doi:10.1111/evo.12099

PHYLOGENOMICS REVEALS EXTENSIVE RETICULATE EVOLUTION IN *XIPHOPHORUS* FISHES

Rongfeng Cui,[1,2,3] Molly Schumer,[4] Karla Kruesi,[5] Ronald Walter,[6] Peter Andolfatto,[4,7] and Gil G. Rosenthal[1,2]

[1]*Department of Biology, Texas A&M University, TAMU, College Station, Texas 77843*

[2]*Centro de Investigaciones Científicas de las Huastecas "Aguazarca", Calnali, Hidalgo, Mexico*

[3]*E-mail: rcui@mail.bio.tamu.edu*

[4]*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey*

[5]*Departamento de Ecología y Recursos Naturales, Universidad Nacional Autónoma de México, Mexico City, Mexico*

[6]*Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas*

[7]*Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey*

Received January 17, 2013
Accepted February 26, 2013
Data Archived: Dryad doi:10.5061/dryad.8bt72

Hybridization is increasingly being recognized as a widespread process, even between ecologically and behaviorally divergent animal species. Determining phylogenetic relationships in the presence of hybridization remains a major challenge for evolutionary biologists, but advances in sequencing technology and phylogenetic techniques are beginning to address these challenges. Here we reconstruct evolutionary relationships among swordtails and platyfishes (*Xiphophorus*: Poeciliidae), a group of species characterized by remarkable morphological diversity and behavioral barriers to interspecific mating. Past attempts to reconstruct phylogenetic relationships within *Xiphophorus* have produced conflicting results. Because many of the 26 species in the genus are interfertile, these conflicts are likely due to hybridization. Using genomic data, we resolve a high-confidence species tree of *Xiphophorus* that accounts for both incomplete lineage sorting and hybridization. Our results allow us to reexamine a long-standing controversy about the evolution of the sexually selected sword in *Xiphophorus*, and demonstrate that hybridization has been strikingly widespread in the evolutionary history of this genus.

KEY WORDS: **Hybridization, incomplete lineage sorting, premating isolation, sexually selected traits, transcriptome sequencing.**

A growing body of work has begun to recognize the importance of hybridization in the evolutionary process (Mallet 2007, 2008; Abbott et al. 2010; Arnold and Martin 2010). There is increasing awareness that many species diversify in the presence of ongoing gene flow, or experience post-speciation gene flow without the collapse of reproductive isolation. Some of the most rapidly diversifying groups, such as African cichlids (e.g., Seehausen 2004; Schwarzer et al. 2012) and *Heliconius* butterflies (Mallet 2005), have weak postzygotic isolation and frequently hybridize. In certain species groups, the spread of adaptive alleles through hybridization is thought to underlie phenotypic diversification (Rieseberg et al. 2003; Heliconius Genome Consortium 2012). It has also been suggested that hybridization can lead to speciation in some cases (Mallet 2007, 2008; Abbott et al. 2010; Arnold and Martin 2010). Understanding the extent of hybridization, and the role of hybridization in evolution, is an important focus of current evolutionary research.

Large, genome-wide datasets can be used both to address phylogenetic relationships between species and examine patterns of incomplete lineage sorting and hybridization. Historically, it has been difficult to determine phylogenetic relationships in groups with hybridization or high levels of incomplete lineage

EVOLUTION INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

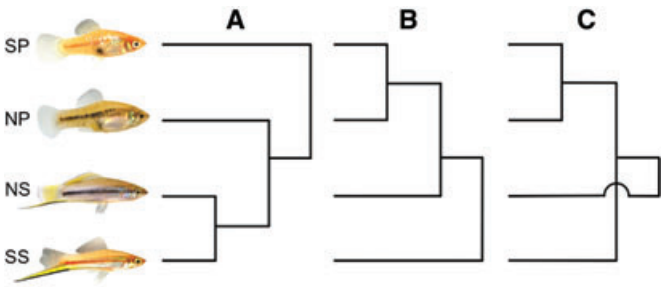

Figure 1. Previous phylogenetic hypotheses for interclade relationships in *Xiphophorus*. Abbreviations: NP, northern platyfishes; SP, southern platyfishes; NS, northern swordtails; SS, southern swordtails. (A) Platyfishes are paraphyletic and basal to genus (Rosen, 1979; Rauchenberger et al. 1990; Basolo 1990). (B) Swordtails are paraphyletic and northern swordtails are grouped with platyfishes (Meyer et al., 2006). (C) Swordtails are paraphyletic and southern swordtails are grouped with platyfishes (Meyer et al., 1994).

sorting (reviewed in Degnan and Rosenberg 2009), but the availability of genome-wide data and new computational techniques (Ané et al. 2007; Kubatko 2009; Pickrell and Pritchard 2012) allows researchers to explicitly account for incomplete lineage sorting and hybridization when constructing species trees (Pollard et al. 2006; Cranston et al. 2009). Previous studies have shown that next-generation sequencing data have the potential to generate well-resolved gene trees for testing hypotheses of hybridization (Hittinger et al. 2010). Despite this, very few phylogenetic studies to date have examined hybridization, particularly in a large group of species. Many phylogenetic studies that have examined hybridization have not explicitly accounted for incomplete lineage sorting (Decker et al. 2009; Schwarzer et al. 2012), used candidate genes (Heliconius Genome Consortium 2012), small numbers of genes (Hailer et al. 2012), cyto-nuclear discordance (Meyer et al. 2006; Aboim et al. 2010; Kang et al. 2013), or been unable to investigate hybridization due to the use of concatenated datasets (Nabholz et al. 2011; dos Reis et al. 2012).

Swordtails and platyfishes (Poeciliidae, genus *Xiphophorus*) are longstanding models of sexual selection (Darwin 1859; Ryan 1990), evolutionary genetics (Schartl 1995; Basolo 2006), and oncology (reviewed in Meierjohann and Schartl 2006). Numerous phylogenetic studies have failed to reach consensus on many aspects of their evolutionary relationships (Rosen 1960, 1979; Rauchenberger et al. 1990; Haas 1992; Marcus and McCune 1999; Morris et al. 2001; Meyer et al. 2006; Kang et al. 2013). Early studies placed the swordtails as a monophyletic group derived from the more basal platyfishes (Fig. 1A). More recent studies have proposed conflicting topologies for three clades: northern swordtails, southern swordtails, and platyfishes (northern and southern; Fig. 1B, 1C; Meyer et al. 1994).

Although the inconsistency among phylogenetic studies may be the result of insufficient data or errors in inference, another likely cause is interspecific hybridization. *Xiphophorus* species have weak postzygotic isolation (Morizot et al. 1991; Walter et al. 2004) and there are contemporary natural hybrid zones between multiple pairs of species (Rosenthal et al. 2003; Schartl 2008). Premating barriers play a primary role in reproductive isolation between species (Schartl 2008). As a rule, females prefer to mate with conspecifics over heterospecifics (McLennan and Ryan 1999; Hankison and Morris 2003), but these preferences are susceptible to ecological disturbances and plasticity in preferences (Fisher et al. 2006; Verzijden and Rosenthal 2011; Willis et al. 2011, 2012). It is therefore likely that episodes of hybridization have played an important role in *Xiphophorus* evolution and may account for conflicting phylogenies.

Uncertainties about the *Xiphophorus* phylogeny have also contributed to controversies about trait evolution in the genus. *Xiphophorus* has been intensively studied in the context of the preexisting bias hypothesis of sexual selection (Basolo 1990; Rosenthal and Evans 1998). The sexually dimorphic "sword" extension of the caudal fin in males of some *Xiphophorus* species was proposed to have evolved in response to a latent bias in females, but this hypothesis hinged on the basal placement of platyfishes and monophyly of the swordtails (Basolo 1990). This view was challenged by subsequent phylogenies that did not support the monophyly of northern and southern swordtails (Meyer et al. 1994); but see (Basolo 1995b). Without a robust phylogenetic framework, it is difficult to draw inferences about whether or not preexisting preferences drove the evolution of the sword ornament in *Xiphophorus*.

Hybridization could also account for the discontinuous distribution of traits in the genus. For example, the geographically separated *Xiphophorus andersi* and *Xiphophorus xiphidium* are the only two platyfishes naturally expressing swords (albeit short and unpigmented). Also, the sympatric swordtail *Xiphophorus birchmanni* and platyfish *Xiphophorus variatus* (Fisher et al. 2009) share several male sexually dimorphic traits, despite no evidence of current hybridization (Rauchenberger et al. 1990; G. G. Rosenthal, pers. obs.). Although these shared male sexually dimorphic traits could be the result of convergent evolution, interfertility between species raises the possibility that gene flow may be responsible for some shared traits.

