A supernumerary "B-sex" chromosome drives male sex determination in the Pachón cavefish, Astyanax mexicanus

Boudjema Imarazene, Kang Du, Séverine Beille, Elodie Jouanno, Romain Feron, Qiaowei Pan, Jorge Torres-Paz, Céline Lopez-Roques, Adrien Castinel, Lisa Gil, et al.

A supernumerary "B-sex" chromosome drives male sex determination in the Pachón cavefish, *Astyanax mexicanus*

Boudjema Imarazene[1,2], Kang Du[3], Séverine Beille[1], Elodie Jouano[1], Romain Feron[1,4,5], Qiaowei Pan[1,4], Jorge Torres-Paz[2], Céline Lopez-Roques[6], Adrien Castinel[6], Lisa Gil[6], Claire Kuchly[6], Cécile Donnadieu[6], Hugues Parrinello[7], Laurent Journot[7], Cédric Cabau[8], Margot Zahm[9], Christophe Klopp[9], Tomáš Pavlica[10,11], Ahmed Al-Rikabi[12], Thomas Liehr[12], Sergey Simanovsky[13], Joerg Bohlen[10], Alexandr Sember[10], Julie Perez[14], Frédéric Veyrunes[14], Thomas D. Mueller[15], John H. Postlethwait[16], Manfred Schartl[3,17], Amaury Herpin[1], Sylvie Rétaux[2*], Yann Guiguen[1*§]

AUTHOR AFFILIATIONS

[1] INRAE, LPGP, 35000 Rennes, France.

[2] Université Paris-Saclay, CNRS, Institut des Neurosciences Paris-Saclay, 91198 Gif sur Yvette, France.

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

[4] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland.

[5] Swiss Institute of Bioinformatics, Lausanne, Switzerland.

[6] INRAE, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France.

[7] Institut de Génomique Fonctionnelle, IGF, CNRS, INSERM, Univ. Montpellier, F-34094 Montpellier, France.

[8] SIGENAE, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France.

[9] SIGENAE, UMIAT, INRAE, Castanet Tolosan, France.

[10] Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, 27721 Liběchov, Czech Republic.

[11] Department of Zoology, Faculty of Science, Charles University, Viničná 7, 12844 Prague, Czech Republic.

[12] University Clinic Jena, Institute of Human Genetics, 07747 Jena, Germany.

[13] Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia.

[14] Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, Université de Montpellier, IRD, 34095 Montpellier, France.

[15] Department of Plant Physiology and Biophysics, Julius-von-Sachs Institute of the University Wuerzburg, D-97082 Wuerzburg, Germany.

[16] Institute of Neuroscience, University of Oregon, Eugene, USA.

[17] Department of Developmental Biochemistry, University of Wuerzburg, Wuerzburg, Germany.

* Equally contributing senior authors. § Lead contact: Yann Guiguen, INRAE, Laboratoire de Physiologie et Génomique des poissons, Campus de Beaulieu, 35042 Rennes cedex, France. Tel: +33 2 99 46 58 09 (yann.guiguen@inrae.fr). Twitter: @Houpss35

Key words: Sex determination, cavefish, B chromosome, sex chromosomes, *gdf6*

SUMMARY

Sex chromosomes are generally derived from a pair of classical type-A chromosomes, and relatively few alternative models have been proposed up to now[1,2]. B chromosomes (Bs) are supernumerary and dispensable chromosomes with non-Mendelian inheritance found in many plant and animal species[3,4], that have often been considered as selfish genetic elements that behave as genome parasites[5,6]. The observation that in some species Bs can be either restricted or predominant in one sex[7–14] raised the interesting hypothesis that Bs could play a role in sex determination[15]. The characterization of putative B master sex-determining (MSD) genes, however, has not yet been provided to support this hypothesis. Here, *in Astyanax mexicanus* cavefish originating from Pachón cave, we show that Bs are strongly male-predominant. Based on a high-quality genome assembly of a B-carrying male, we characterized the Pachón cavefish B sequence and found that it contains two duplicated loci of the putative MSD gene growth differentiation factor 6b (*gdf6b*). Supporting its role as an MSD gene, we found that the Pachón cavefish *gdf6b* gene is expressed specifically in differentiating male gonads, and that its knockout induces male-to-female sex reversal in B-carrying males. This demonstrates that *gdf6b* is necessary for triggering male sex determination in Pachón cavefish. Altogether these results bring multiple and independent lines of evidence supporting the conclusion that the Pachón cavefish B is a "B-sex" chromosome that contains duplicated copies of the *gdf6b* gene which can promote male sex determination in this species.

RESULTS AND DISCUSSION

Pachón cavefish B chromosomes are male-predominant B chromosomes

Supernumerary B chromosomes (Bs) are generally thought to arise from the duplication and assembly of A chromosome sequences[16–19] and their relationship to sex chromosomes has often been suspected and discussed[15]. Some hypotheses state that Bs are derived from sex chromosomes or, alternatively, evolved to become sex chromosomes[15,20–23]. Because Bs have been described in *A. mexicanus*[19,24,25], we performed cytogenetic analyses in 17 males and 11 females of a laboratory population of Pachón cavefish (Pachón) to investigate whether Pachón Bs could be sex-restricted. We found that Pachón Bs are euchromatic mitotically unstable microchromosomes (Figure 1A-B, Figure S1), that are present in one to three copies in most male metaphases (mean number ± SD of Bs per male metaphase = 1.08 ± 0.41), contrasting with a barely detectable B occurrence in female metaphases (mean number ± SD of Bs per female metaphase = 0.05 ± 0.08, Figure 1C and Data S1A). Chromosomal mapping by fluorescence *in situ* hybridization (FISH) using probes generated from microdissected male Bs painted Bs in males and even the rare B in females, supporting that Pachón male-predominant Bs and the low occurrence female Bs share a similar gDNA content (Figure 1D). In addition, weaker FISH signals were also detected on different terminal parts of some A chromosomes (see small white arrows in Figure 1D), suggesting that Pachón Bs are made up of many duplicated fragments of A chromosomes[16–19], and / or that they share repetitive DNAs with the A chromosomes[21,24,26]. Female or male sex-restricted or sex-predominant Bs have been described in many fishes, like for instance in some cichlids[12–14] and characiforms[8,9,27,28]. In *A. mexicanus*, Bs were recently described as being restricted only in some, but not all, males[19], suggesting population differences in the frequency and sex-linkage of Bs that have also been reported in many species[4–6,29]. In addition to metaphase spreads (Figure 1E-F), we also detected Bs in pachytene chromosome spreads from testes of two Pachón males (N=52 and N=25). Single Bs were found in most cells (44/52 and 23/25) and always as unpaired chromosomes (Figure 1G-H). Hence, Pachón Bs are present in meiotic germ cells and they do not pair with A chromosomes in line with what has been found in other species with Bs[30]. The question of whether Pachón Bs can pair to each other remains open because we detected no case of multiple Bs in these pachytene chromosome spreads.

Characterization of the Pachón cavefish B chromosome sequence

Because the publicly available Pachón genome assembly was obtained from a female[31], we sequenced the genome of a B+ Pachón male to assemble its B sequence. Bs are notoriously difficult to assemble[32,33], due to their complex mosaic composition of A chromosome fragments[16–19] and their high-repeat content[17,19], and most of the B sequence information is from short-read sequencing of purified Bs[17,18] or B+ versus B-devoid (B-) individuals[17–19,34]. To accurately assemble a high-quality Pachón B sequence, we used a combination of HiFi PacBio and Oxford nanopore long-reads, 10X genomics Illumina short-linked reads, and a Hi-C chromosome contact map. The resulting whole-genome assembly (see assembly metrics in Data S1B) contains 25 large scaffolds corresponding to the 25 Pachón A-chromosome pairs[19,25] and 170 remaining unplaced scaffolds (2.12% of the total assembly size).

To identify the Pachón B sequence we used its male-predominant feature and a pool-sequencing (pool-seq) approach to contrast whole-genome sequences of a genomic DNA (gDNA) pool of 91 phenotypic males versus a gDNA pool of 81 phenotypic females. By remapping these male and female pool-seq reads on our male Pachón genome assembly, we identified a single 2.97 Mb unplaced scaffold (HiC_Scaffold_28) that displays a clear male-biased read coverage profile (Figure 2A-A'). The sequence analysis of HiC_Scaffold_28 revealed that it is made up from a complex mosaic of numerous duplicated fragments of A chromosomes (Figure 2B) including complete but also truncated duplicates of A chromosome genes (Table S3). The B also displays a repeat content that is markedly different from A chromosomes (Figure 2C-D, Figures S2A-B, Data S1D). Both its sex-biased profile and its sequence characteristics indicate that HiC_Scaffold_28 is the Pachón B. The contribution of many A chromosome regions to the structure of Pachón Bs is in line with the recent findings that *A. mexicanus* Bs contain a large number of transposable elements[19]. The high B proportion of satellite DNA (Figures S2 and Data S1D), was also reported in other species[35]. Our manually curated B gene annotation (Data S1C) identified 63 genes on the Pachón B. Of these, 20 show high-quality annotation over their full length, 11 are truncated compared to their conserved homologs in other fish and one is a chimeric gene. Five genes show multiple copies and constitute almost one-third of the B gene content (Data S1C). These results contrast with earlier studies which reported a much higher number of B genes in *A. mexicanus*[19]. These annotation differences are likely due to indirect assessment of the B gene content based on short-read sequencing of very few B+ versus B- individuals[19]. This comparison clearly illustrates the need for better, complete and high-quality B assemblies like we provided here for Pachón cavefish, to better understand the structure and gene content of Bs generally.

The Pachón cavefish B contains two copies of a putative master sex determining gene

Among the B genes with well-supported annotation evidence (Data S1C), we identified two duplicated loci of the A-chromosome-3 growth differentiation factor 6b (*gdf6b*) gene (located at Chr03:863,919-866,170), that is the teleost ohnolog (teleost whole genome duplication[36,37] paralogous copy) of the *gdf6a* gene (Figure 3A). *Gdf6* genes belong to the TGF-β superfamily within which many master sex-determining (MSD) genes have been found, including TGF-β receptors[38–40] and ligands[41–44]. Of note, *gdf6a* on the Y chromosome (*gdf6aY*) has been characterized as the master sex-determining gene in the turquoise killifish, *Nothobranchius furzeri*[44]. The Pachón B *gdf6b* genes *(B-gdf6b = B-gdf6b-1* and *B-gdf6b-2)* were thus retained as potential candidate B MSD genes. Chromosome FISH hybridization with a *gdf6b* locus probe revealed a *gdf6b* hybridization signal on the Pachón Bs, along with a single *gdf6b* labelling on a single A chromosome pair, both in the male-predominant Bs and the low occurrence female Bs (Figure 1E-F, and inset in Figure 1E).

The two *B-gdf6b* loci are 99.6 % identical in the 21.6 kb region shared by the two genes, and 100% identical in their coding sequences (CDS), with *B-gdf6b-2* being derived from an internal B duplication of the *B-gdf6b-1* locus (Figure S2C). Such internal B duplications / insertions indicate that the origin of the B structure can be more complex than initially thought. Comparison of these *B-gdf6b* loci with the overlapping sequence of their A chromosome counterpart (*A-gdf6b*) revealed numerous differences in their proximal promoters and also their intron that contains two *B-gdf6b* specific insertions (Figure S2D). However, differences within the *gdf6b* CDS were limited to a T-to-C (*A-gdf6b*-to-*B-gdf6b*) synonymous substitution (c.591T>C) and two nonsynonymous substitutions, i.e., a T-to-G (*A-gdf6b*-to-*B-gdf6b*) transversion (c.180T>G) in exon 1 that switches a A-Gdf6b lysine into a B-Gdf6b asparagine (p.Lys60Asn) and a G-to-A (*A-gdf6b*-to-*B-gdf6b*) transition (c.679G>A) in exon 2 that switches the A-Gdf6b serine into a B-Gdf6b glycine (p.Ser227Gly) (Figure S2E). The Lys60Asn does not impact a conserved amino-acid position of Gdf6b proteins, in contrast to the Ser227Gly that impacts a glycine of the TGF-β/BMP propeptide domain that is conserved in Pachón B-Gdf6b and in all vertebrate Gdf6 proteins, but not in Pachón A-Gdf6b (Figure 3B). It is interesting to note that this Ser227Gly modification suggests that the *A-gdf6b* acquired this mutation after the *B-gdf6b* copy was duplicated on the B chromosome. This non-conserved Ser227 engages in a central hydrogen bond network at a looptip region in the TGF-β/BMP propeptide domain (Figure 3C) that could affect the stability of this pro-domain. The Gly227Ser exchange in Pachón A-Gdf6b leads to the gain of several hydrogen bonds that are due to the side chain hydroxyl group of the Ser227 residue (Figure 3D-E). The gain of hydrogen bonds can confer a higher folding stability of the proprotein complex and thereby might affect activation of Gdf6b as this requires release of the mature C-terminal domain from the proprotein complex. The mature C-terminal growth factor domain, however, is likely to be unaffected by this mutation. Whether these conformation differences between the A-Gdf6b and B-Gdf6b proproteins could provide a potential functional explanation for a sex-determining role of the male-predominant Pachón B-Gdf6b remains to be explored, but point mutations in other MSD genes like *amhr2Y* in *Takifugu rubripes* and *amhY* in *Oreochromis niloticus*, have been described to be directly responsible for male sex determination[39,41].

Based on these *B-gdf6b* and *A-gdf6b* loci differences, we developed several B-specific PCR genotyping tests on fin clips (Figure S3A). In our Pachón laboratory population, we found a complete (100%) association between B-specific amplifications and the male phenotype in 723 males, with all the 787 tested females being negative (p-value of association with sex < 2.2e-

16). We also found the same complete association in wild-caught Pachón individuals (20 males and 20 females, recognized by external sex-specific traits without sacrifice; p-value of association with sex = 1.87e-09) (Data S1F) showing that this male-predominant B is not the result of a domestication effect. These results strengthen our cytological observations of a male-predominant B. The absence of B-specific amplifications in females despite the cytogenetic detection of rare Bs in females is probably the result of a PCR sensitivity issue as increasing the number of PCR cycles allows the detection of a faint PCR fragment in Pachón females (Figure S3B).

