Background: Understanding the dynamics of speciation with incomplete reproductive isolation is one of the central challenges of modern evolutionary biology. Increasingly, researchers are recognizing that many taxa have continued gene flow during and after speciation (*1*), and that infrequent hybridization can result in substantial introgression (*2*). Some studies have demonstrated that introgression will homogenize regions of the genome not essential to local adaptation (*3*). High levels of introgression can mask phylogenetic relationships between species, obscuring patterns of speciation and diversification. For example, in the freshwater fish *Xiphophorus*, each phylogeny published to date has proposed different evolutionary relationships between species. *Xiphophorus* is a morphologically and behaviorally diverse clade of twenty-seven species that has been a model system for sexual selection and speciation (*4, 5*). Due to low genetic divergence between species most *Xiphophorus* species are interfertile, even with distantly related species (*6*). Nuclear and mitochondrial phylogenies of this genus have been plagued with inconsistencies, suggesting introgression or incomplete lineage sorting (*4, 5, 7-9*). Ongoing hybridization and introgression has been demonstrated in a number of species (*9*) and at least one species is thought to have originated from an ancient hybridization event (*5*). In the past, phylogenies have used one or a handful of genes to reconstruct relationships between species. Though this is effective for many taxa, disagreement between gene trees is not uncommon (*10*). Advances in technology allow sequencing of thousands of transcripts in lineages of interest, which can be used to assemble a large number of gene trees, from which a consensus phylogeny can be derived (*11*). Though ongoing gene flow and low genetic divergence have made *Xiphophorus* difficult to study in the past, new sequencing techniques make it an ideal system in which to investigate speciation with continued gene flow and the role of sexual selection in driving diversification.

Objectives: The main objective of this study is to use next-generation sequencing to construct a phylogeny for *Xiphophorus* which will allow me to distinguish species relationships even in the face of introgression. Using next-generation sequencing techniques to generate sequence information for all 24 available *Xiphophorus* species will allow me to 1) determine evolutionary relationships between *Xiphophorus* species 2) determine how many species have experienced significant introgression from related species 3) determine whether sister species are more likely to occur in sympatry than allopatry or have divergent matings signals and 4) determine whether female preference drove the evolution of sworded males multiple times in this genus.

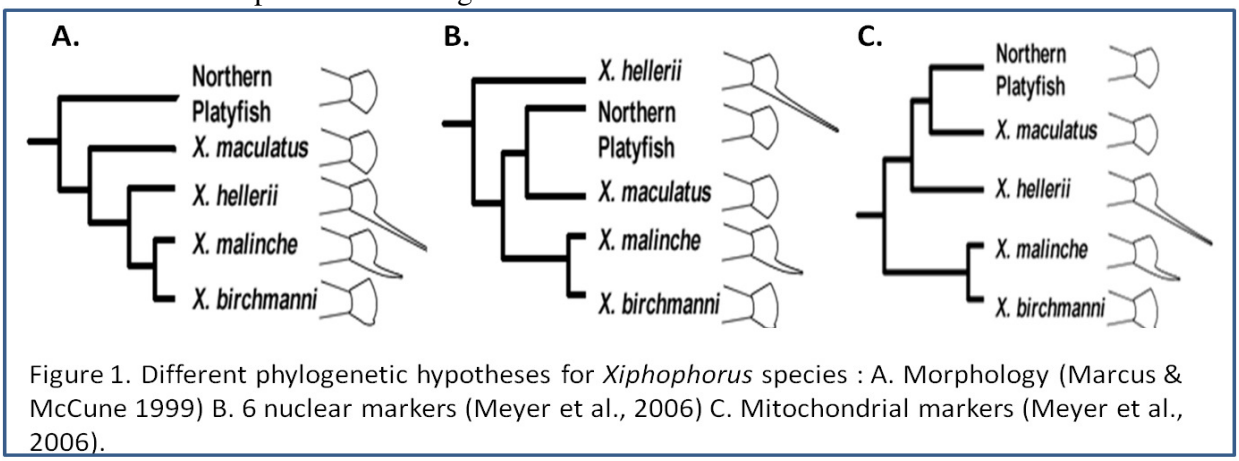

Figure 1. Different phylogenetic hypotheses for *Xiphophorus* species : A. Morphology (Marcus & McCune 1999) B. 6 nuclear markers (Meyer et al., 2006) C. Mitochondrial markers (Meyer et al., 2006).