To evaluate the role of hybridization in shaping the interrelationships among *Xiphophorus* species, we use an RNAseq-based approach to collect transcriptome sequence data for 24 of the 26 described *Xiphophorus* species and two outgroups. Despite their increasing use in population genetics and gene mapping studies, next-generation sequencing techniques have had limited applications in phylogenetics thus far. Although a variety of next-generation sequencing methods can be used to generate

phylogenetic datasets, for recently diverged species such as *Xiphophorus* (∼4–6 mya; Mateos et al. 2002; Kallman and Kazianis 2006), techniques such as RNAseq that generate long alignments are promising (e.g., Kocot et al. 2011; Smith et al. 2011). Although RNAseq datasets have been concatenated to produce highly supported phylogenies (Hittinger et al. 2010; Nabholz et al. 2011), concatenated datasets cannot be used to investigate gene flow. We use RNAseq data to resolve the *Xiphophorus* species tree with high confidence and apply phylogenetic approaches to identify cases of interspecific hybridization. We find evidence for widespread hybridization in *Xiphophorus* and discuss our findings in the context of sexual selection and trait evolution.

Materials and Methods

RNA EXTRACTION AND LIBRARY PREPARATION

We obtained a single live male of each of 24 *Xiphophorus* species and two outgroups from a variety of sources (Table S1). Fresh specimens were photographed and killed with an overdose of tricaine methanesulfate (MS-222) in compliance with Texas A&M University IACUC protocol #2012-164. Whole brains were dissected immediately and preserved in RNAlater (Ambion, Austin, TX) at -20°C until use.

RNA was extracted using a standard TRI Reagent protocol (Ambion) following manufacturer's instructions and quantified with a Nanodrop 1000 (Thermo Scientific, Wilmington, DE). One to four micrograms of total RNA were used to prepare libraries following Illumina's TruSeq mRNA Sample Prep Kit with minor modifications. Briefly, mRNA was purified using a bead-based protocol and chemically fragmented; first- and second-strand cDNA was synthesized from these fragments and end repaired. Following end repair, 3' ends were adenylated and a custom adapter was ligated. This allowed for one of 23 unique indices (Table S2) to be added to each library during PCR amplification (16–18 cycles). Following agarose gel purification of the desired size distribution (350–500 bp) and quality verification on the Bioanalyzer 2100 (Agilent, Santa Clara, CA), this 23-plex library was sequenced on one paired-end Illumina HiSeq 2000 lane at the Lewis-Sigler Institute Microarray Facility (Princeton, NJ). Samples from three additional species that could not be uniquely indexed were sequenced on partial lanes (*X. birchmanni*, *X. malinche*, and *X. nigrensis*). Raw 101 base pair (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).

All raw data have been deposited in NCBI's Sequence Read Archive (accession number: SRA061485).

TRANSCRIPTOME ASSEMBLY

For phylogenetic analysis, we analyzed alignments to the *Xiphophorus maculatus* genome and two transcriptomes to demonstrate that our results are robust to the effects of reference species and reference sequence type (transcriptome or genome). The two transcriptomes were generated using paired-end data for the two species with the highest coverage, *X. birchmanni* and *X. mayae*. A total of 29,535,466 (*X. birchmanni*) and 31,993,860 (*X. mayae*) paired-end reads were assembled using velvet (Zerbino and Birney 2008) with a range of kmers (21, 31, and 41). We initially used the merge assemblies option with Oases (Schulz et al. 2012), but found that this increased the incidence of highly divergent alignments in our later analysis, likely as a result of the formation of chimeric transcripts. We therefore used Oases to construct transcript isoforms without merging and used a custom perl script to select the longest isoform for our transcriptome reference. We compared the quality of different assemblies based on the assembly N50 and total number of bases assembled. In both cases, using a kmer of 31 produced the highest quality transcriptome: *X. birchmanni*—N50 of 2441 bp, total size 108 Mbp; *X. mayae*-–N50 of 3545 bp, total size 147 Mbp.

Using this preliminary transcriptome assembly for each species, we then identified regions that were not unique in the *X. maculatus* genome by BLASTing all transcripts to the *X. maculatus* genome using the blastn algorithm (Camacho et al. 2009). To be conservative, we masked regions of transcripts that had multiple hits to the genome at an *e*-value $< 1e-5$, by converting the bases in that region to *N*s. Many transcripts had short regions that BLASTed with high confidence to multiple sites in the genome. To avoid excluding regions that would not result in incorrect mapping we performed simulations to determine whether reads originating from another transcript would incorrectly map to small homologous regions ($<$100 bp) in our transcriptome. We found that regions with multiple BLAST hits shorter than 70 bp could be included in our analysis (Supporting Information section i). We completed all subsequent steps in the analysis using the *X. birchmanni* transcriptome, but repeated the analyses using both the *X. mayae* transcriptome assembly and *X. maculatus* genome to examine the effect of choice of reference sequence, and effects of using the transcriptome versus the genome assembly (Supporting Information section ii, Fig. S1). We chose the *X. birchmanni* transcriptome assembly for our main analysis because it resulted in more alignments that passed our criteria than the *X. mayae* assembly, and is less likely to combine heterogeneous sequences in gene tree analysis than alignments to the *X. maculatus* genome (see Supporting Information section ii).

ALIGNMENT GENERATION AND TOTAL EVIDENCE PHYLOGENY

Trimmed reads were aligned to the *X. birchmanni* transcriptome sequence and the *X. maculatus* mitochondrial genome (GenBank Accession No.: AP005982.1) using STAMPY v1.0.17 (Lunter and Goodson 2011). The number of reads per species and alignment statistics are summarized in Table S2. Mapped reads were analyzed for variant sites and sequence depth in each species using the samtools / bcftools pipeline (Li et al. 2009) with a mapping quality cutoff of 20. Because the mapping process generates aligned sequences to the reference, traditional gene by gene global alignment is unnecessary.

A custom PHP script was used to generate sequence alignments for all 26 species based on the bcf file that is the output of the samtools / bcftools pipelines. Before analyzing phylogenetic relationships based on these sequence alignments, we performed a number of quality control steps to exclude low-quality and low-information sites from our dataset. For each species, transcripts with average per-site coverage $<5\times$ were excluded. Individual sites within a transcript were masked as N if coverage at that site was $<5\times$ or the variant quality score was <20. Sites containing polymorphism or indels were also masked. After this initial masking, we compared the remaining sites between species. If a site had been masked in 90% or more of the *Xiphophorus* species, or the transcript had been excluded in both outgroup species, we excluded that site from out analysis. We also excluded regions of high divergence (more than 7 character differences from the *X. birchmanni* sequence in 21 bp) using a sliding window. All remaining sites were included in the analysis. This resulted in 10,999 alignments $\geq$500 bp with a total alignment length of 16.85 Mbp (22.91% missing data).

To produce a total evidence phylogeny as a first approximation of the likely species tree, all alignments were concatenated for analysis. Because of the large *a priori* partition numbers we did not allow free model parameters for each partition. Instead, we analyzed the dataset as a single partition with 100 rapid bootstraps (Stamatakis et al. 2008) followed by maximum likelihood tree estimation using the General Reversible Time substitution model (GTR + T) in RAxML 7.2.8 (Stamatakis 2006). Similar methods were used to produce total evidence phylogenies based on alignments to the *X. mayae* transcriptome and *X. maculatus* genome (Supporting Information section ii).

GENE TREE ANALYSIS WITH BUCKy

Gene trees can sometimes produce topologies that are different from the species tree because of incomplete lineage sorting or hybridization (Degnan and Rosenberg 2009). To explicitly account for these factors, we performed gene tree analysis using BUCKy (Ané et al. 2007). We initially attempted gene tree analysis using both the programs BEST (Liu 2008) and BUCKy (Ané et al. 2007) but were unable to use BEST due to computational limitations. The BUCKy program uses Bayesian concordance analysis to estimate the likely species tree topology, and what proportion of loci support the dominant topology. It has the advantage of making no assumptions about the source of the discordance. Therefore, in addition to incomplete lineage sorting, BUCKy accommodates hybridization (Larget et al. 2010). We treat instances of major discordance identified by BUCKy as potential cases of hybridization for further investigation.