Evidence supporting *gdf6b* as a potential master sex determining gene in Pachón cavefish

Sex-specific expression patterns during the sex differentiation period and alteration of gonadal development upon knockout are key arguments for the evaluation of a candidate MSD gene. Due to the high sequence identity of the *B-gdf6b* and *A-gdf6b* cDNAs, we were not able to specifically quantify the *B-gdf6b* expression. Quantification of the expression of *gdf6b* (*B-gdf6b* and/or *A-gdf6b*) showed that it has both a predominant expression in developing and adult male gonads (as well as in male brain, intestine, kidney and swim-bladder; Figure S4A) and a strong sexually dimorphic expression during early differentiation of B+ individuals (Figure 4A). Using *in situ* hybridization, the expression of *gdf6b* during the early differentiating period was restricted to gonads of B+ individuals (at 15, 21, 30, and 60 days post-fertilization), with no strict colocalization with the gonadal soma-derived factor gene (*gsdf*) (Figure 4B-C and Figure S4B). Gsdf is a well-known gonad-restricted, somatic supporting cell lineage marker[45–47] that has also been described as one of the earliest Pachón gonadal sex differentiation marker genes[48]. This result demonstrates that *gdf6b* mRNA (*B-gdf6b* and/or *A-gdf6b*) has an expression profile compatible with a male (B+) MSD function, being expressed in the right place, i.e., only in the differentiating testis, and at the right time, i.e., during early testicular differentiation.

To bring additional and functional evidence that *gdf6b* could act as an MSD gene, we generated *gdf6b* knockouts in Pachón cavefish using the genome-editing CRISPR-Cas9 system with two guide RNAs in order to remove a large part of the *gdf6b* CDS (Figure 4D). This large deletion (~470 bp) includes most of the TGF-β propeptide region and the beginning of the TGF-β like domain resulting in a truncated, likely non-functional, Gdf6b protein (Figure S2F). Among 200 first-generation microinjected individuals (which are mosaic for the genome-edited loci), eighteen B+ individuals (i.e., genotypic males) had a ~470 bp deletion in their *A*- and/or *B*-

gdf6b exon 2 (Figure S3C), and they were all sex-reversed into phenotypic females (Figure 4G). In contrast, all B+ males and B- females without the *gdf6b* deletion developed normal testes (B+) or ovaries (B-) (Figure 4E-F). This result shows that *gdf6b* is necessary to trigger Pachón testicular development in B+ individuals and brings further functional evidence that *gdf6b* could be used as a male MSD gene in Pachón cavefish. However, because of the high similarity of the *B-gdf6b* loci with the *A-gdf6b* locus, we have not been able to specifically knockout the Pachón *B-gdf6b*. Further studies will bring more functional proof supporting the role of the *B-gdf6b* genes in sex determination, including the specific *B-gdf6b* knockout in B+ individuals and the overexpression of *gdf6b* by transgenesis in B- fish.

Conclusions

Altogether our results bring new pieces of evidence to support a role of some B chromosomes in sex determination. The potential implication of Bs in sex determination had been suspected in many species including some fishes[14,15]. Up to now, however, the characterization of potential B MSD genes along with strong functional evidence has only been provided for the bacterial-derived *haploidizer* gene in the jewel wasp, *Nasonia vitripennis* B chromosome (named PSR for paternal sex ratio chromosome)[49], although maleness in this haplo-diploid organism is determined through elimination of the paternal A chromosome set[50]. Here, combining a variety of approaches we discovered that Pachón cavefish of the species *A. mexicanus* carry male-predominant Bs that contain two copies of the *gdf6b* gene, which itself behaves as an excellent MSD candidate gene. This indeed brings the interesting hypothesis that these Pachón Bs could be considered as "B-sex" chromosomes. However, the question remains open whether Pachón Bs are predominant in males because they are eliminated from female tissues or whether they are by themselves necessary and sufficient to trigger maleness. Despite being male-predominant, Pachón Bs are also found in some females albeit only in very few metaphases (21.6 times less abundant than in males on average) and most often as a single B copy. B frequencies have been described as being highly variable between species, sexes, individuals of the same population, and even in different cells of a single individual[4–6,29]. This variation is assumed to result from meiotic and/or mitotic instability of Bs that can be present only in some organs and absent from others[4,29,51–53]. In the plant *Aegilops speltoides*, a mechanism of programmed elimination of Bs occurs specifically in the roots[54]. It results from a B chromatid nondisjunction during mitosis, leading to the micronucleation of Bs and their subsequent degradation at early stages of the proto-root embryonic tissue differentiation[54]. Such a mechanism would potentially explain a specific B elimination in Pachón female organs, as it has been hypothesized in another *Astyanax* species with male-restricted Bs[11]. Further studies are now needed to better understand this sex-specific B drive mechanism, and if it reflects a cause or a consequence of sex determination in Pachón cavefish. Our results also lay a high-quality genome-based foundation in an important emerging fish model for studying the genomic evolution of Bs, including the micro- and macro-evolution of this B chromosome in line with the evolution of sex chromosomes.

ACKNOWLEDGEMENTS

We thank Victor Simon, Stéphane Père, Krystel Saroul, Pierre-Lo Sudan and Amélie Patinote for taking care and handling cavefish, and Manon Thomas and the LPGP TEFOR infrastructure platform for acquisition of the RNAScope confocal images. This project was supported by funds from the "Agence Nationale de la Recherche" (ANR/DFG, PhyloSex project, 2014-2016) to Y.G and M.S. S.R was supported by grants from an Equipe FRM (Fondation pour la Recherche Médicale, DEQ20150331745) and MITI CNRS (Mission pour les Initiatives Transverses et Interdisciplinaires). J.H.P was supported by a NIH grant (R35 GM139635). Sequencing was supported by France Génomique as part of an "Investissement d'avenir" program managed by ANR (contract ANR-10-INBS-09) and by the GET-PACBIO program (« Programme opérationnel FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020 »). The CytoEvol platform at ISEM was supported by the Labex CeMEB. J.B, T.P and A.S were supported by RVO: 67985904 of IAPG CAS, Liběchov. T.P was supported by the projects of the Czech Ministry of Education (SVV 260571/2021). S.S was supported by the Russian Foundation for Basic Research (RFBR) (18-34-00638). B.I PhD fellowship was supported by the Doctoral School of Ecology, Geosciences, Agronomy, Nutrition of the University of Rennes 1 and INRAE. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing help, computing and storage resources. Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, S.R., A.H. and Y.G.; Methodology, B.I., S.R., A.H., A.S., and Y.G.; Formal Analysis, T.D.M., K.D., R.F., Q.P., C.C., M.Z., and C.K.; Investigation, B.I., S.B., E.J., J.T-P.,

C.L-R., A.C., L.G., C.K., C.D., H.P., L.J., T.P., A.A-R., T.L., S.S., J.B., A.S., J.P., F.V.; Writing –Original Draft, B.I., S.R., and Y.G.; Writing –Review & Editing, A.S., J.H.P., T.D.M., and M.S.; Visualization, B.I., K.D., E.J., R.F., Q.P., C.C., A.S., T.D.M., and Y.G.; Funding Acquisition, S.R., M.S., J.H.P., and Y.G.; Supervision, S.R., A.H., and Y.G.

DECLARATION OF INTERESTS

The authors declare no competing interests

FIGURE

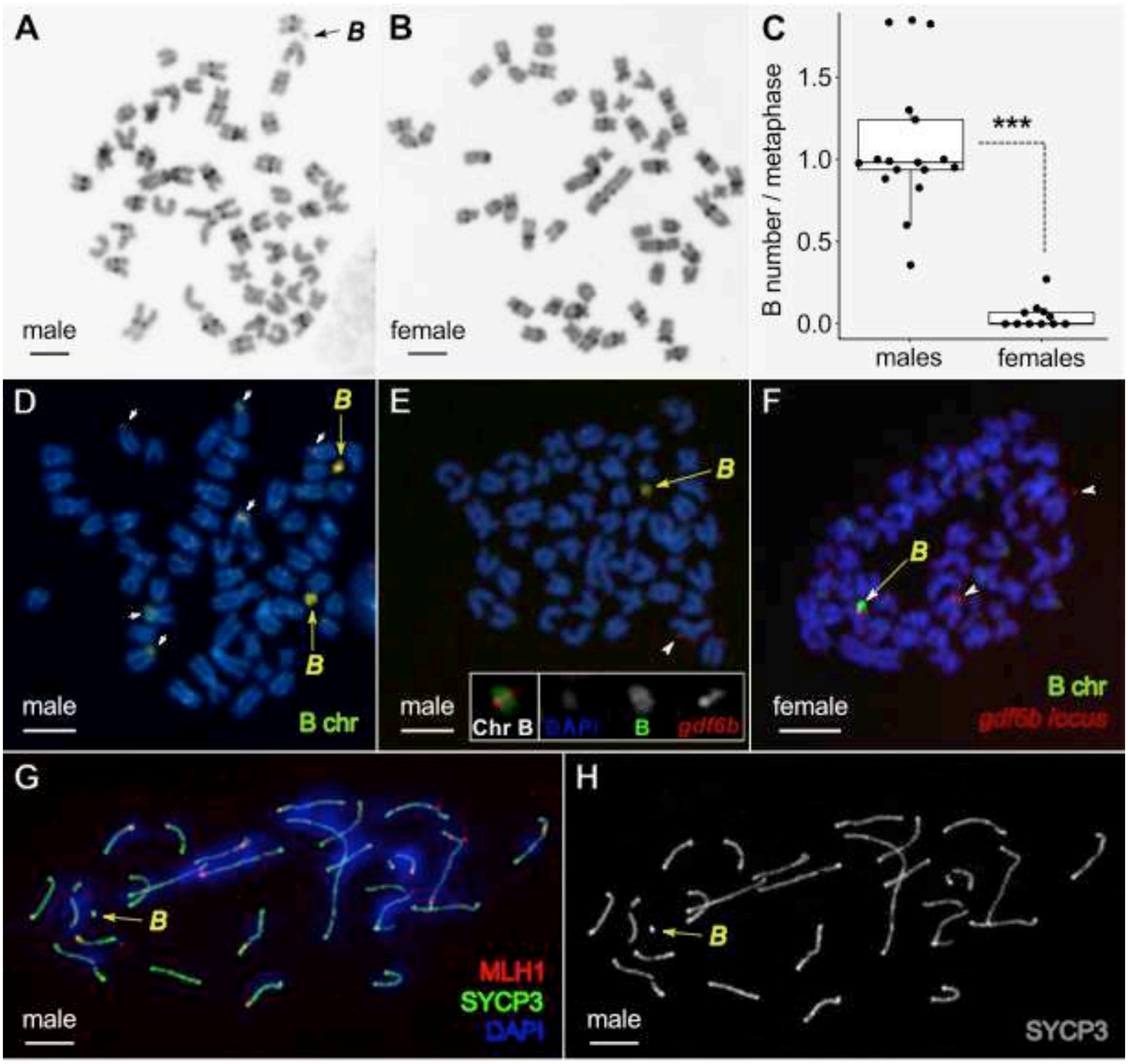

Figure 1: Karyological characterization of male-predominant supernumerary B chromosomes (Bs) in Pachón cave *Astyanax mexicanus*. A-B: Representative C-banding patterns of single B+ male (**A**) and B- female (**B**) from Pachón cave. The Bs (black arrow) lack C-bands, suggesting that Pachón cavefish Bs are largely euchromatic. See also Figure S1. **C:** Boxplots of the average number of Bs per metaphase in Pachón cavefish males and females. Horizontal lines indicate the median, the box indicates the interquartile range (IQR), and the whiskers the range of values that are within 1.5 x IQR. Statistical significance between males and females were tested with the Wilcoxon Rank Test (*** = $P < 0.001$). See also Data S1A. **D:** Fluorescence *in situ* hybridization (FISH) of male Pachón cave mitotic metaphase labeled with combined microdissected B probes. Yellow arrows point to the strong labelling of Bs and the small white arrows to the lighter labelling of some different parts of A chromosomes. **E-F:** FISH co-labelling of a 1B male (**E**) and a 1B female (**F**) metaphase with microdissected B (green) and *gdf6b*-specific (red) probes. Bs are indicated by yellow arrows and the white arrowheads point to pairs of A chromosome sister chromatids labelled by the *gdf6b* probe. Only one A-chromosome *gdf6b* signal was detected in panel **E**. The two *gdf6b* signals (see inset in **E)** that were often visible on male metaphases, cannot be interpreted as the two different *B-gdf6b* loci due to their genomic proximity (see Table S3) and the FISH resolution. **G-H**: Synaptonemal complex (SC) analysis showing that Pachón cave Bs (yellow arrow) do not pair with the 25 fully synapsed standard bivalents of A chromosomes. SCs were visualized by anti-SYCP3 antibody (green), the recombination sites were identified by anti-MLH1 antibody (red) and chromosomes were counterstained by DAPI (blue). **G:** Merged image. **H:** SYCP3 visualization only. Scale bars = 5 μm.

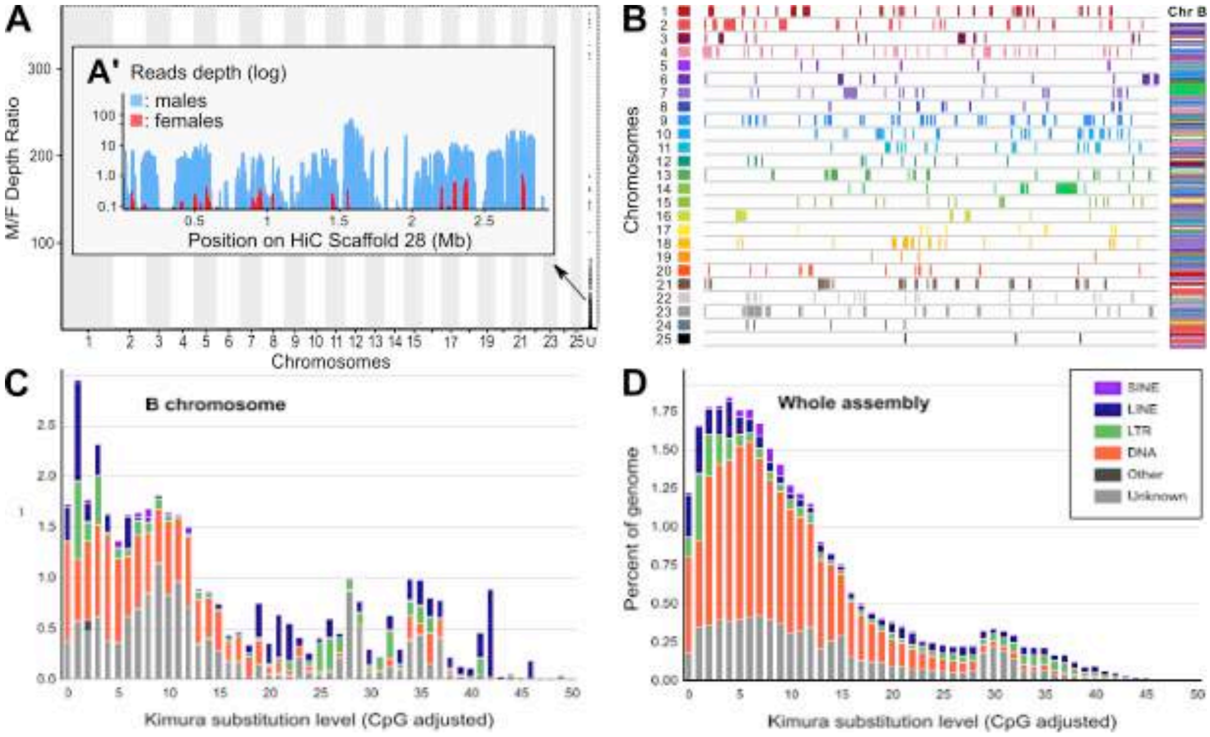

Figure 2: **Genomic characterization of Pachón cavefish B chromosome (B). A:** Read depth ratio of male and female Pachón genomic pools showing a strong coverage bias in a single scaffold, Hi_scaffold_28 (enlarged in **A'** inset showing male and female read coverage). **B:** Karyoplots of the A chromosome regions duplicated on the Pachón cavefish B (Chr B) showing that the Pachón B is made from a complex mosaic of duplicated A chromosomal fragments. **C-D:** Comparison of the repeat landscapes of the Pachón B (**C**) and whole-genome including the B (**D**), showing that the Pachón B has a very different repeat element (color code provided in inset of panel C) content compared to A chromosomes. Short interspersed repeated sequences (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), DNA repeat elements (DNAs), and terminal inverted repeat sequences (TIRs). See also Figure S2A-B and Data S1D for additional details.

