Genome biology of the Darkedged Splitfin, *Girardinichthys multiradiatus* , and the evolution of sex chromosomes and placentation

Kang Du, Martin Pippel, Susanne Kneitz, et al.

Genome Res. published online January 26, 2022
Access the most recent version at doi:10.1101/gr.275826.121

Genome biology of the Darkedged Splitfin, *Girardinichthys multiradiatus*, and the evolution of sex chromosomes and placentation

Kang Du[1], Martin Pippel[2], Susanne Kneitz[3], Romain Feron[4], Irene da Cruz[5], Sylke Winkler[2], Brigitta Wilde[3], Edgar G. Avila Luna[6], Gene Myers[2*], Yann Guiguen[7*], Constantino Macias Garcia[6*], Manfred Schartl[1,5,#]

[1]The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas, TX 78666, USA

[2]Max-Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 10307 Dresden, Germany

[3]Biochemistry and Cell Biology, Biocenter, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany

[4]Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[5]Developmental Biochemistry, Biocenter, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany

[6]Instituto de Ecologia, Universidad Nacional Autónoma de México, Ciudad Universitaria, Circuito exterior s/n anexo al Jardín Botánico C. P. 04510, Mexico City D. F, Mexico

[7]INRAE, LPGP, Rennes, France

[*]Equally contributing senior authors

[#]Corresponding author

Abstract

Viviparity evolved independently about 150 times in vertebrates and over 20 times in fish. Several lineages added to the protection of the embryo inside the body of the mother the provisioning of nutrients and physiological exchange. This often led to the evolution of a placenta. Amongst fish, one of the most complex systems serving the function of the placenta is the embryonal trophotaenia/ovarian luminal epithelium of the Goodeid fishes. For a better understanding of this feature and others of this group of fishes, high quality genomic resources are essential. We have sequenced the genome of the darkedged splitfin, *Girardinichthys multiradiatus*. The assembly is chromosome-level and includes the X and Y Chromosomes. A large male-specific region on the Y was identified covering 80% of Chromosome 20, allowing some first inferences on the recent origin and a candidate male sex determining gene. Genome-wide transcriptomics uncovered sex specific differences in brain gene expression with an enrichment for neurosteroidogenesis and testis genes in males. The expression signatures of the splitfin embryonal and maternal placenta showed overlap with homologous tissues including human placenta, the ovarian follicle epithelium of matrotrophic Poeciliid fish species and the brood pouch epithelium of the seahorse. Our comparative analyses on the evolution of embryonal and maternal placenta indicate that the evolutionary novelty of maternal provisioning development repeatedly made use of genes which already had the same function in other tissues. In this way already pre-existing modules are assembled and repurposed to provide the molecular changes for this novel trait.

Keywords: Goodeids, placenta, trophotaenia, sex chromosome, transcriptome, brain, whole genome sequencing, chromosome assembly.

Introduction

Viviparity, a mode of reproduction that can exacerbate the conflict between sexes over the allocation of resources to offspring, has evolved independently amongst all classes of jawed vertebrates (except birds) about 150 times, 21 of which occurred in fish (for review see (Blackburn 2015)). Viviparous species can be arranged along a continuum from lecithotrophic to matrotrophic based on whether the nutrients required for embryonic development are present in the egg (lecithotrophic development), or whether they are

gradually provided by the mother through gestation (matrotrophic development), for instance via a placenta (placentotrophy) (Blackburn 1999; Ghalambor et al. 2004; Pollux et al. 2009). In spite of the large number of independent origins, our understanding of the evolutionary and functional development of matrotrophic viviparity, especially of that involving a placenta, comes mostly from the study of mammalian matrotrophy, and while vast, our knowledge of its genomic underpinning is limited to only one transition from oviparity to matrotrophic viviparity and a few recent studies on fish (see below) (Lynch et al. 2008; Kin et al. 2016). There is hence a need to incorporate more taxa to the study of the evolution of viviparity, ideally taxa which include both oviparous and viviparous (matrotrophic) species. Fish in the Order Cyprinodontiformes seem appropriate to conduct such comparisons, as different forms of viviparity have evolved independently in several families: Poeciliidae, Jenynsidae, Anablepidae and Goodeidae (Meyer and Lydeard 1993) (Blackburn 2015). Accordingly, comparative analyses revealed that placentation drives a shift from a reliance on pre-zygotic mate choice to increasing levels of polyandry, in conjunction with post-zygotic mechanisms of mate selection (Pollux et al. 2014; Furness and Capellini 2019). More recently, comparative transcriptomic analyses have found parallels between poecilid and mammalian placental viviparity (Guernsey et al. 2020). By focusing on the speciose Poeciliidae, this study explored the consequences of several transitions between lecithotrophy and matrotrophy, although still only covering one transition between oviparity and viviparity. It is likely due to the lack of genomic resources that the family Goodeidae has not been used to investigate the genomic background of the transition from oviparity to placental viviparity. The Goodeidae is a small, yet diverse family of Cyprinodontiformes native to North America (Parenti 1981). The subfamily Empetrichthynae, from Nevada in the USA, includes two genera of oviparous fish, and the subfamily Goodeinae from Central Mexico includes ~40 viviparous species distributed in 16 genera (Webb et al. 2004). This high ratio of genera to species is indicative of a burst of diversification (Mayr 1942), which has been identified as an adaptive radiation following the evolution of viviparity (Ritchie et al. 2005). There are at present no genomic resources that may allow the use of the Goodeidae to explore the genomic transitions that accompanied the evolution of matrotrophic viviparity from oviparity.

Matrotrophy in goodeids is linked to the possession of a dynamic placenta sensu Mossmann (Mossman 1991) made up of a maternal part, the internal ovarian epithelium, and an embryonic component, the trophotaenia (Lombardi and Wourms 1985; Uribe MC 2005; Schindler 2015) (Fig. 1). The latter are extensions of the embryonic hindgut mostly shaped in the form of ribbons or rosettes (Turner 1937; Lombardi and Wourms 1985) made of an absorptive epithelium around a well vascularized core (Mendoza 1972; Lombardi and Wourms 1985). The maternal part of the placenta is represented by the follicular epithelium of the ovary, which is closely associated with the trophotaenia. This is a dynamic system because early embryos rely on an initial supply of yolk, then develop the trophotaenia that reaches out to the modified ovarian lumen epithelium, but starts to regress, possibly through apoptosis, as gestation draws to an end (Iida et al. 2015). Maternally provided proteins, which can include vitellogenin (Vega-López et al. 2007) and also residues of dead siblings (Greven and Grossherr 1992), are uptaken by the trophotaenial epithelium via receptor-mediated endocytosis, and subjected to lysosomal degradation (Schindler and De Vries 1987; Schindler and Kujat 1990; Schindler 2003).

Unlike other viviparous fish such as poeciliids (Constantz 1984), splitfin females do not store sperm (BISAZZA 1997), and they are only receptive for about one week after giving birth, which takes place every two months (Macías-Garcia and Saborío 2004). Additionally, males lack an intromittent organ, thus sperm is deposited in the outer part of the gonopore (BISAZZA 1997; Garcia and Valero 2010; Greven and Brenner 2010). These particularities of splitfin reproductive biology may promote that males adjust their sperm production and mating effort to the probability of finding mates and of facing sperm competition (Simmons and Fitzpatrick 2012), enforcing novel conditions for sperm to reach the ova. Thus in addition to the consequences for females, it is likely that viviparity has also left a genomic footprint on males.

Splitfin males are heavily ornamented (Ritchie et al. 2005; Méndez-Janovitz et al. 2019) and, as they cannot force copulations (see previous paragraph) they perform complex courtship displays to secure mating (Macías-Garcia and Saborío 2004). This preponderance of sexual selection may be related to high rates of differentiation, as previously proposed (Darwin 1896; Panhuis et al. 2001; Ritchie 2007).

Mexican goodeids face a variety of environmental challenges (De La Vega-Salazar et al. 2003; DOMÍNGUEZ-DOMÍNGUEZ et al. 2006), including chemical pollution (Marina et al. 1997), which has made splitfins a model for ecotoxicological research (Arellano-Aguilar and Macías Garcia 2008; Arellano-Aguilar and Macias Garcia 2008; Rueda-Jasso et al. 2014; Guerrero-Estévez and López-López 2016).

Here we report the first high-quality reference genome of a Goodeid, the darkedged splitfin (also known as Amarillo), *Girardinichthys multiradiatus*, assembled at full chromosome level including the X and Y Chromosomes. Using this genomic resource, we uncovered a pronounced sex difference in brain transcriptomes and established the expression signature of the trophotaenial placenta showing considerable overlap with human placenta and RNA-seq data from another livebearing fish and seahorse.

Results

Genome sequencing, chromosome assembly and annotation

To obtain a genome assembly of high contiguity and completeness, we used the pipelines of the international Vertebrate Genome Project that incorporate state-of-the-art sequencing technologies and assembly algorithms (Rhie et al. 2021). Pacific Biosciences (PacBio) continuous long reads (71-fold coverage), 10x Genomics Illumina read clouds (52 fold coverage), Bionano optical maps (1870 fold coverage) and chromosome conformation capture (Hi-C) Illumina read pairs (94 fold coverage) from three male individuals were generated (Supplemental Tables S1, S2). The PacBio reads were assembled into contigs using the customized assembler DAmar (https://github.com/MartinPippel/DAmar). Next, we used the 10x Illumina read-cloud data to correct base errors and phase haplotypes, arbitrarily picking one haplotype in a phased block. Finally, we used Bionano optical maps and then Hi-C data to produce long-range scaffolds. This resulted in a final assembly of 1.1 Gb with a contig N50 of 17 Mb, arranged in 71 scaffolds. Of these, 24 are chromosome-size (Fig. 2) in agreement with the karyotype of *G. multiradiatus* (Uyeno et al. 1983), making up 99.7% of the genome assembly (Supplemental Table S3).

A comparison with other teleost chromosome assemblies disclosed a high degree of conserved synteny not only to other Cyprinodontiformes (*Fundulus*, *Xiphophorus* and medaka), but also with the more distant Perciformes and to a lower extent with mostly small translocations and rearrangements to herring, cavefish and zebrafish (Fig. 3).To annotate protein coding genes, gene evidence from protein homology of other species, RNA-seq transcriptomes from darkedged splitfin and ab-initio predictions were integrated. A total of 23770 protein coding genes were annotated. The BUSCO completeness based on the actinopterygii_odb9 dataset was improved from 96.4% to 99.3% by the annotation process (Supplemental Table S4). Out of the 23770 genes, 23463 (98.7%) have a BLAST hit to the Swiss-Prot/RefSeq database. 2113 (8.9%) of the protein-coding genes are single exon genes. Additionally, 1900 tRNA, 21 rRNA, 279 miRNA and 355 other non-coding RNA genes were annotated (Supplemental Table S5)

Genome evolution

The genome of *G. mulitradiatus* consists of 37.3% of repetitive sequences according to RepeatMasker (Supplemental Table S6). This is in the range of many other teleost fish genomes (Shao et al. 2018) including in the closely related *Fundulus* and Medaka species but considerably higher than the livebearing platyfish. Of note, there is a strong expansion of ERV class I elements, which is not found in other genomes of the order Cyprinodontiformes (2.5% vs 0.01- 0.4%). The transposon landscape history (Supplemental Fig. S1) indicates that darkedged splitfin experienced a strong wave of TE expansion rather recently, in contrast to other teleosts genomes analyzed so far.

