female groups using GENEPOP (Rousset, 2008; permutation test; $P < 0.05$). Theoretically, the maximum F_{ST} expected in a ZW/ZZ system assuming fixation of one allelic variant in the Z and another in the W chromosomes is $F_{ST} = 0.5$ (Kirkpatrick and Guerrero, 2014), and we used threshold to evaluate the magnitude of observed F_{ST}.

Deviation from Hardy-Weinberg (HW) expectations (exact tests; $P < 0.05$) and estimation of the F_{IS} fixation index were obtained for all markers, both computed by GENEPOP. In a ZW/ZZ chromosome pair, excess heterozygosity in the putative non-recombining region might occur due to genetic differentiation between the Z and W chromosomes. However, heterozygote deficiency may arise due to null alleles caused by degeneration of the W chromosome (Mank and Avise, 2009). Furthermore, in highly evolved SD chromosome systems, the heterogametic sex may be hemizygous for Z-linked markers, thus we would not expect heterozygous females at those loci where the W has sufficiently degraded. We also used linkage disequilibrium (LD) to confirm sets of loci putatively linked to the sex chromosome pair, again using GENEPOP (exact test; $P < 0.05$). To accommodate multiple test issues, we compared the proportion of significant LD deviations of a putative sex-linked SNP set *versus* the average genome LD, estimated using 20 samples of 100 SNPs each randomly chosen among the 9,863 identified SNPs. A frequency distribution of the proportion of LD using the 20 random SNP samples was constructed and the confidence intervals obtained. LD between sex-linked loci should be higher than between an average random SNP sample, and further, LD

would likely increase at those regions where recombination between Z and W chromosomes is restricted.

Validation of sex-linked markers

Female-specific RAD sequences were individually validated by PCR screening on an extended panel of 35 females and 25 males from the PR population. Primers were designed using Primer3 (Untergasser *et al.*, 2012) and PCR conditions were those provided by the program (Table S2). Sex-associated SNPs showing high F_{ST} were selected for their validation in the same PR population. Validation was hampered by the short length of the 5' PE (93 bases) and the lack of the species reference genome. Therefore, we scanned the 3' end of assembled contigs with sex-associated SNPs, a region more suitable for primer design (usually > 300 bases), as this end is more likely to be in LD with the sex associated SNPs identified. Primers for the selected SNPs were designed (Table S3) and samples genotyped on the MALDI-TOF mass spectromety platform (Sequenom,San Diego, CA, USA) at the University of Santiago de Compostela.

In order to investigate the inter-population and inter-specific conservation of sex-associated markers, we also amplified our PCR primers on 20 individuals (10 males and 10 females) from the TR *C. gomesi* population (Figure 1), as well as *C. pterostictum* and *C. zebra,* with 5 males and 5 females in each species, using the protocol described above.

Assessment of W-linked markers: microdissection and fluorescent in situ hybridisation (FISH)

A *C. gomesi* W-specific chromosome library was constructed through microdissection of the W chromosome and amplified using the GenomePlex Single Cell Whole Genome Amplification Kit (WGA4, Sigma-Aldrich). This library was then used as the DNA template to verify W-linkage using the screening methodologies outlined above. In addition, the amplified female-specific PCR products were labelled with digoxigenin-11-dUTP (Roche) to verify their location and distribution in *C. gomesi* genome (or W chromosome) through subsequent FISH experiments.

FISH analysis was performed as described in Scacchetti *et al*. (2015). Briefly, slides were incubated with RNAse (50μg/mL) for 1 h at 37°C, and the chromosomal DNA was denatured in 70% formamide/2 × SSC for 5 min at 70°C. For each slide, 30 μl of hybridisation solution (containing 200 ng of labeled probe, 50% formamide, 2xSSC and 10% dextran sulphate) was denatured for 10min at 95 °C, then dropped onto the slides and allowed to hybridise at 37°C in a moist chamber for 36 h. Post-hybridisation, all slides were washed in 0.2 × SSC/15% formamide for 20 min at 42°C, followed by a second wash in 0.1 × SSC for 15 min at 60 °C and a final wash at room temperature in 4 × SSC/0.5% Tween for 10 min. Probe detection was carried out with anti-digoxigenin-rhodamine (Roche), and the chromosomes were counterstained with DAPI (4',6-diamidino-2-phenylindole, Vector Laboratories) and visualised by optical photomicroscopy (Olympus BX61). Images were captured using Image Pro Plus 6.0 software (Media Cybernetics).

Results

RAD sequencing: SNP calling and genotyping

After demultiplexing and filtering for quality control, we recovered 247,743,386 PE reads, 87% of the total raw read count. Two females were removed from further analyses due to the very low number of reads, leaving us with 19 females and 18 males for further analysis with an average of 6,664,512 filtered reads per sample (Table S1).

We generated a catalog of 360,754 unique RAD loci with the Stacks pipeline, of which 89,572 were polymorphic. The number of unique RAD-tags in each sample ranged from 62,945 to 83,463 (Table S1). After applying the final filtering steps, we retained 9,863 putative biallelic SNPs with a minimum depth of 10 reads and genotyped in at least 28 individuals, with similar average number of reads for both sexes (females: 60.8, sd = 29.6; males: 53.1, sd = 26.1).

Sex-linked genetic markers: annotation and characterisation

RAD-tags in PR samples of *C. gomesi* were analysed for two types of sex-linkage. First, we identified 26 female-specific RAD loci sequences in the full panel of 19 females, which are consistent with W-linkage in regions that are sufficiently distinct from the Z chromosome (Table S4a). Interestingly, the average number of reads per individual for the female-specific RAD loci (31.8, sd = 18.1) was very similar to that observed for the whole RAD locus dataset (38.8, sd = 29.8), suggesting that our W-specific set represents a limited proportion of the *C. gomesi* genome.

Seven of the female-specific RAD loci showed sequence similarity with protein related genes, including hydroxysteroid 17-beta dehydrogenase 3 (*hsd17β3*), a steroidogenic factor related to gonad differentiation (Mindnich *et al.*, 2005) (Table S4b). Around half of the female-specific RAD loci were related

to various types of repetitive elements as defined by a range of criteria, with two annotated as transposable elements and seven with ≥ 5 hits in the reference genomes, especially that of the blind cave fish (average = 33.6 hits, sd = 34.4). Additionally, four were annotated as pol-like proteins, suggesting a relationship to retro-elements and two showed very long microsatellite tracts (> 100 bp). Only two of the female-specific RAD loci rendered a unique hit in the blind cave fish genome and one could be anchored on the genetic map to linkage group (LG) 20. Additionally, the *hsd17β3* gene was located in LG6 and LG8 of blind cave fish and zebrafish genomes, respectively. As expected in a female-heterogametic species, we observed no male-specific RAD sequences.

