PLOS ONE

RESEARCH ARTICLE

The Genome of the Trinidadian Guppy, *Poecilia reticulata*, and Variation in the Guanapo Population

Axel Künstner[1,2¤a]*, Margarete Hoffmann[1], Bonnie A. Fraser[1¤b], Verena A. Kottler[1¤c], Eshita Sharma[1¤d], Detlef Weigel[1], Christine Dreyer[1]

1 Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany,
2 Guest Group Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

¤a Current address: Group of Medical Systems Biology, Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany
¤b Current address: School of Life Sciences, University of Sussex, Falmer, Brighton, United Kingdom
¤c Current address: Department of Physiological Chemistry, University of Würzburg, Würzburg, Germany
¤d Current address: Bioinformatics and Statistical Genetics Core, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom
* axel.kuenstner@uni-luebeck.de

OPEN ACCESS

Citation: Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, et al. (2016) The Genome of the Trinidadian Guppy, *Poecilia reticulata*, and Variation in the Guanapo Population. PLoS ONE 11(12): e0169087. doi:10.1371/journal.pone.0169087

Editor: Peng Xu, Xiamen University, CHINA

Received: June 15, 2016

Accepted: December 12, 2016

Published: December 29, 2016

Data Availability Statement: All short read data generated during this project have been made available via the NCBI Short Read Archive (NCBI SRA study accession SRP038017). Accession numbers for genomic libraries can be found in S1 Table, and for population samples in S2 Table. Accession numbers for the samples used to estimate mutation rates are SRR1503964 and SRR1503965 for the parents (female, male), and SRR1503967-SRR1503971 for the five F1. The genome assembly was submitted to NCBI under BioProject PRJNA238429. The entire project has been deposited at DDBJ/EMBL/GenBank under the

Abstract

For over a century, the live bearing guppy, *Poecilia reticulata*, has been used to study sexual selection as well as local adaptation. Natural guppy populations differ in many traits that are of intuitively adaptive significance such as ornamentation, age at maturity, brood size and body shape. Water depth, light supply, food resources and predation regime shape these traits, and barrier waterfalls often separate contrasting environments in the same river. We have assembled and annotated the genome of an inbred single female from a high-predation site in the Guanapo drainage. The final assembly comprises 731.6 Mb with a scaffold N50 of 5.3 MB. Scaffolds were mapped to linkage groups, placing 95% of the genome assembly on the 22 autosomes and the X-chromosome. To investigate genetic variation in the population used for the genome assembly, we sequenced 10 wild caught male individuals. The identified 5 million SNPs correspond to an average nucleotide diversity (π) of 0.0025. The genome assembly and SNP map provide a rich resource for investigating adaptation to different predation regimes. In addition, comparisons with the genomes of other Poeciliid species, which differ greatly in mechanisms of sex determination and maternal resource allocation, as well as comparisons to other teleost genera can begin to reveal how live bearing evolved in teleost fish.

Introduction

The advent of cost-effective and high-throughput sequencing technologies has brought genomics into the field of evolutionary biology. In combination with progress in computational biology, next generation sequencing enables researchers to study closely related populations or species using full genome sequencing (e.g. [1], [2]), to investigate how genomic variation is structured among and within species, and to determine how genomic variation relates to phenotypic

accession AZHG00000000. The version described in this paper is version AZHG01000000. The assembly comprises linkage groups 1 to 23, all unplaced sequences, and the mitochondrial genome. The scaffold accessions appear in the WGS_SCFLD line at the bottom of the WGS master record, AZHG00000000 (CM002706-CM002728 = chromosomes, KK214999-KK218026 = scaffolds).

Funding: This work was supported by a Gottfried Wilhelm Leibniz Award of the Deutsche Forschungsgemeinschaft and the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declare no competing financial interests.

and environmental variation (e.g. [3–5]). With this information we can now address fundamental questions in evolutionary biology, such as how populations adapt to new or changing environments, whether traits are caused by few genes of large effect or many genes of small effect, and what the relative importance of demography and selection are in shaping variation.

The first crucial step required to address these questions is the construction of a high-quality reference genome. While, many 'reference-free' methods for studying genetic variation exist (e.g. *de novo* RAD-seq, *de novo* transcriptome assembly [6, 7]), they have substantial inherent shortcomings, such as limited information about linkage among genes, ortholog-paralog ambiguity, difficulties in differentiating gene loss from insufficient sampling, and mis-annotation [8]. Some of these issues will likely be particularly troublesome in teleost fishes because a whole genome duplication event occurred early in diversification of teleosts, and various lineages have since independently undergone additional genome duplications.

In addition to the basal whole genome duplication being followed by rediploidization, which in turn leads to the loss of extra copies, functional diversification and neo-functionalization of paralogous gene copies are prominent features of genome evolution in teleosts [9, 10] (reviewed in [11]). For this reason, and because of the gigantic number of diverse species in this class, teleost fishes provide a rich resource for studying the evolution of molecular function of genes and whole genome organization. Consequently, a contiguous, annotated reference genome assembly is yet another milestone in fish evolutionary biology [12].

Here we focus on the Trinidadian guppy (*Poecilia reticulata*) as a premier vertebrate model for the study of natural variation and local adaptation. The Trinidadian guppy is a small, live bearing freshwater fish with marked phenotypic dimorphism between the sexes and an XX/XY sex-determination system. In contrast to the inconspicuous reticulate pattern of the larger females, the complex patterns of adult males vary greatly within and between different

Fig 1. Guppy reference genome strain. Female (top) and male (bottom) from the inbred Guanapo strain.

doi:10.1371/journal.pone.0169087.g001

natural populations (Fig 1). Not surprisingly, the guppy was one of the first vertebrate species for which sex-linked inheritance of traits could be demonstrated [13].

Comparative studies have demonstrated that guppies have convergently evolved similar adaptations to life with or without predators in the different rivers that drain the slopes of the Northern Range Mountains in Trinidad. These adaptations include male coloration, mating and schooling behavior, and life history traits [14, 15]. Population genetic studies have shown that natural low predation populations derive from independent colonization events by ancestral down-stream and high predation populations within each river drainage [16, 17]. The predation regime is regarded as a major driving force for adaptation. Some predators in downstream localities prey predominantly on large, mature guppies, favoring the evolution of male guppies with less coloration and covert mating tactics. Barriers exclude these predators from upstream localities. The few predators found there, predominantly the killifish *Rivulus hartii*, eat fewer guppies and tend to prey on smaller, immature fish. Under these conditions, sexual selection in the form of female preferences prevails over natural selection by predators, and males evolve to be more brightly colored [14, 15, 18]. When guppies are transplanted from high to low predation localities, measurable character shifts occur within three to ten generations [19, 20], indicative of selection from standing natural variation in the founder populations [14, 18, 21, 22]. Standing genetic variation provides a rich repertoire of alleles that allows for selection on beneficial alleles that enhance fitness in a changing environment. Owing to the maintenance of ancient genetic variation or current gene flow, beneficial alleles have already passed a 'selection filter' [23]. Furthermore, adaptive alleles may be maintained at high frequencies due to balancing selection. Negative frequency dependent selection has been shown to be operating in guppies, where males with rare color patterns have a higher probability of surviving [24] and a higher mating advantage [25].

The guppy has a haploid complement of 23 chromosomes, with an XX/XY sex-determination system. Early in the 20th century, Winge used crosses among inbred lines of guppies to show that many male color patterns are Y linked [13, 26]. Later investigators used quantitative genetic approaches to show that sires have much larger effects than dams on the inheritance of male size and coloration [27–29] which again argued for these traits being controlled in part by Y-linked genes. More recently, Tripathi and colleagues [30] used a classic F_2 quantitative genetic framework with approximately 800 markers and 2,000 individuals for QTL mapping of size and coloration traits. They found several regions associated with both male size and color to be linked to the sex-determining locus, but also detected a number of additional QTL on different autosomes. Models of sex chromosome evolution predict that genes that benefit only the heterogametic sex, such as genes for conspicuous color patterns, will have increased evolutionary fitness if physically linked to the sex-determining locus through suppression of recombination [31]. Investigation of sex chromosome structure from various guppy populations and synaptonemal complex measurements have revealed polymorphisms between the X and Y chromosomes [32]. Guppy sex chromosomes are considered relatively young because the X and Y chromosomes are in most populations not morphologically distinct, in spite of there being many genes apparently linked to the non-recombining portion of the Y chromosome [33–35]. Consequently extended pseudoautosomal regions continue to be exchanged between X and Y [33, 35].

Both the evolution of sex determination in this species, and the genetic basis of local adaptation would greatly benefit from a high-quality genome assembly. We have therefore sequenced and assembled a female reference genome, using a combination of paired-end, mate-pair, and fosmid libraries. The vast majority of assembly scaffolds was then placed by genetic linkage on the 22 autosomes and the X-chromosome. Additionally, we describe genetic variation found in the reference genome's source population, a high-predation site from the Guanapo River in North West Trinidad.

Material and Methods

Genome assembly and annotation

Founder fish sampling and fish housing. Founder fish for inbreeding were first-generation lab-reared guppies from the Lower Guanapo River (Twin Bridge North West Trinidad, PS 91100 77800) where they are neither endangered nor protected. The specimens were kindly donated by Dr. David Reznick, UC Riverside, in 2009. Fish collection and export of these fish was approved by the Ministry of Agriculture, Land and Marine Resources, Republic of Trinidad and Tobago, conforming to their legislation. Since their collection progeny was kept and bred at the Max Planck Institute for Developmental Biology, Tübingen, according to German legislation. The facility was approved by the Regierungspräsidium Tübingen, registration number 35/9185.46. Fish required for preparation of DNA were anesthetized with a lethal dose of MS222 before being stored in 95% ethanol.

Genome sequencing. DNA from a 5th-generation female was used to prepare Illumina paired-end libraries with insert sizes of 240 to 460 bp and DNA from female offspring of the same lineage at generations 6 to 8 (six individuals total) was used to construct Roche/Illumina hybrid mate-pair libraries of 3 to 20 kb length; for details see [36]. Briefly, mate-pair libraries were prepared by ligating a circularization adaptor (Roche) and fragments were selected. The fragments were then circularized by Cre Recombinase, following the Roche protocol for 454 sequencing. Linear DNA was digested before circularized DNA was fragmented. Illumina paired-end standard adaptors were ligated onto these fragments, mate-pairs (of 180 to 480 bp length) were amplified and sequenced from both ends on the Illumina GA II platform. Long-jump 40 kb fosmid libraries were constructed from a single female offspring at generation 8 using the Nx 40 kb mate-pair cloning kit from Lucigen (Middleton, USA). Fosmid clones were amplified, fosmid DNA extracted in bulk, digested with *BfaI* and end fragments of 8–9 kb size (vector including insert ends) were selected. After recircularization, and digestion of linear DNA, mate-pairs were amplified by PCR and sequenced on the Illumina HiSeq2000 platform. See S1 Table for details about library sizes and sequencing yield.