A phylogenomics reconstruction of the evolutionary relationships of splitfin to other teleosts using a gene set of 1425 high quality orthologs confirmed previous groupings. The tree revealed a split of the Goodeid branch from egg-laying cyprinodonts around 52 MYA, about twice the age estimated from partial mitochondrial sequences (Webb et al. 2004). The last common ancestor of Goodeids with the livebearing Poeciliids dated back almost to the same period, around 64 MYA (Fig. 3, Supplemental Fig. S2).

Gene family analysis revealed that 42 gene families expanded and 128 contracted in the splitfin genome. However, none of these families shows a dynamic exclusive to the lineage of *G. multiradiatus* (Supplemental Fig. S2).

Sex-biased gene expression in brain

The availability of an annotated genome allowed us to look for sexual dimorphism of the brain on a molecular level. Differential gene expression analyses of female and male brain transcriptomes revealed that 324 genes, with a $\log_2$ fold change (FC) >1 and adjusted p-values <0.05, were more expressed in males, while only 42 genes showed a biased expression towards females (Supplemental Table S7). In the female brain *foxl2a*, which is a main transcription factor in female gonad development, was higher expressed than in males. Enrichment for Gene Ontology terms revealed functions related to membrane structure and function including transporter activities (Supplemental Fig. S3A-D). The male biased genes were enriched for several categories related to germ cell development, male meiosis and spermatogenesis (Supplemental Fig. S3E-G). Amongst these are *dmrt1*, an important regulator of male sexual development in fish, and genes that function in the early stage of meiosis, e.g. *spo11*, *sycp1*, *sycp2*, *sycp3*, *hormad1*, or sperm development, e.g. *spata4*, *spata7*, *spata17*, *spata22*, *spef1*, *spag6*, *spag8*. Of note, also genes from the piRNA pathway (*exd1*, *piwil1*, *tdrd5*, *tdrd6*, *tdrd9*), which is particularly active in testes, are predominantly expressed in male brain. Several genes of steroidogenesis pathways were also upregulated in male (*dhrs11b*, *akr1a1b*, *cyp11c1*).

Placental development

The placenta is an organ formed by the sustained apposition or fusion of fetal membranes and parental tissue for physiological exchange (Mossman 1991). Goodeids have been described as having the most complex placental organ amongst livebearing fishes to support the developing embryos, given its large array of histological and cytological modifications of the intra-ovarian epithelium and the throphotaenial absorptive epithelium, which are apposed to each other during pregnancy (Uribe et al. 2010; Schindler 2015). To contribute to the understanding of the evolution, function and development of this outstanding trait, we analyzed transcriptomes of the trophotaenia

(embryonal part of placenta) and the ovary after dissecting out the embryos (maternal part of the placenta). Both datasets were then compared to all other transcriptomes (male and female brains, testis, organ mix, early and late embryos and trophotaeniae or ovary, respectively). Trophotaenia and ovary predominant expression signatures were defined by a $\log_2$ fold change >2 compared to all other transcriptomes in accordance with the definition of "genes preferentially expressed in human placenta" (https://www.proteinatlas.org/humanproteome/tissue/placenta). This revealed 1653 trophotaenia and 1132 ovary enhanced genes (Supplemental Tables S8, S9).

The ovarian transcriptome is enriched for genes with functions for formation of the zona pellucida and extracellular matrix, fertilization, immune response and steroidogenesis (Supplemental Fig. S4). The most upregulated gene in the ovary is *cyp19a1a* ($\log_2$FC=7.46), the ovarian isoform of the estrogen producing enzyme aromatase.

The genes that constitute the expression signature of the trophotaenia are highly enriched for genes with a function in brush border membranes, transport of ions and metabolites, and metabolic pathways (Supplemental Fig. S4) in agreement with the role of the embryonal part of a placenta. Of note, the telomerase reverse transcriptase is highly upregulated ($\log_2$FC=5.82 compared to all other organs), as well as the gene encoding the prolactin receptor ($\log_2$FC=4.41).

Insulin-like growth factor 2 (IGF2) signaling has an important role in regulating placental supply of nutrients in mammals and is a key gene undergoing paternal imprinting (DeChiara et al. 1990; DeChiara et al. 1991). In the darkedged splitfin the *insulin-like growth factor-binding protein1-like* and *insulin-like growth factor-binding protein4-like* genes are predominantly expressed in the trophotaenia and *insulin-like growth factor-binding protein 1* and *3*, *insulin-like growth factor 1*, and *IGF-like family receptor 1* are predominantly expressed in the ovary. However, *igf2* is not expressed in the ovary and also not part of the trophotaenia transcriptome.

We then checked the trophotaenia and ovary transcriptomes for expression of other genes imprinted in mammals (N=255, https://www.geneimprint.com/site/genes-by-status) and

found that 16 of these are expressed in trophotaenia and 10 in ovary (Supplemental Table S10)

In another family of livebearing cyprindontiform fishes, the Poeciliidae, several species have evolved maternal provisioning of nutrients to the developing embryo throughout pregnancy. Here, the follicular epithelium like in splitfins is the maternal part of the placenta supplying support. We compared the transcriptomes of follicles of the matrotrophic species *Poeciliopsis retropinna* (Guernsey et al. 2020) with our dataset. 1640 genes were found to be overexpressed in follicles of *P. retropinna* compared to follicles of the lecithotrophic species *P. turrubarensis*, and among those 59 overlapped with the ovary-specific expression pattern of *G. multiradiatus* (Supplemental Fig. S5; Supplemental Table S11). The shared genes are enriched for extracellular matrix and cytokine signalling related GO terms (Supplemental Fig. S6A).

The lineage of seahorses has also developed livebearing, convergently. In addition, seahorses have evolved provisioning of nutrients to the embryo by a specialization of the ventral skin integument, the male brood pouch epithelium, representing the parental part of the placenta (Griffith and Wagner 2017). A transcriptomic study identified 147 significantly upregulated genes in the pregnant (compared with the nonpregnant) and post-parturition seahorse brood pouch (compared with the late-pregnant). (Whittington et al. 2015). Nine of these are also upregulated in the ovary of splitfin (Supplemental Fig. S7) with functions in calcium and lipid transport and tissue remodeling.

The human placenta is formed from the endometrium as the maternal part and the trophoblast, which develops into the embryonal contribution of the placenta. The available human placenta transcriptome combines the embryonal and maternal part. Thus, we compared it to both parts of the splitfin placenta. The trophotaenia and ovary transcriptomes of the splitfin overlap with 89 of the 493 genes of the human placenta transcriptome(https://www.proteinatlas.org/) (Supplemental Fig. S8A; Supplemental Table S12). These include the vascular endothelial growth factor family member placental growth factor (PGF), the prolactin receptor and many components of the extracellular matrix and genes enriched for GO terms angiogenesis, growth and morphogenesis (Supplemental Fig. S6B,C). A comparison of trophotaenia with the

human trophoblast, which gives rise to the embryonal part of the placenta, revealed 97 genes in common between both tissues (Supplemental Table S12, Supplemental Fig. S8B) including many transporters and transcription factors enriched again for morphogenesis, extracellular matrix and angiogenesis GO terms (Supplemental Fig. S6D).

The evolution of a highly sophisticated placental organ mediating matrotrophy in Goodeids may have required changes in protein structure and function that should be visible as signatures of positive selection. To uncover such genomic traces of natural selection we collected 7102 one-to-one orthologs from *G. multiradiatus* and 11 other teleost species (Supplemental Fig. S2) including the livebearing species from the Poecilidae family and the more distantly related seahorse and tested these genes for positive selection in *G. multiradiatus*. With site class 2a (positive selection in marked branch and conserved in rest) we identified 48 genes and with site class 2b (positive selection in marked branch and relaxed in rest) 122 genes under positive selection specifically in the splitfin lineage (Supplemental Table S13). These genes include 16 from the trophotaenia and 6 from the ovary-specific transcriptomes (Supplemental Table S14).

The positive selection analysis revealed 440 genes (site class 2a, 128 genes, 2b, 312 genes) in seahorse (Supplemental Table S15). In *Poeciliopsis* because of the unavailability of an annotated genome the orthologous genes could only be identified by a BLAST to the genome assembly. This restricted the number of one-to-one orthologs for comparison with splitfin to about one quarter (n=1521) of which 65 genes (site class 2a, 35 genes, 2b, 30 genes) are under positive selection (Supplemental Table S16). Between all three species, there are two genes commonly under positive selection in *Poeciliopsis* and splitfin, 8 genes shared between seahorse and splitfin and one gene, *sterol-O-acyltransferase 1*, which is under positive selection in all three species (Supplemental Fig. S9).

Sex determination and sex chromosomes

To characterize the sex determination system of the darkedged splitfin we first searched for sex-linked markers using Restriction site Associated DNA Sequencing (RAD-seq) (Feron et al. 2021) in 29 females and 27 males. A total of 6,359,801 markers was found, of which 892 were significantly associated with male phenotype ($p < 0.05$, chi-square test with Bonferroni correction) and none were associated with female phenotype (Fig. 4A; Supplemental Table S17). These results suggested a stable male heterogametic genetic sex determination system. The absence of outliers most likely indicates the absence of relevant environmental factors or autosomal modifiers influencing the XY sex chromosomal system.

We then aligned these RAD-seq markers to the genome assembly of darkedged splitfin. Among the 892 markers significantly associated with male phenotype, 675 (76%) were aligned to a unique position with mapping quality higher than 20, and of these good quality markers 654 (97%) were aligned to scaffold 20 in a region ranging from 22 kb to 34.5 Mb (Fig. 4B-C). These results indicate that scaffold 20 is the sex chromosome in darkedged splitfin and that a large region spanning 80% of the sex chromosome is differentiated between the X and Y Chromosomes.

To identify the male specific region and to elucidate the evolution of the sex chromosomes of splitfin, a haplotype-resolved assembly of scaffold 20 was done. Mapping of the male-specific RAD-tags identified the putative Y Chromosome (Fig. 5A), where 645 (71%) had a significant hit.