Because of limitations in computational speed we used a smaller dataset for BUCKy analysis (transcripts $>$ 1.5 kb in length, 7.6 Mb) and excluded both outgroups from analysis due to extensive missing data in *Priapella* (Table S3). As a result of computational constraints, we divided this 2366 transcript dataset into two 1183 transcript datasets for analysis with MrBayes 3.2.1 to obtain a posterior distribution of gene trees for each partition. Stationary stage was determined by inspecting parameter traces using Tracer. The following chain length (and burn-ins) were used: run 1, 31.2 million (10 million); run 2, 28.6 million (10 million). BUCKy was run with a range of α priors, which describes the expected number of unique gene trees given the number of taxa and the total number of genes in the genome. We did not find differences in the topology or concordance factor (CF) values between α values of 1, 2, and 5, and report results obtained with $\alpha = 1$. The CF describes the estimated proportion of the genome significantly supporting a topology and BUCKy systematically underestimates the CF for the major topology, especially with short branch lengths and in the presence of ILS or hybridization (Chung and Ané 2011). Alternative bipartitions with CFs higher than 10% were further investigated for evidence of hybridization. This large CF cutoff was used to identify species with strong evidence for gene tree-species tree discordance. We also repeated the BUCKy analysis using alignments to the *X. maculatus* genome (Supporting Information sections iii and iv).

MITOCHONDRIAL PHYLOGENY

Using the same methods outlined earlier, but with a $20\times$ coverage cutoff, we obtained concatenated mitochondrial alignments of all coding regions for 26 species (15,787 bp, 42.63% missing). We repeated the analysis excluding both outgroups to prevent long-branch attraction. Both datasets were analyzed with RAxML 7.8.2 using GTR + Γ. Nodal support was determined with 100 rapid bootstraps in RAxML.

INVESTIGATING POTENTIAL HYBRIDIZATION

Strong discordance between gene trees and the species tree can be caused by incomplete lineage sorting or hybridization. To investigate whether incomplete lineage sorting can be rejected as the cause of discordance, we used the approximately unbiased (AU) test (Shimodaira 2002; Schumer et al. 2012) and Patterson's

D-statistic (Green et al. 2010) to further investigate all groups in which we found strong evidence of discordance using BUCKy (CF $\geq$ 10%). Both tests were used to measure whether there was significant asymmetry in support for the two minor topologies in a four species tree; significant asymmetry in support for one of the two minor topologies can be a sign of hybridization (Meng and Kubatko 2009; Durand et al. 2011). We only investigated potential hybridization that occurred between extant species and we did not investigate potential hybridization between *X. andersi* and *Xiphophorus milleri* because it was not detected in BUCKy analysis of alignments to the *X. maculatus* genome (Supporting Information section iv). Two potential hybridization events could not be examined using our approach because of species interrelationships; these were the potential hybridization events between *Xiphophorus hellerii* and *X. mayae*, and *Xiphophorus couchianus* and *Xiphophorus meyeri*, because higher discordance occurred between *X. hellerii* and *X. alvarezi* and *X. couchianus* and *X. meyeri*, respectively (see Results).

We exported four (*D*-statistic) or five (AU test) taxon alignments for the eight pairs of species for which we found strong evidence of gene tree discordance from BUCKy (excluding *X. hellerii–X. mayae* and *X. couchianus–Xiphophorus gordoni*), a closely related species, and two outgroups (detailed in Table S2). To investigate a case of mito-nuclear discordance, we also tested for gene flow between *X. birchmanni* and *X. pygmaeus* despite the fact that we failed to detect nuclear discordance between these two species (see Results). Quality criteria for alignments were as described earlier, except missing data at a site in any species resulted in that site being excluded from the analysis.

For the AU test, we enforced monophyly of the two outgroups and tested support for the three possible topologies of an unrooted 4-taxa tree. Site likelihoods were calculated using RAxML 7.8.2 with the General Time Reversible model and a gamma distribution of substitution rates (GTR + GAMMA). These likelihoods were used as input for AU tests implemented in Consel 0.2 (Shimodaira and Hasegawa 2001). The AU *P*-value is the probability that a tree is as likely as the maximum likelihood tree. If a particular topology had an AU *P*-value greater than 0.95 for an alignment, we concluded that the alignment supported that topology. We excluded partitions that had an observed likelihood difference of 0 because these are likely caused by low numbers of informative characters in the alignment (Schmidt 2009). We determined whether there was significant asymmetry in support for the two minor topologies by calculating 95% confidence intervals using 1000 replicates of nonparametric bootstrapping.

As a secondary method, we calculated Patterson's *D*-statistic (Green et al. 2010) for the same alignments using only one outgroup (Table S2). Sites with ABBA (shared sites between species 2 and 3) and BABA (shared sites between species 1 and 3) patterns were counted in the four species alignments. With the null hypothesis of incomplete lineage sorting, the number of ABBA and BABA sites is expected to be equal. Significant deviation of *D* from 0 suggests that incomplete lineage sorting can be rejected as a null hypothesis. Significance of *D* was determined using a two sample *z*-test. Standard error was determined by jack-knife bootstrapping of *D* for each transcript using the bootstrap package in R (R Development Core Team 2010), we then compared the *D*-statistic results with the AU results (Table 1); in cases in which all three methods (BUCKy, AU test, and Patterson's *D*-statistic) support hybridization, we conclude that hybridization likely occurred between the taxa in question.

CHARACTER MAPPING AND INDEPENDENT CONTRASTS

To examine patterns of sword evolution and sword preference evolution in the context of our species tree topology, we compiled a dataset of sword index (sword length / standard length; Basolo 1995a), machinery for sword production (MSP, the ability to generate a sword after androgen treatment: Gordon et al. 1943; Dzwillo 1963, 1964; Zander and Dzwillo 1969), and sword preference in *Xiphophorus* females based on previously published studies (Supporting Information section v, Tables S4 and S5). Although previous researchers have distinguished between the long and pigmented sword found in southern and some northern swordtails, and the short unpigmented sword found in two platyfishes, we do not make that distinction here. However, we recognize that the sword is a complex trait and has many morphological differences between clades.

Sword index and female preference index were coded as continuous variables whereas MSP was coded as a binary character. We traced the characters on the total evidence tree produced by alignment to the *X. birchmanni* transcriptome by maximum parsimony in Mesquite 2.75 (Maddison and Maddison 2011), because maximum likelihood methods cannot be implemented for continuous variables in this program. We used the PDAP-PDTREE package (Midford et al. 2011) in Mesquite to perform independent contrast analysis (Felsenstein 1985) between sword index and preference for swords and linear regressions between sword and preference characteristics and node height.

Results

A HIGH CONFIDENCE SPECIES TREE BASED ON NUCLEAR DATA

We used multiple methods to construct a high confidence species tree for *Xiphophorus*. As an initial approach to determine the likely species tree, we constructed a total evidence phylogeny using RAxML based on alignments to the *X. birchmanni* transcriptome; this concatenated nuclear dataset produced a fully

Table 1. Proportion of the gene trees supporting alternative topologies estimated by approximately unbiased tests (AU tests) and *D*-statistic calculated for all species with discordance > 10% identified by BUCKy. Confidence intervals are calculated by 1000 replications of nonparametric bootstraps for AU tests and jack-knife bootstrapping for the *D*-statistic. Positive values support gene flow between the two species in question. Divergence refers to the average sequence divergence between the two sister species in the genomic regions significantly (AU *P* > 0.95) supporting the indicated sister relationships.