Read filtering and quality trimming. All genomic libraries were retrieved and converted to FASTQ format using the *import* and *convert* commands in SHORE version 0.7.1 [37]. To remove PCR duplicates for a non-random fragment representation, each library was scanned with the *filterPCRdupl.pl* script (Version 1.01) included in ConDeTri version 2.0 [38]. For paired-end libraries, the first 50 bp of both reads of a pair were compared, and for mate-pair libraries the first 35 bp.

Mate-pair libraries were screened for the following adapter sequences:

5′-TCGTATAACTTCGTATAATGTATGCTATACGAAGTTATTACG-3′
5′-CGTAATAACTTCGTATAGCATACATTATACGAAGTTATACGA-3′

Screening was conducted using CUTADAPT version 1.1 [39], with an error rate of 0.15, an overlap of 6 bp, minimum read length of 35 bp, and matching wildcards. The screening resulted in two sets of mate-pairs: one set that contained parts of the adapter sequences and the other set without any adapter sequence. All other libraries were filtered for low quality bases using ConDeTri version 2.0 (using default parameters for paired-end libraries and with parameters–rmN–hq 20 –minLen 25 for mate-pair libraries).

Estimating mate-pair library insert sizes. To make use of mate-pair library information even in the absence of adapter sequences in the reads we first prepared a draft assembly. For this assembly SOAPdenovo version 1.05 [40] (kmer size 27, -d 1, -D 2, -F) was used. In detail, all paired-end libraries were used for contig building, but only mate-pair libraries with adapter sequence were included; these mate-pair reads were added stepwise according to their insert size (smallest to longest) for eight rounds of scaffolding. Mate-pair reads without adapter

sequence were mapped to the assembly using BWA version 0.6.1 [41] with default parameters. Mate-pairs that mapped within 1 kb distance of each other were excluded from further assembly steps to prevent possible paired-end contamination in the data. Insert sizes for the remaining reads were estimated per library based on distances between the reads.

Fosmid 40 kb insert library. The fosmid 40 kb library was treated slightly differently from the mate-pair libraries. First, *phiX* sequences were removed by mapping the library to the *phiX* genome sequence using BWA with default parameters. Then, reads were filtered for PCR duplicates and low quality bases, similar to the mate-pair libraries. Finally, reads were trimmed from the 3' end to keep 50 bp per read due to the low quality towards the 3' end of the reads. Insert sizes of the filtered and trimmed reads were estimated as described above (using the SOAPDENOVO assembly).

Genome assembly. *T*he final genome assembly was built with ALLPATHS-LG version 43668 [42] using default parameters. Due to a quality correction step in the ALLPATHS-LG pipeline, we used PCR filtered libraries for the paired-end and mate-pair libraries, and PCR-filtered and quality-trimmed fosmid libraries. Overlapping paired-end libraries were used for contig building (insert sizes of 240 and 270 bp, library ID 1–5). Longer insert paired-end libraries (insert size 460 bp, library ID 6–8), mate-pair libraries (ID 9–19), and the fosmid library (ID 20) were incorporated in the scaffold building process. ALLPATHS-LG was used to estimate the heterozygosity rate of the reference genome sample.

After assembly, all paired-end reads were mapped back to the genome assembly using BWA version 0.6.2 to estimate the proportion of the genome that was not covered by paired-end reads. Only a very small proportion of the genome assembly was not covered by paired-end reads (224,382 bp or 0.03%). Additionally, the assembly was screened for regions of runs of Ns. These regions occur during the scaffolding process and denote that sequence information between two contigs might be missing.

Contamination removal. To exclude cross contamination in the assembly from other organisms, we aligned the final genome assembly against the NCBI *nt* database (BLASTN version 2.2.21, e-value cutoff 10^{-5} [43]), reporting the best hit only. Scaffolds with hits only against non-vertebrate organisms were excluded from the assembly. This strategy excluded eight scaffolds with combined length of 27,088 bp (0.004% of the assembly) from the final assembly. Additionally, a more stringent contamination screen for adaptors was performed by NCBI resulting in 1,260 bp of potential adaptor sequences that were removed from the assembly. Visual inspection did not reveal any signs of mis-assemblies in these regions.

Genetics map integration. Scaffolds were anchored on linkage groups using a genetic linkage map built from 5,493 RAD-seq markers (for details about the linkage map see S1 File). Markers were aligned against the assembled genome (BLASTN version 2.2.27+, e-value $< 10^{-20}$ [44]) and only markers with unique hits were used to anchor scaffolds using the method described in [1]. Adjacent scaffolds were separated by a character string of 100 Ns. Scaffolds that could not be reliably anchored to one of the linkage groups were grouped into LG Un.

Assembly validation. Guppy expressed sequence tags (ESTs), were downloaded from the NCBI EST database (accessed 2013-09-25) and blasted (BLASTN version 2.2.27+, e-value $< 10^{-10}$ [44]) against the draft genome assembly. Additionally, a set of 454 transcriptome sequences [45] was downloaded (http://www.bio.fsu.edu/kahughes/Databases.html) and blasted (BLASTN, e-value $< 10^{-10}$) against the draft genome sequence. See S1 File for further assembly validation steps.

Repeat content. A guppy specific repeat library was built using the draft guppy genome assembly and REPEATMODELER version open-1.0.5 [46] in combination with AB-BLAST version 2.2.6 [47] (default parameters used for both programs). The resulting repeat-library was

applied to identify and mask repeats in the draft assembly using RepeatMasker version open-3.3.0 [48], with ab-blast used as the search engine (default parameters).

Gene set annotation. Genes were annotated using 'The NCBI Eukaryotic Genome Annotation Pipeline' (http://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/, accessed 2014-09-05). This automated pipeline annotated genes, transcripts and proteins on the draft genome assembly (NCBI Poecilia reticulata Annotation Release 100).

Genes with potential functions in pigment pattern development, vision, growth and sex differentiation. To identify protein coding genes with putative functions in pigment pattern development, vision, growth and sex differentiation, reciprocal blast searches were performed using a sample of 624 published coding sequences (S13 Table), mostly from related fish species. Resulting high-scoring segment pairs (HSPs) were manually scrutinized for e-value ($< 10^{-20}$), percent identity (>90%), length, and reading frame. Further, predicted gene models from scaffolds localized on LG12 were searched against NCBI *nt* and *nr* databases as well as Ensembl (version 71) databases of medaka, stickleback and platyfish. To resolve positions of LWS-1 (A180), LWS-2 (P180 and LWS-3 (S180) in the cluster on scaffold 43, all exons were scrutinized for best e-value ($\leq 10^{-40}$), percent identity and coding strand. Further, we aligned a Cumaná genomic BAC (GenBank HM540108.1) to genomic scaffold 43. A region of 32,500 bp length in the reconstructed BAC sequence corresponds to the opsin gene cluster that was aligned to about 38,300 bp on scaffold 43 of our assembly (96 to 98% identity, excluding few gaps). Reciprocal blast alignment revealed stretches of N in the Allpaths-lg assembly (up to 4,500 bp) as the main reason for length discrepancy. We inspected this region by eye but did not find evidence that points towards a mis-assembly in this region. A potential explanation for the length discrepancy is wrongly estimated insert sizes of the mate pair libraries for this particular region. Another possible explanation is a length difference for this particular region between the Cumaná and Guanapo strains.

Small non-coding RNAs and transfer RNAs (tRNAs). Small non-coding RNA loci were annotated using Infernal version 1.1rc1 [49] (e-value threshold 10^{-4}) in combination with Infernal Rfam database version 1.1. To annotate tRNAs we additionally ran tRNA-Scan version 1.3.1 [50].

Alignments

Pairwise alignments/Synteny analysis. The guppy genome was aligned to repeat-masked versions of the medaka (*Oryzias latipes*) and stickleback (*Gasterosteus aculeatus*) genomes (Ensembl version 71) using NUCmer version 3.1 from the MUMmer package version 3.23 [51]. Alignments were visualized with Circos plot version 0.67–7 [52].

Three-way alignments. Coding sequence annotations for the guppy were downloaded from GenBank (Accession GCF_000633615.1) and coding sequences from platy (*Xiphophorus maculatus*) and medaka (*Oryzias latipes*) were downloaded from Biomart (Ensembl 70). Three-way 1:1:1 orthology sets were identified using ProteinOrtho version 5.11 (parameter settings: minimum similarity for additional hits 0.8, blastp+). In total, 10,840 1:1:1 orthologs were identified. Next, codon sequences were aligned using prank version 140603 [53] with an empirical codon model.

Molecular evolution analyses

Substitution rate estimates. Substitution rates were estimated separately for synonymous (d_S) and nonsynonymous (d_N) substitutions per nucleotide using a maximum likelihood method, implemented in the Codeml program (model = 1, star-like user tree specified according to the phylogeny) of the Paml package v4 [54]. Alignments with $d_S > 2$ along any branch

were excluded to minimize statistical artifacts from short sequences and saturation effects in d_S (no alignment showed an estimated $d_N > 2$). The final data set comprised 9,111 1:1:1 orthologs with mean d_S estimates of 0.052 (±0.033 s.d.), 0.058 (±0.040 s.d.), and 0.949 (±0.262 s.d.) for the platy, guppy, and medaka branches, respectively. Estimates for average d_N were 0.008 (±0.010 s.d.), 0.008 (±0.013 s.d.) and 0.093 (±0.066 s.d.) for the platy, guppy, and medaka branches, respectively.

Mutation rate estimates

Parent-offspring trios were used. We crossed a Quare female (laboratory strain, originally from the Quare River in North East Trinidad) with an EnUlmBL male (laboratory strain, presumably originally from Venezuela). RAD-seq libraries for the parents and five F_1 individuals were prepared using the restriction enzymes *PstI* and *MseI* and 8 unique barcodes for each parent and one unique barcode for each F_1 individual [55]. The RAD-seq libraries, with an approximate insert size of 120–220 bp, were sequenced single-end on an Illumina HiSeq 2000 lane.

The raw reads were obtained from the sequencing platform, converted to FASTQ format and de-multiplexed using SHORE version 0.8.1. Read mapping was performed separately for each individual with BOWTIE2 version 2.1.0 [56]. Mapping results were enhanced by local realignment using GATK version 2.4–9 [57]. Single nucleotide polymorphism (SNP) detection was performed using GATK UNIFIEDGENOTYPER (default parameters). High quality bases were extracted using the *mpileup* command, implemented in SAMTOOLS version 0.1.18 [58] (BAQ score cutoff 20).

To detect *de novo* mutations, SNP calls from the parents were compared with the offspring using only sites that were covered by at least 10 reads. Approximately 16 million bases reached the base quality cutoff and between 0 and 2 new mutations were detected per F_1 individual. None of the *de novo* mutations were detected in more than one individual.

Resequencing

Sampling and sequencing. Ten males from a downstream high-predation population were collected in 2011 (PS 91100 77800, Twin Bridges) from the Guanapo drainage in North West Trinidad. Fish were euthanized with MS222 and stored in 95% ethanol.

