

## RESEARCH ARTICLE

## Genomic anatomy of male-specific microchromosomes in a gynogenetic fish

Miao Ding<sup>1,2</sup> , Xi-Yin Li<sup>1,2</sup> \*, Zhi-Xuan Zhu<sup>1,2</sup>, Jun-Hui Chen<sup>3,4</sup>, Meng Lu<sup>1,2</sup>, Qian Shi<sup>1,2</sup>, Yang Wang<sup>1,2</sup>, Zhi Li<sup>1,2</sup>, Xin Zhao<sup>1,2</sup>, Tao Wang<sup>1,2</sup>, Wen-Xuan Du<sup>1,2</sup>, Chun Miao<sup>1,2</sup>, Tian-Zi Yao<sup>1,2</sup>, Ming-Tao Wang<sup>1,2</sup>, Xiao-Juan Zhang<sup>1,2</sup> , Zhong-Wei Wang<sup>1,2</sup> , Li Zhou<sup>1,2</sup>, Jian-Fang Gui<sup>1,2</sup> 

**1** State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, The Innovative Academy of Seed Design, Hubei Hongshan Laboratory, Chinese Academy of Sciences, Wuhan, China, **2** University of Chinese Academy of Sciences, Beijing, China, **3** BGI Genomics, BGI-Shenzhen, Shenzhen, China, **4** ShenZhen People's Hospital, Shenzhen, China

 These authors contributed equally to this work.

\* [lixiyin@ihb.ac.cn](mailto:lixiyin@ihb.ac.cn)


 OPEN ACCESS

**Citation:** Ding M, Li X-Y, Zhu Z-X, Chen J-H, Lu M, Shi Q, et al. (2021) Genomic anatomy of male-specific microchromosomes in a gynogenetic fish. *PLoS Genet* 17(9): e1009760. <https://doi.org/10.1371/journal.pgen.1009760>

**Editor:** Manfred Scharlt, University of Wuerzburg, GERMANY

**Received:** February 15, 2021

**Accepted:** August 9, 2021

**Published:** September 7, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1009760>

**Copyright:** © 2021 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The whole genome assemblies of *C. gibelio* and *C. auratus* are deposited at GenBank database under the accession number of PRJNA546443 (BioSample SAMN11978084) and PRJNA546444 (BioSample

## Abstract

Unisexual taxa are commonly considered short-lived as the absence of meiotic recombination is supposed to accumulate deleterious mutations and hinder the creation of genetic diversity. However, the gynogenetic gibel carp (*Carassius gibelio*) with high genetic diversity and wide ecological distribution has outlived its predicted extinction time of a strict unisexual reproduction population. Unlike other unisexual vertebrates, males associated with supernumerary microchromosomes have been observed in gibel carp, which provides a unique system to explore the rationales underlying male occurrence in unisexual lineage and evolution of unisexual reproduction. Here, we identified a massively expanded satellite DNA cluster on microchromosomes of hexaploid gibel carp via comparing with the ancestral tetraploid crucian carp (*Carassius auratus*). Based on the satellite cluster, we developed a method for single chromosomal fluorescence microdissection and isolated three male-specific microchromosomes in a male metaphase cell. Genomic anatomy revealed that these male-specific microchromosomes contained homologous sequences of autosomes and abundant repetitive elements. Significantly, several potential male-specific genes with transcriptional activity were identified, among which four and five genes displayed male-specific and male-biased expression in gonads, respectively, during the developmental period of sex determination. Therefore, the male-specific microchromosomes resembling common features of sex chromosomes may be the main driving force for male occurrence in gynogenetic gibel carp, which sheds new light on the evolution of unisexual reproduction.

## Author summary

Unisexual taxa are considered short-lived as the accumulation of deleterious mutations and hindering the creation of genetic diversity. However, the gynogenetic gibel carp (*Carassius gibelio*) containing rare and variable proportions of males in wild populations has

SAMN11978330), respectively. The unassembled sequence of *C. gibelio* and *C. auratus* genome generated in this study have been submitted to the NCBI BioProject database under accession number PRJNA659101. The PacBio-SMRT data of the three male-specific microchromosomes generated in this study have been submitted to the NCBI BioProject database under accession number PRJNA658923, PRJNA658930, and PRJNA658940. The long-reads and short-reads RNA-seq data of genotypic female and male gonads at different developmental stages have been submitted to the NCBI BioProject database under accession number PRJNA658709 and PRJNA658708, respectively. The long-reads RNA-seq data of temperature-dependent female and male gonads at different developmental stages have been submitted to the NCBI BioProject database under accession number PRJNA681130. And the other data is available within the paper and its [Supporting Information](#) files.

**Funding:** This work was supported by the National Key Research and Development Project (2018YFD0900204, L.Z.), the National Natural Science Foundation of China (31873036, X.Y.L.), the Key Program of Frontier Sciences of the Chinese Academy of Sciences (QZDY-SSW-SMC025, J.F.G.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24030104, J.F.G.), China Agriculture Research System of MOF and MARA (CARS-45-07, J.F.G.), the Autonomous Project of the State Key Laboratory of Freshwater Ecology and Biotechnology (2019FBZ04, J.F.G.), the Youth Innovation Promotion Association CAS (2020334, X.Y.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

outlived its predicted time of extinction and exhibited strong environmental adaptation, which provides a special system to investigate the evolution of unisexual reproduction in vertebrates. Our previous studies have revealed that the supernumerary microchromosomes are associated with male determination in gibel carp. Here, we further isolated three male-specific supernumerary microchromosomes and revealed that they contained homologous sequences of autosomes and abundant repetitive elements. Besides, we identified several genes with transcriptional activity on these microchromosomes, especially some genes with male-specific or male-biased expression during the developmental period of sex determination. The male-specific microchromosomes with abundant repetitive elements and active male-specific/male-biased genes display common features of sex chromosomes and may be the main driving forces for male occurrence in gynogenetic gibel carp.

## Introduction

Sexual reproduction is prevalent in vertebrates, while only about 100 taxa have been documented to develop unisexual reproductive ability [1–3] since the first unisexual vertebrate Amazon molly (*Poecilia formosa*) was described in 1932 [4,5]. Unisexual vertebrates produce solely female offspring with nearly identical genetic information, mainly via three modes including parthenogenesis, gynogenesis, or hybridogenesis [2,3,6]. In parthenogenesis, females produce unreduced eggs containing the same chromosome complement as somatic cells, and these eggs develop into offspring spontaneously in the absence of males [7]. In gynogenesis, females also produce unreduced eggs with the same chromosome complement as somatic cells, but sperm are required to stimulate the eggs to initiate embryogenesis using only maternal genetic information [8]. In typical hybridogenesis, females produce reduced eggs that contain only maternal haploid chromosomes, and these eggs must be fertilized by sperm from another species. These hybridogenetic offspring contain both maternal and paternal haploid chromosomes, but only maternal haploid chromosomes remain in the reduced eggs [9].

Unisexual taxa without meiosis and meiotic recombination are supposed to be unable to purge deleterious mutations and create genetic diversity stated by Muller's ratchet, which are preconditions for adaptation to the changing environment [1,5,10,11]. Thus, unisexual lineages are considered to be short-lived, although mating costs can be avoided and high fecundity can be achieved with unisexual reproduction [3]. The predicted extinction time of a strict unisexual vertebrate population is no more than 100,000 generations [12], however, a few unisexual taxa have outlived their predicted time of extinction and exhibited strong environmental adaptation [5,13–17]. The hexaploid gibel carp (*Carassius gibelio*), which was originated from ancestral sexual tetraploid crucian carp (*Carassius auratus*) through autopolyploidy [15], can reproduce via unisexual gynogenesis using the males of sympatric host sexual species [8]. And the gynogenetic *C. gibelio* with higher genetic diversity and wider ecological distribution than its sexual progenitor *C. auratus* [15,16] has existed over 0.5 million years [15]. Besides, variable male proportions ranging from 1.2% to 26.5% have been discovered in wild populations of gynogenetic *C. gibelio* [18,19], which is unlike other unisexual taxa with all-female composition [6]. These characteristics make gynogenetic *C. gibelio* a special system to investigate the rationales underlying male occurrence in unisexual lineage and the evolution of unisexual reproduction.

Supernumerary B chromosomes, which occur in about 15% of eukaryotic species [20], are non-essential karyotypic components with non-Mendelian inheritance in addition to standard

A chromosomes (autosomes and sex chromosomes) [20–23], also known as supernumerary chromosomes, B chromosomes, or extra chromosomes. Genomic analyses have revealed that supernumerary chromosomes arise from A chromosomes and accumulate organelle genome-derived sequences [22,24]. Although supernumerary chromosomes are dispensable for the normal life of host individuals [22,25], they have been revealed to contain genes with expression activity [25–28] and be associated with some phenotypes [29]. Especially, sex-ratio distortions related with the presence of supernumerary chromosomes have been identified in many species [30–34]. In hexaploid *C. gibelio*, supernumerary microchromosomes in males are also associated with male determination [23,30], and these males with supernumerary microchromosomes contribute to the creation of genetic diversity, which is able to counter Muller's ratchet at a certain level [19,31,32]. However, the genomic components of these supernumerary microchromosomes and the underlying mechanisms of male occurrence remain elusive in gynogenetic *C. gibelio*.

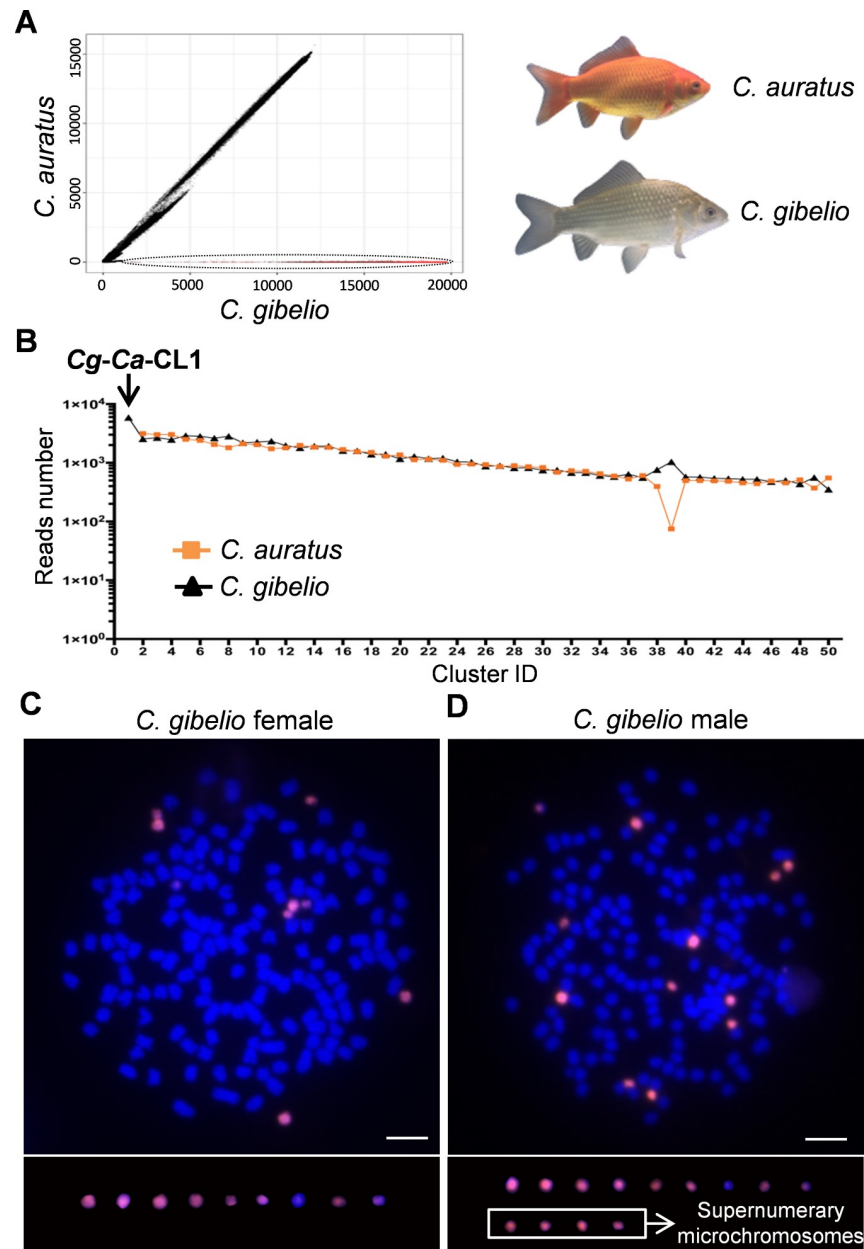
In this study, we analyzed the sequence composition of three microdissected male-specific microchromosomes (MSMs) in hexaploid *C. gibelio*. These MSMs contained sequences homologous to the A chromosomes and abundant repetitive elements. Besides, several genes with transcriptional activity were identified on the MSMs, among which four and five genes showed male-specific and male-biased expression in the gonads, respectively, during the developmental period of sex determination. The features of the MSMs are similar to those of sex chromosomes, including expansion of repetitive elements and accumulation of genes with sex-specific or sex-biased expression. These results suggest that MSMs may be the main driving forces for the male occurrence in gynogenetic *C. gibelio*, which sheds new light on the evolution of unisexual reproduction.