We scanned the 9,863 SNPs for evidence of significant allelic differentiation (F_{ST}) between males and females ($P < 0.05$) (Table S5), which would be indicative of regions where the Z and W chromosomes have started to diverge but which still retain significant homology. The 75% genotyping threshold used to retain RAD-loci (14 for females and males) ensured a minimum number of individuals per sex in order to control the number of false positives in our sample (Brelsford *et al*., 2017). The average sample size used for SNP genotyping was close to the total number, the mean number of females and males analysed per locus being 18.3 and 17.0, respectively. We identified 148 consistent sex-associated markers common to both statistical approaches used (exact G tests and GWAS; $P < 0.05$) (Table S6a). F_{ST} values of this set ranged between 0.090 and 0.299 with a mean of 0.144 (sd = 0.046) (Table S6b). Ten loci showed an $F_{ST} > 0.250$ and SNP with the highest $F_{ST} = 0.299$ suggests notable divergence between the Z and W, given the maximum expected $F_{ST} = 0.500$. The F_{ST} frequency distribution of this sex-associated SNP set and a 500

random non-associated reference set (Figure 3) showed distinct modes, and the small overlap between the distributions confirm the robustness of our statistical approach. In fact, the six non-associated SNPs overlapping at the left end of the sex-associated distribution were suggestive, namely they were identified with only one approach, GWAS or exact G tests.

We searched public databases to identify candidate genes related to sex differentiation for the 148 sex-linked SNPs (Table S6c). A total of 116 sequences were annotated (78.4%), 16 of which (13.8%) were transposable elements (TE) and five included long microsatellite tracts (> 100 bp). Two RAD loci were annotated to genes involved in gonad differentiation and reproduction, including *nectin-2*, a gene that codifies for a junction molecule crucial for spermatogenesis (Zhang and Lui, 2014) and also suggested to play a role in the follicular development of the mouse ovary (Kawagishi *et al.*, 2005), and *tgfb2*, a gene that plays a pivotal role in gonad development (Ergin *et al.*, 2008). *Nectin-2* and *tgfb2* were located at LG23 and LG10 of blind cave fish and at LG19 and LG16 of zebrafish, respectively.

BLASTn searches of the 148 sex-associated RAD loci to the blind cave fish and the zebrafish genomes were undertaken to potentially identify regions of synteny (E-value < 10^{-5}, Table S6c). A total of 24 sex-associated SNPs mapped to unique positions of the *A. mexicanus* genome, and 5 to the *D. rerio* genome (Table S6c), which is consistent with the closer phylogenetic relatedness between *A. mexicanus* and *C. gomesi*. Nineteen out of 24 mapped loci were located in the genetic map of the blind cave fish (Carlson *et al.*, 2015), with LG8, LG10, LG15, LG16 and LG22 each containing two hits. Most mapped loci showed F_{ST} values around the mean (0.144; range: 0.094 - 0.278), except

L37075_37 on LG22 (F_{ST} = 0.278). Gene mining within 1 Mb on either side of the mapped RAD loci identified several genes involved in gonad differentiation, particularly in the blind cave fish genome. Two notable genes were found at LG8 include sperm adhesion molecule 1 gene (*spam1*), which is involved in sperm penetration through the cumulus matrix (Kimura *et al.,* 2009), and stimulated by retinoic acid 6 gene (*stra6*), which is involved in vitellogenesis (Levi *et al.,* 2012). Additionally, a gene that functions in female differentiation, b-catenin1 (*ctnnb1*) (Chassot *et al.,* 2014), was tightly linked to another RAD locus mapping at LG10, whereas two genes related to the steroid hormone mediated signalling pathway, i.e. nuclear receptor subfamily 4 group A member 1 (*nr4a1*; Abdou *et al.,* 2013) and retinoic acid receptor, gamma b (*rargb*), closely mapped in LG22. Other important genes related to gonad development, bone morphogenetic protein 15 (*bmp15;* Han *et al.,* 2015) and cytochrome P450, family 26, subfamily b, polypeptide 1 (*cyp26b1*; Saba *et al.,* 2014) were found in LG16, while the estrogen receptor 2b gene (*esr2b*; Delalande *et al.,* 2015) in zebrafish LG13.

Our analysis is somewhat limited by the relatively small number of individuals screened (37) and the large number of loci analysed (9,863). Nevertheless, we observed several important indicators of differentiation between the Z and W chromosomes. First, we observed a significant deficit of heterozygotes within the 148 sex-associated SNP dataset compared to the genomic average ($P < 0.05$; 19.6% vs 4.2%, respectively; Table S6b), suggesting a higher proportion of null alleles in the sex chromosome pair. This would be expected if the W and Z chromosomes have diverged significantly from one another. Furthermore, we detected 25 loci with high heterozygosity ($H_E \geq 0.4$)

showing extreme heterozygote deficiency in females (F_{IS} = 1) but in HW equilibrium in males (Table S7). Only two markers showing these characteristics were detected in males, which suggest a low proportion of false positives within the 25 SNPs identified in females. These loci might be located in regions where the W has significantly diverged, for which females are hemizygous.

The proportion of linkage disequilibrium (LD) departures ($P < 0.05$) in the 148 sex-associated (8.4% pair-wise LD) and in the 25 female-limited (17.7%) SNP sets was much higher than that observed in the genome background (20 random 100 SNP samples over the 9,863 loci; mean = 1.725% ± 0.062%; 95% CI: 1.595 - 1.855), strongly supporting their linkage. Moreover, pair-wise LD between the 148 and 25 SNP sets was also much higher than the background (9.2%), supporting their linkage to the ZW pair.

Validation of the female-specific and sex-associated markers

The 26 female-specific RAD loci were validated by PCR on the expanded sample of 35 females and 25 males of *C. gomesi* from the PR population. Two RAD loci showed high sequence similarity, and a single primer pair was designed for their amplification (Table S2). Seven out of 26 loci showed different amplification patterns between the sexes (Figure 4). Most of these showed a single prominent band of the expected size in females, while no band or a different banding pattern was observed in males. The remaining 19 RAD loci showed a similar banding pattern in both sexes. Nevertheless, this fact does not preclude diagnostic differences between males and females for these RAD loci, possibly reflecting diagnostic polymorphism within the restriction enzyme site associated with the specific tag.