Paired-end DNA sequencing libraries were prepared according to the "Illumina Paired End Preparation protocol"; using unique barcoded Illumina TruSeq adaptors for each individual. The PCR amplified fragments were size selected on a 2% Low Range Ultra Agarose (Bio-Rad) gel. Libraries were pooled and sequenced on two flowcell lanes with an Illumina HiSeq 2000 instrument, aiming for approximately 10x coverage per individual (101 bp read length).

Data preparation. Libraries were checked out from the sequencing platform using the SHORE *import* command version 0.8.1 to retrieve raw data. Raw reads were converted to FASTQ files using the SHORE *convert* command (see S2 Table for details about sequencing yield per sample).

Paired-end reads were mapped to the reference genome using BOWTIE2 version 2.10, applying the 'end-to-end' mapping option (no read-clipping) in the 'very-sensitive' mode. Discordant alignments for paired reads were suppressed. Mapping was enhanced for each individual by local realignment as implemented in GATK version 2.4–9 (REALIGNERTARGETCREATOR, INDELREALIGNER) and duplicates were marked using PICARD version 1.89 (http://picard.sourceforge.net, last accessed 2014-07-09).

SNP calling. SNPs were called using three different variant calling programs: GATK UNIFIEDGENOTYPER, FREEBAYES version 0.9.9 [59], and SAMTOOLS MPILEUP version 0.1.18. GATK UNIFIEDGENOTYPER and SAMTOOLS MPILEUP were run with standard parameters, FREEBAYES was run

with '—no-indels—no-mnps—no-complex—use-mapping-quality' parameters. SNPs called by all three SNP calling approaches (~0.7 million) were selected as input data for base quality recalibration with the default set of covariants (GATK BaseRecalibrator). Base quality of the mapping files was adjusted using GATK PrintReads and the resulting bam files were then used as input for a second round of SNP calling with GATK UnifiedGenotyper (standard parameters). Variants with a quality by depth annotation above 10.0 (QD>10.0) were selected as the training set for quality recalibration and filtering using GATK (~2.1 million). Next, we conducted a third round of variant calling using GATK UnifiedGenotyper. As there was no significant difference in the number (and location) of SNPs between the second and the third round, the third round SNPs were used as the final variant set for downstream analyses. In the final SNP set, the average transition-to-transversion ratio was 1.35. The predicted effect of each SNP was annotated using SnpEff version 3.3h [60].

Pairwise nucleotide diversity estimation. The genome was divided into non-overlapping windows of 50 kb for each scaffold separately. The last window of a scaffold was disregarded if it was shorter than 50 kb. Scaffolds were ordered and oriented along chromosomes as described above (see section 'Physical assembly'). Windows with at least 25 kb coverage of unique sequences were retained for downstream analyses; unique sequences were defined as the sites within each window that were neither N nor repeat masked. Additionally, we required that these windows were covered by sequence information from at least 7 individuals (i.e., a missing data threshold of 30%). For each retained window, average pairwise nucleotide diversity (π, 0. . .1) was computed for non-repetitive sites using compute, as implemented in the analysis package version 0.8.3 (https://github.com/molpopgen/analysis, last accessed 2014-11-11) found in the libsequence library [61] with default parameters.

Demographic inference. To infer the effective population sizes (N_e) history of the Guanapo high predation population, a coalescent-based hidden Markov model (PSMC) was applied as described in [62]. This method infers ancestral N_e over time, exploiting a probabilistic model of coalescence that accounts for recombination and changes in heterozygosity rates along a single diploid genome. We ran PSMC with standard parameters for each individual using recalibrated bam files (see above for details) on long, repeat masked scaffolds (size >10 kb). To visualize the results, *psmc_plot.pl* was used, assuming a mutation rate of 4.89×10^{-8} bp per generation and a generation time of 0.5 years [20]. Results were plotted for 800 to 25,000 generations ago, which translates to 400 to 12,500 years before present. The lower bound was set because estimates more recent than 800 generations are difficult to predict by the PSMC method [62].

Male-specific sequences

Resequencing and assembly. From each of the 10 males used for resequencing, reads that did not align to the female genome were extracted using the view command from the SAMtools package. Reads from all individuals were pooled together, resulting in 35 million read pairs. To lower the computational burden, ten million read pairs were randomly extracted from this pooled set of unmapped reads and assembled using Trinity version r20140717 [6]. Contigs shorter than 1,000 bp were discarded and only the longest isoform was kept for each assembled component. Other assemblers (SOAPdenovo, Velvet [63] and ABySS [64]) were tested for the male-specific assembly as well, but resulted in higher fragmentation compared to the Trinity assembly (data not shown).

Annotation. We used blastn to compare contigs against *env-nt* and *nt* databases (blastn version 2.2.30+, e-value $< 10^{-5}$, database version 20141104) to screen for contamination. Contigs with hits against *env-nt* longer than 5% of the contig size were removed. Additionally, contigs with no hits against *nt* were removed. The remaining contigs were blasted against the

female genome (BLASTN, e-value $< 10^{-10}$) and all contigs with alignment length longer than one-quarter of the contig length were discarded. The resulting contigs were compared to the *nr* database (BLASTX, e-value $< 10^{-10}$, database version 20151112). Only hits with at least 120 amino acids of target protein coverage, 50% identity and an e-value $< 10^{-40}$ were kept.

Statistical analysis

If not stated differently, all statistical tests were performed using R version 3.1.1 [65]. Where necessary, we used Bonferroni correction to adjust significance thresholds for multiple testing.

Data deposition

All short read data generated from this project have been made available via the NCBI Short Read Archive (NCBI SRA study accession SRP038017). Accession numbers for genomic libraries can be found in S1 Table, and for population samples in S2 Table. Accession numbers for the samples used to estimate mutation rates are SRR1503964 and SRR1503965 for the parents (female, male), and SRR1503967-SRR1503971 for the five F_1.

The genome assembly was submitted to NCBI under BioProject PRJNA238429. The entire project has been deposited at DDBJ/EMBL/GenBank under the accession AZHG00000000. The version described in this paper is version AZHG01000000. The assembly comprises linkage groups 1 to 23, all unplaced sequences, and the mitochondrial genome.

The scaffold accessions appear in the WGS_SCFLD line at the bottom of the WGS master record, AZHG00000000 (CM002706-CM002728 = chromosomes, KK214999-KK218026 = scaffolds).

Results

Reference genome assembly

In order to generate a reference genome for the guppy, we selected a single female and her female descendants from a high-predation population in the Guanapo drainage. This lineage had been inbred by brother-sister matings in the laboratory for five generations to reduce heterozygosity in the genome. We estimated heterozygosity to be about 1 SNP per 400 bp in the individual used for paired end genome sequencing. We used offspring of this female from later generations to produce a range of Illumina libraries, with insert sizes of up to 40 kb, and generated approximately 225 Gb of raw data (see S1 Table for additional information about libraries and insert sizes). After removing PCR duplicates about 148 Gb of sequence data remained, which we assembled into 3,028 scaffolds of a total length of 732 Mb. The longest scaffold (scaffold 0) is over 21 Mb long. Half of the assembly is represented by 43 scaffolds that are at least 5.3 Mb long (N50), and 90% of the assembly by 163 scaffolds larger than 1 Mb (N90; see Table 1 for further assembly details). The assembly size is within previous estimates of 740 to 900 Mb for the guppy genome [66], with $2n = 46$ chromosomes [32, 34]. A *k-mer* frequency approach estimated a genome size of 779.8 Mb for the female genome (S1 File), which is just slightly larger than the assembled genome.

To estimate the completeness of the assembly, we mapped expressed sequence tags (ESTs) [65] and a Roche 454 transcriptome [45] from other guppy strains to our newly constructed reference. The majority of both ESTs (15,579/16,220; 96.0%) and 454 contigs (50,188/54,981; 91.28%) could be located on the genome assembly.

Visual pigment genes (opsins) have been extensively characterized in guppies [67–71]; we therefore searched our genome assembly for these loci as a further validation of the assembly's completeness. Using published opsin cDNA and genomic sequences from guppies [67, 68, 70]

Table 1. Overview of assembly and annotation for the female guppy genome.

Contigs longer than 1kb	44,571
Total length of all contigs	663,389,323 bp
N50 length of contigs	35,577 bp
Scaffolds	3,028
Total length of all scaffolds	731,579,643 bp
Length of unclosed gaps	66,967,969 bp
Median size of gaps in scaffolds	535 bp
Lengths of scaffolds anchored on linkage groups	696,674,853 bp
Scaffolds anchored on linkage groups	284
Longest scaffold	21,430,553 bp
N50 length of scaffolds	5,270,359 bp
N90 length of scaffolds	1,021,883 bp
Scaffolds longer than N50 length	43
Scaffolds longer than N90 length	163
GC content	39.3%
Protein-coding genes	22,982
Pseudogenes	249
Fraction of transposable elements	21.3%

doi:10.1371/journal.pone.0169087.t001

and from closely related poeciliid species [69], we confirmed the presence of rhodopsin and nine cone opsin genes (S3 Table). Seven are in two clusters on LG5, one including the green-sensitive RH2-1 and RH2-2, and the other including the blue sensitive SWS2A and SWS2B and three red/orange-sensitive LWS genes. A fourth, retrotransposed LWS-4 gene is on LG2, and the gene encoding the UV-sensitive SWS-1 is on an unplaced scaffold (LG Un). These results confirmed the high quality of our assembly.

Based on comparing RAD-seq tags in parents and their F_1 offspring, we calculated a mutation rate of 4.9×10^{-8} bp^{-1} $generation^{-1}$. This is in a similar range as for Midas cichlids, which have an estimated mutation rate of 6.6×10^{-8} bp^{-1} $generation^{-1}$ [72]. Our assembly should thus also be useful for detecting *de novo* mutations that may contribute to local adaptation in addition to standing variation.

Genome annotation

A total of 22,982 protein-coding genes and 249 pseudogenes were predicted. Additionally, we annotated 439 tRNA genes for the 20 standard amino acids, 707 microRNA loci, and 160 snoRNA loci (S4 Table).

Repetitive sequences made up approximately 20% (156 Mb) of the assembly (see S5 Table for further details). Given the difficulty of assembling highly repetitive centromeric regions, these may be underrepresented. The average GC content of the genome is 39.3%, without clear signs of isochore organization along the 23 chromosomes (Fig 2), though some chromosomes show slightly elevated GC content towards the ends of the linkage groups.

