**Supporting Information**

Table of Contents

1. **STAMPY simulations of mapping efficiency 1**
2. **Total evidence phylogeny using the *X. maculatus* reference genome and *X. mayae* transcriptome 1**
   1. **Figure S1.** Total evidence phylogenies with *X. maculatus* and *X. mayae* references 17
   2. **Figure S2.** Rooted mitochondrial phylogeny of *Xiphophorus* 18
3. **Genome sequencing of *X. malinche*** **2**
4. **BUCKy analysis using the *X. maculatus* genome 3**
   1. **Figure S3.** Comparisons between BUCKy results for alignments to the *X. birchmanni* transciptome and BUCKy results for alignments to the *X. maculatus* genome 19
   2. **Table S6.** Bayesian concordance factors for BUCKy analyses 14
5. **Compilation of MSP, sword length and sword preference data 4**
   1. **Table S4.** Sword indices compiled from literature 11
   2. **Table S5.** Sword preference of *Xiphophorus* and an outgroup species (*Priapella olmecae*) compiled from literature 12
6. **Other tables referenced in main text**
   1. **Table S1**. Sources of species and strains used in this study with pedigree information where available 7
   2. **Table S2.** List of species included in AU test and D-statistic 8
   3. **Table S3.** Read, mapping and alignment statistics 9
   4. **Table S7.** Divergence estimates 16
7. **STAMPY simulations of mapping efficiency**

To filter our transcriptome to eliminate regions likely to generate errors in mapping (e.g. transcripts from gene families), we BLASTed the transcriptome to the *X. maculatus* genome (see main text *Materials and Methods*). Because many small regions (25-100 bp) of the assembled transcriptome had multiple BLAST hits which passed our e-value threshold, we ran simulations to determine how likely a read is to be incorporated in our analysis pipeline given varying levels of divergence from the reference sequence. For example, a read generated from a non-matching transcript that has 70 bp of overlap with a transcript included in our reference transcriptome will have approximately 25% divergence from the reference transcript. To roughly simulate this effect, we ran 100 replicates of a simulation that generated ten thousand 100 bp reads from Scaffold 0 in the reference genome with a range of mutations rates (20%, 25%, 30%) using a custom perl script, and used STAMPY to map these reads to Scaffold 0 as described for the real data (see main text: Methods). We found that the mean proportion of reads mapped by STAMPY ranged from 58% at 20% divergence to 10% at 30% divergence. However, only a fraction of these reads were retained by samtools during the mpileup step, such that <7% of these reads were used at all three divergence levels. We chose a BLAST overlap threshold of 70 bp for masking (see main text: Methods); given our simulations this suggests that fewer than 2% of reads will map incorrectly to homologous regions.

1. **Total evidence phylogeny using the *X. maculatus* reference genome and *X. mayae* transcriptome**

Trimmed reads were aligned to the *X. maculatus* reference genome (GenBank Assembly ID: [GCA\_000241075.1](http://www.ncbi.nlm.nih.gov/genome/assembly/?term=GCA_000241075.1), Ensembl annotation: http://pre.ensembl.org/Xiphophorus\_maculatus) and the *X. maculatus* mitochondrial genome (GenBank Accession: AP005982.1) using STAMPY v1.0.17 ([Lunter and Goodson 2011](#_ENREF_10)). Mapping followed methods described in the main text; in addition to the RNAseq data, we also aligned genomic reads from three species collected for another project to the *X. maculatus* genome (see below). Mapped reads were analyzed for variant sites using the samtools/bcftools pipeline ([Li et al. 2009](#_ENREF_9)) with a mapping quality cutoff of 20. Methods used for the *X. mayae* transcriptome were the same as used for the *X. birchmanni* transcriptome (see main text Materials and Methods).

For the alignments to the *X. maculatus* genome, a custom PHP script was used to generate sequence alignments based on the output of the samtools/bcftools pipelines. For each species, bases with coverage <5X or variant quality score <20 were masked; sites containing polymorphism or indels were also masked. After this initial masking, we compared sites between species. If a particular site was coded as N in 50% or more of the *Xiphophorus* species, or both outgroup species had an N at that position, we excluded that site from our analysis. We also excluded regions of high divergence (more than 7 character differences from the *X. maculatus* reference sequence in 21 bp) using a sliding window. We then concatenated alignments if they were separated by less than 1 kb; all alignments where both outgroup species had fewer than 1.5 kb non-informative characters were excluded from the analysis. This concatenation step simplifies analysis but could have consequences as a result of combining heterogenous sites. This resulted in 4,819 alignments with a total alignment length of 16.54 Mbp (12.56% missing). Methods used for screening of the sequences aligned to the *X. mayae* transcriptome were the same as used for the *X. birchmanni* analysis (see main text Materials and Methods). This resulted in 1,111 alignments with a total alignment length of 11.68 Mbp (26.85% missing). Due to initial observation of long branch attraction between outgroups and the *X. mayae* reference, we excluded both outgroups (resulted in 24.2% missing data). Total evidence phylogenies for both datasets were produced as described in the main text.

