The Genetic Architecture of Speciation in a Primate Hybrid Zone

by

Marcella Baiz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2019

Doctoral Committee:

Research Associate Professor Liliana Cortés Ortiz, Co-Chair
Professor Priscilla Tucker, Co-Chair
Assistant Professor Regina Baucom
Assistant Professor Jacob Mueller

Marcella Baiz

baizm@umich.edu

ORCID iD: 0000-0002-1629-6737

DEDICATION

For Mom, Steph, Carly, Jenna, & JJ

ACKNOWLEDGEMENTS

This dissertation would not be possible without the following people and grants. Thank you to Raquel Marchán-Rivadeneira, Sonal Singhal, Iris Holmes, Amanda Haponski, Lucy Tran, Ryan Bracewell, Beatriz Otero Jiménez, and Lisa Walsh for helpful discussion, suggestions, and comments. This research was supported by the National Science Foundation (DEB-0640519, BCS-0962807 & BCS-1517701 to LCO). I was supported in part by the University of Michigan Genetics Training Program (T32-GM07544). This work was also supported by a Grant-In-Aid of Research from the American Society of Mammalogists and a block grant from the Department of Ecology & Evolutionary Biology.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ABSTRACT

Speciation is the fundamental evolutionary process that generates biodiversity. A major goal is to identify regions of the genome that underlie this process, including regions conferring selective advantages to diverging lineages and regions associated with the maintenance of reproductive isolation between them. Advancements in sequencing technology and methods along with falling costs of genome-wide sequencing has led to great insights in speciation research, however, investigation has been biased to model systems often studied under unnatural laboratory conditions (e.g., *Drosophila*, *Mus*). To help address this gap, I use differential introgression in a natural howler monkey hybrid zone system to identify candidate genomic regions that may underlie adaptation and reproductive isolation. In hybrid zones, regions that are advantageous on the genomic or ecological background of the opposite species may be selected for and reach high frequency when they enter the population (i.e., adaptive introgression may occur). However, regions with interspecific alleles that are deleterious on the genomic or ecological background of the other species will exhibit reduced introgression due to selection against unfit hybrids that carry combinations of such incompatible alleles. Using these regions identified here, I investigate (1) the functional role of genes associated with directional and reduced introgression, (2) the role of the X chromosome in reproductive isolation, and (3) the role of selection in shaping genomic regions with reduced introgression. My results are consistent with a weak signature of functional organization shaping patterns of gene introgression, an important role of the X chromosome in reproductive isolation that may be consistent across primate systems, and the influence of multiple forms of selection on the evolution of reproductive isolation in this system.

CHAPTER I

Differential introgression in the *Alouatta* hybrid zone reveals candidate genes for adaptation and reproductive isolation

Introduction

Hybridization occurs when members of evolutionarily distinct lineages, or species, interbreed and produce offspring of mixed ancestry (i.e., hybrids). This process can result in a number of different outcomes depending on the level of reproductive isolation between the species and the fitness of hybrids. If reproductive isolation is not complete and hybrids are able to breed with members of either parental population (i.e., backcross), then introgression, or inter-species gene flow, may occur. However, different genomic regions are not expected to exhibit the same pattern of introgression because interspecific alleles become shuffled in admixed genomes over many generations of recombination allowing selection to act in a heterogeneous manner (Barton & Hewitt 1985). This process can introduce novel alleles from heterospecific populations, which can be selectively neutral, deleterious, or adaptive.

Interspecific alleles that are advantageous on the genomic or ecological background of the opposite species may be selected for and reach high frequency when they enter the population (i.e., adaptive introgression may occur). The contribution of interspecific gene flow to adaptation has long been recognized in the evolution of plants, but has recently gained attention in animal systems (Taylor & Larson 2019). There is a small number of well-documented examples of adaptive introgression that use genomic data to demonstrate selection on

heterospecific alleles (see Taylor & Larson 2019 for a recent review). Most such case studies are investigations of one or a small number of candidate genes presumed to underlie a phenotype of interest (e.g., Song et al. 2011, Huerta-Sánchez et al. 2014). Some of these studies have been able to demonstrate strong evidence for selection on the introgressed variant that is linked to environmental variation (e.g., Jones et al. 2018).

On the other hand, loci with interspecific alleles that are deleterious on the genomic or ecological background of the other species ("barrier loci") will exhibit reduced introgression due to selection against unfit hybrids that carry combinations of such incompatible alleles. A major goal in speciation research is to identify regions of the genome that underlie reproductive isolation between taxa (i.e., 'barrier loci') (Noor & Feder 2006, Butlin et al. 2012, Nosil & Schluter 2011). Barrier loci have been of particular interest to evolutionary biologists who investigate mechanisms of reproductive isolation in the lab using classical genetic approaches to assay hybrid sterility or inviability phenotypes (e.g., Dobzhansky 1936, Masly & Presgraves 2007, Good et al. 2010, Brekke et al. 2016). Some important insights from classical genetic approaches are that, 1) hybrid male sterility traits often map to the X chromosome. This is consistent with Coyne & Orr's (1989) second "rule" of speciation: The X chromosome has a disproportionately large effect on isolation. 2) Hybrid male sterility has a complex genetic basis and likely involves more than just a few genes (Masly & Presgraves 2007, Good et al. 2010, Turner & Harr 2014). 3) Interspecific regulatory divergence can disrupt both gene expression and genomic imprinting in hybrids (Good et al. 2010, Brekke et al. 2016). 4) Reproductive isolation and the number of hybrid incompatibilities increases with divergence time between species (Coyne & Orr 1997, Matute et al. 2010).

Despite the success of laboratory experiments in identifying candidate barrier loci in particular taxa, these regions may not have played a causal role in divergence in nature. For example, the histone methyltransferase *Prdm9* was shown to cause F1 hybrid sterility for inbred mouse strains crossed in the lab (Mihola et al. 2009), but in the Bavarian house mouse hybrid zone, sterility phenotypes vary in severity with complete sterility appearing to be rare or absent (Turner et al. 2012) and sterility traits map to multiple loci (Turner & Harr 2014). Thus, to better understand the genetic basis of reproductive isolation it is important to study its genomic architecture in natural populations, where barriers to gene flow are acting.

Areas where natural hybridization occurs (i.e., hybrid zones) can be used to infer the contribution of introgressive hybridization to evolutionary processes like adaptation and reproductive isolation (Harrison 1990). In such systems, population genetic data can be used to infer differential patterns of gene introgression (e.g., Larson et al. 2013, Janoušek et al. 2015). Candidate genes that underlie adaptation and reproductive isolation are expected to exhibit extreme patterns of introgression (increased and reduced introgression, respectively) and can thus be distinguished from neutral markers, which should exhibit a limited amount of introgression across the hybrid zone proportional to a cline in the degree of genomic admixture (i.e., hybrid index) (Barton 1983, Barton & Hewitt 1985).

Among study systems used to explore the genomic architecture of adaptation and reproductive isolation, natural hybrid zones offer many advantages. Many generations of backcrossing and recombination yield a variety of recombinant genotypes beyond F1 and F2 which may allow for fine-scale mapping of regions of interest (e.g., Turner & Harr 2014). Further, unlike in laboratory populations, selective pressures of the natural environment acting in hybrid zones can be directly observed. Although hybrid zone studies do not have the benefit of

directly manipulating crosses for hypothesis testing, admixture mapping approaches do not require knowledge of phenotypes associated with markers and can be thought of as taking a more objective approach to identifying phenotypes of interest using reverse genetics (Buerkle & Lexer 2008). One limitation to the study of natural hybrid zones has been the lack of genome assemblies for non-model species. Thanks to recent technological advances in sequencing technology and reduced costs associated with obtaining genome-wide sequence data for a large number of individuals, we are now overcoming this limitation.

Although limited in number, recent studies of natural hybrid zones using genomic data have yielded important insight on the genetics of hybridization and speciation. Genes with reduced introgression in the house mouse hybrid zone in Europe were found to be expressed intracellularly and involve functions like DNA binding while genes with increased introgression were found to be expressed in the cell periphery with functions in signal transduction (Janoušek et al. 2015). The authors hypothesized that this may reflect constraint on genes expressed intracellularly, which tend to be more interconnected within gene networks, whereas genes expressed at the cell periphery tend to be less interconnected. Consistent with this, genes with reduced introgression in a rabbit hybrid zone are more involved in protein-protein interactions than by chance (Rafati et al. 2018). These observations support the hypothesis for an important role of the functional organization of the genome in speciation.

Here, we use genomic cline analysis with reduced-representation sequencing data to identify genes with extreme introgression in a bimodal howler monkey hybrid zone (*Alouatta palliata* x *A. pigra*). Throughout most of their ranges, *A. palliata* and *A. pigra* are allopatric, but they form a narrow contact zone in Tabasco, Mexico (Figure 1.1). Hybridization has been confirmed in the contact zone using molecular markers and initial surveys suggest that parental

types are nearly equally abundant and no F1s have been observed (Cortés-Ortiz et al. 2007; Kelaita et al. 2013; Cortés-Ortiz et al. 2019).

The parental species diverged 3 MA (Cortés-Ortiz et al. 2003) and have many relevant differences in their morphology (Smith 1970), cytogenetics (Steinberg et al. 2008), and social systems. In parental populations, individuals live in either uni- or multi-male/multi-female groups. In allopatry, group size for *A. palliata* (6-20+ individuals, Chapman & Balcomb 1998) tends to be larger than in *A. pigra* (4-9 individuals, Chapman & Belcomb 1998, Van Belle & Estrada 2006), and *A. palliata* groups tend to have a more female-biased sex ratio (Chapman & Balcomb 1998). There are also differences between the parental species in their characteristic loud vocalizations (Bergman et al. 2016) that give the name to members of this genus.

Previously, we investigated genetic admixture and introgression in the hybrid zone using a limited set of 28 markers and found reduced introgression for X-linked markers and several autosomal microsatellite loci, and a complete lack of introgression for the Y chromosome (Cortés-Ortiz et al. 2019). This indicates that differential introgression can be used to identify specific loci that may be associated with reproductive isolation, and suggests an important role for the sex chromosomes. The lack of introgression for the Y chromosome is consistent with absence or sterility of first generation hybrid males (i.e., Haldane's Rule) (Haldane 1922), since they would not be able to backcross.

This study builds upon this work by investigating differential introgression for a large set of variants in protein coding regions with the goal of understanding the functional basis of differential introgression. Specifically, we use reduced-representation sequencing data to identify genes that exhibit extreme introgression relative to the genomic background. Further, we perform functional enrichment analyses for regions of extreme introgression to test the hypothesis of a

functional role in driving speciation. Because we have observed restricted introgression for sex chromosomal markers (Cortés-Ortiz et al. 2019), we predict that genes with reduced introgression are enriched for reproductive functions, particularly for male functions. We also predict that genes with reduced introgression will be associated with intracellular functions while genes with increased or directional introgression will be associated with functions at the cell periphery.

Materials and Methods

Sampling

We obtained blood samples from 181 wild individuals captured between 1998 and 2012 following procedures described in Kelaita et al. (2011) and Kelaita & Cortés-Ortiz (2013). Our samples included 81 individuals from the hybrid zone in Tabasco, Mexico, 35 allopatric *A. palliata* from Veracruz, Mexico, and 32 allopatric *A. pigra* (24 from Campeche, Mexico, eight from Dolores, Guatemala, and 12 from Quintana Roo, Mexico) (Figure 1.1). To avoid sampling individuals more than once, most individuals were tattooed with unique IDs. Non-tattooed individuals were either captured in localities visited only once or were sampled during the same expedition from distinct groups. During sample collection, 2ml of blood was drawn from the caudal vein of chemically immobilized individuals and mixed in 10ml of lysis buffer (Seutin et al. 1991). Samples were kept on ice in the field and stored at -20°C upon arrival in the laboratory.

Genomic DNA was extracted with the QIAGEN DNeasy tissue kit (Qiagen Inc., Valencia, CA) following the manufacturer's protocol for animal tissue extractions with the following modifications: 1) we added 100 µl of whole blood solution to 100 µl buffer ATL, 2)

we eluted DNA in 70 µl of water at 55 °C twice (re-using the same spin column) to maximize DNA yields, after incubating for 5 minutes at room temperature.

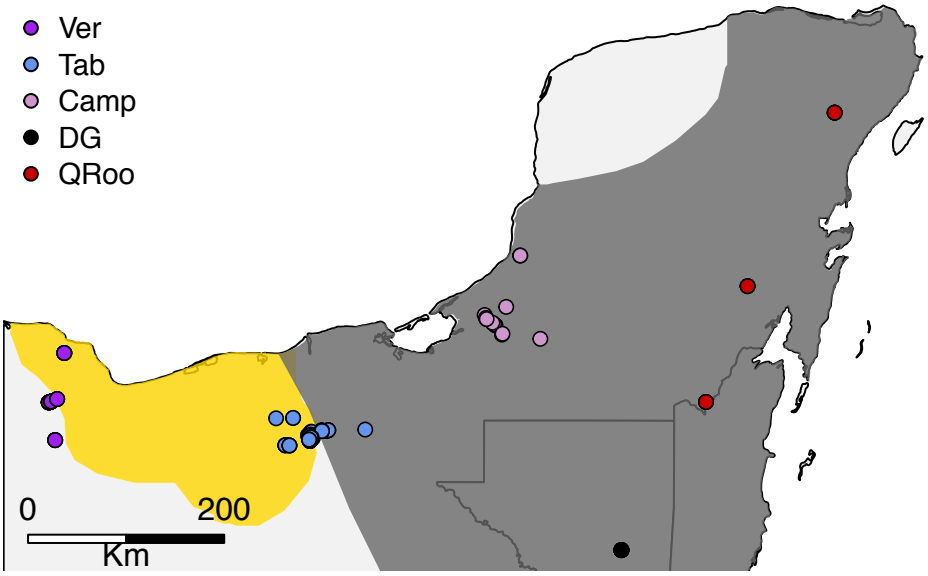

Figure 1.1. Map of sampling sites used in this study. Ver=Veracruz, MX, Tab=Tabasco, MX (hybrid zone), Camp=Campeche, MX, DG=Dolores, Guatemala, QRoo=Quintana Roo, MX.

ddRADseq and genotyping

We prepared and sequenced four ddRAD libraries, following the Peterson et al. (2012) protocol, each library containing DNA from 48 individuals. Briefly, we used the restriction enzymes SphI and MluCI to digest 200-300ng DNA per sample, size selected fragments between 150-350bp, and sequenced libraries on an Illumina HiSeq 4000 machine at the University of Michigan Sequencing Core to obtain 150bp paired-end reads.

We demultiplexed our data using *pyRAD* (Eaton et al. 2014), merged read pairs that overlapped using *FLASH* (Magoč & Salzberg 2011), and aligned both successfully merged reads and unmerged reads (which were expected due to our size selection window) to the draft *Alouatta palliata* genome assembly (accession ID PVKV00000000) using *BWA-MEM* (Li 2013). We then called variants and generated a VCF file using *samtools mpileup* and *bcftools call* (Li et

al. 2009). After removing SNPs within 5bp of an indel and retaining variants with a minimum quality score of 20, we obtained 6,415,368 loci.

Because we were interested in the functional significance of loci associated with reproductive isolation, we limited our genotyping to sites within protein-coding genes. Because the reference assembly for *A. palliata* is not annotated, we used AUGUSTUS v3.2.2 (Stanke & Waack 2003) to identify locations of predicted genes using human as the training species. AUGUSTUS is an *ab initio* gene prediction tool that predicts locations of protein-coding genes in the assembly based solely on its sequence and the known gene structure for the training species. After calling raw genotypes, we retained only biallelic SNPs that fell within AUGUSTUS predicted genes. We further filtered loci to exclude sites with a minimum mean depth across individuals of less than eight. We also excluded sites with a minor allele frequency of less than 0.05, thinned sites within 1kb, and discarded sites with missing data for 80% or more of the individuals in either parental population. Finally, within predicted genes that had multiple variable positions, we retained only the first site. This resulted in 3,242 SNPs that we used in genomic cline analysis. The total length of contigs containing these SNPs (~558 Mb) represents ~18% of the *Alouatta* genome assembly. All filtering was done using vcftools v0.1.14 (Danecek et al. 2011).

Genomic cline analysis

To quantify introgression and identify candidate variants with evidence for adaptation and reproductive isolation, we used genomic cline analysis implemented in *bgc* (Gompert & Buerkle 2012). Genomic cline analysis uses differential introgression to identify loci that are more or less likely than the genome-wide average (assumed to be neutral) to introgress between

populations. *bgc* uses two cline parameters to quantify introgression at each locus. The α parameter describes the direction of introgression, in this case with $\alpha<0$ indicating excess of *A. palliata* ancestry and $\alpha>0$ indicating excess of *A. pigra* ancestry. The β parameter describes the amount of introgression at a locus, with $\beta<0$ indicating greater than expected introgression and $\beta>0$ indicating reduced introgression with respect to the genome-wide average. Loci that are advantageous in both parental populations are expected to exhibit directional introgression (i.e., a shift in cline center, $\alpha>0$ or $\alpha<0$), while barrier loci are expected to have reduced introgression ($\beta>0$) due to selection against hybrids (Gompert & Buerkle 2012).

We ran genomic cline analyses under the genotype uncertainty model in *bgc* for 5 independent chains, each with a burn-in of 30,000 for 50,000 steps, and thinned samples by 20. We then merged outputs and identified outliers with respect to both α and β from MCMC output as loci with a 95% credible interval that does not overlap zero.

Functional annotation

To test whether functional characteristics of outlier genes differ from genes that introgress neutrally, we performed comparative functional enrichment analyses for genes in each beta category. To obtain functional annotations for each variant, we first created a protein database from human proteins accessed from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/protein/protein.fa.gz). We used blastp to query *Alouatta* protein sequences from the AUGUSTUS output for the 3,242 predicted genes in our dataset in the human protein database to identify homologous human proteins. We obtained the top hit and top alignment for each sequence and otherwise used default blastp settings. There were no matches for 412 query sequences, suggesting limited but some protein sequence

divergence between *Alouatta* and *Homo*. To limit our analyses to close matches, we retained proteins with blastp hits where the alignment length was greater than 40% of the query length and percent identity was greater than 60%.

Next, we used refseq accession IDs from the blastp results as query values in biomaRt (Durinck et al. 2005, 2009) to obtain the human HGNC symbol, chromosomal location, and gene ontology IDs for each blast hit. We retrieved annotation information for 1,659 loci, which we used in the following analyses.

Identification of testis-expressed variants

We obtained a list of genes previously identified to have elevated expression (tissue-enriched, group-enriched, or tissue-enhanced) in the human testis (N=2,200) from The Human Protein Atlas (Uhlén et al. 2015) at www.proteinatlas.org (accessed on November 7, 2017). Of these, 119 testis-expressed genes are represented in our dataset occurring on 21 human chromosomes (Figure A1.1).

Functional enrichment

We used two methods to test whether functional characteristics of outlier loci differ from loci that introgress neutrally. We first retrieved Gene Ontology (GO) term classifications (Ashburner et al. 2000) for each gene. The GO classifications are broken down into three independent ontologies to reflect unique aspects of gene functions: molecular function (MF), biological process (BP), and cellular component (CC). Molecular function describes the biochemical activity of a gene product (e.g., enzyme, ligand, binding). Biological process describes the biological outcome to which the gene product contributes (e.g., translation, cell

growth and maintenance). Cellular component describes the location in the cell or in the extracellular space where the gene product is active (e.g., ribosome, nucleus, mitochondrion). We then conducted GO term enrichment analyses for each of the three ontologies (MF, BP, CC) for genes exhibiting different patterns of introgression using the 'parentchild' algorithm and the Fisher test statistic implemented in *topGO* (Alexa & Rahnenfuhrer 2016). For each gene category, we identified the top 20 GO and adjusted p-values for multiple testing.

To test the prediction that genes with reduced introgression are enriched for testis-specific expression, we retrieved expression data for genes in each introgression category using TissueEnrich (Jain 2018). We calculated the proportion of tissue-specific genes (TSG) as the number of tissue-specific genes per tissue divided by the total number of tissue-specific genes in each category.

Results

Introgression

We found a small percentage of genes with cline parameters that were significantly different from zero (Table 1.1), consistent with non-neutral introgression. A similar proportion of genes exhibited reduced and increased introgression, but those with increased introgression had a greater deviation from zero (mean β=-1.42) than those with reduced introgression (mean β=1.01). Genes with increased introgression ($\beta<0$) are not symmetrically centered around α=0, as they tend to be for genes with decreased introgression ($\beta>0$), particularly for outliers (Figure 1.2). This is consistent with the idea that genes that may underlie reproductive isolation are involved in DMIs involving alleles from both species in a balanced manner.

More genes had a non-neutral direction of introgression (15.4%) than a non-neutral amount of introgression (12.7%). Direction of introgression was skewed, with a greater proportion of outliers having excess ancestry from *A. pigra* ($\alpha>0$). Deviation from zero was also greater for these genes (mean $\alpha=1.41$) compared to genes with excess ancestry from *A. palliata* (mean $\alpha=-1.03$). There were 121 genes (3.7%) that were outliers for both cline parameters. All of these genes had increased introgression ($\beta<0$) and all but five of them had excess ancestry from *A. pigra* ($\alpha>0$). These results suggest a stronger signal of directional introgression from *A. pigra* to *A. palliata*, particularly for genes that may underlie adaptive introgression (i.e., those with excess *A. pigra* ancestry that also have an increased rate of introgression).

Table 1.1. Number of genes that are consistent with the genome-wide average (zero) and that are outliers for genomic cline parameters α (direction of introgression) and β (amount of introgression). Positive α indicates genes with excess *A. pigra* ancestry, while negative α indicates genes with excess *A. palliata* ancestry. Positive β indicates genes with reduced introgression, while negative β indicates genes with increased introgression.