Species pair	CF	*D*-statistic (jack-knife SE, *P*-value)	Percent of AU test support for the two minor topologies (95% confident intervals)	Divergence
X. nezahualcoyotl –X. montezumae	0.25	0.38 (0.07, $P<3.1e-14$)	([*X. nezahualcoyotl, X. montezumae*], *X. cortezi*): 32% (30–33%)	0.36%
			([*X. montezumae, X. cortezi*], *X. nezahualcoyotl*): 3% (2–3%)	0.36%
X. signum–X. mayae	0.21	0.56 (0.03, $P<4.4e-65$)	([*X. signum, X. mayae*], *X. hellerii*): 51% (49–53%)	0.40%
			([*X. signum, X. hellerii*], *X. mayae*): 7.6% (7–8.5%)	0.40%
X. hellerii–X. alvarezi	0.21	−0.41 (0.01, $P<6.4e-100$)	([*X. hellerii*, *X. alvarezi*], *X. mayae*): 16% (15–17%)	0.32%
			([*X. hellerii*, *X. mayae*], *X. alvarezi*): 13% (12–14%)	0.36%
X. variatus–X. xiphidium	0.11	0.31 (0.014, $P<1.5e-86$)	([*X. xiphidium, X. variatus*], *X. evelynae*): 30% (29–32%)	0.37%
			((*X. xiphidium, X. variatus*), *X. evelynae*): 10% (9–11%)	0.44%
X. milleri– X. evelynae	0.11	0.09 (0.013, $P<0.00043$)	((*X. evelynae*, *X. milleri*), *X. couchianus*): 16% (15–18%)	0.43%
			([*X. milleri*, *X. couchianus*], *X. evelynae*): 14% (13–15%)	0.43%
X. couchianus–X. meyeri	0.29	0.11 (0.04, $P<0.0030$)	([*X. meyeri, X. couchianus*], *X. gordoni*): 26% (24–29%)	0.06%
			([*X. couchianus, X. gordoni*], *X. meyeri*): 17% (15–19%)	0.08%
X. xiphidium–X. andersi	0.13	0.03 (0.014, $P<0.041$)	([*X. andersi*, *X. xiphidium*], *X. meyeri*): 19% (18–21%)	0.54%
			([*X. meyeri*, *X. andersi*], *X. xiphidium*): 11% (9–12%)	0.55%
X. evelynae–X. variatus	0.14	0.18 (0.013, $P<5.3e-28$)	([*X. variatus*, *X. evelynae*], *X. couchianus*): 19% (18–20%)	0.34%
			([*X. evelynae*, *X. couchianus*], *X. variatus*): 10% (10–11%)	0.36%
X. birchmanni– X. pygmaeus	<0.10	−0.03 (0.017, $P=0.19$)	([*X. pygmaeus*, *X. malinche*], *X. birchmanni*): 3% (3–4%)	0.57%
			([*X. birchmanni*, *X. pygmaeus*)), *X. malinche*): 3% (3–3%)	0.61%

resolved phylogeny with 100% bootstrap support for all nodes (Fig. 2A).

To confirm that our total evidence topology was not dependent on the reference sequence used for assembly, we repeated the same analysis using the *X. maculatus* genome and the *X. mayae* transcriptome as reference sequences (Supporting information section ii). The total evidence topology produced from alignment to the *X. maculatus* genome was nearly identical to the topology produced from alignment to the *X. birchmanni* transcriptome; this topology placed *X. nezahualcoyotl* as sister to *X. cortezi* rather than *X. montezumae* (Supporting Information section ii, Fig. S1A). The total evidence topology produced from alignment to the *X. mayae* transcriptome also resulted in a highly similar topology, albeit with lower bootstrap support; for details, see online Supporting Information section ii, Figure S1B.

We also investigated the likely species tree with gene tree analysis using BUCKy. With this method we infer a nearly identical species tree topology to the topology produced by the analysis

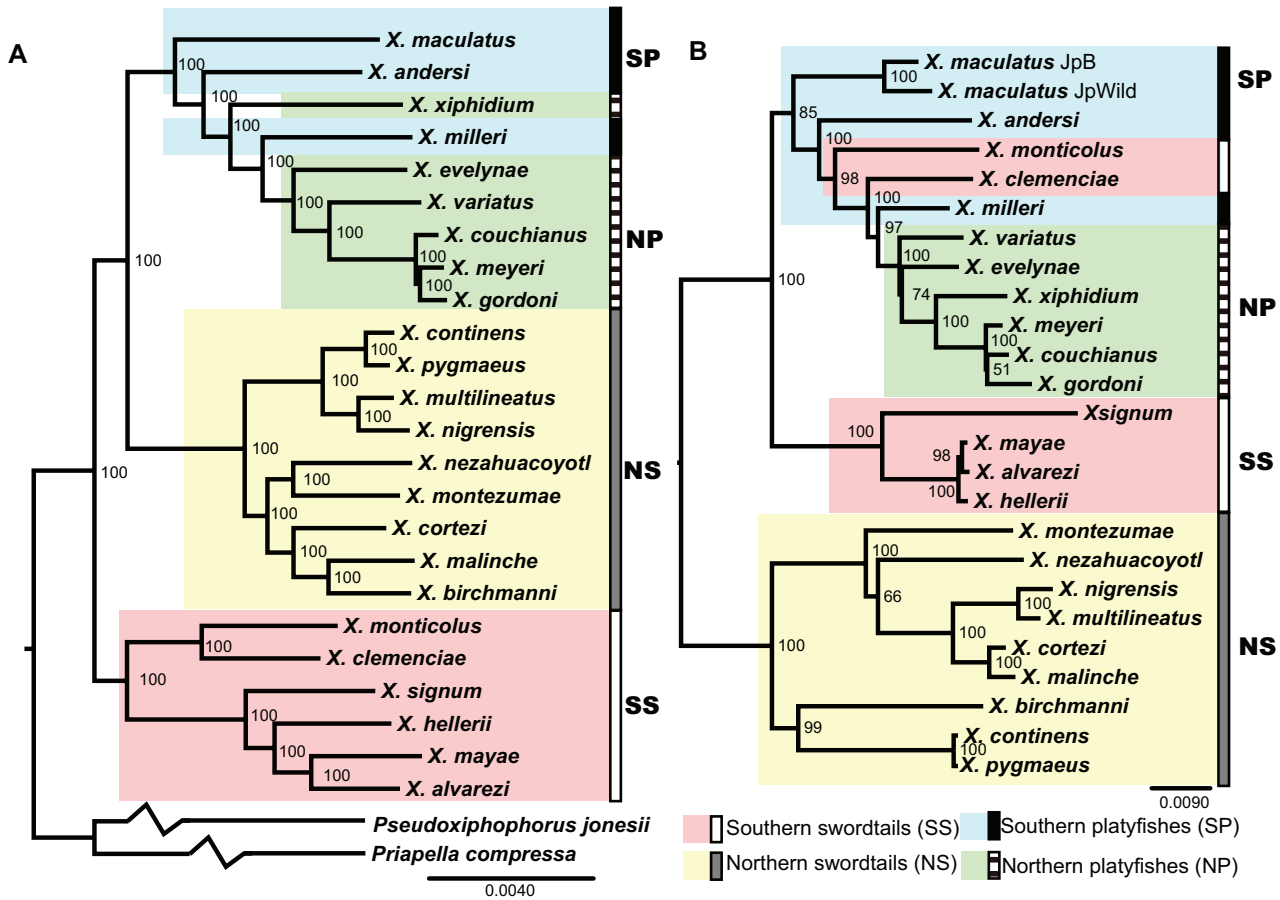

Figure 2. **(A) Total evidence nuclear phylogeny produced by concatenating 10,999 transcripts totaling 16,855,549 (22.91% missing) sites using RAxML 7.2.8 with GTR + Γ model. Nodal support generated by 100 rapid bootstraps with GTR + CAT. Log-likelihood = −33380190.53. (B) Unrooted mitochondrial tree (coding regions only, 15,787 bp, 42.63% missing) of 24 *Xiphophorus* species excluding two outgroup species. Log-likelihood = −40371.19.**

of the concatenated *X. birchmanni* aligned dataset with RAxML (Fig. 3; see Table S6 for CFs). As with the topology produced by aligning to the *X. maculatus* genome, the only difference between the species tree inferred by BUCKy and the concatenated topology was the placement of *X. nezahualcoyotl* sister to *X. cortezi* rather than *X. montezumae*. Given that the species trees produced by the different analyses are nearly identical, we focus subsequent discussion on the species tree determined by BUCKy.

Our results resolve many of the uncertainties of the *Xiphophorus* phylogeny (Figs. 2, 3). The monophyly of the three major *Xiphophorus* clades (northern swordtails, platyfish, and southern swordtails) is strongly supported. Our results also strongly support the placement of platyfish and northern swordtails as sister clades, with southern swordtails at the base of the genus (corresponding to Fig. 1B). In within clade relationships, we also find a number of major differences from previous studies. *X. xiphidium*, which is a northern sworded platyfish, is resolved at a more basal position that is more closely related to the southern platyfish *X. andersi* than in previous studies (Meyer et al. 2006). Excluding *X. xiphidium*, southern platyfishes are paraphyletic whereas northern platyfishes are monophyletic. Out of three hypothesized clades within the northern swordtails (Rauchenberger et al. 1990), two (the *cortezi* and *pygmaeus* clades) were found to be paraphyletic, whereas the *montezumae* clade is polyphyletic. *X. montezumae* was resolved as the sister to the *cortezi* clade, which here included *X. nezahualcoyotl*, whereas *X. continens* was supported as sister of *X. pygmaeus* (but see Table S1).