A total of 15 of the 148 initially identified sex-associated RAD loci were selected for validation in the expanded PR sample (Table S3). Thirteen primer sets of the 15 selected RAD loci were successfully genotyped in the PR population (Tables 1 and S8). Nearly half of these SNPs showed lower polymorphism than their 5' counterparts (H_E < 0.2), diminishing statistical power to detect differences between male and female subpopulations. Of the seven remaining loci (H_E > 0.25), two showed significantly greater differentiation between males and females than the 5' SNP counterpart (F_{ST} > 0.290), supporting sex-linkage. Interestingly, 33.3% positive F_{IS} values were detected in females but not in males (P < 0.05), supporting the presence of null alleles linked to the W chromosome.

Trans-population and trans-specific conservation of female-specific and sex-associated markers

Four out of seven validated female-specific markers were consistent in both the PR and TR populations. The remaining three loci showed a similar banding pattern in males and females (Figure 4). Notably, none of the primer sets for female-specific markers produced female-specific PCR products in *C. pterostictum* and *C. zebra*, and assays of the 13 validated sex-associated markers in the PR population failed or were monomorphic in these two species.

Location of sex validated markers on the C. gomesi genome

Two female-specific markers and one sex-associated marker produced PCR products of the expected size when amplified on the microdissected W chromosome library (Figure 4), confirming their location on this chromosome.

The seven PCR amplified female-specific markers were used as probes for fluorescent in situ hybridisation (FISH) on *C. gomesi* metaphase plates to ascertain their location, but no specific signals of hybridisation were detected. In keeping with their known TE sequence composition, three of these generated weak scattered signals throughout the karyotype of the species, suggesting multiple locations (Figure S1).

Discussion

Teleosts show a high turnover of sex-determining mechanisms (Martínez *et al.*, 2014), including sex chromosomes. This rapid origin makes fish a useful clade for studying the early stages of sex chromosome evolution. In this study, we characterised the sex chromosomes at both population and species level in the genus *Characidium* using a reduced representation genome sequencing strategy (RAD-seq). Our study is based on 9,863 filtered SNPs, equivalent to roughly 1 SNP per 100 kb and 1 RAD locus per 3 kb, based on the reported genome size of the species of roughly1 Gb (Carvalho *et al*., 1998).

Clear heteromorphism between the Z and W chromosomes has been reported for both size and C-banding pattern in *C. gomesi* (Maistro *et al*., 1998), and this is consistent with the proportion of sex-linked RAD sequences and SNPs we identified. The W chromosome of *C. gomesi* shows C-positive banding in mitotic plates suggesting a condensation estate (Figure 2; Maistro *et al.*, 1998) theoretically related to the accumulation of repetitive sequences (Charlesworth, 1991). Thus, the genomes of species with heteromorphic sex chromosomes, like *C. gomesi*, are expected to have more sex-specific restriction sites (and hence sex-specific RAD loci) than species with homomorphic sex chromosomes

(Gamble *et al.*, 2015). Consistent with the female-heterogametic sex chromosome system reported in this species, we recovered female-specific RAD loci, but not male-specific loci, and roughly half of the female-specific RAD loci showed typical features of repetitive elements.

An increasing number of studies using RAD-seq methods have been carried out in a range of organisms either to identify the sex determination region by identifying sex-associated molecular markers at the population or species level. In some cases, linkage mapping has been used to identify the sex determining locus. The first approach can be applied to wild populations based on linkage disequilibrium between polymorphic loci and the SD region, while the second method requires a mapping family panel (Gamble and Zarkower, 2014; Fowler and Buonaccorsi, 2016; Gamble, 2016; Brelsford *et al*., 2017; Robledo *et al*., 2017).

Inevitably, with small numbers of sexed individuals being screened for a very large number of SNP markers, false positive associations between sex and markers are likely to arise. While limited numbers of individuals were available for our study (19 females and 18 males), analyses from similar studies in other species suggest that data from 12-14 random individuals of each sex should be sufficient for relatively robust sex-linked analysis (Lambert et *al.*, 2016; Brelsford *et al.*, 2017). We validated approximately one third of the female-specific RAD loci in a broader sample of the PR population, thereby proving the usefulness of RAD-seq to develop sex-specific markers in non-model species (Gamble and Zarkower, 2014; Fowler and Buonaccorsi, 2016). Moreover, two of these loci were amplified by PCR from the W-chromosome library, confirming W-linkage. The failure of several other sex-specific marker assays to

amplify products from the W chromosome library may reflect methodological limitations of W-DNA library construction due to the gross process of microdissection of mitotic chromosomes and their subsequent random amplification.

Interestingly, some of the validated markers were population-specific, suggesting that a fraction of W-specific sequences may evolve rapidly, and populations located in a rather small geographic area may display different W-specific RAD-tags repertoires. Sex chromosomes often display higher rates of molecular evolution compared autosomes (Bachtrog, 2005; Filatov, 2005; Berlin *et al*., 2006; Shikano *et al*., 2011), and this indicates that differentiation between species should be higher for sex-linked regions. Consistent with that, we found that none of the *C. gomesi* female-specific loci amplified in the two other species evaluated (*C. zebra* and *C. pterostictum*). This finding could be due to several factors, including the rapid differentiation of the W chromosome, as reported recently in molecular cytogenetic studies of other Characiforms (Yano *et al*., 2016); the phylogenetic distance among the analysed species (Pansonato-Alves *et al*., 2014); or even possible transitions of sex-determining systems between *C. gomesi* and *C. zebra*.

It is worth noting that we detected a relatively small proportion of female-specific RAD loci (0.0007%), which is a much smaller fraction than expected based on the fact that the W chromosome constitutes 3% of the *C. gomesi* karyotype in metaphase plates. This is most likely due to a small repertoire of repetitive elements that lack SbfI sites on the W chromosome.

We identified 148 consistent sex-associated SNPs in *C. gomesi* in our PR population samples, and the highest F_{ST} value we observed between males and

females suggests close proximity of this marker to the putative SD gene. A subset of these markers was additionally validated, and some confirmed high differentiation. Furthermore, the RAD locus with the highest differentiation was shown to be W-linked, and LD evidence indicates that most of these markers are located on the same linkage group.