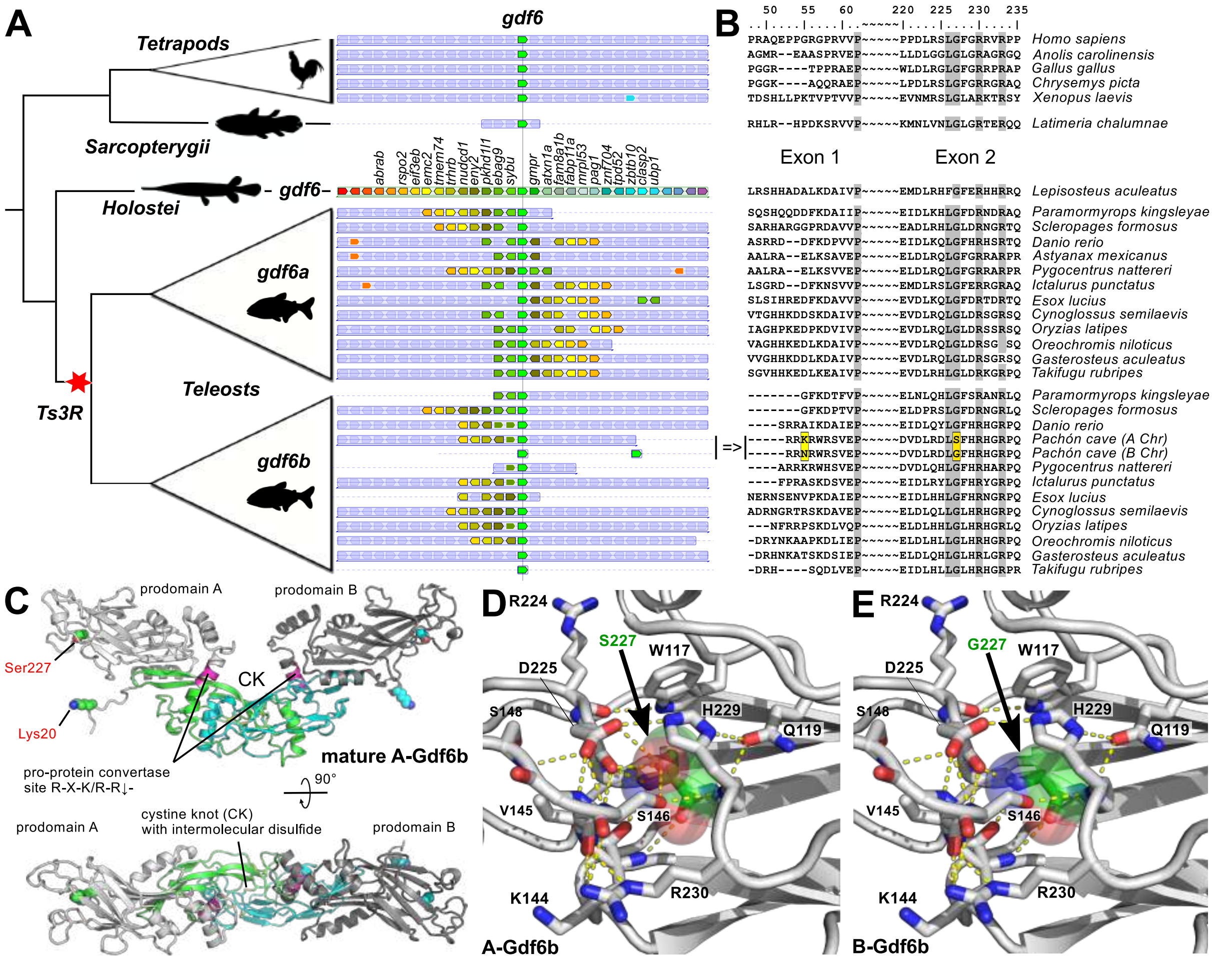

Figure 3: **Gdf6b protein evolution and structure. A:** Simplified phylogeny (left panel) and synteny (middle panel) relationships of the *gdf6a* and *gdf6b* genes, with the additional *B-gdf6b* Pachón cavefish duplication, showing that they are duplicated paralogues stemming from the teleost whole genome duplication (Ts3R). Species names are given on the right side of panel B. **B:** Corresponding multiple alignments of Gdf6 protein-coding sequences around the A-Gdf6b lysine to B-Gdf6b asparagine switch in Exon 1 (p.Lys60Asn) and the A-Gdf6b serine to B-Gdf6b glycine switch in Exon 2 (p.Ser227Gly). **C:** Ribbon plot homology model of A-Gdf6b proprotein dimer (bottom panel view is rotated by 90° around the x-axis). The prodomains are shown in light and dark grey, the furin processing site (R274-R-K-R-R278) is indicated in magenta, the activity-containing mature C-terminal domain is shown in green and cyan. The two residues differing between A-Gdf6b and B-Gdf6b are presented as spheres with their carbon atoms colored in green and cyan. **D-E:** Comparative magnifications of the structure of A-Gdf6b (**D**) and B-Gdf6b (**E**) around the p.Ser227Gly switch. Amino acid residues interacting with Ser227 or Gly227 are shown as sticks, and hydrogen bonds as yellow stippled lines. As shown, the side chain hydroxyl group of Ser227 (shown with carbon atoms colored in green and transparent spheres highlighting the van der Waals spheres of the atoms)

engages in several hydrogen bonds with surrounding residues, e.g. Ser146, Asp225 thereby stabilizing the tertiary and secondary structure in this region. Upon exchange of Ser227 with a glycine as in B-Gdf6b these hydrogen bonds are lost thereby potentially destabilizing this region in the prodomain.

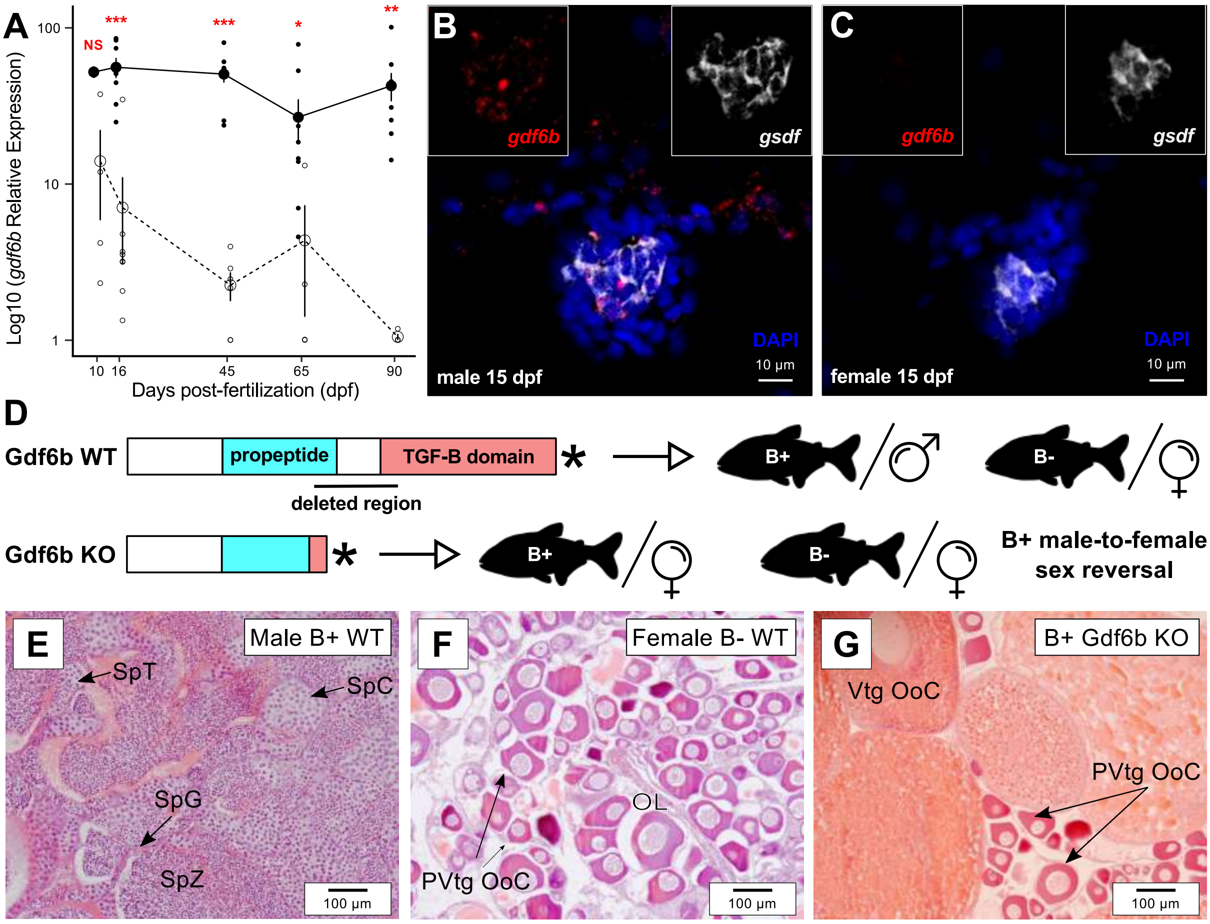

Figure 4: **Gene expression and functional evidence supporting a role of *gdf6b* as a potential master sex-determining gene in Pachón cavefish. A:** Expression profiles of *gdf6b* in male and female trunks during early development from 10 to 90 days post-fertilization (dpf, males: solid line; females: dashed line) showing a significant over-expression in males compared to females starting from 16 dpf. Results are presented as $\log_{10}$ mean ± standard errors; black and white dots represent the individual values of relative expression in males and females, respectively. Statistical significance between males and females was tested with the Wilcoxon Rank Sum Test (Wilcoxon-Mann-Whitney Test) and only significant differences are shown (*** = P < 0.01; ** = P < 0.01; * = P < 0.05). **B-C:** Gonadal expression of *gdf6b* (in red) and the Sertoli and granulosa supporting cell marker *gsdf* (in white) in male (**B**) and female (**C**) differentiating gonads at 15 dpf showing that *gdf6b* is specifically expressed in male gonads with no strict colocalization with *gsdf* in male. See Figure S6B for additional stages of

development. Nuclei were stained with DAPI (in blue). Scale bar = 10 µm. **D:** Schematic representation of the wild-type (WT) and knock-out (KO) Gdf6b proteins and the resulting phenotypes of B+ males and B- females. **E-F:** Representative gonadal histology of WT males (**E**), WT females (**F**), and *Gdf6b KO* B+ males showing that *Gdf6b* KO induces male-to-female sex reversal (**G**) with ovaries containing vitellogenic (Vtg Ooc) and previtellogenic oocytes (PVtg Ooc), like WT ovaries (**F**), contrasting with the testis in WT males (**E**). Ol: Ovarian lamellae. SpC: Spermatocytes; SpG; Spermatogonia; SpT: Spermatids; SpZ: Spermatozoa. Scale bar = 100 µm. See also Figure S4.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yann Guiguen (yann.guiguen@inrae.fr).

Materials availability

To request Pachón fish lines or constructs created in this study, please contact the lead contact.

Data and code availability

Raw sequences and the whole genome assembly of Pachón cavefish have been deposited in the National Center for Biotechnology Information DDBJ/ENA/GenBank databases under the BioProject PRJNA734455. This accession number is listed in the key resources table. This study did not generate new unique code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cavefish breeding and sampling

Laboratory stocks of *A. mexicanus* Pachón cavefish were obtained in 2004 from the Jeffery laboratory at the University of Maryland, College Park, MD. Fish were raised as previously described. Fertilized eggs were provided by CNRS cavefish experimental facilities (Gif sur Yvette, France) and maintained at 24°C until the hatching stage occurring around 24 ± 2 hours post-fertilization (hpf)[55]. Subsequently, larvae were transferred and raised in the Fish Physiology and Genomics laboratory experimental facilities (LPGP, INRAE, Rennes, France) under standard photoperiod (12 h light / 12 h dark) and at two different temperatures: 21 ± 1 °C and 28 ± 1°C. Animals were fed twice a day, firstly with live artemia (Ocean Nutrition) until 15 days post-fertilization (dpf), then with a commercial diet (BioMar) until adult stage. For animal dissections and organ sampling, fish were euthanized with a lethal dose of tricaine methanesulfonate (MS 222, 400 mg/l), supplemented by 150 mg/l of sodium bicarbonate. Phenotypic sex of individuals was determined at 4 months and more, either by macroscopical examination of the gonads when they were enough differentiated, or by histology when gonads were not totally differentiated[48]. Caudal fin clips were collected from all individuals and stored in ethanol 90% at 4°C before genomic DNA (gDNA) extraction. For the chromosome contact map (Hi-C), 80 µl of blood was sampled from three males using a syringe rinsed with EDTA 2%. The fresh blood was slowly frozen in a Freezing Container (Mr. Frosty, Nalgene®) after addition of 15% of dimethyl sulfoxide (DMSO). Karyotypic analyses were carried out in 17 males and 11 females of Pachón cave *Astyanax mexicanus*. Fin samples (a narrow strip of the tail fin) were taken from the live specimens anesthetized by MS-222 (Merck KGaA, Darmstadt, Germany), while for direct preparations (chromosomes from kidneys and gonads), fishes were euthanized first using 2-phenoxyethanol (Sigma-Aldrich, St. Louis, MO, USA).