Synteny with other fish genomes

Using a high-density linkage map comprising 5,493 markers, more than 95% of the assembly could be anchored to 23 linkage groups (LGs), which corresponds to the guppy's haploid set of 23 chromosomes. Estimated chromosome sizes range from 18 and 46 Mb (Figs 2 and 3, S6 Table). The longest chromosome, LG2, is the product of a fusion between ancestral chromosomes that

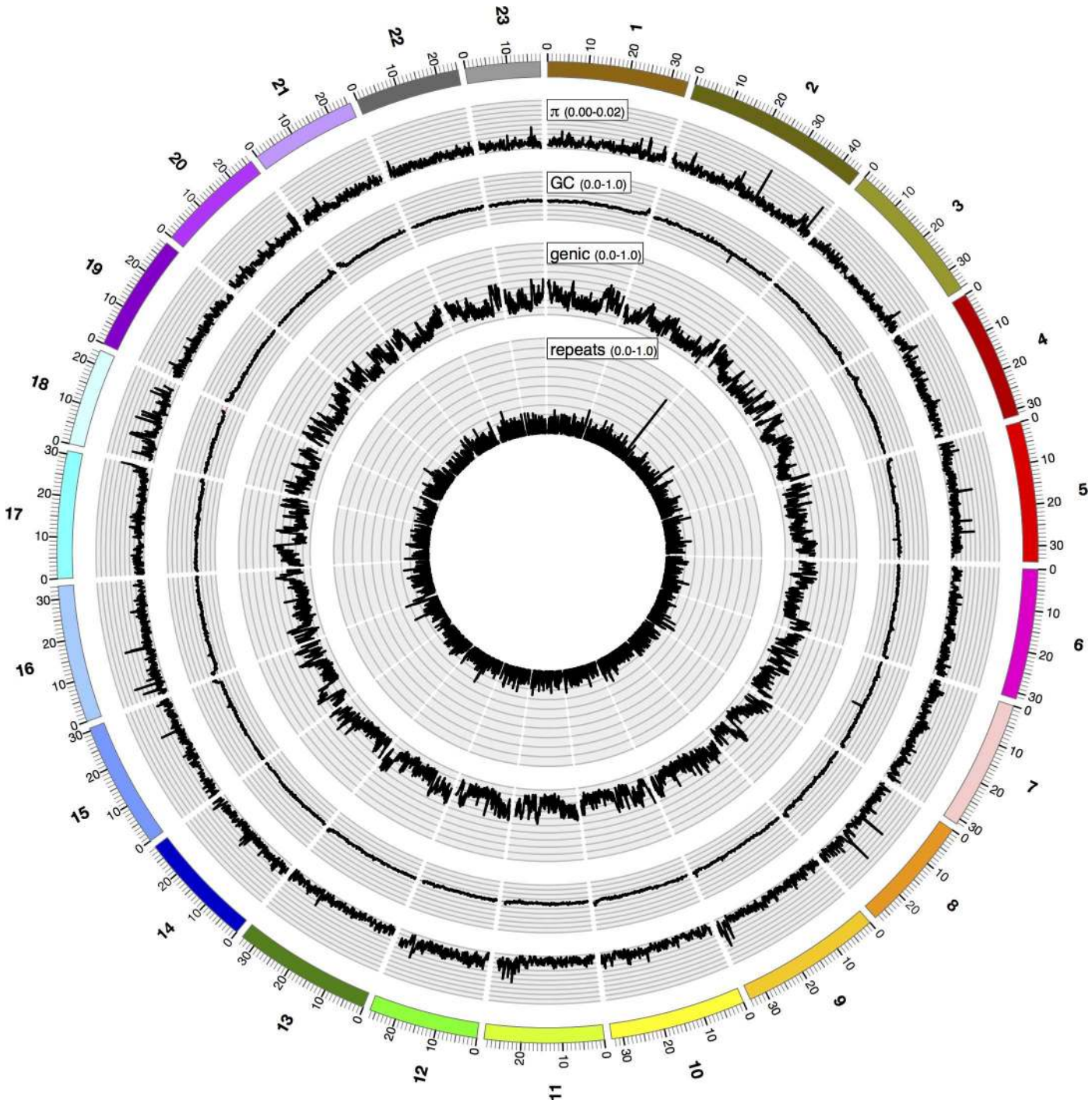

Fig 2. Sequence characteristics for each linkage group. Linkage groups are indicated on the outside. Small numbers indicate distances along each linkage group in Mb. Estimates for nucleotide diversity, π, GC content, and genic (exons and introns) and repeat density are averaged in 50 kb windows. Note that repeats can be located in genic regions.

doi:10.1371/journal.pone.0169087.g002

correspond to chromosomes 2 and 21 of medaka (*Oryzias latipes*), and groups II and XVI of stickleback (*Gasterosteus aculeatus*) (Fig 4, S1 Fig). Almost all linked scaffolds could be oriented, as they contained at least two genetic markers with recombination events between them (Fig 3).

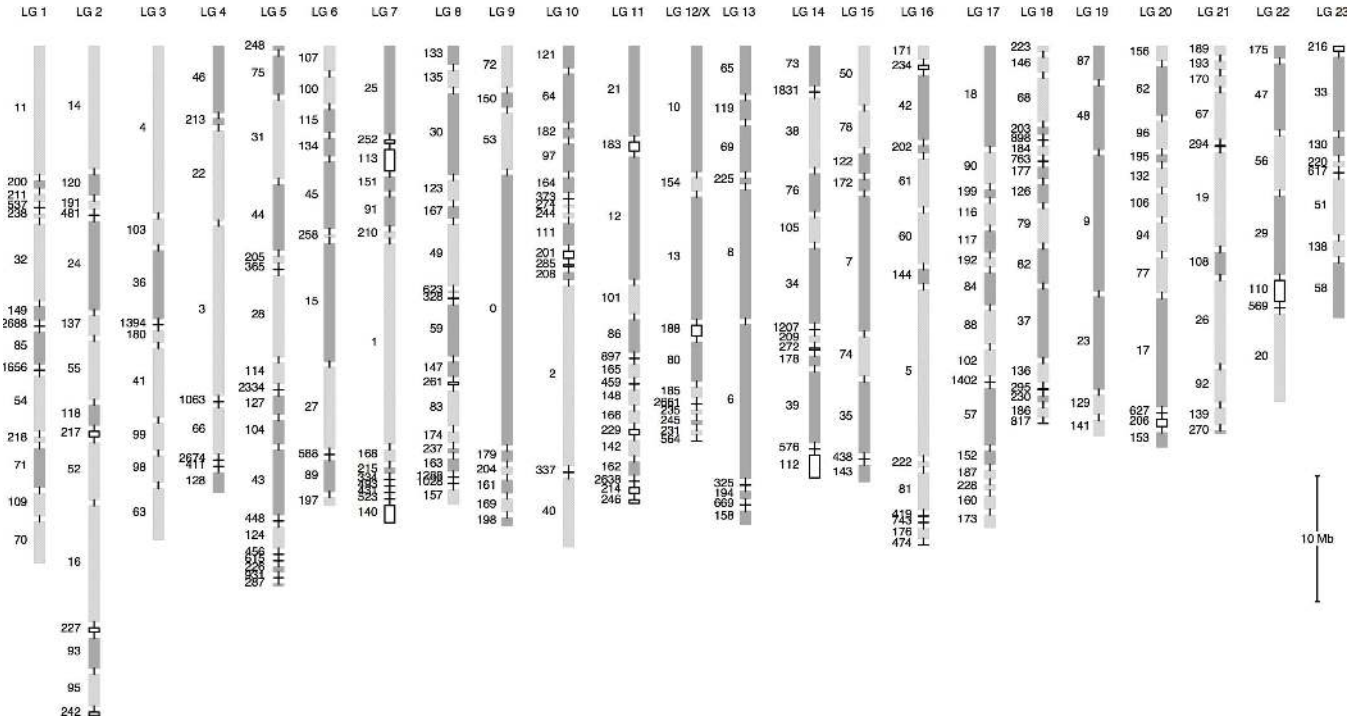

Fig 3. Distribution of anchored scaffolds along linkage groups. Grey outlined boxes denote scaffolds with at least two markers. Scaffolds in the forward orientation are solid grey and in the reverse orientation are dashed light grey. Black outlined boxes and horizontal black bars denote scaffolds with just one marker and unknown orientation. Spacing between scaffolds was set arbitrarily to 500 kb.

doi:10.1371/journal.pone.0169087.g003

An alignment of the guppy and medaka genomes confirmed extensive synteny (Fig 4), as had been previously deduced from mapping homologs of guppy genetic markers to the medaka genome [30].

Sex chromosomes

LG 12 corresponds to the X-chromosome, and the genetic map allowed us to assign 26.4 Mb of assembled sequence to this chromosome (Fig 3, S6 Table). Full genome alignments revealed synteny of guppy LG12 to chromosome 12 of medaka and group XIV of stickleback (S1B and S1C Fig). Gene and repeat content do not significantly differ between the X-chromosome and autosomes (Table 2) and overall GC content is very similar as well (39.0% *vs*. 39.4%). Rates of protein coding evolution, as estimated from the ratio of nonsynonymous to synonymous substitutions, d_N/d_S, are not significantly different in the guppy branch between the X-chromosome and autosomes (Mann-Whitney *U* test, p = 0.0677; Fig 5, Table 2), but average within-species pairwise nucleotide diversity (π), determined from re-sequencing 10 males of the Guanapo population, is significantly higher on the X-chromosome ($p < 0.001$; Table 2).

Numerous lines of evidence point to both X- and Y-chromosomes harboring genes for male size and color (see Introduction), including a previous QTL study that used approximately 800 SNP markers [30]. We therefore explored the QTL regions in more detail. Of three markers on the proximal end of LG12 that explained variation in male size [30], we localized two, marker-30 and marker-61, on scaffold 10 at positions 6,071,651 and 6,634,572. This region (6.0–6.7 Mb), harbors two genes encoding growth-related proteins, epidermal growth factor-like protein 7 (*egf17*) [73], and growth arrest-specific protein 1-like (*GAS1*) [74], (S7 Table). From BLAST searches using known candidate genes as queries as well as alignments of