Annotation of the sex-linked RAD loci identified single-copy genes associated with sex differentiation. Among the most relevant, the hydroxysteroid 17-β dehydrogenase 3 (*hsd17β3*), a steroidogenic marker that catalyses the conversion of androstenedione to testosterone, was detected in the female-specific RAD loci. This gene is almost exclusively expressed in the testis of mammals, but it shows high ovarian expression in zebrafish (Mindnich *et al.*, 2005). Among the sex-associated RAD loci, we identified *nectin-2*, a gene involved in gonadal differentiation (Kawagishi *et al.*, 2005; Zhang and Lui, 2014), and *tgfb2*, an important mediator of growth and differentiation involved in germ cell and gonadal development (Ergin *et al.*, 2008; Ozgüden-Akkoç and Ozer, 2012). *Tgfb2* belongs to the *Tgb*-β superfamily whose members have been associated with important ovarian and testicular functions (Drummond, 2005; Fan *et al.*, 2012) and different components of the TGF-β signalling pathway have been found to be sex determining genes in several fish species (reviewed by Martínez *et al.*, 2014).

We used a comparative mapping strategy interrogating the reference genomes of other fish species to gain insight on the genomic organization of the *C. gomesi* SD region and, specifically, to identify putative SD candidates. The assembled genome and important genomic resources of zebrafish, a species of the superorder Ostariophysi, and the draft genome of blind cave fish, with fewer

genomic resources than zebrafish, but within the family *Characidae*, were used for this purpose. As expected, a much higher number of significant hits were retrieved in our study from the blind cave fish genome than the zebrafish. All three species have a 2n = 50 karyotype (Sola and Gornung, 2001; Kavalco and Almeida-Toledo, 2007), but there have been important genomic reorganizations between the blind cave fish and zebrafish based on synteny, and only two chromosomes of their karyotype show a consistent macrosyntenic pattern (Carlson *et al*., 2015). LG10 and LG22 show features, including number of hits, candidate genes and highest F_{ST}, suggesting that they may contain important regions orthologous to the *C. gomesi* sex chromosomes. However, our data indicate a complex origin, involving reorganizations of different chromosomes.

Degeneration of the W chromosome causes heterozygote deficiency in females, as the loss of restriction sites on the W chromosome would give rise to null alleles. In regions where the W has only begun to diverge from the Z, we expect a significantly higher frequency of W-specific null alleles, manifesting as significant deviations from the null hypothesis ($F_{IS} = 0$), as observed in our data (19.6% deviations for sex-associated loci vs 4.2% across the whole dataset). In regions where the W chromosome is highly diverged, this situation would be extreme ($F_{IS} = 1$) and females will be hemizygous for the Z chromosome, and we observed 25 SNPs with a pattern of female hemizygosity. This suggests heterogeneity in recombination suppression on the *C. gomesi* sex chromosomes, with an older stratum (25 markers), where recombination was suppressed earlier and the W chromosome is more degraded, and a relatively younger stratum (148 markers) where the W has only recently started to diverge. This is consistent with sex chromosomes in other fish species, including *G. aculaeatus* (Schulteiß

et al., 2015) and *P. reticulata* (Wright *et al.*, 2017), both of which show sex chromosome strata of differing ages. Importantly, we observed strong evidence of elevated LD among our female-specific SNPs, consistent with the loss of recombination on the W chromosome.

Taken together, our results indicate significant genetic differentiation between the Z and W sex chromosomes in the genus *Characidium*, illustrating the utility of RAD-seq for sex chromosome characterization in wild species. Both female-specific markers, expected in a ZZ/ZW system, as well as two types of sex-associated markers, suggesting different steps in the process of chromosome differentiation, were identified. The validation of sex-related markers in other populations and species of the genus *Characidium* suggested a quick evolution of sex chromosome associated sequences, as previously reported in other vertebrates. The existence of the genome of the blind cave fish, a closely related species included in the family *Characidae*, enabled the identification of SD candidate genes and suggested a complex evolution of the sex chromosome pair. From a practical perspective, the identified markers will be valuable for developing a suitable molecular tool for non-invasive sex identification in future population dynamics or ecological studies.

Acknowledgements

The authors thank Renato Devidé for help with obtaining samples, and Érica A. Serrano-Freitas and José C. Pansonato-Alves for providing tissue and cell suspensions. This study was supported by grants from the Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the European Research Council (grant agreements 26023 and 680951) and research funding from the Xunta de Galicia local government (GRC2014/010). MB and JBT were partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (Grant reference HR09011) and contributing institutions.

Supplementary information is available at Heredity's website.

References

Abdou HS, Villeneuve G, Tremblay JJ (2013). The calcium signaling pathway regulates leydig cell 358 steroidogenesis through a transcriptional cascade involving the nuclear receptor NR4A1 and the 359 steroidogenic acute regulatory protein. *Endocrinology* **154:** 511-520.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**: 1294–1296.

Bachtrog D (2005). Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. *Genome Res* **15**: 1393-1401.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R, Perrin N, Ross L, Valenzuela N, Vamosi

JC, Tree of Sex Consortium (2014). Sex determination: why so many ways of doing it? *PLoS Biol* **12**: e1001899.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, *et al*. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**: e3376.

Berlin S, Brandström M, Backström N, Axelsson E, Smith NG, Ellegren H (2006). Substitution rate heterogeneity and the male mutation bias. *J Mol Evol* **62**: 226-233.

Böhne A, Wilson CA, Postlethwait JH, Salzburger W (2016). Variations on a theme: Genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* **17**: 883.

Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N (2017). Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. *Mol Ecol Resour* **17**: 752-759.

Bull JJ (1983). Evolution of Sex Determining Mechanisms. Benjamin Cummings Press.

Carlson BM, Onusko SW, Gross JB (2015). A high-density linkage map for *Astyanax mexicanus* using genotyping-by-sequencing technology. *G3 (Bethesda)* **5**: 241–251.

Carvalho ML, Oliveira C, Navarrete MC, Froehlich O, Foresti F (1998) Nuclear DNA content determination in Characiformes fish (Teleostei, Ostariophysi) from the Neotropical region. *Genet Mol Biol* **25:** 49-55.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1:** 171-182.

Charlesworth B (1991). The evolution of sex chromosomes. *Science* **251**: 1030-1033.

Chassot AA, Gillot I, Chaboissier MC (2014). R-spondin1, WNT4, and the CTNNB1 signaling pathway: strict control over ovarian differentiation. *Reproduction* **148**: R97-110.