All animal protocols were carried out in strict accordance with the French and European legislations (French decree 2013-118 and directive 2010-63-UE) applied for ethical use and care of laboratory animals used for scientific purposes. Sylvie Retaux and CNRS institutional authorizations for maintaining and handling *A. mexicanus* in experimental procedures were 91-116 and 91272105, respectively. For karyotypic analysis, all handling of fish individuals followed European standards in agreement with §17 of the Act No. 246/1992 to prevent fish suffering. The procedures involving fish were supervised by the Institutional Animal Care and Use Committee of the Institute of Animal Physiology and Genetics CAS, v.v.i., the supervisor´s permit number CZ 02361 certified and issued by the Ministry of Agriculture of the Czech Republic. Several sampling campaigns were carried out in the field in Mexico between 2013 and 2019, resulting in a collection of tail fin clips from wild-caught individuals from the Pachón cave. In the field, the phenotypic sex of animals was determined by checking the presence or absence of denticles on the anal fins as described previously[56], a small fin clip was gently taken and fish were rapidly returned to their natural pond. In addition, pictures of each individual sampled were taken to confirm the phenotypic sex, back in the laboratory, based on the morphological criteria described previously[56]. The permits for field sampling (02241/13, 02438/16, 05389/17 and 1893/19) were delivered by the Mexican authorities (Mexican Secretaría del Medio 16 Ambiente y Recursos Naturales) to Sylvie Rétaux and Patricia Ornelas-Garcia (UNAM, Mexico).

METHOD DETAILS

DNA extraction

For fish genotyping, gDNA was extracted from fin clips stored in 90% ethanol, after lysis in 5% chelex[57] and 10 mg Proteinase K at 55 °C for 2 h, followed by 10 min at 99 °C. Following extraction, samples were centrifuged and the supernatant containing the gDNA was transferred in clean tubes and stored at -20 °C. For pool-sequencing and TaqMan assay, gDNA was extracted using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) according to the supplier's recommendations. For long-read male genome sequencing, high molecular weight (HMW) gDNA was extracted from a mature testis grounded in liquid nitrogen and lysed in TNES-Urea buffer (TNES-Urea: 4M urea; 10 mM Tris-HCl, pH 7.5; 120 mM NaCl; 10 mM EDTA; 5% SDS) for two weeks at room temperature. For HMW gDNA extraction the TNES-Urea solution was supplemented with Proteinase K at a final concentration of 150 µg/ml and incubated at 37 °C overnight. HMW gDNA was extracted with a modified phenol-chloroform protocol as previously described[43]. The gDNA concentrations for both pool-seq and genome sequencing were quantified with Qubit3 fluorometer (Invitrogen, Carlsbad, CA) and HMW gDNA quality and purity were assessed using spectrophotometry, fluorometry, and capillary electrophoresis.

Primers and probe design

All primers used in this study including PCR genotyping, qPCR gene expression and cDNA cloning were designed using Primer3web software[58] version 4.1.0 and are listed in Data S1E.

B polymerase chain reaction (PCR) genotyping

Genetic sex of Pachón cave individuals was determined by PCR tests using the fact that males have a B-predominant chromosome with three different sets of primers (see Table S6 for primer sequences) based on differences between the *A-gdf6b* and *B-gdf6b* loci. Three sets of primers (P) were designed (see Figure S3A) with two primer sets designed to amplify specifically the two *B-gdf6b* copies based either on a single base variation between the A/B *gdf6b* CDS at position 679 bp (P1-P2), or based on primers located on both sides of the A/B breakpoints downstream of the *B-gdf6b* gene (P3-P4). The third set of primers (P5-P6) was designed on gaps/indels variations between *A-gdf6b* and the two *B-gdf6b* genes in the proximal promoter of *gdf6b* genes. Another set of primers (P9-P10) was designed as a PCR positive control with primers located on both sides of the A/B breakpoints downstream of the *A-gdf6b* gene. Primer sets P1-P2 and P3-P4 produce a single PCR fragment only in males (B+), primer set P5-P6 amplifies two bands in males (all B+) and only a single band in females (all B-), and primer set P9-P10 amplifies a single band in all individuals. For PCR reactions with the P1-P2 primer set, HiDi Taq DNA polymerase (myPOLS Biotec) was used for detecting a single nucleotide variation. PCRs were performed in a total volume of 12.5 µl containing 0.2 µM of each primer, a final concentration of 20 ng/µl gDNA, 200 µM of dNTPs mix, 1X of HiDi buffer (10X), and 2.5 U per reaction of HiDi DNA polymerase. Cycling conditions were as follows: 95 °C for 2 min, then 35 cycles of (95 °C for 15 seconds (secs) + 60 °C for 10 sec + 72 °C for 30 sec), and 72 °C for 5 min. For PCR reactions with the P3-P4 and P9-P10 primer sets, PCR reactions were performed in a final volume of 50 µl containing 0.5 µM of each primer, a final concentration of 20 ng/µl gDNA, 10 µM dNTPs mix, 1X of 10X AccuPrime™ buffer, and 0.5 µl per reaction of AccuPrime™ Taq DNA polymerase. Cycling conditions were as follows: 94 °C for 2 min, then 35 cycles of (94 °C for 30 sec + 58 °C for 30 sec + 68 °C for 1 min and 30 sec). For PCR reactions with the P5-P6 primer set, PCR reactions were performed in a final volume of 25 µl containing 0.5 µM of each primer, a final concentration of 20 ng/µl gDNA, 10 µM dNTPs mixture, 1X of Jumpstart™ buffer (10X), and 0.5 µl per reaction of Jumpstart™ Taq DNA polymerase. Cycling conditions were as follows: 95 °C for 2 min, then 35 cycles of (95 °C for 1 min + 60 °C for 30 sec + 72 °C for 1 min), and 72 °C for 5 min.

10× Genomics sequencing

10X Chromium Library was prepared according to 10X Genomics protocols using the Genome Reagent Kits v2. Optimal performance has been characterized on input gDNA with a mean length greater than 50 kb (~144Kb). GEM reactions were performed on 0.625 ng of genomic DNA, and DNA molecules were partitioned and amplified into droplets to introduce 16-bp partition barcodes. GEM reactions were thermally cycled (30 °C for 3 h and 65 °C for 10 min; held at 4 °C) and after amplification, the droplets were fractured. P5 and P7 primers, read 2, and sample index were added during library construction. The library was amplified using 10 cycles of PCR and the DNA was subsequently size selected to 450 bp by performing a double purification on AMPure Xp beads. Library quality was assessed using a Fragment Analyzer and quantified by qPCR using the Kapa Library Quantification Kit. Sequencing has been performed on an Illumina HiSeq3000 using a paired-end read length of 2x150 bp with the Illumina HiSeq3000 sequencing kits.

Oxford nanopore genome sequencing

High molecular weight gDNA purification steps were performed using AMPure XP beads (Beckman Coulter). Library preparation and sequencing were performed using Oxford Nanopore (Oxford Nanopore Technologies) Ligation Sequencing Kit SQK-LSK109 according to manufacturer's instructions "1D gDNA selecting for long reads (SQK-LSK109)". Five µg of DNA was purified then sheared at 20 kb using the megaruptor1 system (diagenode). A one-step DNA damage repair + END-repair + dA tail of double-stranded DNA fragments was performed on 2 µg of sample. Then adapters were ligated to the library. Library was loaded onto 1 R9.4.1 flowcell and sequenced on a PromethION instrument at 0.02 pM within 72 h.

PacBio HiFi genome sequencing

Library preparation and sequencing were performed according to the manufacturer's instructions "Procedure & Checklist Preparing HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0". Fifteen µg of DNA were purified and then sheared at 15 kb using the Megaruptor3 system (Diagenode). Using SMRTbell Express Template prep kit 2.0, a Single strand overhangs removal and then a DNA and END damage repair steps were performed on 10 µg of the sample. Then blunt hairpin adapters were ligated to the library. The library was treated with an exonuclease cocktail to digest unligated DNA fragments. A size selection step using a 12 kb cutoff was performed on the BluePippin Size Selection system (Sage Science) with "0.75% DF Marker S1 3-10 kb Improved Recovery" protocol. Using Binding kit 2.0 kit and sequencing kit 2.0, the primer V2 annealed and polymerase 2.0 bounded library was sequenced by diffusion loading onto 2 SMRTcells on Sequel2 instrument at 50 pM with a 2 h pre-extension and a 30 h movie.

Hi-C sequencing

Hi-C data was generated using the Arima-HiC kit (Ref. 510008), according to the manufacturer's protocols using 10 µl of blood as starting material, the Truseq DNA PCR-Free kit and Truseq DNA UD Indexes (Illumina, ref. 20015962, ref. 20020590), and the KAPA library Amplification kit (Roche, ref. KK2620). Hi-C library was sequenced in paired-end 2x150 bp mode on Novaseq6000 (Illumina), using half a lane of a SP flow cell (ref. 20027464).

Image analyses and base calling were performed using the Illumina NovaSeq Control Software and Real-Time Analysis component (v3.4.4). Demultiplexing was performed using Illumina's conversion software (bcl2fastq v2.20). The quality of the raw data and potential contaminants was assessed using FastQC (v0.14.0)[59] from the Babraham Institute and the Illumina software SAV (Sequencing Analysis Viewer).

Genome assembly

Pacbio HiFi reads were assembled with hifiasm[60] version 0.9 using standard parameters. The genome assembly fasta file was extracted from the principal gfa assembly graph file using an awk command line. This assembly was then scaffolded using Hi-C and 10X as a source of linking information. 10X reads were aligned using Long Ranger v2.1.1 (10x Genomics). Hi-C reads were aligned to the draft genome using Juicer[61] with default parameters. A candidate assembly was then generated with 3D de novo assembly (3D-DNA) pipeline[62] with the -r 0 and --polisher-input-size 100000 parameters. Finally, the candidate assembly was manually reviewed using the Juicebox assembly tools[63]. Due to the specific structure of the Pachón cave B chromosome, both Hi-C and 10X signals show some uncertainties in the order and orientation of the contigs. To improve the quality of the B chromosome assembly, ONT reads were then aligned to the final version of the genome using minimap2[64] v2.11 with -x map-ont parameter. Both reads spanning contig junctions and reads showing supplementary alignments linking contigs belonging to the B chromosome were analyzed to resolve these ambiguities.

Genome annotation

The cavefish whole genome assembly was annotated using a pipeline adapted from previous studies[65,66]. In brief, RepeatModeler, RepeatProteinMask, and RepeatMasker (open-4.0.7, http://www.repeatmasker.org/) were first used to scan the genome and mask out repeats. Then protein-coding genes were annotated by collecting gene evidence from homology alignment, RNA-seq mapping, and *ab initio* prediction. For homology alignment, 464,144 protein sequences collected from NCBI were aligned to the assembly using Genewise[67] and Exonerate respectively. RNA-seq data were mapped to the assembly in two independent parallel steps. First Hisat[68] was used to align RNA-seq reads and then StringTie[69] was used to predict the gene models; in the other step reads were first assembled into transcript sequences using Trinity[70] and then PASA[71] was used to map the transcripts to the assembly and model the gene structures. For ab initio prediction and final integrating, Augutus[72] was first trained using the high-quality gene models and then ran in a hint-guide model.

B chromosome annotation

First, repeats were identified and masked from the B chromosome using RepeatModeler, RepeatProteinMask, and RepeatMasker (open-4.0.7, http://www.repeatmasker.org/). To annotate protein-coding genes, we collected all protein sequences of *A. mexicanus* annotated by NCBI (Genome ID: 13073) and Ensembl (release-104), and aligned them onto the repeat-masked B sequence using GeneWise[67,73] and Exonerate respectively. For each query, the best hit was kept. To determine the best gene model when multiple ones compete for a splice site, we introduced RNA-seq data to evaluate the quality of these homology gene models. Hence RNA-seq data of *A. mexicanus* from the previous study[74] were aligned to B using HISAT[68] and parsed using StringTie[69]. RNA-score of each homology gene model was then calculated as the match-extend of splice sites to that of StringTie prediction. When multiple homology gene models compete for a splice site, those with lower RNA-score were discarded. In cases when some genes failed to be identified using homology alignment, we also implemented an *ab initio* gene prediction using Augustus[72], where all the homology and transcriptome evidence were used as hints. The predicted results were included into the final gene set if 1) the splice sites are not occupied and 2) the splice sites match 100% to that of StringTie predictions (RNA-score =100). To further evaluate the quality of the final gene set, we blasted their protein sequences to SWISSPROT (https://www.uniprot.org/) and NR (https://www.ncbi.nlm.nih.gov/), and took the alignment to the best hit to check how much of the query and subject was aligned, respectively (query coverage & subject coverage). Genes with query and subject coverage both >90% were considered as being of good quality.

To characterize the A chromosome content of the B chromosome, sequences of the B chromosome assembly were aligned to the sequences of the 25 A chromosomes with minimap2[64] (v2.11) and the best match of each contig fragment was retained. Overlapping matches were manually filtered considering match lengths and similarities (cigarline and edit distance) in order to build the best non-overlapping matching list. Karyoplots were then plotted using the R package karyoploteR[75] (https://bernatgel.github.io/karyoploter_tutorial/). The median, minimum, and maximum sizes of the 628 B best matches on A chromosomes were respectively 1,087 bp, 44 bp and 41,908 bp.

Male and female Pool-sequencing

DNA was collected from 91 phenotypic Pachón cave males and 81 phenotypic Pachón cave females and was pooled as male and female pools separately. Before pooling, the DNA concentration was normalized in order to obtain an equal amount of each individual genome in the final pool. Pool-sequencing libraries were prepared using the Illumina TruSeq Nano DNA HT Library Prep Kit (Illumina, San Diego, CA) according to the manufacturer's protocol. After the fragmentation of each gDNA pool (200 ng/pool) by sonication using an M220 Focused-ultrasonicator (COVARIS), the size selection was performed using SPB beads retaining fragments of 550 bp. Following the 3′ ends of blunt fragments mono-adenylation and the ligation to specific paired-end adaptors, the amplification of the construction was performed using Illumina-specific primers. Library quality was verified with a Fragment Analyzer (Advanced Analytical Technologies) and then quantified by qPCR using the Kapa Library Quantification Kit (Roche Diagnostics Corp, Indianapolis, IN). The enriched male and female pool libraries were then sequenced using a paired-end multiplexed sequencing mode on a NovaSeq S4 lane (Illumina, San Diego, CA), combining the two pools on the same lane and producing 2×150 bp with Illumina NovaSeq Reagent Kits according to the manufacturer's instructions. Sequencing produced 288 million paired reads and 267 million paired reads for the male and female pool libraries, respectively.