## Results

### Expanded satellite cluster on microchromosomes

Gynogenetic hexaploid gibel carp (*C. gibelio*) was originated from ancestral sexual tetraploid crucian carp (*C. auratus*) via autopolyploidy, and gynogenetic *C. gibelio* contained several microchromosomes, which was absent in sexual *C. auratus* [15]. In order to find repetitive elements on the microchromosomes, we firstly identified the repetitive sequences in *C. gibelio* and *C. auratus* respectively, through all-to-all similarity comparison [33] using the same number of reads (1,220,000) obtained from Illumina sequencing. Subsequently, via inter-species pairwise comparative analysis, remarkable similarities in repetitive sequence composition were found between *C. gibelio* and *C. auratus* (Fig 1A). However, a bunch of repetitive sequences was found only in *C. gibelio* genome (Fig 1A), which were mainly composed of the most expanded satellite cluster (*Cg-Ca-CL1*) (Fig 1B). Besides, *Cg-Ca-CL1* was also the most abundant repetitive sequence in *C. gibelio*, which accounted for 1.49% of whole genome size.

Intriguingly, the 137 bp consensus sequence of satellite cluster *Cg-Ca-CL1* (S1 Table) was the same as the previously identified repeats in a male-specific sequence (*Cg-M-s*) (GenBank accession number KT260068). Five intact and four fragmental repeats of *Cg-Ca-CL1* were distributed in *Cg-M-s* (S1A and S1B Fig). Subsequently, we performed a co-localization analysis of *Cg-Ca-CL1* and *Cg-M-s* via fluorescence *in situ* hybridization (FISH), and found out that the FISH signals of *Cg-Ca-CL1* resided on all microchromosomes in both female and males, which were exactly co-localized with the *Cg-M-s* signals (S2 Fig). These results indicated that the FISH signals of *Cg-M-s* might be mainly derived from these intact or fragmental repeats of *Cg-Ca-CL1*, although *Cg-M-s* has some male-specific sites other than *Cg-Ca-CL1* repeats (S1A Fig) [23]. To reduce procedures of microchromosome identification, we designed three



**Fig 1. Repetitive sequence expansion on microchromosomes.** (A) Pairwise comparison of all analyzed reads between hexaploid *C. gibelio* and tetraploid *C. auratus*. The X-axis and Y-axis show the numbers of similarity hits for each read in *C. gibelio* and *C. auratus*, respectively. Each spot corresponds to one read. The black ellipse indicates the repetitive sequences with expansion in *C. gibelio* compared to *C. auratus*. The red dots represent the reads of the most expanded satellite cluster (*Cg-Ca-CL1*). (B) The top 50 largest repeat clusters generated by *C. gibelio*-*C. auratus* (*Cg-Ca*) pairwise comparative analysis. The Y-axis shows the reads number in clusters, and the X-axis shows the cluster ID. (C, D) FISH analysis of satellite repeat cluster *Cg-Ca-CL1* in female metaphase (C) and male metaphase (D) of *C. gibelio*. Scale bar = 5  $\mu$ m. The white square indicates supernumerary microchromosomes with a male determination role in *C. gibelio*.

<https://doi.org/10.1371/journal.pgen.1009760.g001>

peptide nucleic acid (PNA) probes according to the *Cg-Ca-CL1* satellite cluster (S1C Fig) for FISH analysis. As expected, nine and thirteen microchromosomes could be well identified by these PNA probes in females and males, respectively (Fig 1C and 1D).

*Cg-Ca-CL1* was not detected in tetraploid *C. auratus* during repetitive sequence analysis (Fig 1A and 1B) and no positive *Cg-Ca-CL1* signal was observed through FISH analysis (S3A and S3B Fig). However, polymerase chain reaction (PCR) showed a few low-copy amplicons in the genome of *C. auratus*, compared with abundant high-copy products in *C. gibelio* (S2C Fig). These findings suggest that the satellite cluster *Cg-Ca-CL1* in gynogenetic hexaploid *C. gibelio* might originate from ancestral tetraploid *C. auratus* and undergo substantial expansion along with the evolution of microchromosomes after autopolyploidy [15].

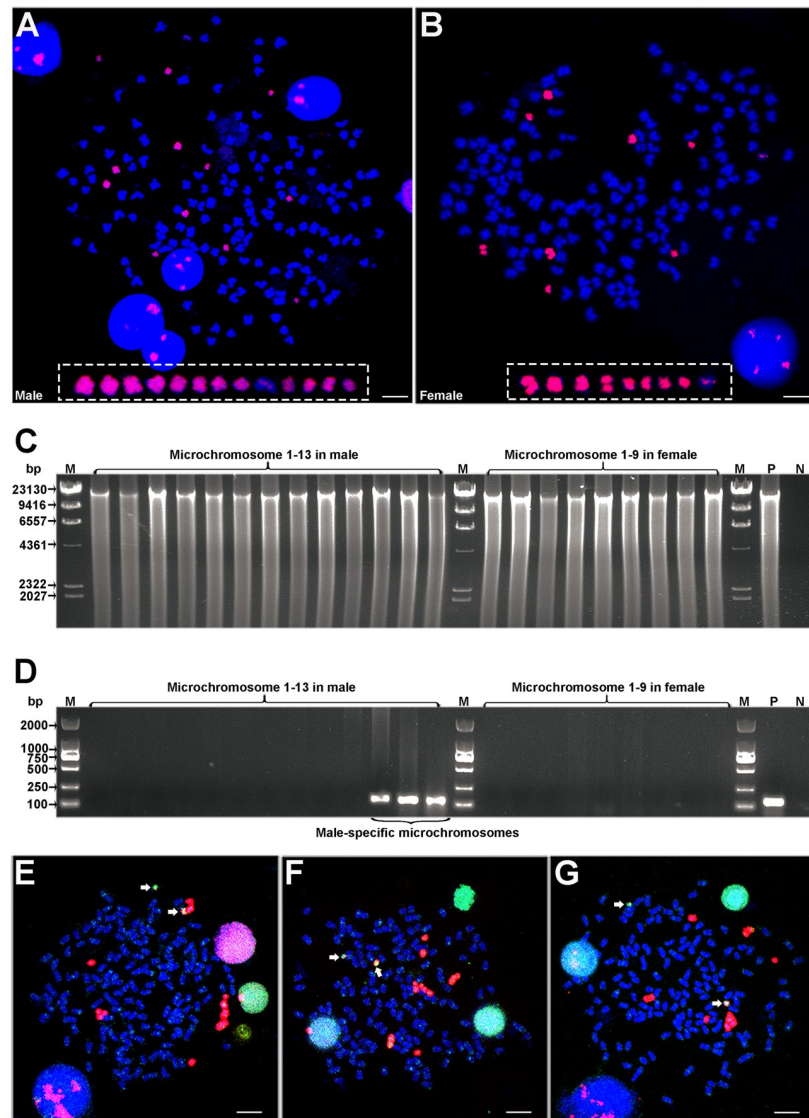
### Microchromosome microdissection and male-specific microchromosome (MSM) identification

To uncover the genomic composition of MSMs, we developed a single chromosomal fluorescence microdissection technique (S4 Fig) based on FISH analysis using PNA probes and microdissection under both fluorescent and white light (see Materials and methods). We microdissected all the 13 microchromosomes from one male metaphase cell (Fig 2A) and all the 9 microchromosomes from one female metaphase cell (Fig 2B), and individually amplified these microdissected chromosomal samples via multiple displacement amplification (MDA) (Fig 2C). To examine the sex specificity of these microdissected microchromosomes, a male-specific marker derived from *Cg-M-s* (S1A and S5 Figs) was used to scan the products of each isolated microchromosome. All the microchromosomes (nine microchromosomes) from the female metaphase cell and most microchromosomes (ten microchromosomes) from the male metaphase cell had no male-specific marker, while the rest three microchromosomes from the male metaphase cell were detected to contain the male-specific marker (Fig 2D). These three microdissected microchromosomes with the male-specific marker were defined as MSMs (Fig 2D). The male metaphase cell had four extra microchromosomes than the female metaphase cell (Fig 2A and 2B), but only three extra microchromosomes in the male metaphase cell were identified as MSMs (Fig 2C and 2D). Maybe, the male metaphase has one extra microchromosome without male specificity, but we also cannot exclude the possibility that the MDA-based DNA amplification of a single microchromosome has not amplified the entire DNA.

Subsequently, the amplified DNAs of three microdissected MSMs were used as probes along with microchromosome-specific PNA probes for FISH co-localization analysis. The signals of amplified DNA were mainly localized on almost all microchromosomes in male metaphases and two microchromosomes exhibited intensive signals (Fig 2E–2G), which indicated that the technical process of single chromosomal fluorescence microdissection is accurate. Besides, some weak signals from amplified DNA were also observed in some autosomes (Fig 2E–2G), possibly due to some other repetitive elements without male specificity.

### Genomic sequences of male-specific microchromosomes

The amplified products of three microdissected MSMs were sequenced via the continuous long-read (CLR) of Sequel II (PacBio platform). After filtering low quality reads and removing adaptor sequences, a total of 6.92 Gb data were obtained from 1,545,433 clean reads with 7,974 bp N50 length. MSM 1, MSM 2, and MSM 3 generated 624,541 clean reads, 505,551 clean reads and 415,341 clean reads with data sizes of 2.56 Gb, 2.34 Gb, and 2.02 Gb, respectively (S2 Table). The clean reads of each MSM were self-corrected by CANU (S3 Table) and then assembled via CANU (S4 Table), SPAdes (S5 Table), and SMARTdenovo (S6 Table), respectively. Subsequently, the genome assembly of a female *C. gibelio* [34] and the full-length gonadal transcriptomes (S7 Table) were used as references respectively to assess the assemblies of MSMs (see Materials and methods). All the contigs and corrected reads displayed much lower alignment level to the references in sharp contrast with the clean reads (S6 Fig), which might be

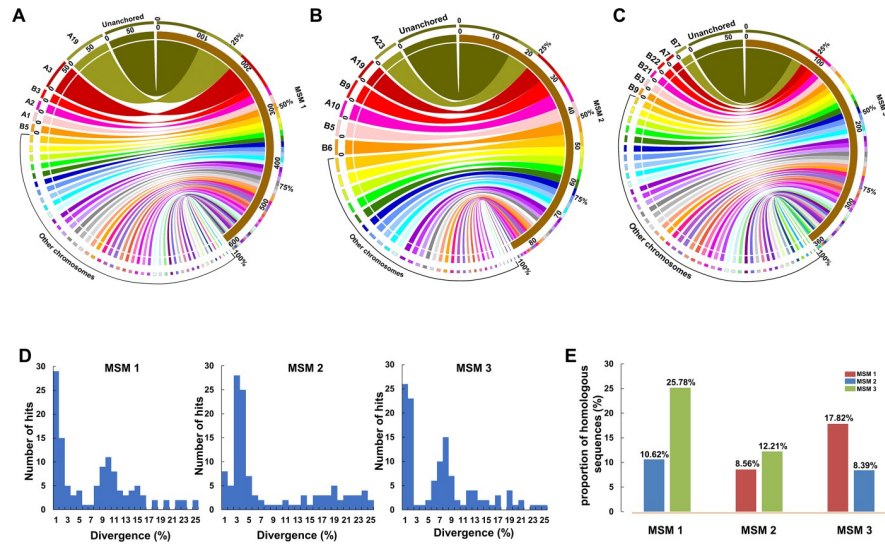


**Fig 2. Identification of three male-specific microchromosomes.** (A, B) FISH analysis on the male metaphase (A) and female metaphase (B) using PNAs as probe. The white rectangle indicates microchromosomes. Scale bar = 5  $\mu$ m. (C) Electrophoresis of the amplified products from microdissected microchromosomes. (D) PCR detection of these amplified products using male-specific primers. M, marker; P, positive control using genomic DNA as template; N, negative control using water as template. (E-G) Co-localization of amplified DNAs and microchromosome-specific PNA probes. The amplified DNAs of MSM 1 (E), MSM 2 (F), and MSM 3 (G) were labeled with Digoxin, and the PNAs were labeled with Biotin, which appeared green and red fluorescence respectively. Scale bar = 5  $\mu$ m. Mitotic cells of three males were used as replication for co-localization analysis.