Cline parameter	Positive	Zero	Negative
α			
N genes (%)	290 (8.9)	2,742 (84.6)	210 (6.5)
mean ± SD	1.41 ± 0.52	-0.10 ± 0.51	-1.03 ± 0.54
β			
N genes (%)	199 (6.1)	2,831 (87.3)	212 (6.5)
mean ± SD	1.01 ± 0.14	0.03 ± 0.52	-1.42 ± 0.58

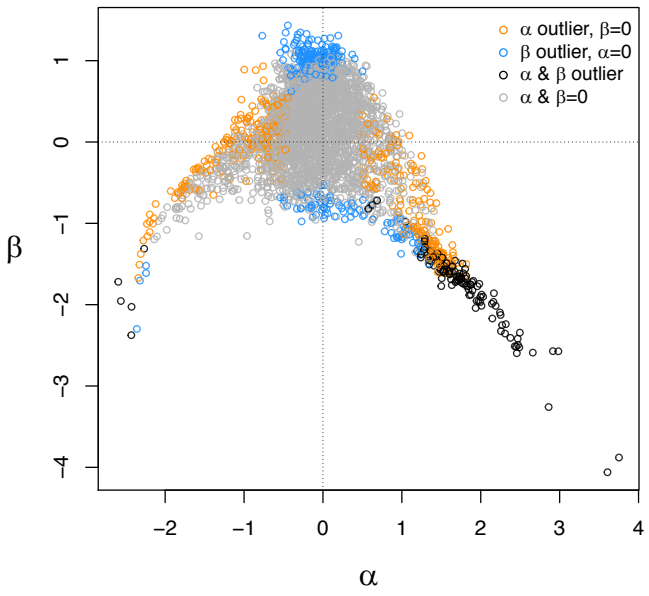

Figure 1.2. Locus-specific point estimates for genomic cline parameters α (x-axis) and β (y-axis), with beta outliers in blue, alpha outliers in orange, and outliers for both cline parameters in black.

Genomic architecture of differential introgression

We examined the distribution in the human genome for the genes in our dataset for which we were able to retrieve functional annotation information. We excluded 22 genes that were mapped to scaffolds not placed on human chromosomes (e.g., CHR_HG2066_PATCH). The number of genes mapped to each chromosome is consistent with expectations based on chromosome size, with a general negative trend between chromosome number and gene number (Figure 1.3). This suggests that although we used reduced-representation sequencing, our genotyping may be representative of the entire genome and not overly biased to particular genomic regions.

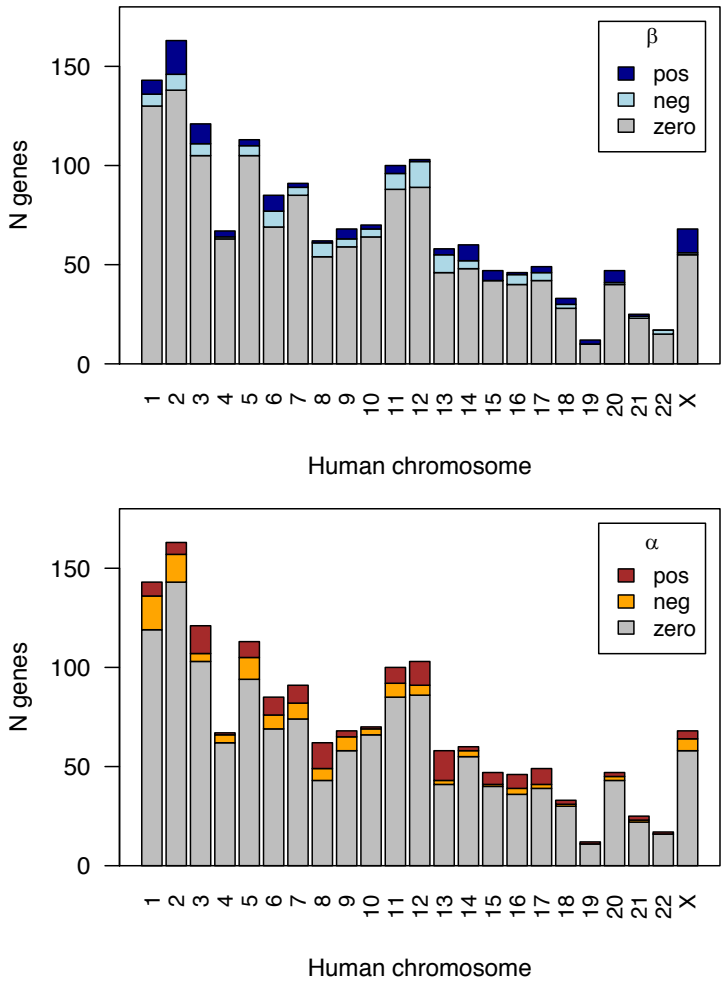

Figure 1.3. Distribution of genes annotated in our dataset in the human genome (N=1,659).

The number of outliers per chromosome for both cline parameters follows a similar trend, where larger chromosomes tend to have more outliers than smaller chromosomes (Figure 1.3). For both parameters, each chromosome contains at least one and up to 25 outlier genes. This broad distribution of outliers across the genome is consistent for positive and negative outliers suggesting a widespread genomic architecture of reproductive isolation and directional introgression.

Notably, most of the genes on the X chromosome (81%) exhibit a neutral amount of introgression. However, with the exception of one gene, all X-linked beta outliers are consistent with reduced introgression ($\beta>0$). The direction of introgression for X-linked genes tends toward excess *A. palliata* ancestry (mean α_X=-0.16). There are six X-linked alpha outliers with excess *A. palliata* ancestry and four with excess *A. pigra* ancestry. This is in contrast to the pattern for autosomal genes (mean α_A=-0.03), particularly for alpha outliers which tend toward excess *A. pigra* ancestry. There are 108 autosomal genes with excess *A. palliata* ancestry, but 141 with excess *A. pigra* ancestry.

Tissue-specific expression

We tested whether genes with non-neutral introgression are enriched for functional classifications. To examine if our data are consistent with hybrid male sterility, we asked whether genes with reduced introgression were enriched for tissue-specific expression in human male tissues (prostate, seminal vesicle, testis). For each male-specific tissue, the proportion of genes with reduced introgression was less than that for genes with both increased and neutral introgression (Figure 1.4). None of the genes with reduced introgression had prostate-specific or seminal vesicle-specific expression. The proportion of prostate- and seminal vesicle-specific genes was low for genes in each beta category, but slightly greatest for genes with a neutral amount of introgression (prostate: $\beta>0$ proportion TSG=0, $\beta<0$ proportion TSG=0.02, $\beta=0$ proportion TSG=0.02, seminal vesicle: $\beta>0$ proportion TSG=0, $\beta<0$ proportion TSG=0.02, $\beta=0$ proportion TSG=0.03). A larger proportion of genes with reduced introgression had testis-specific expression ($\beta>0$ proportion TSG=0.10) compared to other male-specific tissues, but this was lower than the proportion TSG for testis-expressed genes in the other beta categories. For

testis-expressed genes, the proportion TSG was greater for genes with neutral introgression (proportion TSG=0.13) and greatest for genes with increased introgression (proportion TSG=0.19). These results may suggest that most genes with primary functions in male-specific tissues do not underlie reproductive isolation.

For the direction of introgression (Figure 1.4), the proportion TSG for each male-specific tissue was similar to that for the amount of introgression. The proportion TSG for testis-specific genes was greatest for genes with excess *A. pigra* ancestry, lowest for genes with *A. palliata* ancestry and intermediate for genes with a neutral direction of introgression. There were seven testis-specific genes with increased introgression ($\beta<0$) that also had a non-neutral direction of introgression, and six of them had excess ancestry from *A. pigra* ($\alpha>0$).

We noted some other interesting patterns in the proportion TSG for tissues without male-specific expression. Tissues with the greatest difference in proportion TSG between genes with reduced introgression and increased or neutral introgression were the lymph node and placenta, respectively. This may indicate that genes with functions in immunity and female reproduction underlie reproductive isolation in this system. For genes with directional introgression, tissues with the greatest difference in the proportion TSG between genes in different alpha categories were the testis and cerebral cortex.

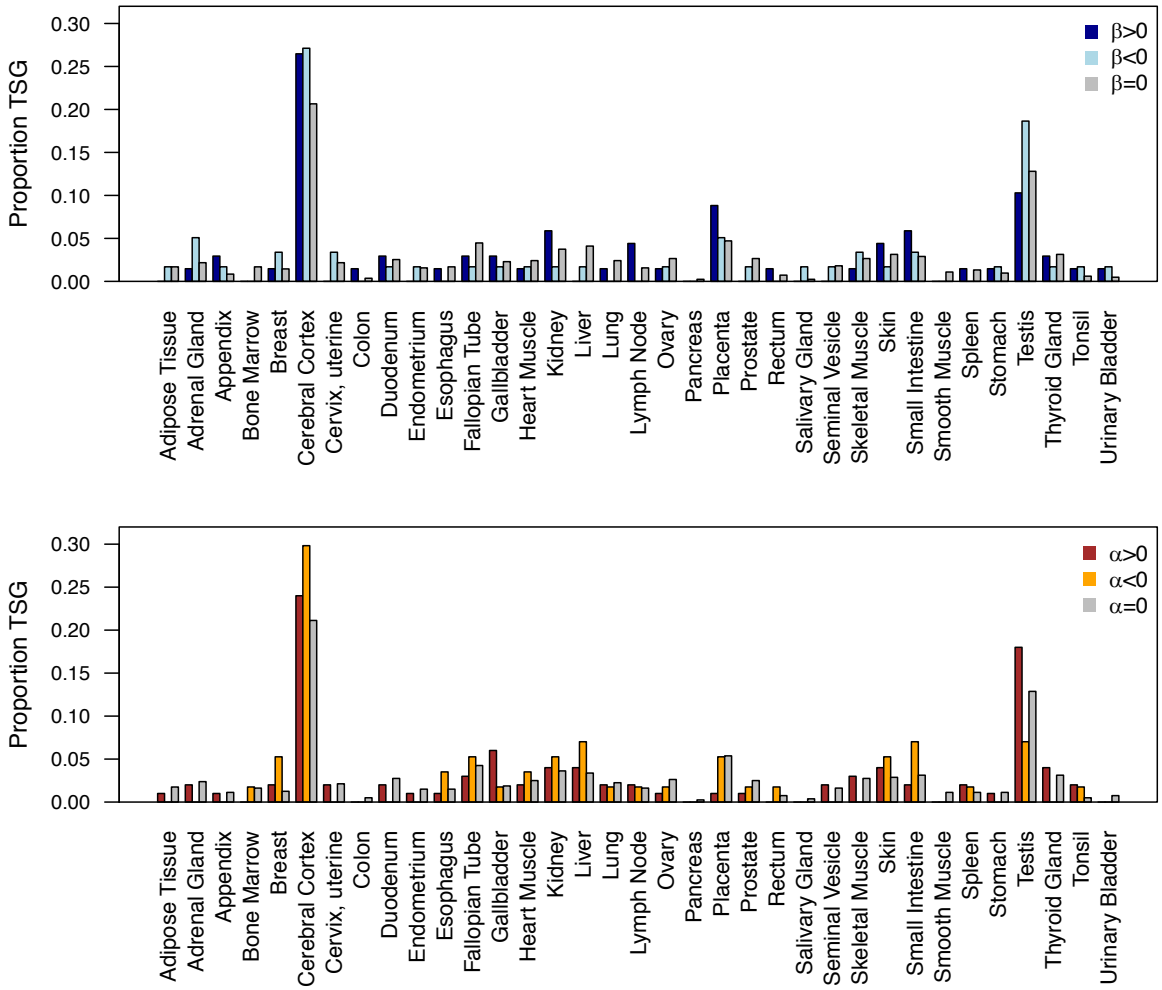

Figure 1.4. The proportion of genes with tissue-specific expression in humans (proportion TSG, Y-axis) for genes with different patterns of introgression. Top panel: Amount of introgression, neutral genes are in gray (β=0), barrier genes are in dark blue (β>0), and genes with increased introgression are in light blue (β<0). Bottom panel: Direction of introgression: neutral genes are in gray (α=0), genes with excess ancestry from *A. pigra* are in red (α>0), genes with excess ancestry from *A. palliata* are in orange (α<0).

Gene ontology enrichment

We found some evidence for functional enrichment of genes with extreme introgression (Table A1.1). Genes with reduced introgression were enriched only for GO terms associated with Biological Process, while genes with increased introgression were enriched for GO terms associated with Biological Process and Cellular Component (Table A1.1b). There was no overlap

between gene sets with reduced and increased introgression in enriched BP GO terms, suggesting these genes perform distinct functions which may underlie their extreme introgression. Notably, genes with reduced introgression were enriched for biological processes involving gene expression (GO:0043628, GO:0042795, GO:0000469, GO:0017148), pigmentation (GO:0042440), behavior (GO:0007626), and other functions. Enrichment for functions related to gene expression were also represented among genes with a neutral amount of introgression (GO:0000122, GO:0010629, GO:0045892) and genes with increased introgression (GO:0098787, GO:0034395), suggesting that either this general process is not involved in reproductive isolation or if it is, the subset of genes involved in gene expression that do contribute have very specific functions.

Genes with increased introgression were enriched for biological processes involving response to stimulus (GO:0034976, GO:0032102, GO:0035360, GO:0035967, GO:1904976, GO:0002832), developmental processes (GO:0021915, GO:0016331, GO:0035148, GO:0072175, GO:0007281, GO:0048854), and other functions. Genes with increased introgression were also enriched for Cellular Component (CC) terms, having functions related to neurons and synapses (GO:0099572, GO:0014069, GO:0044456, GO:0098984, GO:0045202, GO:0098794, GO:0043198), and membranes and membrane proteins (GO:0098590, GO:0060170, GO:0008328). Among genes with increased introgression, there were more enriched terms with intracellular functions than with extra cellular functions or functions at the cell periphery. This finding together with the absence of enriched CC terms for genes with reduced introgression may indicate an absence of functional organization within the cellular component driving reproductive isolation.

Genes with a non-neutral direction of introgression were enriched for terms under Biological Process and Cellular Component ontologies (Table A1.1a). Genes with excess ancestry from *A. pigra* shared similar enriched BP terms with genes with increased introgression: this included functions related to synaptic signaling (GO:0099536, GO:1990709), and response to stimulus (GO:0046578, GO:0009414, GO:0007165). Genes with excess ancestry from *A. palliata* were also enriched for BP terms associated with response to stimulus (GO:1990090, GO:1990089, GO:0050906), and with the developmental process (e.g., GO:0051962, GO:0048706, GO:0050769), and other functions. There were also some similarities in enrichment for CC terms between genes with excess ancestry from *A. pigra* and *A. palliata*. Terms associated with intracellular organelles were highly prevalent for both species, including mitochondrial membrane-specific terms (GO:0005741, GO:0031966, GO:0005740, the latter of which is enriched for genes with excess *A. pigra* and *A. palliata* ancestry).

Discussion

We examined differential introgression using reduced-representation sequence data to detect genes that may be associated with reproductive isolation and directional introgression. The majority of genes exhibited neutral introgression with respect to both the direction and amount of introgression, but a small percentage of markers deviated from neutral introgression (Table 1.1). Our results are consistent with a widespread genomic basis for reproductive isolation and adaptation, with a potential role of the X chromosome in reproductive isolation. Our results are also consistent with skewed directional introgression from *A. pigra* to *A. palliata* for autosomal markers but in the opposite direction for X-linked genes. We detected some functional differences between variants that had neutral and extreme patterns of introgression.

Barrier genes are not enriched for male-specific function

Genes with reduced introgression (barrier genes) were symmetrically distributed around a neutral cline center (α=0), indicating an equal contribution in ancestry from both parental species. This pattern would be expected for barrier loci involved in Dobzhansky-Muller Incompatibilities (Dobzhansky 1936, Muller 1942), since such incompatibilities are caused by interactions between loci with alleles of interspecific origin.

Barrier genes were represented on all human autosomes (except chromosome 22) and on the X chromosome. Due to many chromosomal rearrangements, *Alouatta* chromosomes are not the same as human chromosomes (Steinberg et al. 2014), so we cannot infer chromosomal locations for these markers using human chromosomes. However, the fact that these genes are distributed across human chromosomes suggests that the same is likely true for their distribution across howler monkey chromosomes. This is consistent with observations in other natural hybrid zones (e.g., Janoušek et al. 2015, Rafati et al. 2018), and with a complex genomic basis of reproductive isolation involving many genes. Because *A. pigra* and *A. palliata* diverged ~3 MA (Cortés-Ortiz et al. 2003), this may reflect the accumulation of barriers over the duration of divergence in this species pair (i.e., the snowball effect, Matute et al. 2010). Thus, it is unclear which barrier loci contributed to reproductive isolation during the initial stages of speciation and whether barrier loci may be under selection to strengthen reproductive isolation in the hybrid zone.

Our results are also consistent with a role for the X chromosome in reproductive isolation. In a previous analysis, we found differential introgression of autosomal markers, but a complete lack of introgression for a Y-linked marker and very limited to no introgression for three X-linked markers (Cortés Ortiz et al. 2019). The data analyzed here are representative of a

larger proportion of the X chromosome, lending insight to a more nuanced understanding of potential large X effects in this system. Most of the X-linked genes exhibited neutral introgression, but with the exception of one, all X-linked genes with a non-neutral amount of introgression had reduced introgression. This pattern may be consistent with a large role of the X chromosome in reproductive isolation since the proportion of genes with reduced introgression on the X is greater than that for similar sized autosomes (Figure 1.3). This may suggest that postzygotic isolation is strong in the hybrid zone (Coyne & Orr 1989), although reduced introgression of X-linked markers has been observed in other systems where postzygotic isolation is not particularly strong (e.g., Larson et al. 2014). It will be necessary to use observational data on hybrid fitness to test this hypothesis. To explicitly test for a large X-effect in this system, we will need to validate X-linkage for genomic sequences since the *A. palliata* reference genome is not assembled to chromosome level. Although X chromosome gene content and sequence is known to be well conserved across mammals (Delgado et al. 2009), the rate of chromosomal rearrangements is high in Neotropical primates and sex chromosome rearrangements have occurred in *Alouatta* (Steinberg et al. 2014). Also, sampling for tests of differential introgression including autosomal and X-linked markers should be restricted to female individuals since males are hemizygous for the X. Inclusion of males may bias cline parameter estimates by inflating homozygosity for X-linked SNPs relative to observed genome-wide heterozygosity (e.g., Larson et al. 2014).

Compared to genes with neutral and increased introgression, barrier genes were not particularly enriched for expression in male-specific tissues, which may be inconsistent with the hypothesis that hybrid male sterility is an important postzygotic barrier in this system. However, there were seven barrier genes with testis-specific expression and it is possible that they (and

others not sequenced here) underlie male fertility traits. The genomic architecture of hybrid male sterility phenotypes has only been mapped in select animal species (e.g., Turner & Harr 2014, Masly & Presgraves 2007) so the relationship between the number and types of genes that underlie hybrid male sterility is not well understood. Notably, the *Alouatta* hybrid zone system could contribute to this gap by pairing genomic data like that used in this study with male fertility phenotype data (e.g., sperm morphology traits) to map these traits.

Previous studies of differential introgression in the house mouse and European rabbit hybrids zone found evidence that broad functional organization of the genome drives speciation (Janoušek et al. 2015, Rafati et al. 2018). This hypothesis would be consistent with a large role for DMI's underlying reproductive isolation, since genes with reduced introgression tend be highly interconnected. In contrast, our results do not seem fully consistent with this explanation. Enriched terms for genes with increased introgression included more "intracellular" than "cell periphery" parent terms and genes with reduced introgression were not enriched for terms describing cellular component.

Genes with increased introgression were also distributed across most human chromosomes, with the exception of chromosomes 15 and 19 which only contained barrier genes. In contrast to barrier genes, a larger proportion of genes with increased introgression had testis-specific expression than genes with neutral introgression. This may suggest that interspecific male fertility alleles are not always disadvantageous and that introgression is actually increased for some of these genes because they are adaptive in both species. Genes with increased introgression were highly skewed in their direction of introgression, with most of them tending toward excess *A. pigra* ancestry. Outlier genes with increased introgression that are also outliers for directional introgression may be considered candidate genes for adaptation since they

may be passing through the hybrid zone at an accelerated rate. We discuss this idea further below.

Direction of introgression is asymmetric

We observed asymmetry in the direction of introgression for a subset of genes (Table 1.1). Similar to beta outliers, these genes with directional introgression were not particularly clustered in the human genome (Figure 1.3), suggesting that genes that may be under selection for adaptive introgression are not tightly linked.

Genes with directional introgression had the most pronounced disproportionate tissue-specific expression in the testis and cerebral cortex. All of the testis-specific genes with increased introgression also had directional introgression with excess *A. pigra* ancestry. This is consistent with the idea that instead of being disproportionately involved in hybrid male sterility, testis-specific genes may disproportionately have alleles that are advantageous on both genomic backgrounds (particularly *A. pigra* alleles) and exhibit adaptive introgression in this hybrid zone. For cerebral cortex-specific genes with directional introgression, there was a greater proportion of genes with *A. palliata* ancestry. This may suggest adaptive introgression of alleles that underlie complex phenotypes like social behavior and memory.

There were some similarities in gene ontology enrichment functions between genes with opposing directions of introgression, including a high prevalence of organelle-related (especially for the mitochondria) cellular component terms and synaptic signaling/response to stimulus biological process terms. This may suggest that it is easier for genes associated with these functions to be exchanged across species boundaries, regardless of direction. This may also be

consistent with the finding of Janoušek et al. (2015) that genes with increased introgression tend to be involved in signal transduction.

In general, a larger proportion of genes with directional introgression had excess *A. pigra* ancestry than *A. palliata* ancestry. This pattern could result from several mechanisms, including (1) a bias in the direction of backcrossing in the hybrid zone, (2) a bias in the direction of adaptive introgression, (3) and hybrid zone movement. Previously, we investigated admixture in the hybrid zone using a set of microsatellite markers genotyped for more than three times the number of individuals analyzed here (N=254 individuals), and found that the occurrence of *A. palliata*-like and *A. pigra*-like hybrids were nearly equal (Cortés-Ortiz et al. 2019). This suggests that backcrossing occurs in both directions with similar frequencies.

It is possible that the asymmetry observed here is the result of skewed adaptive introgression if *A. pigra* alleles are generally more favorable than *A. palliata* alleles in the hybrid zone. Our results are consistent with this pattern, especially for genes with increased introgression. The set of genes with both increased and directional introgression may be considered as candidate genes for adaptive introgression. However, more explicit tests will need to be done on these genes to determine if sequence variation is consistent with positive selection in the recipient population (e.g., Huerta-Sánchez et al. 2014). Further, to robustly demonstrate evidence for adaptive introgression, it will be necessary to demonstrate a fitness advantage for individuals that carry the interspecific allele. This can be quite difficult to accomplish in many systems, particularly for those of involving species with long generation times and for natural populations where organisms cannot undergo experimentation (Taylor & Larson 2019)

The asymmetry in direction of introgression may also be caused by movement of the hybrid zone (e.g., Janoušek et al. 2015). If the hybrid zone has moved from East to West (as we

suspect based on our findings of considerable admixture East of the hybrid zone, data not shown), we would expect to see the pattern observed here. As hybrid zones move, alleles that are not involved in reproductive isolation are left behind. These explanations for asymmetrical introgression are not mutually exclusive. For example, hybrid zone movement would result in both neutral and adaptive alleles being left behind and would thus require further analysis to distinguish between them. In general, it may be possible to distinguish neutral alleles from those that have introgressed adaptively using methods to detect signatures of positive selection in the genome (Taylor & Larson 2019).