Methods: I will use an RNAseq-based method to collect sequence information of the most highly expressed transcripts in the brains of 24 *Xiphophorus* species and two outgroups. This tissue was chosen because preliminary data demonstrated that it yielded the most *Xiphophorus*-specific sequence (fewest sequences from parasites or bacteria) and most repeatable transcript pool. Total RNA will be extracted using a standard Trizol protocol, and prepared for sequencing with the Illumina Tru-Seq kit according to manufacturer's instructions. The use of in-house flow cell indeces will allow me to multiplex 24 species in one paired-end lane, resulting in approximately 20 million reads per species. Two additional species will be pooled and sequenced in a paired-end control lane at a depth of 40 million reads per species. All sequencing will be performed at the Lewis-Sigler Center for Integrative Genomics (Princeton, NJ).

Raw data will be parsed by index and trimmed to exclude low quality (QV < 20) and short reads (consecutive bases < 30). The two high coverage species will be used to develop high-quality contigs which will be used as a scaffold to align reads from other species. I will use velvet with a kmer length of 31 for de novo assembly of reads into transcripts. Oases will be used to assemble contigs into transcript isoforms, and a custom perl script will be used to select the longest assembly of each transcript. Raw reads for all 26 species will be aligned to the reference scaffolds using bwa and the samtools will be used to identify sites that differ between the query and reference and depth of coverage at each site. Sites with coverage of 30 or greater will be used. A reference genome is also available (Figure 2, Schartl & Walter, unpublished data) if de novo assembly produces highly variable results in different individuals. Loci with suspected paralogs (receiving divergent intraspecific reads) will be excluded. To detect phylogenetic breakpoints, a moving window of 10 kb will be used to extract alignments along the *Xiphophorus* genome for phylogenetic analyses. For each 10 kb partition, RAxML will be used to obtain the ML tree using the optimal model determined by jModelTest. Approximately unbiased test (AU test, *12*) implemented in Consel will be used to detect conflicting phylogenetic signals between neighboring partitions. To distinguish hybridization events from incomplete lineage sorting I will use a maximum likelihood and Bayesian approach as described previously (*13, 14*).

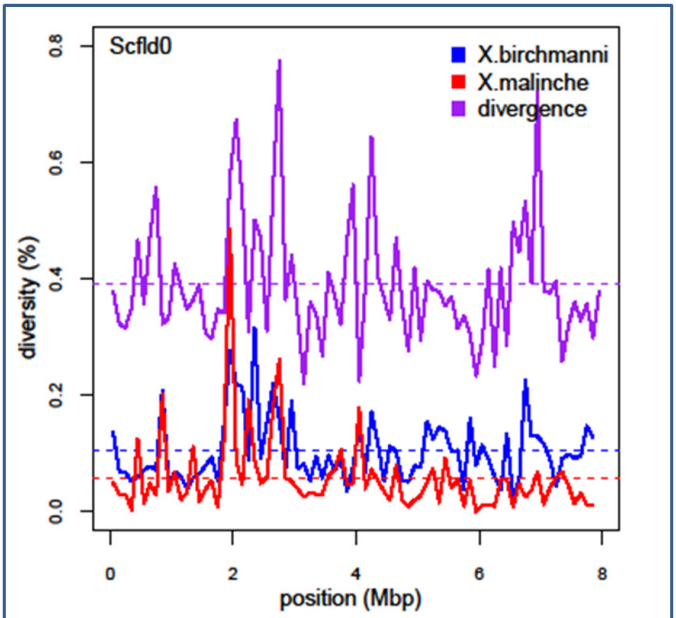

Figure 2. Polymorphism versus divergence in two *Xiphophorus* species: *X. malinche* and *X. birchmanni*

Preliminary results: In preparation for this project I have collected specimens of 24 species of *Xiphophorus* (every known species with the exception of *X. mixei* and *X. kallamani*) and two outgroups (*Gambusia* and *Heterandia* species). In addition I have extracted RNA and prepared libraries for three *Xiphophorus* species. Using data generated from *X. nigrensis* (~40 million single end reads) I assembled transcripts and calculated coverage using the pipeline described above. Through this, 46,451 transcripts were identified with an average length of 580 basepairs, similar to the number published for the transcriptome of another *Xiphophorus* species (*15*). Of these, 5,558 had an average coverage of 30 or greater. This demonstrates that RNA derived from brain tissue yields sequence information of sufficient depth to analyze a large number of sequences, even given variability in expression and coverage. Two other species known to be closely related are currently being analyzed confirm that there is informative divergence between closely related species in highly expressed transcripts. Polymorphism is low relative to divergence in species analyzed to date (Figure 2).