<https://doi.org/10.1371/journal.pgen.1009760.g002>

caused by the removal of repetitive sequences during the process of correcting and assembling. Thus, to better reflect the reality of sequence composition, we only used the clean reads for subsequent analyses.

Mapping the MSM clean reads to the reference genome of a female *C. gibelio* [34] revealed that a total of 46.35% MSM 1 sequences, 38.84% MSM 2 sequences, and 40.01% MSM 3 sequences were homologous to the sequences from almost all linkage groups (Fig 3A–3C). With the exception of unanchored sequences, the largest number of reads from MSM 1, MSM 2, and MSM 3 were mapped to linkage groups A19, A23, and B7 respectively in *C. gibelio* (Fig



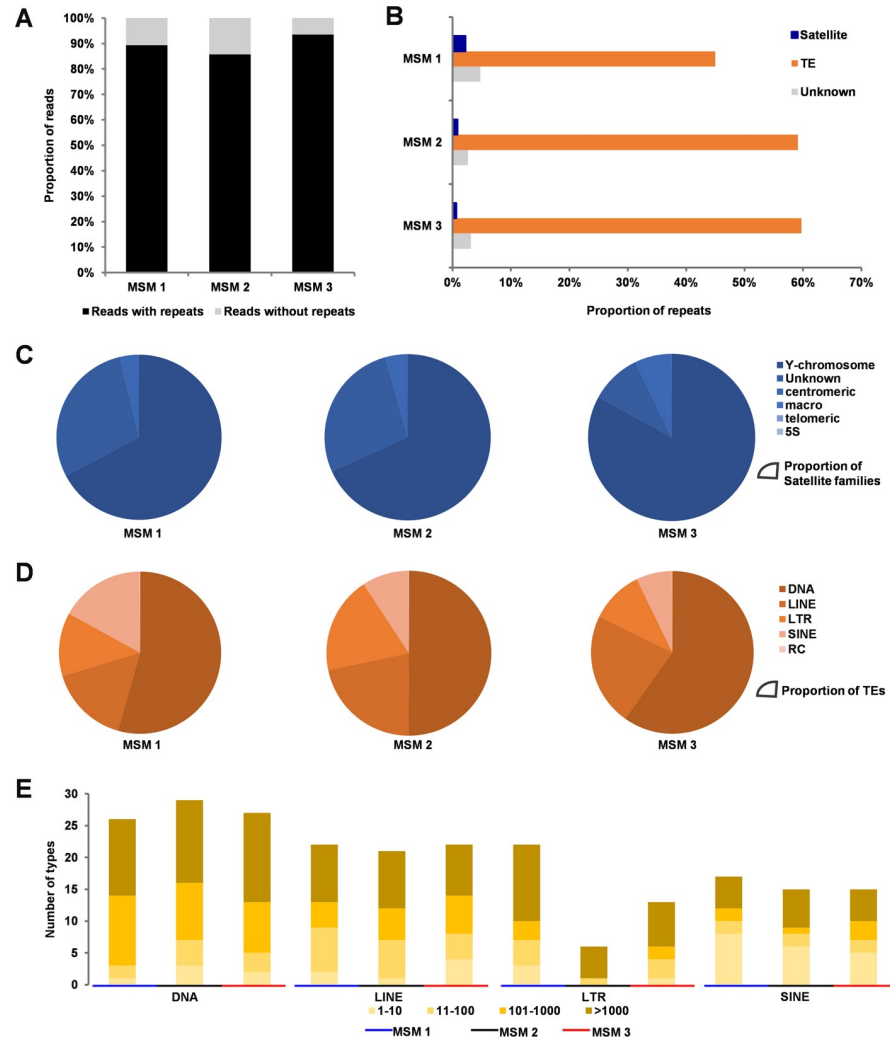
**Fig 3. Comparative genomic analysis.** (A–C) Comparative analysis between MSM sequences and the reference genome of *C. gibelio*. MSM 1 (A), MSM 2 (B), and MSM 3 (C) blocks are shown at the right and the corresponding linkage groups of reference genome are displayed at the left. (D) Kimura divergence between the reference genome and homologous sequences of MSMs. The X axis represents the divergence and the Y axis displays the number of hits. (E) Pairwise comparative analysis among three MSMs. The X axis indicates the compared objects and the Y axis indicates the proportion of homologous sequences.

<https://doi.org/10.1371/journal.pgen.1009760.g003>

3A–3C). Kimura distance analysis (see [Materials and methods](#)) was used to indicate the evolutionary divergence between the reference genome and homologous sequences of MSMs. Referring to the genome assembly, the aligned sequences showed two main distribution peaks in both MSM 1 (1% and 10% divergence) and MSM 3 (1% and 8% divergence), while the aligned sequences of MSM 2 were concentrated in only one main peak with 3% divergence ([Fig 3D](#)). Besides, pairwise comparative analysis revealed that the proportion of homologous sequences between MSM 1 and MSM 3 is over 2-fold higher than that between MSM 2 and MSM 1 or between MSM 2 and MSM 3 ([Fig 3E](#)).

### Characterization of repetitive elements on male-specific microchromosomes

To search repetitive elements, we first constructed a TE database specific to the hexaploid *C. gibelio* via combined prediction of the signature and homology (see [Materials and methods](#)). Afterward, the constructed *C. gibelio*-specific TE database and the known metazoan repetitive database (RepBase 23.07) were used to identify repetitive elements on MSMs. We found that 89.40%, 85.76%, and 93.58% of reads contained at least one repetitive element on MSM 1, MSM 2, and MSM 3, respectively ([Fig 4A](#)). Further, 52.00%, 62.57%, and 63.49% of the sequences of MSM 1, MSM 2, and MSM 3 were identified as repetitive elements including satellite, TE, and unknown repeats ([Fig 4B](#)), which were 1.2- to 1.5-fold higher than repeat content in the whole genome assembly of a female individual containing microchromosomes (42.6%) [34]. Among the three MSMs, Y-chromosome satellites were the most abundant satellite sequences ([Fig 4C](#)), and DNA transposons were the most frequent repeats of TEs ([Fig 4D](#)). Besides, most of the TE families on these MSMs amplified to a medium (11–100 copies) or high (>100 copies) copy number ([Fig 4E](#)). These results indicate that numerous repetitive elements have been accumulated on the MSMs during the evolutionary process.



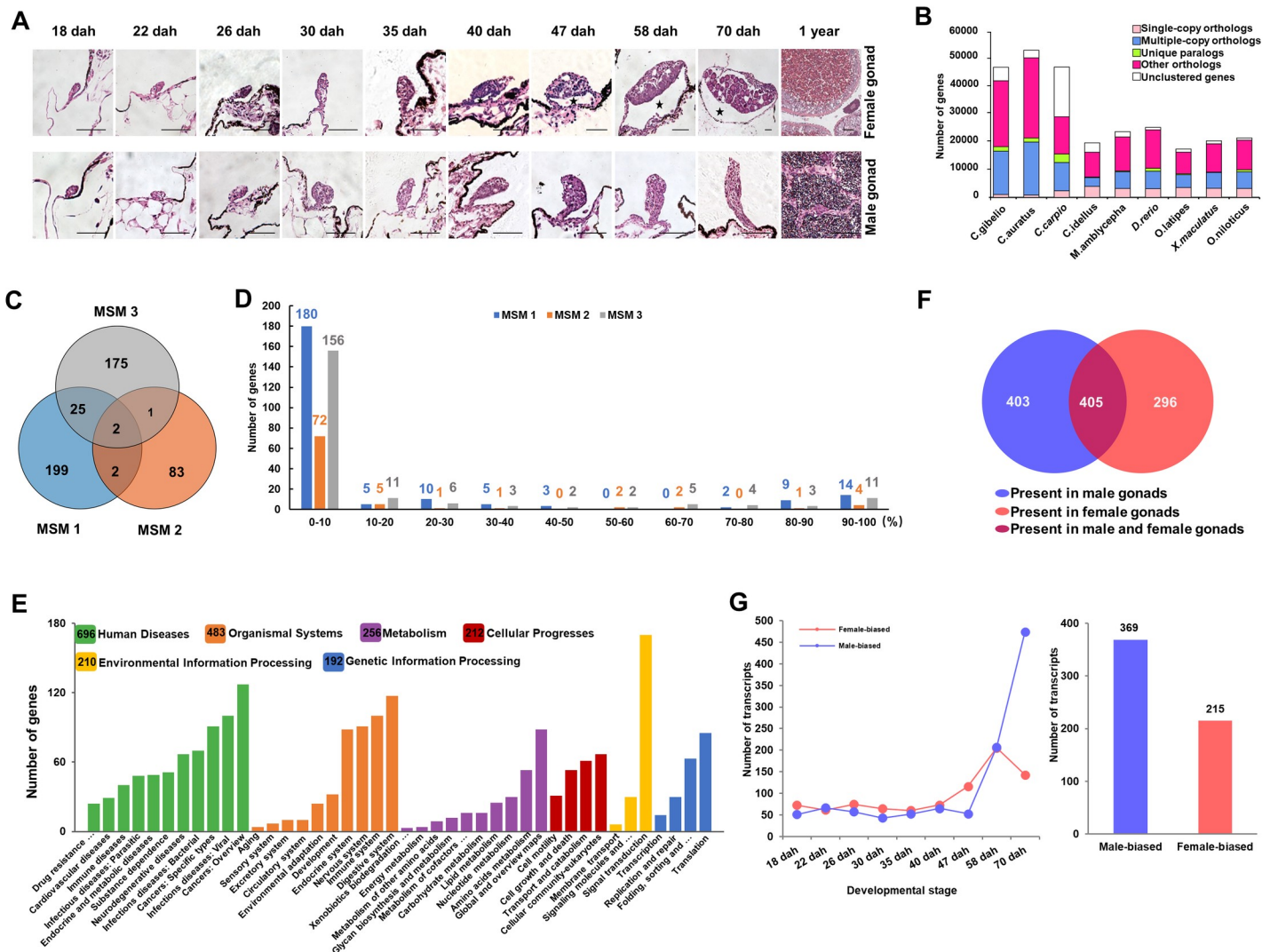
**Fig 4. Analysis of repetitive elements.** (A) The proportion of sequencing reads containing repetitive elements. The Y axis indicates the proportion of reads. (B) The proportion of repetitive elements including satellite, TE, and unknown. (C, D) Types and proportion of satellite repeats (C) and TEs (D). (E) Copy numbers of different types of TEs.

<https://doi.org/10.1371/journal.pgen.1009760.g004>

### Identification of genes derived from male-specific microchromosomes

We performed gene annotation of MSMs via Minimap2 and the Basic Local Alignment Search Tool (BLAST) using gonadal transcriptomes at ten developmental stages of hexaploid *C. gibelio* (Fig 5A and S7 Table) and the coding sequences of nine fish species (Fig 5B) as references. A total of 487 genes (228, 88, and 203 on MSM 1, MSM 2, and MSM 3, respectively) were identified on MSMs, among which 2 genes were shared by all the three MSMs, and 25 genes were shared by MSM 1 and MSM 3 (Fig 5C). Meanwhile, MSM 1 and MSM 2 shared only 1 gene, and MSM 2 and MSM 3 shared 2 genes (Fig 5C). A total of 1.45%, 0.02%, and 0.51% clean reads were revealed to contain gene sequences on MSM 1, MSM 2, and MSM 3, respectively (S7 Fig). And these identified gene sequences account for 1.067%, 0.017%, and 0.439% total sequences of MSM 1, MSM 2, and MSM 3, respectively. Subsequently, we performed analysis of gene integrity and found out that the majority annotated genes had the integrity scores less than 10%, which were accounted for 78.95% (180 genes), 81.82% (72 genes), and 76.85% (156 genes) of annotated genes in MSM 1, MSM 2, and MSM 3, respectively. Meanwhile, there are





**Fig 5. Identification of genes and the corresponding transcripts.** (A) The hematoxylin-eosin staining of genotypic female and male gonads at 10 developmental stages. Scale bar = 50  $\mu$ m. (B) The distribution of different types of gene families among nine species. The X axis shows nine species and the Y axis indicates the number of genes. (C) Numbers of identified genes on MSM 1, MSM 2 and MSM 3. (D) Results of gene integrity analysis. X axis indicates each integrity percentage bins and Y axis shows the number of genes. (E) KEGG analysis of the MSM-linked transcripts. (F) The distribution of MSM-linked transcripts in male and female gonads. (G) Sex-biased transcripts during the early stages of gonadal development. Dah, days after hatch.

