Evolution and conservation of Characidium sex chromosomes

R Utsunomiya1,8, PC Scacchetti1,8, M Hermida2, R Fernández-Cebrián3, X Taboada4, C Fernández2, M Bekaert5, NJ Mendes1, D Robledo4,6, JE Mank7, JB Taggart3, C Oliveira4, F Foresti1 and P Martinez2

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 9863 single-nucleotide polymorphisms (SNPs), corresponding to ~ 1 marker/100 kb distributed across the genome for sex-linked variation. With this data set, 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. In addition, 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.

Heredit (2017) 119, 237–244; doi:10.1038/hdy.2017.43; published online 26 July 2017

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) heterogamy 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 SD 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 intraspecific cytogenetic variation in size and heteromorphism (Scacchetti et al., 2015).

In order to characterise 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 characterisation 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., 2016) 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.

MATERIALS 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 minimise 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 adaptors and library amplification PCR primer sequences, followed Houston et al. (2012). In brief, 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 adaptors, each with a unique 5 or 7 base barcode (Supplementary Table S1) were ligated to the SbfI digested DNA, at 22 °C for 90 min, by adding 0.5 μl 100 nm P1 adaptor, 0.12 μl 100 mn rATP (Promega), 0.2 μl 10 × Reaction Buffer 2 (NEB), 0.1 μl T4 ligase (NEB, 2 × U ml⁻¹) 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
sequence and estimation of the FIS and FST were selected for their validation in the PR population (Figure 1), as well as C. pterostictum (10 males and 5 females in each species, and 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 (Supplementary Table S3) and samples genotyped on the matrix assisted laser desorption ionization-time of flight mass spectrometry platform (Sequonem, 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 (WG44, 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 labeled 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). In brief, slides were incubated with RNase (50 μg ml⁻¹) 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 labelled probe, 50% formamide, 2×SSC and 10% dextran sulphate) was denatured for 10 min 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 (NCBI BioProject PRJNA391395; Accession numbers: SRR5738929-SRR5738965). 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 (Supplementary Table S1).

We generated a catalogue 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 (Supplementary Table S1). After applying the final filtering steps, we retained 9863 putative biallelic SNPs with a minimum depth of 10 reads and genotyped in at least 28 individuals (Dryad doi: 10.5061/dryad.tr3d8), with similar average number of reads for both sexes (females: 60.8, s.d. = 29.6; males: 53.1, s.d. = 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 (Supplementary Table S4a). Interestingly, the average number of reads per individual for the female-specific RAD loci (31.8, s.d. = 18.1) was very similar to that observed for the whole RAD locus data set (38.8, s.d. = 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 (hsd17B3), a steroidogenic factor related to gonad differentiation (Mindnich et al., 2005; Supplementary 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, s.d. = 34.4). In addition, 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. In addition, the hsd17B3 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 9863 SNPs for evidence of significant allelic differentiation (FST) between males and females (P < 0.05; Supplementary 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 genome-wide association; P < 0.05; Supplementary Table S6a). FST values of this set ranged between 0.090 and 0.299 with a mean of 0.144 (s.d. = 0.046) (Supplementary Table S6b). Ten loci showed an FST > 0.250 and SNP with the highest FST = 0.299 suggests notable divergence between the Z and W, given the maximum expected FST = 0.500. The FST 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, genome-wide association or exact G tests.

We searched public databases to identify candidate genes related to sex differentiation for the 148 sex-linked SNPs (Supplementary Table S6c). A total of 116 sequences were annotated (78.4%), 16 of which (13.8%) were transposable elements 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 codes for a junction molecule crucial for spermatogenesis (Zhang and Lui, 2014), and tgfβ2, a gene that plays a pivotal role in gonad development (Erkin et al., 2008). Nectin-2 and tgfβ2 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; Supplementary 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 (Supplementary 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 (span1), 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). In addition, 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, that is, 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 oestrogen 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 (9863). 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 data set compared to the genomic average ($P<0.05$; 19.6% vs 4.2%, respectively; Supplementary 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 (Supplementary 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 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 9863 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 (Supplementary 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 (Supplementary Table S3). Thirteen primer sets of the 15 selected RAD loci were validated by PCR on the expanded sample of 35 females and 25 males of C. gomesi (Carlson et al., 2015), with LG8, LG10, LG15, LG16 and LG22 each containing two hits. The proportion of 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 9863 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.