Significance: This project will significantly contribute to our understanding of speciation with ongoing gene flow, resolve evolutionary relationships in an evolutionary and behavioral model system, and address whether sexual selection can drive diversification and speciation.

Understanding the maintenance of reproductive isolation in the presence of gene flow is of major interest to evolutionary biologists, especially given the newly recognized ubiquity of hybridization in animals. This study will allow me to estimate the extent of introgression at sites throughout the genome in the *Xiphophorus* species group. Some researchers have proposed an "islands of divergence" model of speciation with gene flow, as opposed to a limited introgression model. Since hybridization and introgression are known to occur in *Xiphophorus* this dataset will allow me to compare multiple hypotheses of gene flow in the face of divergence.

Each molecular phylogeny constructed for *Xiphophorus* to date has suggested different evolutionary relationships between species *(5-9)*. Morphological phylogenies are plagued with potential

convergent evolution of male traits driven by shared ancestral phenotypic preferences (*4*), while broad interferility and resulting hybridization creates conflicting molecular phylogenies. A large multi-locus dataset will allow me to reconstruct the evolutionary history of this genus with high confidence (*16*).

Not only will construction of this phylogeny inform the evolutionary history of this unresolved lineage, it will also answer questions about trait evolution in *Xiphophorus*. Female preference for sworded males is basal to the genus and predates the evolution of this trait. Some molecular phylogenies (*4, 7*) suggest that preference for sworded males drove the evolution of the sword multiple times; a high confidence phylogeny of the genus will resolve the question of whether latent female preferences can drive rapid morphological diversification. It will also be informative to compare whether sister species are more likely to have divergent sexual signals, or be allopatrically distributed.

In addition, this will be one of the first studies to use an RNAseq approach to reconstruct phylogenetic relationships in species groups in which hybridization is common or speciation recent. Refinement of computational and methodological techniques associated with using RNAseq data for phylogeny construction in these groups will be another contribution of this project.

Schedule: I am a first year PhD student and this phylogeny comprises one of the major components of my thesis research investigating the role of sexual selection in speciation of *Xiphophorus* fishes. Construction of this phylogeny will be crucial in informing future projects. I have already collected samples and tested the computational pipeline that will be used in this project. I will spend one month constructing high quality contigs from two species (*X. malinche* and *X. birchmanni*) currently being sequenced and verifying that a sufficient number of ancestry informative markers are produced. I will spend three months preparing samples for sequencing and four months processing and analyzing data.

Literature cited:

1. J. Mallet, Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229. (2005).
2. S. J. Goodman, N. H. Barton, G. Swanson et al. Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus Cervus) in Argyll, Scotland. *Genetics* **152**, 355 (1999).
3. T. L. Turner, M. W. Hahn, S. V. Nuzhdin, Genomic islands of speciation in Anopheles gambiae. *Plos Biology* **3**, 1572 (2005).
4. A. Meyer, J. M. Morrissey, M. Schartl, Recurrent origin of a sexually selected trait in Xiphophorus fishes inferred from a molecular phylogeny. *Nature* **368**, 539 (1994).
5. A. Meyer, W. Salzburger, M. Schartl, Hybrid origin of a swordtail species (Teleostei : Xiphophorus clemenciae) driven by sexual selection. *Molecular Ecology* **15**, 721 (2006).
6. R. B. Walter, Z. Ju, A. Martinez et al. Genomic Resources for Xiphophorus Research. *Zebrafish* **3**, 11 (2006).
7. A. L. Basolo, Female preference predates the evolution of the sword in swordtail fish. *Science* **250**, 808 (1990).
8. J. M. Marcus, A. R. McCune, Ontogeny and phylogeny in the northern swordtail clade of Xiphophorus. *Systematic Biology* **48**, 491 (1999).
9. Z. W. Culumber et al., Replicated hybrid zones of Xiphophorus swordtails along an elevational gradient. *Molecular Ecology* **20**, 342 (2011).
10. R. Nichols, Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16**, 358 (2001).
11. J. E. H. McCormack, S. M. Zellmer, A. J. Carstens et al. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, in press.
12. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**, 492 (2002).
13. C. Meng, L. S. Kubatko, Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology* **75**, 35 (2009).
14. Kubatko, L., B. C. Carstens, L. L. Knowles. STEM: Species Tree Estimation using Maximum likelihood for gene trees under coalescence, Bioinformatics, doi: 10.1093/bioinformatics/btp079, (2009).
15. Z. Zhang *et al.*, Transcriptome Analysis of Female and Male Xiphophorus maculatus Jp 163 A. *Plos One* **6**, (2011).
16. C. T. Hittinger, M. Johnston, J. T. Tossberg, et al. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 1476 (2010)