Delalande C, Goupil AS, Lareyre JJ, Le Gac F (2015). Differential expression patterns of three aromatase genes and of four estrogen receptors genes in the testes of trout (*Oncorhynchus mykiss*). *Mol Reprod Dev* **82**: 694-708.

Drummond AE (2005). TGFbeta signalling in the development of ovarian function. *Cell Tissue Res* **322**: 107-115.

Ellegren H (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet* **12**: 157 – 166.

Ergin K, Gürsoy E, Başimoğlu Koca Y, Başaloğlu H, Seyrek K (2008). Immunohistochemical detection of insulin-like growth factor-I, transforming growth factor-beta2, basic fibroblast growth factor and epidermal growth factor-receptor expression in developing rat ovary. *Cytokine* **43**: 209-214.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011). Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE* **6**: e18561.

Fan YS, Hu YJ, Yang WX (2012). TGF-β superfamily: how does it regulate testis development. *Mol Biol Rep* **39**: 4727-4741.

Filatov DA (2005). Substitution rates in a new *Silene latifolia* sex-linked gene, SlssX/Y. *Mol Biol Evol* **22**: 402-408.

Fowler BLS, Buonaccorsi VP (2016). Genomic characterisation of sex-identification markers in *Sebastes carnatus* and *Sebastes chrysomelas* rockfishes. *Mol Ecol* **25**: 2165-2175.

Gamble T (2016). Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Mol Ecol* **25**: 2114-2116.

Gamble T, Zarkower D (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol Ecol Resour* **14**: 902-913.

Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D (2015). Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among Gecko sex-determining systems. *Mol Biol Evol* **32**: 1296-1309.

Han H, Lei Q, Zhou Y, Gao J, Liu W, Li F, *et al*. (2015). Association between BMP15 gene polymorphism and reproduction traits and its tissues expression characteristics in chicken. *PLoS One* **10**: e0143298.

Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A, *et al*. (2012). Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**: 244.

Huang X, Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res* **9:** 868-877.

Kavalco KF, De Almeida-Toledo LF (2007). Molecular cytogenetics of blind mexican tetra and comments on the karyotypic characteristics of genus *Astyanax* (Teleostei, Characidae). *Zebrafish* **4**: 103-111.

Kawagishi R, Tahara M, Morishige K, Sakata M, Tasaka K, Ikeda W, *et al*. (2005). Expression of nectin-2 in mouse granulosa cells. *Eur J Obstet Gynecol Reprod Biol* **121:** 71-76.

Kimura M, Kim E, Kang W, Yamashita M, Saigo M, Yamazaki T, *et al*. (2009). Functional roles of mouse sperm hyaluronidases, HYAL5 and SPAM1, in fertilisation. *Biol Reprod* **81**: 939-947.

Kirkpatrick M, Guerrero RF (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* **197**:531-541.

Lambert MR, Skelly DK, Ezaz T (2016). Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DArTseq provide early insight into sex chromosome evolution. *BMC Genomics* **17**: 844.

Levi L, Ziv T, Admon A, Levavi-Sivan B, Lubzens E (2012). Insight into molecular pathways of retinal metabolism, associated with vitellogenesis in zebrafish. *Am J Physiol Endocrinol Metab* **302**: E626-E644.

Maistro EL, Mata EP, Oliveira C, Foresti F (1998). Unusual occurrence of a ZZ/ZE sex-chromosome system and supernumerary chromosomes in *Characidium* cf. *fasciatum* (Pisces, Characiformes, Characidiinae). *Genetica* **104**: 1-7.

Mank JE, Avise JC (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex Dev* **3:** 60-67.

Martínez P, Viñas AM, Sánchez L, Díaz N, Ribas L, Piferrer F (2014). Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front Genet* **5**: 340.

Mindnich R, Haller F, Halbach F, Moeller G, Hrabé de Angelis M, Adamski J (2005). Androgen metabolism via 17beta-hydroxysteroid dehydrogenase type

3 in mammalian and non-mammalian vertebrates: comparison of the human and the zebrafish enzyme. *J Mol Endocrinol* **35**: 305-316.

Ohno S (1967). *Sex Chromosomes and Sex Linked Genes*. Springer-Verlag, Berlin.

Oliveira C, Foresti F, Hilsdorf AWS (2009). Genetics of Neotropical fish: from chromosomes to populations. *Fish Physiol Biochem* **35**: 81-100.

Ozgüden-Akkoç CG, Ozer A (2012). Immunohistochemical localization of transforming growth factor β1 and β2 in mouse testes during postnatal development. *Biotech Histochem* **87**: 154-159.

Pan Q, Anderson J, Bertho S, Herpin A, Wilson C, Postlethwait JH, *et al*. (2016). Vertebrate sex-determining genes play musical chairs. *Comptes Rendus Biologies* **339**: 258-262.

Pansonato-Alves JC, Serrano EA, Utsunomia R, Camacho JPM, Costa-Silva GJ, Vicari MR, *et al*. (2014). Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus *Characidium. PLoS One* **9**: e107169

Robledo D, Palaiokostas C, Bargelloni L, Martínez P, Houston R (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev Aquacult* (doi: 10.1111/raq.12193).

Rousset F (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* **8:** 103-106.

Saba R, Wu Q, Saga Y (2014). CYP26B1 promotes male germ cell differentiation by suppressing STRA8-dependent meiotic and STRA8-independent mitotic pathways. *Dev Biol* **389**: 173-181.

Scacchetti PC, Utsunomia R, Pansonato-Alves JC, Vicari MR, Artoni RF, Oliveira C, *et al*. (2015). Chromosomal mapping of repetitive DNAs in

Characidium (Teleostei, Characiformes): genomic organisation and diversification of ZW sex chromosomes. *Cytogenet Genome Res* **146**: 136-143.

Schulteiß R, Viitaniemi HM, Leder EH (2015). Spatial dynamics of evolving dosage compenastion in a Young sex chromosome system. *Genome BiolEvol* **7:** 581-590

Shikano T, Natri HM, Shimada Y, Merilä J (2011). High degree of sex chromosome differentiation in stickleback fishes. *BMC genomics* **12**: 474.

Smit AFA, Hubley R, Green P (2015). RepeatMasker Open-4.0. <http://www.repeatmasker.org>

Sola L, Gornung E (2001). Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. *Genetica* **111**: 397–412.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, *et al*. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40:** e115.

Wilson CA, High SK, McCluskey BM, Amores A, Yan YL, Titus TA, *et al*. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* **198**: 1291-1308.

Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, *et al*. (2017). Convergent recombination suppression suggests a role of sexual selection in the guppy sex chromosome formation. *Nat Commun* **8**: 14251

Yano CF , Bertollo LAC, Ezaz T, Trifonov V, Sember A, Liehr T *et al*. (2016). Highly conserved Z and molecularly diverged W chromosomes in the fish genus Triportheus (Characiformes, Triportheidae). *Heredity* **118:** 276-283.

Zhang X, Lui WY (2014). Dysregulation of nectin-2 in the testicular cells: an explanation of cadmium-induced male infertility. *Biochim Biophys Acta* **1839**: 873-884.

Titles and legends to figures

Figure 1: Geographic map of sampling sites for *Characidium* species and populations used in this study.

Figure 2: Mitotic plate of *Characidium gomesi* female after C-banding, showing a strong C-band positive pattern on W chromosome.

Figure 3: Distribution of genetic differentiation (F_{ST}) between male and female subpopulations for the 148 sex-associated SNPs (dark gray) and a random sample of 500 SNPs not associated with sex (light gray).

Figure 4: Banding patterns of validated female-specific markers identified in *Characidium gomesi* in Paranapanema (PR) and Tiete (TR) populations. Amplification on the W chromosome DNA library is also shown.

Figure S1: Fluorescent *in situ* hybridisation (FISH) of *Characidium gomesi* mitotic plates with female-specific RAD locus probes.

Tables

Table 1: Genetic diversity and differentiation between male and female subpopulations for the 13 sex-associated SNPs of *Characidium gomesi*.

Table S1: RAD-seq methodology and filtering applied to identify RAD loci and call SNPs in *Characidium gomesi*.

Table S2: PCR conditions of the 26 female-specific markers identified through RAD-seq in *Characidium gomesi*.

Table S3: Primers and PCR conditions for the 15 sex-associated SNPs in *Characidium gomesi*.

Table S4: Main features of the female-specific RAD loci detected in *Characidium gomesi*.

Table S5: Population parameters for the 9,863 called SNPs in the female and male subpopulations of *Characidium gomesi*

Table S6: Main features of the 148 sex-associated SNPs of *Characidium gomesi* identified in this study

Table S7: Main features of the 25 SNPs compatible with female hemizyosity in *Characidium gomesi*.

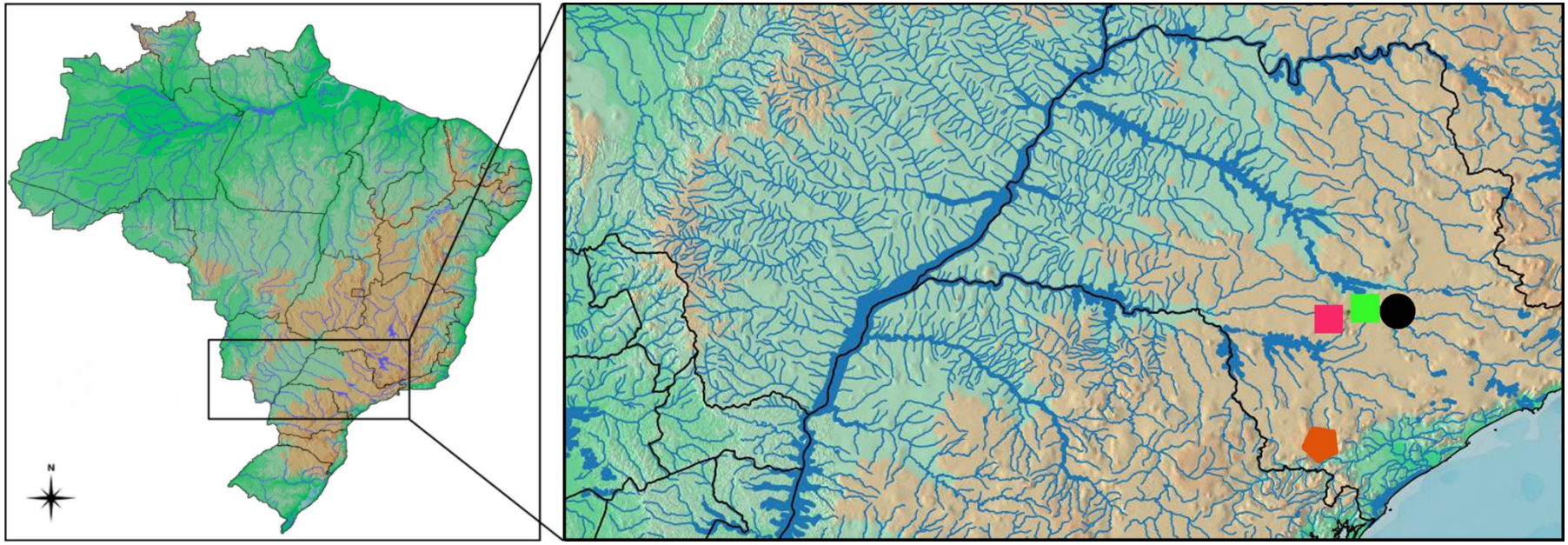
Table S8: Genotypes of the 13 evaluated sex-associated SNPs in the Paranapanema population (PR) of *Characidium gomesi* using Sequenom

Table 1: Genetic diversity and differentiation between male and female subpopulations for the 13 new 3' end validated SNPs detected in *Characicium gomesi*

	Females				Males					
SNP code	**K**	**Ho**	**He**	**Fis**	**K**	**SNP code**	**Ho**	**He**	**Fis**	**Fst**	**Fst 5' end**
L17552_28	2	0.185	0.171	-0.083	2	17552_28	0.059	0.059	0	-0.023	0,150
L23581_62	2	0	0.073	1,000*	1	23581_62	0	0	0	-0.016	0.192
L26007_70	2	0.074	0.307	0,763***	2	26007_70	0.111	0.203	0.46	-0.019	0.299
L29522_74	2	0.522	0.468	-0.128	2	29522_74	0.321	0.353	0.118	**0,298*** **	0.226
L30946_43	2	0.182	0.512	0,650**	2	30946_43	0.333	0.434	0,233	0.035	0.164
L37075_37	2	0.208	0.51	0,597**	2	37075_37	0,200	0.186	-0.077	**0,292*** **	0.278
L38879_69	2	0.16	0.15	-0.068	2	38879_69	0.333	0.286	-0.172	0.015	0,173
L53050_55	1	0	0	na	1	53050_55	0	0	na	na	0.183
L55368_62	2	0.963	0.509	-0.926	2	55368_62	1	0.514	-1	0	0.236
L59864_80	2	0,320	0.327	0,020	2	59864_80	0.167	0.246	0.329	-0,015	0,177
L7660_75	2	0.522	0.502	-0.039	2	7660_75	0.588	0.499	-0.185	-0.022	0,280
L80118_33	2	0.185	0.171	-0.083	2	80118_33	0.056	0.056	0	0.012	0,263
L8459_59	2	0.037	0.037	0	2	8459_59	0.056	0.056	0	-0.022	0,130

* P<0,0,5; ** P<0,01; *** P<0,001; in gray background loci with significant and high Fst in the 3' end tested SNP

Fig 1.