Conclusion

We investigated differential introgression across a howler monkey hybrid zone using reduced-representation sequence data from regions of the genome predicted to be coding regions. We identified a subset of genes with non-neutral introgression that may be candidates for reproductive isolation and adaptive introgression. Contrary to our predictions (and predictions of the large X effect on postzygotic isolation), barrier genes were not enriched for male-specific functions. One gene on the X chromosome had increased introgression and testis-specific genes were more highly represented among genes with increased rather than reduced and neutral introgression, suggesting that hybrid male sterility may not be a prominent reproductive barrier in this system.

Similarly, our results are mostly inconsistent with observations in the house mouse hybrid zone in which genes with reduced introgression tended to be expressed intracellularly while genes with increased introgression tended to be expressed in the cell periphery. However, functional properties of genes with directional introgression may be partially consistent with the

explanation for this pattern (and with introgression in the European rabbit hybrid zone) as these genes are associated with GO terms associated with response to stimulus.

Because our sequencing is representative of a small portion of the genome (~18%), our results must be interpreted with caution and it will be important to investigate the patterns uncovered here using whole genome data. However, genes identified here with non-neutral introgression represent a candidate list for further investigation and this work represents an important step in characterizing the functional basis of reproductive isolation in natural non-model systems.

References

Alexa A., Rahnenfuhrer J. (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.28.0.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, **25**, 25–29.

Barton, N. H. (1983). Multilocus clines. *Evolution*, **37**, 454–471.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics,* **16**, 113–148.

Bergman, T. J., Cortés-Ortiz, L., Dias, P. A. D., Ho, L., Adams, D., Canales-Espinosa, D., & Kitchen, D. M. (2016). Striking differences in the loud calls of howler monkey sister species (*Alouatta pigra* and *A. palliata*). *American Journal of Primatology*, **78**, 755–766.

Brekke T.D., Henry L.A., Good J.M. (2016) Genomic imprinting, disrupted placental expression, and speciation. *Evolution* **70**, 2690–2703.

Buerkle, C. A., & Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends in ecology & evolution*, *23*(12), 686-694.

Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., Castillo, R. C., ... & Hoikkala, A. (2012). What do we need to know about speciation? *Trends in ecology & evolution*, **27**, 27–39.

Chapman, C. A., & Balcomb, S. R. (1998). Population characteristics of howlers: ecological conditions or group history. *International Journal of Primatology*, **19**, 385–403.

Cortés-Ortiz, L., Bermingham, E., Rico, C., Rodrıguez-Luna, E., Sampaio, I., & Ruiz-Garcıa, M. (2003). Molecular systematics and biogeography of the Neotropical monkey genus, *Alouatta*. *Molecular Phylogenetics and Evolution*, **26**, 64–81.

Cortés-Ortiz, L., Duda, T. F., Canales-Espinosa, D., García-Orduña, F., Rodríguez-Luna, E., & Bermingham, E. (2007). Hybridization in large-bodied New World primates. *Genetics,* **176**, 2421–2425.

Cortés-Ortiz, L., Nidiffer, M. D., Hermida-Lagunes, J., García-Orduña, F., Rangel-Negrín, A., Kitchen, D. M., … Canales-Espinosa, D. (2019). Reduced introgression of sex chromosome markers in the Mexican howler monkey (*Alouatta palliata* x *A. pigra*) hybrid zone. *International Journal of Primatology,* **40***,* 114–131.

Coyne JA, Orr HA. in Speciation and its Consequences (eds Otte D, Endler J) 180-207 (Sinauer, Sunderland, Massachusetts, 1989).

Coyne JA, Orr HA (1997) "Patterns of speciation in Drosophila" revisited. Evolution, **51**, 295–303.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 156–158.

Delgado, C. L. R., Waters, P. D., Gilbert, C., Robinson, T. J., & Graves, J. A. M. (2009). Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Research*, **17**, 917–926.

Dobzhansky, T. H. (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, **21**, 113–135.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **4**, 1184–1191.

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Gompert, Z., & Buerkle, C. A. (2012a). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Good, J. M., Giger, T., Dean, M. D., & Nachman, M. W. (2010). Widespread over-expression of the X chromosome in sterile F1 hybrid mice. *PLoS genetics*, **6**, e1001148.

Haldane, J. B. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics*, **12**, 101–109.

Harrison R. G. (1990) Hybrid zones: windows on evolutionary process. *Oxford surveys in evolutionary biology* **7**, 69–128.

Huerta-Sánchez, E., Jin, X., Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., ... & Wang, B. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, **512**, 194–197.

Jain, A, Tuteja, G. (2018) TissueEnrich: Tissue-specific gene enrichment analysis. Bioinformatics, bty890. 10.1093/bioinformatics/bty890

Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, **32**, 1208–1220.

Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J., ... & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, **360**, 1355–1358.

Kelaita, M., Dias, P. A. D., Aguilar-Cucurachi, M., Canales-Espinosa, D., & Cortés-Ortiz, L. (2011). Impact of intrasexual selection on sexual dimorphism and testes size in the Mexican howler monkeys *Alouatta palliata* and *A. pigra*. American Journal of Physical Anthropology, **146**, 179–187.

Kelaita, M. A., & Cortés-Ortiz, L. (2013). Morphological variation of genetically confirmed *Alouatta pigra* × *A. palliata* hybrids from a natural hybrid zone in Tabasco, Mexico. *American Journal of Physical Anthropology*, **150**, 223–234.

Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution*, **67**, 3653–3661.

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Molecular Ecology*, **23**, 1668-1678.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997.*

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS biology*, **5**, e243.

Matute D. R., Butler I. A., Turissini D. A., Coyne J. A. (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, **329**, 1518–1521.

Mihola O., Trachtulec Z., Vlcek C., Schimenti J. C., Forejt J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science,* **323**,373–375.

Muller, H. (1942). Isolating mechanisms, evolution, and temperature. In *Biol. Symp.* (Vol. 6, pp. 71-125).

Noor, M. A., & Feder, J. L. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics*, **7**, 851–861.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135.

Rafati, N., Blanco-Aguiar, J. A., Rubin, C. J., Sayyab, S., Sabatino, S. J., Afonso, S., ... & Andersson, L. (2018). A genomic map of clinal variation across the European rabbit hybrid zone. *Molecular ecology*, **27**, 1457–1478.

Seutin, G., White, B. N., Boag, P.T., 1991. Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology, **69**, 82–90.

Smith, J. D. (1970). The systematic status of the black howler monkey, *Alouatta pigra* Lawrence. *Journal of Mammalogy*, **51**, 358–369.

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F. R., Shih, C. H., ... & Kohn, M. H. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296–1301.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, **19**, ii215–ii225.

Steinberg, E. R., Cortés-Ortiz, L., Nieves, M., Bolzán, A. D., García-Orduña, F., Hermida-Lagunes, J., ... Mudry, M. D. (2008). The karyotype of *Alouatta pigra* (Primates: Platyrrhini): mitotic and meiotic analyses. *Cytogenetic and Genome Research*, **122**, 103–109.

Steinberg, E. R., Nieves, M., & Mudry, M. D. (2014). Multiple sex chromosome systems in howler monkeys (Platyrrhini, *Alouatta*). *Comparative cytogenetics*, **8**, 43–69.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature ecology & evolution*, **3**, 170 –177.

Turner, L. M., Schwahn, D. J., Harr, B. (2012) Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone. *Evolution*, **66**, 443–458.

Turner, L. M., Harr, B. (2014) Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife*, **3**, e02504.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Olsson, I. (2015). Tissue-based map of the human proteome. *Science*, **347**, 1260419.

Van Belle, S., & Estrada, A. (2006). Demographic features of *Alouatta pigra* populations in extensive and fragmented forests. In *New perspectives in the study of Mesoamerican primates* (pp. 121-142). Springer, Boston, MA.

APPENDIX A1

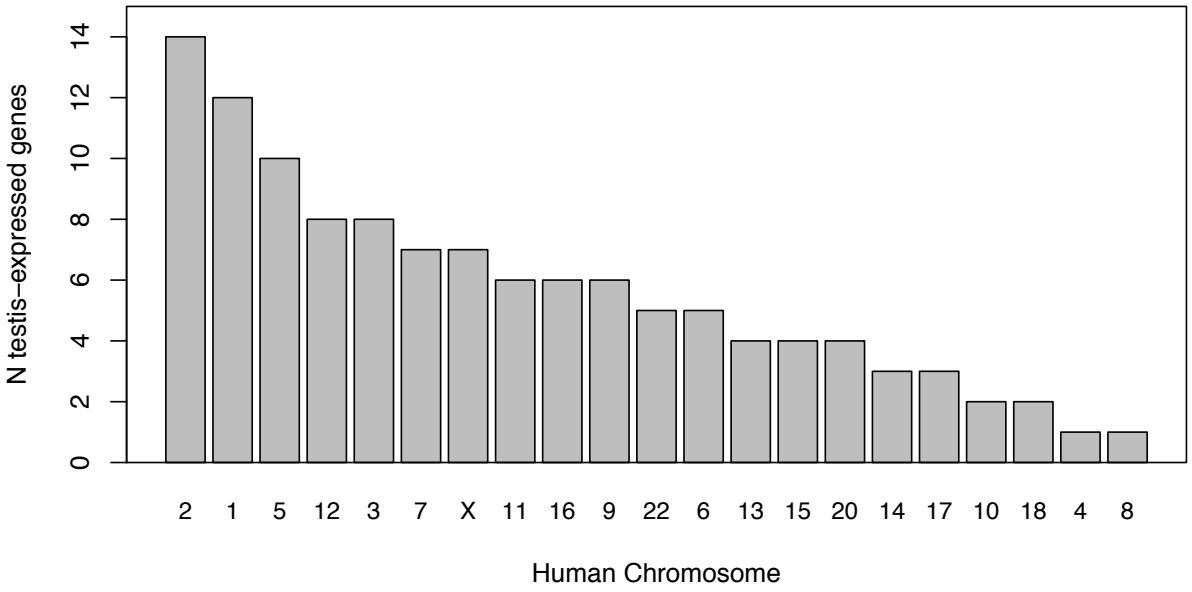

Figure A1.1 Distribution of the subset of testis-expressed genes that are represented in our dataset across human chromosomes.

Table A1.1a. Significantly enriched GO terms (BP=biological process, CC=cellular component, MF=molecular function) for genes that exhibit excess *A. pigra* ancestry (α=pos), excess *A. palliata* ancestry (α=neg), and non-directional introgression (α=zero).

GO.ID	Term	Ann.	Sig.	Exp.	p.CF	p.BH	GO	α
GO:0051955	regulation of amino acid transport	3	3	0.28	0.001	0.014	BP	pos
GO:0099536	synaptic signaling	69	12	6.48	0.002	0.014	BP	pos
GO:0003008	system process	161	26	15.11	0.003	0.014	BP	pos
GO:0046578	regulation of Ras protein signal transdu...	24	6	2.25	0.004	0.014	BP	pos
GO:0032890	regulation of organic acid transport	4	3	0.38	0.004	0.014	BP	pos
GO:0031076	embryonic camera-type eye development	5	3	0.47	0.004	0.014	BP	pos
GO:0007154	cell communication	541	63	50.78	0.007	0.015	BP	pos
GO:1903523	negative regulation of blood circulation	2	2	0.19	0.007	0.015	BP	pos
GO:0050891	multicellular organismal water homeostas...	2	2	0.19	0.008	0.015	BP	pos
GO:1990709	presynaptic active zone organization	2	2	0.19	0.008	0.015	BP	pos
GO:0032892	positive regulation of organic acid tran...	2	2	0.19	0.009	0.015	BP	pos
GO:0009414	response to water deprivation	2	2	0.19	0.010	0.015	BP	pos
GO:0045932	negative regulation of muscle contractio...	2	2	0.19	0.011	0.015	BP	pos

GO:0051590	positive regulation of neurotransmitter ...	2	2	0.19	0.011	0.015	BP	pos
GO:0050879	multicellular organismal movement	10	4	0.94	0.012	0.015	BP	pos
GO:0060840	artery development	13	3	1.22	0.013	0.015	BP	pos
GO:0023052	signaling	541	63	50.78	0.015	0.015	BP	pos
GO:0045088	regulation of innate immune response	36	7	3.38	0.015	0.015	BP	pos
GO:0007165	signal transduction	505	58	47.4	0.015	0.015	BP	pos
GO:0021602	cranial nerve morphogenesis	3	2	0.28	0.015	0.015	BP	pos
GO:0031968	organelle outer membrane	15	6	1.33	0.001	0.012	CC	pos
GO:0019867	outer membrane	15	6	1.33	0.001	0.012	CC	pos
GO:0005741	mitochondrial outer membrane	14	6	1.24	0.002	0.014	CC	pos
GO:0097386	glial cell projection	2	2	0.18	0.008	0.033	CC	pos
GO:0005883	neurofilament	2	2	0.18	0.008	0.033	CC	pos
GO:0098588	bounding membrane of organelle	160	20	14.2	0.010	0.033	CC	pos
GO:0005740	mitochondrial envelope	47	7	4.17	0.015	0.043	CC	pos
GO:0048786	presynaptic active zone	7	3	0.62	0.017	0.044	CC	pos
GO:0051962	positive regulation of nervous system de...	56	12	3.91	5.80E-05	0.001	BP	neg
GO:1990090	cellular response to nerve growth factor...	7	4	0.49	0.001	0.006	BP	neg
GO:1990089	response to nerve growth factor	7	4	0.49	0.001	0.007	BP	neg
GO:0050906	detection of stimulus involved in sensor...	9	3	0.63	0.002	0.010	BP	neg
GO:0048706	embryonic skeletal system development	7	3	0.49	0.003	0.010	BP	neg
GO:0031346	positive regulation of cell projection o...	38	8	2.65	0.004	0.012	BP	neg
GO:0050769	positive regulation of neurogenesis	51	11	3.56	0.004	0.012	BP	neg
GO:0008544	epidermis development	31	6	2.17	0.005	0.012	BP	neg
GO:0007606	sensory perception of chemical stimulus	10	3	0.7	0.006	0.013	BP	neg
GO:0051960	regulation of nervous system development	97	14	6.78	0.007	0.014	BP	neg
GO:0010720	positive regulation of cell development	56	11	3.91	0.009	0.014	BP	neg
GO:0048468	cell development	214	25	14.95	0.010	0.014	BP	neg
GO:0038179	neurotrophin signaling pathway	7	3	0.49	0.010	0.014	BP	neg
GO:0034645	cellular macromolecule biosynthetic proc...	393	34	27.46	0.011	0.014	BP	neg
GO:0006968	cellular defense response	3	2	0.21	0.012	0.014	BP	neg
GO:0072310	glomerular epithelial cell development	2	2	0.14	0.012	0.014	BP	neg
GO:0006518	peptide metabolic process	80	11	5.59	0.012	0.014	BP	neg
GO:0098900	regulation of action potential	7	3	0.49	0.013	0.014	BP	neg
GO:0031280	negative regulation of cyclase activity	3	2	0.21	0.014	0.015	BP	neg
GO:0010810	regulation of cell-substrate adhesion	19	3	1.33	0.015	0.015	BP	neg
GO:0044391	ribosomal subunit	21	6	1.56	0.003	0.033	CC	neg
GO:1904949	ATPase complex	3	2	0.22	0.003	0.033	CC	neg
GO:0098827	endoplasmic reticulum subcompartment	76	9	5.65	0.010	0.041	CC	neg
GO:0031966	mitochondrial membrane	44	8	3.27	0.011	0.041	CC	neg
GO:0070603	SWI/SNF superfamily-type complex	3	2	0.22	0.011	0.041	CC	neg

GO:0030425	dendrite	74	11	5.5	0.013	0.041	CC	neg
GO:0005840	ribosome	27	6	2.01	0.014	0.041	CC	neg
GO:1902493	acetyltransferase complex	6	2	0.45	0.020	0.045	CC	neg
GO:0097440	apical dendrite	2	2	0.15	0.020	0.045	CC	neg
GO:0005740	mitochondrial envelope	47	8	3.49	0.022	0.045	CC	neg
GO:0016192	vesicle-mediated transport	153	141	130.22	0.002	0.036	BP	zreo
GO:0050678	regulation of epithelial cell proliferat...	29	29	24.68	0.006	0.036	BP	zreo
GO:0015807	L-amino acid transport	5	5	4.26	0.008	0.036	BP	zreo
GO:0044419	interspecies interaction between organis...	69	65	58.73	0.008	0.036	BP	zreo
GO:0050673	epithelial cell proliferation	30	30	25.53	0.011	0.036	BP	zreo
GO:0034284	response to monosaccharide	11	11	9.36	0.013	0.036	BP	zreo
GO:0043161	proteasome-mediated ubiquitin-dependent ...	34	32	28.94	0.017	0.036	BP	zreo
GO:0098901	regulation of cardiac muscle cell action...	5	5	4.26	0.018	0.036	BP	zreo
GO:0000398	mRNA splicing, via spliceosome	30	28	25.53	0.019	0.036	BP	zreo
GO:0034599	cellular response to oxidative stress	22	22	18.72	0.020	0.036	BP	zreo
GO:0006979	response to oxidative stress	34	33	28.94	0.020	0.036	BP	zreo
GO:0000904	cell morphogenesis involved in different...	86	79	73.19	0.025	0.036	BP	zreo
GO:0071407	cellular response to organic cyclic comp...	61	56	51.92	0.025	0.036	BP	zreo
GO:0007179	transforming growth factor beta receptor...	24	23	20.43	0.029	0.036	BP	zreo
GO:0071559	response to transforming growth factor b...	27	26	22.98	0.030	0.036	BP	zreo
GO:0006259	DNA metabolic process	61	57	51.92	0.031	0.036	BP	zreo
GO:0097485	neuron projection guidance	32	31	27.24	0.032	0.036	BP	zreo
GO:0022610	biological adhesion	154	139	131.07	0.033	0.036	BP	zreo
GO:0033036	macromolecule localization	252	223	214.48	0.034	0.036	BP	zreo
GO:0000375	RNA splicing, via transesterification re...	30	28	25.53	0.037	0.037	BP	zreo

Table A1.1b. Significantly enriched GO terms (BP=biological process, CC=cellular component, MF=molecular function) for genes that exhibit reduced (β=pos), increased (β=neg), and neutral introgression (β=zero).

GO.ID	Term	Ann.	Sig.	Exp.	p.CF	p.BH	GO	β
GO:0010837	regulation of keratinocyte proliferation	5	3	0.35	0.003	0.026	BP	pos
GO:0043616	keratinocyte proliferation	6	3	0.41	0.005	0.026	BP	pos
GO:1902115	regulation of organelle assembly	12	3	0.83	0.007	0.026	BP	pos
GO:0033013	tetrapyrrole metabolic process	3	2	0.21	0.010	0.026	BP	pos
GO:0043085	positive regulation of catalytic activit...	135	8	9.34	0.011	0.026	BP	pos
GO:0043628	ncRNA 3'-end processing	5	3	0.35	0.012	0.026	BP	pos
GO:0042795	snRNA transcription by RNA polymerase II	4	2	0.28	0.013	0.026	BP	pos
GO:1902017	regulation of cilium assembly	3	2	0.21	0.013	0.026	BP	pos
GO:1902905	positive regulation of supramolecular fi...	24	3	1.66	0.013	0.026	BP	pos

GO:0007626	locomotory behavior	19	4	1.31	0.016	0.026	BP	pos
GO:0051606	detection of stimulus	23	5	1.59	0.016	0.026	BP	pos
GO:0017148	negative regulation of translation	7	2	0.48	0.018	0.026	BP	pos
GO:0042440	pigment metabolic process	4	2	0.28	0.019	0.026	BP	pos
GO:0048284	organelle fusion	6	2	0.41	0.020	0.026	BP	pos
GO:0032007	negative regulation of TOR signaling	6	2	0.41	0.022	0.026	BP	pos
GO:0043396	corticotropin-releasing hormone secretio...	1	1	0.07	0.023	0.026	BP	pos
GO:0051345	positive regulation of hydrolase activit...	82	7	5.67	0.024	0.026	BP	pos
GO:0034330	cell junction organization	33	5	2.28	0.024	0.026	BP	pos
GO:0021781	glial cell fate commitment	1	1	0.07	0.026	0.026	BP	pos
GO:0000469	cleavage involved in rRNA processing	4	2	0.28	0.026	0.026	BP	pos
GO:0051290	protein heterotetramerization	3	3	0.19	0.003	0.028	BP	neg
GO:0034976	response to endoplasmic reticulum stress	21	4	1.35	0.004	0.028	BP	neg
GO:0021915	neural tube development	14	4	0.9	0.004	0.028	BP	neg
GO:0016331	morphogenesis of embryonic epithelium	16	4	1.03	0.009	0.033	BP	neg
GO:0035148	tube formation	17	4	1.09	0.013	0.033	BP	neg
GO:0032102	negative regulation of response to exter...	30	5	1.93	0.013	0.033	BP	neg
GO:0006379	mRNA cleavage	1	1	0.06	0.014	0.033	BP	neg
GO:0072175	epithelial tube formation	15	4	0.96	0.019	0.033	BP	neg
GO:0007281	germ cell development	17	3	1.09	0.022	0.033	BP	neg
GO:0048854	brain morphogenesis	6	2	0.39	0.022	0.033	BP	neg
GO:0098787	mRNA cleavage involved in mRNA processin...	1	1	0.06	0.023	0.033	BP	neg
GO:0035360	positive regulation of peroxisome prolif...	1	1	0.06	0.025	0.033	BP	neg
GO:0035967	cellular response to topologically incor...	14	3	0.9	0.027	0.033	BP	neg
GO:0048871	multicellular organismal homeostasis	29	5	1.86	0.027	0.033	BP	neg
GO:1904976	cellular response to bleomycin	1	1	0.06	0.028	0.033	BP	neg
GO:0002832	negative regulation of response to bioti...	4	2	0.26	0.028	0.033	BP	neg
GO:0044260	cellular macromolecule metabolic process	692	44	44.44	0.030	0.033	BP	neg
GO:1900119	positive regulation of execution phase o...	1	1	0.06	0.030	0.033	BP	neg
GO:0009247	glycolipid biosynthetic process	1	1	0.06	0.031	0.033	BP	neg
GO:0034395	regulation of transcription from RNA pol...	1	1	0.06	0.033	0.033	BP	neg
GO:0098590	plasma membrane region	128	16	7.92	0.004	0.041	CC	neg
GO:0099572	**postsynaptic specialization**	39	7	2.41	0.009	0.041	CC	neg
GO:0014069	**postsynaptic density**	37	7	2.29	0.011	0.041	CC	neg
GO:0060170	ciliary membrane	9	3	0.56	0.014	0.041	CC	neg
GO:0044456	**synapse part**	101	12	6.25	0.018	0.041	CC	neg
GO:0098984	**neuron to neuron synapse**	38	8	2.35	0.021	0.041	CC	neg
GO:0008328	ionotropic glutamate receptor complex	11	3	0.68	0.023	0.041	CC	neg
GO:0000794	condensed nuclear chromosome	5	2	0.31	0.025	0.041	CC	neg
GO:0045202	**synapse**	118	13	7.3	0.026	0.041	CC	neg