Sequence difference between X and Y was determined in 10 kb sliding windows as the percentage of SNP and indels. It uncovered a large region from 0 - 34.5 Mb on the Y, where the nucleotide difference was uniformly in the range of 1.4% (Fig. 5B). Beyond this point the overall sequence difference between X and Y was close to zero. The d_S values were high in the XY-differential region (Median 0.01) and low in the terminal segment (Supplemental Fig. S10A). The d_N/d_S values in the terminal region are predominantly around 0.2 indicating purifying selection. Most genes in the XY-differential region have on average twice as high values and many close to 1 or even higher, indicating relaxation from purifying selection (Supplemental Fig. S10B). Together this identifies a possible terminal pseudoautosomal region of 7 Mb with about

200 genes and a large region on the Y of about 34.5 Mb containing around 900 genes, where it is different from the X (Supplemental Fig. S10C).

To estimate the age of the Y we used the d_S distance between the X and Y in the differential region and calibrated it to the d_S based divergence time estimate of splitfin to *Fundulus* (52.6 Mya, see Supplemental Fig. S2). This revealed that the Y Chromosome of splitfin originated only about 2 Mya (Supplemental Fig. S11).

The region on the Y which is different from the X does not have a higher content of repeats and TE's. Also, the transposon landscape is similar to the corresponding region on the X and all autosomes except for a slightly higher percentage of LINE elements (Supplemental Fig. S12). The region on the Y has a higher content of pseudogenes (20.4%) than the corresponding region on the X (13.7%) and all autosomes (16.1%, Supplemental Methods: Comparison of the X and Y Chromosomes). Mapping the gene contents of the X and the Y to each other revealed that no gene from the X is missing from the Y, while there are twelve gene models annotated that are present only on the male sex chromosome. Manual inspection revealed all of these 12 genes are either corrupted, pseudogenes or transposon derived sequences (Supplemental Table S18).

New sex determining genes arise either from local gene duplications or allelic diversification (Herpin and Schartl 2008). As we did not find any hint of duplicated Y-specific genes, we considered that the splitfin male determining gene would show allelic differentiation between the X and Y copies. We manually checked the 948 genes of the differentiated region on the Y for the recurrently appearing sex determining genes in vertebrates. From these (TGFß signalling, *dmrt1* and *sox* family members) only *sox8a*, *sox9a* and *sox10a* are on scaffold 20. While *sox9a* and *sox10a* have d_N/d_S values of 0.32 and 0.23, which indicate a purifying selection, *sox8a* has a value of 0.81. The X and Y *sox8a* copies differ by 12 (2.6%) amino acids. Overlaying the computer-modeled protein structures revealed that none of these differences would compromise the three-dimensional structure of Sox8a as a whole, suggesting that both copies encode functional transcription factors (Supplemental Fig. S13). The Y copy presents two amino acid changes at a highly conserved position (L364M, S360N). The S360N exchange adds a ß-sheet at this place to the secondary structure. Further functional studies are required to

elucidate what could be the impact of these changes on Sox8a activity in splitfin. Of note, the Y allele is preferentially expressed in testis, however, in male brain, the X and Y alleles are equally expressed (Supplemental Table S19).

Discussion

As part of the Vertebrate Genomes Project (VGP), a project of the G10K Consortium, which aims to generate near error-free reference genome assemblies of ~70,000 extant vertebrate species, the darkedged splitfin was selected to represent the group of Goodeid fishes. The availability of a chromosome-level genome assembly of high completeness allowed us to perform genome-wide analyses to contribute to a better understanding of several highly interesting characteristics of Goodeids and also benefit ecotoxicological research.

We identified a differentiated XY sex chromosome system in *G. multiradiatus* and succeeded in assembling X and Y separately. The comparison of these two sex chromosomes revealed a large MSY with uniform levels of sequence differentiation over its whole length. Such distribution of male-specific SNPs would be expected if recombination stopped simultaneously in the Y-specific part by a large chromosomal inversion. Manual inspection of the Hi-C maps did not support such a single-large inversion, however, several smaller inversions are possible. Further studies are needed to explain this pattern.

The differentiation of the Y occurred in less than two million years since its origin, about six MY later than the split between *G. multiradiatus* and its sister species *G. viviparus* (Webb et al 2004). The two species are different in the expression of epigamic characters, with *G. multiradiatus* males being more ornamented (larger and colourful median fins) and having a large repertoire of courtship displays, whereas *G. viviparus* has smaller dark fins and very simple courtship behaviour (González Zuarth and Macías Garcia 2006; Méndez-Janovitz and Garcia 2017). A full assembly of the genome of *G. viviparous* is needed to better understand sex chromosome evolution in the genus *Girardinichthys*. Comparison of the sex chromosome-specific region to the corresponding region of the *G. multiradiatus* genome can then be used to search for genes involved in male ornaments.

Our brain transcriptome analyses uncovered a large number of genes which were more expressed in males and which are known to fulfill their functions in testis development and spermatogenesis. This is in agreement with RNA-seq data from human, mouse and zebrafish (Guo et al. 2003; Guo et al. 2005; Sreenivasan et al. 2008). A closer relationship of female and male brains to their respective gonad was also reported (Sreenivasan et al. 2008), however, at least to our knowledge such a strong bias as in splitfin of male brain towards testis has not been noted. The biological meaning of the abundance of testis gene expressions in the male brain is unknown as well as its molecular mechanisms. Shared characteristics between neurons and sperm, including the importance of the exocytotic process and the presence of similar receptors and signalling pathways have been hypothesized (Matos et al. 2021). In humans it has been shown that meiosis genes expressed in the brain are not translated. A possible reason for the absence of translation of such transcripts to proteins may be connected to the lack of Dazl, which is a key regulator of translation of spermatogenesis genes (Li et al. 2019). In splitfin brain, *dazl* is expressed only at low levels in male, not at all in females, and abundantly only in testis. In splitfin, we lack the tools to analyze protein expression, but it may well be that a large fraction of the testis genes is not translated due to the lack of Dazl (Supplemental Fig. S14). Even if not, all expressed “testis” genes may have a function in brain they might reflect a co-regulatory expression mechanism, which upregulates genes that are of importance in both organs, e.g. steroidogenesis genes. The reproductive biology of splitfins requires a complex male-specific behavior including searching for receptive females, a whole suite of courtship behaviors (Garcia and Valero 2010) and constantly adjusting sperm production to suit the current availability of partners and of potential sperm competitors (Garcia et al. 1998; Simmons and Fitzpatrick 2012) which is orchestrated in the brain and may be regulated by neurosteroids. We found upregulation of genes involved in steroidogenesis specifically in the brain of males. Further studies have to show if these enzymes are indeed involved in the synthesis of neurosteroids.

So far, from the other placental fish, seahorse and *Poeciliopsis*, only the parental part of the placenta was analysed. Here in splitfin, for the first time information also from the embryonal part could be generated. The placental organ of *G. multiradiatus* shows a high similarity with the expression signature of human placenta. This is indicative of a

convergent evolution of organs that serve embryonic nutrition although they are phylogenetically quite distant and developmentally totally unrelated. The mammalian placenta decends from the maternal endometrium and the trophoblast, which have no counterpart in fish. The goodeid trophotaeniae, however, are outgrowths of the hindgut that interact with the maternal ovarian wall. The similarities to the human placenta transcriptome, which is derived from the whole composite organ, are equally shared between the maternal and embryonal part of the *G. multiradiatus* placenta.

The trophotaeniae exhibit a high expression of the prolactin receptor indicating that this organ is responsive to prolactin stimulation. Prolactin, the lactation hormone of mammals, is a hormone that plays multiple roles in parental care in fish including: mouth breeding, nutrient provision by mucus or fluids, parental care behavior and viviparity (Whittington and Wilson 2013; Dobolyi et al. 2020). Expression of the prolactin receptor gene has also been observed in the follicular epithelium of matrotrophic *Poeciliopsis* species (Guernsey et al. 2020). The *prolactin* gene of *G. multiradiatus* is not only expressed in brain but also at similar levels in trophotaenia and ovary, indicating not only endocrine but also juxtacrine hormone signaling.

The trophotaenia is an embryonic organ and thus high telomerase expression is within expectation (Wright et al. 1996; Anchelin et al. 2011), in splitfin, its level of expression achieve 15 – 50 fold higher than in the embryo proper. In the human placenta dysregulation of telomerase activity is connected to a variety of pregnancy complications (Fragkiadaki et al. 2016).

The absence of *igf2* in the ovary and the trophotaenia of *G. multiradiatus* is intriguing (Lawton et al. 2005). Parent-of-origin specific expression of this growth factor gene constitutes an iconic example of imprinting linked to sexual conflict. In mammalian embryos, the paternal copy can be overexpressed, hence promoting embryo growth beyond the optimum for the mother, and the maternally inherited copy silenced through methylation, thus resisting the effects of the paternally-inherited allele. Conversely, the maternal copy of the cation-independent mannose-6 phosphate receptor, which binds the growth factor and transports it to lysosomes, is overexpressed, and the paternal allele is silenced (DeChiara et al. 1990; DeChiara et al. 1991). *Igf2* is expressed in fish embryos

(Yuan et al. 2011), including those of *G. multiradiatus*, where there is evidence suggesting that, as in mammals, the paternal copy is expressed, and the maternal allele is silenced through methylation (Saldivar Lemus et al. 2017). Thus, we would have predicted its absence in the ovary transcriptome but expected it to be expressed in the trophotaenia. At present we have no information of where in the embryo *igf2* is being expressed, or why it is not being expressed in the trophotaenia, although it is still possible that its expression in the trophotaenia depends on the developmental stage of the embryo and that we collected trophotaeniae from embryos of 3-4 and 5-6 weeks of development, which were, at that stages, not producing this growth factor.

Livebearing and maternal or paternal provisioning through placenta-like structures has evolved independently several times in fish. We asked the question if this physiological and morphological convergent evolution is connected to similar convergent changes on the molecular level. Proteins that may have adapted to new functional properties required for placentation should show signatures of positive selection in the respective lineage. Indeed, about one fifth of the genes identified as positively selected in splitfin are specifically expressed in the placenta. An even higher fraction was noted in the matrotrophic Poecillid species *Heterandria formosa* (van Kruistum et al. 2019). However, comparing positively selected genes of seahorse, Poeciliopsis and splitfin revealed only a single gene in common, *sterol-O-acyltransferase 1*. This gene encodes a lipid metabolism enzyme, which has a key role as the rate-limiting enzyme for regulation of the transport of circulating cholesterol and fatty acids, their metabolism and the turnover of lipoproteins. Perturbations of the placental lipid metabolism have been connected to pregnancy complications like preeclampsia (Hu et al. 2021). The low overlap between seahorse and the other two taxa could be due to the different sexes contributing to the placenta.