The total evidence phylogeny produced by aligning reads to the *X. maculatus* genome closely matches the phylogeny produced based on alignment to the *X. birchmanni* transcriptome, with only *X. nezahualcoyotl* changing placement (Fig. S1). The unrooted total evidence phylogeny produced based on alignment to the *X. mayae* transcriptome changes the placement of *X. mayae*, but is otherwise identical to the above, with some low bootstrap support values (Fig. S1).

1. **Genome sequencing of *X. malinche*, *X. birchmanni* and *X. clemenciae***

*X. birchmanni* and *X. clemenciae* genome sequence data was obtained from a previous project (NCBI SRA Acc # SRA060275). To generate genome sequence data for *X. malinche*, one *X. malinche* individual (Arroyo Xontla at Chicayotla, Hidalgo, Mexico) was obtained from a wild population. Genomic DNA was extracted from fin clips using the Agencourt bead-based DNA purification kit (Beckman Coulter Inc., Brea, CA) following manufacturer’s protocol with slight modifications. Fin clips were incubated in a 55 °C shaking incubator (100 rpm) overnight in 94 μl of lysis buffer with 3.5 μl 40 mg/mL proteinase K and 2.5 DTT, followed by bead binding and purification. One μg of genomic DNA was then sheared with a Covaris sonicator (Covaris, Woburn, MA) to approximately 500 bp. Briefly, the sheared DNA was end-repaired, and an A-tail was added to facilitate adapter ligation. After adapters were ligated, the product was run on a 2% agarose gel and fragments between 350-500 bp were selected, purified, and PCR amplified for 14 cycles. Purified samples were analyzed for quality and size distribution on a Bioanalyzer 2100 (Agilent, Santa Clara, CA) and sequenced on an Illumina HiSeq 2000 sequencer at the Lewis-Sigler Institute Sequencing Facility (Princeton University, Princeton, NJ).

Raw 101 bp reads were trimmed to remove low quality bases (Phred quality score<20) and reads with fewer than 30 bp of contiguous high quality bases were removed using the script TQSfastq.py (<http://code.google.com/p/ngopt/source/browse/trunk/SSPACE/tools/TQSfastq.py>). For *X. birchmanni*,320,296,082 were obtained ([Schumer et al. 2012](#_ENREF_21)), 352,437,337 reads were obtained for *X. malinche*, and 138,371,315 reads were obtained for *X. clemenciae* ([Schumer et al. 2012](#_ENREF_21)). Between 98%-99% of reads mapped to the *X. maculatus* reference. Raw sequences are available on the NCBI Sequence Read Archive (Acc # SRA060275, SRA061485).

1. **BUCKy analysis using the *X. maculatus* genome**

Methods used for BUCKy analysis of 4,468 alignments > 1.5 kb to the *X. maculatus* genome followed methods described in the main text with slight modifications. Due to computational limitations of MrBayes 3.2.1, we split the dataset into 4 runs. Within each run, tree topologies, branch lengths, gene-specific rate multiplier were unlinked and we linked the GTR matrix, gamma and proportion of invariant sites to avoid over-parameterization. We ran the chains to stationary, sampling every 2000 generations. The following chain lengths and burn-in values were used: dataset 1 41.39 million (10 million burn-in), dataset 2 37.84 million (10 million burn-in), dataset 3 44.59 million (33 million burn-in), dataset 4 62.13 million (36 million burn-in). Initial analysis using the full dataset for BUCKy failed due to computational limitations. We then randomly reduced the dataset by retaining 50% (2,234) of the genes for BUCKy. BUCKy was run for 500,000 generations with an additional 50,000 as burn-ins.  We rooted the major concordance tree between southern swordtails and (platyfishes, northern swordtails) according to the rooting in the total evidence tree.

The major concordance tree is identical to the topology of the total evidence tree produced by the genome reference and the major concordance tree produced by BUCKy using the *X. birchmanni* reference. The concordance factors (CFs) of the major “splits” (or instances of major discordance) are consistently higher than those produced by the *X. birchmanni* reference (average difference 3%, paired t-test, p=6.8e-05), likely due to less missing data (Table S3). All 19 instances of major discordance identified in this analysis were also found in the BUCKy analysis using the *X. birchmanni* reference (Fig. S3).

1. **Compilation of MSP, sword length and sword preference data**

We compiled a dataset of sword length, sword production ability, and sword preference in *Xiphophorus*. *Xiphophorus* species naturally possessing a sword were deemed to have machinery for sword production (or MSP). In addition, *X. maculatus* and *X. milleri* produce short swords in response to androgen treatment and as a result are both considered to have MSP ([Dzwillo 1963](#_ENREF_6), [1964](#_ENREF_7); [Zander and Dzwillo 1969](#_ENREF_23); [Offen 2008](#_ENREF_14)). Average sword indices (normalized to standard length) were compiled from the literature (Table S4). The metric sword index includes the ventral length of the caudal fin. Unsworded species were assigned a sword index of 0.275; since no caudal fin lengths have been reported for platyfishes, this number was based on the caudal fin length without the sword reported in one study on swordtails ([Kallman et al. 2004](#_ENREF_8)).