<https://doi.org/10.1371/journal.pgen.1009760.g005>

only 10.96% (25 genes), 10.23% (9 genes), and 12.32% (25 genes) annotated genes with integrity score over 50% in MSM 1, MSM 2, and MSM 3, respectively (Fig 5D). All the annotated 487 genes corresponded to 1104 transcripts (MSM-linked transcripts), which belonged to six Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to “Human Disease” (696 transcripts), “Organismal Systems” (483 transcripts), “Metabolism” (256 transcripts), “Cellular Progresses” (212 transcripts), “Environmental Information Processing” (210 transcripts) and “Genetic Information Processing” (192 transcripts) (Fig 5E). Among these 1104 MSM-linked transcripts, 403 and 296 transcripts were only detected in the transcriptome of male and female gonads, respectively (Fig 5F). Further, sex-biased transcripts were identified via differential expression analysis (false discovery rate < 0.05,  $|\log_2\text{FoldChange}| > 1$ ) at each developmental stage. A total of 369 male-biased transcripts and 215 female-biased transcripts

were identified from nine developmental stages, specifically at 18, 22, 26, 30, 35, 40, 47, 58, and 70 days after hatch (dah) (Fig 5G).

### Gene fragments with male-specific or male-biased expression

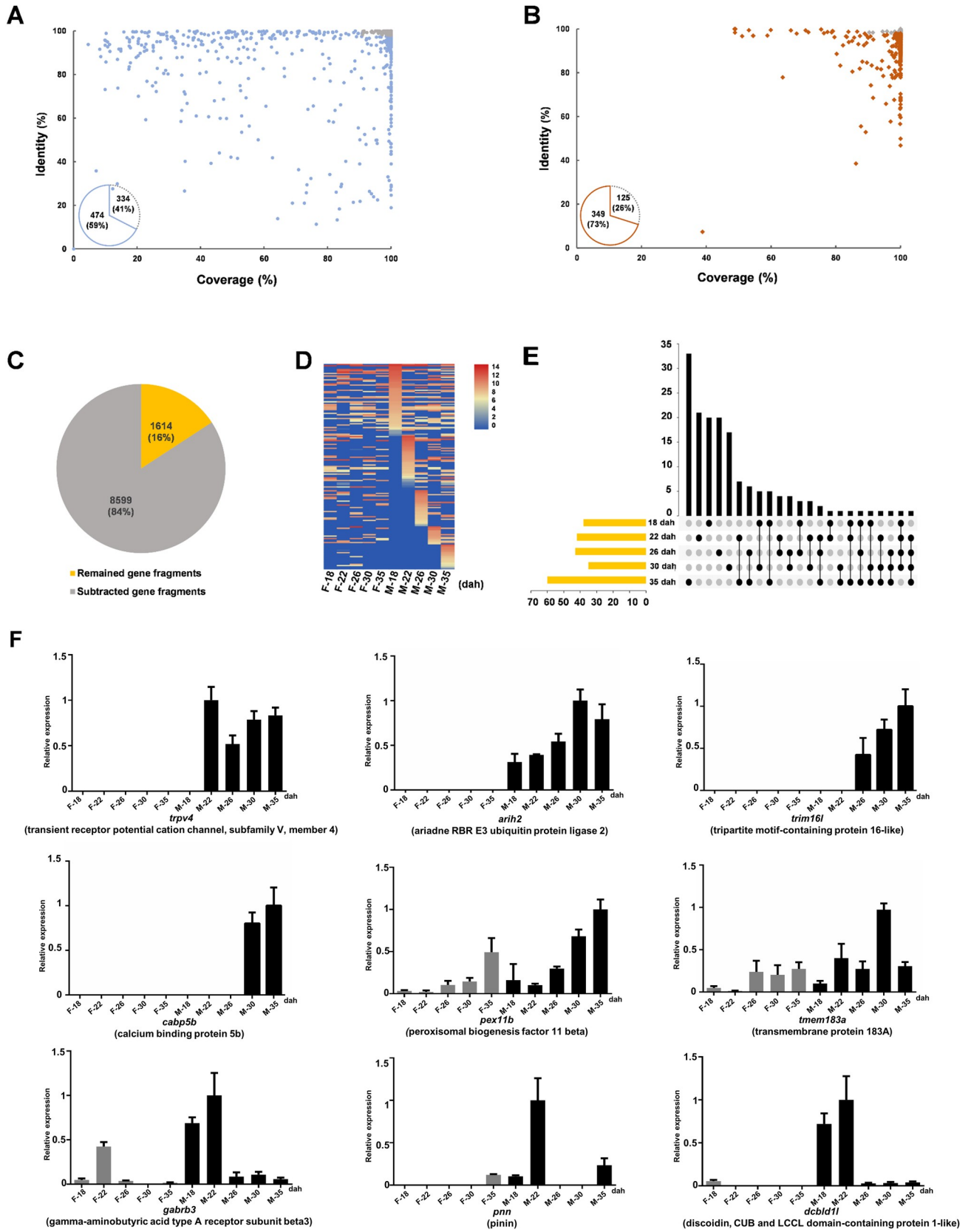
To identify MSM genes with male-specific expression, 808 MSM-linked full-length transcripts presenting in male gonads (Fig 5F) were selected for subsequent bioinformatic subtraction (see [Materials and methods](#)). Firstly, two transcriptomes derived from individuals without MSMs were used for transcriptome subtraction, and the 334 transcripts that had high identity with these two transcriptomes were excluded (Fig 6A). After transcriptome subtraction, the remaining 474 transcripts were then mapped to the reference genome of a female *C. gibelio* [34] for genome subtraction. And then 125 transcripts that had high identity with the female genome were excluded and 349 transcripts were obtained after genome subtraction (Fig 6B). Commonly the full-length transcripts cannot continuously be aligned to the sequences of MSMs, not only because of the intron sequences on MSMs but also as many genes have been duplicated on MSMs as partial truncated genes [20,28,35–37]. Thus, we mapped the remaining 349 transcripts to the sequences of MSMs via BLAST with default parameter, and 10,213 well-aligned sequences of MSMs (identity > 75% and aligned length  $\geq$  200 bp) were defined as gene fragments. To be more specific, the second round of transcriptome/genome subtractions was performed on these gene fragments, and 8,599 gene fragments were excluded. Finally, 1,614 gene fragments remained as the potential male-specific gene fragments on MSMs (Fig 6C).

To identify the genes crucial for sex determination, we screened the gonadal expression of these 1,614 potential male-specific gene fragments on MSMs before gonadal morphological differentiation (40 dah) (Fig 5A), via Illumina sequencing data analysis (S8 Table). We found out that a total of 159 gene fragments had higher transcription level in the male gonads than in the female gonads at least at one gonadal developmental stage before 40 dah (38, 42, 43, 35, and 60 gene fragments with higher transcriptional expression in male gonads than in the female gonads at 18, 22, 26, 30, and 35 dah, respectively) (Fig 6D and 6E). After manual subtraction of the highly overlapped fragments via pairwise sequence alignment, only 42 unique gene fragments remained, and 10 gene fragments were identified to have a conserved coding sequence compared to their homolog in A chromosomes of *C. gibelio* or other species (S9 Table).

Subsequently, relative quantitative real-time PCR was performed to analyze the gonadal expression patterns of all the 42 gene fragments before gonadal morphological differentiation, which was the developmental period of sex determination (Figs 6F and S8). At last, four (*trpv4*, *arih2*, *trim16l*, and *capb5b*) and five (*pex11b*, *tmem183a*, *gabrb3*, *pnn*, and *dcbl11*) gene fragments were confirmed to display male-specific and male-biased expression respectively (Fig 6F), among which two gene fragments (*trim16l* and *tmem183a*) contained a potential coding sequence. Thus, these MSMs containing genes with male-specific or male-biased expression might be beneficial for male occurrence.

### Discussion

The consequences of polyploidy are frequently related to unisexual reproduction modes such as gynogenesis, parthenogenesis, hybridogenesis, and kleptogenesis [2,38–40]. Hexaploid *C. gibelio* was originated from ancestral tetraploid *C. auratus* at about 0.5 Mya via autopolyploidy [15,16,41], and the newly formed hexaploid *C. gibelio* broke through the reproduction bottleneck via unisexual gynogenesis [8,42]. Although gynogenesis has the ability to avoid mating costs and obtain high fecundity, gynogenetic taxa without meiosis cannot purge deleterious mutation and create genetic diversity, which will lead to the eventual extinction stated by Muller's ratchet [1–3,10,11]. However, the gynogenetic *C. gibelio* has higher genetic diversity and



**Fig 6. Potential male-specific gene fragments with transcriptional activity.** (A, B) Transcriptome subtraction (A) and subsequent genome subtraction (B) performed on MSM-linked transcripts. The grey dots represent transcripts with the coverage > 90% and identity > 98%. The pie charts indicate the number of discarded (grey) and remained transcripts (white and brown) (C) The pie chart of the second round of transcriptome/genome subtractions performed on gene fragments. (D) Heatmap of the gene fragments with higher expression in male gonads than in female gonads at least at one developmental stage from 18 to 35 dah. (E) The distribution of the gene fragments with higher expression in male gonads than in female gonads. The axis indicates the number of gene fragments. (F) qPCR detection of gene fragments with male-specific or male-biased expression at early gonadal developmental stages including 18, 22, 26, 30, and 35 dah. The X axis represents the stages of gonad development. The Y axis represents the relative expression, and the highest expression level of each gene fragment was used as control and defined as 1. F, female; M, male.

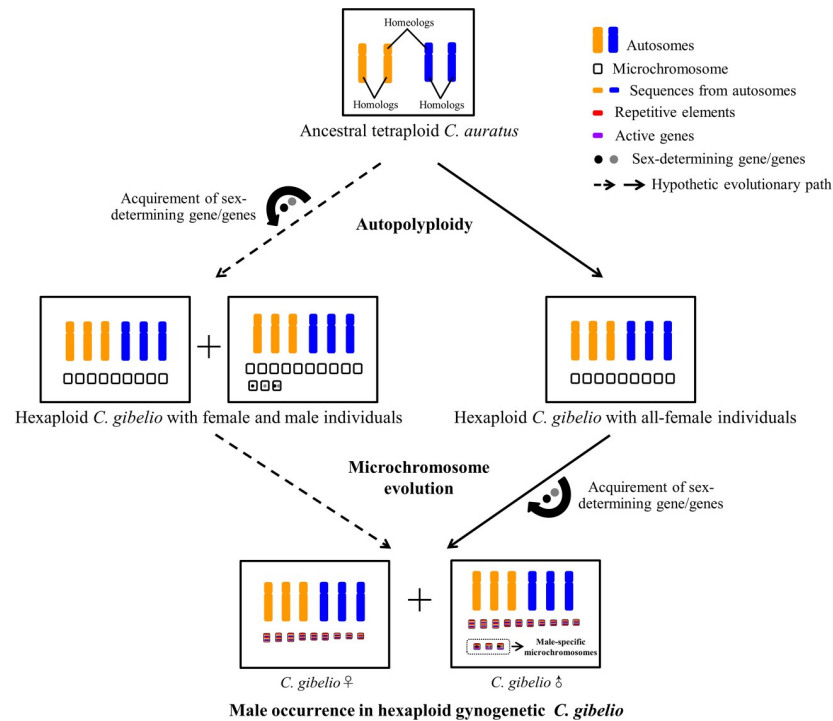
<https://doi.org/10.1371/journal.pgen.1009760.g006>

wider geographic distributions than its sexual progenitor *C. auratus* [16,19]. Besides, given a generation time of 1–2 years, the gynogenetic *C. gibelio* has existed for about 250,000–500,000 generations [15], which has exceeded the predicted extinction generation of a strict unisexual reproduction population (100,000 generations) [5,12]. In contrast with other unisexual vertebrates composed of all females, males containing MSMs have been observed in wild populations and artificially propagated strains of *C. gibelio* [18,23]. These males with MSMs can initiate a variant of gynogenesis along with male occurrence, accumulation of microchromosomes, and creation of genetic diversity in the offspring [31,32], which can contribute to the environmental adaptation and evolutionary long existence of gynogenetic *C. gibelio*.