GO:0000123	histone acetyltransferase complex	6	2	0.37	0.029	0.041	CC	neg
GO:0031983	vesicle lumen	24	4	1.49	0.029	0.041	CC	neg
GO:0005856	cytoskeleton	207	19	12.81	0.030	0.041	CC	neg
GO:0060205	cytoplasmic vesicle lumen	24	4	1.49	0.031	0.041	CC	neg
GO:0005793	endoplasmic reticulum-Golgi intermediate...	5	2	0.31	0.032	0.041	CC	neg
GO:1904813	ficolin-1-rich granule lumen	6	2	0.37	0.035	0.041	CC	neg
GO:0044322	endoplasmic reticulum quality control co...	5	2	0.31	0.036	0.041	CC	neg
GO:0044430	cytoskeletal part	150	15	9.28	0.038	0.041	CC	neg
GO:0072562	blood microparticle	6	2	0.37	0.039	0.041	CC	neg
GO:0098794	**postsynapse**	74	9	4.58	0.039	0.041	CC	neg
GO:0043198	**dendritic shaft**	5	2	0.31	0.041	0.041	CC	neg
GO:0046872	metal ion binding	364	326	317.95	0.001	0.028	MF	zreo
GO:0051050	positive regulation of transport	78	76	67.93	0.002	0.019	BP	zreo
GO:0000122	negative regulation of transcription by ...	76	75	66.18	0.002	0.019	BP	zreo
GO:0048468	cell development	214	200	186.36	0.004	0.019	BP	zreo
GO:0010629	negative regulation of gene expression	159	150	138.47	0.004	0.019	BP	zreo
GO:0051347	positive regulation of transferase activ...	55	53	47.9	0.005	0.019	BP	zreo
GO:0097190	apoptotic signaling pathway	55	53	47.9	0.007	0.019	BP	zreo
GO:2001233	regulation of apoptotic signaling pathwa...	33	33	28.74	0.009	0.019	BP	zreo
GO:0007275	multicellular organism development	511	459	445.01	0.009	0.019	BP	zreo
GO:0019693	ribose phosphate metabolic process	39	39	33.96	0.010	0.019	BP	zreo
GO:0008152	metabolic process	925	820	805.54	0.011	0.019	BP	zreo
GO:0045934	negative regulation of nucleobase-contai...	123	117	107.12	0.011	0.019	BP	zreo
GO:0042327	positive regulation of phosphorylation	76	72	66.18	0.012	0.019	BP	zreo
GO:0051253	negative regulation of RNA metabolic pro...	117	111	101.89	0.013	0.019	BP	zreo
GO:0042330	taxis	49	48	42.67	0.013	0.019	BP	zreo
GO:0045892	negative regulation of transcription, DN...	109	104	94.92	0.015	0.019	BP	zreo
GO:0070848	response to growth factor	74	71	64.44	0.017	0.019	BP	zreo
GO:0044237	cellular metabolic process	870	772	757.64	0.017	0.019	BP	zreo
GO:0044283	small molecule biosynthetic process	69	67	60.09	0.018	0.019	BP	zreo
GO:0051130	positive regulation of cellular componen...	113	106	98.41	0.019	0.019	BP	zreo
GO:0030030	cell projection organization	165	153	143.69	0.019	0.019	BP	zero

CHAPTER II

X chromosome introgression in the *Alouatta* hybrid zone mirrors archaic ancestry in human genomes

Introduction

Recent advances in sequencing technology have allowed the exploration of genomic signatures of speciation in many organisms. A major observation that has come out of this research is that sex chromosomes are important drivers of reproductive isolation across many animal systems (e.g., birds: Sætre et al. 2002, Carling & Brumfield 2008, *Drosophila*: Presgraves 2008, *Mus*: Good et al. 2008a,b, 2010, Janoušek et al. 2012, fish: Kitano & Peichel 2012). Sequencing of archaic hominin DNA and its comparison with that of modern humans indicates that humans interbred with archaic lineages, resulting in the production of admixed offspring (Reich et al. 2010, 2011, Green et al. 2010, Prüfer et al. 2014). Comparison of archaic hominin genomes with modern human genomes revealed that individuals in Europe and East Asia interbred with Neanderthals and Denisovans, and that some modern humans with ancestors from outside of Africa carry as much as ~5% archaic DNA as a result of this ancient admixture (Sankararaman et al. 2014, 2016). However, there are many regions of reduced ancestry from archaic hominins in modern human genomes (i.e., "deserts" of archaic ancestry), which are especially enriched on the X chromosome (Sankararaman et al. 2014, 2016). Further, Neanderthal Y chromosome sequence has not been discovered in modern human genomes

(Mendez et al. 2016), suggesting that despite introgressive hybridization, F1 males may have been unfit. These findings suggest the X and Y chromosomal regions may have been involved in hybrid incompatibilities during the time of hybridization and thus archaic ancestry was rapidly purged by selection acting on unfit hybrids, limiting introgression on sex chromosomes.

Experimental model systems have proven to be powerful to interrogate the mechanisms underlying the large role of the sex chromosomes in speciation. However, empirical investigations of this process using natural populations in taxa with XY sex chromosome systems have been limited in taxonomic breadth (Presgraves 2018) due to challenges associated with extensive sampling and limited genomic resources. Thus, the generality of this pattern within taxonomic groups (e.g., mammals, Primates) is not clear due to a lack of empirical investigation of gene introgression in natural systems, and whether a common genetic architecture (i.e., underlying genetic basis) of reproductive isolation is shared among systems. Thus, it is important to broaden investigations to determine the generality of this pattern and its underlying genetic basis, especially among natural systems. Natural hybrid zones offer unique opportunities to test predictions of the rules of speciation. When admixture occurs between divergent lineages, novel combinations of alleles in hybrids are tested by selection (Dobzhanksy 1936, Muller 1942). Recombination shuffles composite genomes with many generations of backcrossing, and loci associated with reproductive isolation are expected to have restricted introgression as a result of selection against unfit hybrids carrying incompatible alleles (Barton & Hewitt 1985, Gompert & Buerkle 2011).

Here, we investigate interspecific gene flow in a howler monkey hybrid zone (*Alouatta palliata* x *Alouatta pigra*) to test for reduced introgression of X-linked markers, a prediction of the large-X effect on postzygotic isolation. The parental species diverged ~3 MA (Cortés-Ortiz et

al. 2003), and the contact zone is likely the result of secondary contact after periods of isolation and expansion (Cortés-Ortiz et al. 2003, Ford 2006, Ellsworth & Hoelzer 2006). We previously analyzed introgression in this system with a limited number of loci and found differential introgression for autosomal markers (i.e., some had reduced introgression, others had neutral or directional introgression), but markers on the X and Y chromosomes had restricted introgression (Cortés-Ortiz et al. 2019). This observation is consistent with a role for the sex chromosomes in reproductive isolation in this system, but because we only used 3 X-linked markers we were not able to determine the extent of reduced introgression across the X chromosome or identify the candidate regions driving this reduced introgression.

Specifically, we identify and validate X chromosome sequence in the *A. palliata* genome assembly and use reduced-representation sequence data in genomic cline analysis to test for a large X-effect in this system. If the X chromosome is disproportionately involved in postzygotic reproductive isolation, we should observe reduced introgression of X-linked markers compared to autosomal markers. Specifically, we ask (1) whether introgression of X-linked SNPs is reduced compared to autosomal SNPs, and, if so (2) which X-linked regions underlie this pattern? We also use the validated X chromosome sequence to assess the degree of its sequence conservation in *A. palliata*. We discuss our results in the context reproductive isolation in other primate systems.

Materials and methods

Overview

We used the draft *A. palliata* genome assembly (PVKV00000000) to call SNPs from reduced-representation sequence data for our differential introgression analyses to test for

reduced introgression of the X chromosome (see below). Since the *A. palliata* genome assembly is a de novo draft assembly and contigs have yet to be assigned to chromosomes, we first performed a mapping experiment with whole-genome re-sequencing data to identify X-linked contigs. To identify X-linked sequence, we analyzed differences in contig-specific sequencing coverage between female and male *A. palliata* individuals with the assumption that X-linked contigs will show significantly greater coverage for females than for males because in XY species, females have two copies of the X while males only have one. This method has been used in several study systems to successfully identify sex chromosome sequence (e.g., Chen et al. 2012, Carvalho & Clark 2013, Gamble et al. 2015, Vicoso & Bachtrog 2015, Mongue et al. 2017, Bracewell et al. 2017). We also take advantage of the high degree of conservation of the mammalian X chromosome (Ohno 1967, Quilter et al. 2002, Raudsepp et al. 2004, Murphy et al. 2007, Delgado et al. 2009) and use sequence homology of our putative X-linked regions to primate X chromosomes. We likewise used quantitative PCR (qPCR), to validate X-linkage of *A. palliata* sequences. See Figure 2.1 for an overview of our strategy.

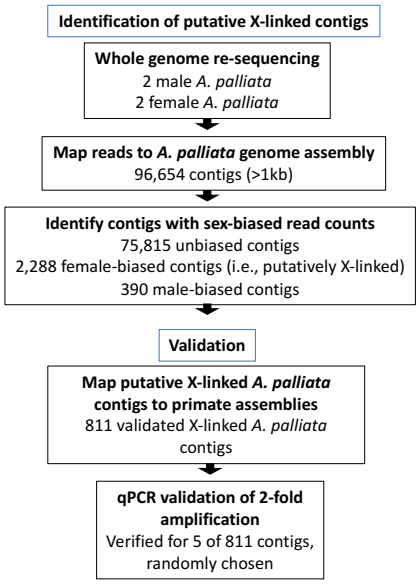

Figure 2.1. Overview of methods used in this study to identify and validate X-linkage for *A. palliata* assembly contigs.

Whole genome re-sequencing

Between 2001-2012, we obtained blood samples from four wild *A. palliata* individuals (two males, two females) sampled in Veracruz, Mexico (Table A2.1) and stored them in lysis buffer at -80ºC. We extracted genomic DNA using the QIAGEN DNeasy tissue kit (Qiagen Inc., Valencia, CA) as described in Baiz et al. (2019). Sex was determined in the field by visual assessment and verified using genetic data by amplifying known genes on the sex chromosomes (following Cortés-Ortiz et al. 2019). Specifically, we amplified the Y-linked *SRY* locus to verify the presence of a PCR product for males and absence of a PCR product for females, and genotyped an X-linked microsatellite locus to verify the hemizygous genotype for males (Table A2.1).

Our libraries for whole genome sequencing were generated and sequenced by the Sequencing Core at the University of Michigan. For each sample, libraries were constructed

using the Swift Biosciences PCR-Free DNA Library Kit with a target insert size of 350bp following the manufacturers protocol. Libraries were multiplexed and sequenced in a single lane using Illumina HiSeq 4000 to obtain 150bp paired-end reads.

We obtained between ~54 M–123 M reads per individual from the sequencer (Table A2.1). We used *Trim Galore!* v0.4.2 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to trim adapters and low quality bases (Q<20) from raw reads and retained only read-pairs where each read was ≥100bp in length after trimming.

Detection of X-linked contigs

To avoid false identification of X-linked contigs due to differences in sequencing effort among males and females, we subsampled our post-trimmed fastq files to standardize the number of reads across individuals before mapping using *seqtk* v1.2 (https://github.com/lh3/seqtk) by randomly sampling 50 million read pairs per individual. We used these subsampled files as input for sequence alignment.

Because short read data can potentially map to multiple genomic locations due to the expansion of repetitive sequence, we generated a repeat-masked version of the *A. palliata* genome assembly in order to map reads to unique sequences. To do this, we used RepeatMasker v4.0.6 (Smit et al. 2013) on *A. palliata* contigs 1 kb and larger (N=96,654, 87% of the assembly sequence) using the 'primates' RepeatMasker repeat library (i.e., a library of known primate repeats) to perform a low-sensitivity search (option –qq).

We then used bwa-MEM (Li 2013) to align the subsampled reads to the repeat-masked *A. palliata* genome. For each individual, we used samtools idxsats (Li et al. 2009) to count the

number of reads mapped to each masked contig. We then used exact tests in edgeR (Robinson et al. 2010) to detect contigs for which there was a significant difference in read counts between the sexes, where X-linked contigs are expected to have an average male-to-female $\log_2$ fold-change (logFC) of -1 and autosomal contigs are expected to have a logFC of 0. For this analysis, we excluded contigs with low counts across samples as they provide limited power to detect significant differences between groups. Thus, we only retained contigs where two or more individuals had a count per million (CPM) >0.2, corresponding to ~15 reads (N=78,493 contigs).

Validation of X-linkage

We aligned contigs for which we detected sex-differences in read count to the marmoset (CaJac3) and human (hg38) genomes to identify homologous regions. We used these results as a second line of evidence for X-linkage of female-biased contigs since X chromosome sequence is highly conserved across mammals. To do this, we downloaded the masked version of each assembly from the University of California Santa Cruz Genome Browser and used a custom script to remove scaffolds that have not yet been assigned to any chromosome (i.e., sequences with a header containing "chrUn"). We then used *lastz* v.1.04.00 (a program designed for efficient alignment of long genomic sequences, Harris 2007) to align the repeatmasked *A. palliata* contigs to each repeatmasked assembly, requiring at least 50% of the query to be included in the alignment block (--coverage=50) and using a distance of 20bp between potential seeds (--step=20). To assess the ability of our method to detect X-linked sequence, we also aligned a set of 2,288 unbiased, likely autosomal contigs for comparison (i.e., the same number of female-biased contigs), randomly chosen from the list of contigs that did not have a significant difference in read counts between the sexes.

For sex-biased and unbiased *A. palliata* contigs, a larger proportion aligned to the marmoset genome (>55% for each contig type) than to the human genome (Figure A2.1). For male-biased and unbiased *A. palliata* contigs, the proportion that aligned to autosomes (~94-99%) or sex chromosomes (~1-5%) was similar for the marmoset and human genome (Figure A2.1). Further, the majority (74%) of the female-biased contigs that aligned to the human X chromosome also aligned to the marmoset X chromosome. This pattern is not surprising given a greater divergence time between *Alouatta* and *Homo* than between *Alouatta* and *Callithrix* (Perelman et al. 2011). Thus, we considered female-biased contigs that aligned to the marmoset X chromosome to be X-linked for *Alouatta* in this study.

We used qPCR to test for the expected 2-fold amplification of putative X-linked regions in female individuals compared to male individuals. We randomly selected five putative X-linked contigs and one putative autosomal contig for comparison. The five putative X-linked contigs were selected from *A. palliata* female-biased contigs that mapped to the marmoset X chromosome. The putative autosomal contig was randomly selected from *A. palliata* unbiased contigs that mapped to a marmoset autosome. The putative X-linked contigs mapped to positions that spanned the length of the marmoset X (between X:6.6–X:104 Mb) and the putative autosomal contig mapped to marmoset chromosome 1 (Table A2.2). For qPCR validation, we designed a primer pair targeting each of the six contigs (Table A2.2) using Primer 3 v. 0.4.0 (Koressaar & Remm 2007, Untergasser et al. 2012) using the 'human' setting to avoid designing primers in repetitive sequence, target size between 100-200bp, and otherwise default settings. We first verified amplification using standard PCR in at least three *A. palliata* individuals using the following cycling conditions: initial denaturation at 95ºC for 3 min, followed by 34 cycles of 95 ºC for 30 sec, annealing temperature of 55ºC for 30 sec, 72ºC for 30 sec, followed by 72ºC for

5 min. Amplifications were carried out using a reaction volume of 25 µl, containing 0.63 µl each of forward and reverse primer (10 µM), 0.125 µl GoTaq and 5 µl 5X green GoTaq buffer (Promega), 5 µl dNTPs (2µM each), 17.2 µl water, and 1 µl DNA extract. We included one negative control per reaction to ensure no contamination of our PCR reagents. We visualized PCR products on a 2% agarose gel to ensure amplification of a single band of the expected size.

For each primer pair, we then prepared 20 µl reactions for qPCR using 0.6 µl each of forward and reverse primer (10 µM), 10 µl Power SYBR Green PCR Master Mix (Applied Biosystems), 6.6 µl water, and 2.2 µl genomic DNA (~44 ng). For each primer set per run, we included three technical replicate amplifications each for one male and one female *A. palliata* individual. We also included one no-template DNA negative control per run using a single primer pair to ensure no contamination of PCR reagents. Amplifications were run on an ABI 7500 Fast Real-Time PCR machine with the following cycling conditions: 50ºC for 2 min, 95ºC for 10 min, followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min. We calculated relative fold-change in template DNA between the sexes using the $2^{-\Delta\Delta C}{}_{T}$ method (Livak & Schmittgen 2001), where the gene-of-interest was the presumed X-linked marker, the normalizing gene was the autosomal marker, and female was the 'experimental' condition while male was the 'control' condition.

Overview of introgression analysis

We performed reduced-representation sequencing on a geographically broad sample of individuals from allopatric ranges of *A. palliata* and *A. pigra* and from the hybrid zone and mapped sequence data to the non-masked *A. palliata* assembly to generate genotype data for genomic cline analysis.

ddRADseq and SNP calling

We extracted DNA from whole blood samples from individuals captured from the wild (see Baiz et al. 2019 for details). Because the X chromosome is hemizygous in males and X-linked SNPs will appear to be homozygous, biasing genomic cline estimates, we only included sequence data from female individuals in genomic cline analyses to avoid bias in our X-autosome comparison of differential introgression. Thus, our analysis included 88 female individuals, 48 of which were sampled from the hybrid zone in Tabasco, Mexico, 17 from the allopatric range of *A. palliata* and 23 from the allopatric range of *A. pigra* (Figure 2.2). All allopatric individuals included here have been previously shown to be non-admixed (Baiz et al. 2019). We used double digest restriction site associated DNA sequencing (ddRADseq, Peterson et al. 2012) to generate reduced-representation genome sequence data for these individuals, as described in Baiz et al. (2019).

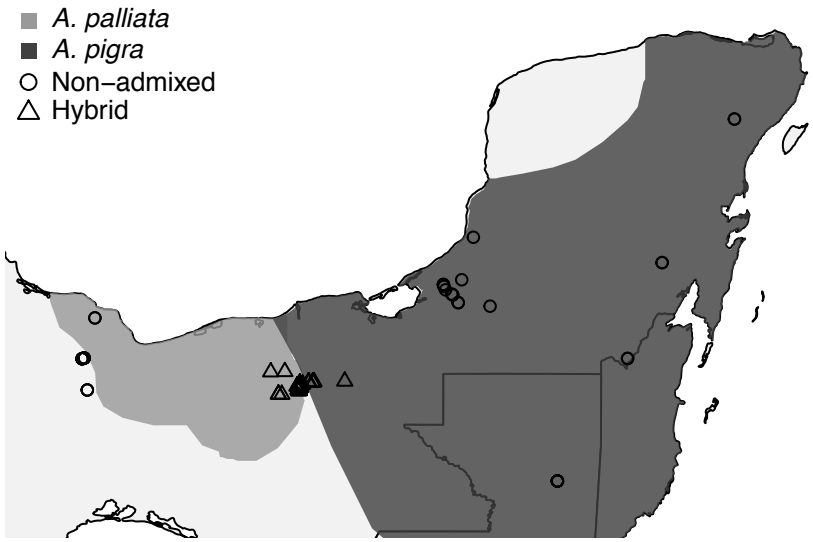

Figure 2.2. Map of sampling sites used in this study. The allopatric range of *A. palliata* is in light gray and the allopatric range of *A. pigra* is in dark gray. Non-admixed individuals are represented with circles, and individuals sampled from the hybrid zone are represented with triangles.

We retained only biallelic SNPs with a minor allele frequency of at least 0.05, a minimum mean read depth of 12 across all individuals, and sites present in 14 or more individuals in either parental population. To reduce the effects of linkage among markers, we retained only one SNP at random per contig. This resulted in 10,353 SNPs used in the genomic cline analysis. The combined length of these *Alouatta* contigs with SNPs used in our analysis represents ~39% of the assembly sequence. We considered X-linked SNPs to be those on female-biased contigs that mapped to the marmoset X chromosome (N_{SNPs}=97) and autosomal SNPs to be those on contigs that had no significant difference in read counts between the sexes (N_{SNPs}=10,256). The set of X-linked and autosomal SNPs represent approximately equal genotyping densities on female-biased (1.9×10^{-5} SNPs/Mb) and unbiased contigs (8.4×10^{-6} SNPs/Mb). All filtering steps were carried out using *bcftools* v.1.3.1, *vcftools* 0.1.14 (Danecek et al 2011), and custom scripts.

Genomic cline analysis

To analyze the pattern of introgression for X-linked and autosomal SNPs, we calculated genomic clines for each locus using *bgc* (Gompert & Buerkle 2011, Gompert & Buerkle 2012), as described in Baiz et al. (2019). This analysis is used to identify loci that are more or less likely than the genome-wide average, which is assumed to be neutral, to introgress between parental populations. Two cline parameters are used to summarize the amount (β) and direction (α) of introgression. Loci associated with reproductive isolation are expected to have reduced introgression (β>0), while loci with increased introgression (β<0) may be candidates for adaptive introgression. Loci with a shift in cline center reflect directional movement of alleles into *A.*

palliata (excess *A. pigra* ancestry, $\alpha>0$) or movement into *A. pigra* (excess *A. palliata* ancestry, $\alpha<0$).

We ran *bgc* analyses using the genotype uncertainty model and ran five independent chains, each with a burn-in of 30,000 for 50,000 steps, and thinned samples by 20. We then merged outputs and identified outlier SNPs with respect to both β and α from MCMC output as SNPs with a 95% credible interval that does not overlap zero.