We obtained data from previous studies on sword preference for 5 northern swordtails, 2 southern swordtails, 2 platyfishes and 1 outgroup (Table S5). Sword preference index is calculated as (Ts-Tu)/(Ts+Tu), where Ts is the association time with sworded stimuli and Tu is the association time with unsworded stimuli. Association time is used as a proxy for mating preference in *Xiphophorus* ([Cummings and Mollaghan 2006](#_ENREF_5)). When multiple experiments were performed, we used the average of reported values.

**Supplementary References**

Basolo, A. L. 1990. Female preference predates the evolution of the sword in swordtail fish. Science 250:808-810.

Basolo, A. L. 1995a. A further examination of a pre-existing bias favouring a sword in the genus *Xiphophorus*. Anim. Behav. 50:365-375.

Basolo, A. L. 1995b. A further examination of a pre-existing bias favouring a sword in the genusXiphophorus. Animal Behaviour 50:365-375.

Basolo, A. L. 1995c. Phylogenetic evidence for the role of a pre-existing bias in sexual selection. Proc. R. Soc. Lond. B 259:307-311.

Cummings, M. and D. Mollaghan. 2006. Repeatability and consistency of female preference behaviours in a northern swordtail, *Xiphophorus nigrensis* Anim. Behav. 72:217-224.

Dzwillo, M. 1963. Einfluß von Methyltestosteron auf die Aktivierung sekundärer Geschlechtsmerkmale Über den arttypischen Ausbildungsgrad hinaus. Ver Dent Zool Gesell 1962:152-159.

Dzwillo, M. 1964. Sekundäre Geschlechtsmerkmale an der Caudalflosse einiger Xiphophorini unter dem Einfiuß von Methyltestosteron. Mitt Ham Zool Mus 61:15-22.

Kallman, K. D., R. B. Walter, D. C. Morizot, and S. Kazianis. 2004. Two new species of *Xiphophorus* (Poeciliidae) from the Isthmus of Tehuantepec, Oaxaca, Mexico, with a discussion of the distribution of the *X. clemenciae* clade. Am. Mus. Novit.:1-34.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Lunter, G. and M. Goodson. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:936-939.

Meyer, A., W. Salzburger, and M. Schartl. 2006. Hybrid origin of a swordtail species (Teleostei : *Xiphophorus clemenciae*) driven by sexual selection. Mol. Ecol. 15:721-730.

Meyer, M. K. and M. Schartl. 1979. Eine neue *Xiphophorus*-Art aus Vera Cruz, Mexiko. Senckenbergiana biol. 60:147-151.

Meyer, M. K. and M. Schartl. 2002. *Xiphophorus mayae*, a new species of swordtail from Guatemala (Teleostei: Poeciliidae). Ichthyol. Explor. Freshwaters 13:59-66.

Offen, N. 2008. The molecular basis of development of the sword, a sexual selected trait in the genus *Xiphophorus*. Mathematisch-naturwissenschaftliche Sektion Fachbereich Biologie. Universität Konstanz, Konstanz, Germany.

Rauchenberger, M., K. D. Kallman, and D. C. Morizot. 1990. Monophyly and geography of the Río Pánuco Basin swordtails (genus *Xiphophorus*) with descriptions of four new species. Am. Mus. Novit. no. 2975:1-41.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala : revisionary studies and comparative geography. Bull. Am. Mus. Nat. Hist. 162:267-376.

Rosenthal, G. and M. Ryan. 2000. Visual and acoustic communication in non-human animals: a comparison. Journal of Biosciences 25:285-290.

Rosenthal, G. G. 2000. The behavioral ecology of visual signaling in swordtails. Zoology. University of Texas at Austin, Austin, Texas.

Rosenthal, G. G. and C. S. Evans. 1998. Female preference for swords in *Xiphophorus helleri* reflects a bias for large apparent size. Proc. Natl. Acad. Sci. 95:4431-4436.

Rosenthal, G. G., W. E. Wagner Jr, and M. J. Ryan. 2002. Secondary reduction of preference for the sword ornament in the pygmy swordtail *Xiphophorus nigrensis* (Pisces: Poeciliidae). Anim. Behav. 63:37-45.

Schumer, M., R. Cui, B. Boussau, R. Walter, G. G. Rosenthal, and P. Andolfatto. 2012. An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. Evolution doi: 10.1111/evo.12009.

Wong, Bob B. M. and Gil G. Rosenthal. 2006. Female disdain for swords in a swordtail fish. Am. Nat. 167:136-140.

Zander, C. D. and M. Dzwillo. 1969. Untersuchungen zur Entwicklung und Vererbung des Caudalfortsatzes der *Xiphophorus*-Arten (Pisces). Zeitschr Wissensch Zool 178:275-315.