Pool-sequencing analysis

Characterization of genomic regions enriched for sex-biased signals between Pachón cave males and females, consisting of coverage and Single Nucleotide Variations (SNVs) was performed as described previously[38,43]. Pachón cave *A. mexicanus* paired-end reads from male and female pool-seq pools were mapped onto our own Pachón cave genome assembly using BWA mem version 0.7.17[76]. The resulting BAM files were sorted and the duplicate reads due to PCR amplification during library preparation were removed using Picard tools version 2.18.2 (http://broadinstitute.github.io/picard) with default parameters. Then, for each pool and each genomic position, a file containing the nucleotide composition was generated using samtools mpileup[77] version 1.8, and popoolation2[78] mpileup2sync version 1201. This file was then analyzed with custom software (PSASS version 2.0.0; doi: 10.5281/zenodo.2615936) to compute: (a) the position and density of sex-specific SNVs, defined as SNVs heterozygous in one sex but homozygous in the other, and (b) the average read depths for male and female pools along the genome to look for regions present in one sex but absent in the other (i.e., sex-specific insertions). All PSASS analyses were run with default parameters except for the range of frequency for a sex-linked SNV in the homogametic sex, --range-hom, that was set to 0.01 instead of 0.05, and the size of the sliding window, --window-size, that was set at 50,000 instead of 100,000.

Chromosome conventional cytogenetics

Mitotic or meiotic chromosome spreads were obtained either from regenerating caudal fin tissue as previously described[79], with slight modifications[80] and altered time of fin regeneration (one week), or by direct preparation from the cephalic kidney and gonads[81]. In the latter, the quality of chromosomal spreading was enhanced by a previously described dropping method[82]. Chromosomes were stained with 5% Giemsa solution (pH 6.8) (Merck, Darmstadt, Germany) for conventional cytogenetic analyses, or left unstained for other methods. For FISH, slides were dehydrated in an ethanol series (70%, 80% and 96%, 3 min each) and stored at -20 °C before analysis. Constitutive heterochromatin was visualized by C-banding[83], with chromosomes being counterstained by 4′,6-diamidino-2-phenolindole (DAPI), 1.5 µg/ml in antifade (Cambio, Cambridge, United Kingdom).

gdf6b probe synthesis for FISH mapping

gDNA was extracted using NucleoSpin Kits for Tissue (Macherey-Nagel, Duren, Germany) as described above. A *gdf6b* fragment comprising the two exons, the intron and 2,260 bp of the proximal promoter (with a total size of 4,368 bp) was amplified by PCR in a total volume of 50 µl. The mixture contained 0.5 µM of each primer, a final concentration of 20 ng/µl gDNA, 1X of 10X AccuPrime™ PCR Buffer II, 1U/reaction of AccuPrime™ Taq DNA Polymerase, High Fidelity (Thermofisher), was adjusted to 50 µl with autoclaved and distilled water. Cycling conditions were as follows: 94 °C for 45 sec, then 35 cycles of (94 °C for 15 sec + 64 °C for 30 sec + 68 °C for 5 min and 30 sec), and 68 °C for 5 min. The resulting PCR product was cloned into TOPO TA cloning Kit XL (Thermofisher) and after sequence verification it was purified using NucleoSpin plasmid DNA purification kit (Machery-Nagel, Düren, Germany) according to the supplier's indications. This Pachón cave *gdf6b* cloned DNA fragment was labeled by nick translation with Cy3-dUTP using Cy3 NT Labeling Kit (Jena Bioscience, Jena, Germany). The optimal fragment size of the probe (approx. 200-500 bp) was achieved after 30 min of incubation at 15 °C.

Chromosome microdissection and FISH mapping

Twelve copies of B chromosome from *A. mexicanus* male individual (male 10) and twelve copies encompassing two B chromosomes (per cell) from a male individual (male 9) were manually microdissected as previously described[84] under an inverted microscope (Zeiss Axiovert 135) using a sterile glass needle attached to a mechanical micromanipulator (Zeiss). The chromosomes were subsequently amplified by degenerate oligonucleotide primed-PCR

(DOP-PCR) following previously described protocols[85]. One µl of the resulting amplification product was used as a template DNA for a labeling DOP-PCR reaction, with Spectrum Orange-dUTP and Spectrum Green-dUTP, for Male 9 (2B) and Male 10 (1B), respectively (both Vysis, Downers Grove, USA). The amplification was done in 30 cycles, following previously described protocols[86]. Depending on the experimental scheme, the final probe mixture contained i) both painting probes (200 ng each) or ii) a single painting probe (200 ng) and a labeled 4,368 bp long fragment containing *gdf6b* gene and its promoter (300 ng; see below). To block the shared repetitive sequences, the probe also contained 4-5 µg of unlabelled competitive DNA prepared from female gDNA (on male preparations) or male gDNA (on female preparations). Male and female gDNAs were isolated from liver and spleen using MagAttract HMW DNA kit (Qiagen) and C_0t-1 DNA (i.e., fraction of gDNA enriched with highly and moderately repetitive sequences) was then generated from them according to previously described protocols[87]. The complete B probe mixture was dissolved in the final volume 20 µl (in case of two painting probes) or 14 µl (in case of one painting probe and a *gdf6b* gene probe) of hybridization mixture (50% formamide and 10% dextran sulfate in 2× SSC).

FISH and whole-chromosome painting

The FISH (Fluorescence in situ hybridization on chromosomes) experiments were done using a combination of two previously published protocols[80,88], with slight modifications. Briefly, the aging of slides took place overnight at 37 °C and then 60 min at 60 °C, followed by treatments with RNase A (200 µg/ml in 2× SSC, 60–90 min, 37 °C) (Sigma-Aldrich) and then pepsin (50 µg/ml in 10 mM HCl, 3 min, 37 °C). Subsequently, the slides were incubated in 1% formaldehyde in PBS (10 min) to stabilize the chromatin structure. Denaturation of chromosomes was done in 75% formamide in 2× SSC (pH 7.0) (Sigma-Aldrich) at 72 °C, for 3 min. The hybridization mixture was denatured for 8 min (86 °C) and then pre-hybridized at 37 °C for 45 min to outcompete the repetitive fraction. After application of the probe cocktail on the slide, the hybridization took place in a moist chamber at 37 °C for 72 h. Subsequently, non-specific hybridization was removed by post-hybridization washes: two times in 1× SSC (pH 7.0) (65 °C, 5 min each) and once in 4× SSC in 0.01% Tween 20 (42°C, 5 min). Slides were then washed in PBS (1 min), passed through an ethanol series, and mounted in antifade containing 1.5 µg/ml DAPI (Cambio, Cambridge, United Kingdom).

Synaptonemal complex immunostaining

Pachytene chromosome spreads from two Pachón cave *A. mexicanus* males (male 2 and 4) were prepared from testes following the protocol for *Danio rerio*[89,90] with some modifications. Briefly, dissected testes were suspended in 200-600 µl (based on the testes size and cell density) of cold PBS. Cell suspensions were applied onto poly-l-lysine slides (ThermoFisher), with 1:30 (v/v) dilution in hypotonic solution (PBS: H_2O, 1:2 v/v). After 20 min at room temperature (RT), slides were fixed with freshly prepared cold 2% formaldehyde (pH 8.0 – 8.5) for 3 min at RT. Slides were then washed three times in 0.1% Tween-20 (pH 8.0-8.5), 1 min each, and left to dry (1 h). Afterwards, immunofluorescence analysis of synaptonemal complexes took place, using antibodies against the proteins SYCP3 (lateral elements of synaptonemal complexes) and MLH1 (mismatch repair protein; marker for visualization of recombination sites). The primary antibodies – rabbit anti-SYCP3 (1:300; Abcam, Cambridge, UK) and mouse anti-MLH1 (1:50, Abcam) – were diluted (v/v) in 3% BSA (bovine serum albumin) in 0.05% Triton X-100/ PBS. After application onto the slides, the incubation was carried out overnight in a humid chamber at 37°C. Next day, slides were washed three times in 0.1% Tween-20 in PBS, 10 min each and secondary antibodies, diluted (v/v) in 3% BSA (bovine serum albumin) in 0.05% Triton X-100/ PBS, were applied. Specifically, we used goat anti-rabbit Alexa 488 (1:300; Abcam) and goat anti-mouse Alexa555 (1:100; Abcam), and the slides were incubated for 3 h at 37 °C. Then, after washing in 0,1% Tween-20 in PBS (10 min) and brief washing in 0.01% Tween-20 in distilled H_2O, slides were mounted in antifade containing DAPI, as described above.

Microscopy and image analysis

At least 50 metaphase spreads per individual were analyzed to confirm the diploid chromosome number (2n), karyotype structure, and FISH results. Giemsa-stained preparations were analyzed under Axio Imager Z2 microscope (Zeiss, Oberkochen, Germany), equipped with an automatic Metafer-MSearch scanning platform. Photographs of the chromosomes were captured under 100× objective using CoolCube 1 b/w digital camera (MetaSystems, Altlussheim, Germany). The karyotypes were arranged using Ikaros software (MetaSystems, Altlussheim, Germany). Chromosomes were classified according to their centromere positions[91], modified as metacentric (m), submetacentric (sm), subtelocentric (st), or acrocentric (a). FISH preparations were inspected using an Olympus BX53 epifluorescence microscope (Olympus, Tokyo, Japan), equipped with an appropriate fluorescence filter set. Black-and-white images were captured under 100× objective for each fluorescent dye with a cooled DP30BW CCD camera (Olympus) using Olympus Acquisition Software. The digital images were then pseudo-colored (blue for DAPI, red for Cy3, green for FITC) and merged in DP Manager (Olympus). Composed images were then optimized and arranged using Adobe Photoshop CS6.

Phylogeny and synteny of Gdf6 proteins

Phylogeny and synteny relationships of *Gdf6* genes were inferred from a Genomicus instance[92] in which synteny and phylogeny have been reconciled with the Scorpios pipeline[93] on 4 tetrapod species, one sarcopterygii species, one holostei species and 13 teleosts (see Figure 3 for species names).

Three-dimensional protein modelling

Three-dimensional models for the proprotein of *A. mexicanus* A-Gdf6b and B-Gdf6b were obtained by homology modeling. The amino acid sequences of full-length A-Gdf6b and B-Gdf6b were submitted to automated homology modeling using the hm_build macro of the software package YASARA[94]. Modeling includes alignment of the two target sequences against sequences from Uniprot *via* PSI-Blast to build a position-specific scoring function matrix/profile, which is then used to search the PDB structure data bank for suitable modeling targets. Five templates were identified with sufficiently high scores, i.e. PDB entries 4YCG (proprotein complex of GDF2 (alternative naming BMP9))[95], 5HLY (proprotein complex of Activin A)[96], 6Z3J (the mature C-terminal growth factor domain of GDF5 in complex with repulsive guidance molecule B)[97] as well as 6Z3M (which is like 6Z3J but includes neogenin in complex with GDF5 and the repulsive guidance molecule B)[97], and 3QB4 (which is only the C-terminal growth factor domain of GDF5)[98]. Several initial models were built on the basis of these template structures, missing sequence elements were modeled in an automated procedure through YASARA on the basis of an indexed PDB structure database. By this scheme 13 models were built, three only covered the C-terminal growth factor domain comprising residues 293 to 398, while 10 models covered the proprotein complex consisting of residues 3 to 398 (models on the basis of the 5HLY entry) and of residues 66 to 398 (models on the basis of template 4YCG). These models were individually refined by a short molecular dynamics simulation in explicit water to optimize hydrogen bonding and protein packing. Due to the overall low Z-score of the individual models, YASARA used the various models to form a hybrid homology model combining the elements with the highest-scoring factors into a single 3D model. This hybrid model was then used for further analysis.

Expression analysis by Real- Time PCR

For gene expression studies, mRNA transcripts levels were quantified during gonadal development from 10 dpf to 90 dpf, male and female gametogenesis stages, and finally in 10 adult organs including gonads as described previously[48]. All samples were frozen in liquid nitrogen and stored at -80 °C until RNA extraction. Total RNA extraction from gonads, trunks, and adult organs, followed by cDNA synthesis, and expression analysis by RT-PCR were carried out as previously described[48]. Specific primers were designed for *gdf6b* in the most divergent sequence regions between the two paralogous *gdf6a* and *gdf6b* genes.

RNAScope in situ hybridization of gdf6b

RNA *in situ* hybridization (ISH) assays have been carried out using the RNAscope® technology (ACD Biotechne™) on 7 µm cross-sections of 15, 21, 30 and 60 dpf Pachon cavefish fixed in paraformaldehyde 4% overnight at 4 °C and embedded in paraffin after serial dehydration in increasing methanol solutions. Specific probes for Pachon cavefish *gdf6b* and *gsdf* were synthesized by ACD Biotechne™. Sections were collected on Super frost+ slides, heated at 60 °C for 1 h and dewaxed 2 x 5 min in xylene followed by 2x 2 min in Ethanol 100% at RT. Fluorescent ISH was carried out with the Multiplex Fluorescent Reagent Kit v2 (ACD Biotechne™, ref: 323100) according to manufacturer's protocol. Following hybridization, nuclei were labelled with DAPI (4′,6-diamidino-2-phenylindole) staining and slides were mounted with ProLong™ Gold Antifade Mountant (Invitrogen™) and observed with a Leica TCS SP8 laser scanning confocal microscope.

Knockout of Pachón A. mexicanus gdf6b

Pachón cave *A. mexicanus* inactivated for *gdf6b* were generated using the CRISPR/Cas9 method. Guide RNAs (sgRNAs) targeting two sites located in exon 2 of *gdf6b* were designed using ZiFiT software[99] (http://zifit.partners.org/ZiFiT/Disclaimer.aspx). DR274 vector (Addgene #42250) containing the guide RNA universal sequence was first linearized with Bsa1, electrophoresed in a 2% agarose gel and purified. PCR amplifications were then performed using linearized DR274 as a template and two primers for each sgRNA. Forward primers containing sgRNAs target sequences (#site 1 and #site 2) (bolded and underlined) between the T7 promoter sequence in the 5' end and the conserved tracrRNA domain sequence were as follows. Forward primer (#site 1): 5'-GAAATTAATACGACTCACTATA**GGAGTCTGAAACCGGTTCTG**GTTTTAGAGCTAGAAATAGCAAG-3'. Forward primer (#site 2): 5'-GAAATTAATACGACTCACTATA**GGGAGCTGGGCTGGGACGAC**GTTTTAGAGCT

AGAAATAGCAAG-3’. Universal Reverse primer: 5'-AAAAGCACCGACTCGGTGCCACT-3'. Subsequently, residual plasmid was digested with Dpn1 (renewed once) at 37°C for 3 hours. The final product was purified and used as a DNA template for transcription. The sgRNAs were transcribed using the MAXIscript™ T7 Transcription Kit (Ambion) according to the manufacturer’s instructions. The sgRNAs were precipitated in 200 µl of isopropanol solution at -20°C, centrifuged and the supernatant was removed. The precipitated sgRNAs were resuspended in RNase-free water. The sgRNAs were co-injected with Cas9 protein. Synthesized RNAs were then injected into 1-cell stage *A. mexicanus* Pachón cave embryos at the following concentrations: 72 ng/µl for each sgRNA and 216 ng/µl for the Cas9 protein (kindly provided by Tacgene, MNHN, Paris). Genotyping was performed on gDNA from caudal fin-clips of adult fishes. CRISPR-positive fish were screened for mutations using a set of PCR primers (P7-P8) (Figure S3A, Data S1E) flanking the sgRNAs target sites leading to a ~470 bp deletion on the exon 2 of the *gdf6b* genes (Figure S10). The genetic sex of the mutants was determined by specific primers (P5-P6) on the *gdf6b* promoter with gap/indel variation between *A-gdf6b* and *B-gdf6b* (see STAR methods above and Figure S3A).