C. gomesi - Paranapanema River

C. gomesi - Tietê River

C. zebra

C. pterostictum

Fig 2.

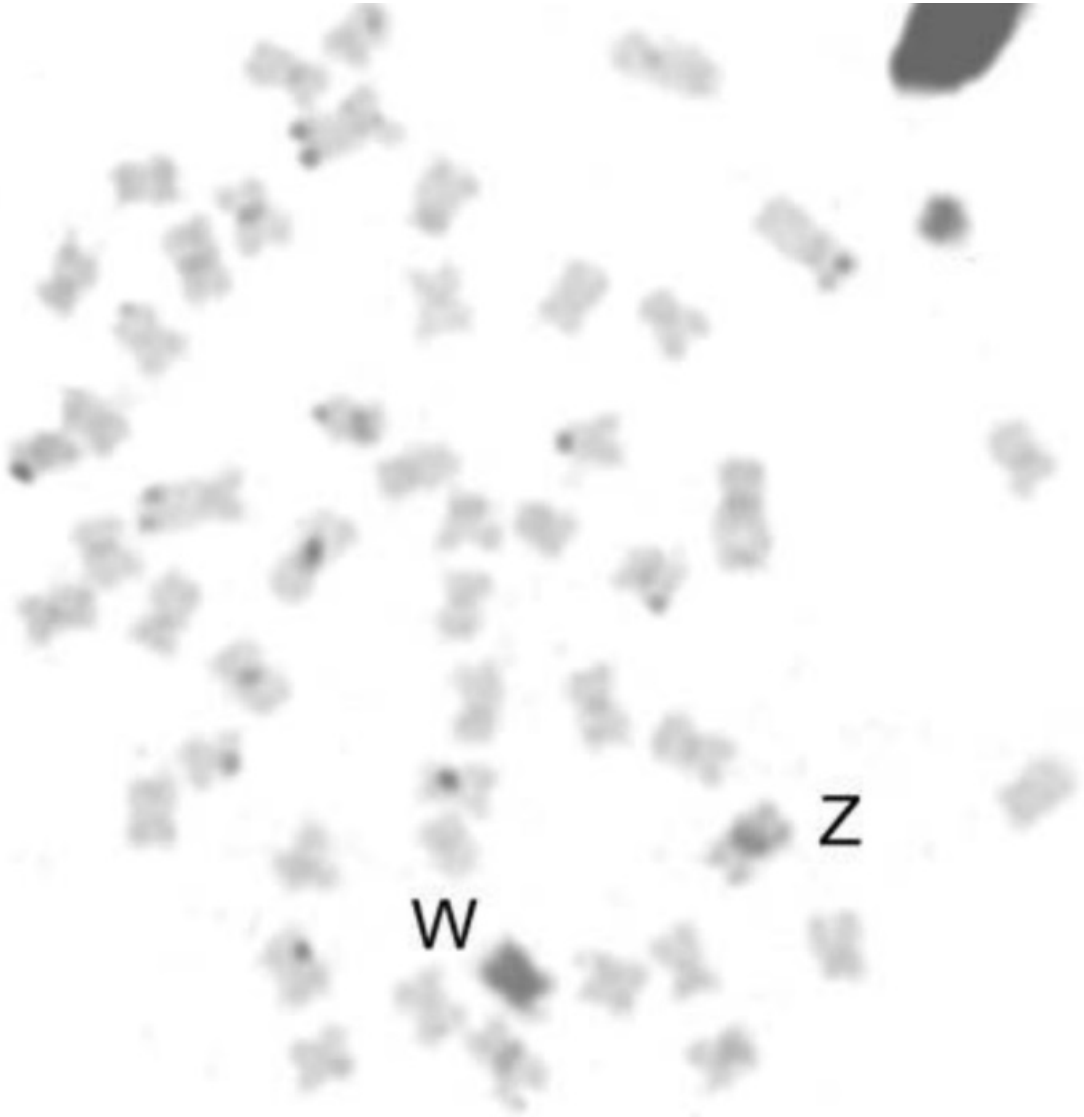


Fig 3.