X-autosomal comparison of introgression

To test if X-linked SNPs have a distinct pattern of introgression, we tested for significant differences in cline parameters between X-linked and autosomal SNPs using permutation tests in R. We constructed 10,000 permuted datasets from the autosomal data by sampling without replacement from the distribution of cline parameter point estimates for both α and β. For each permuted dataset, we sampled 97 autosomal SNPs, ensuring comparisons were made using a sample size equal to the set of X-linked SNPs (N=97). We compared the mean of the observed cline parameter for X-linked SNPs to the mean cline parameter of each permuted autosomal dataset and considered the pattern of introgression for X-linked SNPs to be distinct if the observed mean exceeded the mean in >95% of the permuted datasets.

Genomic basis of non-neutral introgression of the X chromosome

To identify genes on contigs containing SNPs with non-neutral introgression, we queried the marmoset X chromosome for genes using *biomaRt* v2.36.1 (Durnick et al. 2005, 2009). To do this, we input the marmoset alignment block coordinates for each X-linked *bgc* outlier contig expanded by 500kb on both ends to obtain marmoset genes within each region. We also report

human gene homologs within each region, as the human gene annotation is more complete than marmoset.

To determine if the previously observed "deserts" of archaic hominin ancestry in the human genome are homologous to the regions of reduced introgression we observed here, we plotted cline parameter estimates along the human X chromosome for X-linked contigs in our *bgc* dataset that mapped to the human X using our alignment criteria (N=52) with *karyoPloteR* v1.6.2 (Gel & Serra 2017).

Results

Identification of X-linked contigs in A. palliata

Upon mapping our re-sequencing data to the masked *A. palliata* assembly, we found that read counts for most contigs (97%) were not significantly different between males and females (i.e., we refer to these as "unbiased" contigs), suggesting they are autosomal (Table 2.1). Thus, we used this set of contigs to call autosomal SNPs for our genomic cline analyses. We detected 2,288 contigs with read counts greater in females than in males (i.e., "female-biased" contigs), with an average logFC of ~1, as expected for X-linked contigs. We also detected 390 contigs with higher read counts in males than in females (i.e., "male-biased" contigs), which was unexpected since reads were mapped to a reference genome generated from a female individual (Broad Institute, personal communication). LogFC was much more variable for these male-biased contigs, many of which went beyond the expected two-fold difference in read count between the sexes (Table 2.1, Figure 2.3).

Table 2.1. Summary of mapping experiments to identify X-linked contigs in the *A. palliata* assembly. N contigs=number of contigs detected to be biased or unbiased in edgeR, Mean logFC=mean log2-fold-change of read counts for male data relative to female data.

Read count bias	**N contigs (%)**	**Mean logFC (M:F) ± SD**	**N contigs mapped to marmoset (%)**
Unbiased	75,815 (97%)	-0.003 ± 0.30	1,392† (60.1%)
Female	2,288 (2.9%)	-0.992 ± 0.38	1,077 (47.1%)
Male	390 (0.5%)	2.129 ± 1.53	179 (45.9%)

†Note that the number of contigs mapped to marmoset for unbiased contigs was 2,288, randomly chosen to match the sample size of female-biased contigs.

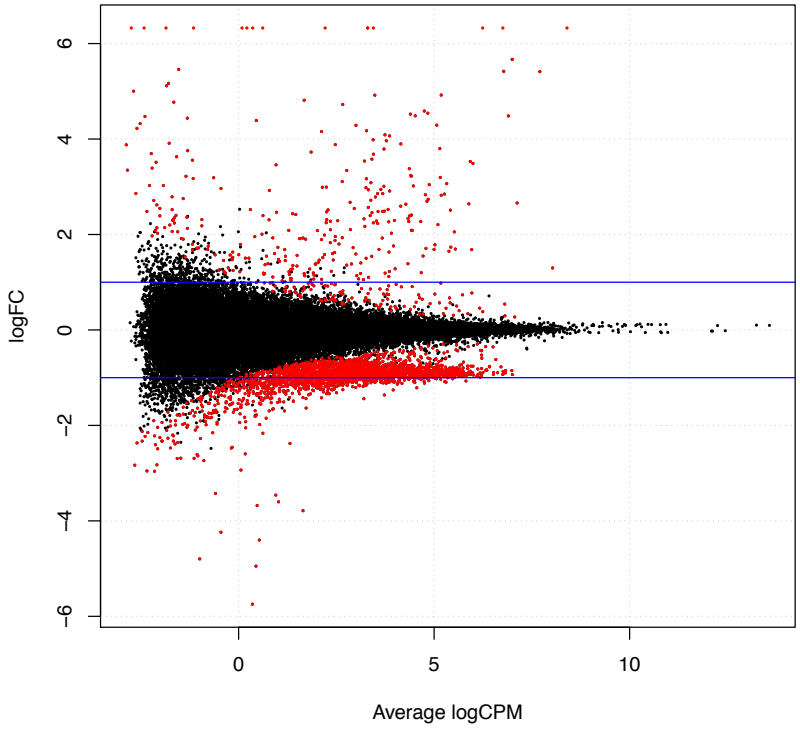

Figure 2.3. Summary of sex differences in read mapping count for *A. palliata* genome assembly contigs (N=96,654). Contigs in black show no significant difference in read count between the sexes and are likely autosomal, while contigs in red show greater read counts for females (logFC<0) or males (logFC>0). Blue horizontal lines indicate a 2-fold difference in read count between the sexes. LogFC is log2-fold-change and Average logCPM is log2-counts-per-million, a measure of the number of reads mapped averaged across samples.

Comparative sequence analysis and qPCR validation of X-linkage

Of the set of 2,288 *A. palliata* female-biased contigs, 1,077 could be mapped to the marmoset genome using our mapping criteria (Table 2.1). Of these contigs, 811 (75.3%) mapped to the marmoset X chromosome, 277 (25.7%) mapped to marmoset autosomes, and one mapped to the marmoset Y chromosome. This enrichment of hits to the marmoset X is consistent with the results of our comparative read count analysis (see above), indicating that these regions are likely to be X-linked in *A. palliata*. Comparatively, a much smaller proportion of the unbiased contigs (3.4%) and male-biased contigs (1.1%) mapped to the marmoset X while the majority mapped to autosomes (Figure A2.1). Further, logFC for the subset of the 811 female-biased contigs that mapped to the marmoset X was less variable and closer to the expected -1 for X-linked sequences compared to the larger pool of 2,288 female-biased contigs (Figure A2.2). Thus, this set of 811 female-biased contigs that mapped to the marmoset X constitutes our set of validated X-linked regions used in further analyses.

As proof of concept, we randomly chose five of the 811 female-biased contigs that mapped to the marmoset X chromosome to confirm two-fold amplification in females relative to males (as expected for X chromosome sequence) using qPCR. Consistent with this, mean fold-change was 2.19±0.49 (Table 2.2). These observations are consistent with high conservation of the X chromosome in primates and further corroborate our method of identifying X-linked versus autosomal sequence.

Table 2.2. qPCR validation of five *A. palliata* X-linked contigs. ΔΔCt is relative quantification of template DNA for each female-biased contig (gene-of-interest) compared to an unbiased (i.e., autosomal) marker (normalizing gene).

Contig	ΔΔCt	Fold-change
26402	0.92	1.89
35197	1.55	2.94
92787	1.14	2.20
118733	1.17	2.25
60023	0.73	1.65

Because a portion of our female-biased contigs mapped to autosomes, it is possible that we detected X chromosome sequence in *A. palliata* that is not shared with other primates (i.e., lineage-specific translocations to the X chromosome). To explore this, we looked at the mapping positions of the 277 female-biased contigs that mapped to marmoset autosomes and compared them to the mapping positions of male-biased and unbiased contigs for the autosomes with the most hits (Figure A2.3A). After the X chromosome, chromosome 7 had the highest number of hits for female-biased contigs. However, chromosome 7 also had the highest number of hits for both male-biased and unbiased contigs, and for all contig types, the hits were clustered around positions 29.6 MB and 74.3 MB (Figure A2.3B). Similarly, chromosome 21 had the third highest number of hits for female-biased contigs, high mapping numbers for male-biased and unbiased contigs, and a clustering of mapping positions around 17.7 MB (Figure A2.3C). These mapping positions are not unique to sex-biased contigs, which may be due to multiple factors including misassembly. This may also indicate that these sequences are not unique to the sex chromosomes. Thus, we dropped female-biased contigs that map to marmoset autosomes and male-biased contigs from further analyses to account for this uncertainty.

Distinct introgression of X-linked SNPs

Genomic cline parameters for most SNPs were consistent with neutral introgression, but we detected several outliers (Table 2.3). Among outlier autosomal SNPs, the majority had increased introgression ($\beta<0$) and excess ancestry from *A. pigra* ($\alpha>0$). On the other hand, outlier X-linked SNPs had reduced introgression ($\beta>0$) and excess ancestry from *A. palliata* ($\alpha<0$) (Figure 2.4).

Table 2.3. Number of X-linked (type=X) and autosomal (type=A) SNPs with neutral (zero) and extreme introgression (outliers). The cline parameter β is a measure of the amount of introgression, where negative outliers have increased introgression and positive outliers have reduced introgression. The cline parameter α measures the direction of introgression where negative outliers have excess *A. palliata* ancestry and positive outliers have excess *A. pigra* ancestry.

Cline parameter	**Type**	**Negative outlier**	**Zero (neutral)**	**Positive outlier**
β	X	0	95	2
	A	194	10,047	15
	Total	194	10,142	17
α	X	3	94	0
	A	273	9,513	470
	Total	276	9,607	470

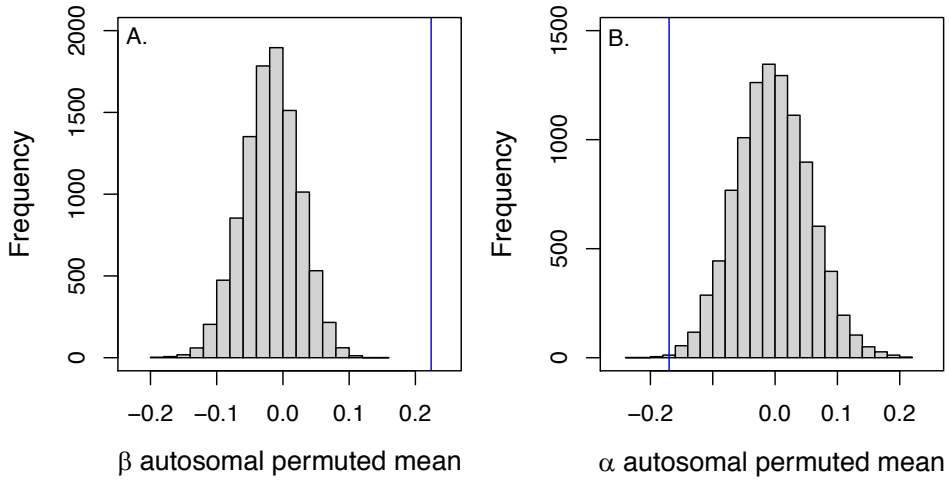

Figure 2.4. Histogram of means of 10,000 permuted autosomal SNP datasets (gray bars) for A) the amount of introgression (β) and B) the direction of introgression (α). In each case, the vertical blue line is the observed mean for X-linked SNPs, which is more extreme than the mean of the permuted data set in >95% samples indicating X-linked SNPs have a distinct pattern of introgression with respect to both cline parameters. Reduced introgression is indicated by $\beta>0$ and increased introgression by $\beta<0$. Excess *A. pigra* ancestry is indicated by $\alpha>0$ and excess *A. palliata* ancestry by $\alpha<0$.

Our permutation tests indicated that cline parameters are more extreme for X-linked than for autosomal SNPs, suggesting a distinct pattern of introgression for the X chromosome in this system. For X-linked SNPs, the amount of introgression was significantly reduced compared to autosomal SNPs (mean β_X=0.22, mean β_A=-0.02, P<0.001) and the direction of introgression was more negative (mean α_X=-0.17, mean α_A=-0.003, P<0.001), indicating excess ancestry from *A. palliata* (Figure 2.4), consistent with the signal for outlier loci.

Genomic basis of the large X-effect

After adding 500kb to each end of the alignment block within the marmoset X chromosome for alignments of contigs with outlier SNPs, two regions overlapped in the marmoset assembly (X: 46475367: 47494965, X:47487523:48488951). Thus, we report this as a

single region (region 1), which had the greatest gene content in comparison to the other X chromosomal regions containing outlier loci (Table 2.4). The two contigs containing SNPs with reduced introgression mapped to the distal end of the long arm of the marmoset X, and the contigs containing SNPs with excess *A. palliata* ancestry mapped more proximally, two to the short arm and one to the long arm of the marmoset X (Table 2.4).

Table 2.4. Alignment positions to the marmoset genome and gene content of X-linked *A. palliata* contigs containing SNPs with non-neutral introgression. CalJac3=the coordinates of the biomaRt query which includes an extension of 500kb on each end of the alignment block, N genes=the number of genes within each region, and Outlier type=*bgc* cline parameter, where α is direction and β is amount of introgression.

Contig	Length (kb)	CalJac3	Region	N genes	Outlier type
49400	23.4	X:46475367:47494965	1	50	α<0
151667	1.4	X:47487523:48488951			α<0
30014	54.1	X:67467391:68496613	2	4	α<0
32694	48.1	X:113940392:114968526	3	3	β>0
54333	18.8	X:135586677:136602592	4	3	β>0

Of the 97 X-linked contigs represented in our SNP dataset, 53 mapped to the human X chromosome, which included three of the five X-linked contigs containing SNPs with non-neutral introgression (Table A2.3). Of the two contigs containing SNPs with reduced introgression, one (region 3) mapped to a position within one of the previously described human "deserts" for ancestry from both Neanderthals and Denisovans (Sankararaman et al. 2016) (Figure 2.5). The other contig with reduced introgression (region 4) mapped just distally, but outside of the same desert. Finally, a contig with excess *A. palliata* ancestry mapped to the proximal end of the short arm of the human X. These results are consistent with our mapping analysis using the marmoset X chromosome (Table 2.4), and with a high degree of conservation in X chromosome sequence among primates.

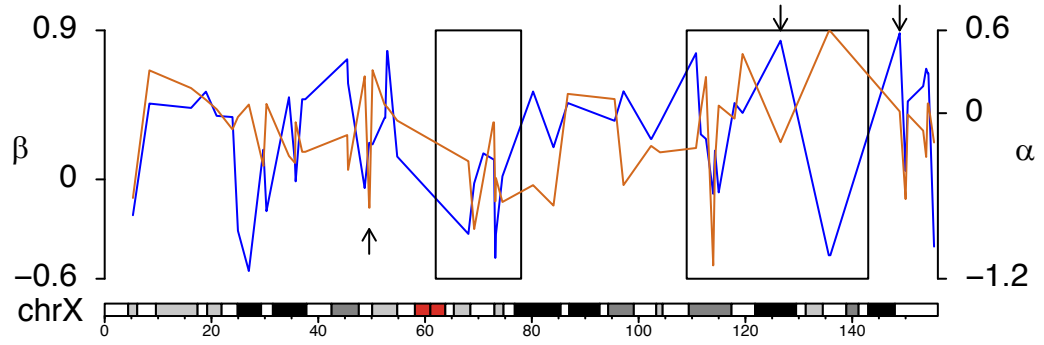

Figure 2.5. Cline parameter estimates for SNPs within *Alouatta* contigs that mapped to the human X chromosome. The direction of introgression is measured by α (orange, right axis) and the amount of introgression is measured by β (blue, left axis). The two previously described "deserts" of archaic ancestry (Sankararaman et al. 2016) are enclosed in boxes and mapping positions of contigs with outlier loci are shown with arrows. Shaded regions along the human X chromosome are cytobands and the centromere is shown in red.

Discussion

We identified X-linked contigs in the draft *A. palliata* genome assembly by analyzing differences in coverage between males and females using whole genome re-sequencing data and used homology to the marmoset X chromosome and qPCR to validate our results. We then performed genomic cline analysis with reduced-representation sequencing data and compared the pattern of introgression for SNPs on these contigs to SNPs on putatively autosomal contigs. We found that X-linked SNPs have a pattern of introgression that is distinct from introgression at autosomal loci. Cline parameter estimates revealed that compared to autosomal SNPs, X-linked SNPs have reduced introgression across hybrid genomes. This pattern is consistent with a large-X effect on postzygotic isolation. X-linked SNPs also had an excess of ancestry from *A. palliata*, consistent with limited, but directional introgression of X-linked regions. Using sequence homology to the marmoset and human X chromosomes, we identified candidate sequence on the *A. palliata* X chromosome that may underlie X-linked postzygotic isolation and directional introgression.

Discovery of X chromosome sequence in A. palliata

Using whole genome-resequencing data, we identified 811 regions in the *A. palliata* assembly to be X-linked based on greater mapped read counts in females compared to males and alignment to the marmoset genome. Consistent with this, we observed approximately two-fold amplification of a subset of these regions in females as compared to males using qPCR. This is the first study to identify extensive, contiguous sex chromosome sequence in *A. palliata.* However, considering the combined length of these regions is 26.2 Mb, they likely represent <20% of the *A. palliata* X (assuming the size of the *A. palliata* X is similar to the human and marmoset X, which are ~156 Mb and ~142 Mb, respectively). Although, we expect that our dataset includes an underrepresentation of the X chromosome since we limited our analysis to non-repetitive sequence (by repeatmasking the genome). Still, considering only the non-masked proportion of the human and marmoset X chromosomal sequence, the *A. palliata* regions we identified to be X-linked here represent <50% of the expected size of the non-masked X chromosome. Further, the pseudoautosomal region of the X chromosome (3 Mb in humans) would not be detected using our methods since it is not hemizygous in males. The density of SNPs analyzed in this study was similar for X-linked and autosomal contigs (see results) so this should not bias our results. Regardless, our ability to detect X-linked sequence was limited by several factors, including the contiguity of the *A. palliata* reference assembly. We only used contigs greater than 1kb in size in our analyses to avoid downstream complications from working with difficult to assemble sequence with likely high repeat content. We also discarded contigs from our analyses with low mapped read counts across samples due to low power to detect sex differences. Thus, our set of X-linked contigs should not be treated as definitive.

Because a large proportion of our X-linked contigs mapped to both the marmoset and human X chromosomes in similar positions, our results are consistent with a high degree of sequence and gene order conservation in primates following high conservation of the eutherian mammal X (Quilter et al. 2002, Raudsepp et al. 2004, Murphy et al. 2007, Delgado et al. 2009). This result is also consistent with cytomolecular studies that identified a high degree of similarity between the human and *A. palliata* X using chromosome painting probes (Steinberg et al. 2014). This is in contrast, however, to the high frequency of chromosomal rearrangements that have occurred in New World primates (Wienberg & Stanyon 1998, Müller 2006, de Oliviera et al. 2012). This suggests that despite the propensity for chromosomal rearrangements within *Alouatta*, including an autosome-Y translocation in *A. palliata* (Solari & Rahn 2005, Steinberg et al. 2014), selection for conservation of the X chromosome remains strong.

Some of the female-biased contigs we identified in *A. palliata* mapped to marmoset autosomes (~26%). This pattern would be expected for autosomal regions that have been translocated from autosomes to the X chromosome in *Alouatta*. However, mapping positions for these contigs were shared and also common among male-biased and unbiased contigs. Thus, it may be more likely that these regions are not unique to the X chromosome and represent false positives in our mapping experiment. Investigation of what caused this putatively false positive pattern and identification of any autosome-to-sex chromosome translocated regions are beyond the scope of this study and remain avenues for further research. For example, it may be possible to detect the breakpoint for the autosome-Y translocation and male-specific sequence using an approach similar to our approach here pending the development of a male *A. palliata* genome assembly.

Distinct pattern of introgression for the X chromosome

We identified several SNPs with non-neutral introgression (Table 2.3). Overall, these results are similar to our previous analyses using the same dataset with different genotyping parameters (Baiz et al. 2019). However, in this work, we detected fewer outlier SNPs which is surprising since we used more loci in the analysis. We suspect that this is due to reduced power associated with including fewer individuals in our analyses (i.e., since we used only females). Given the clear signal in the difference of cline parameter estimates between autosomal and X-linked SNPs (Figure 2.4), we do not believe this has limited our ability to test for a large X-effect.

We found that, compared to autosomal SNPs, X-linked SNPs had reduced introgression (Figure 2.4). These results are consistent with our previous analyses on differential introgression in this system (Cortés-Ortiz et al. 2019). Because our previous analysis used both males and females, (likely overinflating cline parameters for X-linked markers since males are hemizygous for the X chromosome), and only three X-linked markers, this study provides more rigorous evidence for a large X-effect in this system. This is consistent with anecdotal observations indicating that hybrid males with intermediate admixture proportions (i.e., hybrid index ~0.5) may be sterile (LCO, personal observation). We sampled a group in the hybrid zone multiple times containing a male with intermediate admixture proportions (Q=0.46, Baiz et al. 2019), and even though he was the only reproductively mature male in the group for 7 years, no offspring were observed to be sired. Although this male is not an F1 individual (he carries the *A. pigra* haplotype for both mtDNA and *SRY*), he may carry combinations of incompatible alleles that hinder the production of sperm capable of fertilization. Future studies should connect the signature of the large X-effect we observed here to phenotypic evidence for postzygotic reproductive isolation in this system.

Compared to autosomal SNPs, X-linked SNPs also have excess ancestry from *A. palliata*. This result is consistent with our previous observations of limited, but directional introgression for some X-linked markers resulting in the acquisition of *A. palliata* alleles for some *A. pigra*-like hybrids (Cortés-Ortiz et al. 2019, CHAPTER I). Together, these results indicate that for the X chromosome, *A. palliata* alleles may be more favorable than *A. pigra* alleles in the hybrid zone when they do pass the species boundary. Outlier region 1 is particularly gene-rich (Table A2.3) and contains genes with varied functions, including functions related to the immune system (e.g., FOXP3, WAS, CFP), neuron function, (e.g., ELK1, SYN1, SYP), and gene regulation (FOXP3, SSX1/SSX4B, UXT).

Because we used reduced-representation data, our ability to pinpoint regions driving the patterns we observe is limited. Given that our genotype data represents a small portion of the genome, it is likely these regions were not sequenced in our library and we are detecting effects of linkage to nearby genes under selection for postzygotic isolation or directional introgression. Future studies using whole genome sequence data that represents the full scope of variation in these species will be needed to pinpoint specific regions underlying non-neutral introgression. It will also be important to link directional introgression of X-linked markers to the introgression of *A. palliata* phenotypes in the hybrid zone, a prediction that follows from our results. Previous studies using many of the same individuals sequenced here found that the parental species are morphologically distinct for several phenotypic characters, and genetically intermediate hybrids are highly variable in their morphology (Kelaita et al. 2011, 2013). Thus, it may be possible to link key phenotypes (e.g., testis volume, pelage coloration) to the X (and/or autosomes) using association mapping (e.g., Turner & Harr 2014).