Changes in expression levels, e.g. of proteins involved in trafficking molecular cargo between compartments, would be another common adaptation connected to placentation. Indeed, about 20% of the genes preferentially expressed in human placenta are also found in the placental transcriptomes of splitfin. Nevertheless, all transcriptomes were enriched for similar pathways and gene categories, in particular those related to transport across

membranes, extracellular matrix, cell interactions and metabolism, but our data do not point to specific “placentation genes”, which are specifically recruited repeatedly. We hypothesize that the results may instead indicate that the independent development of parental provisioning repeatedly made use of genes which already had the same function in other tissues (e.g. in the transport of metabolites or interaction of between cells and organs) when it became necessary to make up a novel organ for embryo nutrition. In this way already pre-existing modules could have been assembled to provide the molecular changes for the novel trait. This is consistent with the idea that regulatory evolution is crucial to major evolutionary innovation (Emerson and Li ; Carroll 2008). We have generated genomic tools for *G. mulitradiatus* as a representative of the livebearing placental goodeids and could identify molecular changes that are connected to this form of maternal care. Further analysis will benefit from the inclusion of genomes and transcriptomes of lecithotropic splitfin species and the basal oviparous Goodeidae; the Empetrichthynae, from which the viviparous splitfins (Goodeinae) diverged 16.8 MYA (Webb et al 2004).

Methods

Experimental Animals

Most samples were from laboratory-born descendants of fish collected in 2011 at Salazar Lagoon (19°18′26″ N, 99°23′20″ W, 3000 masl) under license DGOPA/05332/071007.2437 from Dirección General de Gestión de Pesca y Acuicultura. We also used, for genome annotation, one sample from an aquarium stock originating from Zempoala (19°03′00″ N, 99°18′42″ W) and collected under permit DGOPA A/01584/110310.0785.

Live fish and preserved tissues were exported under permit B00.02.04.-1277.2019 from the Animal Health Direction (DGSA; SENASICA) of the National Ministry of Agriculture and Rural Development (SAGARPA).

Adult fish were sedated by placing them in ice-cold water, then quickly sacrificed by decapitation and immediately dissected. Embryos were classified as “early” (≤4 weeks of development) or “late” (five to six weeks of development; gestation lasts around 8 weeks

494 in this species). Embryos were separated from the single ovarian lumen. The
495 corresponding remaining ovary tissue and trophotaeniae were collected and processed
496 separately, hence we used whole ovaries from pregnant females, which were free of any
497 embryonic tissue. For RNA-seq tissues were promptly placed in RNA-later™ (2019). For
498 genome sequencing, dissected organs were flash frozen and stored at -80°C and for RAD-
499 tag sequencing tissues were stored in 100% ethanol.

500 For whole genome sequencing and assembly three individuals from the same family were
501 used (Supplemental Table S2).

502 As transcript evidence for genome annotation RNA-seq reads from a total RNA organ
503 mix of one female (brain, eyes, gills, muscle, skin, kidney, liver, embryos) and ovary
504 from a second female originating from Zempoala were used.

505 For expression analysis pools of organs and embryos were used (Supplemental Table
506 S20).

507 **DNA extractions and genome sequencing**

508 Ultralong genomic DNA of different organs of two splitfin individuals was extracted with
509 the agarose plug based Bionano PrepTM Animal tissue kit following the manufacturer's
510 instructions. Residual pigments were precipitated by adding Triton X-114, SDS and NaCl
511 to the genomic DNA (Ma et al. 2012). The fragment size of all genomic DNAs was
512 controlled by pulse-field gel electrophoresis before library construction. Two size-
513 selected PacBio continuous long read libraries of 24 and 29 kb in size were run on the
514 SEQUEL system with 10-hour movie times. Linked Illumina reads were generated with
515 the 10x Genomics ChromiumTM genome protocol following the manufacturer's
516 instructions. These libraries were sequenced on the Illumina NovaSeq with a 150bp
517 paired-end regime. For Bionano optical mapping, genomic DNA was labeled following
518 the Bionano DLS protocol. Labeled genomic DNAs were run on the Bionano Saphyr
519 instrument to 200x genome coverage. Hi-C conformation capture of a third individual
520 was performed following Arima Genomics Hi-C protocol (Arima Hi-C user guide for
521 Animal tissues, v01), an Illumina library of the enriched regions was prepared with the

KAPA HyperPrep Kit, and sequencing was done on an Illumina NovaSeq applying 150 bp paired-end conditions.

Details of all protocols are provided in Supplemental Methods: DNA extractions and genome sequencing.

Genome assembly

We assembled the PacBio long reads into contigs using a customized assembler we termed DAmar, a hybrid of the earlier Marvel (Nowoshilow et al. 2018), Dazzler and Daccord (Tischler and Myers 2017) systems. Next, we used 10x Illumina read-cloud data to correct base errors and phase haplotypes, arbitrarily picking one haplotype in a phased block. Afterwards we used Bionano optical maps and then Hi-C data to produce long-range scaffolds. A manual curation step was performed that uses the Hi-C reads. By inspecting the Hi-C interaction contact matrix with HiGlass (Kerpedjiev et al. 2018) remaining false joins were split and curated scaffolds were joined into chromosomes. Finally, the curated chromosomes were phased by applying an adapted version of the DipAsm pipeline (Garg et al. 2021).

Details of the procedures are provided in Supplemental Methods: Genome assembly.

Genome annotation

The genome was annotated using a pipeline adapted from our previous study (Du et al. 2019; Du et al. 2020; Powell et al. 2020). The repeats of the genome were identified and masked using RepeatModeler, RepeatProteinMask and RepeatMasker (http://www.repeatmasker.org). Then the protein coding genes were annotated by collecting and synthesizing gene evidence from homology, transcriptome and ab-initio predictions. Annotated genes were blasted to database InterProscan, Swiss-Prot and RefSeq for identifying the protein domain and assigning gene symbol and name. At last, Non-coding RNAs (ncRNAs) were annotated using the method adapted from Ensembl (http://ensemblgenomes.org/info/data/ncrna).

Details of the procedures are provided in Supplemental Methods: Genome annotation.

Orthology assignment

Orthology of genes were assigned by a clustering of sequence similarity followed by gene-tree construction in each gene cluster. Sequence similarity was indexed by BLAST (Camacho et al. 2009) score between protein sequences of each two genes. Clustering was implemented by hc_cluster (Ruan et al. 2007) (Supplemental Fig. S15). Then the gene tree of each cluster was constructed using TreeBeST (Ponting 2007).

Orthology between chromosomes from two species was depicted using Circos diagram (Krzywinski et al. 2009) based on orthologous genes with conserved synteny.

Details of the procedures are provided in Supplemental Methods: Orthology assignment.

Phylogenetic analysis

Phylogenetic tree was reconstructed first using maximum-likelihood method implemented in RAxML (Stamatakis 2014), then confirmed the topology using bayesian inference implemented in MrBayes (Ronquist et al. 2012). The input data were protein sequences of 1425 one-to-one orthologs with conserved synteny. We also inferred the divergence time of species using MCMCTree under a relaxed-clock model (Yang 2007; Inoue et al. 2011). Two time-calibrations were set: *O. latipes–T. nigroviridis* (~96.9–150.9Ma) and Clupeiformes-Cypriniformes (~185-225Ma) (Near et al. 2012; Lin et al. 2016; Hughes et al. 2018).

Details of the procedures are provided in Supplemental Methods: Phylogenetic analysis.

Dynamics of gene family size

Gene groups clustered by Hcluster_sg were taken as gene families. Together with the phylogenetic tree built by RAxML and MCMCTree, they were transferred to CAFE5 for gene family size analysis (Zenodo https://doi:10.5281/zenodo.3625141, as developed on GitHub). To estimate the birth-death parameter λ, we only selected gene families that were present in more than seven species and in each species had less than 100 gene family members. Then with the previously estimated λ, gene family contraction and expansion dynamics were accessed for the remaining families. Before λ estimation, an error model was estimated to account for genome assembly and annotation error. λ was estimated using Gamma modelling with two categories.

RAD-tag sequencing and analysis of sex-specific markers

RAD-tag libraries were built from genomic DNA of 29 females and 27 males, and sequenced on the HiSeq 2500 platform. The reads were demultiplexed and then determined present or not in each female and male individual in RADSex (Computing 2013; Feron et al. 2021). A tile plot describing the number of reads in the number of female/male individuals was then generated and used to reveal the sex-determination system of the species. Reads that present only in male were then aligned to the genome to located the sex-determination region.

Details of the procedures are provided in Supplemental Methods: RAD-tag sequencing and analysis of sex-specific markers.

RNA extractions and transcriptome sequencing

Total RNA was isolated using TRIzol Reagent (Thermo Fisher Scientific, Waltham, USA) according to the supplier's recommendation. Custom sequencing (BGI, Shenzen, China) of TruSeq libraries generated 30-35 million 150bp paired end clean reads for each sample on the BGISEQ platform.

Differential gene expression analysis

Clean reads were aligned to genome assembly using STAR (–runMode alignReads –quantMode GeneCounts) (Dobin et al. 2013) and counted and normalized for each gene using DESeq2 (Love et al. 2014). Gene with normalized reads count >10 in a tissue was defined as "expressbased divergence time estimate of splitfined". Expression with upregulation of $\log_2$ fold change >2 in one tissue than in the other was termed as "enhanced expression" in the former tissue. We pooled the data of early and late stage embryos and trophotaeniae considering they are tightly grouped together in both hierarchical clustering and correspondence analysis (Supplemental Fig. S16). For cross-comparison of detected genes, dataset of human, Poeciliopsis and seahorse were retrieved from previous studies (for details how the datasets were produced see Supplemental table S21).

Details of the procedures are provided in Supplemental Methods: Differential gene expression analysis.

Positive selection

Protein and CDS sequences of genes were downloaded from NCBI (Supplemental Table S22). Sequences were aligned using MUSCLE (Edgar 2004) and the format was transformed using an in-house script (Supplemental Code). Positive selection on a gene of the species was detected using 'Environment for Tree Exploration' (ETE3) toolkit (Huerta-Cepas et al. 2016), which automates CodeML and Slr analyses by using pre-configured evolutionary models, the species tree and alignment of the orthologous coding sequences. A gene was identified as a positively-selected gene when it contains sites (probability > 0.95) of either class 2a (positive selection in marked branch and conserved in rest) or class 2b (positive selection in marked branch and relaxed in rest).

Details of the procedures are provided in Supplemental Methods: Positive selection.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA768329 and PRJNA745519. The genome assembly generated in this study has been submitted to the NCBI Assembly database (https://www.ncbi.nlm.nih.gov/assembly) under accession number PRJNA768329. The genome annotation has been submitted to figshare (https://figshare.com/articles/dataset/Girardinichthys_multiradiatus_201028_tar_gz/15173181).

Competing interest statement

The authors declare that they have no competing interests.

Acknowledgements

We gratefully acknowledge the contribution of the Long Read Team of the DRESDEN-concept Genome Center, DFG NGS Competence Center, part of the Center for Molecular

and Cellular Bioengineering (CMCB), Technische Universität Dresden and MPI-CBG. MS was supported by the Deutsche Forschungsgemeinschaft (SCHA 408/10-1). CMG received funding from PAPIIT (PAPIIT IN210718), the science funding agency of the Universidad Nacional Autonoma de Mexico. MP was funded by the Federal Ministry of Education and Research (grant 01IS18026C).