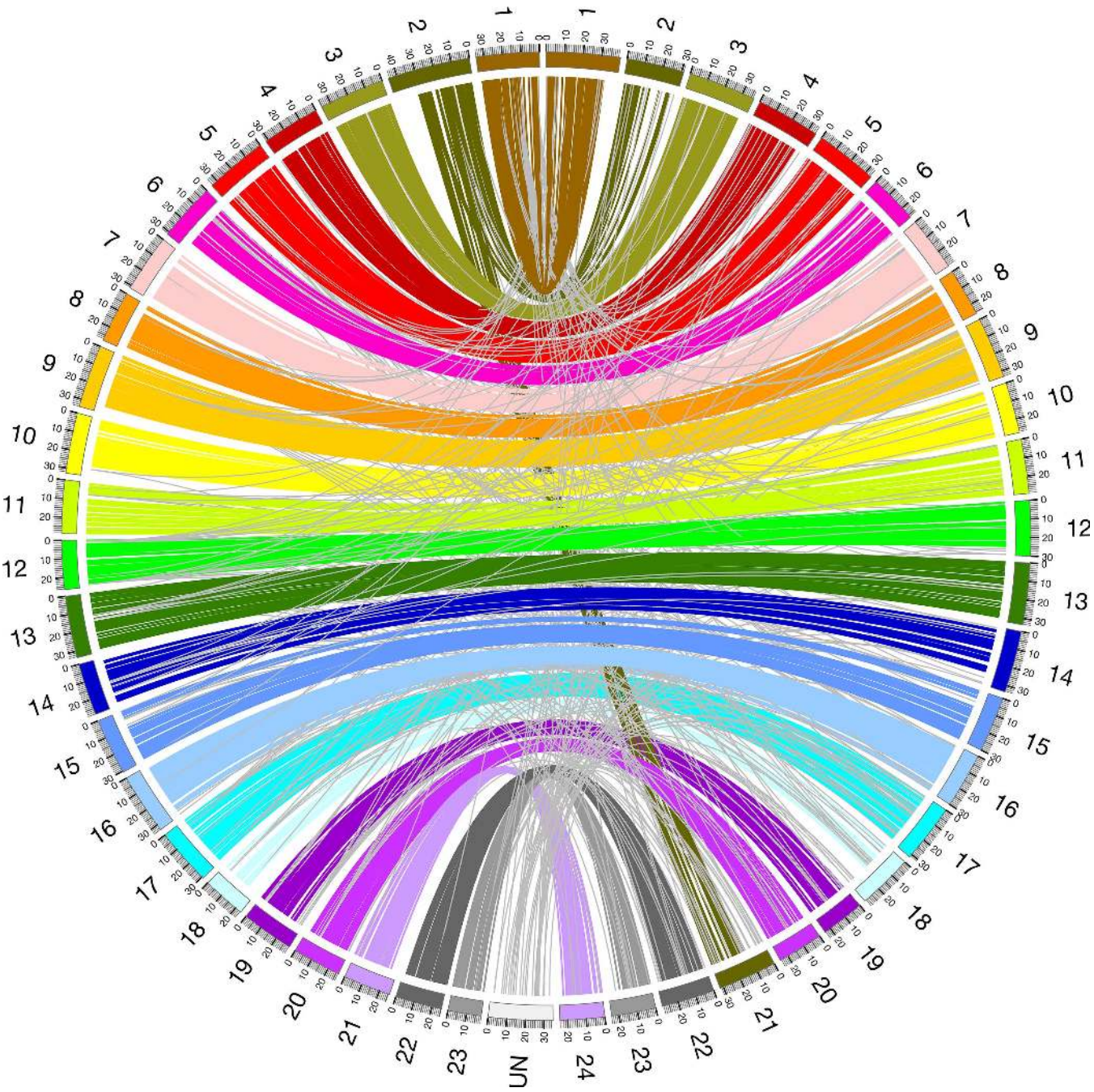

Fig 4. Whole genome alignment between Guppy and Medaka. CIRCOS plot of syntenic relationship between guppy (left) and medaka (right) chromosomes. Minimum block length 500 bp. Light grey lines indicate non-syntenic alignment blocks or blocks not assigned to any guppy linkage group (UN).

doi:10.1371/journal.pone.0169087.g004

all LG12 scaffolds to public databases we found another six candidates for growth across this chromosome (S7 Table). Tripathi and colleagues [30] also found markers on LG12 explaining color variation [30]; we could locate three of them, marker_691, marker_423, and marker_210, on scaffold 13 at positions 4,566,793, 5,622,647 and 7,057,973. Within this region (4.6–7.1 Mb) we identified one possible coloration gene candidate, *mlana* [75]. Additional reciprocal blast

Table 2. Characteristics of autosomes and X chromosome (LG12). Estimates are from 50 kb windows across each linkage group. Measurements are d_N (rate of nonsynonymous substitutions per nonsynonymous site), d_S (rate of synonymous substitutions per synonymous site) and the ratio of d_N/d_S (mean estimates are approximated by $\Sigma d_N/\Sigma d_S$), π (nucleotide diversity within populations). Coverage was estimated using pooled resequencing data and is shown as average coverage per base. Statistical testing was carried out using non-parametrical Mann-Whitney *U* test (n.s., not significant, ***, p < 0.001).

Feature	Autosomes	LG12 (Chromosome X)	MW*U* Test
d_N	0.0086	0.0085	n.s
d_S	0.0579	0.0652	n.s
d_N/d_S	0.1855	0.1619	n.s
π	0.0025	0.0039	***
Coverage	110	110	n.s.
GC content	39.40%	39.04%	***
Genic	28.34%	27.86%	n.s.
Repeat content	7.62%	7.54%	n.s.

doi:10.1371/journal.pone.0169087.t002

searches revealed four other coloration gene candidates distributed throughout LG12: Solute Carrier Family 45 Member A2 (*slc45a2*, *aim1*) [76], Superkiller Viralicidic Activity 2-Like 2 (*skiv2l2*) [77], prepromelanin concentrating hormone (*pro-MCH-like*) [78], and Sepiapterin Reductase (*spra*) [79] (S7 Table).

Tripathi and colleagues [30] had mapped the sex-determining locus to the distal-most position of LG12, a region possibly not included in our female genome assembly. Of note, three genes related to sex differentiation in fishes (*parm1* [80], 5-hydroxytryptamine receptor 1A-beta-like [81], and *gadd45gamma* [82]) and one to sex-linked behavior (*5ht receptor 1a* [81]) were found within 120 kb of each other on one of the short distal scaffolds (scaffold 185; S7 Table).

Reads from the 10 Guanapo wild-caught male individuals (details about these individuals are given in the next subsection and in Methods section) that could not be aligned to the female reference genome were assembled separately. This resulted in 1,462 contigs between 1 and 7.5 kb length (mean 1.6 kb), summing up to 2.3 Mb of sequence that potentially represent male-specific regions from the Y-like differentiated part of the sex chromosomes. These sequences include 72 protein-coding genes with and 34 without putative function, each of which covers at least 40% of a homolog in the *nr* database (S8 Table).

Population resequencing

To investigate sequence diversity within the Guanapo population, the origin of the strain used for genome assembly, we resequenced 10 wild-caught males, with a mean coverage of 12x per individual (range 8.5x to 14.0x). Mapping of the male reads to the female reference genome identified almost 5 million single nucleotide polymorphisms (SNPs). About 80% of SNPs were detected in at least two alleles (S9 Table). On average, slightly more than 2 million SNPs were identified per individual and the average ratio of heterozygous to homozygous SNPs was 1.76 (±0.16, median 1.74; S10 Table). 10% of SNPs were located in coding regions, with 173,485 nonsynonymous substitutions, and 2,520 nonsense changes relative to the reference (S11 Table). Average pairwise nucleotide diversity (π) was 0.0025 (±0.0013, median 0.0024) with a very homogeneous distribution across the entire genome, and limited within-chromosome variation along LG2, LG5, LG8, LG16 and LG18 (Fig 2). Inspecting the distribution of the uppermost 1% of π windows estimates ($\pi > 0.009$) further showed no spatial clustering along the chromosomes (Kolmogorov-Smirnov test D = 0.2145, p = 0.2324).

Nucleotide composition (measured as GC content averaged across 50 kb windows) was weakly negatively correlated with π (Spearman's p = -0.0216, p = 0.0299). Inspecting this correlation further showed that for low and average (35%-41%) GC content the correlation was

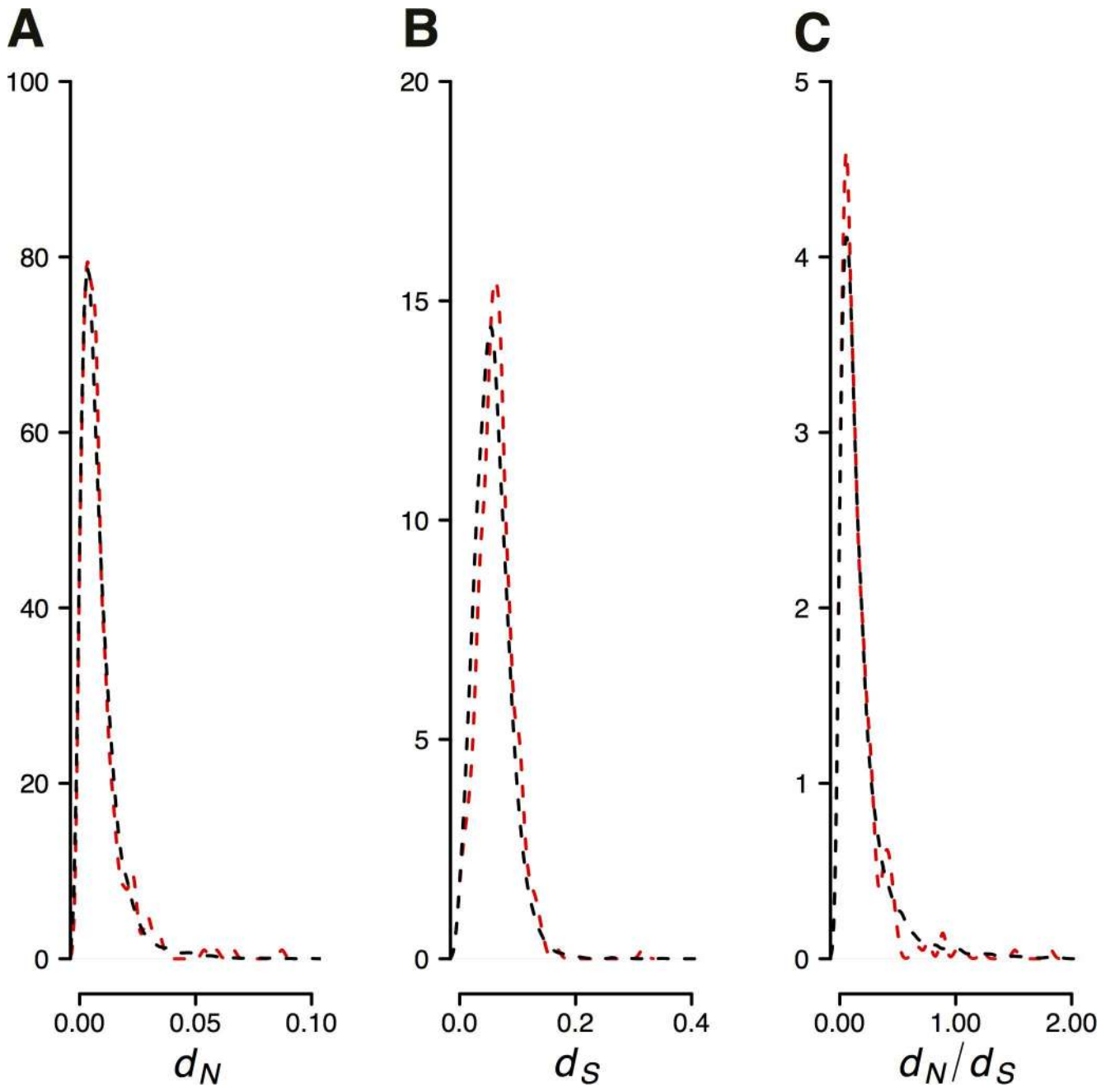

Fig 5. Evolutionary rates along X-chromosome (LG12) and autosomes. Density plots (A) of the rates of nonsynonymous (d_N), (B) synonymous substitutions per nucleotide (d_S) and (C) the ratios of d_N and d_S (d_N/d_S) between LG12 (red dashed lines) and autosomes (black dashed lines).

doi:10.1371/journal.pone.0169087.g005

negative, whereas for higher GC content (>41%) the correlation showed a positive trend (S2 Fig). We found no correlation between π and number of genic sites per window (p = -0.0146, p = 0.1419), but there was a small negative correlation between π and repeat content (p = -0.0295, p = 0.0030).