Genetic basis of large-X effect

Regions with reduced introgression mapped to the long arm of the marmoset X in areas with a relatively low-density of genes. If these regions are involved in postzygotic reproductive isolation, particularly hybrid male sterility, a prediction would be that genes in these regions are enriched for elevated expression in the testis as observed for genes with reduced archaic ancestry in humans (Sankararaman et al. 2014, 2016). Our analyses indicate that region 3 includes three protein-coding genes (Table A2.3), two of which are likely paralogs of the DDB1 and Cul4 associated factor 12-like genes DCAF12L1 and/or DCAF12L2. In humans, both of these genes have elevated expression in the testis. The other gene, PRR32, has low absolute gene expression overall, and does not have elevated expression in any tissue. Region 4 includes three protein-coding genes (Table A2.3), two of which are associated with human homolog AFF2, and one that is unnamed. In humans, AFF2 is expressed at low levels in the testis, but has elevated expression in the cerebellum. To our knowledge, none of these genes has a direct role in mammalian sperm production or function and it is unknown how mutations in these genes affect postzygotic isolation in *Alouatta*. Again, it is possible that the regions driving reduced introgression of the X chromosome were not sequenced in our ddRAD library and we only detected linked regions in our queries.

It is interesting, however, that one of these regions (region 3) falls within a known "desert" of both Neanderthal and Denisovan ancestry on the human X (Sankararaman et al. 2016), while the other (region 4) maps just distally (Figure 2.5). Because this desert spans a large section of the human X (34 Mb) and the contig for region 3 mapped in the central portion of that window, it is possible this window includes the true region of interest. This may indicate that this region underlies postzygotic reproductive isolation in both systems and thus may be important to the

genetic architecture of speciation in primates. To address this question, it will be highly informative to compare these results to those from other primate systems. To our knowledge, the *Aloautta* system is the only non-human primate system that has been used specifically to identify genomic barrier regions. However, there are many known primate hybrid zone systems that could be used to similar ends, some of which have genetic data that have previously been generated (e.g., baboons: Wall et al. 2016, chimpanzees: de Manuel et al. 2016, South American howlers: Mourthe et al. 2018, marmosets: Malukiewicz et al. 2015). A recent study detected historical introgressive hybridization between bonobos and chimpanzees and found that gene exchange was restricted on the X chromosome (de Manuel et al. 2016). However, the authors did not report whether any X-linked regions were more or less resistant to introgression. Future studies on the genetics of hybridization and speciation in primates should report such detail so results can be compared across studies to test the hypothesis that a general genetic architecture of reproductive isolation that drives speciation is shared across primate taxa.

References

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics,* **16**, 113–148.

Bracewell, R. R., Bentz, B. J., Sullivan, B. T., & Good, J. M. (2017). Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. *Nature Communications*, **8**, 1593.

Carling, M. D., & Brumfield, R. T. (2008). Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. *Evolution: International Journal of Organic Evolution*, **62**, 2600–2615.

Carvalho A. B., Clark A. G. 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome research* **23** 1894–1907.

Chen N., Bellott D. W., Page D. C., Clark A. G. 2012. Identification of avian W-linked contigs by short-read sequencing. *BMC genomics* **13,** 183.

Cortés-Ortiz, L., Bermingham, E., Rico, C., Rodrıguez-Luna, E., Sampaio, I., & Ruiz-Garcıa, M. (2003). Molecular systematics and biogeography of the Neotropical monkey genus, *Alouatta*. *Molecular phylogenetics and evolution*, **26**, 64–81.

Cortés-Ortiz, L., Nidiffer, M. D., Hermida-Lagunes, J., García-Orduña, F., Rangel-Negrín, A., Kitchen, D. M., ... & Canales-Espinosa, D. (2019). Reduced Introgression of Sex Chromosome Markers in the Mexican Howler Monkey (*Alouatta palliata* × *A. pigra*) Hybrid Zone. *International Journal of Primatology*, **40***,* 114–131.

Coyne J. A., Orr H. A. (1989) Two rules of speciation. In: Otte D, Endler J, editors. Speciation and its consequences. Sunderland (Massachusetts): Sinauer Associates. pp. 180–207.

Coyne J. A., Orr H. A. (2004) Speciation. Sunderland (Massachusetts): Sinauer Associates. 545p.

Coyne, J. A. (2018). "Two rules of speciation" revisited. *Molecular ecology*, **27**, 3749–3752.

Delgado, C. L. R., Waters, P. D., Gilbert, C., Robinson, T. J., & Graves, J. A. M. (2009). Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Research*, **17**, 917–926.

de Oliveira, E. H. C., Neusser, M., & Müller, S. (2012). Chromosome evolution in new world monkeys (Platyrrhini). *Cytogenetic and genome research*, **137**, 259–272.

de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., ... & Schmidt, J. M. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, **354**, 477–481.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dobzhansky, T. 1936. Studies on hybrid sterility II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21**, 113–135.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W. (2005). "BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis." *Bioinformatics*, **21**, 3439–3440.

Durinck, S., Spellman, P., Birney, E., Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **4**, 1184–1191.

Ellsworth, J. A., & Hoelzer, G. A. (2006). Genetic evidence on the historical biogeography of Central American howler monkeys. In *Primate biogeography* (pp. 81-103). Springer, Boston, MA.

Ford, S. M. (2006). The biogeographic history of Mesoamerican primates. In *New perspectives in the study of Mesoamerican primates* (pp. 81-114). Springer, Boston, MA.

Gamble, T., Coryell, J., Ezaz, T., Lynch, J., Scantlebury, D. P., Zarkower D. (2015). Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Molecular Biology and Evolution* **32**, 1296–1309.

Gel, B., Serra, E. (2017). karyoploteR: an R / Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, **33**, 3088–3090.

Gompert, Z., & Buerkle, C. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Good, J.M., Handel, M.A., Nachman M.W. (2008a). Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution*, **62**, 50–65.

Good, J.M., Dean, M.D., Nachman MW (2008b). A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics*, **179**, 2213–2228.

Good, J.M., Giger, T., Dean, M.D., Nachman, M.W. (2010). Widespread over-expression of the X chromosome in sterile F1 hybrid mice. *PLoS Genetics*, **6**, e1001148.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Hansen, N. F. (2010). A draft sequence of the Neandertal genome. *Science*, **328**, 710–722.

Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., ... & Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus. *Molecular Ecology*, **21**, 3032–3047.

Kelaita, M. A., & Cortés-Ortiz, L. (2013). Morphological variation of genetically confirmed *Alouatta Pigra*× *A. palliata* hybrids from a natural hybrid zone in Tabasco, Mexico. *American journal of physical anthropology*, **150**, 223–234.

Kelaita, M., Dias, P. A. D., Aguilar-Cucurachi, M. D. S., Canales-Espinosa, D., & Cortés-Ortiz, L. (2011). Impact of intrasexual selection on sexual dimorphism and testes size in the Mexican

howler monkeys *Alouatta palliata* and *A. pigra*. *American journal of physical anthropology*, **146**, 179–187.

Kitano, J., & Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental biology of fishes*, **94**, 549–558.

Koressaar, T., Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics,* **23**, 1289–91.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv,* 1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods*, **25**, 402–408.

Malukiewicz, J., Boere, V., Fuzessy, L. F., Grativol, A. D., e Silva, I. D. O., Pereira, L. C., ... & Stone, A. C. (2015). Natural and anthropogenic hybridization in two species of eastern Brazilian marmosets (Callithrix jacchus and C. penicillata). *PloS one*, **10**, e0127268.

Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS biology*, **5**, e243.

Mendez, F. L., Poznik, G. D., Castellano, S., & Bustamante, C. D. (2016). The divergence of Neandertal and modern human Y chromosomes. *The American Journal of Human Genetics*, **98**, 728–734.

Mongue, A. J., Nguyen, P., Voleníková, A., Walters, J. R. 2017. Neo-sex Chromosomes in the Monarch Butterfly, *Danaus plexippus*. *G3: Genes, Genomes, Genetics*, **7**, 3281–3294.

Mourthe, I., Trindade, R. A., Aguiar, L. M., Trigo, T. C., Bicca-Marques, J. C., & Bonatto, S. L. (2018). Hybridization Between Neotropical Primates with Contrasting Sexual Dichromatism. *International Journal of Primatology*, **40**, 99–113.

Muller, H. J. (1942). Isolating mechanisms, evolution and temperature. *Biological Symposia* **6**, 71–125.

Müller, S. Primate chromosome evolution, in Lupski JR, Stankiewicz P (eds): Genomic Disorders: The Genomic Basis of Disease, pp 111–152 (Humana Press, Totowa 2006).

Murphy, W. J., Davis, B., David, V. A., Agarwala, R., Schäffer, A. A., Wilkerson, A. J. P., ... & Menotti-Raymond, M. (2007). A 1.5-Mb-resolution radiation hybrid map of the cat genome and comparative analysis with the canine and human genomes. *Genomics*, **89**, 189–196.

Ohno, S. Sex Chromosomes and Sex-linked Genes. Springer; Berlin: 1967.

Orr, H. A. (1993). Haldane's rule has multiple genetic causes. *Nature*, **361**, 532–533.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., ... & Schneider, M. P. C. (2011). A molecular phylogeny of living primates. *PLoS genetics*, **7**, e1001342.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, **7**, e37135.

Presgraves, D. C. (2008). Sex chromosomes and speciation in Drosophila. *Trends in Genetics*, **24**, 336–343.

Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. *Molecular Ecology*, **27**, 3822–3830.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., ... & Li, H. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.

Quilter, C. R., Blott, S. C., Mileham, A. J., Affara, N. A., Sargent, C. A., & Griffin, D. K. (2002). A mapping and evolutionary study of porcine sex chromosome gene. *Mammalian Genome*, **13**, 588–594.

Raudsepp, T., Lee, E. J., Kata, S. R., Brinkmeyer, C., Mickelson, J. R., Skow, L. C., ... & Chowdhary, B. P. (2004). Exceptional conservation of horse–human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. *Proceedings of the National Academy of Sciences*, **101**, 2386–2391.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., ... & Maricic, T. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, **468**, 1053–1060.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., ... & Saitou, N. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics*, **89**, 516–528.

Robinson, M. D., McCarthy, D. J., Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* **26**, 139–140.

Sætre, G. P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C., & Syvänen, A. C. (2003). Sex chromosome evolution and speciation in Ficedula flycatchers. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 53–59.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., ... & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507**, 354–357.

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology*, **26**, 1241–1247.

Smit, A. F. A, Hubley, R. & Green, P. *RepeatMasker Open-4.0*. 2013–2015 <http://www.repeatmasker.org>.

Solari, A. J., & Rahn, M. I. (2005). Fine structure and meiotic behaviour of the male multiple sex chromosomes in the genus *Alouatta*. *Cytogenetic and genome research*, **108**, 262–267.

Steinberg, E. R., Nieves, M., & Mudry, M. D. (2014). Multiple sex chromosome systems in howler monkeys (Platyrrhini, *Alouatta*). *Comparative cytogenetics*, **8**, 43–69.

Turner, L. M., & Harr, B. (2014). Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife*, **3**, e02504.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Research,* **40**, e115

Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*ogy **13**, p.e1002078.

Wall, J. D., Schlebusch, S. A., Alberts, S. C., Cox, L. A., Snyder-Mackler, N., Nevonen, K. A., ... & Tung, J. (2016). Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. *Molecular ecology*, **25**, 3469–3483.

Wienberg, J., & Stanyon, R. (1998). Comparative chromosome painting of primate genomes. *ILAR journal*, **39**, 77–91.

APPENDIX A2

Table A2.1. Sample information for the *A. palliata* individuals used in the mapping experiment to identify X-linked contigs. Phenotypic sex = presumed sex based on visual assessment in the field, *SRY* = *SRY* (Y-linked gene) haplotype (NA = no amplification, Apm = *A. palliata* haplotype), *HAM80* = X-linked microsatellite genotype (allele sizes in bp). See Cortés-Ortiz et al. 2019 for details. N raw reads = number of paired reads obtained from the sequencer.

Sample ID	**Phenotypic sex**	***SRY* (Y)**	***HAM80* (X)**	**N raw reads**
S173	F	NA	138/138	89,770,052
S145	F	NA	132/136	123,076,486
S608	M	Apm	136	54,368,866
S618	M	Apm	132	90,755,417

Table A2.2. Primer information for *Alouatta* assembly contigs used in validation of X-linkage (A=autosomal contig, X=presumed X-linked contig, Marmoset position is the chromosomal coordinate for the start of the alignment block).

Contig	**Contig length (bp)**	**Marmoset position**	**Primer sequences**	**Product length (bp)**
84001[A]	4,955	1: 182,469,879	ACGAATGCTTCAGGCTGAGT AGGCAAGACCCACTGGTATG	163
26402[X]	63,021	X: 104,175,832	CAGAGGCTAAATGGCTTTGC TCTTGGCTGTTTGCATGAAG	102
35197[X]	43,086	X: 83,726,463	CCCTCCCTGGAGAAAGAATC CTTGGTTGCTTTGCAGATGA	113
92787[X]	3,984	X: 52,026,941	TGCTTGTCATCCCAACACAT GATTACAGACGCCCACCACT	144
118733[X]	2,288	X: 6,638,526	ATGGGCTAGCAAGACTGCAT TAGGAAGTGGGTTCCTGTGG	162
60023[X]	14,492	X: 18,870,743	CACCTGTTGATGGACACTGG GCCTTTGGCATTCAGATCAT	139

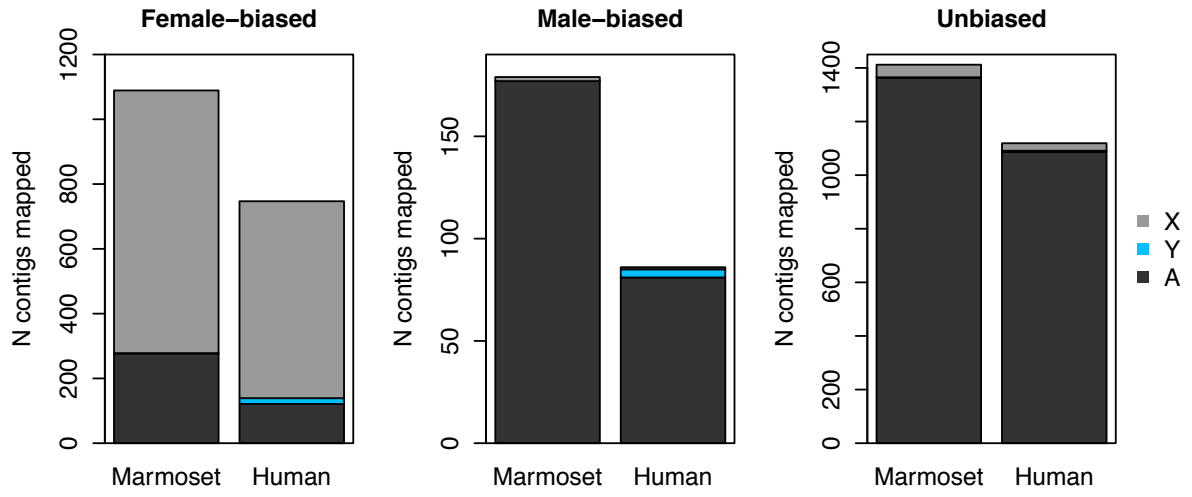

Figure A2.1. Number of contigs that mapped to the marmoset and human genome, for female-biased contigs, male-biased contigs, and unbiased contigs. Color denotes mapping position to either the X chromosome (light gray), the Y chromosome (blue), or to autosomes (dark gray).

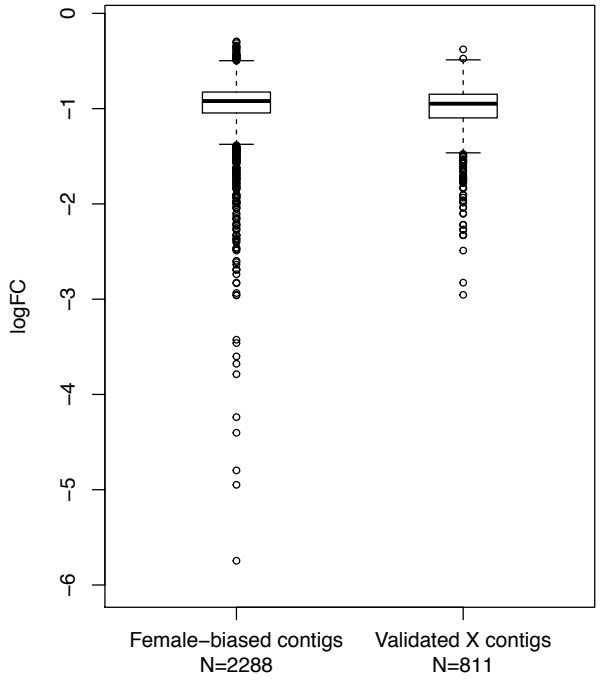

Figure A2.2. Male-to-female (log2) fold change in read mapping count for female-biased contigs and the subset of which mapped to the marmoset X chromosome (i.e., validated X contigs).

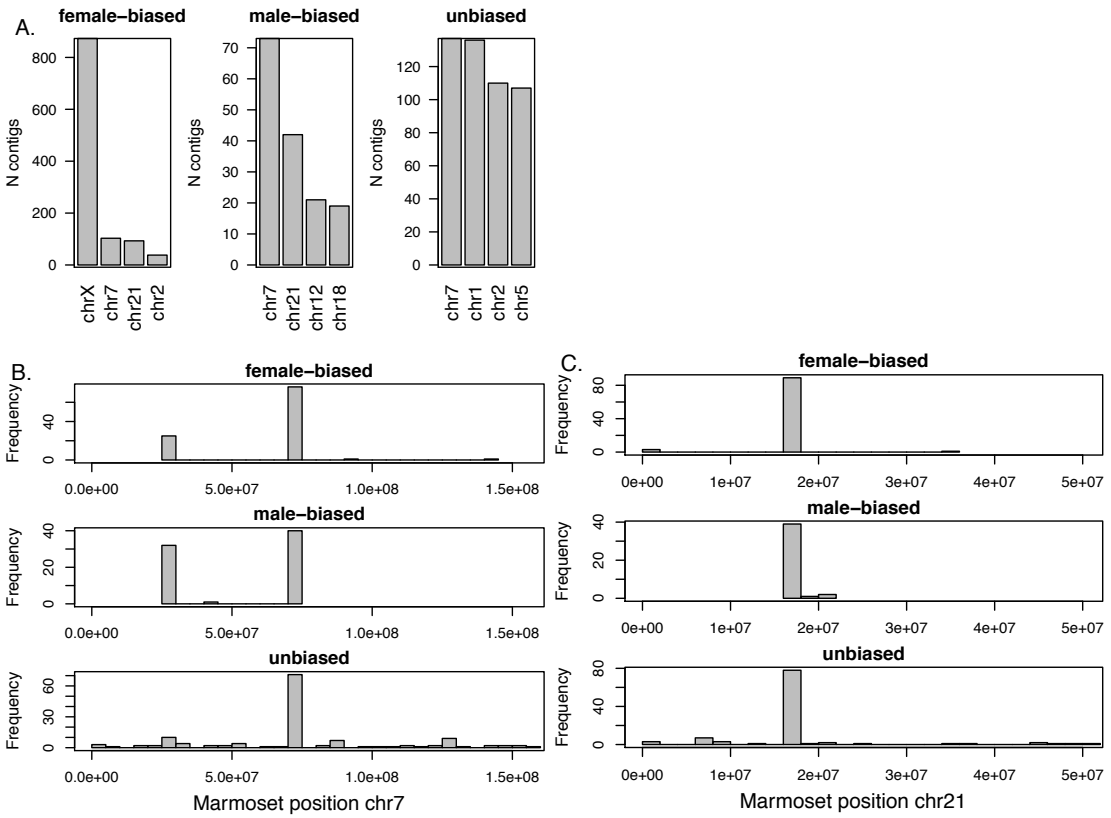

Figure A2.3. Mapping positions to the marmoset genome for sex-biased and unbiased contigs, for A) the top four chromosomes with most hits, B) chromosome 7, and C) chromosome 21.

Table A2.3. Genes present in regions associated with *Alouatta* outlier X-linked SNPs that map to the human X chromosome. R=region, calJac3=mapping position in the marmoset genome, bgc=outlier type (α=direction of introgression: α>0=excess *A. pigra* ancestry, α<0=excess *A. palliata* ancestry. β=amount of introgression: β>0=reduced introgression, β<0=increased introgression).