EXTENSIVE CYTONUCLEAR CONFLICT

Previous phylogenetic studies in many taxa have historically relied heavily on mitochondrial DNA (mtDNA) as a marker. We constructed mtDNA relationships based on concatenated expressed mitochondrial sequences using RAxML. We find low concordance between mitochondrial and nuclear relationships in *Xiphophorus* (Fig. 2). The placement of *X. maculatus* was not well resolved in the mtDNA tree. When outgroups are included,

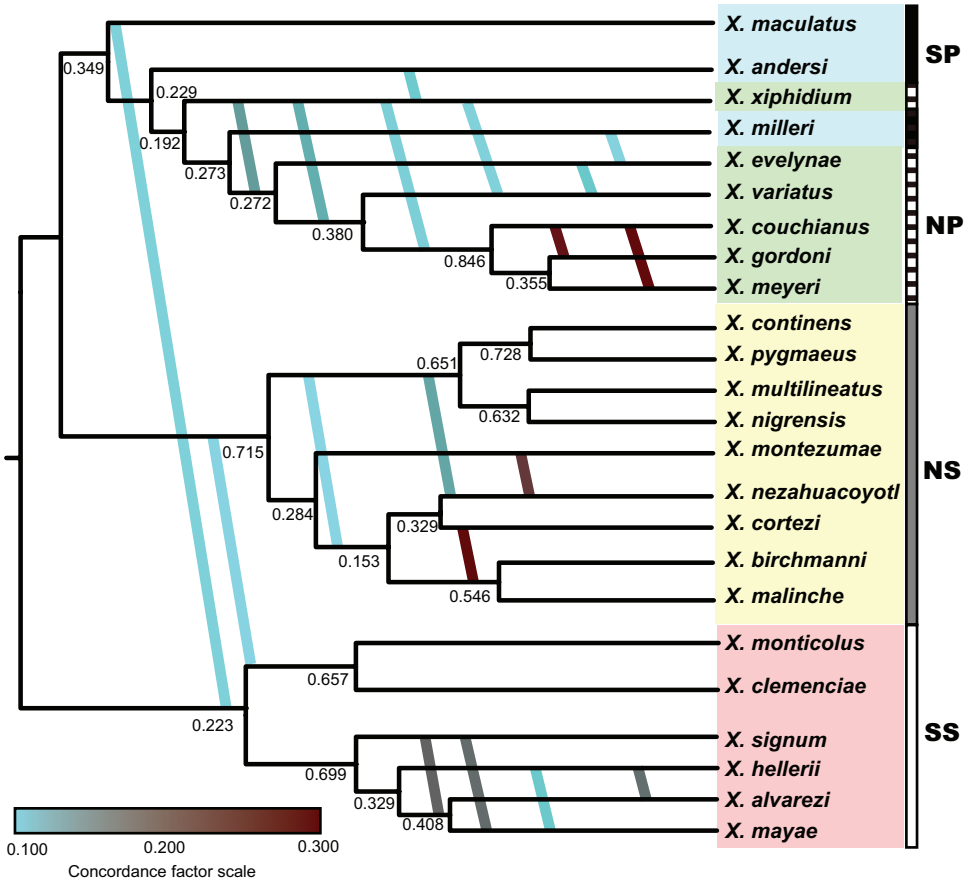

Figure 3. **A primary concordance tree produced by BUCKy ($\alpha = 1$) from 2366 gene trees inferred by MrBayes 3.2.1. Nodal values are Bayesian concordance factors (CFs). Discordance which is detected using both the *X. birchmanni* transcriptome and the *X. maculatus* genomic reference (Fig. S3) with CFs > 10% is indicated with solid lines using a color gradient. See Table S6 for CFs and confidence intervals.**

X. maculatus resides at the base of a clade including all southern swordtails and other platyfishes, with low bootstrap support (Fig. S2). Excluding both outgroups places *X. maculatus* sister to other platyfishes, but again with low bootstrap support (Fig. 2B). In the rooted trees, mtDNA sequences support monophyly of southern swordtails and platyfishes, in direct conflict with results from nuclear data. As reported previously (Meyer et al. 2006; Kang et al. 2013), *X. clemenciae* and *X. monticolus* are nested within the platyfishes, also in contradiction to the species tree inferred from nuclear genome sequences.

HYBRIDIZATION INFERRED BY BUCKy, AU TESTS, AND THE *D*-STATISTIC

Gene tree discordance can result from incomplete lineage sorting (Pollard et al. 2006), reconstruction errors, or from hybridization (Degnan and Rosenberg 2009). We used BUCKy to identify instances of major discordance between individual nuclear gene trees and the consensus species tree (Figs. 3, 4). A large number of *Xiphophorus* species had high levels of nuclear gene tree discordance based on the results of our BUCKy analysis. We found greater than 10% discordance in all three major groups: northern swordtails (*X. nezahualcoyotl, Xiphophorus montuzumae*), platyfishes (*X. xiphidium*, *X. andersi*, *Xiphophorus evelynae*, *X. variatus*, *X. couchianus, X. milleri*, *X. meyeri*, *X. gordoni*), and southern swordtails (*X. hellerii*, *X. signum*, *X. alvarezi*, and *X. mayae*). We find nearly identical patterns of discordance in both the alignments to the *X. birchmanni* transcriptome assembly and the alignments to the *X. maculatus* genome (Supporting Information sections iii and iv; Fig. S3).

For all species that showed high levels of discordance (i.e., $\geq$10%; Fig. 3), we performed two additional analyses to differentiate between incomplete lineage sorting and hybridization. To evaluate the evidence for hybridization, we applied the AU test (Shimodaira 2002) and Patterson's *D*-statistic (Green et al. 2010) to these species (Table S2). Both tests allow us to determine whether incomplete lineage sorting can be rejected as the cause of discordance by testing for asymmetry of the two less frequent topologies. Given agreement between all three analyses (BUCKy, the AU test, and Patterson's *D*-statistic), we conclude that hybridization is a likely contributor to gene tree discordance among