The historical demography of the resequenced individuals was examined using a coalescent-based Hidden Markov model [62] for each of the resequenced individuals. The ancestral

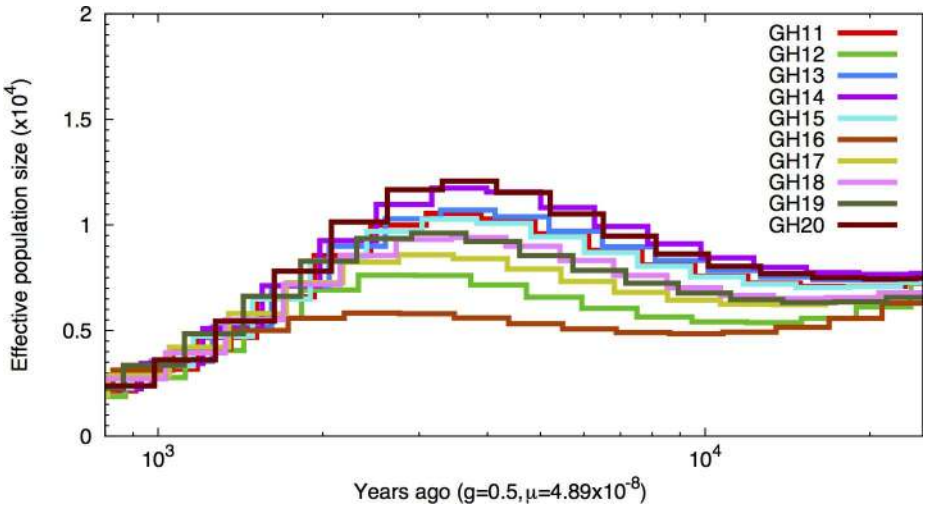

Fig 6. Inference of population size changes over time. PSMC results for the high-predation population for each individual. Each color represents a single individual. Time scale on the x-axis is calculated assuming a mutation rate of 4.89×10^{-8} bp^{-1} generation^{-1} and a generation time of 0.5 years.

doi:10.1371/journal.pone.0169087.g006

effective population (N_e) size was estimated to be highest about three to four thousand years ago (N_e ~10,000) and to have declined since then at a fairly constant rate (Fig 6). To verify the observed pattern, we bootstrapped the analysis. The bootstrap results (50 bootstraps) confirmed the demographic history of the high-predation Guanapo guppy population (shown only for individual GH13, see Fig 7).

Discussion

Here, we present a high-quality assembled reference genome for the evolutionary and ecological model system, the Trinidadian guppy. The total assembly size is approximately 732 Mb,

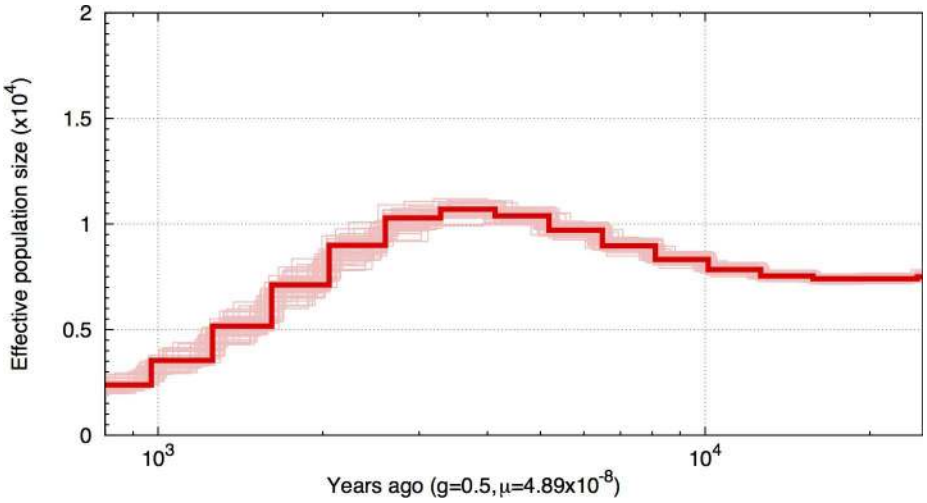

Fig 7. Bootstrapped inference of population size change over time. Plots of bootstrapped PSMC results for single individuals representing the high-predation population (GH13). The solid red line depicts the average estimate, the light red lines the 50 bootstrap results. Time scale on the x-axis is calculated assuming a mutation rate of 4.89×10^{-8} bp^{-1} generation^{-1} and a generation time of 0.5 years.

doi:10.1371/journal.pone.0169087.g007

close to the predicted genome size of 740 to 900 Mb derived from flow cytometry and Feulgen stain densitometry [66]. The assembly is highly contiguous compared to other published teleost genomes (S12 Table), with half the assembly represented by 43 scaffolds that are at least 5.3 Mb long. We achieved a high quality assembly, even though inbreeding was relatively limited. For comparison, over 100 generations of inbreeding preceded efforts to generate a whole genome assembly for *Xiphophorus maculatus*, another species from the same family as the guppy [83]. Using a high-density linkage-map, we oriented 219 scaffolds (94% of the assembled genome) along 23 chromosomes. We found a high amount of synteny with medaka and stickleback, and verified that LG2 is the result of a fusion between two ancestral chromosomes [30], further confirming the high degree of karyotypes stability in percomorph fishes (e.g. [83]). We also fully assembled the mitochondrial genome (Supplementary Text, S3 Fig), which can be used to better understand the phylogenetic relationships among distantly related fishes [84]. Together, the female reference genome greatly enhances the molecular resources that have been developed for this system [45, 85–87]. The new reference assembly has already helped to determine the molecular basis of color mutations [86] and it has informed comparisons of natural and experimental populations [88].

The guppy reference genome will help to uncover the genetic basis of adaptive phenotypes in the guppy. A previous QTL analysis provided preliminary evidence that several loci involved in male size and coloration are located on the sex chromosome, with the sex-determining locus situated at the most distal end [30]. After anchoring these markers in the assembled genome, we searched for candidate genes in their vicinity and identified several genes with known functions in pigmentation, growth, and sex-determination in other taxa (see S7 Table for references). Likewise, several research groups interested in the relationship between female mate choice and male-specific coloration attempted to characterize the visual pigment (opsin) gene complement of the guppy, with special focus on the multiple and diversified LWS genes, but results varied across studies (e.g., depending on whether guppy genomic DNA or eye RNA was screened) [70, 89, 90]. Our results validate the combined results of these previous studies, identifying four LWS opsins, two RH2 opsins, two SWS2 opsins, one SWS1 opsin, and one RH1 opsin. Comparison of this Guanapo guppy genome assembly to the previous genomic Cumana guppy BAC sequencing results agree in LWS copy number and arrangement, including a retrotransposed LWS4 within the gepherin gene. Watson *et al.* [68] have previously compared genomic structure of all LWS genes between the Cumana guppy and *Xiphophorus helleri* by genomic BAC sequencing and concluded that the arrangement of the LWS cluster as well as the LWS-4 retrotransposition event occurred before the split between *Xiphophorus* and *Poecilia* clades [68, 69].

We are particularly interested in the evolution of the sex chromosomes, LG12, because of its known variation underlying traits important in local adaptation such as male color and size. Sex chromosomes are predicted to evolve differently compared to autosomes because of their differences in transmission and ploidy, and the resulting differences in effective population size [91]. We have, however, not found evidence for an elevated rate of X-chromosome evolution in the guppy. This may be due to the fact that the majority of the assembled chromosome is pseudoautosomal and freely recombines with the Y-chromosome while the diversified differentiated distal region might still be missing in the assembly. Moreover, sex determination is a rapidly evolving trait in fishes including the Poeciliids, and estimates of evolutionary rates using medaka and platy as outgroups are almost certainly dominated by time periods when LG12 sequences were autosomal. Within-species nucleotide diversity was higher on the X chromosome than on the autosomes, which suggests that molecular evolutionary differences between the guppy sex chromosomes and autosomes do become apparent at shorter timescales.

Since we cannot tell whether a female counterpart of the differentiated region of the Y-chromosome exists in females, we separately assembled reads from ten male samples, and identified contigs that did not map to the female reference assembly. Although we assembled one thousand such contigs, none contained obvious candidates for sex determination, pigmentation, or growth. These male-only contigs were short in length (the longest just over 7 kb) and harbored many truncated open reading frames, which may be due to incomplete assembly, or because they had undergone pseudogenization. The absence of recombination in Y-chromosomes is predicted to reduce natural selection in this region and, in turn, increase the rate of pseudogenization, gene-loss, and repeat expansion [92, 93]. A third alternative is that sequences are mostly shared between X and Y, with male-specific sequences interspersed between shared sequences, rather than long blocks of male-specific sequences.

We have already exploited the reference assembly to investigate genetic diversity in the Guanapo river, a high predation locality in Northwest Trinidad that was the source population for our reference strains. Diversity was fairly homogeneous across the genome and was not strongly correlated with other genomic features such as GC or genic content. Coverage could be another potential confounding factor with respect to pairwise nucleotide diversity. We found that with higher coverage slightly fewer alternative alleles were called ($p = -0.1147$, $p < 0.001$, S4 Fig) but the correlation was not linear and did not significantly change the proportion of alternative alleles called between regions of different coverage. While the correlation appeared positive for regions of lower coverage (70x-130x), it was negative for regions of higher coverage (130x-300x).

Using a pairwise sequential Markovian coalescent model [62] that uses local density of heterozygous sites in individual diploids, we estimated that our source population had a large current effective population size (~2,500). This was comparable to other estimates of effective population sizes of a high predation population estimated from the same river (~1,300 [16]) and other lowland high predation populations (~2,000–14,000 [88], ~700–4,200 [16]). Estimates of N_e grew with increasing coalescence age, which is expected if the population is well connected via gene flow to a larger meta-population. This extensive genetic variation found in lowland populations, could be a major contributor to the rapid and repeatable adaptation of colonizers to novel predation regimes seen in further upstream locations [19, 20].

Our reference assembly presents an important step in providing a much-needed resource for the study of evolutionary genetics in the guppy. Future studies can make use of our reference assembly and explore the many aspects of guppy biology that make it a model system in understanding evolutionary biology and ecology, including life history evolution, maternal provisioning, and invasion success. The limitations of our short-read based reference assembly, however, also highlight that new genome sequencing and assembly approaches are needed to reveal the complete sequence of the sex chromosomes in this species.