**Table S1**. Sources of species and strains used in this study with pedigree information where available. Source of the specimen is indicated as: XGSC—*Xiphophorus* Genetic Stock Center, Texas State University, BFL—Brackenridge Field Laboratory University of Texas at Austin, Wild—wild caught individual, or Hobby—obtained from a fish collector. The asterisk designates that the specimen comes from the type locality. Hobby strains were directly obtained from Greg Sage (<http://selectaquatics.com/>).

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Species** | **Strain** | **Pedigree #** |
| Hobby | *Xiphophorus alvarezi* | "yellow" collection from Dave Macallister | N/A |
| XGSC | *X. andersi* | andC, Río Atoyac, Veracruz | 11480 |
| Wild | *X. birchmanni* | Río Garces, Hidalgo | N/A |
| XGSC | *X. clemenciae* | FincaII, San Carlos, Oaxaca | 11316 |
| XGSC | *X. continens*1 | contiIV, Río Ojo Frío, San Luis Potosí \* | 11520 |
| XGSC | *X. cortezi* | cortezi, Hidalgo | 11220 |
| XGSC | *X. couchianus* | Xc, Huasteca Canyon, Nuevo León | N/A |
| XGSC | *X. evelynae* | eve, lake near Necaxa, Hidalgo | 11394 |
| XGSC | *X. gordoni* | gordoni, Laguna Santa Tecla, Coahuila | 11692 |
| XGSC | *X. hellerii strigatus* | Sara, Río Sarabia near Oaxaca | N/A |
| XGSC | *X. maculatus* | Río Jamapa drainage, Veracruz | 11615 |
| Wild | *X. malinche* | Arroyo Xontla, Chicayotla, Hidalgo | N/A |
| Hobby | *X. mayae* | Rio Bellaire, Honduras2 | N/A |
| XGSC | *X. meyeri* | meyeri, Melchor Muzquiz, Coahuila | 11523 |
| XGSC | *X. milleri* | mil82, Catemaco, Veracruz | 11305 |
| XGSC | *X. montezumae* | Rascon, Río Ojo Frio in the Río Gallinas system, Damian Carmona, north of Rascon, San Luis Potosí \* | 11333 |
| XGSC | *X. monticolus* | Tej, El Tejón | 11355 |
| BFL | *X. multilineatus* | Río Coy at federal Highway 85, San Luis Potosí \* | N/A |
| BFL | *X. nezahualcoyotl* | Arroyo Gallitos, Tamaulipas \* | N/A |
| BFL | *X. nigrensis* | Río Choy, San Luis Potosí \* | N/A |
| BFL | *X. pygmaeus* | Nacimiento Río Huichihuayán, San Luis Potosi \* | N/A |
| XGSC | *X. signum* | Signum, from Dr. J. H. Schroder, Munich September 28, 1993 | 11294 |
| Wild | *X. variatus* | Río Calnali, Hidalgo. | N/A |
| XGSC | *X. xiphidium* | SC, Sierra San Carlos, Tamaulipas | 11557 |
| Wild | *Pseudoxiphophorus jonesii* | Río Calnali, Hidalgo | N/A |
| Wild | *Priapella compressa* | Río El Azufre, permit # DGOPA.00093.120110.-0018 | N/A |

**1** The individual identified as *X. continens* is closely related to *X. pygmaeus*, contradicting previous phylogenetic placements as sister to *X. montezumae*.Based on morphological similarity between *X. continens* and *X. pygmaeus*, misidentification of this individual is possible and these results are interpreted with caution.

**2** Controversy exists over whether distinct populations designated as *X. mayae* are in fact different species (Kallman & Kazianis, 2006)

**Table S2.** List of species included in AU test and D-statistic. In the topology ((species 1, species 2), species 3), species 4), the D-statistic tests gene flow between species 3 and species 1 or species 2, while species 4 is the outgroup. An additional outgroup (species 5) was used in AU tests.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species 1 | Species 2 | Species 3 | Species 4 | Species 5 |
| *X. hellerii* | *X. mayae* | *X. signum* | *X.birchmanni* | *X. malinche* |
| *X. mayae* | *X. alvarezi* | *X. hellerii* | *X.birchmanni* | *X. malinche* |
| *X. cortezi* | *X. montezume* | *X. nezahuacoyotl* | *X. hellerii* | *X. maculatus* |
| *X. meyeri* | *X. xiphidum* | *X. andersi* | *X. hellerii* | *X. maculatus* |
| *X. evelynae* | *X. milleri* | *X. andersi* | *X. hellerii* | *X. maculatus* |
| *X. evelynae* | *X. variatus* | *X. xiphidium* | *X. hellerii* | *X. maculatus* |
| *X. couchianus* | *X. evelynae* | *X. milleri* | *X. hellerii* | *X. maculatus* |
| *X. couchianus* | *X. variatus* | *X. evelynae* | *X. hellerii* | *X. maculatus* |
| *X. gordoni* | *X. meyeri* | *X. couchianus* | *X. hellerii* | *X. maculatus* |
| *X. malinche* | *X. birchmanni* | *X.pygmaeus* | *X. hellerii* | *X. maculatus* |