Authors' contributions

M. S, C. M. G. and G. M. conceived and designed the project. E. A. L. and C. M. G. provided the materials. B. W. and S. W. extracted DNA and RNA samples. S. W. conducted the sequencing, M. P. did the assembly and K. D. the annotation. K. D, S. K, I.dC. and M. S. analyzed the sequencing data. Y. G. and R. F. did the RADSex analyses. K. D, R. F, S. K. S. W. and M. P. prepared figures and tables. C. M. G, M. S. and K. D. wrote the manuscript draft, and Y. G. revised it. All authors read and approved the final manuscript.

References

Anchelin M, Murcia L, Alcaraz-Pérez F, García-Navarro EM, Cayuela ML. 2011. Behaviour of telomere and telomerase during aging and regeneration in zebrafish. *PloS one* **6**: e16955.

Arellano-Aguilar O, Macías Garcia C. 2008. Correction for Arellano-Aguilar and Macías Garcia, Exposure to pesticides impairs the expression of fish ornaments reducing the availability of attractive males. *Proceedings of the Royal Society B: Biological Sciences* **275**: 2899-2899.

Arellano-Aguilar O, Macias Garcia C. 2008. Exposure to pesticides impairs the expression of fish ornaments reducing the availability of attractive males. *Proceedings of the Royal Society B: Biological Sciences* **275**: 1343-1351.

BISAZZA A. 1997. Sexual selection constrained by internal fertilization in the livebearing fishXenotoca eiseni. *Animal behaviour* **54**: 1347-1355.

Blackburn DG. 1999. Are viviparity and egg-guarding evolutionarily labile in squamates? *Herpetologica*: 556-573.

Blackburn DG. 2015. Evolution of vertebrate viviparity and specializations for fetal nutrition: a quantitative and qualitative analysis. *Journal of Morphology* **276**: 961-990.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC bioinformatics* **10**: 421.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25-36.

Computing R. 2013. R: A language and environment for statistical computing. *Vienna: R Core Team*.

669 Constantz GD. 1984. Sperm competition in poeciliid fishes. *Sperm competition and*
670 *the evolution of animal mating systems*: 465-485.
671 Darwin C. 1896. *The descent of man and selection in relation to sex*. D. Appleton.
672 De La Vega-Salazar MY, Avila-Luna E, Macías-Garcia C. 2003. Ecological evaluation of
673 local extinction: the case of two genera of endemic Mexican fish,
674 Zoogoneticus and Skiffia. *Biodiversity & Conservation* **12**: 2043-2056.
675 DeChiara TM, Efstratiadis A, Robertson EJ. 1990. A growth-deficiency phenotype in
676 heterozygous mice carrying an insulin-like growth factor II gene disrupted by
677 targeting. *Nature* **345**: 78-80.
678 DeChiara TM, Robertson EJ, Efstratiadis A. 1991. Parental imprinting of the mouse
679 insulin-like growth factor II gene. *Cell* **64**: 849-859.
680 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
681 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*
682 **29**: 15-21.
683 Dobolyi A, Oláh S, Keller D, Kumari R, Fazekas EA, Csikós V, Renner É, Cservenák M.
684 2020. Secretion and function of pituitary prolactin in evolutionary
685 perspective. *Frontiers in Neuroscience* **14**: 621.
686 DOMÍNGUEZ-DOMÍNGUEZ O, MARTÍNEZ-MEYER E, Zambrano L, DE LEÓN GPP.
687 2006. Using ecological-niche modeling as a conservation tool for freshwater
688 species: live-bearing fishes in central Mexico. *Conservation Biology* **20**: 1730-
689 1739.
690 Du K, Stöck M, Kneitz S, Klopp C, Woltering JM, Adolfi MC, Feron R, Prokopov D,
691 Makunin A, Kichigin I. 2020. The sterlet sturgeon genome sequence and the
692 mechanisms of segmental rediploidization. *Nature Ecology & Evolution* **4**:
693 841-852.
694 Du K, Wuertz S, Adolfi M, Kneitz S, Stöck M, Oliveira M, Nóbrega R, Ormanns J, Kloas
695 W, Feron R. 2019. The genome of the arapaima (Arapaima gigas) provides
696 insights into gigantism, fast growth and chromosomal sex determination
697 system. *Scientific Reports* **9**: 5293.
698 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
699 throughput. *Nucleic acids research* **32**: 1792-1797.
700 Emerson J, Li W-H. The genetic basis of evolutionary change in gene expression.
701 Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, Herpin A, Journot L,
702 Parrinello H, Klopp C. 2021. RADSex: A computational workflow to study sex
703 determination using restriction site-associated DNA sequencing data.
704 *Molecular Ecology Resources* **21**: 1715-1731.
705 Fragkiadaki P, Tsoukalas D, Fragkiadoulaki I, Psycharakis C, Nikitovic D, Spandidos
706 DA, Tsatsakis AM. 2016. Telomerase activity in pregnancy complications
707 (Review). *Mol Med Rep* **14**: 16-21.
708 Furness AI, Capellini I. 2019. The evolution of parental care diversity in amphibians.
709 *Nature communications* **10**: 1-12.
710 Garcia CM, Saborío E, Berea C. 1998. Does male-biased predation lead to male
711 scarcity in viviparous fish? *Journal of Fish Biology* **53**: 104-117.

712 Garcia CM, Valero A. 2010. Sexual conflict and sexual selection in the Goodeinae, a
713 clade of viviparous fish with effective female mate choice. *Advances in the*
714 *study of behavior* **42**: 1-54.
715 Garg S, Fungtammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P,
716 Hatas E, Ghurye J. 2021. Chromosome-scale, haplotype-resolved assembly of
717 human genomes. *Nature biotechnology* **39**: 309-312.
718 Ghalambor CK, Reznick DN, Walker JA. 2004. Constraints on adaptive evolution: the
719 functional trade-off between reproduction and fast-start swimming
720 performance in the Trinidadian guppy (Poecilia reticulata). *The American*
721 *Naturalist* **164**: 38-50.
722 González Zuarth C, Macías Garcia C. 2006. Phenotypic differentiation and pre-mating
723 isolation between allopatric populations of Girardinichthys multiradiatus.
724 *Proceedings of the Royal Society B: Biological Sciences* **273**: 301-307.
725 Greven H, Brenner M. 2010. How to copulate without an intromittent organ: the
726 external genital structures and mating behaviour of Xenotoca eiseni
727 (Goodeidae). *Viviparous Fishes II New Life Publications, Homestead, Florida*:
728 446-450.
729 Greven H, Grossherr M. 1992. Adelphophagy and oophagy in Ameca splendens
730 Miller & Fitzsimons, 1971 (Goodeidae. Teleostei). *Z Frischk* **1**: 193-197.
731 Griffith O, Wagner G. 2017. The placenta as a model for understanding the origin
732 and evolution of vertebrate organs. *Nature Ecology & Evolution* **1**: 0072.
733 Guernsey MW, van Kruistum H, Reznick DN, Pollux BJ, Baker JC. 2020. Molecular
734 signatures of placentation and secretion uncovered in Poeciliopsis maternal
735 follicles. *Molecular biology and evolution* **37**: 2679-2690.
736 Guerrero-Estévez S, López-López E. 2016. Effects of endocrine disruptors on
737 reproduction in viviparous teleosts with intraluminal gestation. *Reviews in*
738 *Fish Biology and Fisheries* **26**: 563-587.
739 Guo J, Zhu P, Wu C, Yu L, Zhao S, Gu X. 2003. In silico analysis indicates a similar gene
740 expression pattern between human brain and testis. *Cytogenet Genome Res*
741 **103**: 58-62.
742 Guo JH, Huang Q, Studholme DJ, Wu CQ, Zhao Z. 2005. Transcriptomic analyses
743 support the similarity of gene expression between brain and testis in human
744 as well as mouse. *Cytogenet Genome Res* **111**: 107-109.
745 Herpin A, Schartl M. 2008. Regulatory putsches create new ways of determining
746 sexual development. *EMBO reports* **9**: 966-968.
747 Hu M, Li J, Baker PN, Tong C. 2021. Revisiting preeclampsia: a metabolic disorder of
748 the placenta. *The FEBS Journal*.
749 Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and
750 visualization of phylogenomic data. *Molecular biology and evolution* **33**:
751 1635-1638.
752 Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arcila D, Betancur-R
753 R, Li C, Becker L. 2018. Comprehensive phylogeny of ray-finned fishes
754 (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of*
755 *the National Academy of Sciences* **115**: 6249-6254.

Iida A, Nishimaki T, Sehara-Fujisawa A. 2015. Prenatal regression of the trophotaenial placenta in a viviparous fish, Xenotoca eiseni. *Scientific reports* **5**: 1-5.

Inoue J, Dos Reis M, Yang Z. 2011. A step-by-step tutorial: Divergence time estimation with approximate likelihood calculation using MCMCTREE in PAML. Citeseer.

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome biology* **19**: 1-12.

Kin K, Maziarz J, Chavan AR, Kamat M, Vasudevan S, Birt A, Emera D, Lynch VJ, Ott TL, Pavlicev M et al. 2016. The Transcriptomic Evolution of Mammalian Pregnancy: Gene Expression Innovations in Endometrial Stromal Fibroblasts. *Genome Biology and Evolution* **8**: 2459-2473.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* **19**: 1639-1645.

Lawton BR, Sevigny L, Obergfell C, Reznick D, O'Neill RJ, O'Neill MJ. 2005. Allelic expression of IGF2 in live-bearing, matrotrophic fishes. *Development genes and evolution* **215**: 207-212.

Li H, Liang Z, Yang J, Wang D, Wang H, Zhu M, Geng B, Xu EY. 2019. DAZL is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* **6**: 455-468.

Lin Q, Fan S, Zhang Y, Xu M, Zhang H, Yang Y, Lee AP, Woltering JM, Ravi V, Gunter HM. 2016. The seahorse genome and the evolution of its specialized morphology. *Nature* **540**: 395.

Lombardi J, Wourms JP. 1985. The trophotaenial placenta of a viviparous goodeid fish. II. Ultrastructure of trophotaeniae, the embryonic component. *Journal of morphology* **184**: 293-309.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.

Lynch VJ, Tanzer A, Wang Y, Leung FC, Gellersen B, Emera D, Wagner GP. 2008. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proceedings of the National Academy of Sciences* **105**: 14928-14933.

Ma R, Zhao J, Du H-C, Tian S, Li L-W. 2012. Removing endotoxin from plasmid samples by Triton X-114 isothermal extraction. *Analytical biochemistry* **424**: 124-126.

Macías-Garcia C, Saborío E. 2004. Sperm competition in a viviparous fish. *Environmental Biology of Fishes* **70**: 211-217.

Marina Y, Tabche LMn, García CMa. 1997. Bioaccumulation of methyl parathion and its toxicology in several species of the freshwater community in Ignacio Ramirez dam in Mexico. *Ecotoxicology and environmental safety* **38**: 53-62.