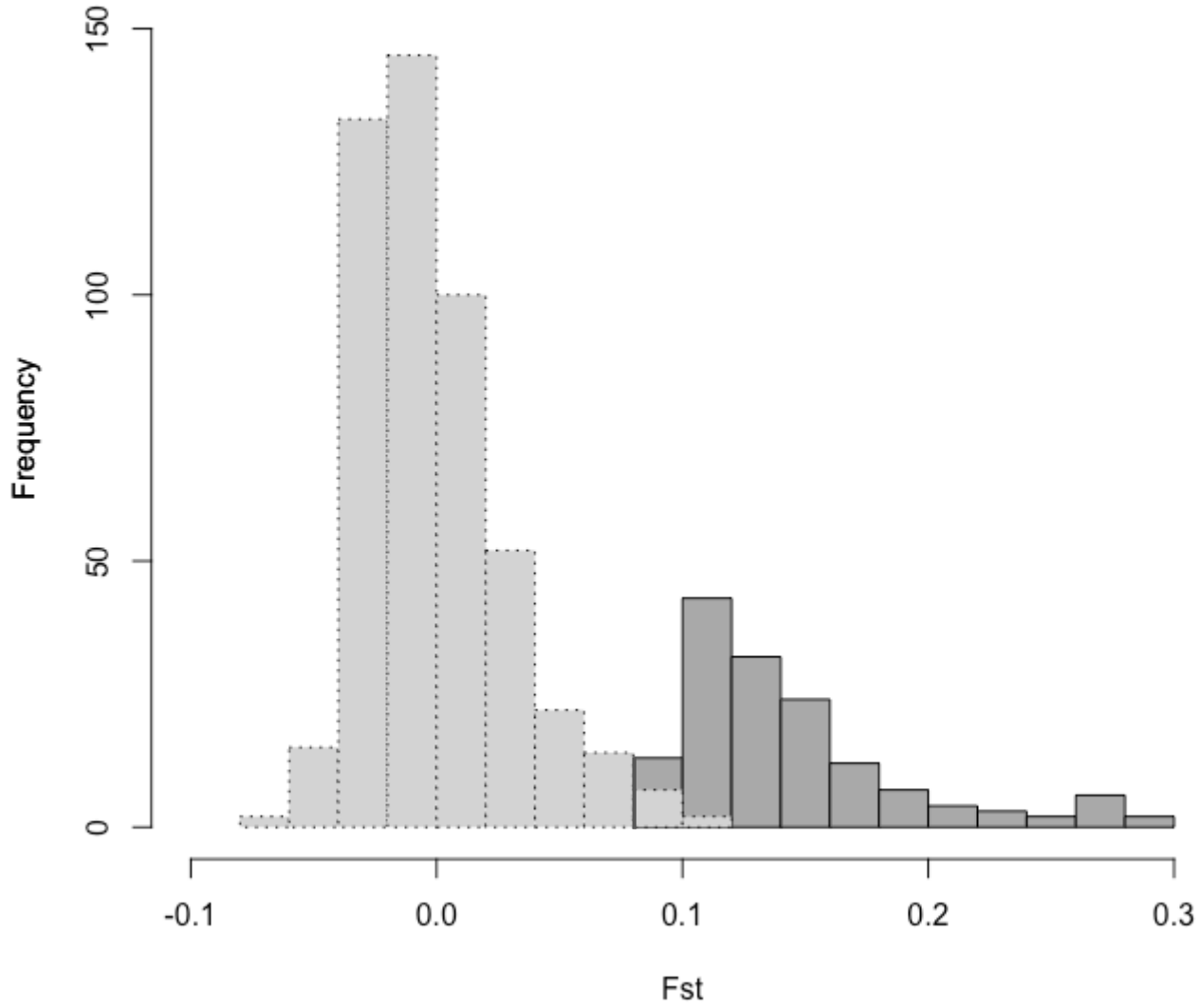
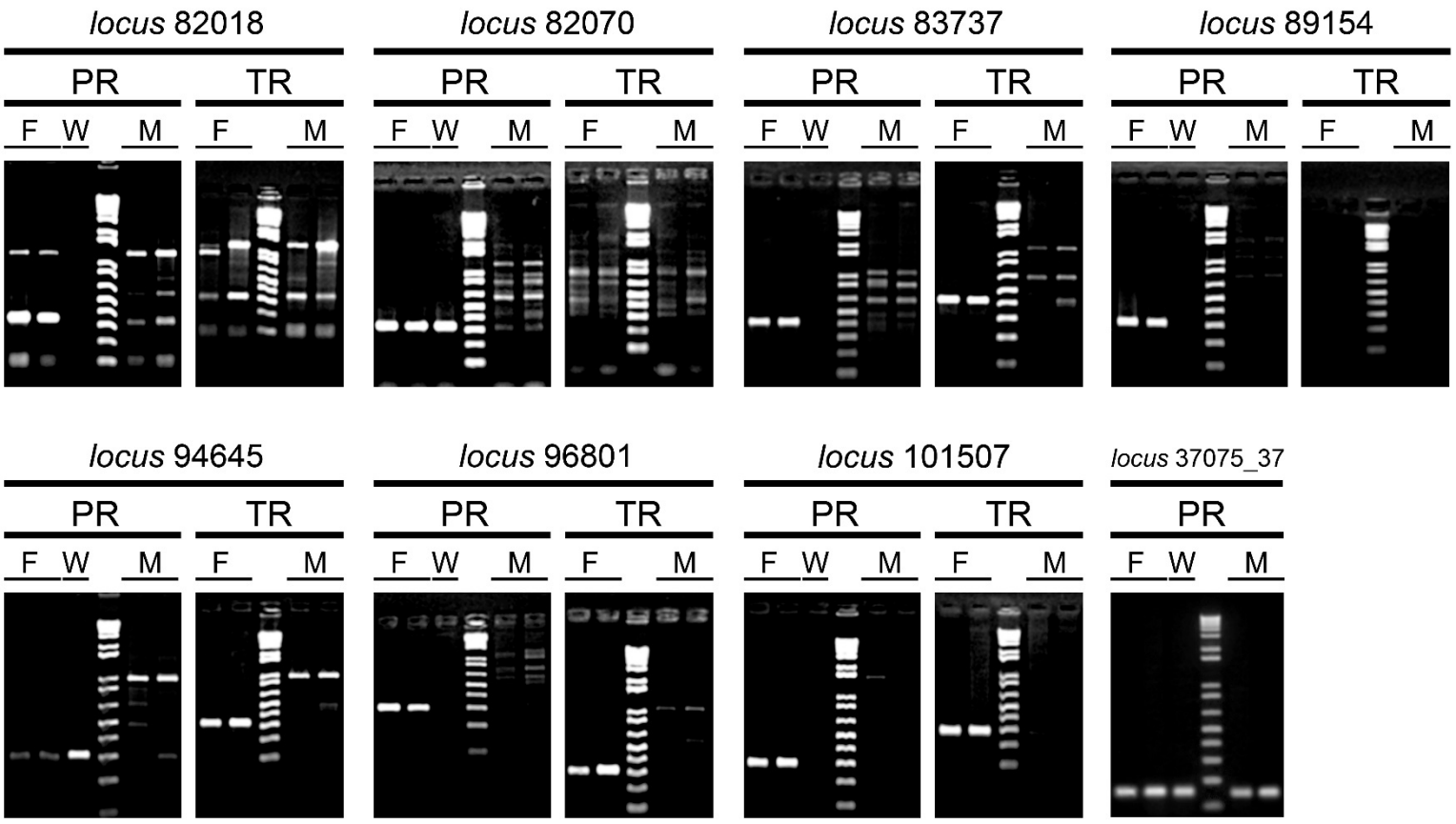


Fig 4.


locus 82018
PR
TR
F W M F M
locus 82070
PR
TR
F W M F M
locus 83737
PR
TR
F W M F M
locus 89154
PR
TR
F W M F M
locus 94645
PR
TR
F W M F M
locus 96801
PR
TR
F W M F M
locus 101507
PR
TR
F W M F M
locus 37075_37
PR
F W M