**Table S3.** Total number of reads and number of reads mapped for each species used in our analysis. Percent of reads mapped is given in terms of number of reads that mapped to the masked *X. birchmanni* transcriptome; with the exception of *Priapella*,90-99% of reads mapped to the unmasked transcriptome. Approximately 40% of the *Priapella* reads mapped to the unmasked transcriptome; however, over 90% of the reads mapped to the *X. maculatus* genome, suggesting that the mapping discrepancy with *Priapella* is due to a distinct gene expression profile in this sample. The same pattern is seen with *X. nigrensis*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **Number of mapped reads (total number of reads)** | **Percent of reads mapped** | **Index used** | **% missing data in BUCKy alignments**  **(*X. birchmanni* ref.)** | **% missing data in BUCKy alignments**  **(*X. maculatus* genome ref.)** |
| *X. alvarezi* | 4764604 (6727173) | 71% | CAGATC | 24.6% | 9.7% |
| *X. andersi* | 9605053 (13586654) | 71% | CTTGTA | 6.3% | 1.9% |
| *X. birchmanni* | 13152734 (20535466) | 64% | ACAGTG | 3.1% | 5.0% |
| *X. clemenciae* | 10543154 (14781899) | 71% | TAGCTT | 5.2% | 4.1% |
| *X. continens* | 12385269 (17219728) | 72% | GAGTGG | 4.0% | 1.3% |
| *X. cortezi* | 9832543 (13944987) | 71% | ACAGTG | 5.4% | 2.5% |
| *X. couchianus* | 8117587 (11460410) | 71% | GTCCGC | 8.3% | 3.3% |
| *X. evelynae* | 11833353 (16553199) | 71% | CGTACG | 5.0% | 1.8% |
| *X. gordoni* | 8137716 (11315115) | 72% | GGTAGC | 12.1% | 6.5% |
| *X. hellerii* | 8418680 (11917210) | 71% | GGCTAC | 8.0% | 2.7% |
| *X. maculatus* | 4988699 (6919889) | 72% | AGTCAA | 21.7% | 8.1% |
| *X. malinche* | 13350213 (17799782) | 75% | GCCAAT | 4.9% | 5.4% |
| *X. mayae* | 22854015 (32988520) | 69% | GCCAAT | 1.1% | 0.4% |
| *X. meyeri* | 11269787 (15795805) | 71% | GTGAAA | 4.8% | 1.8% |
| *X. milleri* | 6872855 (9411474) | 73% | ATCACG | 14.5% | 7.0% |
| *X. monticolus* | 5486910 (7700997) | 71% | TTAGGC | 19.6% | 10.3% |
| *X. montezumae* | 9149897 (12811964) | 71% | GTTTCG | 6.8% | 2.4% |
| *X. multilineatus* | 11612602 (16301024) | 71% | GTGGCC | 5.0% | 2.0% |
| *X. nigrensis* | 9213054 (21460448) | 43% | CGATGT | 8.2% | 4.2% |
| *X. nezahualcoyoytl* | 9244520 (12836405) | 72% | TGACCA | 17.0% | 3.7% |
| *X. pygmaeus* | 12950063 (18378794) | 70% | ATGTCA | 3.0% | 1.0% |
| *X. signum* | 12329297 (17820816) | 69% | AGTTCC | 3.4% | 1.1% |
| *X. variatus* | 8624103 (12262391) | 70% | ACTTGA | 7.5% | 2.6% |
| *X. xiphidium* | 4196660 (5927653) | 70% | GATCAG | 26.9% | 11.7% |
| *Pseudoxiphophorus jonesii* | 7443246 (11026293) | 68% | CCGTCC | 14.5% | 6.9% |
| *Priapella compressa* | 3611982 (11706270) | 31% | GTAGAG | 85.1% | 76.6% |

**Table S4.** Sword index (sword length/standard length) compiled from literature. Unsworded species were assigned a sword length value of 0.275, corresponding to the approximate relative caudal fin length.