Matos B, Publicover SJ, Castro LFC, Esteves PJ, Fardilha M. 2021. Brain and testis: more alike than previously thought? *Open Biology* **11**: 200322.

Mayr E. 1942. Systematics and theOrigin of Species. *Columbia UniversityPress, New York*.

Méndez-Janovitz M, Garcia CM. 2017. Do male fish prefer them big and colourful? Non-random male courtship effort in a viviparous fish with negligible paternal investment. *Behavioral Ecology and Sociobiology* **71**: 1-12.

Méndez-Janovitz M, Gonzalez-Voyer A, Macías Garcia C. 2019. Sexually selected sexual selection: Can evolutionary retribution explain female ornamental colour? *Journal of evolutionary biology* **32**: 833-843.

Mendoza G. 1972. The fine structure of an absorptive epithelium in a viviparous teleost. *Journal of morphology* **136**: 109-129.

Meyer A, Lydeard C. 1993. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene X-src. *Proceedings of the Royal Society of London Series B: Biological Sciences* **254**: 153-162.

Mossman H. 1991. Comparative morphogenesis of the fetal membranes and accessory uterine structures. *Placenta* **12**: 1-5.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences* **109**: 13698-13703.

Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AW, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**: 50-55.

Panhuis TM, Butlin R, Zuk M, Tregenza T. 2001. Sexual selection and speciation. *Trends in ecology & evolution* **16**: 364-371.

Parenti LR. 1981. A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History*.

Pollux B, Meredith R, Springer M, Garland T, Reznick D. 2014. The evolution of the placenta drives a shift in sexual selection in livebearing fish. *Nature* **513**: 233-236.

Pollux B, Pires M, Banet A, Reznick D. 2009. Evolution of placentas in the fish family Poeciliidae: an empirical study of macroevolution. *Annu Rev Ecol Evol Syst* **40**: 271-289.

Ponting C. 2007. TreeBeST: Tree building guided by Species Tree.

Powell DL, García-Olazábal M, Keegan M, Reilly P, Du K, Díaz-Loyo AP, Banerjee S, Blakkan D, Reich D, Andolfatto P. 2020. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science* **368**: 731-736.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**: 737-746.

Ritchie MG. 2007. Sexual selection and speciation. *Annu Rev Ecol Evol Syst* **38**: 79-102.

Ritchie MG, Webb SA, Graves JA, Magurran AE, Macias Garcia C. 2005. Patterns of speciation in endemic Mexican Goodeid fish: sexual conflict or early radiation? *Journal of Evolutionary Biology* **18**: 922-929.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* **61**: 539-542.

Ruan J, Li H, Chen Z, Coghlan A, Coin LJM, Guo Y, Heriche J-K, Hu Y, Kristiansen K, Li R. 2007. TreeFam: 2008 update. *Nucleic acids research* **36**: D735-D740.

Rueda-Jasso RA, los Santos-Bailón D, Fuentes-Farias AL, Gutierrez-Ospina G. 2014. Toxicidad letal y subletal del fosfato de sodio dibásico y efectos en branquias y conducta de las crías del pez goodeido Skiffia multipunctata. *Hidrobiológica* **24**: 207-214.

Saldivar Lemus Y, Vielle-Calzada JP, Ritchie MG, Macías Garcia C. 2017. Asymmetric paternal effect on offspring size linked to parent-of-origin expression of an insulin-like growth factor. *Ecol Evol* **7**: 4465-4474.

Schindler J, Kujat R. 1990. Structure and function of the trophotaenial placenta in the goodeid teleost, Skiffia bilineata. *Jahrbuch für Morphologie und mikroskopische Anatomie 2 Abteilung, Zeitschrift für mikroskopisch-anatomische Forschung* **104**: 241-257.

Schindler JF. 2003. Scavenger receptors facilitate protein transport in the trophotaenial placenta of the goodeid fish, Ameca splendens (Teleostei: Atheriniformes). *Journal of Experimental Zoology Part A: Comparative Experimental Biology* **299**: 197-212.

Schindler JF. 2015. Structure and function of placental exchange surfaces in goodeid fishes (Teleostei: Atheriniformes). *Journal of morphology* **276**: 991-1003.

Schindler JF, De Vries U. 1987. Protein uptake and transport by trophotaenial absorptive cells in two species of goodeid embryos. *Journal of Experimental Zoology* **241**: 17-29.

Shao F, Wang J, Xu H, Peng Z. 2018. FishTEDB: a collective database of transposable elements identified in the complete genomes of fish. *Database* **2018**.

Simmons LW, Fitzpatrick JL. 2012. Sperm wars and the evolution of male fertility. *Reproduction* **144**: 519.

Sreenivasan R, Cai M, Bartfai R, Wang X, Christoffels A, Orban L. 2008. Transcriptomic analyses reveal novel genes with sexually dimorphic expression in the zebrafish gonad and brain. *PLoS One* **3**: e1791.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Tischler G, Myers EW. 2017. Non hybrid long read consensus using local de Bruijn graph assembly. *bioRxiv*: 106252.

Turner C. 1937. Reproductive cycles and superfetation in poeciliid fishes. *The Biological Bulletin* **72**: 145-164.

Uribe M, Aguilar-Morales M, De la Rosa-Cruz G, García-Alarcón A, Campuzano-Caballero J, Guerrero-Estévez S. 2010. Ovarian structure and embryonic traits associated with viviparity in poeciliids and goodeids. *Viviparous fishes II*: 211-229.

Uribe MC DlR-CG, García-Alarcón A. 2005. The ovary of viviparous teleost. Morphological differences between the ovaries of Goodea atripinnis and

Ilyodon whitei (Goodeidae). In *Viviparous Fishes*, (ed. GH Uribe MC), pp. 217-235. New Life Publications, Homestaed, Florida.

Uyeno T, Miller RR, Fitzsimons JM. 1983. Karyology of the Cyprinodontoid Fishes of the Mexican Family Goodeidae. *Copeia* **1983**: 497-510.

van Kruistum H, van den Heuvel J, Travis J, Kraaijeveld K, Zwaan BJ, Groenen MA, Megens H-J, Pollux BJ. 2019. The genome of the live-bearing fish Heterandria formosa implicates a role of conserved vertebrate genes in the evolution of placental fish. *BMC evolutionary biology* **19**: 1-11.

Vega-López A, Ortiz-Ordóñez E, Uría-Galicia E, Mendoza-Santana EL, Hernández-Cornejo R, Atondo-Mexia R, García-Gasca A, García-Latorre E, Domínguez-López ML. 2007. The role of vitellogenin during gestation of Girardinichthys viviparus and Ameca splendens; two goodeid fish with matrotrophic viviparity. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **147**: 731-742.

Webb SA, Graves JA, Macias-Garcia C, Magurran AE, Foighil DÓ, Ritchie MG. 2004. Molecular phylogeny of the livebearing Goodeidae (Cyprinodontiformes). *Molecular phylogenetics and evolution* **30**: 527-544.

Whittington CM, Griffith OW, Qi W, Thompson MB, Wilson AB. 2015. Seahorse brood pouch transcriptome reveals common genes associated with vertebrate pregnancy. *Molecular Biology and Evolution* **32**: 3114-3131.

Whittington CM, Wilson AB. 2013. The role of prolactin in fish reproduction. *General and comparative endocrinology* **191**: 123-136.

Wright WE, Piatyszek MA, Rainey WE, Byrd W, Shay JW. 1996. Telomerase activity in human germline and embryonic tissues and cells. *Developmental genetics* **18**: 173-179.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**: 1586-1591.

Yuan XN, Jiang XY, Pu JW, Li ZR, Zou SM. 2011. Functional conservation and divergence of duplicated insulin-like growth factor 2 genes in grass carp (Ctenopharyngodon idellus). *Gene* **470**: 46-52.

Figure legends

Fig. 1. Images of a pregnant female adult of darkedged splitfin, one laid embryo with the trophotaeniae and a schematic drawing of the ovary/placental organ. For detail of the organ structure, see Figures 31-33 in Uribe, Aguilar-Morales et al. 2010. For comparison of the placental organs in mammals, seahorse and poeciliid fishes, see Figure 3 in Griffith&Wagner 2017.

Fig. 2. Hi-C contact heat map of the scaffolded portion of the splitfin genome assembly.

Fig. 3. Species tree and Chord diagrams showing orthology between *Girardinichthys multiradiatus* and other teleost species. For each species, chromosome orthology to *G.*

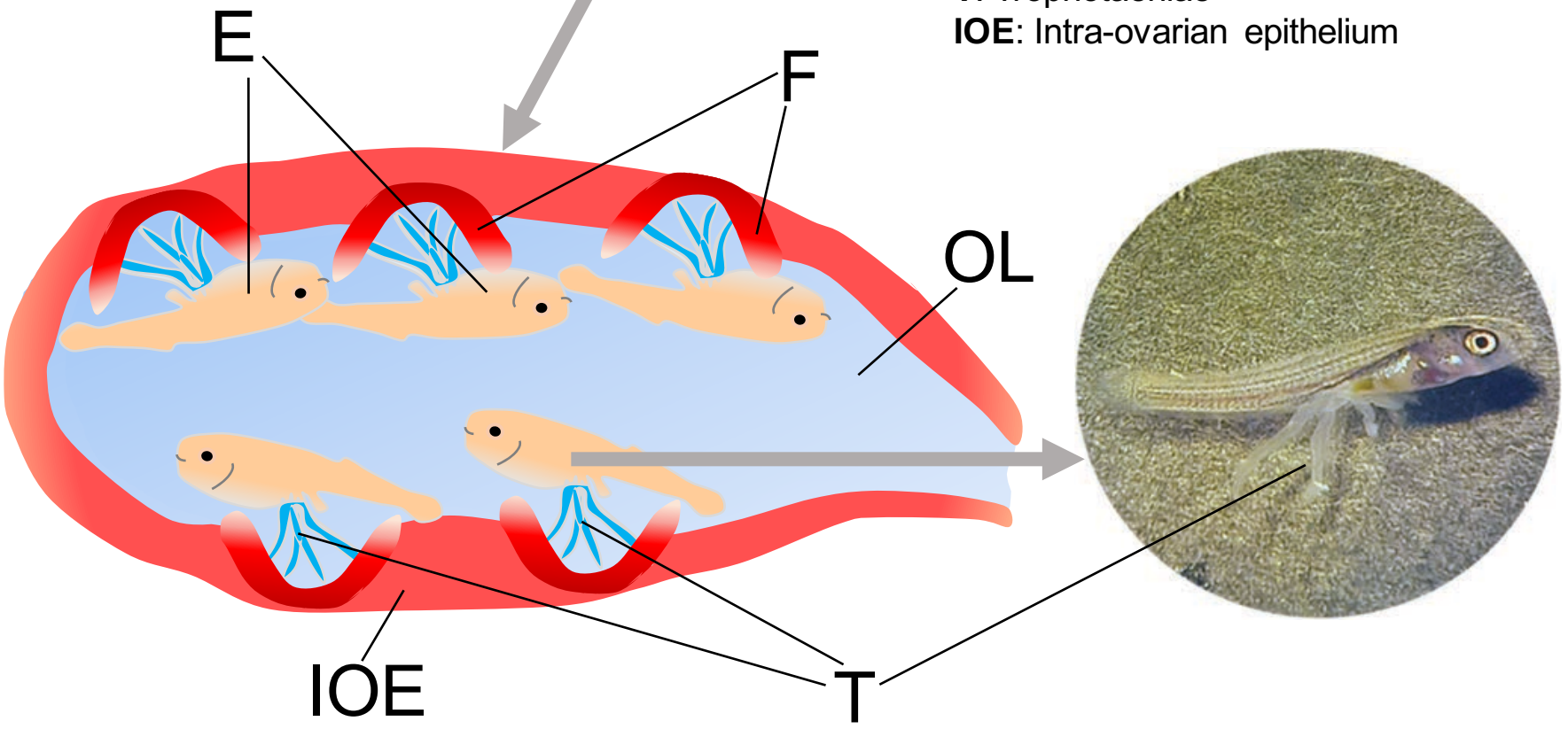
930 *multiradiatus* is shown in a chord diagram, where blocks on circumference represent
931 chromosomes, coloured threads link orthologous genes with conserved synteny.
932 Chromosomes of *G. multiradiatus* are represented by coloured blocks while those of
933 other species are in black. Bootstrap support values are labeled on each node. Scale bar
934 indicates average substitutions per site. Branches leading to live-bearing species are
935 marked with purple four-pointed stars.