The close association between the presence of supernumerary chromosomes and sex-ratio distortion is not only observed in *C. gibelio* [23,42], but also has been demonstrated in many other taxa [43–46]. Besides, the supernumerary chromosomes have also been revealed to have a functional effect on sex determination in *Lithochromis rubripinnis* [35], *Nasonia vitripennis*, and *Trichogramma kaykai* [47,48]. In *C. gibelio*, male-specific supernumerary microchromosomes accumulate numerous repetitive elements (Figs 1 and 4) and contain active genes with male-specific or male-biased expression during the developmental period of sex determination, which are usually accompanied with the evolution of sex chromosomes [49–53]. Therefore, we could deduce that the MSMs resembling common features of sex chromosomes may be the main driving force for male occurrence in the gynogenetic *C. gibelio*.

There are two possibilities for the origin of MSMs (Fig 7). The first one is that the MSMs might emerge along with autopolyploidy at about 0.5 Mya. These microchromosomes acquired the sex-determining gene/genes from the A chromosomes of sexual progenitor during autopolyploidy and male *C. gibelio* emerged at the beginning of hexaploid *C. gibelio* formation. The second possibility is that the MSMs might form during the evolutionary process after autopolyploidy, in which no male *C. gibelio* emerged at the beginning of hexaploid *C. gibelio* formation. The sex-determining gene/genes were acquired on MSMs during the evolutionary trajectory of gynogenesis. And the sex determinant might be derived from the duplicates of A chromosomes of gynogenetic *C. gibelio* or the DNA introgression of host sexual species. Thus, further identification of sex-determining gene/genes on MSMs is essential for unveiling the origin and evolution of MSMs in the gynogenetic *C. gibelio*.

Although supernumerary chromosomes are nonessential genetic elements, many supernumerary chromosomes have the intrinsic ability to transmit themselves at frequencies above that predicted by Mendelian rules [54,55]. Similarly, the microchromosomes derived from genotypic males of *C. gibelio* also can be accumulated in the offspring [18,31]. So there is a possibility that the MSMs may be transmitted steadily across generations as a selfish genomic parasite, and the gynogenetic *C. gibelio* will maintain the present status with MSMs for male determination. Meanwhile, we also cannot exclude another possibility that the MSMs may evolve to sex chromosome or provide materials for sex chromosome evolution, as sex chromosomes in some species have been demonstrated to be evolved from supernumerary chromosomes, such as the Y chromosome in *Drosophila* species [56,57], W chromosome in Lepidoptera [58], and sex chromosomes in some cichlid fish species [46,59].



**Fig 7. Schematic diagram of MSM origin and male occurrence in *C. gibelio*.** Hexaploid *C. gibelio* was originated from ancestral tetraploid *C. auratus* via autopolyploidy, and the newly formed hexaploid *C. gibelio* reproduced via unisexual gynogenesis. Two possible evolutionary trajectories of MSM origin and male occurrence in *C. gibelio* were indicated by the dashed arrow and solid arrow respectively. One possibility is that MSMs and males might emerge at the beginning of *C. gibelio* formation via autopolyploidy, and the sex-determining gene/genes might be accumulated from the A chromosomes of sexual progenitor during autopolyploidy (dashed arrow). The other possibility is that MSMs and males did not emerge at the beginning of *C. gibelio* formation but formed during the evolutionary process after autopolyploidy. And the sex-determining gene/genes might be acquired from the duplicates of A chromosomes of *C. gibelio* or the DNA introgression of host sexual species (solid arrow).

<https://doi.org/10.1371/journal.pgen.1009760.g007>

In conclusion, we isolated three MSMs in gynogenetic *C. gibelio* and characterized abundant repetitive elements and some genes with male-specific/male-biased expression on MSMs. Our present results indicate that MSMs could be responsible for male occurrence in *C. gibelio* [18,23,31], which can facilitate further identification of the sex-determining gene/genes on MSMs. Besides, continued investigations on sex determination mechanism and reproduction mode of the gynogenetic *C. gibelio* with male occurrence will provide insights into the evolution of unisexual reproduction.

## Materials and methods

### Ethics statement

Animal experiments and treatments were performed according to the Guide for Animal Care and Use Committee of Institute of Hydrobiology, Chinese Academy of Sciences (IHB, CAS, Protocol No. 2016–018).

### Experimental fish

Experimental fish species including hexaploid gibel carp (*C. gibelio*), tetraploid crucian carp (*C. auratus*), and red common carp (*Cyprinus carpio*) were provided by the National Aquatic

Biological Resource Center (NABRC), Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China.

### Illumina sequencing and identification of repetitive sequences

One female *C. auratus* and one female *C. gibelio* were used for re-sequencing, which was performed by Illumina (HiSeq X-ten) platform and yielded quality filtered paired-end reads with a length of 150 bp. The same amount of reads (1,220,000) were randomly selected from each fish sample for pairwise comparison to identify the repetitive sequences. Repetitive sequences were analyzed by all-to-all similarity comparisons and graph-based clustering at the RepeatExplorer platform with default parameters on a Galaxy server [33].

### Fluorescence *in situ* hybridization (FISH)

Chromosome preparation was performed as previously described [60]. Repetitive sequence cluster *Cg-Ca-CL1* was labeled by Biotin-Nick Translation Mix (Roche) and used as the probe. The male-specific sequence (*Cg-M-s*) (GenBank accession number KT260068) and the amplified products of microdissected microchromosome were labeled by DIG-Nick Translation Mix (Roche) and used as the probe. FISH analysis was performed as previously described [61].

### Single chromosomal fluorescence microdissection

One male individual and one female individual were used for single chromosomal fluorescence microdissection. Three PNA probes, which were designed according to the repetitive sequence cluster *Cg-Ca-CL1* and labeled with Cy3 (S1 Fig), were used for FISH analysis with some modifications. The 60  $\mu$ l hybridization mix contained 50% deionized formamide, 0.5  $\mu$ g/ $\mu$ l sheared salmon sperm DNA, 0.1% SDS, 20 $\times$ SSC, 20% dextran sulphate, and 3  $\mu$ l PNA probes (each probe 1  $\mu$ l, 50 ng/ $\mu$ l). The mixture was denatured at 73°C for 3.5 min and immediately transferred into ice. After cooling, the hybridization mix was placed on the slides without any bubble and kept at 37°C overnight for hybridization. Single chromosome microdissection was performed as described previously [62] with minor modifications. All the equipment, microneedles, and RNase/DNase-free tubes were treated with UV irradiation for 30 min to decrease the DNA contaminant [63]. The single microchromosome was scraped from metaphase after FISH analysis with a glass microneedle under both fluorescent and white light, using the Eppendorf TransferMan 4 micromanipulator and Nikon Ti-E microscope. The scraped microchromosome was transferred into a 200  $\mu$ l RNase/DNase-free tube on ice immediately.

We microdissected all the 13 microchromosomes from one male metaphase cell and all the 9 microchromosomes from one female metaphase cell. After all the microchromosomes were isolated respectively, the microdissected microchromosomes were individually amplified via multiple displacement amplification (MDA) using the REPLI-g Single Cell Kit (Qiagen), following the manufacturer's protocol [64]. To eliminate the contamination of extraneous DNA, all the operations were performed in a vertical clean bench, and all the instruments were treated with UV irradiation for 30 min. RNase/DNase-free water and 25 pg genomic DNA were used as negative and positive control respectively for amplification *in vitro*. The amplified products were analyzed by 1% agarose gel electrophoresis and then purified via the AxyPrep PCR Cleanup Kit (Axygen) with 30  $\mu$ l eluent, whose concentration was measured using NanoDrop 2000 (Thermo). And all the products were stored at -20°C until use. Mitotic cells of three males were used as replication for co-localization of amplified DNAs and microchromosome-specific PNA probes.

## Sequencing and assembling of MSMs

The MDA products of three microdissected MSMs were purified and then used to construct libraries with 4.5–10 kb inserts following the protocol of the PacBio template preparation kit. CLR technology was used and sequencing reactions were performed on the PacBio Sequel II platform instrument with Sequel Sequencing Kit 2.1 (Pacific Biosciences) and Sequel SMRT Cell 1M v2 Tray, at BGI-shenzhen, China. After filtering low quality reads and removing adaptor sequences, the remained clean reads were self-corrected using CANU [65] (version 1.4) with parameters genomeSize = 20m, errorRate = 0.013, corOutCoverage = 60, corMhapSensitivity = normal, min-ReadLength = 500, and corMinEvidenceLength = 500. And the corrected reads were subjected to three widely-used PacBio assembler, including CANU (version 1.4) with default parameters, SPAdes (SPAdes-3.12.0) with default parameters, and SMARTdenovo (smartdenovo-170825) with parameters of -S 4, -k 21, -z 12, -Z 300, -U, -l, -m 0.6, -A 1000. And all the contigs derived from these three methods were polished by Illumina short reads using Pilon program [66] (version:2.3).

To assess the assemblies of MSMs, two databases were used as references respectively including the genome assembly of a female *C. gibelio* [34] and the full-length gonadal transcriptomes of genotypic females and males at ten developmental stages (S7 Table). The clean reads, corrected reads, CANU contigs, SPAdes contigs, and SMARTdenovo contigs of MSMs were mapped to the two references respectively, via BLAST (default parameters). Subsequently, the aligned hits (identity >75% and aligned length  $\geq$  100 bp) were used to generate genome blocks and transcriptome blocks using our custom python script.

## Production of genotypic male offspring and RNA-seq

To obtain a high proportion of genotypic male offspring in *C. gibelio*, we firstly produced sex-reversed physiological females from the genotypic males (with MSMs) via estradiol treatment [23]. Subsequently, the sex-reversed physiological female (with MSMs) was mated with a normal genotypic male (with MSMs), and the proportion of genotypic males in the offspring was 96.1% (S9A Fig). Meanwhile, we also reproduce all-female offspring via unisexual gynogenesis as the previous description, that the ovulated eggs from a female *C. gibelio* were inseminated with the sperm from another species red common carp (S9B Fig) [23]. All the lava in these two families were reared at normal water temperature about 20°C ( $\pm$ 1°C), and these two families were constructed in the strain DA of *C. gibelio* as the sex of individuals in strain DA could not be easily affected by rearing temperature [18].

Genotypic female gonads were obtained from the gynogenetic family with all-female offspring (S9B Fig), while genotypic male gonads were obtained from the family with a high proportion of male offspring (S9A Fig), in which female individuals were excluded via analysis of male-specific marker identified previously [23]. Gonads at ten developmental stages from 198 genotypic females and 198 genotypic males were sampled for subsequent morphological observation and transcriptome analysis. And a total of 30, 30, 20, 20, 20, 20, 10, 10, 5, and 3 gonads were pooled for RNA isolation at the stage of 18, 22, 26, 30, 35, 40, 47, 58, 70 dah, and mature gonads with 1 year old respectively. And 3 female gonads and 3 male gonads were sampled at each stage for morphological observation.

Ten RNA samples of genotypic female gonads at different stages were pooled into one female sample, and ten RNA samples of genotypic male gonads at different stages were also pooled into one male sample. These two pooled samples were sequenced on the PacBio Sequel platform respectively (S7 Table). IsoSeq v3 (<https://github.com/ben-lerch/IsoSeq-3.0>) was used to cluster and polish isoforms with default parameters. TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>) was used to identify candidate coding regions within the final polished isoforms.

Moreover, nine RNA samples of genotypic female gonads and nine RNA samples of genotypic male gonads at 18, 22, 26, 30, 35, 40, 47, 58, and 70 dah were sequenced respectively via Illumina HiSeq platform at BGI-shenzhen, China (S8 Table). Illumina clean reads were mapped to PacBio full-length transcriptome, and Fragments per Kilobase Million (FPKM) was used to quantify gene expression, which was calculated via RSEM (Version 1.2.12) [67,68]. To normalize gene expression, if the FPKM was 0, it would be modified to 0.001.