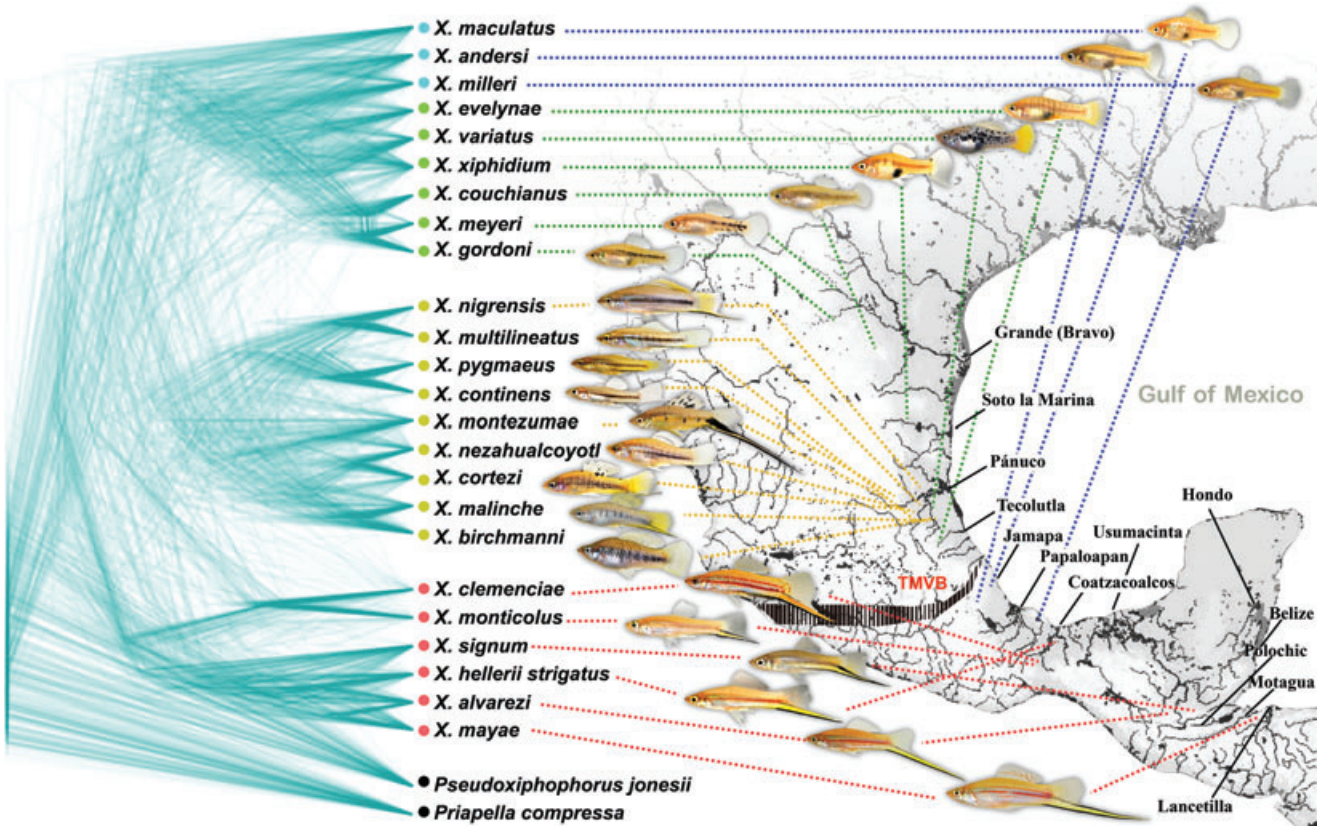

Figure 4. **Localities and photographs of the specimens used in this study mapped on a DensiTree produced by overlaying 160 gene trees inferred by MrBayes 3.2.1. Black vertical lines indicate the location of a major geographical barrier, the Trans-Mexican Volcanic Belt (TMVB). Blue lines, southern platyfishes; green lines, northern platyfishes; yellow lines, northern swordtails; pink lines, southern swordtails.**

the species in question. Specifically, we find that 7 of 8 tested cases of discordance are likely the result of hybridization (Table 1). Some species had levels of hybridization consistent with extensive admixture or hybrid speciation. For example, the majority of the *X. xiphidium* alignments (70%) cluster with *X. meyeri* in AU tests, but a significant proportion (19%) are grouped with the southern platyfish *X. andersi*. In *X. nezahualcoyotl*, 66% of alignments were most closely related to *X. cortezi* but 32% are more closely related to *X. montezumae*. This hybridization is unlikely to be recent however, because divergence between species was still high in regions supporting minor topologies (Table 1, compared to Table S7).

CHARACTER MAPPING AND INDEPENDENT CONTRASTS

We used the total evidence tree to investigate evolutionary trends for sword length, sword preference, and whether female preference for the sword is correlated with sword length (Supporting Information section v). Independent contrasts reveal a marginally nonsignificant positive correlation between male sword length and female preference ($r = 0.55$, degree of freedom [df] $= 9$, $P = 0.083$). Female preference significantly decreased with node height ($r = -0.71$, df $= 8$, $P = 0.022$), suggesting an evolutionary trend for decreased sword preference. Sword length, however, was not correlated with node height in the species tree ($r = -0.27$, df $= 23$, $P = 0.185$).

Discussion

Using sequence information generated from one paired-end Illumina lane, we resolve phylogenetic relationships in a group of species with a highly contested evolutionary history. We show that there are high levels of hybridization between many species, potentially due to weak postzygotic isolation in *Xiphophorus* (Schartl 2008; Rosenthal and Ryan 2011). Extensive admixture has occurred in lineages within the platyfish, northern, and southern swordtail clades, whereas cases of introgression have occurred within all major clades. A high-confidence molecular phylogeny of *Xiphophorus* allows us to address questions about the role of sexual selection and hybridization in trait evolution.

PHYLOGENETIC RELATIONSHIPS IN *XIPHOPHORUS*

We find strong support for monophyly of the three previously identified *Xiphophorus* clades: the southern swordtails, the northern swordtails, and the platyfishes. The rooting of our phylogeny (most closely corresponding to Fig. 1B) confirms results from previous nuclear datasets that northern swordtails and southern swordtails are not monophyletic (e.g., Meyer et al. 2006). This rooting conflicts with early studies using morphology (Rosen 1979) and mitochondrial DNA (Meyer et al. 1994). We also find major differences in intraclade relationships compared with previous studies, particularly in the northern swordtails (Rauchenberger et al. 1990; Morris et al. 2001; Kang et al. 2013).

RETICULATE EVOLUTION IN *XIPHOPHORUS*

Based on our analysis of nuclear sequences, we find evidence of gene flow between multiple species within platyfishes, northern swordtails, and southern swordtails (Table 1; Figs. 3, 4). Only eight of the species surveyed show no evidence for discordance of nuclear gene trees. Our results demonstrate that hybridization has been extensive within clades, even between nonsister species, but there is little evidence of hybridization between the major clades. The extent of historical hybridization in *Xiphophorus* is especially interesting given that contemporary natural hybrids are relatively rare. Premating isolation via sexual selection is the major barrier to hybridization in this genus (Schartl 2008; Rosenthal and Ryan 2011) but reproductive isolation via mating preferences may be particularly vulnerable to perturbation due to demographic fluctuations (Willis et al. 2011) or environmental disruption of sexual communication (Fisher et al. 2006). The vulnerability of female preference as the major barrier to reproductive isolation is apparent in our results, which indicate rampant genetic exchange throughout the *Xiphophorus* phylogeny. Hybridization is likely to be even more widespread than we report, because our method will only detect gene flow when it is extensive enough to cause discordance $> 10\%$. For example, historical hybridization between the widely sympatric *X. maculatus* and *X. hellerii* has been recently reported (Schumer et al. 2012) but the level of shared ancestry ($\sim$6%) is lower than the detection threshold used here.

Mitochondrial DNA relationships have historically been used for phylogenetic studies and as a method of detecting past hybridization, but accumulating evidence casts doubt on the utility of mitochondrial genes in studying evolutionary relationships (Ballard and Rand 2005). Our mtDNA results differ strikingly from relationships suggested by nuclear sequences. In the mtDNA phylogeny, the southern swordtails are grouped as sister to the platyfishes, whereas nuclear sequences group platyfishes and northern swordtails (Fig. 2). Past research in *Xiphophorus* has inferred potential cases of hybridization based on mitonuclear incongruence (Meyer et al. 1994, 2006; Kang et al. 2013). However, the mtDNA topology does not support the same hybridization patterns detected in our nuclear gene tree analysis. Instead, we find mtDNA incongruence can occur even when the nuclear genome shows no evidence of hybridization. For example, we did not find evidence of gene flow between platyfish and *X. monticolus* as recently suggested (Kang et al. 2013). Similarly, our previous study (Schumer et al. 2012) using whole genome sequencing of *X. clemenciae* did not support extensive gene flow with *X. maculatus*. In northern swordtails, even though the sister relationship between the *X. pygmaeus* clade and *X. birchmanni* is highly supported by mtDNA, there is no evidence for gene flow between these species based on nuclear gene trees. Other factors such as incomplete lineage sorting or selection could be an alternate explanation of the observed mtDNA patterns (e.g., Schumer et al. 2012).

EVIDENCE FOR EXTENSIVE ADMIXTURE IN ALL MAJOR GROUPS

We find evidence based on genome-wide data that *X. nezahualcoyotl*, *X. xiphidium*, and multiple species in the *X. hellerii* clade have significant proportions of their genomes derived from hybridization. These species have levels of discordance ranging from 19% to 42% based on AU tests (Table 1), suggestive of extensive historical admixture or even hybrid speciation. Given current species distributions, hybridization between *X. xiphidium–X. andersi* and *X. nezahualcoyotl–X. montezumae* has not been previously suspected, suggesting that historical species distributions or migration allowed hybridization between these species.