|  |  |  |
| --- | --- | --- |
| **Species** | **Sword index** | **Source** |
| *Xiphophorus alvarezi* | 0.650 | ([Rosen 1979](#_ENREF_16)) |
| *X. andersi* | 0.350 | ([Meyer and Schartl 1979](#_ENREF_12)) |
| *X. birchmanni* | 0.275 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. clemenciae* | 0.564 | ([Kallman et al. 2004](#_ENREF_8)) |
| *X. continens* | 0.300 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. cortezi* | 0.370 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. couchianus* | 0.275 | N/A |
| *X. evelynae* | 0.275 | N/A |
| *X. gordoni* | 0.275 | N/A |
| *X. hellerii* | 0.641 | ([Rosen 1979](#_ENREF_16); [Kallman et al. 2004](#_ENREF_8)) |
| *X. maculatus* | 0.275 | N/A |
| *X. malinche* | 0.370 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. mayae* | 0.700 | ([Meyer and Schartl 2002](#_ENREF_13)) |
| *X. meyeri* | 0.275 | N/A |
| *X. milleri* | 0.275 | N/A |
| *X. montezumae* | 1.030 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. monticolus* | 0.517 | ([Kallman et al. 2004](#_ENREF_8)) |
| *X. multilineatus* | 0.400 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. nezahualcoyotl* | 0.480 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. nigrensis* | 0.370 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. pygmaeus* | 0.300 | ([Rauchenberger et al. 1990](#_ENREF_15)) |
| *X. signum* | 0.600 | ([Rosen 1979](#_ENREF_16)) |
| *X. variatus* | 0.275 | N/A |
| *X. xiphidium* | 0.300 | ([Basolo 1995b](#_ENREF_3)) |
| *Pseudoxiphophorus jonesii* | 0.275 | N/A |
| *Priapella compressa* | 0.275 | N/A |

**Table S5.** Sword preference of *Xiphophorus* and an outgroup species (*Priapella olmecae*) compiled from literature.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Source** | **Treatment** | **# male pairs** | **Mean sword difference**  **(mm)d** | **N** | **Interaction time** | | **Assoc. time difference (sec)** | **Sword preferencee** | **Statistical test** | **Test statistic** | ***p*** |
| **Sword**  **(sec)** | **no sword**  **(sec)** |
| *X. nigrensis* | [Rosenthal et al. (2002](#_ENREF_20)) | Sword/no-sword | 2 | 11.5b | 21 | 520 | 554 | -34 | -0.03 | Wilcoxon matched-pairs signed-rank test (WSR) | 0.50 (*Z*) | 0.61 |
| Animation: sword/no-sword | \* | 4.6b | 37 | 363 | 452 | -89 | -0.11 | Paired *t*-test | 1.04 (*t*) | 0.30 |
| Animation  *X. helleri*-like | \* | 5.2b | 16 | 463 | 622 | -159 | -0.15 | WSR | 1.97 (*Z*) | 0.05 |
| *X. helleri* | [Rosenthal and Evans (1998](#_ENREF_19)) | Animation: sword/no-sword | \* | 33.0a,c | 14 | 67 | 3 | 64 | 0.91 | WSR | -2.97 (*Z*) | <0.01 |
| *X. clemenciae* | [Meyer et al. (2006](#_ENREF_11)) | *X. helleri/X. maculatus* | *?* | 29.3a | 28 | 390 | 150 | 240 | 0.44 | WSR | -2.86 (*Z*) | <0.01 |
| *X. maculatus* | [Basolo (1990](#_ENREF_1)) | Yellow/clear | 6 | 24.0b | 9-16 (84) | 689 | 419 | 270 | 0.24 | WSR | Not available | <0.05 |
| *X. variatus* | [Basolo (1995c](#_ENREF_4)) | Yellow/clear | 6 | 20.5a | 6-13 (59) | 700 | 394 | 306 | 0.28 | Two-tailed binomial test | Not available | 0.03 |
| *X. birchmanni* | [Wong and Rosenthal (2006](#_ENREF_22)) | Animation:  sword/no-sword | \* | 7.3b | 18 | 150 | 300 | -150 | -0.33 | WSR | 2.59 (*Z*) | 0.01 |
| *X. malinche* | Rosenthal, unpublished data | Animation:  sword/no-sword | \* | 16.5a | 12 | 202 | 327 | -125 | -0.24 | Paired *t*-test | 1.33 (*t*) | 0.21 |
| *X. pygmaeus* | [Rosenthal (2000](#_ENREF_18)) | Animation:  sword/no-sword | \* | 4.4b | 208 | 218 | 228 | -10 | -0.02 | ANCOVA | 0.04/1.27/0.09 (*F*) | >0.05 |
| *X. multilineatus* | [Rosenthal and Ryan (2000](#_ENREF_17)) | Animation:  sword/no-sword | \* | 4.4b | 40 | 236 | 282 | -46 | -0.08 | ANCOVA | 0.62/0.34 (*F*) | >0.05 |
| *X. montezumae* | Unpublished data | Sword/no-sword | 1 | 67.1b | 6 | 353 | 210 | 143 | 0.19 | WSR | -0.73 (*F*) | 0.46 |
| *Priapella olmeca* | [Basolo (1995a](#_ENREF_2)) | Yellow/clear | 11 | 31.6b | 14 | 567 | 216 | 351 | 0.43 | Two-tailed binomial test | Not available | <0.05 |

a original study reported sword length (including length of caudal fin)

b sword measured as sword extension length (excluding caudal fin length);

c median (N=200) of the length of sword of *X. hellerii* from Basolo and Wagner (2004)

d Mean sword differences between the two male stimuli presented to the subjects in a binary choice test paradigm.

e Sword preference is calculated as the ratio of net association time and total trial duration.