Supporting Information

S1 Fig. Whole genome alignment between guppy and stickleback. (A) Circos plot showing the syntenic relationship between guppy linkage groups 1–23 and UN for unassigned scaffolds and stickleback chromosomes I-XXI. **(B)** Circos plot highlighting alignments between guppy LG2 (left) and medaka chromosomes. **(C)** Circos plot for alignments between selected regions from guppy (LG2 and LG12) and stickleback. Each line represents an alignment block of at least 500 bp.
(PDF)

S2 Fig. Nucleotide composition (GC-content) in correlation with average nucleotide diversity (π). GC content and π were estimated from 50 kb windows. The orange line represents

the LOWESS regression; dashed lines denote mean GC content and mean π, respectively.
(PDF)

S3 Fig. Guppy mitochondrial genome. CIRCOS plot of annotation of mitochondrial genome. The outermost circle denotes genes and rRNAs/tRNAs transcribed from the leading strand, and the second outermost circle from the lagging strand. The innermost circle represents the GC content per every 5 bp; the darker lines are, the higher the GC content.
(PDF)

S4 Fig. Coverage in correlation with proportion of alleles called. Mean coverage was estimated from 50 kb windows. The number of reference and non-reference sites per SNP was counted and the proportion per 50 kb window was estimated for the reference allele (Proportion Ref Allele) and for the non-reference allele (Proportion Alt Allele). The orange line represents the LOWESS regression; the vertical dashed line denotes mean coverage.
(PDF)

S1 Table. Overview of sequencing data used for assembling the female guppy genome. Raw data denotes sequencing yield after *phiX* removal and filtered data denotes sequence data after removing PCR duplicates. PE = Paired-end library, MP = Mate pair library, FM = Fosmid library.
(PDF)

S2 Table. Overview of sequencing data used for population resequencing and population. Sequencing yield for high predation population individuals. Column 'Reads mapping to reference' refers to the percentage of reads mapping to the female genome. The column 'Uncovered in Mb' refers to the uncovered sequences in Mb after mapping the paired-end genomic libraries to the reference genome.
(PDF)

S3 Table. Rhodopsin and opsin genes found in the genome assembly.
(PDF)

S4 Table. Predictions of small RNA loci using INFERNAL and predicted tRNAs and potential pseudogenes coding tRNAs using tRNA-SCAN.
(PDF)

S5 Table. Repeat content of the female guppy genome as identified by REPEATMASKER. In total, 156,122,771 bp (~21.3%) of the assembly were classified as repeats.
(PDF)

S6 Table. Linkage group sizes after anchoring scaffolds. Linkage group Un contains all unanchored scaffolds.
(PDF)

S7 Table. Protein coding genes related to (A) growth, (B) pigment pattern and (C) sex differentiation on LG12.
(PDF)

S8 Table. Best hit protein-coding genes retrieved for male specific sequences.
(XLSX)

S9 Table. Frequency of alleles different from the reference in resequencing populations (single nucleotide polymorphisms only).
(PDF)

S10 Table. Number of single nucleotide polymorphisms per individual and ratio of heterozygous to homozygous SNPs.
(PDF)

S11 Table. Single nucleotide polymorphisms by type and region for resequencing data inferred by SnpEff. Note SNPs can be counted in two or more categories.
(PDF)

S12 Table. Selected published fish genome assemblies. Comparison of 10 different published fish genomes and the guppy genome. Chromosomes refer to whether the assembly is available as chromosomes from GenBank.
(PDF)

S13 Table. Protein coding genes with putative functions in pigment pattern development, vision, growth and sex differentiation.
(XLSX)

S1 File. Supplementary Text. Additional information about procedures and analyses related to the guppy genome project.
(PDF)

Acknowledgments

We thank D. Reznick for the founder fish from the Guanapo, E. Ruell and C. Ghalambor for the population samples, A. Furness for assistance with selecting sample sites, S. Topuz for fish care, M. Zaidem for helping with measuring mutation rate, C. Lanz for assistance in sequencing, C. Dreischer for updating the genetic map and sharing scripts, S. Henz for providing scripts, and R. Burri, A. Nolte, L. Smeds, P. I. Ólason, D. Koenig, R. Neher, and F. Zanini for discussion. We are grateful to two anonymous reviewers whose comments helped to improve the manuscript.

Author Contributions

Conceptualization: CD.

Data curation: AK.

Formal analysis: AK, BAF, ES.

Funding acquisition: DW.

Investigation: AK, BAF.

Methodology: AK.

Project administration: CD.

Resources: BAF, VAK, MH.

Software: AK.

Supervision: CD, DW.

Validation: AK.

Visualization: AK.

Writing – original draft: AK, BAF, DW, CD.

Writing – review & editing: AK, BAF, DW, CD.

References

1. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in Ficedula flycatchers. Nature. 2012; 491(7426):756–60. doi: 10.1038/nature11584 PMID: 23103876
2. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 2014.
3. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 2015.
4. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick insect genomes reveal natural selection's role in parallel speciation. Science. 2014; 344(6185):738–42. doi: 10.1126/science.1252136 PMID: 24833390
5. Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun. 2013; 4:1827. doi: 10.1038/ncomms2833 PMID: 23652015
6. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 2011; 29(7):644–52. doi: 10.1038/nbt.1883 PMID: 21572440
7. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genet. 2010; 6(2): e1000862–. doi: 10.1371/journal.pgen.1000862 PMID: 20195501
8. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends in genetics: TIG. 2008; 24(3):142–9. PubMed Central PMCID: PMCPMC2680276. doi: 10.1016/j.tig.2007.12.006 PMID: 18262676
9. Gabaldon T, Koonin EV. Functional and evolutionary implications of gene orthology. Nat Rev Genet. 2013; 14(5):360–6. doi: 10.1038/nrg3456 PMID: 23552219
10. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010; 11(2):97–108. doi: 10.1038/nrg2689 PMID: 20051986
11. Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends in genetics: TIG. 2004; 20(10):481–90. doi: 10.1016/j.tig.2004.08.001 PMID: 15363902
12. Braasch I, Peterson SM, Desvignes T, McCluskey BM, Batzel P, Postlethwait JH. A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo. J Exp Zool B Mol Dev Evol. 2015; 324(4):316–41. PubMed Central PMCID: PMCPMC4324401. doi: 10.1002/jez.b.22589 PMID: 25111899
13. Winge Ö. One-sided masculine and sex-linked inheritance in Lebistes reticulatus. Journal of Genetics. 1922; 12:146–62.
14. Endler JA. Multiple-trait coevolution and environmental gradients in guppies. Trends Ecol Evol. 1995; 10(1):22–9. Epub 1995/01/01. PMID: 21236940
15. Magurran AE. Evolutionary Ecology: The Trinidadian Guppy: Oxford University Press; 2005 Aug 25. 206 p.
16. Barson NJ, Cable J, Van Oosterhout C. Population genetic analysis of microsatellite variation of guppies (Poecilia reticulata) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. J Evol Biol. 2009; 22(3):485–97. doi: 10.1111/j.1420-9101.2008.01675.x PMID: 19210594
17. Alexander HJ, Taylor JS, Wu SS, Breden F. Parallel evolution and vicariance in the guppy (Poecilia reticulata) over multiple spatial and temporal scales. Evolution. 2006; 60(11):2352–69. PMID: 17236426
18. Endler JA. Natural selection in color patterns in Poecilia reticulata. Evolution. 1980; 34:76–91.
19. Gordon SP, Reznick D, Arendt JD, Roughton A, Ontiveros Hernandez MN, Bentzen P, et al. Selection analysis on the rapid evolution of a secondary sexual trait. Proc Biol Sci. 2015; 282(1813):20151244. doi: 10.1098/rspb.2015.1244 PMID: 26290077
20. Reznick DN, Shaw FH, Rodd FH, Shaw RG. Evaluation of the Rate of Evolution in Natural Populations of Guppies (Poecilia reticulata). Science. 1997; 275(5308):1934–7. Epub 1997/03/28. PMID: 9072971
21. Reznick DN. Life history evolution in guppies (Poecilia reticulata): 1. Phenotypic and genetic changes in an introduction experiment. Evolution (NY). 1987; 41:1370–85.

22. Reznick DN, Ghalambor CK, Crooks K. Experimental studies of evolution in guppies: a model for understanding the evolutionary consequences of predator removal in natural communities. Mol Ecol. 2008; 17(1):97–107. Epub 2007/08/30. doi: 10.1111/j.1365-294X.2007.03474.x PMID: 17725576

23. Barrett RD, Schluter D. Adaptation from standing genetic variation. Trends Ecol Evol. 2008; 23(1):38–44. doi: 10.1016/j.tree.2007.09.008 PMID: 18006185

24. Olendorf R, Rodd FH, Punzalan D, Houde AE, Hurt C, Reznick DN, et al. Frequency-dependent survival in natural guppy populations. Nature. 2006; 441(7093):633–6. doi: 10.1038/nature04646 PMID: 16738659

25. Hughes KA, Houde AE, Price AC, Rodd FH. Mating advantage for rare males in wild guppy populations. Nature. 2013:1–10.

26. Lindholm A, Breden F. Sex chromosomes and sexual selection in poeciliid fishes. Am Nat. 2002; 160 Suppl 6:S214–24. Epub 2008/08/19.

27. Brooks R, Endler JA. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (Poecilia reticulata). Evolution. 2001; 55(5):1002–15. PMID: 11430637

28. Houde AE. Sex, Color, and Mate Choice in Guppies: Princeton University Press; 1997. 210 p.

29. Hughes KA, Rodd FH, Reznick DN. Genetic and environmental effects on secondary sex traits in guppies (Poecilia reticulata). J Evol Biol. 2005; 18(1):35–45. doi: 10.1111/j.1420-9101.2004.00806.x PMID: 15669959

30. Tripathi N, Hoffmann M, Willing EM, Lanz C, Weigel D, Dreyer C. Genetic linkage map of the guppy, Poecilia reticulata, and quantitative trait loci analysis of male size and colour variation. Proceedings of the Royal Society B: Biological Sciences. 2009; 276(1665):2195–208. PubMed Central PMCID: PMCPMC2677598. doi: 10.1098/rspb.2008.1930 PMID: 19324769

31. Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb). 2005; 95(2):118–28.

32. Nanda I, Schories S, Tripathi N, Dreyer C, Haaf T, Schmid M, et al. Sex chromosome polymorphism in guppies. Chromosoma. 2014.

33. Lisachov AP, Zadesenets KS, Rubtsov NB, Borodin PM. Sex chromosome synapsis and recombination in male guppies. Zebrafish. 2015; 12(2):174–80. doi: 10.1089/zeb.2014.1000 PMID: 25608108

34. Traut W, Winking H. Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. Chromosome Res. 2001; 9(8):659–72. PMID: 11778689

35. Tripathi N, Hoffmann M, Weigel D, Dreyer C. Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (Poecilia reticulata). Genetics. 2009; 182(1):365–74. PubMed Central PMCID: PMCPMC2674833. doi: 10.1534/genetics.108.098541 PMID: 19299341

36. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat Genet. 2014:1–6.

37. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 2008:gr.080200.108-.

38. Smeds L, Künstner A. CONDETRI—A Content Dependent Read Trimmer for Illumina Data. PLoS ONE. 2011; 6(10):e26314. doi: 10.1371/journal.pone.0026314 PMID: 22039460

39. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011; 17(1):pp. 10–2.

40. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010; 20(2):265–72. doi: 10.1101/gr.097261.109 PMID: 20019144

41. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009; 25(14):1754–60.

42. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA. 2011; 108(4):1513–8. PubMed Central PMCID: PMCPMC3029755. doi: 10.1073/pnas.1017351108 PMID: 21187386

43. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25 (17):3389–402. PMID: 9254694

44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10(1).

45. Fraser BA, Weadick CJ, Janowitz I, Rodd FH. Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome. BMC Genomics. 2011; 12:202. doi: 10.1186/1471-2164-12-202 PMID: 21507250

46. Smit A, Hubley R. RepeatModeler Open-1.0. 2010.

47. Gish W. http://blast.advbiocomp.com 1996–2009.

48. Smit A, Hubley R, Green P. RepeatMasker Open-3.0. 2010.

49. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics (Oxford, England). 2013; 29(22):2933–5.

50. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25(5):955–64. PubMed Central PMCID: PMCPMC146525. PMID: 9023104

51. Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5(2).

52. 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):1639–45. PubMed Central PMCID: PMCPMC2752132. doi: 10.1101/gr.092759.109 PMID: 19541911

53. Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A. 2005; 102(30):10557–62. PubMed Central PMCID: PMCPMC1180752. doi: 10.1073/pnas.0409137102 PMID: 16000407

54. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007; 24(8):1586–91. doi: 10.1093/molbev/msm088 PMID: 17483113

55. Poland JA, Brown PJ, Sorrells ME, Jannink J- L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. 2012; 7(2): e32253. PubMed Central PMCID: PMCPMC3289635. doi: 10.1371/journal.pone.0032253 PMID: 22389690

56. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012; 9 (4):357–9. doi: 10.1038/nmeth.1923 PMID: 22388286

57. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. doi: 10.1101/gr.107524.110 PMID: 20644199

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25(16):2078–9.

59. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012.

60. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6(2):80–92. PubMed Central PMCID: PMCPMC3679285.

61. Thornton K. Libsequence: a C++ class library for evolutionary genetic analysis. 2003.

62. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011; 475(7357):493–6. doi: 10.1038/nature10231 PMID: 21753753

63. Zerbino D, Birney E. Velvet: Algorithms for De Novo Short Read Assembly Using De Bruijn Graphs. Genome Research. 2008; 18:821–9. doi: 10.1101/gr.074492.107 PMID: 18349386

64. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Research. 2009; 19:1117–23. PubMed Central PMCID: PMCPMC2694472. doi: 10.1101/gr.089532.108 PMID: 19251739

65. R Development Core Team. R: A Language and Environment for Statistical Computing. 2014.

66. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. Nucl Acids Res. 2007; 35(Database issue):D332–8. PubMed Central PMCID: PMCPMC1669731. doi: 10.1093/nar/gkl828 PMID: 17090588

67. Laver CR, Taylor JS. RT-qPCR reveals opsin gene upregulation associated with age and sex in guppies (Poecilia reticulata)—a species with color-based sexual selection and 11 visual-opsin genes. BMC Evol Biol. 2011; 11:81. Epub 2011/03/31. PubMed Central PMCID: PMC3078887. doi: 10.1186/1471-2148-11-81 PMID: 21447186

68. Watson CT, Gray SM, Hoffmann M, Lubieniecki KP, Joy JB, Sandkam BA, et al. Gene duplication and divergence of long wavelength-sensitive opsin genes in the guppy, Poecilia reticulata. J Mol Evol. 2011; 72(2):240–52. Epub 2010/12/21. doi: 10.1007/s00239-010-9426-z PMID: 21170644

69. Watson CT, Lubieniecki KP, Loew E, Davidson WS, Breden F. Genomic organization of duplicated short wave-sensitive and long wave-sensitive opsin genes in the green swordtail, Xiphophorus helleri.

BMC Evol Biol. 2010; 10:87. Epub 2010/04/01. PubMed Central PMCID: PMC3087554. doi: 10.1186/1471-2148-10-87 PMID: 20353595

70. Hoffmann M, Tripathi N, Henz SR, Lindholm AK, Weigel D, Breden F, et al. Opsin gene duplication and diversification in the guppy, a model for sexual selection. Proc Biol Sci. 2007; 274(1606):33–42. PubMed Central PMCID: PMCPMC1679887. doi: 10.1098/rspb.2006.3707 PMID: 17015333

71. Sandkam B, Young CM, Breden F. Beauty in the eyes of the beholders: colour vision is tuned to mate preference in the Trinidadian guppy (Poecilia reticulata). Mol Ecol. 2015; 24(3):596–609. doi: 10.1111/mec.13058 PMID: 25556876

72. Recknagel H, Elmer KR, Meyer A. A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (Amphilophus spp.) obtained by massively parallel DNA sequencing (ddRADSeq). G3 (Bethesda). 2013; 3(1):65–74. PubMed Central PMCID: PMC3538344.

73. Shen MM, Schier AF. The EGF-CFC gene family in vertebrate development. Trends in genetics: TIG. 2000; 16(7):303–9. Epub 2000/06/20. PMID: 10858660

74. Airaksinen MS, Holm L, Hatinen T. Evolution of the GDNF family ligands and receptors. Brain, behavior and evolution. 2006; 68(3):181–90. Epub 2006/08/17. doi: 10.1159/000094087 PMID: 16912471

75. Du J, Miller AJ, Widlund HR, Horstmann MA, Ramaswamy S, Fisher DE. MLANA/MART1 and SILV/PMEL17/GP100 are transcriptionally regulated by MITF in melanocytes and melanoma. Am J Pathol. 2003; 163(1):333–43. PubMed Central PMCID: PMCPMC1868174. doi: 10.1016/S0002-9440(10)63657-7 PMID: 12819038

76. Fukamachi S, Asakawa S, Wakamatsu Y, Shimizu N, Mitani H, Shima A. Conserved function of medaka pink-eyed dilution in melanin synthesis and its divergent transcriptional regulation in gonads among vertebrates. Genetics. 2004; 168(3):1519–27. Epub 2004/12/08. PubMed Central PMCID: PMC1448775. doi: 10.1534/genetics.104.030494 PMID: 15579703

77. Yang CT, Hindes AE, Hultman KA, Johnson SL. Mutations in gfpt1 and skiv2l2 cause distinct stage-specific defects in larval melanocyte regeneration in zebrafish. PLoS Genet. 2007; 3(6):e88. Epub 2007/06/05. PubMed Central PMCID: PMC1885281. doi: 10.1371/journal.pgen.0030088 PMID: 17542649

78. Kawauchi H, Baker BI. Melanin-concentrating hormone signaling systems in fish. Peptides. 2004; 25(10):1577–84. Epub 2004/10/13. doi: 10.1016/j.peptides.2004.03.025 PMID: 15476924

79. Negishi S, Fujimoto K, Katoh S. Localization of sepiapterin reductase in pigment cells of Oryzias latipes. Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society. 2003; 16(5):501–3. Epub 2003/09/03. PMID: 12950727

80. Park JY, Jang H, Curry TE, Sakamoto A, Jo M. Prostate androgen-regulated mucin-like protein 1: a novel regulator of progesterone metabolism. Molecular endocrinology. 2013; 27(11):1871–86. Epub 2013/10/03. PubMed Central PMCID: PMC3805850. doi: 10.1210/me.2013-1097 PMID: 24085821

81. Loveland JL, Uy N, Maruska KP, Carpenter RE, Fernald RD. Social status differences regulate the serotonergic system of a cichlid fish, Astatotilapia burtoni. The Journal of experimental biology. 2014; 217(Pt 15):2680–90. Epub 2014/05/24. doi: 10.1242/jeb.100685 PMID: 24855673

82. Johnen H, Gonzalez-Silva L, Carramolino L, Flores JM, Torres M, Salvador JM. Gadd45g is essential for primary sex determination, male fertility and testis development. PLoS One. 2013; 8(3):e58751. Epub 2013/03/22. PubMed Central PMCID: PMC3596291. doi: 10.1371/journal.pone.0058751 PMID: 23516551

83. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, et al. The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat Genet. 2013.

84. Fischer C, Koblmuller S, Gully C, Schlotterer C, Sturmbauer C, Thallinger GG. Complete mitochondrial DNA sequences of the threadfin cichlid (Petrochromis trewavasae) and the blunthead cichlid (Tropheus moorii) and patterns of mitochondrial genome evolution in cichlid fishes. PLoS One. 2013; 8(6):e67048. PubMed Central PMCID: PMCPMC3691221. doi: 10.1371/journal.pone.0067048 PMID: 23826193

85. Sharma E, Künstner A, Fraser BA, Zipprich G, Kottler VA, Henz SR, et al. Transcriptome assemblies for studying sex-biased gene expression in the guppy, Poecilia reticulata. BMC Genomics. 2014; 15:400. PubMed Central PMCID: PMC4059875. doi: 10.1186/1471-2164-15-400 PMID: 24886435

86. Kottler VA, Kunstner A, Koch I, Flotenmeyer M, Langenecker T, Hoffmann M, et al. Adenylate cyclase 5 is required for melanophore and male pattern development in the guppy (Poecilia reticulata). Pigment Cell Melanoma Res. 2015; 28(5):545–58. doi: 10.1111/pcmr.12386 PMID: 26079969

87. Dreyer C, Hoffmann M, Lanz C, Willing EM, Riester M, Warthmann N, et al. ESTs and EST-linked polymorphisms for genetic mapping and phylogenetic reconstruction in the guppy, Poecilia reticulata. BMC Genomics. 2007; 8:269. PubMed Central PMCID: PMCPMC1994688. doi: 10.1186/1471-2164-8-269 PMID: 17686157

88. Fraser BA, Kunstner A, Reznick DN, Dreyer C, Weigel D. Population genomics of natural and experimental populations of guppies (Poecilia reticulata). Mol Ecol. 2015; 24(2):389–408. doi: 10.1111/mec.13022 PMID: 25444454

89. Ward MN, Churcher AM, Dick KJ, Laver CR, Owens GL, Polack MD, et al. The molecular basis of color vision in colorful fish: four long wave-sensitive (LWS) opsins in guppies (Poecilia reticulata) are defined by amino acid substitutions at key functional sites. BMC Evol Biol. 2008; 8:210. Epub 2008/07/22. doi: 10.1186/1471-2148-8-210 PMID: 18638376

90. Weadick CJ, Chang B. Long-wavelength sensitive visual pigments of the guppy (Poecilia reticulata): six opsins expressed in a single individual. BMC Evol Biol. 2007.

91. Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, et al. Are all sex chromosomes created equal? Trends in genetics: TIG. 2011; 27(9):350–7. doi: 10.1016/j.tig.2011.05.005 PMID: 21962970

92. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013; 14(2):113–24. PubMed Central PMCID: PMCPMC4120474. doi: 10.1038/nrg3366 PMID: 23329112

93. Ellegren H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet. 2011; 12(3):157–66. doi: 10.1038/nrg2948 PMID: 21301475