Histology

Gonads were fixed in Bouin’s fixative solution for 48 h and then dehydrated serially in aqueous 70% and 95% ethanol, ethanol/butanol (5∶95), and butanol. Tissues were embedded in paraffin blocks that were cutted serially into 5 µm sections, and were stained with hematoxylin-eosin-safran (HES) (Microm Microtech, Brignais, France).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses

For the sex genotyping marker based on the heterozygous and specific site of the B chromosome on the exon 2 (position 679 bp of the *B-gdf6b* CDS), the significance of the correlation between this polymorphism and the male phenotypic sex was tested with the Pearson's Chi-squared test with Yates' continuity correction. For gene expression, normality of data residuals, homogeneity of variances and homoscedasticity were verified before performing parametric or non-parametric tests. Consequently, statistical analyses were carried out only with non-parametric tests using RStudio (Open Source version) considering the level of significance at $p<0.05$. For comparisons between two groups, we used Wilcoxon signed rank test. All data are shown as Mean ± Standard Error of the Mean (SEM).

LEGEND OF SUPPLEMENTARY DATAFILE

Data S1: Supplementary information on metaphase numbers (A), genome assembly metrics and annotation (B, C, D), primer names (E) and sex linkage of B-gdf6b (F) in Pachón cavefish. Related to STAR Methods and Figures 1 and 2. A) Description of Data S1A. Number of metaphases (NM) containing B chromosomes (Bs) in males and females of Pachón cave *A. mexicanus*. % M = percentage of metaphases. Only very few male metaphases did not have a B. In contrast, Bs were not detected in most females (7/11), and when present in females (4/11), they were detected only in very few metaphases and most often as a unique B copy. **B) Description of Data S1B.** Genome assembly metrics of the Pachón cavefish *A. mexicanus* male assembly (HiFi PacBio) and comparison with previous publicly available surface (Astyanax_mexicanus-2.0) and Pachón cave (Astyanax_mexicanus-1.02) genome assemblies. **C) Description of Data S1C.** Gene annotation of Pachón cavefish B chromosome. **D) Description of Data S1D.** Transposable elements in the Pachón cavefish genome. **E) Description of Data S1E.** Primer names, sequences, target genes, and their corresponding experiments. **F) Description of Data S1F.** Association between *B-gdf6b* specific amplifications and sex phenotypes with different *B-gdf6b* primer sets in a laboratory stock and in wild-caught Pachón cavefish. P-value of *B-gdf6b* association with sex is based on the Pearson's Chi-squared test with Yates' continuity correction.

REFERENCES

1. Wright, A.E., Dean, R., Zimmer, F., and Mank, J.E. (2016). How to make a sex chromosome. Nature Communications *7*, 12087.
2. Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biology and Evolution *12*, 750–763.
3. D'Ambrosio, U., Alonso-Lifante, M.P., Barros, K., Kovařík, A., Xaxars, G.M. de, and Garcia, S. (2017). B-chrom: a database on B-chromosomes of plants, animals and fungi. New Phytologist *216*, 635–642.
4. Jones, N. (2017). New species with B chromosomes discovered since 1980. Nucleus *60*, 263–281.
5. Camacho, J.P.M. (2005). CHAPTER 4 - B Chromosomes. In The Evolution of the Genome, T. R. Gregory, ed. (Academic Press), pp. 223–286.
6. Camacho, J.P.M., Sharbel, T.F., and Beukeboom, L.W. (2000). B-chromosome evolution. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences *355*, 163–178.
7. Beladjal, L., Vandekerckhove, T.T.M., Muyssen, B., Heyrman, J., de Caesemaeker, J., and Mertens, J. (2002). B-chromosomes and male-biased sex ratio with paternal inheritance in the fairy

shrimp Branchipus schaefferi (Crustacea, Anostraca). Heredity (Edinb) *88*, 356–360.

8. Favarato, R.M., Ribeiro, L.B., Ota, R.P., Nakayama, C.M., and Feldberg, E. (2019). Cytogenetic Characterization of Two Metynnis Species (Characiformes, Serrasalmidae) Reveals B Chromosomes Restricted to the Females. CGR *158*, 38–45.

9. Néo, D.M., Filho, O.M., and Camacho, J.P.M. (2000). Altitudinal variation for B chromosome frequency in the characid fish Astyanax scabripinnis. Heredity *85*, 136–141.

10. Vicente, V.E., Moreira-Filho, O., and Camacho, J.P. (1996). Sex-ratio distortion associated with the presence of a B chromosome in Astyanax scabripinnis (Teleostei, Characidae). Cytogenet Cell Genet *74*, 70–75.

11. Stange, E. a. R., and Almeida-toledo, L.F. (1993). Supernumerary b chromosomes restricted to males in astyanax scabripinnis (pisces , characidae). Revista brasileira de genetica *16*, 601–615.

12. Clark, F.E., Conte, M.A., Ferreira-Bravo, I.A., Poletto, A.B., Martins, C., and Kocher, T.D. (2017). Dynamic Sequence Evolution of a Sex-Associated B Chromosome in Lake Malawi Cichlid Fish. Journal of Heredity *108*, 53–62.

13. Clark, F.E., and Kocher, T.D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. Scientific Reports *9*, 20213.

14. Yoshida, K., Terai, Y., Mizoiri, S., Aibara, M., Nishihara, H., Watanabe, M., Kuroiwa, A., Hirai, H., Hirai, Y., Matsuda, Y., et al. (2011). B Chromosomes Have a Functional Effect on Female Sex Determination in Lake Victoria Cichlid Fishes. PLOS Genetics *7*, e1002203.

15. Camacho, J.P.M., Schmid, M., and Cabrero, J. (2011). B Chromosomes and Sex in Animals. SXD *5*, 155–166.

16. Hanlon, S.L., and Hawley, R.S. (2018). B Chromosomes in the Drosophila Genus. Genes (Basel) *9*.

17. Martis, M.M., Klemme, S., Banaei-Moghaddam, A.M., Blattner, F.R., Macas, J., Schmutzer, T., Scholz, U., Gundlach, H., Wicker, T., Šimková, H., et al. (2012). Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. PNAS *109*, 13343–13346.

18. Valente, G.T., Conte, M.A., Fantinatti, B.E.A., Cabral-de-Mello, D.C., Carvalho, R.F., Vicari, M.R., Kocher, T.D., and Martins, C. (2014). Origin and evolution of B chromosomes in the cichlid fish Astatotilapia latifasciata based on integrated genomic analyses. Mol Biol Evol *31*, 2061–2072.

19. Ahmad, S.F., Jehangir, M., Cardoso, A.L., Wolf, I.R., Margarido, V.P., Cabral-de-Mello, D.C., O'Neill, R., Valente, G.T., and Martins, C. (2020). B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. BMC Genomics *21*, 656.

20. Pansonato-Alves, J.C., Serrano, É.A., Utsunomia, R., Camacho, J.P.M., da Costa Silva, G.J., Vicari, M.R., Artoni, R.F., Oliveira, C., and Foresti, F. (2014). Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus Characidium. PLoS One *9*, e107169.

21. Serrano-Freitas, É.A., Silva, D.M.Z.A., Ruiz-Ruano, F.J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., Camacho, J.P.M., and Foresti, F. (2020). Satellite DNA content of B chromosomes in the characid fish Characidium gomesi supports their origin from sex chromosomes. Mol Genet Genomics *295*, 195–207.

22. Conte, M.A., Clark, F.E., Roberts, R.B., Xu, L., Tao, W., Zhou, Q., Wang, D., and Kocher, T.D. (2021). Origin of a Giant Sex Chromosome. Molecular Biology and Evolution *38*, 1554–1569.

23. Zhou, Q., Zhu, H., Huang, Q., Zhao, L., Zhang, G., Roy, S.W., Vicoso, B., Xuan, Z., Ruan, J., Zhang, Y., et al. (2012). Deciphering neo-sex and B chromosome evolution by the draft genome of Drosophila albomicans. BMC Genomics *13*, 109.

24. Piscor, D., and Parise-Maltempi, P.P. (2016). Microsatellite Organization in the B Chromosome and A Chromosome Complement in Astyanax (Characiformes, Characidae) Species. Cytogenet

Genome Res *148*, 44–51.

25. Kavalco, K.F., and De Almeida-Toledo, L.F. (2007). Molecular Cytogenetics of Blind Mexican Tetra and Comments on the Karyotypic Characteristics of Genus Astyanax (Teleostei, Characidae). Zebrafish *4*, 103–111.

26. Ebrahimzadegan, R., Houben, A., and Mirzaghaderi, G. (2019). Repetitive DNA landscape in essential A and supernumerary B chromosomes of Festuca pratensis Huds. Scientific Reports *9*, 19989.

27. Mizoguchi, S.M.H.N., and Martins-Santos, I.C. (1997). Macro- and Microchromosomes B in Females of Astyanax scabripinnis (Pisces, Characidae). Hereditas *127*, 249–253.

28. Portela-Castro, A.L. de B., Júnior, H.F.J., and Nishiyama, P.B. (2000). New occurrence of microchromosomes B in Moenkhausia sanctaefilomenae (Pisces, Characidae) from the Paraná River of Brazil: analysis of the synaptonemal complex. Genetica *110*, 277–283.

29. Houben, A. (2017). B Chromosomes – A Matter of Chromosome Drive. Front. Plant Sci. *8*.

30. Jones, R.N., González-Sánchez, M., González-García, M., Vega, J.M., and Puertas, M.J. (2008). Chromosomes with a life of their own. Cytogenet Genome Res *120*, 265–280.

31. McGaugh, S.E., Gross, J.B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, W.R., Keene, A., Ma, L., et al. (2014). The cavefish genome reveals candidate genes for eye loss. Nat Commun *5*, 5307.

32. Ahmad, S.F., and Martins, C. (2019). The Modern View of B Chromosomes Under the Impact of High Scale Omics Analyses. Cells *8*.

33. Blavet, N., Yang, H., Su, H., Solanský, P., Douglas, R.N., Karafiátová, M., Šimková, L., Zhang, J., Liu, Y., Hou, J., et al. (2021). Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. Proc Natl Acad Sci U S A *118*, e2104254118.

34. Navarro-Domínguez, B., Ruiz-Ruano, F.J., Cabrero, J., Corral, J.M., López-León, M.D., Sharbel, T.F., and Camacho, J.P.M. (2017). Protein-coding genes in B chromosomes of the grasshopper Eyprepocnemis plorans. Sci Rep *7*, 45200.

35. Ruiz-Ruano, F.J., Cabrero, J., López-León, M.D., Sánchez, A., and Camacho, J.P.M. (2018). Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust. Chromosoma *127*, 45–57.

36. Christoffels, A., Koh, E.G.L., Chia, J., Brenner, S., Aparicio, S., and Venkatesh, B. (2004). Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. Molecular Biology and Evolution *21*, 1146–1151.

37. Vandepoele, K., De Vos, W., Taylor, J.S., Meyer, A., and Van de Peer, Y. (2004). Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A *101*, 1638–1643.

38. Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Eché, C., Valière, S., Donnadieu, C., Haffray, P., et al. (2020). Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, Perca flavescens. Mol Ecol Resour *20*, 531–543.

39. Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu). PLoS Genet. *8*, e1002798.

40. Rafati, N., Chen, J., Herpin, A., Pettersson, M.E., Han, F., Feng, C., Wallerman, O., Rubin, C.-J., Péron, S., Cocco, A., et al. (2020). Reconstruction of the birth of a male sex chromosome present in Atlantic herring. PNAS *117*, 24359–24368.

41. Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., et al. (2015). A Tandem Duplicate of Anti-Müllerian Hormone with a Missense SNP on the Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, Oreochromis niloticus. PLOS Genetics *11*, e1005678.

42. Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strüssmann, C.A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. Proc. Natl. Acad. Sci. U.S.A. *109*, 2955–2959.

43. Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., et al. (2019). Identification of the master sex determining gene in Northern pike (Esox lucius) reveals restricted sex chromosome differentiation. PLOS Genetics *15*, e1008013.

44. Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., et al. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. Cell *163*, 1527–1538.

45. Gautier, A., Sohm, F., Joly, J.-S., Le Gac, F., and Lareyre, J.-J. (2011). The Proximal Promoter Region of the Zebrafish gsdf Gene Is Sufficient to Mimic the Spatio-Temporal Expression Pattern of the Endogenous Gene in Sertoli and Granulosa Cells1. Biology of Reproduction *85*, 1240–1251.

46. Nakamura, S., Kobayashi, D., Aoki, Y., Yokoi, H., Ebe, Y., Wittbrodt, J., and Tanaka, M. (2006). Identification and lineage tracing of two populations of somatic gonadal precursors in medaka embryos. Dev. Biol. *295*, 678–688.

47. Sawatari, E., Shikina, S., Takeuchi, T., and Yoshizaki, G. (2007). A novel transforming growth factor-β superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (Oncorhynchus mykiss). Developmental Biology *301*, 266–275.

48. Imarazene, B., Beille, S., Jouanno, E., Branthonne, A., Thermes, V., Thomas, M., Herpin, A., Rétaux, S., and Guiguen, Y. (2021). Primordial Germ Cell Migration and Histological and Molecular Characterization of Gonadal Differentiation in Pachón Cavefish Astyanax mexicanus. Sex Dev, 1–18.

49. Benetta, E.D., Antoshechkin, I., Yang, T., Nguyen, H.Q.M., Ferree, P.M., and Akbari, O.S. (2020). Genome elimination mediated by gene expression from a selfish chromosome. Science Advances *6*, eaaz9808.

50. Nur, U., Werren, J.H., Eickbush, D.G., Burke, W.D., and Eickbush, T.H. (1988). A “selfish” B chromosome that enhances its transmission by eliminating the paternal genome. Science *240*, 512–514.

51. Jones, R.N. (2018). Transmission and Drive Involving Parasitic B Chromosomes. Genes (Basel) *9*, E388.

52. Bernardino, A.C.S., Cabral-de-Mello, D.C., Machado, C.B., Palacios-Gimenez, O.M., Santos, N., and Loreto, V. (2017). B Chromosome Variants of the Grasshopper Xyleus discoideus angulatus Are Potentially Derived from Pericentromeric DNA. Cytogenet Genome Res *152*, 213–221.

53. Stevens, J.P., and Bougourd, S.M. (1994). Unstable B-chromosomes in a European population of Allium schoenoprasum L. (Liliaceae). Biological Journal of the Linnean Society *52*, 357–363.