R	calJac3	Gene	Gene_biotype	Human homolog	Human homolog %id	Human homolog orthology type	bgc
1	X:46524184:46524588		protein_coding				α<0
1	X:46559441:46570184	ARAF	protein_coding	ARAF	97.7	ortholog_one2one	α<0
1	X:46570165:46617942	SYN1	protein_coding	SYN1	98.9	ortholog_one2one	α<0
1	X:46583380:46587609		protein_coding				α<0
1	X:46601350:46601456		snRNA				α<0
1	X:46622530:46628236	CFP	protein_coding	CFP	87.6	ortholog_one2one	α<0
1	X:46634262:46645920	ELK1	protein_coding	ELK1	96.0	ortholog_one2one	α<0
1	X:46647105:46653594	UXT	protein_coding	UXT	97.6	ortholog_one2one	α<0
1	X:46783582:46856973	ZNF81	protein_coding	ZNF81	96.1	ortholog_one2one	α<0
1	X:46805656:46805760		rRNA				α<0
1	X:46912743:46943825	ZNF182	protein_coding	ZNF182	96.5	ortholog_one2one	α<0
1	X:46943345:47077150	SLC38A5	protein_coding	SLC38A5	90.2	ortholog_one2one	α<0
1	X:46944183:46949397		protein_coding	SPACA5	91.2	ortholog_one2many	α<0
1	X:46944183:46949397		protein_coding	SPACA5B	91.2	ortholog_one2many	α<0
1	X:47008402:47008505		snoRNA				α<0
1	X:47043931:47053167		protein_coding	SSX1	58.5	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX4	58.5	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX4B	58.5	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX3	58.0	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX7	57.4	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX5	56.4	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX2	48.4	ortholog_one2many	α<0
1	X:47043931:47053167		protein_coding	SSX2B	48.4	ortholog_one2many	α<0
1	X:47056142:47056261		rRNA	RNA5SP503	77.5	ortholog_one2many	α<0
1	X:47056142:47056261		rRNA	RNA5SP504	77.5	ortholog_one2many	α<0
1	X:47088308:47090170	FTSJ1	protein_coding	FTSJ1	94.8	ortholog_one2one	α<0
1	X:47116194:47129261	PORCN	protein_coding	PORCN	97.2	ortholog_one2one	α<0
1	X:47130218:47138922	EBP	protein_coding	EBP	96.5	ortholog_one2one	α<0
1	X:47152036:47171588	TBC1D25	protein_coding	TBC1D25	98.5	ortholog_one2one	α<0
1	X:47170685:47170785		snoRNA				α<0
1	X:47182903:47183225		pseudogene				α<0
1	X:47183584:47186730		protein_coding				α<0
1	X:47183584:47184464		pseudogene				α<0

1	X:47192501:47192596		misc_RNA	Y_RNA	84.4	ortholog_one2many	α<0
1	X:47210660:47223652	WDR13	protein_coding	WDR13	99.6	ortholog_one2one	α<0
1	X:47251356:47252178		protein_coding				α<0
1	X:47266935:47268993		pseudogene				α<0
1	X:47304180:47313473	WAS	protein_coding	WAS	94.6	ortholog_one2one	α<0
1	X:47305340:47305474		snoRNA				α<0
1	X:47310054:47310349		misc_RNA				α<0
1	X:47321721:47332971	SUV39H1	protein_coding	SUV39H1	94.7	ortholog_one2one	α<0
1	X:47396131:47412004	GLOD5	protein_coding	GLOD5	68.2	ortholog_one2one	α<0
1	X:47447336:47472570	HDAC6	protein_coding	HDAC6	91.4	ortholog_one2one	α<0
1	X:47478309:47479575	ERAS	protein_coding	ERAS	93.6	ortholog_one2one	α<0
1	X:47524542:47530403	TIMM17B	protein_coding	TIMM17B	98.8	ortholog_one2one	α<0
1	X:47530215:47535840	PQBP1	protein_coding	PQBP1	98.5	ortholog_one2one	α<0
1	X:47536135:47547272	SLC35A2	protein_coding	SLC35A2	97.4	ortholog_one2one	α<0
1	X:47548628:47554630	PIM2	protein_coding	PIM2	97.4	ortholog_one2one	α<0
1	X:47558863:47592157		protein_coding	OTUD5	94.0	ortholog_one2many	α<0
1	X:47594241:47594345		snRNA	RNU6-722P	86.7	ortholog_one2many	α<0
1	X:47597647:47602686	KCND1	protein_coding	KCND1	93.9	ortholog_one2one	α<0
1	X:47609181:47638593	GRIPAP1	protein_coding	GRIPAP1	78.2	ortholog_one2one	α<0
1	X:47631097:47631185		snoRNA	SNORA40	91.0	ortholog_one2one	α<0
1	X:47648691:47649616		pseudogene				α<0
1	X:47658718:47673447	TFE3	protein_coding	TFE3	87.4	ortholog_one2one	α<0
1	X:47691970:47703316	CCDC120	protein_coding	CCDC120	98.0	ortholog_one2one	α<0
1	X:47706609:47709461	PRAF2	protein_coding	PRAF2	91.5	ortholog_one2one	α<0
1	X:47710584:47716044	WDR45	protein_coding	WDR45	96.8	ortholog_one2one	α<0
1	X:47726915:47727499		pseudogene				α<0
1	X:47756195:47766273	GPKOW	protein_coding	GPKOW	91.4	ortholog_one2one	α<0
1	X:47781741:47781853		misc_RNA	Y_RNA	86.7	ortholog_one2one	α<0
1	X:47783521:47783814		misc_RNA	RN7SL262P	84.4	ortholog_one2one	α<0
1	X:47794754:47799714	MAGIX	protein_coding	MAGIX	90.1	ortholog_one2one	α<0
1	X:47804303:47807461	PLP2	protein_coding	PLP2	97.4	ortholog_one2one	α<0
1	X:47808048:47817687	PRICKLE3	protein_coding	PRICKLE3	91.2	ortholog_one2one	α<0
1	X:47817279:47817428		protein_coding				α<0
1	X:47830633:47840358	SYP	protein_coding	SYP	96.2	ortholog_one2one	α<0
1	X:47846758:47877754	CACNA1F	protein_coding	CACNA1F	95.3	ortholog_one2one	α<0
1	X:47866258:47866359		misc_RNA				α<0
1	X:47879139:47897726	CCDC22	protein_coding	CCDC22	95.1	ortholog_one2one	α<0
1	X:47898060:47912844		protein_coding	FOXP3	90.2	ortholog_one2one	α<0
1	X:47917171:47935050	PPP1R3F	protein_coding	PPP1R3F	92.3	ortholog_one2one	α<0
1	X:47949517:47951180		protein_coding				α<0

1	X:47949852:47949935		miRNA				α<0
1	X:47970163:48051622		protein_coding				α<0
1	X:47970163:48051622		protein_coding				α<0
1	X:47974726:47975839		pseudogene				α<0
1	X:48126558:48133698		protein_coding	PAGE1	42.1	ortholog_one2one	α<0
1	X:48298011:48302715	PAGE4	protein_coding	PAGE4	82.4	ortholog_one2one	α<0
1	X:48352872:48354800	USP27X	protein_coding	USP27X	72.6	ortholog_one2one	α<0
1	X:48398918:48566118	CLCN5	protein_coding	CLCN5	99.0	ortholog_one2one	α<0
1	X:48476370:48476460		miRNA				α<0
1	X:48476732:48476799		miRNA				α<0
1	X:48479887:48479964		miRNA				α<0
1	X:48481665:48481748		miRNA				α<0
1	X:48482210:48482287		miRNA				α<0
1	X:48483790:48483872		miRNA				α<0
2	X:67480740:67481444		protein_coding	MAGEE2	85.8	ortholog_one2one	α<0
2	X:67594298:67595150		protein_coding	ARL5A	79.1	ortholog_one2one	α<0
2	X:67696303:67701697		protein_coding	PBDC1	95.3	ortholog_one2many	α<0
2	X:68113106:68113192		miRNA				α<0
2	X:68288442:68288560		rRNA	RNA5SP509	91.6	ortholog_one2one	α<0
2	X:68480659:68481183		processed_pseudogene				α<0
3	X:114052741:114054497		protein_coding	DCAF12L2	94.4	ortholog_many2many	β>0
3	X:114052741:114054497		protein_coding	DCAF12L1	85.3	ortholog_many2many	β>0
3	X:114422500:114423885		protein_coding	DCAF12L2	83.3	ortholog_many2many	β>0
3	X:114422500:114423885		protein_coding	DCAF12L1	80.5	ortholog_many2many	β>0
3	X:114626321:114626427		snRNA				β>0
3	X:114657456:114659354	PRR32	protein_coding	PRR32	91.9	ortholog_one2one	β>0
3	X:114756176:114756222		miRNA				β>0
4	X:135646720:135647448		protein_coding				β>0
4	X:135691416:136093653		protein_coding	AFF2	97.4	ortholog_one2many	β>0
4	X:136008198:136008478		misc_RNA	RN7SKP267	85.1	ortholog_one2one	β>0
4	X:136141861:136179092		protein_coding	AFF2	89.4	ortholog_one2many	β>0

CHAPTER III

Multiple Forms of Selection Shape Speciation in the *Alouatta* Hybrid zone

Preamble: The contents of this chapter have been published in *Molecular Ecology*. The published version appears as: Baiz, M. D., Tucker, P. K., & Cortés-Ortiz, L. (2019). Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular ecology* **28**, 1056–1069.

Introduction

Natural selection is considered to play an important role in driving speciation (Funk et al. 2006, Sobel et al. 2010). Divergent selection can contribute to speciation when allopatric populations encounter different habitats with different selective pressures (Schluter 2001, 2009). Under such circumstances, it is expected that loci that underlie local adaptation will show allele frequency differences in populations under different environments (Schluter 2009). Similarly, if allopatric populations encounter environments with similar selection pressures (i.e., uniform selection), different adaptive mutations may be selected for (Schluter 2001, 2009) because different mutations may result in similar optimal phenotypes. If populations that are experiencing either divergent or uniform selection maintain geographic isolation, thus restricting gene flow between them, divergence will proceed and the populations can become reproductively isolated

over time. In either case, the rate of divergence will be contingent upon the rate of migration, the strength of selection, and the initial allele frequencies. It is widely assumed that reproductive isolation can result as a by-product of such divergence in allopatry (Schluter 2001, Wu 2001, Sobel et al. 2010). This idea has rarely been tested empirically (Payseur & Rieseberg 2016, but see Kilias et al. 1980, Dodd 1989, Nosil et al. 2012a, Gompert et al. 2012b, Parchman et al. 2013, Janoušek et al. 2015).

In sympatric populations, selection can directly favor reproductive isolation. This can occur in hybrid zones when hybrids are less fit than parental types and as a consequence, individuals who mate with conspecifics have greater reproductive success than individuals who mate with heterospecifics (Butlin 1987). This process, called reinforcement, has traditionally been considered to result in a strengthening of prezygotic barriers that prevent the formation of unfit hybrids (Butlin 1987, Servedio & Noor 2003). However, it has recently been extended to include the evolution of any additional barrier effect in sympatry, including postzygotic isolation, as a form of adaptive coupling of reproductive barriers (Butlin & Smadja 2018). The frequency at which reinforcement occurs and its importance in shaping species diversity are open questions (Servedio & Noor 2003, Servedio 2004).

Hybrid zones offer a unique opportunity to test hypotheses about the contribution of different forms of selection that shape reproductive isolation over the course of the speciation process (e.g., Nosil et al. 2012b). They are particularly suited to empirical investigation of the genetic basis of reproductive isolation as population genetic data can be used to infer differential patterns of introgression. Barrier loci, i.e. loci associated with reproductive isolation, should have a signature of reduced introgression relative to the neutral expectation, which is caused by limited gene flow as a consequence of selection against hybrids (Barton & Hewitt 1985,

Gompert & Buerkle 2011a). If the genetic differences that contribute to reproductive isolation in the hybrid zone involve loci under selection in allopatric parental populations, we should expect to see higher differentiation in allopatric populations for barrier loci compared to neutral markers (Payseur & Rieseberg 2016). If reinforcing selection shaped barrier loci in the hybrid zone, we should expect to see greater differentiation in sympatry than in allopatry for these markers (e.g., Nosil et al. 2012b, Wang et al. 2014).

Here, we examined locus-specific differentiation and introgression using reduced-representation sequencing data from a bimodal howler monkey hybrid zone (*Alouatta palliata* x *A. pigra*) (Cortés-Ortiz et al. 2015) and from allopatric parental populations to test predictions about the forms of selection acting on loci associated with reproductive isolation. The parental species diverged ~3 MA (Cortés-Ortiz et al. 2003) and have many important differences in their morphology (Smith 1970, Kelaita et al. 2011), cytogenetics (Steinberg et al. 2008), social systems (Chapman & Balcomb 1998, Ho et al. 2014), and loud vocalizations (Bergman et al. 2016). Throughout most of their ranges, *A. palliata* and *A. pigra* are allopatric, but their ranges overlap in a narrow contact zone (~20 km, Cortés-Ortiz et al. 2007, Cortés-Ortiz et al. 2019) in Tabasco, Mexico (Figure 3.1). It is likely that the contact zone in Tabasco is the result of secondary contact between the parental species after periods of isolation and range expansion (Cortés-Ortiz et al. 2003, Ford 2006, Ellsworth & Hoelzer 2006). Despite the relatively large degree of divergence between *A. palliata* and *A. pigra*, reproductive isolation is incomplete, as hybridization has been confirmed in the contact zone using molecular markers. Initial surveys showed that multigenerational backcrossed hybrids into each parental species are nearly equally abundant, there are few intermediate hybrids, and no putative F1s (Cortés-Ortiz et al. 2007; Kelaita & Cortés-Ortiz 2013). We have previously shown that there is a lack of introgression for

SRY (the Y-linked sex determination gene), suggesting that F1 males may be infertile or inviable (Cortés-Ortiz et al. 2019). Consistent with this, anecdotal evidence suggests that there may be a cost to hybridization as a previously identified intermediate hybrid male did not produce offspring despite living as the only adult male in a group with two reproductively mature females for a period of seven years (LCO personal observation). We also found reduced introgression for X-linked markers (Cortés-Ortiz et al. 2019), consistent with the "large X effect", which suggests the X chromosome plays a disproportionate role in speciation (Coyne & Orr 1989).

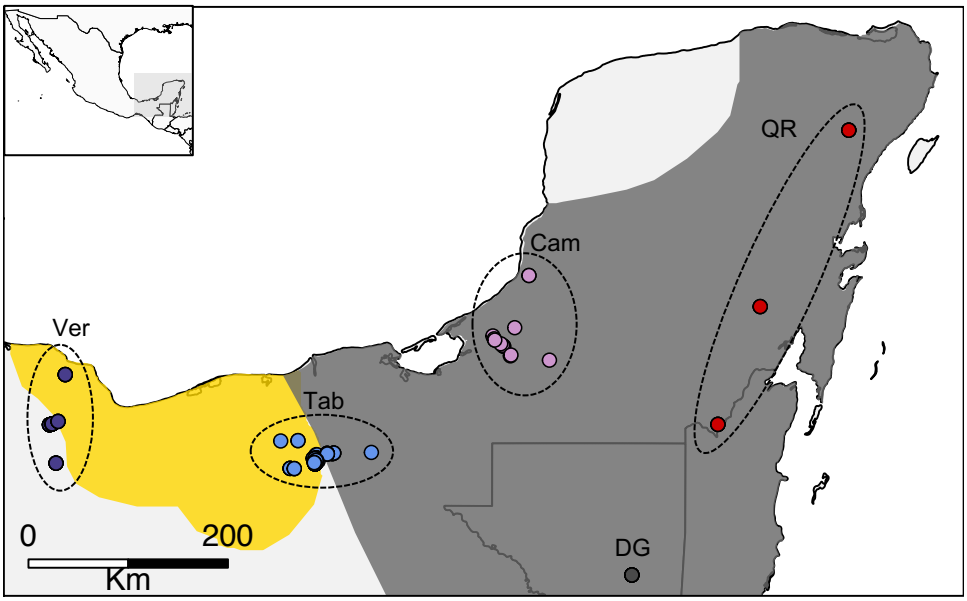

Figure 3.1. Map of sampling sites used in this study. The range for *A. palliata* is shown in yellow and the range for *A. pigra* is in gray. Ver=Veracruz, Mexico, Tab=Tabasco, Mexico, Cam=Campeche, Mexico, DG=Dolores, Guatemala, QR=Quintana Roo, Mexico.

For this study, we used ddRADseq data to assess the extent of genomic admixture and the distribution of admixed genotypes across the *Alouatta* contact zone and identified loci that show a pattern of reduced introgression (the pattern expected for loci associated with reproductive

isolation) relative to the genomic background, which is assumed to be mostly neutral. By exploring the relationship between locus-specific differentiation and introgression in the hybrid zone, we tested the hypothesis that reproductive isolation results as a by-product of divergence in allopatry and that reinforcing selection in sympatry shapes reproductive barriers. We also performed functional annotation for loci that showed the strongest evidence for divergent selection in sympatry to associate these regions with putative phenotypes under reinforcement and evaluated the pattern of introgression for putatively X-linked markers to test for a large X effect. Our results are consistent with signatures of divergence in allopatry and reinforcement in sympatry, indicating that multiple forms of selection have shaped the evolution of reproductive isolation in this system.

Materials and Methods

Sampling

Our sampling included 181 wild individuals captured between 1998 and 2012 following procedures described in Kelaita et al. (2011, 2013) and adhered to the University of Michigan's Institutional Animal Care and Use Program standards (UCUCA permit #09319). Individuals were chosen from a larger pool of previously collected samples to maximize the geographic distribution of *Alouatta* in Mexico, as well as the representation of admixed genotypes present in the hybrid zone (i.e., individuals backcrossed into *A. palliata*, intermediate hybrids, individuals backcrossed into *A. pigra*) as determined by hybrid index measured from 29 microsatellite markers (Cortés-Ortiz et al. 2019). Thus, our samples included 99 individuals from the hybrid zone in Tabasco, Mexico (Tab), 38 allopatric *A. palliata* from Veracruz, Mexico (Ver), and 44

allopatric *A. pigra* including 24 from Campeche, Mexico (Cam), 12 from Quintana Roo, Mexico (QR), and 8 from Dolores, Guatemala (DG) (Figure 3.1).

Samples were kept on ice in the field and stored at -20°C upon arrival in the laboratory. Genomic DNA was extracted with the QIAGEN DNeasy tissue kit (Qiagen Inc., Valencia, CA) following the manufacturer's protocol for animal tissue extractions with the following modifications: 1) we added 100 µl of whole blood in lysis buffer solution (1:5 concentration) to 100 µl buffer of ATL, 2) we eluted DNA in 70 µl of water at 55 °C twice (re-using the same spin column) to maximize DNA yields, after incubating for 5 minutes at room temperature.

ddRADseq and genotyping

We prepared and sequenced four ddRAD libraries, following the Peterson et al. (2012) protocol, each library containing DNA from 48 individuals (two libraries also included individuals sequenced for use in other projects). Briefly, we used the restriction enzymes SphI and MluCI to digest 200–300ng DNA per sample, size selected fragments between 150–350bp using a 2% Pippen Prep gel (Sage Science, Beverly, MA), and sequenced libraries on an Illumina HiSeq 4000 machine at the University of Michigan Sequencing Core to obtain 150bp paired-end reads.

We demultiplexed our data using *pyRAD* (Eaton et al. 2014), merged read pairs that overlapped using *FLASH* (Magoč & Salzberg 2011), and aligned both successfully merged reads and unmerged reads (which were expected due to our size selection window) to the draft *Alouatta palliata* genome assembly (accession ID PVKV00000000) with *BWA-MEM* (Li 2013) using default settings. This genome assembly is part of the 200 Mammals Project of the Broad Institute (unpublished, shared by Kerstin Lindblad-Toh with permission to publish our genotype

data). We then called variants and generated a VCF file using *samtools mpileup* (including options -u -g -t DP, DPR) and *bcftools call* (options -v -m -O v) (Li et al. 2009). After removing SNPs within 5bp of an indel and retaining variants with a minimum quality score of 20, we obtained 6,415,368 loci (including SNPs and indels) that were subsequently filtered in further analyses (Table A3.1).

Admixture and population structure

We first used *fastStructure* (Raj et al. 2014) to quantify admixture proportions and to assign admixed (hybrid) or non-admixed status to individuals. Because we were interested in detecting hybrid individuals, we ran ten replicates using the simple prior to infer admixture proportions (Q) with the number of clusters (K) equal to two, reflecting the parental species. Here, admixture proportions are an estimate of the proportion of the genome inherited from each parental species (i.e., where Q_1 is the proportion from parental species 1 and Q_2 is the proportion from parental species 2 and $Q_1 + Q_2 = 1$). For simplicity, we only report Q_1 as the proportion of the genome inherited from *A. pigra*.

Because our sampling sites are geographically widespread and there may be some within-population structuring, we ran an additional ten replicates for each K between 3 to 8 to ensure that imposing K=2 on this system did not affect our ability to detect hybrids. We used the *fastStructure* script 'chooseK.py' to detect the number of clusters that best fit our data. We also examined the correlation between admixture proportion Q_1 and hybrid index as calculated by *bgc* (see below) for hybrid zone individuals using the Pearson method in R.

We limited our *fastStructure* analysis to biallelic loci (those with different alleles between species or intraspecific polymorphism) with low missing data across individuals. To do this, we

used *bcftools* and *vcftools* (Danecek et al. 2011) to filter loci by removing indels, non-biallelic sites, sites with a minor alleles frequency of ≤ 0.01, and sites with a minimum mean depth across individuals of less than 10. Because we did not apply a depth-per-read threshold for calling genotypes, we discarded sites with missing genotypes in more than 50% of individuals to only include sites with high read depth. To reduce effects of linkage and comply with assumptions of the *fastStructure* model-based approach, we also thinned sites within 200bp of each other and discarded sites out of Hardy-Weinberg equilibrium in either of the allopatric parental populations. This resulted in 74,448 SNPs, which we included in the *fastStructure* analysis (Table A3.1). Using this dataset, we dropped 23 individuals from further analyses due to a high frequency (>80%) of missing genotype data, which can affect confidence in *fastStructure* results. In further analyses, we identified hybrids as individuals with $0.05<Q_1>0.95$, non-admixed *A. palliata* individuals as $Q_1<0.05$, and non-admixed *A. pigra* individuals as individuals with $Q_1>0.95$. With this dataset, we also visualized structure among sampling sites using principle component analysis (PCA) implemented in *SNPRelate* (Zheng et al. 2012).

Since our fastStructure analyses identified individuals in allopatric populations with some level of admixture (N=11, Table A3.2), we dropped these individuals from further analyses to avoid complications from including admixed individuals outside the hybrid zone in differentiation and introgression analyses. This reduced our dataset to include 81 individuals from the hybrid zone (Tabasco), 32 allopatric *A. pigra* individuals, and 34 allopatric *A. palliata* individuals (Table A3.1). All further analyses only included these individuals.

Genomic cline analysis

For our analyses of genomic clines and genetic differentiation, we filtered loci to increase confidence in genotype calls and to maximize information about ancestry. To do this, we retained biallelic loci with a minimum mean depth across individuals of 30, loci with a minor allele frequency of ≥ 0.05, and loci that were present in at least 80% of individuals in either parental population. This resulted in 5,763 loci (Table A3.1) distributed on 2,883 contigs between 80.2 Kb–1.28 Mb in size (representing 18.8% of the total reference assembly).

To quantify introgression across loci and identify candidate variants with evidence for an association with reproductive isolation (i.e., reduced introgression compared to neutral expectations), we used genomic cline analysis implemented in *bgc* (Gompert & Buerkle 2011a, Gompert & Buerkle 2012a). Genomic cline analysis uses differential introgression to identify loci that are more or less likely than the genome-wide average (assumed to be neutral) to introgress between populations. For each locus, *bgc* uses the cline parameter β to quantify the amount of introgression, with $\beta<0$ indicating greater than expected introgression and $\beta>0$ indicating reduced introgression with respect to the genome-wide average. Loci showing evidence for an association with reproductive isolation (barrier loci) are expected to have reduced introgression ($\beta>0$) due to selection against hybrids.