Although all three of these groups show high levels of hybridization, extensive admixture in *X. nezahualcoyotl* is also consistent with morphological data and previous research. Based on such high levels of nuclear gene tree discordance ($\sim$30%), it is possible that *X. nezahualcoyotl* is the product of admixture between its sister species *X. cortezi* and *X. montezumae.* The sword of *X. nezahualcoyotl* is of intermediate length of *X. montezumae* and *X. cortezi* (sword index: 0.48, *X. montezumae* 1.0, *X. cortezi* 0.37; see Table S4 and Supporting Information section v), and *X. nezahualcoyotl* shares morphological characteristics with both species (pigment patterns shared with *X. cortezi* and lateral stripes shared with *X. montezumae*). Previous phylogenetic analyses could not resolve relationships between *X. nezahualcoyotl, X. montezumae*, and *X. cortezi* (Rauchenberger et al. 1990; Morris et al. 2001). We suggest that these difficulties stem from the extensive hybridization in the evolutionary history of these species. A more complete survey throughout the range of *X. nezahualcoyotl*, and whole genome sequencing of both species and *X. cortezi* is required to further evaluate this prediction.

IMPLICATIONS FOR THE EVOLUTION OF SEXUALLY SELECTED TRAITS

Early research proposed that platyfishes were basal to the genus and the evolution of the sword was driven by latent female

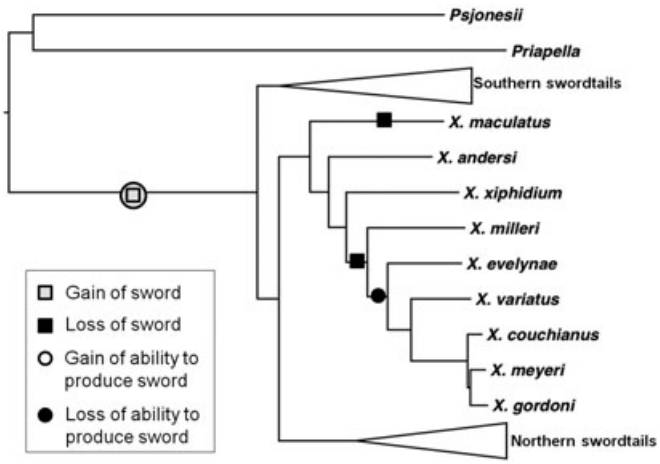

Figure 5. Sword traits (presence of a sword and ability to produce a sword following androgen induction) mapped onto the total evidence phylogeny with southern and northern swordtails collapsed. This topology shows a single loss of the ability to produce a sword in the northern platyfishes and multiple losses of the natural expression of sword.

preferences (Basolo, 1995a,b). Based on the rooting of the genus in our phylogeny, we find that the ability to produce a sword was likely present in the common ancestor of *Xiphophorus* and secondarily lost in some platyfishes (as observed in Meyer et al. 1994; Fig. 5; Meyer et al. 2006). However, preferences for swords in outgroup taxa suggest that the sword could have evolved in response to a preexisting bias prior to the diversification of *Xiphophorus* (as suggested by Basolo 1995b).

Our phylogeny changes the placement of *X. xiphidium*, a sworded platyfish, in comparison to previous studies. Because *X. maculatus* and *X. milleri* both have the capacity to produce a short sword if induced by androgens (Gordon et al. 1943; Dzwillo 1963, 1964; Zander and Dzwillo 1969; Offen 2008), and *X. andersi* and *X. xiphidium* are both sworded, our phylogeny demonstrates that the ability to produce a short sword was present in the common ancestor of platyfishes, and thus a synapomorphy of the genus. Based on our results, the genetic pathway for sword production was not completely lost in platyfishes until after the diversification of *X. milleri* (Fig. 5). This finding has important implications for research into the genetic basis of the sword; because this trait arose only once it is likely that the production of the sword is regulated by the same mechanisms in all *Xiphophorus* species. Despite the likely presence of the sword in the common ancestor of *Xiphophorus*, sword preference is negatively correlated with node height, suggesting an evolutionary trend for a reduction in sword preference in *Xiphophorus*.

GENE FLOW AND SEXUALLY SELECTED TRAITS

Only two platyfish species have a short sword, *X. xiphidium* and *X. andersi*. The platyfish sword is significantly reduced compared to sword ornaments in northern and southern swordtails, and lacks pigmentation. An alternate explanation for the presence of short swords in *X. andersi* and *X. xiphidium* is a single loss of the trait in platyfishes followed by introgression. We consider this unlikely because we do not find evidence of swordtail ancestry in *X. xiphidium* or *X. andersi*, although we caution that lower levels of hybridization (less than our 10% detection threshold) could have occurred between swordtails and these two platyfishes. However, we find evidence of hybridization between *X. andersi* and *X. xiphidium*, raising the possibility that loci underlying the short sword could be shared in these two species. Identifying the genetic basis of the sword is a crucial next step in determining whether hybridization has played a role in the phylogenetic distribution of this trait (e.g., Heliconius Genome Consortium 2012).

Conclusions

Previous investigations of the relationships among *Xiphophorus* species have been fraught with inconsistencies, likely as a result of gene flow among taxa. Our approach highlights the utility of genome-scale data in resolving patterns of gene flow in addition to determining species trees with high confidence. Our results demonstrate that evolution of the sword predated the diversification of *Xiphophorus* and that the ability to produce the sword was not completely lost until after the diversification of northern platyfishes. In addition, we find that hybridization in *Xiphophorus* has been historically common, and we document evidence of extensive admixture, potentially indicative of hybrid speciation, in the northern swordtail *X. nezahualcoyotl*, southern swordtails in the *X. hellerii* clade, and the platyfish *X. xiphidium*. Our results suggest that species boundaries primarily maintained by behavioral premating isolation may be particularly porous and that gene flow can be widespread without resulting in species collapse. Hybridization is likely to play a major role in the evolutionary history of many species without strong postmating isolation.

ACKNOWLEDGMENTS

The authors thank the *Xiphophorus* Genetic Stock Center for providing many of the *Xiphophorus* samples. M. Tobler, V. Smith, and M. Squire collected outgroup species from the wild. We are grateful to S. Monks, the Mexican federal government, and the government of the state of Hidalgo, Mexico, for permission to collect samples. This study was supported by a Rosemary Grant Award (Society for the Study of Evolution) and National Science Foundation (NSF) DDIG (0246456100001) to RC, a Graduate Student Research Award (Society for Systematic Biology) and NSF GRFP (DGE0646086) to MS, and was partially supported by National Institutes of Health, NCRR, Division of Comparative Medicine grant R24OD011120 (RW) including an ARRA supplement to this award. Computation was performed on the Brazos HPC cluster at Texas A&M University and the Cetus Cluster at Princeton University. The authors also thank B. Boussau, M. Mateos, S. Cordoba, our associate editor, and two

anonymous reviewers for critically reading the manuscript. The authors declare no conflict of interest.

LITERATURE CITED

Abbott, R. J., M. J. Hegarty, S. J. Hiscock, and A. C. Brennan. 2010. Homoploid hybrid speciation in action. Taxon 59:1375–1386.

Aboim, M. A., J. Mavárez, L. Bernatchez, and M. M. Coelho. 2010. Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. J. Evol. Biol. 23:817–828.

Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412–426.

Arnold, M. L., and N. H. Martin. 2010. Hybrid fitness across time and habitats. Trends Ecol. Evol. 25:530–536.

Ballard, J. W. O., and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Ann. Rev. Ecol. Evol. Syst. 36:621–642.

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

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

———. 1995b. Phylogenetic evidence for the role of a pre-existing bias in sexual selection. Proc. R. Soc. Lond. B 259:307–311.

———. 2006. Genetic linkage and color polymorphism in the southern platyfish (*Xiphophorus maculatus*): a model system for studies of color pattern evolution. Zebrafish 3:65–83.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Chung, Y., and C. Ané. 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60:261–275.

Cranston, K. A., B. Hurwitz, D. Ware, L. Stein, and R. A. Wing. 2009. Species trees from highly incongruent gene trees in rice. Syst. Biol. 58:489–500.

Darwin, C. 1859. On the origin of species. John Murray, London.

Decker, J. E., J. C. Pires, G. C. Conant, S. D. McKay, M. P. Heaton, K. Chen, A. Cooper, J. Vilkki, C. M. Seabury, A. R. Caetano, et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. 106:18644–18649.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

dos Reis, M., J. Inoue, M. Hasegawa, R. J. Asher, P. C. J. Donoghue, and Z. Yang. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc. R. Soc. B: Biol. Sci. 279:3491–3500.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

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

———. 1964. Sekundäre Geschlechtsmerkmale an der Caudalflosse einiger Xiphophorini unter dem Einfiuß von Methyltestosteron. Mitt. Ham. Zool. Mus. 61:15–22.