### 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 micro-dissected W chromosome library (Figure 4), confirming their location on this chromosome. The seven PCR amplified female-specific markers were used as probes for FISH on C. gomesi metaphase plates to ascertain their location, but no specific signals of hybridisation were detected. In keeping with their known transposable elements sequence composition, three of these generated weak scattered signals throughout the karyotype of the species, suggesting multiple locations (Supplementary Figure S1).

**DISCUSSION**

Teleosts show a high turnover of sex-determining mechanisms (Martinez 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 9863 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 roughly 1 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 SD 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 LD 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 to 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 to 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 FST 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 LG.

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 tgfβ2, an important mediator of growth and differentiation involved in germ cell and gonadal development (Ergin et al., 2008; Ozgüden-Akkoç and Ozer, 2012). Tgfβ2 belongs to the Tgf-β superfamily, and the 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 Martinez et al., 2014).

We used a comparative mapping strategy interrogating the reference genomes of other fish species to gain insight on the genomic organisation 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 Characidiae, 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 De Almeida-Toledo, 2007), but there have been important genomic reorganisations 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 FST, suggesting that they may contain important regions orthologous to the C. gomesi sex chromosomes. However, our data indicate a complex origin, involving reorganisations 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 (FWS = 0), as observed in our data (19.6% deviations for sex-associated loci vs 4.2% across the whole data set). In regions where the W chromosome is highly diverged, this situation would be extreme (FWS = 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 Gasterosteus aculeatus (Schulteß 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 characterisation in wild species. Both female-specific markers, expected in a ZW/ZZ 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 Characidiae, 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.

DATA ACCESSIBILITY
Raw data: NCBI BioProject PRJNA391395. Raw sequences: SRR5738929-SRR5738965. Filtered RADtags Dryad doi:10.5061/dryad.tr3d8.

DATA ARCHIVING
There were no data to deposit.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
We 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) and 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.
Abdou HS, Vilain E, 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.

Achtschel 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, Manke J, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL et al. (2014). Sex determination: why so many ways of doing it? *PloS Biol* 12: e1001899.

Bailly NA, Elmay-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, Brandstöm M, Oskarsson N, Axelsson E, Smith NG, Ellegren H (2006). Molecular cytogenetics of the sex chromosomes from the European eel *Anguilla anguilla*. *Genome Res* 16: 1410–1418.

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

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

Bull JJ (1983). Evolution of Sex Determining Mechanisms. Benjamin Cummings Press: Menlo Park, CA, USA.

Carlson BM, Oursou SW, Gross JB (2015). A high-density linkage map for Asyntan 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 and karyotype of the male and female reproduction biology in the Southern stingray (*Dasyatis americana*). *Sci Mar* 62: 257–262.

Chaffin KJ, Filatov DA (2005). Substitution rates in a new sex chromosome system. *Mol Ecol* 14: 2114–2116.

Chang YW, Yang X, Li JY, Yang WX (2012). Tgf-f and Tgf-j superfamily: how does it regulate testis development. *Mol Biol Rep* 39: 4727–4741.

Chen J (2003). Substitution sites in a new sex chromosome system. *Mol Biol Evol* 20: 1203–1212.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.

Chen J, Wang H, Zhang X, Zhang S, Chen S, Yuan Y (2010). Genetic diversity and origins of ZW sex chromosomes in the Chinese alligator (*Alligator sinensis*). *Genet Mol Biol* 33: 875–886.