**How did the guppy Y chromosome evolve?**

**Accepted manuscript: PGENETICS-D-21-00396R1**

**Authors:** Deborah Charlesworth ([https://orcid.org/0000-0002-3939-9122](https://orcid.org/0000-0002-9006-3448)), Roberta Bergero, Chay Graham, Jim Gardner and Karen Keegan

**Abstract**: The study used genetic mapping and coverage data in genome sequences of multiple male and female individuals of *M. picta* from multiple natural populations to investigate genetic degeneration of the Y chromosome, and quantify gene loss from the Y. The files include coverage results from the sex chromosome that were (i) used for sexing the sequenced individuals, and (ii) combined with autosomal results to analyze M/F, M/A and F/A depth of coverage ratios. Genetic mapping was also used to validate sex linkage, and the data set includes files with genotypes of genetic markers.

**Usage Notes:** The readme file contains an explanation of each of the variables in the dataset, its measurement units, and -if it concerns a derived variable (displayed in grey) - the way it was calculated from the primary data (displayed in black). #NA =  values not available. Information on how the measurements were done can be found in the associated manuscript referenced above.

NOTE: an assembly of the M. picta genome is available in the European Nucleotide Archive under accession number ERZ1744533 (<https://www.ebi.ac.uk/ena/browser/text-search?query=PRJEB43222>).

**Files**

There are 4 Excel files, and the contents are indicated in title rows and column headings, or, for the Coverage data file, in an initial Methods page.

* + File “Coverage data” includes coverage data in genome sequences of male and female individuals from natural populations listed in Supplementary Table S1. The table shows results for all chromosomes. The results from the sex chromosome were used for sexing the sequenced individuals, and (combined with autosomal results) for analyses of M/F, M/A and F/A depth of coverage ratios (Figure 2 and Supplementary Figures S2, 3, 7 and 8). Details of genetic markers used for genetic mapping of the guppy LG12 (sex chromosome pair). The genetic mapping results are in an unpublished paper, Bergero et al., 2019 (Bergero, R., J. Gardner and D. Charlesworth, 2019 Evolution of a Y chromosome from an X chromosome. Unpublished manuscript 10.2139/ssrn.3417937. doi: 10.2139/ssrn.3417937), and are referred to in the PLoSGenetics paper above

NOTE 1: As the coverage values rely on genome sequence information from the guppy, *Poecilia reticulatae*, we provide a table of recall values and comments on the quality of the sequences used, in a Word file “Recall values tables”.

NOTE 2 on **Supplementary Table 1 and Supplementary Methods:** We used 13 females from the CAR\_H population when making locally re-assembled genes to estimate divergence between *M. picta* and *P. reticulata* species. However, when analysing coverage in *M. picta*, one female from the CAR\_H population (AWCSU02 N705 AK403) was omitted.

* + File “LG12 SNPs\_mapped” includes SNPs in coding regions (with very few exceptions, as indicated) chosen for high-throughput genotyping, their positions in the guppy female assembly (Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, et al. The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. PLOS ONE. 2017;11(12):e0169087. doi: 10.1371/journal.pone.0169087), and the genes in which these SNPs are located.
	+ File “SeqSNP LG12targets (both experiments)” indicates which SNPs were mapped in the two high-throughput experiments mentioned in the PLoS Genetics paper, and described in detail in the unpublished paper, Bergero et al., 2019 (see reference above).
	+ File “Microsatellite genotypes for mapping LG12” provides the genotypes for microsatellite markers used for genetic mapping of the guppy sex chromosome in a family mentioned in the PLoS Genetics paper. The data are also in Table S7 of Bergero et al., 2019.
	+ There are also 6 files relating to the assembly of the sex chromosome, LG12 (which relates closely to the genetic mapping results). These were used mainly in a related paper (see below), but were also used in this study.
* There are 4 script files and 2 input files
* Script files
1. File ‘generate\_LG12\_sequences\_for\_picta\_BLAT’ is a Python script for BLAT searches of genes in the female guppy genome assembly in M. picta contigs.
2. File ‘generate\_male\_chr12\_sequences\_for\_picta\_BLAT’ is a Python script for BLAT searches of genes in the male guppy genome assembly in M. picta contigs.
3. File ‘split\_large\_picta\_contigs\_1500bp-windows’ performs BLAT is a Python script to search for sequences in long contigs with very few genes (the analysis is described in the PLoS Genetics paper, and indicates that these contigs are autosomal, noy=t part of the sex chromosome of M. picta).
4. picta\_long\_contig\_BLATs is an R script to prepare the output data from script 3 for plotting Supplementary Figure S8 of the manuscript under revision for Genome Biology and Evolution, revision submitted as GBE-210408.R1.
* Input files
* guppy\_LG12\_genes.txt  [input file for ‘generate\_LG12\_sequences\_for\_picta\_BLAT.py’]
* guppy\_male\_chr12\_genes.txt [input file for ‘generate\_male\_chr12\_sequences\_for\_picta\_BLAT.py’]

**Keywords**: Descriptive words that may help others discover your dataset. We recommend that you determine whether your discipline has an existing controlled vocabulary from which to choose your keywords. Please enter as many keywords as applicable.

* + Sex chromosome, coverage, genetic mapping, *Poecilia*, *Micropoecilia*

**Methods**: The methods used in the study are described in detail in the unpublished paper Bergero et al., 2019 (see reference above). The genome sequence assembly of the species studied is described in a related paper (reference 26 of the present paper): Charlesworth D, Graham C, Trivedi U, Gardner J, Bergero R. 2021 PromethION sequencing and assembly of the genome of *Micropoecilia picta*, a fish with a highly degenerated Y chromosome. This manuscript has been revised for Geneome Biology and Evolution.

**Funding Information**: Name of the funding organization that supported creation of the resource, including applicable grant number(s).

* + European Research Council (ERC), grant number 695225 (GUPPYSEX)

**Related works**: Code files relating to the assembly of the guppy sex chromosome, LG12, are also related to a manuscript that is currently in revision for Genome Biology and Evolution — GBE-210408.R1.