### Production of temperature-dependent male offspring and RNA-seq

Genotypic sex determination (GSD) and temperature-dependent sex determination (TSD) were revealed to coexist in the hexaploid *C. gibelio* [18,23,42]. The temperature-dependent males, which were determined by the ambient temperature during larval development, contained no MSMs. To obtain temperature-dependent all-male offspring without MSMs, the gynogenetic larvae were raised at 32°C ( $\pm 1^\circ\text{C}$ ) since first feeding for 30 days as previously reported [18,32]. Meanwhile, the gynogenetic larvae from the same family were raised at normal temperature 20°C ( $\pm 1^\circ\text{C}$ ), which generated all-female offspring. The temperature-dependent all-male offspring and the corresponding all-female offspring were constructed in the strain A<sup>+</sup> of *C. gibelio*, as the sex of individuals in strain A<sup>+</sup> could be easily affected by rearing temperature [18].

Gonads sampled from 93 temperature-dependent males and 93 females were used for transcriptome analysis. And a total of 30, 30, 20, 10, 3 gonads were pooled for RNA isolation at the stage of 6 dah, 16 dah, 30 dah, 60 dah, and mature gonads with 1 year old respectively. The RNA samples of temperature-dependent males and the corresponding females were pooled into one sample, and sequenced on the PacBio Sequel platform (S10 Table).

### Comparative analysis

Clean reads of MSMs were mapped against the female *C. gibelio* genome (GenBank assembly accession: PRJNA546443), using Minimap2 with default parameters [69]. Self-python was used to extract matched sequences (MapQ >10), which were merged to blocks according to overlap region. The alignments between the identified blocks of each MSM and female *C. gibelio* genome were displayed by the Circos software (version: 0.69–6) [70]. Kimura distance analysis [71] was used to infer the sequence divergence between the MSMs and A chromosome sequence, based on calculating the pairwise divergence of the aligned hits. The aligned hits of each MSM were randomly selected in different length sections, and a total of 120 aligned hits were selected for each MSM. The Kimura value of each aligned hits were calculated by MEGA [72]. Pairwise comparisons among these three MSMs were performed by BLAST with default parameters.

### Construction of transposable element database and identification of repetitive elements

The assembly of *C. gibelio* genome (GenBank accession: PRJNA546443) was used to construct a specific transposable element (TE) database for hexaploid *C. gibelio* by a combinatory prediction of signature and homology as follows. During the process of the signature-based prediction, several programs were implemented for given classes of TEs with particular features. Full-length LTR retrotransposons were identified according to Ray et al. [73], with modifications. The candidate LTR elements were ultimately aligned with sequence identity parameters 0.9 by CD-HIT-EST [74] in order to reduce the redundancy. Helitron transposons were collected by HelitronScanner based on a two-layer local combinational variable (LCV) algorithm [75]. The input genomes were scanned using LCVs that were extracted by known Helitrons,



and putative Helitrons were drawn using the parameters “-ht 5 -tt 10” to avoid high false positives and missing true Helitrons. Non-autonomous DNA transposons were identified using MITE-Hunter [76] with the “-P 0.1” parameters. Autonomous non-LTR transposons were identified and classified using MGEScan-non-LTR with default parameters [77]. SINE transposons were predicted on the basis of structural features by using SINE-Finder [78], and then aligned using Needle in EMBOSS package (version 6.6.0) [79], with the output used for assigning the SINE elements to respective families if they shared 60% (or more) similarities.

The TEs identified from the way of signature-based prediction were then used to mask *C. gibelio* genome by RepeatMasker (<http://www.repeatmasker.org>, version 4.0.7) with WUBlast engine. The unmasked portion of genomes was subsequently used for homology-based prediction. Superfamilies of the putative TE elements were classified according to the conserved domains or the sequences of terminal invert repeats (TIR) and target site duplication (TSD). The open reading frame (ORF) within the full length LTR element, which was obtained by EMBOSS Getorf program [79], was used to search for known coding domains (e.g. Gag, Protease (PR), Reverse Transcriptase (RT), Ribonuclease H (RN), Integrase (INT), Envelope (ENV), Transposase (TR), etc.), through HMMER (version 3.1b) [80] with pHMMs model downloaded from GyDB (Gypsy database 2.0) [81]. Classification of LTR elements with complete gag-pol structures was based on the order of RT and INT, which was the distinction of two main LTR superfamilies *Gypsy* and *Copia* [82]. The rest of LTR elements were categorized into *LARDs* (> 4 kb) and *TRIMs* (< 4 kb). A homology search of The MITE elements against RepBase (version 23.07) was performed by RepeatMasker. TIRs and TSDs identified by MITE-Hunter were compared with the previously characterized TIRs and TSDs in plants and animals [83–85] through a custom Perl script. The output of RepeatMasker and TIR-TSD searching were the bases of MITEs classification. The unmasked MITEs with ambiguous (or unknown) TIRs and TSDs will be classified as unknowns. Classification of TEs identified by RepeatModeler was performed as described previously [73].

The constructed *C. gibelio*-specific TE database and the known metazoan repetitive database (RepBase 23.06) were used to identify repetitive elements on MSMs, using RepeatMasker (version: version open-4.0.6) with parameters -nolow -no\_is -norna -engine ncbi -parallel 1.

### Identification of genes on MSMs and their corresponding transcripts

To screen the MSMs-linked genes, two databases including gonadal full-length transcriptomes at ten developmental stages of hexaploid *C. gibelio* (S7 Table) and the coding sequence (CDS) pool of nine fish species were used as references. The coding sequence pool of nine species consisted of *C. gibelio* (GenBank accession: PRJNA546443), *C. auratus* (GenBank accession: PRJNA546444), *C. carpio* ([http://www.carpbase.org/download\\_home.php](http://www.carpbase.org/download_home.php)), *Ctenopharyngodon idellus* (*C. idellus*) (<http://www.ncgr.ac.cn/grasscarp/>), *Megalobrama amblycephala* (*M. amblycephala*) (<http://bream.hzau.edu.cn/page/species/download.html#1>), *Danio rerio* (*D. rerio*) ([ftp://ftp.ensembl.org/pub/release-96/fasta/danio\\_rerio/](ftp://ftp.ensembl.org/pub/release-96/fasta/danio_rerio/)), *Oryzias latipes* (*O. latipes*) ([ftp://ftp.ensembl.org/pub/release-96/fasta/oryzias\\_latipes/](ftp://ftp.ensembl.org/pub/release-96/fasta/oryzias_latipes/)), *Xiphophorus maculatus* (*X. maculatus*) ([ftp://ftp.ensembl.org/pub/release-96/fasta/xiphophorus\\_maculatus/](ftp://ftp.ensembl.org/pub/release-96/fasta/xiphophorus_maculatus/)) and *Oreochromis niloticus* (*O. niloticus*) ([ftp://ftp.ensembl.org/pub/release-96/fasta/oreochromis\\_niloticus/](ftp://ftp.ensembl.org/pub/release-96/fasta/oreochromis_niloticus/)), which were clustered using OrthoMCL based on an all-to-all BLASTP strategy with the default parameters [86,87]. The sequences of MSMs were mapped to the references via Minimap2 (MapQ  $\geq$  10 and aligned length  $\geq$  200 bp) and BLAST (identity >75% and aligned length  $\geq$  200 bp).

To identify the corresponding transcripts of genes identified by the CDS pool, the reference CDSs were mapped against the gonadal transcriptomes of *C. gibelio* (Fig 5A and S7 Table) via

BLAST. And the following criteria of BLAST were applied that cumulative identity percentage (CIP)  $\geq 60\%$  and cumulative alignment length percentage (CALP)  $\geq 70\%$ . CIP corresponds to the cumulative percentage of sequence identity obtained for all of the high scoring pairs (HSPs) ( $CIP = \sum ID \text{ by HSP/AL} * 100$ ). CALP represents the sum of the HSP aligned length (AL) for all of the HSPs divided by the length of the query sequence ( $CALP = \sum AL / \text{query length}$ ).

TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>) was used to identify potential coding sequence of the annotated genes. And the sequences of MSMs were aligned to the full-length coding sequence by BLAT (-t = dna -q = dna -oneOff = 2 -minIdentity = 80) to analyze the gene integrity.

All the transcripts were used for the subsequent analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG). MARS model of DEGseq2 (Version: 1.4.5) (P-value =  $1e^{-3}$ , zscore = 4, q-value = 0.001, ThresholdKind = 5) was used to detect differentially expressed genes (DEGs). Sex-biased transcripts at each developmental stage were identified separately from DEGs under the conditions:  $|\log_2 \text{FoldChange}| \geq 1$  and  $FDR \leq 0.05$  and the final number of the sex biased transcripts were calculated by discarded the duplicated transcripts.

### Sequence subtraction

A set of subtraction approaches were used to identify the potential male specific transcripts as previously described [88–90] with some modifications. Two transcriptomes were used as a reference for transcriptome subtraction, including one female gonadal transcriptome of ten developmental stages from strain DA (S7 Table) and one gonadal transcriptome of females and temperature-dependent males (without MSMs) from strain A<sup>+</sup> (S10 Table). After transcriptome subtraction, the remaining transcripts were mapped to the female reference genome of hexaploid *C. gibelio* [34] for genome subtraction. The MSMs-linked transcripts were mapped to the transcriptomes and genome for subtraction using BLAT [87], and the sequences were discarded when the aligned length/full length is over 90% and the identity is higher than 98%. After subtraction, the remaining transcripts were aligned to the sequences of MSMs via BLAST with default parameters, and the well-aligned sequences of MSMs (identity > 75% and aligned length  $\geq 200$  bp) were defined as gene fragments. Subsequently, these gene fragments were used for the second round of transcriptome/genome subtractions as described above. Gene fragments were annotated using the coding sequence of the female reference genome of hexaploid *C. gibelio* (GenBank accession: PRJNA546443), the non-redundant protein sequences database (<https://www.ncbi.nlm.nih.gov>), and the database of transcripts/splice variants (<http://www.ensembl.org>) orderly.

### RNA isolation and relative quantitative real-time PCR (qPCR)

Genotypic female and male gonads at 18, 22, 26, 30, and 35 dah were pooled respectively for total RNA isolation using SV Total RNA Isolation System (Promega). RNAs were reverse-transcribed via M-MLV reverse system, and qPCR was then performed on CFX96 Real-Time System (BioRad) with iQ SYBR Green Supermix (BioRad) as described previously [30].  *$\beta$ -actin* was used as the internal reference. Each sample was analyzed in triplicates, and the relative expression level of target gene was calculated with  $2^{-\Delta\Delta CT}$  method. The highest expression level of each gene fragment was used as control and defined as 1. PCR cycling conditions were: 95°C for 1 min; 40 cycles of 15 s at 95°C, 20 s at 58°C, and 30 s at 72°C in a 20  $\mu$ l reaction mix. PCR primers were designed following the subsequent rules. If the MSM fragment had homolog sequences (identity > 75%) in the reference genome without MSMs, we designed the PCR primers according to the MSM fragment and made the 3' end nucleotide of each primer locate at the different sites between MSM fragment and homologous sequence of reference genome.

If the MSM fragment had no homolog sequence in the reference genome without MSM, we randomly designed the PCR primers according to the MSM fragment. Sequences of PCR primers are given in [S11 Table](#).

## Supporting information

**S1 Fig. Schematic diagram of the relationship between male-specific sequence *Cg-M-s* and satellite repeat cluster *Cg-Ca-CL1*.** (A) *Cg-M-s* contains several intact and fragmental repeats of *Cg-Ca-CL1*. The sites of male-specific primers including *Cg-M-s-F* and *Cg-M-s-R* are marked by black arrows. (B) Sequence alignment between the consensus sequence of *Cg-Ca-CL1* and the repeats of *Cg-Ca-CL1* in *Cg-M-s*. (C) Peptide nucleic acid (PNA) probes used for fluorescence in situ hybridization (FISH). The sequences of PNA probes are indicated by yellow background. Each probe is labeled with three Cy3.

(TIF)

**S2 Fig. Co-localization of satellite repeat cluster *Cg-Ca-CL1* and male-specific sequence *Cg-M-s* in *C. gibelio*.** (A-H) The *Cg-Ca-CL1* probe and *Cg-M-s* probe were labeled with Biotin and Digoxin respectively, and red and green fluorescence were produced accordingly. FISH analysis was performed in metaphases of female *C. gibelio* (A-D) and male *C. gibelio* (E-H). Chromosomes were counterstained with DAPI and appeared blue. Scale bar = 5  $\mu\text{m}$ .