Felsenstein, J. 1985. Phylogenies and the comparative method. University of Chicago Press, Chicago, IL.

Fisher, H. S., B. B. M. Wong, and G. G. Rosenthal. 2006. Alteration of the chemical environment disrupts communication in a freshwater fish. Proc. R. Soc. Lond. B 273:1187–1193.

Fisher, H. S., S. J. Mascuch, and G. G. Rosenthal. 2009. Multivariate male traits misalign with multivariate female preferences in the swordtail fish, *Xiphophorus birchmanni*. Anim. Behav. 78:265–269.

Gordon, M., H. Cohen, and R. F. Nigrelli. 1943. A hormone-produced taxonomic character in *Platypoecilus maculatus* diagnostic of wild *P. xiphidium*. Am. Nat. 77:569–572.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, and M. H. Y. Fritz. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Haas, V. 1992. *Xiphophorus* phylogeny, reviewed on the basis of courtship behavior. Pp. 279–288 *in* J. H. Schroder, J. Bauer, and M. Schartl, eds. Trends in ichthyology, and international perspective. Blackwell Scientific, Oxford, U.K.

Hailer, F., V. E. Kutschera, B. M. Hallström, D. Klassert, S. R. Fain, J. A. Leonard, U. Arnason, and A. Janke. 2012. Nuclear genomic sequences reveal that Polar Bears are an old and distinct bear lineage. Science 336:344–347.

Hankison, S. J., and M. R. Morris. 2003. Avoiding a compromise between sexual selection and species recognition: female swordtail fish assess multiple species-specific cues. Behav. Ecol. 14:282–287.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

Hittinger, C. T., M. Johnston, J. T. Tossberg, and A. Rokas. 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. Proc. Natl. Acad. Sci. 107:1476–1481.

Kallman, K. D., and S. Kazianis. 2006. The genus *Xiphophorus* in Mexico and Central America. Zebrafish 3:271–285.

Kang, J. H., M. Schartl, R. Walter, and A. Meyer. 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. BMC Evol. Biol. 13:25.

Kocot, K. M., J. T. Cannon, C. Todt, M. R. Citarella, A. B. Kohn, A. Meyer, S. R. Santos, C. Schander, L. L. Moroz, B. Lieb, et al. 2011. Phylogenomics reveals deep molluscan relationships. Nature 477:452–456.

Kubatko, L. S. 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58:478–488.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910–2911.

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.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542–2543.

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

Maddison, W., and D. Maddison. 2011. MESQUITE: a modular system for evolutionary analysis.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–237.

———. 2007. Hybrid speciation. Nature 446:279–283.

———. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 363:2971–2986.

Marcus, J. M., and A. R. McCune. 1999. Ontogeny and phylogeny in the northern swordtail clade of *Xiphophorus*. Syst. Biol. 48:491–522.

Mateos, M., O. I. Sanjur, and R. C. Vrijenhoek. 2002. Historical biogeography of the livebearing fish genus *Poeciliopsis* (Poeciliidae: Cyprinodontiformes). Evolution 56:972–984.

McLennan, D. A., and M. J. Ryan. 1999. Interspecific recognition and discrimination based upon olfactory cues in northern swordtails. Evolution 53:880–888.

Meierjohann, S., and M. Schartl. 2006. From Mendelian to molecular genetics: the *Xiphophorus* melanoma model. Trends Genet. 22:654–661.

Meng, C., and L. S. Kubatko. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor. Popul. Biol. 75:35–45.

Meyer, A., J. M. Morrissey, and M. Schartl. 1994. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. Nature 368:539–542.

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.

Midford, P. E., T. G. Jr., and W. P. Maddison. 2011. PDAP: PDTREE package for Mesquite. Version 1.16. Available at http://mesquiteproject.org/pdap_mesquite/.

Morizot, D. C., S. A. Slaugenhaupt, K. D. Kallman, and A. Chakravarti. 1991. Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei: Poeciliidae). Genetics 127:399–410.

Morris, M. R., K. de Queiroz, D. C. Morizot, and J. D. McEachran. 2001. Phylogenetic relationships among populations of northern swordtails (*Xiphophorus*) as inferred from allozyme data. Copeia 2001:65–81.

Nabholz, B., A. Künstner, R. Wang, E. D. Jarvis, and H. Ellegren. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. Mol. Biol. Evol. 28:2197–2210.

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.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Pollard, D. A., V. N. Iyer, A. M. Moses, and M. B. Eisen. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genetics 2:e173.

R Development Core Team. 2010. R version 2.15.2. R Project for Statistical Computing, Vienna, Austria.

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. 2975:1–41.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216.

Rosen, D. E. 1960. Middle-American poeciliid fishes of the genus *Xiphophorus*. Bull. Fla. Mus. Nat. Hist. 5:58–242.

———. 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. 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., and M. J. Ryan. 2011. Conflicting preferences within females: sexual selection versus species recognition. Biol. Lett. 7:525–527.

Rosenthal, G. G., X. F. de la Rosa Reyna, S. Kazianis, M. J. Stephens, D. C. Morizot, M. J. Ryan, F. J. García de León, and R. E. Wood. 2003. Dissolution of sexual signal complexes in a hybrid zone between the swordtails *Xiphophorus birchmanni* and *Xiphophorus malinche* (Poeciliidae). Copeia 2003:299–307.

Ryan, M. J. 1990. Sexual selection, sensory systems and sensory exploitation. Oxf. Surv. Evol. Biol. 7:157–195.

Schartl, M. 1995. Platyfish and swordtails: a genetic system for the analysis of molecular mechanisms in tumor formation. Trends Genet. 11:185–189.

———. 2008. Evolution of Xmrk: an oncogene, but also a speciation gene? BioEssays 30:822–832.

Schmidt, H. A. 2009. Testing tree topologies. Pp. 381–395 *in* P. Lemey, M. Salemi, and A.-M. Vandamme, eds. The phylogenetic handbook—a practical approach to phylogenetic analysis and hypothesis testing. Cambridge Univ. Press, Cambridge, U.K.

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

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: *In press.* doi: 10.1111/evo.12009.

Schwarzer, J., B. Misof, and U. K. Schliewen. 2012. Speciation within genomic networks: a case study based on *Steatocranus* cichlids of the lower Congo rapids. J. Evol. Biol. 25:138–148.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.

Smith, S. A., N. G. Wilson, F. E. Goetz, C. Feehery, S. C. S. Andrade, G. W. Rouse, G. Giribet, and C. W. Dunn. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480:364–367.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.

Verzijden, M. N., and G. G. Rosenthal. 2011. Effects of sensory modality on learned mate preferences in female swordtails. Anim. Behav. 82:557–562.

Walter, R. B., J. D. Rains, J. E. Russell, T. M. Guerra, C. Daniels, D. A. Johnston, J. Kumar, A. Wheeler, K. Kelnar, V. A. Khanolkar, et al. 2004. A microsatellite genetic linkage map for *Xiphophorus*. Genetics 168:363–372.

Willis, P. M., M. J. Ryan, and G. G. Rosenthal. 2011. Encounter rates with conspecific males influence female mate choice in a naturally hybridizing fish. Behav. Ecol. 22:1234–1240.

Willis, P. M., G. G. Rosenthal, and M. J. Ryan. 2012. An indirect cue of predation risk counteracts female preference for conspecifics in a naturally hybridizing fish *Xiphophorus birchmanni*. PloS one 7:e34802.

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

Zerbino, D., and E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome. Res. 18:821–829.

Associate Editor: T. Streelman

Supporting Information

Additional Supporting information may be found in the online version of this article at the publisher's website:

Table S1. Sources of species and strains used in this study with pedigree information where available.

Table S2. List of species included in AU tests and *D*-statistic.

Table S3. Total number of reads and number of reads mapped for each species used in our analysis.

Table S4. Sword index (sword length/standard length) compiled from literature.

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

Table S6. Bayesian concordance factors and 95% CI for bipartitions in the species tree (marked with #) and the alternative bipartitions (marked with *).

Table S7. Pairwise distance (GTR + Γ) between partial sample pairs ranked by distance.

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.

Figure S2. Rooted mitochondrial phylogeny of *Xiphophorus* using 20× coverage cutoff (coding regions only, 15,787 bp, 37.8% missing).

Figure S3. Comparisons between BUCKy results for alignments to the *X. birchmanni* transciptome and BUCKy results for alignments to the *X. maculatus* genome.