**Table S6.** Bayesian concordance factors and 95% CI for bipartitions in the species tree (marked with #) and the alternative bipartitions (marked with \*). Only the bipartitions common to both references are shown. Abbreviations: *couchianus* group – *X. couchianus*, *X. meyeri*, *X. gordoni*. See main text for other abbreviations.

|  |  |  |
| --- | --- | --- |
| **Monophyletic clade** | **CF *X. birchmanni* reference [95% CI]** | **CF *X. maculatus* reference [95% CI]** |
| *\* X. maculatus*, Southern Swordtails | 0.114 [0.107, 0.121] | 0.112 [0.105, 0.119] |
| *\** Platyfishes, *hellerii* group | 0.101 [0.095, 0.107] | 0.105 [0.099, 0.112] |
| *\* X. andersi*, *X. xiphidium* | 0.131 [0.124, 0.139] | 0.143 [0.136, 0.150] |
| *\* X. xiphidium*, Northern Platys | 0.177 [0.169, 0.185] | 0.197 [0.190, 0.205] |
| *\* X. xiphidium*, *X. variatus* and *X. couchianus* group | 0.162 [0.154, 0.169] | 0.175 [0.167, 0.182] |
| *\* X. xiphidium*, *X. couchianus* group | 0.118 [0.110, 0.125] | 0.127 [0.120, 0.135] |
| *\* X. xiphidium, X. variatus* | 0.111 [0.104, 0.118] | 0.115 [0.108, 0.123] |
| *\* X. milleri, X. evelynae* | 0.109 [0.102, 0.117] | 0.105 [0.099, 0.112] |
| *\* X. evelynae, X. variatus* | 0.137 [0.128, 0.145] | 0.146 [0.137, 0.155] |
| *\* X. couchianus, X. gordoni* | 0.278 [0.262, 0.293] | 0.281 [0.263, 0.299] |
| *\* X. couchianus, X. meyeri* | 0.294 [0.279, 0.310] | 0.302 [0.284, 0.320] |
| *\** All Northern Swordtails except *X. montezumae* | 0.100 [0.092, 0.109] | 0.138 [0.127, 0.149] |
| *\* X. nezahualcoyotl, nigrensis* group | 0.169 [0.163, 0.175] | 0.137 [0.132, 0.142] |
| *\* X. nezahualcoyotl, X. montezumae* | 0.251 [0.243, 0.260] | 0.203 [0.195, 0.210] |
| *\* X. cortezi, X. birchmanni, X. malinche* | 0.292 [0.283, 0.301] | 0.219 [0.210, 0.228] |
| *\* X. signum, X. alvarezi, X. mayae* | 0.220 [0.210, 0.230] | 0.248 [0.237, 0.258] |
| *\* X. signum, X. mayae* | 0.213 [0.204, 0.222] | 0.195 [0.186, 0.204] |
| *\* X. mayae, X. hellerii* | 0.135 [0.126, 0.143] | 0.127 [0.119, 0.136] |
| *\* X. hellerii, X. alvarezi* | 0.211 [0.202, 0.220] | 0.204 [0.195, 0.214] |
| # *X. clemenciae, X. monticolus* | 0.657 [0.648, 0.667] | 0.681 [0.672, 0.690] |
| # *hellerii* group | 0.699 [0.691, 0.708] | 0.763 [0.755, 0.772] |
| *# X. hellerii, X. alvarezi, X. mayae* | 0.329 [0.319, 0.340] | 0.363 [0.353, 0.374] |
| *# X. alvarezi, X. mayae* | 0.408 [0.398, 0.418] | 0.456 [0.447, 0.466] |
| *# X. continens, X. pygmaeus* | 0.728 [0.718, 0.738] | 0.757 [0.747, 0.767] |
| *# nigrensis* group | 0.651 [0.643, 0.659] | 0.721 [0.713, 0.728] |
| *# X. nigrensis, X. multilineatus* | 0.632 [0.620, 0.643] | 0.720 [0.709, 0.731] |
| *#* Northern swordtails | 0.715 [0.706, 0.724] | 0.708 [0.700, 0.716] |
| *# montezumae* group | 0.284 [0.274, 0.293] | 0.329 [0.318, 0.340] |
| *# cortezi group,* including *X. nezahualcoyotl* | 0.153 [0.145, 0.161] | 0.284 [0.275, 0.293] |
| *# X. nezahualcoyotl, X. cortezi* | 0.329 [0.321, 0.337] | 0.447 [0.438, 0.455] |
| *# X. birchmanni, X. malinche* | 0.546 [0.535, 0.557] | 0.560 [0.549, 0.570] |
| *#* Platyfishes | 0.349 [0.339, 0.358] | 0.361 [0.352, 0.370] |
| *#* Platyfishes except *X. maculatus* | 0.229 [0.219, 0.239] | 0.248 [0.239, 0.258] |
| *#* Platyfishes except *X. maculatus* & *X. andersi* | 0.192 [0.184, 0.199] | 0.202 [0.194, 0.210] |
| *# X. milleri,* Northern Platyfishes | 0.273 [0.266, 0.280] | 0.302 [0.295, 0.309] |
| *#* Northern Platyfishes | 0.272 [0.263, 0.281] | 0.300 [0.290, 0.310] |
| *# X. variatus*, *couchianus* group | 0.380 [0.369, 0.390] | 0.402 [0.391, 0.414] |
| *# couchianus* group | 0.846 [0.838, 0.854] | 0.897 [0.889, 0.904] |
| *# X. gordoni, X. meyeri* | 0.355 [0.339, 0.370] | 0.369 [0.351, 0.386] |