Budget and Budget Justification

Item	Amount
Cost of 23 reactions using the Illumina Tru-Seq kit	$720.00
Cost of 1 paired-end Illumina lane	$2,000.00
Total	$2,720.00
Total requested:	**$2,000.00**

The items listed above will allow me to prepare and sequence the remaining 21 *Xiphophorus* samples and two outgroups. To reduce costs, I have validated the use of half reactions of the Illumina Tru-Seq kit, and custom indeces produced by our lab also allow for higher multiplexing, enabling all 23 unsequenced samples to be sequenced in one lane. Funding for the sample preparation is covered by a Princeton EEB through Summer Research Awards available to first year graduate students.

CV

Education

Reed College, Portland, Oregon, 2005-2009. B.A. in Biology, GPA: 3.98
Princeton University, Princeton, New Jersey, 2011-present. Ph.D. student

Awards and Grants

Centennial Fellowship in the Sciences and Engineering, Princeton University, 2011-2015
National Science Foundation Graduate Research Fellowship, 2011-2014
Significant Classroom Gains, Teach for America, 2009-2010; 2010-2011
Phi Beta Kappa, 2009
SICB Best Student Poster, 2009
Reed College Opportunity Grant, 2008
Reed College Biology Undergraduate Research Program Grant, 2008-2009
James F. and Marion L. Miller Foundation Award Recipient, 2007
Goldwater Scholarship, 2007

Research Experience

Graduate Research: Hybridization, sexual selection, and speciation in *Xiphophorus* fish, 2011-present, (research adviser: Professor Peter Andolfatto).
Summer Research Assistant: Plasticity in sex-specific aggression and gene expression in *J. transcriptus*, 2010, (research advisers: Professor Suzy C.P. Renn and Professor Albyn Jones).
Undergraduate Thesis Research: Gene Expression, Hormones, and Behavior in a Sex-Role Conventional and Sex-Role Reversed Cichlid Species Pair, 2008-2009 (research adviser: Professor Suzy C.P. Renn).
Roswell Park Cancer Institute Research Assistant: The mechanisms of BMAL-dependent CLOCK phosphorylation, 2008, (research adviser: Dr. Mary Spengler).
Independent Research: rRNA gene copy-number is correlated with the environment in a perennial wildflower, *Delphinium nuttallii*, 2007-2008, (research adviser: Professor Keith Karoly)
Miller Foundation Summer Student Fellowship: Gene Duplication and Gene Divergence among the Genus *Julidochromis,* 2007, (research adviser: Professor Suzy C.P. Renn).

Publications

Schumer, M., Wood, K., and Renn, S.C.P. Behavioral plasticity induces novel gene expression patterns in an African Cichlid fish. In preparation.
Zhen, Y., Aardema, M.L., Medina, E.M., **Schumer, M.**, Andolfatto, P. Parallel molecular evolution in a herbivore community. In review at *Science.*
Schumer, M., Krishnakant, K., and Renn, S.C.P. (2011). Comparative gene expression profiles for highly similar aggressive phenotypes in male and female cichlid fishes (*Julidochromis*). *Journal of Experimental Biology* 214:3269-3278.
Spengler, M., Kuropatwinski, K., **Schumer, M.** and Antoch, M. (2009). A serine cluster mediates BMAL1-dependent CLOCK phosphorylation and degradation. *Cell Cycle* 8:24, 4138-4146.

Selected Presentations

Schumer, M., and Renn, S.C.P. (2011). Comparative gene expression profiles for highly similar aggressive phenotypes in male and female cichlid fishes. Contributed talk: Presented at Behavior 2011 (Bloomington, IN).
Schumer, M., Wood, K., and Renn, S.C.P. (2010) Genomic Basis for Sex-Biased Behavior. Presented at Evolution (Portland, OR).
Schumer, M. and Renn, S.C.P. (2009) hCGH Detects Genomic Architecture Among African Cichlid Species of the Genus *Julidochromis.* Presented at the Society for Integrative and Comparative Biology (Boston, MA).