54. Ruban, A., Schmutzer, T., Wu, D.D., Fuchs, J., Boudichevskaia, A., Rubtsova, M., Pistrick, K., Melzer, M., Himmelbach, A., Schubert, V., et al. (2020). Supernumerary B chromosomes of Aegilops speltoides undergo precise elimination in roots early in embryo development. Nature Communications *11*, 2764.

55. Hinaux, H., Pottin, K., Chalhoub, H., Père, S., Elipot, Y., Legendre, L., and Rétaux, S. (2011). A developmental staging table for Astyanax mexicanus surface fish and Pachón cavefish. Zebrafish *8*, 155–165.

56. Elipot, Y., Legendre, L., Père, S., Sohm, F., and Rétaux, S. (2014). Astyanax transgenesis and husbandry: how cavefish enters the laboratory. Zebrafish *11*, 291–299.

57. Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Høyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., et al. (2006). A linkage map for brown trout (Salmo trutta): chromosome homeologies and comparative genome organization with other salmonid fish. Genetics

172, 2405–2419.

58. Kõressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., and Remm, M. (2018). Primer3_masker: integrating masking of template sequence with primer design software. Bioinformatics *34*, 1937–1938.

59. Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res *7*.

60. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods *18*, 170–175.

61. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst *3*, 95–98.

62. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95.

63. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. cels *3*, 99–101.

64. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

65. Powell, D.L., García-Olazábal, M., Keegan, M., Reilly, P., Du, K., Díaz-Loyo, A.P., Banerjee, S., Blakkan, D., Reich, D., Andolfatto, P., et al. (2020). Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. Science *368*, 731–736.

66. Du, K., Stöck, M., Kneitz, S., Klopp, C., Woltering, J.M., Adolfi, M.C., Feron, R., Prokopov, D., Makunin, A., Kichigin, I., et al. (2020). The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. Nat Ecol Evol *4*, 841–852.

67. Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988–995.

68. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357–360.

69. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290–295.

70. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644–652.

71. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res *31*, 5654–5666.

72. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res *34*, W435-439.

73. She, R., Chu, J.S.-C., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res *19*, 143–149.

74. Pasquier, J., Cabau, C., Nguyen, T., Jouanno, E., Severac, D., Braasch, I., Journot, L., Pontarotti, P., Klopp, C., Postlethwait, J.H., et al. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics *17*, 368.

75. Gel, B., and Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics *33*, 3088–3090.

76. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-

MEM.

77. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

78. Kofler, R., Pandey, R.V., and Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics *27*, 3435–3436.

79. Völker, M., and Ràb, P. (2015). Direct Chromosome Preparations from Embryos and Larvae. In Fish Cytogenetic Techniques (CRC Press), p. 7.

80. Sember, A., Bohlen, J., Šlechtová, V., Altmanová, M., Symonová, R., and Ráb, P. (2015). Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. BMC Evol Biol *15*, 251.

81. Ràb, P., and Roth, P. (1988). Cold-blooded vertebrates. In Methods of Chromosome Analysis (Czechoslovak Biological Society Publishers), pp. 115–124.

82. Bertollo, L., Cioffi, M., and Moreira-Filho, O. (2015). Direct Chromosome Preparation from Freshwater Teleost Fishes (CRC Press).

83. Haaf, T., and Schmid, M. (1984). An early stage of ZW/ZZ sex chromosome differentiation in Poecilia sphenops var. melanistica (Poeciliidae, Cyprinodontiformes). Chromosoma *89*, 37–41.

84. Ahmed, A.-R., Leon, B.L., and Thomas, L. (2020). Glass needle-based chromosome microdissection—How to set up probes for molecular cytogenetics? Video Journal of Clinical Research *2*, 1.

85. Yang, F., Trifonov, V., Ng, B.L., Kosyakova, N., and Carter, N.P. (2009). Generation of Paint Probes by Flow-Sorted and Microdissected Chromosomes. In Fluorescence In Situ Hybridization (FISH) — Application Guide Springer Protocols Handbooks., T. Liehr, ed. (Springer), pp. 35–52.

86. Yang, F., and Graphodatsky, A.S. (2009). Animal Probes and ZOO-FISH. In Fluorescence In Situ Hybridization (FISH) — Application Guide Springer Protocols Handbooks., T. Liehr, ed. (Springer), pp. 323–346.

87. Zwick, M.S., Hanson, R.E., Islam-Faridi, M.N., Stelly, D.M., Wing, R.A., Price, H.J., and McKnight, T.D. (1997). A rapid procedure for the isolation of C0t-1 DNA from plants. Genome *40*, 138–142.

88. Yano, C.F., Bertollo, L. a. C., Ezaz, T., Trifonov, V., Sember, A., Liehr, T., and Cioffi, M.B. (2017). Highly conserved Z and molecularly diverged W chromosomes in the fish genus Triportheus (Characiformes, Triportheidae). Heredity (Edinb) *118*, 276–283.

89. Kochakpour, N. (2009). Immunofluorescent microscopic study of meiosis in zebrafish. Methods Mol Biol *558*, 251–260.

90. Kochakpour, N., and Moens, P.B. (2008). Sex-specific crossover patterns in Zebrafish (Danio rerio). Heredity *100*, 489–495.

91. Levan, A., Fredga, K., and Sandberg, A.A. (1964). Nomenclature for Centromeric Position on Chromosomes. Hereditas *52*, 201–220.

92. Louis, A., Muffato, M., and Roest Crollius, H. (2013). Genomicus: five genome browsers for comparative genomics in eukaryota. Nucleic Acids Res. *41*, D700-705.

93. Parey, E., Louis, A., Cabau, C., Guiguen, Y., Crollius, H.R., and Berthelot, C. (2020). Synteny-guided resolution of gene trees clarifies the functional impact of whole genome duplications. Mol. Biol. Evol.

94. Krieger, E., and Vriend, G. (2014). YASARA View—molecular graphics for all devices—from smartphones to workstations. Bioinformatics *30*, 2981–2982.

95. Mi, L.-Z., Brown, C.T., Gao, Y., Tian, Y., Le, V.Q., Walz, T., and Springer, T.A. (2015).

Structure of bone morphogenetic protein 9 procomplex. Proc Natl Acad Sci U S A *112*, 3710–3715.

96. Wang, X., Fischer, G., and Hyvönen, M. (2016). Structure and activation of pro-activin A. Nat Commun *7*, 12052.

97. Malinauskas, T., Peer, T.V., Bishop, B., Mueller, T.D., and Siebold, C. (2020). Repulsive guidance molecules lock growth differentiation factor 5 in an inhibitory complex. PNAS *117*, 15620–15631.

98. Klammert, U., Mueller, T.D., Hellmann, T.V., Wuerzler, K.K., Kotzsch, A., Schliermann, A., Schmitz, W., Kuebler, A.C., Sebald, W., and Nickel, J. (2015). GDF-5 can act as a context-dependent BMP-2 antagonist. BMC Biol *13*, 77.

99. Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., and Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res *38*, W462-468.

100. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol *31*, 227–229.

101. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Supplementary figures

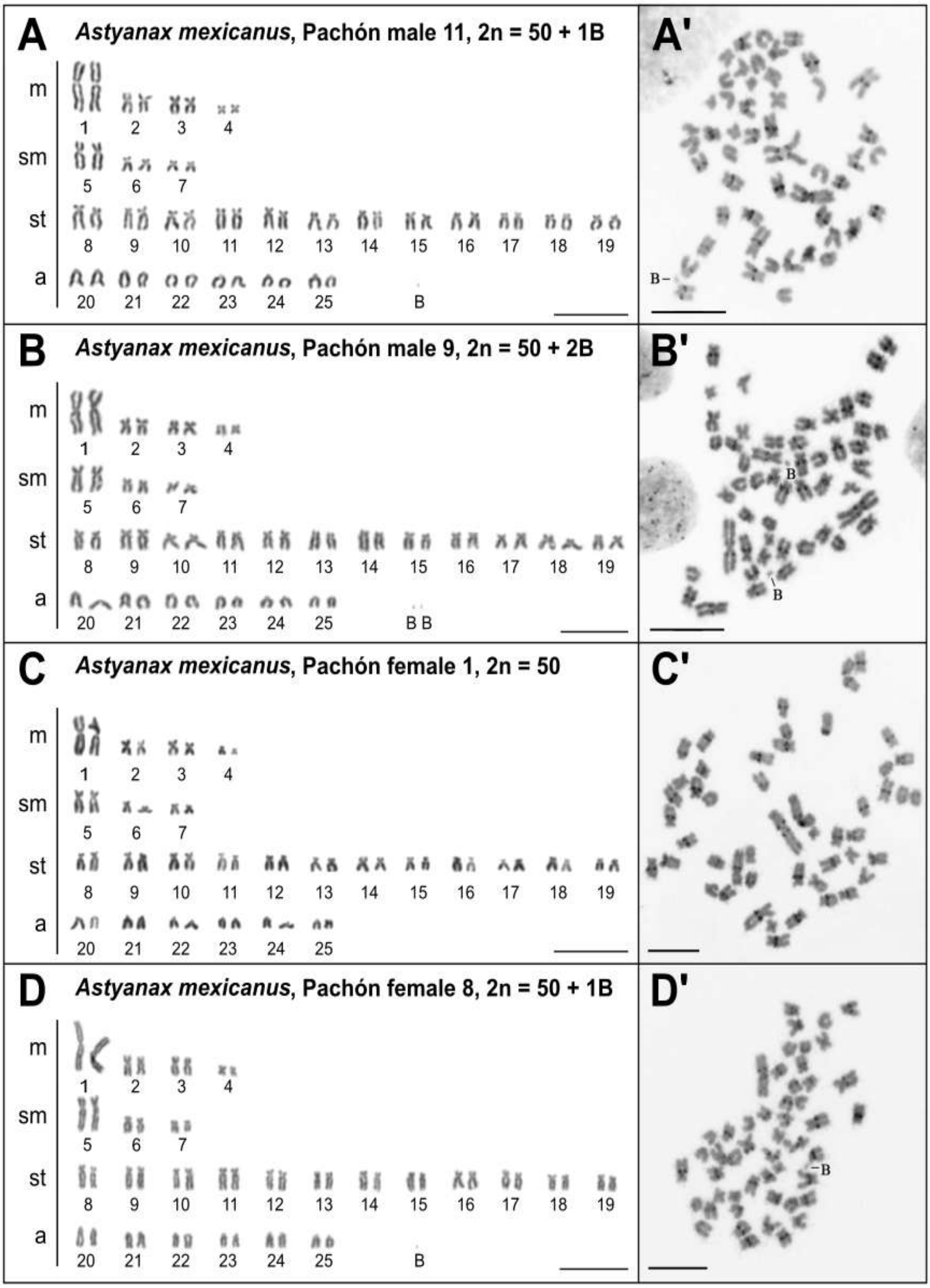

Figure S1. Karyotypes and corresponding C-banded mitotic metaphases of Pachón cave *Astyanax mexicanus*, with different male and female B chromosome (B) constitution. Related to Figure 1. Representative male and female Pachón cave karyotypes arranged from Giemsa-stained mitotic chromosomes (panels **A-D**) and their corresponding C-banding patterns (panels **A'-D'**). B numbers were found to be variable among individuals (from 0 to 3 Bs) with all males having a single (panels **A-A'**) or multiple Bs (panels **B-B'**) in most of their metaphases, most females having no B (panels **C-C'**), and only a few females having rare B positive metaphases (panels **D-D'**) Notice also the lack of C-bands on Bs suggesting that these Pachón cavefish male-predominant Bs are largely euchromatic. Scale bar = 10 µm. Male and female numbers referred to Data S1A.