**Table S7.** Pairwise distance (GTR+ Γ) between partial sample pairs ranked by distance. Distance calculated from concatenating 1490 partitions (each >500bp, total length 2.56Mb) produced by mapping to *X. maculatus* genome with a stringent variant filter (5x cutoff, 0.17% missing data). These distance estimates can be compared to the divergence estimates in Table 1. Strain abbreviations indicate sampling location: JpB—Jamapa B, JpWild—Jamapa Wild, CHIC—Chicayotla, COAC—Coacuilco, GARC—Garces.

|  |  |  |  |
| --- | --- | --- | --- |
| **Relationship** | **Sample 1** | **Sample 2** | **ML Dist** |
| Same population | *X. malinche* CHIC1 | *X. malinche* CHIC2 | 0.000416 |
| Same population | *X. maculatus* JpWild | *X. maculatus* JpB | 0.000549 |
| Sister species | *X. meyeri* | *X. gordoni* | 0.000678 |
| Non-sister species | *X. meyeri* | *X. couchianus* | 0.000701 |
| Non-sister species | *X. couchianus* | *X. gordoni* | 0.000726 |
| Sister species | *X. continens* | *X. pygmaeus* | 0.000932 |
| Sister species | *X. nigrensis* | *X. multilineatus* | 0.001276 |
| Different population | *X. birchmanni* GARC | *X. birchmanni* COAC | 0.001276 |
| Non-sister species | *X. pygmaeus* | *X. multilineatus* | 0.002267 |
| Non-sister species | *X. nigrensis* | *X. pygmaeus* | 0.002389 |
| Non-sister species | *X. continens* | *X. multilineatus* | 0.002520 |
| Non-sister species | *X. continens* | *X. nigrensis* | 0.002648 |
| Different population | *X. clemenciae* FincaII | *X. clemenciae* Grande | 0.002655 |
| Sister species | *X. malinche* CHIC2 | *X. birchmanni* GARC | 0.002808 |
| Sister species | *X. malinche* CHIC1 | *X. birchmanni* COAC | 0.002842 |
| Sister species | *X. malinche* CHIC2 | *X. birchmanni* COAC | 0.003046 |
| Sister species | *X. malinche* CHIC1 | *X. birchmanni* GARC | 0.003062 |
| Sister species | *X. mayae* | *X. alvarezi* | 0.003427 |
| Non-sister species | *X. malinche* CHIC2 | *X. cortezi* | 0.003728 |
| Sister species | *X. cortezi* | *X. nezahualcoyotl* | 0.003856 |
| Non-sister species | *X. hellerii* | *X. alvarezi* | 0.003896 |
| Non-sister species | *X. malinche* CHIC1 | *X. cortezi* | 0.003979 |

**Supplementary Figures**

**Figure S1.** A) Total evidence phylogeny constructed from alignments to the *X. maculatus* genome. Differences from the total evidence phylogeny presented in the main text are highlighted in blue. Additional genome sequence data has been added for *X. clemenciae*, *X. malinche*, and *X. birchmanni*. B) Total evidence phylogeny constructed from alignments to the *X. mayae* transcriptome. Differences from the total evidence phylogeny presented in the main text are highlighted in blue. Nodal support generated by 100 rapid bootstraps with GTR+CAT.

FigureS1.tif

**Figure S2.** Rooted mitochondrial phylogeny of *Xiphophorus* using 20x coverage cutoff (coding regions only, 15,787 bp, 37.8% missing).Nodal support generated by 100 rapid bootstraps with GTR+CAT. The placement of *X. maculatus* differs from the unrooted mitochondrial phylogeny (Fig. 2), but with weak bootstrap support.

**K:\Ray\learning\lab\phylogenetics\ms\phylogenetics_manuscript\upload\FigureS2.tif**

**Figure S3.** Comparisons between BUCKy results for alignments to the *X. birchmanni* transciptome and BUCKy results for alignments to the *X. maculatus* genome reveal highly similar patterns of discordance. Bayesian concordance factors are marked in pink for the *X. maculatus* reference dataset and blue for the *X. birchmanni* reference dataset. Alternative splits recovered at CFs>10% in both analyses are marked black. Splits marked in blue were only found using the *X. birchmanni* as reference; no splits were unique to the *X. maculatus* genome analysis. Refer to Table S6 for confidence intervals for CFs.