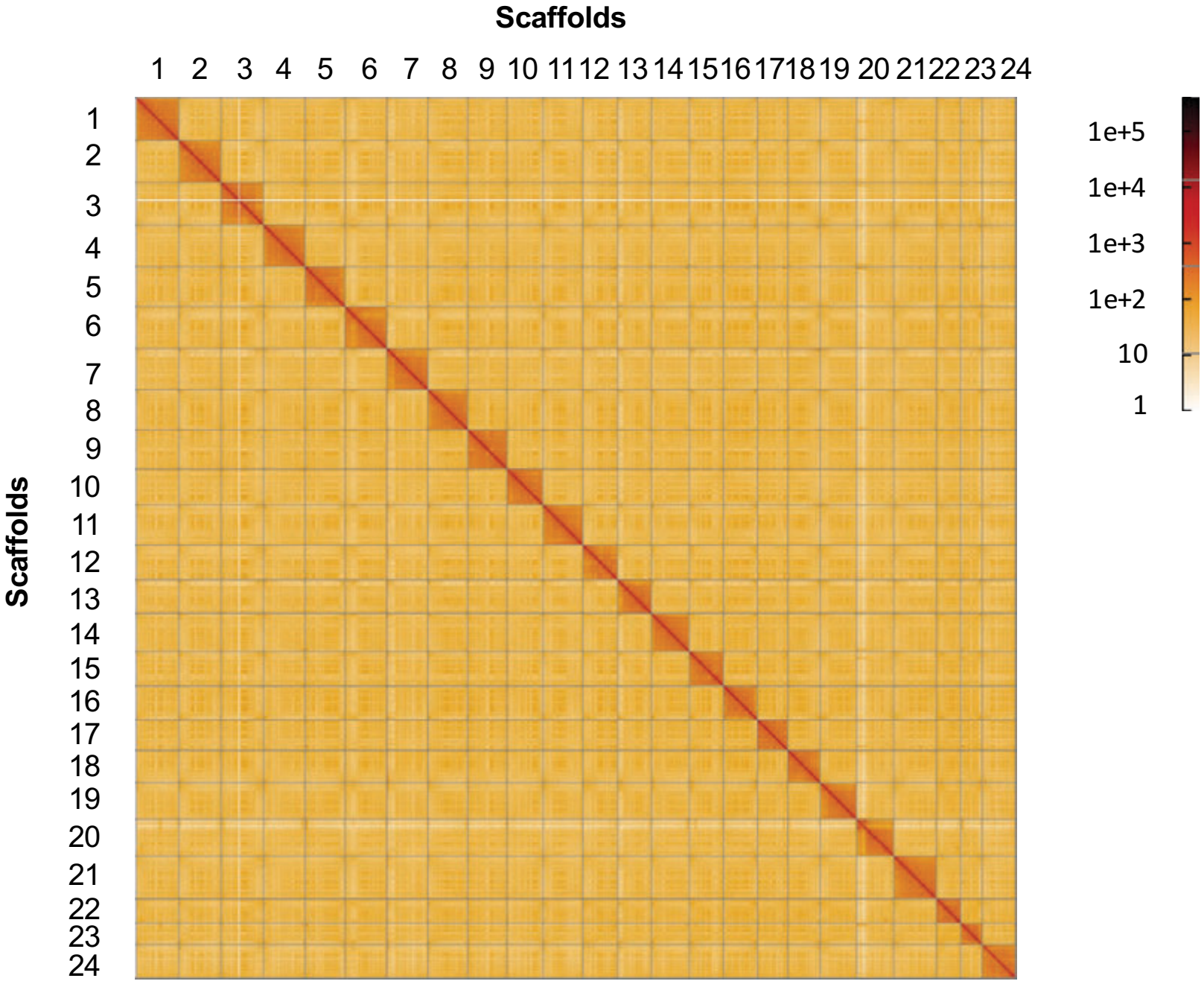
936 Fig. 4. Identification of the sex chromosomes by RAD-tag sequencing. (A) Distribution
937 of RADSex markers between males (horizontal axis) and females (vertical axis) using a
938 minimum depth of 1 to consider a marker present in an individual. There were 892
939 markers significantly associated with male phenotypes in the tiles highlighted in red. (B)
940 Manhattan plot showing the negative log of p-value of association with sex for all
941 RADSex markers aligned to the genome. The vast majority of aligned markers
942 significantly associated with sex (97%) was aligned to scaffold_20. (C) Negative log of
943 p-value of association with sex for all RADSex markers aligned to scaffold_20. Markers
944 significantly associated with sex aligned to a continuous region spanning from the start of
945 the chromosome to ~32.5 Mb.

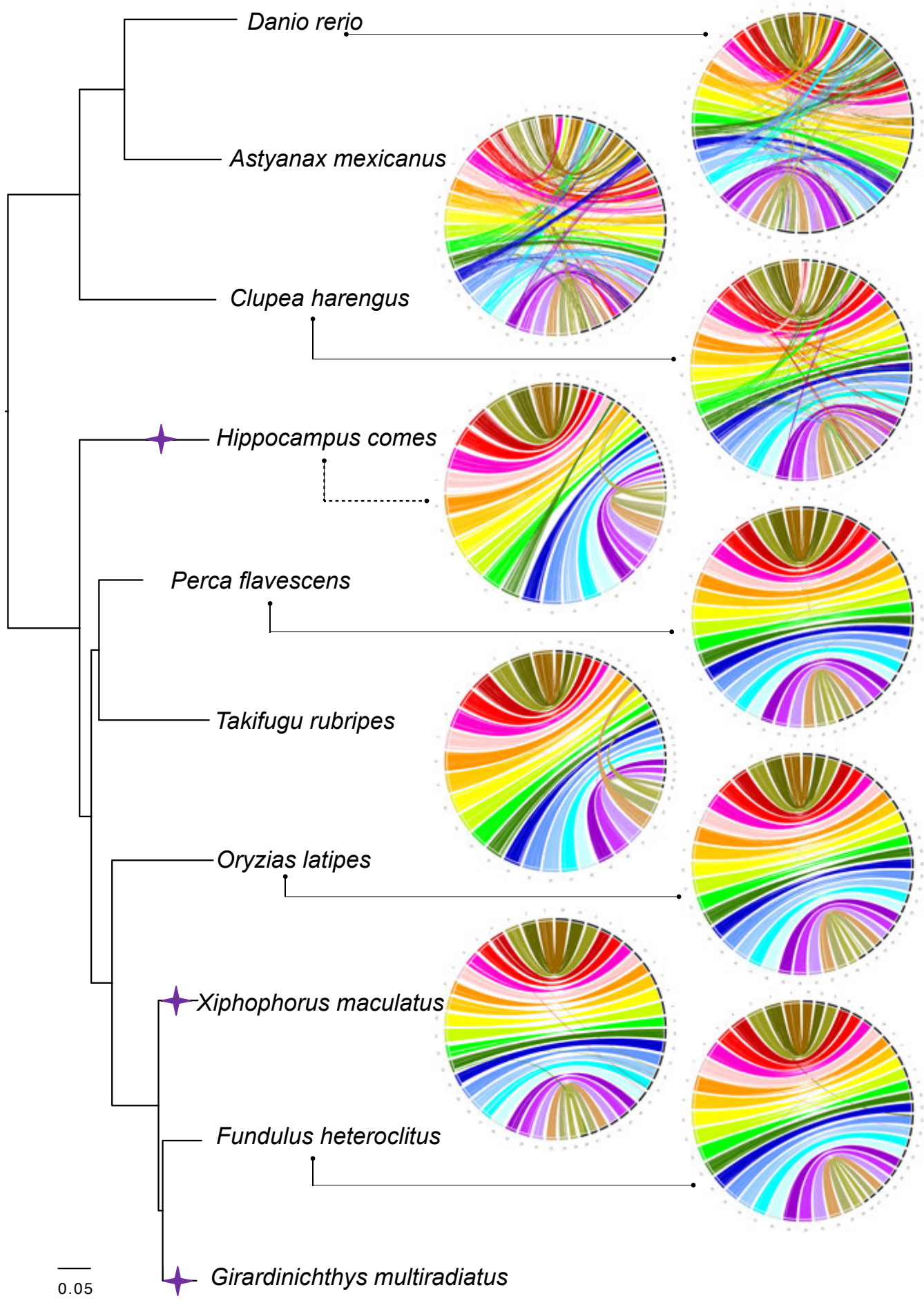
946 Fig. 5. Circos and dot plots showing synteny and sequence difference between
947 Chromosome X and Y. (A) Circos plot of Chromosomes 18 to 24 and both haplotypes of
948 scaffold 20 (X and Y) showing repeat content (gray histogram), GC value (black
949 histogram), male-specific RAD-tag hits (red bars) and synteny (linking ribbon) on
950 chromosomes; with in middle the image of a male darkedged splitfin adult. (B) Dot plots
951 showing the sequence difference (percentage of SNPs and indels in windows of size
952 100k) between X and Y.