(TIF)

**S3 Fig. FISH analysis and PCR detection of satellite repeat cluster *Cg-Ca-CL1*.** (A, B) FISH analysis of satellite repeat cluster *Cg-Ca-CL1* (red) in metaphases of female *C. auratus* (A) and male *C. auratus* (B). (C) PCR assay of satellite repeat cluster *Cg-Ca-CL1* in *C. gibelio* and *C. Carassius*. ♀, female; ♂, male. Scale bar = 5  $\mu\text{m}$ .

(TIF)

**S4 Fig. Single chromosomal fluorescence microdissection.** (A, C) FISH detection on a male metaphase using PNA probes before (A) and after (C) microdissection. (B, D) The enlarged images of the white squares in (A) and (C), respectively. The red signals from PNA probes highlighted microchromosomes, and chromosomes were counterstained with DAPI and appeared blue. The white arrows indicate chromosome before (A and B) and after (C and D) isolation. Scale bar = 5  $\mu\text{m}$ .

(TIF)

**S5 Fig. PCR detection of the male-specific marker.** (A-C) Male-specific marker in 8 randomly-picked males and 8 randomly-picked females in the offspring as well as the parental individuals from family 1 (A), family 2 (B), and family 3 (C). ♀, female; ♂, male; M, maternal individual; P, paternal individual.

(TIF)

**S6 Fig. Assessments of MSM sequence assemblies.** (A, B) Number (A) and total length (B) of aligned blocks referring to the female genome. (C, D) Number (C) and total length (D) of aligned blocks referring to the full-length transcriptomes. Different colors represent different lengths of aligned blocks. The X axis represents different datasets including clean reads, corrected reads, contigs assembled by CANU, contigs assembled by SPAdes, and contigs assembled by SMARTdenovo.

(TIF)

**S7 Fig. Analysis of gene content.** (A) Gene content of three MSMs. (B) Number of clean reads containing gene sequences. The X axis represents the proportion of gene sequences. The

Y axis indicates the number of clean reads.  
(TIF)

**S8 Fig. qPCR analysis of gene fragments.** qPCR detection of gene fragments at early gonadal developmental stages including 18, 22, 26, 30, and 35 dah (days after hatch). The X axis represents the stages of gonad development. The Y axis represents the relative expression, and the highest expression level of each gene fragment was used as control and defined as 1. F, female; M, male.  
(TIF)

**S9 Fig. Schematic diagrams of the family establishment.** (A) Establishment of a family containing a high proportion of male offspring. (B) Establishment of a family with all-female offspring. ♀, female; ♂, male; (+), with the male-specific genetic marker; (-) without the male-specific genetic marker.  
(TIF)

**S1 Table. The consensus sequence of satellite DNA *Cg-Ca-CL1*.**  
(DOCX)

**S2 Table. Sequencing summary of male-specific microchromosomes.**  
(DOCX)

**S3 Table. The summary of the corrected sequences of MSMs.**  
(DOCX)

**S4 Table. The summary of sequence assembly of MSMs by CANU.**  
(DOCX)

**S5 Table. The summary of sequence assembly of MSMs by SPAdes.**  
(DOCX)

**S6 Table. The summary of sequence assembly of MSMs by SMARTdenovo.**  
(DOCX)

**S7 Table. PacBio sequencing summary of female and male gonads.**  
(DOCX)

**S8 Table. Illumina sequencing summary of female and male gonads.**  
(DOCX)

**S9 Table. Summary of 42 unique potential male-specific gene fragments.** # Red color indicates the gene fragments with a conserved coding sequence.  
(DOCX)

**S10 Table. PacBio sequencing summary of female and male gonads without MSMs.**  
(DOCX)

**S11 Table. Primers that used in this study.**  
(DOCX)

## Acknowledgments

We thank Fang Zhou and Yan Wang (the Center for Instrumental Analysis and Metrology, Institute of Hydrobiology, Chinese Academy of Science) for providing confocal and microdissection services.

## Author Contributions

**Conceptualization:** Xi-Yin Li, Jian-Fang Gui.

**Formal analysis:** Miao Ding, Zhi-Xuan Zhu, Jun-Hui Chen, Meng Lu, Qian Shi, Yang Wang, Zhi Li, Xin Zhao, Tao Wang, Wen-Xuan Du, Chun Miao, Tian-Zi Yao, Ming-Tao Wang, Xiao-Juan Zhang, Zhong-Wei Wang, Li Zhou.

**Funding acquisition:** Xi-Yin Li, Li Zhou, Jian-Fang Gui.

**Investigation:** Zhi-Xuan Zhu, Jun-Hui Chen, Meng Lu, Qian Shi, Yang Wang, Zhi Li, Xin Zhao, Tao Wang, Wen-Xuan Du, Chun Miao, Tian-Zi Yao, Ming-Tao Wang, Xiao-Juan Zhang, Zhong-Wei Wang, Li Zhou.

**Methodology:** Miao Ding.

**Resources:** Miao Ding, Xi-Yin Li.

**Writing – original draft:** Xi-Yin Li.

**Writing – review & editing:** Xi-Yin Li, Jian-Fang Gui.

## References

1. Neaves WB, Baumann P. Unisexual reproduction among vertebrates. *Trends Genet.* 2011; 27(3):81–88. <https://doi.org/10.1016/j.tig.2010.12.002> PMID: 21334090
2. Avise JC. Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proc Natl Aca Sci U S A.* 2015; 112(29):8867–8873. <https://doi.org/10.1073/pnas.1501820112> PMID: 26195735
3. Burke NW, Bonduriansky R. Sexual conflict, facultative asexuality, and the true paradox of sex. *Trends Ecol Evol.* 2017; 32(9):646–652. <https://doi.org/10.1016/j.tree.2017.06.002> PMID: 28651895
4. Hubbs CL, Hubbs LC. Apparent parthenogenesis in nature, in a form of fish of hybride origin. *Science.* 1932; 76:628–630. <https://doi.org/10.1126/science.76.1983.628> PMID: 17730035
5. Warren WC, García-Pérez R, Xu S, Lampert KP, Chalopin D, Stöck M, et al. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nat Ecol Evol.* 2018; 2(4):669–679. <https://doi.org/10.1038/s41559-018-0473-y> PMID: 29434351
6. Schlupp Ingo. The evolutionary ecology of gynogenesis. *Annu Rev Ecol Evol Syst.* 2005; 36(1):399–417. <https://doi.org/10.1146/annurev.ecolsys.36.102003.152629>
7. Lutes AA, Neaves WB, Baumann DP, Wieggraebe W, Baumann P. Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. *Nature.* 2010; 464(7286):283–286. <https://doi.org/10.1038/nature08818> PMID: 20173738
8. Gui J, Zhou L. Genetic basis and breeding application of clonal diversity and dual reproduction modes in polyploid *Carassius auratus gibelio*. *Sci China Life Sci.* 2010; 53(4):409–415. <https://doi.org/10.1007/s11427-010-0092-6> PMID: 20596906
9. Lavanchy G, Schwander T. Hybridogenesis. *Curr Biol.* 2019; 29(1):R9–R11. <https://doi.org/10.1016/j.cub.2018.11.046> PMID: 30620918
10. Muller HJ. The relation of recombination to mutational advance. *Mutat Res.* 1964; 1(1):2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8) PMID: 14195748
11. Strotz LC, Simões M, Girard MG, Breikreuz L, Kimmig J, Lieberman BS. Getting somewhere with the Red Queen: chasing a biologically modern definition of the hypothesis. *Biol Lett.* 2018; 14(5):20170734. <https://doi.org/10.1098/rsbl.2017.0734> PMID: 29720444
12. Lynch M, Gabriel W. Mutation load and the survival of small populations. *Evolution.* 1990; 44:1725–1737. <https://doi.org/10.1111/j.1558-5646.1990.tb05244.x> PMID: 28567811
13. Loewe L, Lamatsch DK. Quantifying the threat of extinction from Muller’s ratchet in the diploid Amazon molly (*Poecilia formosa*). *BMC Evol Biol.* 2008; 8(1):88. <https://doi.org/10.1186/1471-2148-8-88> PMID: 18366680
14. Bi K, Bogart JP. Time and time again: unisexual salamanders (genus *Ambystoma*) are the oldest unisexual vertebrates. *BMC Evol Biol.* 2010; 10:238. <https://doi.org/10.1186/1471-2148-10-238> PMID: 20682056

15. Li XY, Zhang XJ, Li Z, Hong W, Liu W, Zhang J, et al. Evolutionary history of two divergent *Dmrt1* genes reveals two rounds of polyploidy origins in gibel carp. *Mol Phylogenet Evol.* 2014; 78:96–104. <https://doi.org/10.1016/j.ympev.2014.05.005> PMID: 24859683
16. Liu XL, Jiang FF, Wang ZW, Li XY, Li Z, Zhang XJ, et al. Wider geographic distribution and higher diversity of hexaploids than tetraploids in *Carassius* species complex reveal recurrent polyploidy effects on adaptive evolution. *Sci Rep.* 2017; 7(1):5395. <https://doi.org/10.1038/s41598-017-05731-0> PMID: 28710383
17. Bogart JP, Bogart J. A family study to examine clonal diversity in unisexual salamanders (genus *Ambystoma*). *Genome.* 2019; 62(8):549–561. <https://doi.org/10.1139/gen-2019-0034> PMID: 31172800
18. Li XY, Liu XL, Zhu YJ, Zhang J, Ding M, Wang MT, et al. Origin and transition of sex determination mechanisms in a gynogenetic hexaploid fish. *Heredity.* 2018; 121(1):64–74. <https://doi.org/10.1038/s41437-017-0049-7> PMID: 29391565
19. Jiang FF, Wang ZW, Zhou L, Jiang L, Zhang XJ, Apalikova OV, et al. High male incidence and evolutionary implications of triploid form in northeast Asia *Carassius auratus* complex. *Mol Phylogenet Evol.* 2013; 66(1):350–359. <https://doi.org/10.1016/j.ympev.2012.10.006> PMID: 23099150
20. Ahmad SF, Martins C. The modern view of B chromosomes under the impact of high scale omics analyses. *Cells.* 2019; 8(2):156. <https://doi.org/10.3390/cells8020156> PMID: 30781835
21. Camacho JPM. B chromosomes. Elsevier Academic Press. 2005. pp. 223–286. <https://doi.org/10.1016/B978-012301463-4/50006>
22. Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN. Evolution and biology of supernumerary B chromosomes. *Cell Mol Life Sci.* 2014; 71(3):467–478. <https://doi.org/10.1007/s00018-013-1437-7> PMID: 23912901
23. Li XY, Zhang QY, Zhang J, Zhou L, Li Z, Zhang XJ, et al. Extra microchromosomes play male determination role in polyploid gibel carp. *Genetics.* 2016; 203(3):1415–1424. <https://doi.org/10.1534/genetics.115.185843> PMID: 27017622
24. Clark FE, Conte MA, Kocher TD. Genomic characterization of a B chromosome in Lake Malawi cichlid fishes. *Genes.* 2018; 9(12). <https://doi.org/10.3390/genes9120610> PMID: 30563180
25. Banaei-Moghaddam AM, Martis MM, Macas J, Gundlach H, Himmelbach A, Altschmied L, et al. Genes on B chromosomes: old questions revisited with new tools. *BBA Gene Regul Mech.* 2015; 1849(1):64–70. <https://doi.org/10.1016/j.bbagrm.2014.11.007> PMID: 25481283
26. Huang W, Du Y, Zhao X, Jin W. B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC Plant Biol.* 2016; 16:88. <https://doi.org/10.1186/s12870-016-0775-7> PMID: 27083560
27. Ma W, Gabriel TS, Martis MM, Gursinsky T, Schubert V, Vrána J, et al. Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its A chromosome paralog. *New Phytol.* 2017; 213(2):916–928. <https://doi.org/10.1111/nph.14110> PMID: 27468091
28. Ahmad SF, Jehangir M, Cardoso AL, Wolf IR, Martins C. B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. *BMC Genomics.* 2020; 21(1):656. <https://doi.org/10.1186/s12864-020-07072-1> PMID: 32967626
29. Scharti Manfred, Nanda Indrajit. Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish. *Nature.* 1995; 373(6509):68–68
30. Li XY, Liu XL, Ding M, Li Z, Zhou L. A novel male-specific SET domain-containing gene *setdm* identified from extra microchromosomes of gibel carp males. *Sci Bull.* 2017; 62:528–536. <https://doi.org/10.1016/j.scib.2017.04.002>
31. Zhao X, Li Z, Ding M, Wang T, Wang MT, Miao C, et al. Genotypic males play an important role in the creation of genetic diversity in gynogenetic gibel carp. *Front Genet.* 2021. <https://doi.org/10.3389/fgene.2021.691923> PMID: 34122529
32. Zhu YJ, Li XY, Zhang J, Li Z, Ding M, Zhang XJ, et al. Distinct sperm nucleus behaviors between genotypic and temperature-dependent sex determination males are associated with replication and expression-related pathways in a gynogenetic fish. *BMC Genomics.* 2018; 19(1):437. <https://doi.org/10.1186/s12864-018-4823-6> PMID: 29866041
33. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics.* 2013; 29(6):792–793. <https://doi.org/10.1093/bioinformatics/btt054> PMID: 23376349
34. Wang Y, Li XY, Wu B, Xu M, Wang ZW, Li Z, et al. Tetraploid and hexaploid genomes reveal evolutionary insights of recurrent autotripleidy and reproduction success in a polyploid species complex. *Natl Sci Rev.* 2021. UnderReview

35. Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, et al. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genet.* 2011; 7(8): e1002203. <https://doi.org/10.1371/journal.pgen.1002203> PMID: 21876673
36. Jehangir M, Ahmad SF, Cardoso AL, Ramos E, Martins C. De novo genome assembly of the cichlid fish *Astatotilapia latifasciata* reveals a higher level of genomic polymorphism and genes related to B chromosomes. *Chromosoma.* 2019; 128(2):81–96. <https://doi.org/10.1007/s00412-019-00707-7> PMID: 31115663
37. Valente GT, Conte MA, Fantinatti BE, Cabral-de-Mello DC, Carvalho RF, Vicari MR, et al. Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol.* 2014; 31(8):2061–2072. <https://doi.org/10.1093/molbev/msu148> PMID: 24770715
38. Choleva L, Janko K. Rise and persistence of animal polyploidy: evolutionary constraints and potential. *Cytogenet Genome Res.* 2013; 140(2–4):151–70. <https://doi.org/10.1159/000353464> PMID: 23838539
39. Zhou L, Gui J. Natural and artificial polyploids in aquaculture. *Aquaculture and Fisher.* 2017; 2(3):103–11. <https://doi.org/10.1016/j.aaf.2017.04.003>
40. Spoelhof JP, Keeffe R, McDaniel SF. Does reproductive assurance explain the incidence of polyploidy in plants and animals? *New Phytol.* 2020; 227(1):14–21. <https://doi.org/10.1111/nph.16396> PMID: 31883115
41. Luo J, Gao Y, Ma W, Bi Xy, Wang Sy, Wang J, et al. Tempo and mode of recurrent polyploidization in the *Carassius auratus* species complex (Cypriniformes, Cyprinidae). *Heredity.* 2014; 112(4):415–427. <https://doi.org/10.1038/hdy.2013.121> PMID: 24398883
42. Li XY, Gui JF. Diverse and variable sex determination mechanisms in vertebrates. *Sci China Life Sci.* 2018; 61(12):1503–1514. <https://doi.org/10.1007/s11427-018-9415-7> PMID: 30443862
43. Beladjal L, Vandekerckhove T, Muysen B, Heyrman JCJ, Mertens J. B-chromosomes and male-biased sex ratio with paternal inheritance in the fairy shrimp *Branchipus schaefferi* (Crustacea, Anostraca). *Heredity.* 2002; 88:356–360. <https://doi.org/10.1038/sj.hdy.6800061> PMID: 11986871
44. Camacho JPM, Schmid M, Cabrero J. B chromosomes and sex in animals. *Sex Dev.* 2011; 5(3):155–166. <https://doi.org/10.1159/000324930> PMID: 21430369
45. Rafael Bueno, Noleto Marcelo, Ricardo Vicari, Marta Margarete, Cestari Roberto. Variable B chromosomes frequencies between males and females of two species of pufferfishes (*Tetraodontiformes*). *Rev Fish Biol Fisher.* 2011; 22(1): 343–349. <https://doi.org/10.1007/s11160-011-9231-9>
46. Clark FE, Kocher TD. Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Sci Rep.* 2019; 9(1). <https://doi.org/10.1038/s41598-019-55774-8> PMID: 31882583
47. Beukeboom LW, Werren JH. Transmission and expression of the parasitic paternal sex ratio (PSR) chromosome. *Heredity.* 1993; 70(4):437–443. <https://doi.org/10.1038/hdy.1993.61>
48. Werren JH, Stouthamer R. PSR (paternal sex ratio) chromosomes: the ultimate selfish genetic elements. *Genetica.* 2003; 117(1). <https://doi.org/10.1023/a:1022368700752> PMID: 12656576
49. Bao L, Tian C, Liu S, Zhang Y, Elasmad A, Yuan Z, et al. The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *BMC Biol.* 2019; 17(1):6. <https://doi.org/10.1186/s12915-019-0627-7> PMID: 30683095
50. Pan Q, Feron R, Yano A, Guyomard R, Jouanno E, Vigouroux E, et al. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genet.* 2019; 15(8):e1008013. <https://doi.org/10.1371/journal.pgen.1008013> PMID: 31437150
51. Hobza R, Kubat Z, Cegan R, Jesionek W, Vyskot B, Kejnovsky E. Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Res.* 2015; 23(3):561–570. <https://doi.org/10.1007/s10577-015-9496-2> PMID: 26474787
52. Garrido-Ramos MA. Satellite DNA: an evolving topic. *Genes.* 2017; 8(9). <https://doi.org/10.3390/genes8090230> PMID: 28926993
53. Lower SS, McGurk MP, Clark AG, Barbash DA. Satellite DNA evolution: old ideas, new approaches. *Curr Opin Genet Dev.* 2018; 49:70–78. <https://doi.org/10.1016/j.gde.2018.03.003> PMID: 29579574
54. Gregory D, Hurst D, Werren JH. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet.* 2001; 2(8): 597. <https://doi.org/10.1038/35084545> PMID: 11483984
55. Benetta ED, Antoshechkin I, Yang T, Nguyen H, Akbari OS. Genome elimination mediated by gene expression from a selfish chromosome. *Science Adv.* 2020; 6(14). <https://doi.org/10.1126/sciadv.aaz9808> PMID: 32284986
56. Hackstein JH, Hochstenbach R, Hauschteck-Jungen E, Beukeboom LW. Is the Y chromosome of *Drosophila* an evolved supernumerary chromosome? *BioEssays.* 1996; 18(4):317–323. <https://doi.org/10.1002/bies.950180410> PMID: 8967900

57. Bernardo Carvalho A, Koerich LB, Clark AG. Origin and evolution of Y chromosomes: *Drosophila* tales. *Trends Genet.* 2009; 25(6):270–277. <https://doi.org/10.1016/j.tig.2009.04.002> PMID: 19443075
58. FraiSse C, Picard M, Vicoso B. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat Commun.* 2017; 8(1):1486. <https://doi.org/10.1038/s41467-017-01663-5> PMID: 29133797
59. Conte MA, Clark FE, Roberts RB, Xu L, Tao W, Zhou Q, et al. Origin of a giant sex chromosome. *Mol Biol Evol.* 2020. 13; 38(4):1554–1569. <https://doi.org/10.1093/molbev/msaa319> PMID: 33300980
60. Zhu HP, Gui JF. Identification of genome organization in the unusual allotetraploid form of *Carassius auratus gibelio*. *Aquaculture.* 2007; 265:109–117. <https://doi.org/10.1016/j.aquaculture.2006.10.02651>
61. Lu M, Li XY, Li Z, Du WX, Zhou L, Wang Y, et al. Regain of sex determination system and sexual reproduction ability in a synthetic octoploid male fish. *Sci China Life Sci.* 2021; 64(1):77–87. <https://doi.org/10.1007/s11427-020-1694-7> PMID: 32529288
62. Yi MS, Li YQ, Liu JD, Zhou L, Yu QX, Gui JF. Molecular cytogenetic detection of paternal chromosome fragments in allogynogenetic gibel carp, *Carassius auratus gibelio* Bloch. *Chromosome Res.* 2003; 11(7):665–671. <https://doi.org/10.1023/a:1025985625706> PMID: 14606628
63. Dernburg AF. Microdissection of *Drosophila* polytene chromosomes for DOP-PCR. *Cold Spring Harbor Protocols.* 2012; 2012(3):376–379. <https://doi.org/10.1101/pdb.prot067298> PMID: 22383634
64. Deleye L, Vander Plaetsen A-S, Weymaere J, Deforce D, Van Nieuwerburgh F. Short tandem repeat analysis after whole genome amplification of single B-lymphoblastoid cells. *Sci Rep.* 2018; 8(1):1255. <https://doi.org/10.1038/s41598-018-19509-5> PMID: 29352241
65. Koren Sergey, Walenz Brian P, Berlin Konstantin, et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. 2017; 27(5):722–736. <https://doi.org/10.1101/gr.215087.116> PMID: 28298431
66. Walker B, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE.* 2015; e112963. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509
67. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011; 12:323. <https://doi.org/10.1186/1471-2105-12-323> PMID: 21816040
68. Yu P, Wang Y, Yang WT, Li Z, Gui JF. Upregulation of the PPAR signaling pathway and accumulation of lipids are related to the morphological and structural transformation of the dragon-eye goldfish eye. *Sci China Life Sci.* 2021. <https://doi.org/10.1007/s11427-020-1814-1> PMID: 33428077
69. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018; 34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242
70. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19(9):1639–1645. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911
71. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Mol Biol Evol.* 1980; 16(2):111–120. <https://doi.org/10.1007/BF01731581> PMID: 7463489
72. Sudhir K, Glen S, Koichiro T. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; 33(7):1870–4. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
73. Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li LT, et al. Genome of the long-living sacred lotus (*Nelumbo nucifera Gaertn.*). *Genome Biol.* 2013; 14(5):R41. <https://doi.org/10.1186/gb-2013-14-5-r41> PMID: 23663246
74. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010; 26(5):680–682. <https://doi.org/10.1093/bioinformatics/btq003> PMID: 20053844
75. Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc Natl Acad U S A.* 2014; 111(28):10263–10268. <https://doi.org/10.1073/pnas.1410068111> PMID: 24982153
76. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 2010; 38(22):e199–e199. <https://doi.org/10.1093/nar/gkq862> PMID: 20880995
77. Rho M, Tang H. MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. *Nucleic Acids Res.* 2009; 37(21):e143. <https://doi.org/10.1093/nar/gkp752> PMID: 19762481
78. Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in



- plant genomes. *Plant Cell*. 2011; 23(9):3117–3128. <https://doi.org/10.1105/tpc.111.088682> PMID: 21908723
79. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000; 16(6):276–277. [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2) PMID: 10827456
  80. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011; 7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195> PMID: 22039361
  81. Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res*. 2010. <https://doi.org/10.1093/nar/gkq1061> PMID: 21036865
  82. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nature Rev Genet*. 2007; 8(12):973–982 <https://doi.org/10.1038/nrg2165> PMID: 17984973
  83. Yuan YW, Wessler SR. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad U S A*. 2011; 108(19):7884–7889. <https://doi.org/10.1073/pnas.1104208108> PMID: 21518873
  84. Böhne A, Zhou Q, Darras A, Schmidt C, Scharf M, Galiana-Arnoux D, et al. Zisupton—a novel superfamily of DNA transposable elements recently active in fish. *Mol Biol Evol*. 2012; 29(2):631–645. <https://doi.org/10.1093/molbev/msr208> PMID: 21873630
  85. Zhou M, Tao G, Pi P, Zhu Y, Bai Y, Meng X. Genome-wide characterization and evolution analysis of miniature inverted-repeat transposable elements (MITEs) in moso bamboo (*Phyllostachys heterocycla*). *Planta*. 2016; 244(4):775–787. <https://doi.org/10.1007/s00425-016-2544-0> PMID: 27160169
  86. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003; 13(9):2178–2189. <https://doi.org/10.1101/gr.1224503> PMID: 12952885
  87. Kent WJ. BLAT—The BLAST-like alignment tool. *Genome Res*. 2002; 12(4):656–664. <https://doi.org/10.1101/gr.229202> PMID: 11932250
  88. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*. 2014; 514(7497):494. <https://doi.org/10.1038/nature13206> PMID: 24759411
  89. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 2014; 508(7497):488–493. <https://doi.org/10.1038/nature13151> PMID: 24759410
  90. Mahajan S, Bachtrog D. Convergent evolution of Y chromosome gene content in flies. *Nat Commun*. 2017; 8(1):785. <https://doi.org/10.1038/s41467-017-00653-x> PMID: 28978907