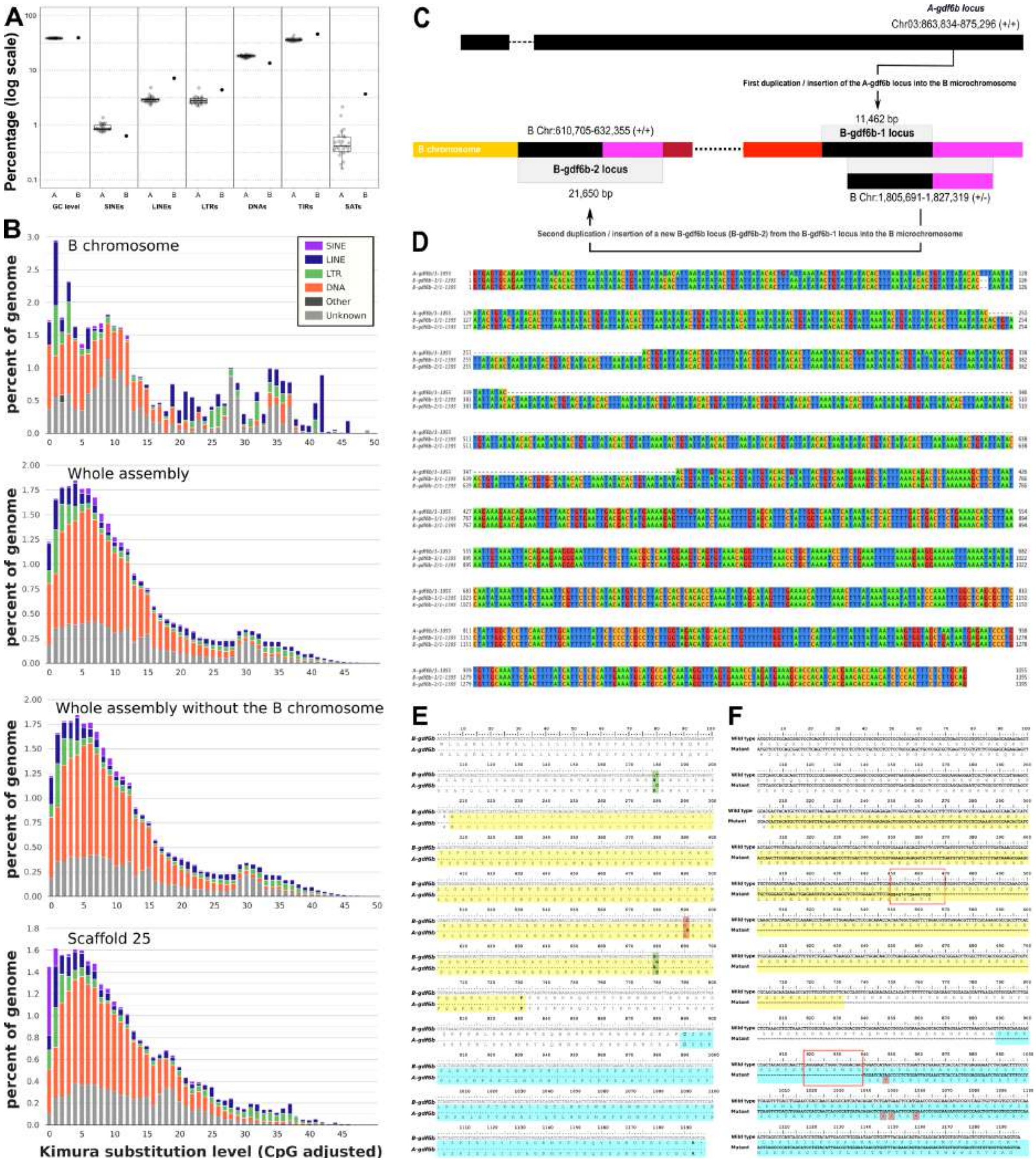

Figure S2: Genomic repeated elements, *gdf6b* evolution and *gdf6b* sequence alignments. Related to Figure 2. A. Comparison of the GC content and repeated elements in Pachón cavefish B chromosome (B) versus Pachón cavefish A chromosomes. Percentage (in log scale) of GC content (GC level), short interspersed repeated sequences (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), DNA repeat elements (DNAs), terminal inverted repeat sequences (TIRs), and satellite DNA (SATs) in the 25 A Pachón chromosomes (boxplots, A) versus Pachón B (black dots, B). **B.** Comparison of the repeat landscapes of the Pachón B chromosome (B), the whole assembly with (whole assembly) and without B and of Scaffold_25. Color code for repeat elements is provided in the top right inset of the figure. B repeat landscape is different from the whole assembly with or without the B chromosome and to the repeat landscape of scaffold_25. The B tends to have a higher content of interspersed repeats, mainly due to the expansion of long interspersed nuclear elements (LINEs). Short interspersed repeated sequences (SINEs), long terminal repeats (LTRs), DNA repeat elements (DNAs), and terminal inverted repeat sequences (TIRs). See Data S1D for additional details. **C.** The *B-gdf6b* loci on the Pachón cave *A. mexicanus* B chromosome (B) stemmed from two successive A and B duplications. Schematic representation of the duplication / insertion history of the two *B-gdf6b* loci on the Pachón cave B (HiC_scaffold_28) with a two-steps duplication hypothesis scheme suggesting that the *B-gdf6b-1* locus originated from an initial duplication of a 11.4 kb region surrounding the *A-gdf6b* locus and inserted in the B, followed by a second independent duplication of a 21.6 kb region surrounding the *B-gdf6b-1* locus that was inserted also in the B. This second independent duplication is probably very recent as the two *B-gdf6b* loci are 99.6 % identical in the 21.6 kb region shared by the two genes. Locations of the duplicated regions are given with respect to the 5'-3' orientation of the *gdf6b* cDNA. The colors of the schematic B fragments depict their different A chromosome origin. **D**: Multiple sequence alignment of Pachón cavefish intron 1 of *A-gdf6b* with *B-gdf6b-1* and *B-gdf6b-2*. Sequences were extracted from the whole genome Pachón cavefish assembly with the following coordinates: For *A-gdf6b* = HiC_scaffold_3:864791:865845:-, For *B-gdf6b-1* = HiC_scaffold_28:617028:618422:+ and for *B-gdf6b-2* = HiC_scaffold_28:1832027:1833421:-. The percentage of identity between the *A-* and *B-gdf6b* is respectively 1.7 % for intron 1 (18 differences in 1054 bp after gaps and indels collapsed to 1 bp) compared to 0.58 % for their proximal promoter (7 differences in 1200 bp after gaps and indels collapsed to 1 bp; proximal promoter defined from the ATG to the 3'end of the adjacent gene, ~ 1400 bp). This could suggest different evolutionary constraints between these two different *gdf6b* regions. **E.** Coding sequence and protein alignments of *A-gdf6b* and *B-gdf6b*. Sequences were aligned with CLUSTALW[S1]. As the *gdf6b-B1* and *gdf6b-B2* are 100% identical from their ATG to STOP codons the alignment only shows differences between *B-gdf6b* (*B-gdf6b-1* and *B-gdf6b-2*) and *A-gdf6b* coding sequences (CDS). The three nucleotide variations between the *B-gdf6b* and *A-gdf6b* CDS are boxed (green for nonsynonymous sites and red for synonymous sites) at positions 180 bp, 591 bp and 679 bp positions of the CDS. Regions highlighted in yellow and blue indicated respectively the Pfam "TGF-β propeptide" and "TGF-β-like" domains[S2]. ".": Identical nucleotides; "*": stop codon. **F.** Alignment of nucleotide (*B-gdf6b*) and translated protein sequences (B-Gdf6b) in wild type and one representative example of a mutant male showing a 470 bp deletion in its *B-gdf6b* gene in F0 fish in the exon 2 encoding for the Pfam "TGF-β propeptide" (yellow) and the "TGF-β-like" (blue) domains[S2]. The resulting fish displayed a frame-shifted and truncated Gdf6b-B protein with premature stop codon (red). The positions of the guide RNAs selected to inactivate *gdf6b* in Pachón cavefish are boxed in red. Similar 470 bp deletions have been also within the *A-gdf6b* gene in other fish.

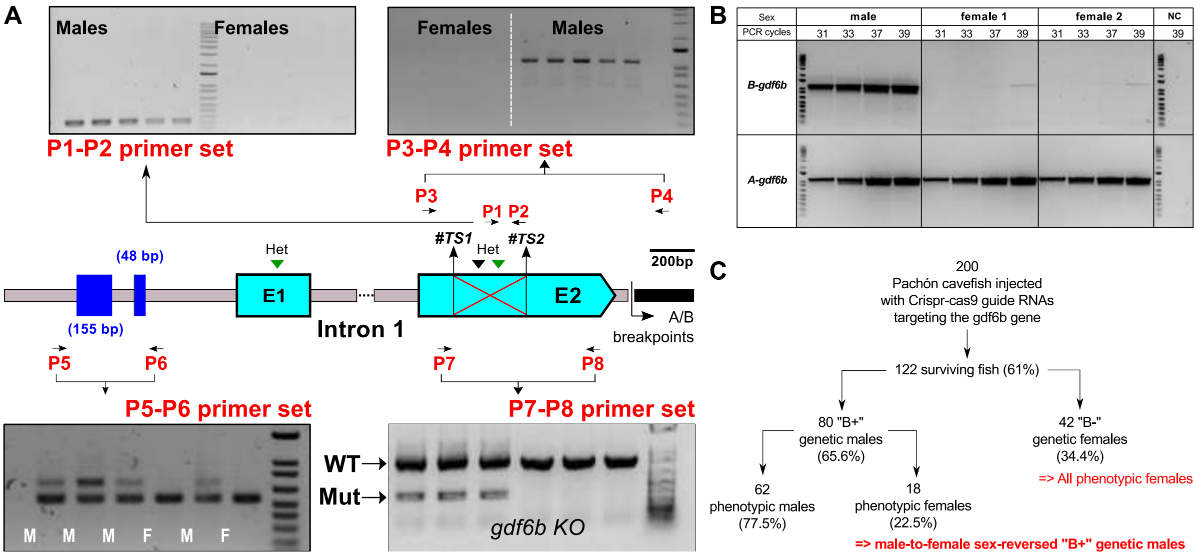

Figure S3: PCR genotyping of the Pachón cavefish male-predominant B chromosome and the *gdf6b* knockout mutant individuals. Related to Figure 4 and STAR Methods. A. Absence / presence of the Pachón cavefish B chromosome was detected with different primer pairs based on differences between the *A-gdf6b* and *B-gdf6b loci*. Three sets of primers (P) were designed with two primer sets designed to amplify specifically the *B-gdf6b* copy based either on a single base variation between the A/B *gdf6b* CDS at position 679 bp (P1-P2 primer set), or based on primers located on both sides of the A/B breakpoints downstream of the *B-gdf6b* gene (P3-P4 primer set). The third set of primers (P5-P6 primer set) was designed on gaps/indels variations between *A-gdf6b* and *B-gdf6b* in the proximal promoter of *gdf6b* genes. Primer sets P1-P2 and P3-P4 produce a single PCR fragment only in B+ individuals, and primer set P5-P6 amplifies two bands in B+ individuals and only a single band in B- individuals. Gene knockout (KO) was performed by genome editing using the CRISPR-cas9 method with two guide RNAs target sites (#TS1 and #TS2) designed in order to target *gdf6b* exon 2 (E2). KO individuals were genotyped using primers (P7-P8 primer set) flanking the 2 target sites that induced a 470 bp deletion in *gdf6b* mutant individuals (Mut) compared to the wild type (WT) *gdf6b* sequence. **B.** Increasing PCR cycle numbers allows the detection of a faint PCR fragment specific to the B chromosome in Pachón cavefish females. PCRs were carried out in one male and two female Pachón cavefish with the protocol described for the P3-P4 primer set and with increasing PCR cycle numbers. The B chromosome was detected using primers specifically designed to amplify the B-gdf6b loci (primer set P3-P4) and a control was incorporated with primers specifically designed to amplify the *A-gdf6b* locus (primer set P9-P10). NC: PCR negative control. **C.** Numbers of *gdf6b* knockout generated by CRISPR-Cas9 method including the number of fish injected, and the number and percentage of sex-reversed males obtained (in bold red type). Out of the 200 micro-injected eggs (at 1 cell stage), 122 adult fishes were obtained including 80 genetic males and 42 genetic females. Among the 80 genetic males, we found 62 phenotypic males (77.5%), and 18 phenotypic females (22.5%) displaying a 470 bp deletion on the exon 2 of their *A-gdf6b* and/or *B-gdf6b* gene.

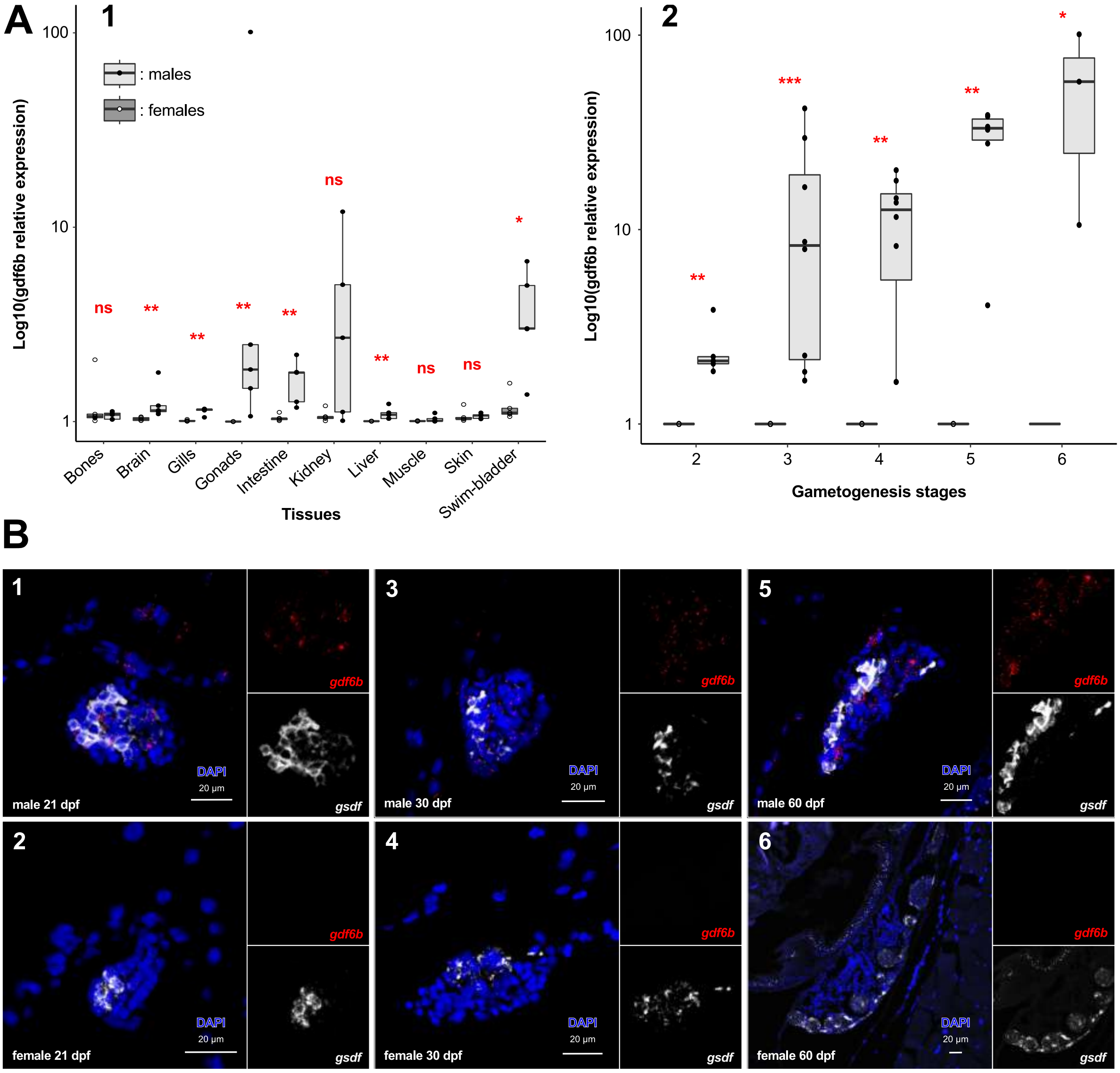

Figure S4. Expression patterns of *gdf6b* in adult Pachón cave *Astyanax mexicanus*. Related to Figure 4. A. Expression profiles of *gdf6b* (*A-gdf6b* and *B-gdf6b*) in different adult tissues (**A1**) in males (light grey) and females (dark grey) and during male (light grey) and female (dark grey) gametogenesis (**A2**). Gametogenesis stages 2 to 6 were defined as previously described[S3]. Results are presented as boxplots with individual expression values displayed as dots, the expression median as a line, and the box displaying the first and third quartiles of expression. Statistical significance between males and females were tested with the Wilcoxon Rank Sum Test (Wilcoxon-Mann-Whitney Test) and significant differences are *** = P < 0.01; ** = P < 0.01; * = P < 0.05. **B.** Gonadal expression of *gdf6b* (in red) and the Sertoli and granulosa supporting cell marker *gsdf* (in white) in male (B1, B3 and B5) and female (B2, B4 and B6) differentiating gonads. At all stages i.e., 21, 30- and 60-days post-fertilization (dpf) *gdf6b* is specifically expressed in male gonads with no strict colocalization with *gsdf* in males. Nuclei were stained with DAPI (in blue). Scale bar = 20 µm.

SUPPLEMENTAL REFERENCES

S1. Higgins, D.G., and Sharp, P.M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene *73*, 237–244.

S2. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res *47*, D427–D432.

S3. Imarazene, B., Beille, S., Jouanno, E., Branthonne, A., Thermes, V., Thomas, M., Herpin, A., Rétaux, S., and Guiguen, Y. (2021). Primordial Germ Cell Migration and Histological and Molecular Characterization of Gonadal Differentiation in Pachón Cavefish Astyanax mexicanus. Sex Dev, 1–18.