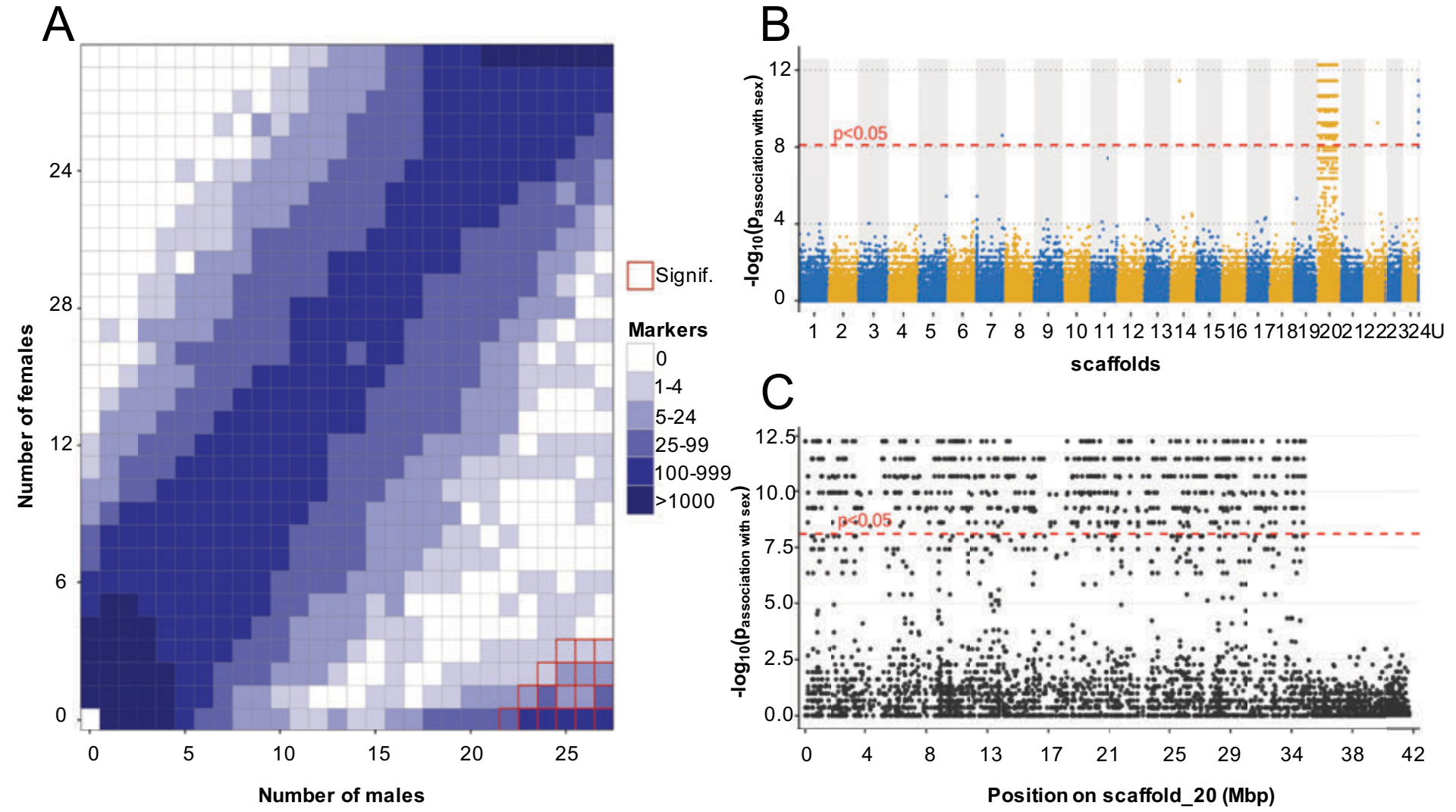
953

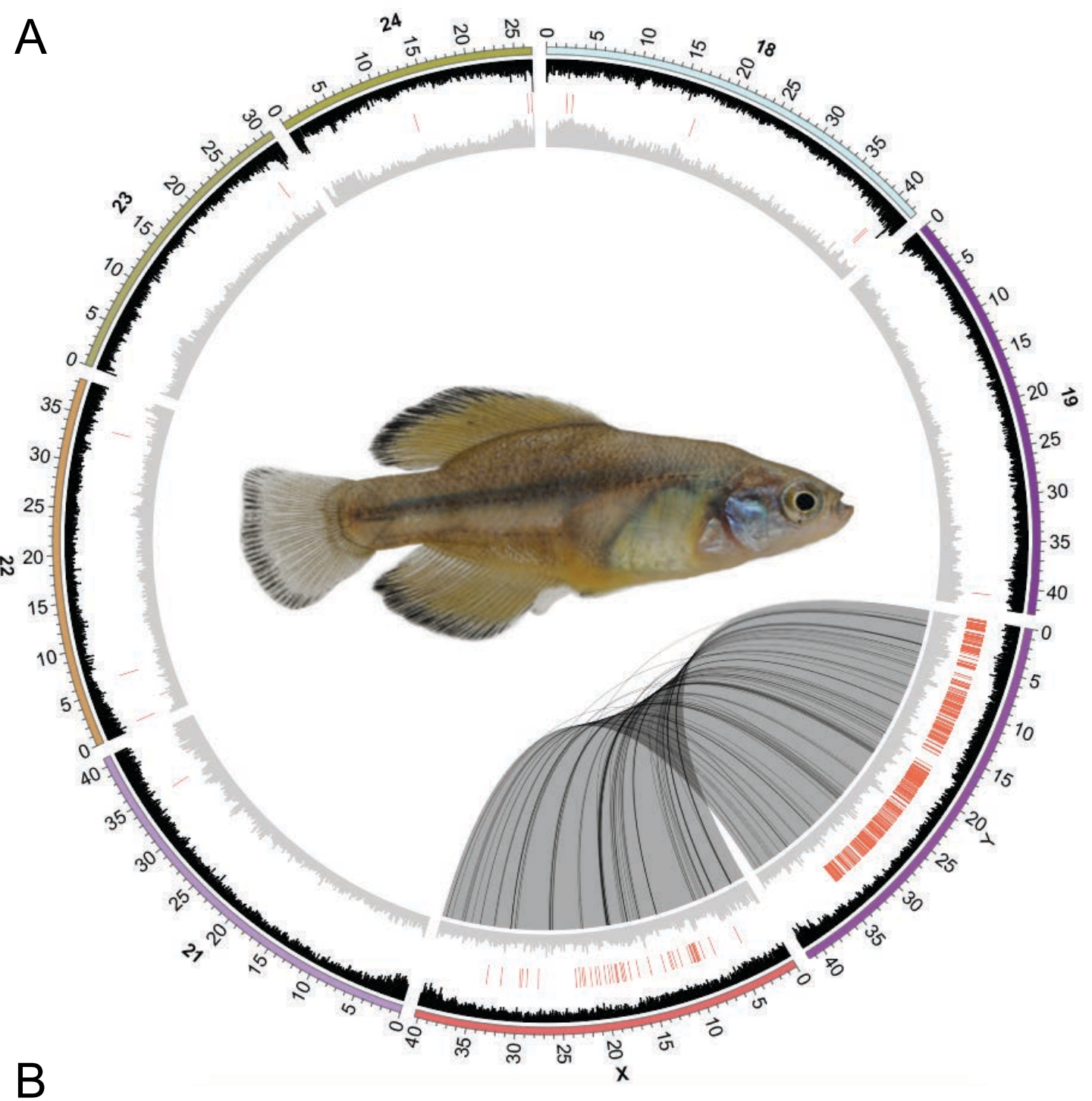
E: Embryos
F: Folicle (ruptured -ovulation- shortly after fertilisation)
OL: Ovarian lumen
T: Trophotaeniae
IOE: Intra-ovarian epithelium

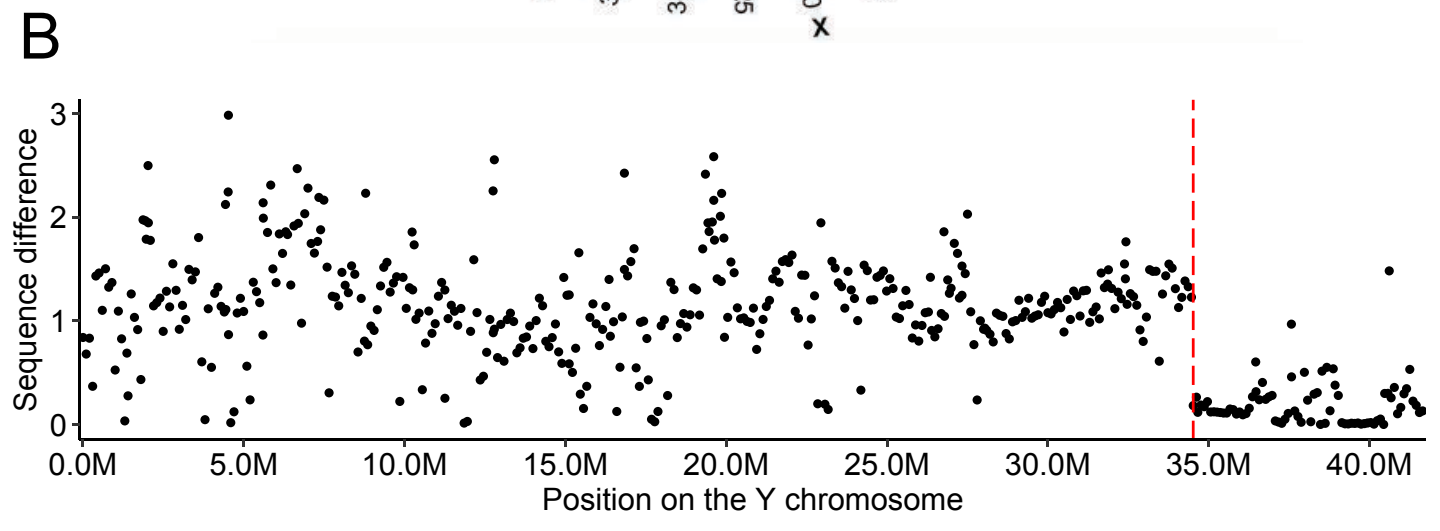



Scaffolds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Scaffolds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1e+5
1e+4
1e+3
1e+2
10
1


Danio rerio
Astyanax mexicanus
Clupea harengus
Hippocampus comes
Perca flavescens
Takifugu rubripes
Oryzias latipes
Xiphophorus maculatus
Fundulus heteroclitus
Girardinichthys multiradiatus
0.05


A
Number of females
Number of males
0
6
12
24
28
0
5
10
15
20
25
Signif.
Markers
0
1-4
5-24
25-99
100-999
>1000
B
-log10(passociation with sex)
p<0.05
0
4
8
12
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 U
scaffolds
C
-log10(passociation with sex)
p<0.05
0.0
2.5
5.0
7.5
10.0
12.5
0 4 8 13 17 21 25 29 34 38 42
Position on scaffold_20 (Mbp)


A
24
18
23
19
22
21
X
Y


B
Sequence difference
0
1
2
3
0.0M
5.0M
10.0M
15.0M
20.0M
25.0M
30.0M
35.0M
40.0M
Position on the Y chromosome