We ran genomic cline analyses under the genotype uncertainty model (appropriate for next-generation sequence data, Gompert et al. 2012b) in *bgc* for five independent chains, each with a burn-in of 30,000 for 50,000 steps, and thinned samples by 20. We then merged outputs and identified β outliers from MCMC output as loci with a 95% credible interval that does not overlap zero.

Identifying putative X chromosome markers

To allow us to test for restricted introgression of the X chromosome, we used NCBI BLASTN (Altschul et al. 1997) to associate *Alouatta* contigs with genes on the X chromosome in humans. We expect genes on the human X to be X-linked in *Alouatta* since gene content and order is highly conserved across mammals (Delgado et al. 2009), and the *Alouatta* X appears to be highly similar to the human X (Steinberg et al. 2014). We downloaded unique, unspliced human X chromosome genes (GRCh38.p12) from Ensembl (Zerbino et al. 2017) using the online BioMart tool. We assumed any locus to be putatively X-linked if any of the human X gene sequences had a BLASTN best hit (–max_hsps 1 –max_target_seqs 1, otherwise default settings) to the same *Alouatta* contig and the percent identity of the aligned sequence was >85%. Using these criteria, we identified 191 putatively X-linked loci on 90 contigs in our data set. To determine the pattern of introgression for the X chromosome, we visually inspected the pattern of introgression for this subset of putatively X-linked loci.

Genetic differentiation

Loci influenced by divergent selection are expected to show elevated differentiation (Beaumont & Balding 2004, Gompert & Buerkle 2011b). We measured locus-specific differentiation with the method of Weir & Cockerham (1984) implemented in *vcftools* (Danecek et al. 2011) using the same dataset as for our cline analysis. For allopatric parental populations, we tested for differences in the distribution of F_{ST} values between loci that showed reduced introgression in the hybrid zone, those with neutral introgression, and those with increased introgression using ANOVA and detected pair-wise differences between each category using the Tukey Honest Significant Difference method, both implemented in base R v3.4.1 (R Core Team

2017). For analyses of differentiation in sympatry, we calculated F_{ST} between *A. palliata*-like (mean Q_I<0.5, N=54) and *A. pigra*-like individuals (mean Q_I>0.5, N=27) in the hybrid zone. We included the same set of loci for each comparison, but when we separated the samples into allopatric and sympatric populations, some loci were no longer polymorphic. Thus, for comparisons where sites were monomorphic and at sites where there was more variation within than between populations (F_{ST}<0), we report F_{ST}=0. We compared the distribution means of F_{ST} in sympatry and F_{ST} in allopatry for loci with reduced and neutral introgression using a Wilcoxon Rank Sum test and by fitting a linear model to the relationship between these two variables in R. If reinforcing selection contributed to reproductive isolation in sympatry, we would expect loci to have higher F_{ST} in sympatry than in allopatry and to see a relationship between the two variables that differs from a 1:1 linear relationship, which would be assumed if divergence in sympatry is equal to divergence in allopatry (i.e., no reinforcement).

Genomic basis of reinforcement

To explore potential functions of loci with reduced introgression that showed strong evidence for divergent selection in sympatry, we identified homologous human protein-coding genes near these loci. We first identified the set of loci with reduced introgression (β>0) that had greater F_{ST} in sympatry than in allopatry ($F_{STsympatry} - F_{STallopatry}$>0, N=104), i.e., those with a signature of reinforcement. To focus our search on regions showing only evidence for reinforcement and not divergence in allopatry or adaptive introgression, we excluded contigs that also contained β>0 loci with greater F_{ST} in allopatry than in sympatry and contigs that also contained β<0 outliers. This resulted in 93 loci on 79 contigs. We then ranked loci by the difference in $F_{STsympatry} - F_{STallopatry}$ and took the top 10% of loci with the greatest difference as

candidate loci showing strong evidence for selection in sympatry. For each locus, we extracted the entire contig from the *Alouatta* genome assembly and used the UCSC genome browser online BLAT tool (Kent 2002) to identify its position in the human genome (version GRCh38/hg38). We BLAT searched the first and last 25kb of sequence separately for each *Alouatta* contig and took the outermost coordinates from each alignment so that we could identify human genes that occur between regions for each contig (Table A3.3). For each region, we ensured that alignment orientation, length, and span of human genomic positions were consistent with each *Alouatta* contig, suggesting that the human genomic regions are collinear and we can assume these genes are also present on the *Alouatta* contigs. We then used biomaRt (Durnick et al. 2005, 2009) to obtain all human protein coding genes within each region. Finally, we identified mammalian phenotypes associated with each gene using the Mouse Genome Informatics (MGI) batch query tool online (URL: http://www.informatics.jax.org).

Results

Structure and admixture

Across ten replicate *fastStructure* runs for each K between 2–8, maximum likelihood scores were highest for K=2 (Figure A3.1A). Model complexity that maximizes marginal likelihood was equal to two in each replicate, and the number of model components used to explain structure in the data was equal to two in four replicates and equal to three in six replicates (Figure A3.1B). Admixture proportions using K=2 and K=3 were very similar due to extremely low assignment values for each individual to the third cluster (each $Q_3<0.0001$) in nine of 10 K=3 replicates (Figure A3.2). Further, hybrid index scores inferred from *bgc* were closely correlated with *fastStructure*'s Q_1 at K=2 ($r=0.996$, $P<2.2X10^{-16}$) (Figure 3.2B).

Together, these results indicate that K=2 best describes our data and that our use of admixture proportion Q_1 was appropriate in assigning hybrid status to individuals. Thus, we report mean Q_1 scores across our ten K=2 replicates (Figure 3.2, Table A3.4).

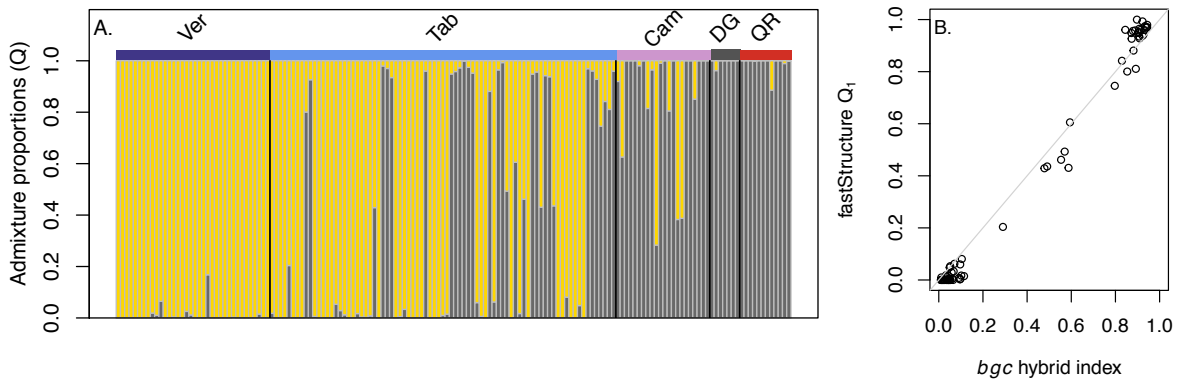

Figure 3.2. A) *fastStructure* plot at K=2 showing the geographical distribution of non-admixed individuals and hybrids. Individuals are arranged from West (left) to East (right). *A. palliata* ancestry is shown in yellow and *A. pigra* ancestry is shown in gray. B) Admixture proportion Q_1 is closely correlated with *bgc* hybrid index for individuals in the hybrid zone (Tabasco, only representing individuals under the blue bar in the *fastStructure* plot).

Concordant with our previous analyses using microsatellite markers (Cortés-Ortiz et al. 2019), our *fastStructure* analysis at K=2 shows that most individuals in the contact zone are multigenerational backcrosses to either parental species and there are few intermediate hybrids (Figure 3.2). Out of 81 individuals, we identified five with an intermediate admixture proportion ($0.4>Q_1<0.6$). Although admixture is mainly restricted to the contact zone, several individuals in Campeche also appear to be admixed, along with a single individual in Quintana Roo and two individuals in Veracruz (Figure 3.2, Table A3.4). In Campeche and Quintana Roo, most admixed individuals are *A. pigra*-like ($Q_1>0.6$), concordant with the geographic range of the parental species that inhabits those locations. Similarly, the admixed individuals in Veracruz have predominantly *A. palliata* ancestry ($Q_1<0.2$).

The PCA results are largely concordant with the *fastStructure* K=2 analysis, suggesting that the set of variants used to detect hybrids robustly discriminates the parental species from each other and from hybrids (Figure 3.3A). PC 1 explains 55% of the genetic variation among individuals and clearly separates allopatric populations (with the exception of admixed individuals detected outside the contact zone). Thus, not surprisingly, PC 1 is strongly correlated with the *fastStructure* admixture proportion Q_1 (r=0.98, P<2.2 x 10^{-16}, Figure 3.3B). PC 2 explains 2.4% of the genetic variation among individuals and seems to primarily be associated with population structure among sampling sites within *A. pigra*.

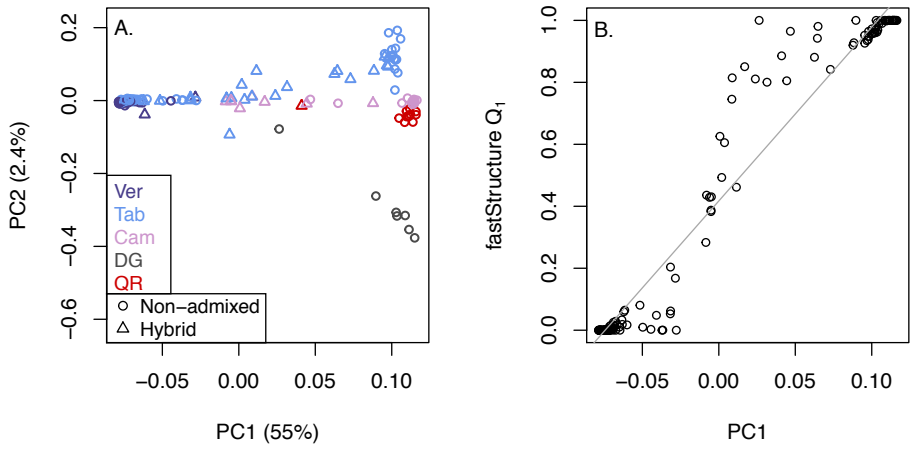

Figure 3.3. A) PCA summarizing population structure among sampling sites. PC1 explains 55%, and PC2 explains 2.4% of the genetic variation among individuals. Open circles are non-admixed individuals and triangles are hybrid individuals as determined by their admixture proportion (Q_1) in *fastStructure*. B) Admixture proportion Q_1 is closely correlated with PC1. The gray line is a linear model fit to the data.

Differential introgression across loci

We found a small percentage of loci that were β outliers (Figure 3.4) consistent with non-neutral introgression. There were 255 loci (4.4%) that showed reduced introgression (β>0) distributed on 206 contigs (1.2 loci/contig) and 319 loci (5.5%) with increased introgression

(β<0) distributed on 248 contigs (1.3 loci/contig). Only six contigs had loci with both reduced and increased introgression. The remaining 5,189 loci (90%) were consistent with neutral introgression (β=0). Of the 191 putatively X-linked loci, 183 (96%) had neutral introgression (Figure A3.3). Five loci showed reduced introgression, three of which were tightly linked on the same contig (within 12bp). Three loci showed increased introgression, all of which were on different contigs.

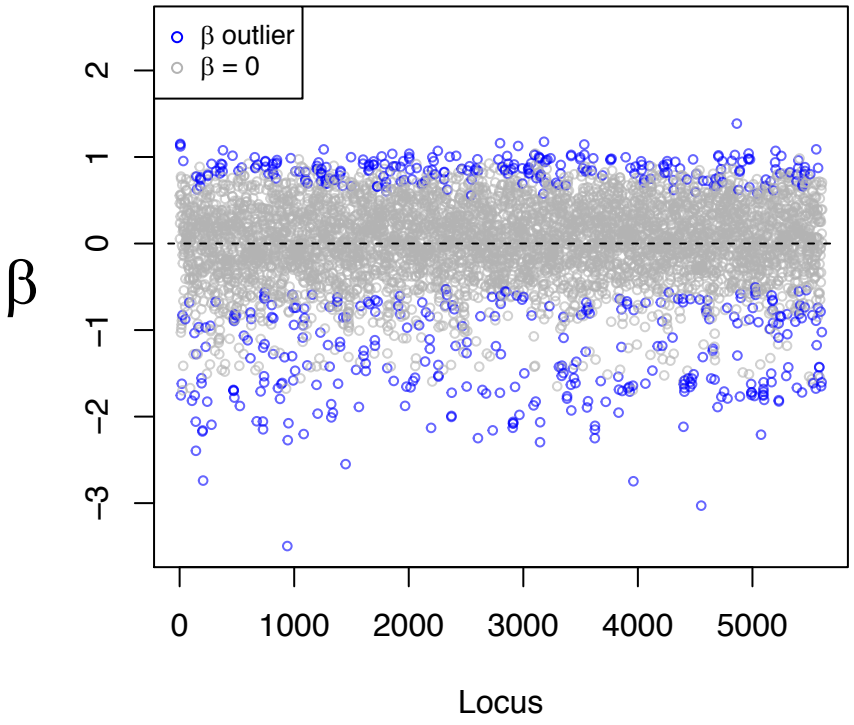

Figure 3.4. Locus-specific point estimates for the genomic cline parameter β (amount of introgression) with β outliers in blue. β>0 indicates reduced introgression, β<0 indicates increased introgression. β=0 indicates neutral introgression.

Genetic differentiation and its relationship with introgression

Locus-specific genetic differentiation between allopatric parental species was high overall (mean F_{ST}=0.65) and ranged from 0–1 with a seemingly bimodal distribution with peaks near F_{ST} = 0 and F_{ST} = 0.9 (Figure 3.5A). Of the 5,763 loci analyzed, 117 had fixed differences (F_{ST}=1) between allopatric parental species. Overall, differentiation was positively correlated with the amount of introgression (β) in the hybrid zone, but the relationship was weak (r=0.08, P=6.21 x 10^{-10}).

Mean F_{ST} for allopatric parental populations was not equal among β categories (F=85.93, P<2.2 X 10^{-16}). Post hoc comparisons indicated that the distributions of F_{ST} within each β category were significantly different from each other (Figure 3.5B, Table A3.5), with β>0 loci having the highest F_{ST} (mean=0.85, range=0.31–1), and loci with β=0 having the lowest F_{ST} (mean=0.63, range=0–1). β<0 loci had an intermediate F_{ST} (mean=0.78, range=0–1).

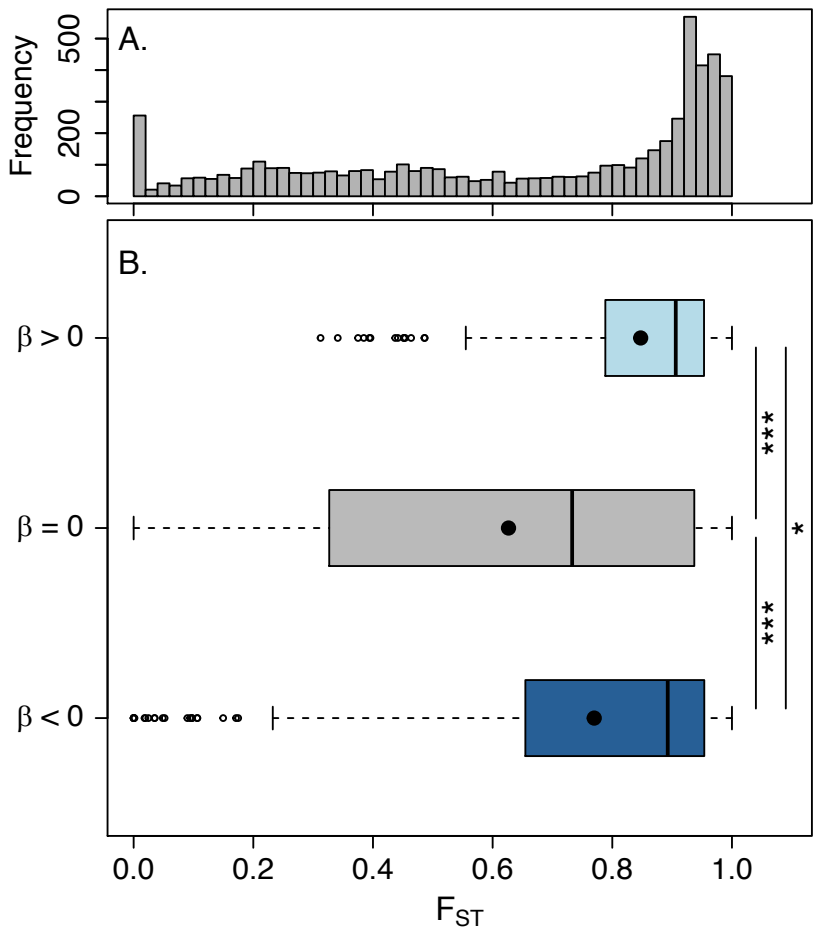

Figure 3.5. Genetic differentiation between allopatric parental species. A) Distribution of F_{ST} for all loci. B) Boxplot showing F_{ST} for loci in each β category. Within each box, distribution medians are denoted by the vertical line and means are denoted with a black circle. Box height is equal to the 1st–3rd interquartile range. *P<0.05, ***P<0.001.

Comparison of differentiation in sympatry and allopatry

Genetic differentiation across loci was lower in sympatry (mean $F_{STsympatry}$ = 0.55) than in allopatry (mean $F_{STallopatry}$ = 0.65). When loci are partitioned across β categories, F_{ST} was significantly higher in allopatry than in sympatry for markers with neutral and increased introgression (β=0: mean $F_{STsympatry}$=0.544, mean $F_{STallopatry}$=0.628, P<2.2 X 10^{-16}, β<0: mean $F_{STsympatry}$=0.373, mean $F_{STallopatry}$=0.778, P<2.2 X 10^{-16}). However, we found that for loci with reduced introgression (β>0), F_{ST} was significantly higher in sympatry than in allopatry (mean

$F_{STsympatry}$=0.852, mean $F_{STallopatry}$=0.850, P=6.37 X 10^{-6}). Although the magnitude of the difference is small, this pattern seems to be driven by loci with intermediate differentiation in allopatry having higher differentiation in sympatry, while loci with high differentiation in allopatry tended to also have high differentiation in sympatry (Figure 3.6A). The fit of linear models to the data in each beta category showed that confidence intervals for the slope of the line did not encompass one (Table 3.1), but was closer to one for loci with neutral introgression (Figure 3.6B).

Table 3.1. Summary of linear models fit to the relationship between F_{ST} in sympatry and F_{ST} in allopatry for loci with reduced introgression (β>0), neutral introgression (β=0), and increased introgression (β<0).

Pattern of introgression	**Adjusted r^2**	**Slope**	**Confidence interval 2.5%**	**Confidence interval 97.5%**	**P**
β>0	0.53	0.36	0.32	0.40	2.2 X 10^{-16}
β=0	0.76	0.73	0.71	0.74	2.2 X 10^{-16}
β<0	0.40	0.57	0.49	0.65	2.2 X 10^{-16}

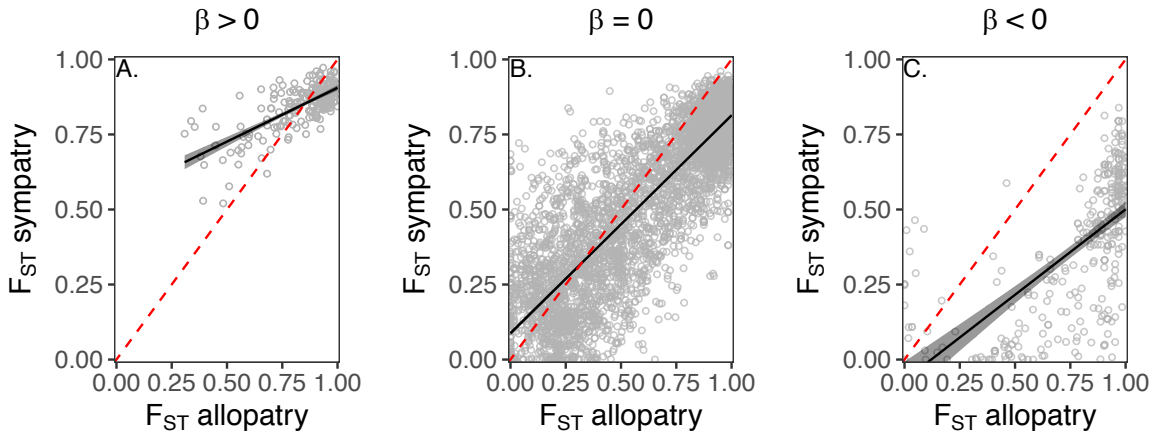

Figure 3.6. The relationship between locus-specific differentiation in allopatry and sympatry for A) loci with reduced introgression (β>0) (r^2 = 0.53, P<2.2 X 10^{-16}), B) loci with neutral introgression (β = 0) (r^2 = 0.76, P < 2.2 X 10^{-16}), and C) loci with increased introgression (β<0) (r^2=0.57, P<2.2 X 10^{-16}). In each case, the linear model fit to the data is represented by a solid black line with gray shading showing the 95% confidence interval of the slope, and the dashed red line indicates a 1:1 relationship.

Nine loci were included in the top 10% of β>0 loci that showed the greatest difference in F_{ST} between sympatry and allopatry (Table A3.6). We identified regions of human chromosomes 3, 4, 7, 8, 11, and 16 that seem to be homologous with the *Alouatta* contigs containing these loci. The human regions contained 42 protein-coding genes, of which 28 could be associated with 420 mammalian phenotypes (MPs) in the MGI database (Table A3.6). Notably, several genes were associated with behavior (*SCARB2*, *BRPF1*, *SLC5A2*, *KMT2A*), abnormal embryonic/fetal development or lethality (*SHROOM3*, *BRPF1*, *CRELD1*, *TADA3*, *ARL13B*, *PROS1*, *KMT2A*), hair texture (*ARPC4*, *KMT2A*), facial morphology (*SHROOM3*, *CRELD1*, *KMT2A*), and the immune system (*ITGAD*, *ITGAX*, *CD3C*, *CD3E*, *CD3G*, *KMT2A*).

Discussion

We used reduced-representation sequencing to examine admixture, population structure, introgression, and its relationship with locus-specific differentiation in a natural primate hybrid zone system. Our results are consistent with the hypothesis that reproductive isolation results as a byproduct of divergence in allopatry and we detected a genomic signature of reinforcement in sympatry, indicating that multiple forms of selection have shaped speciation in this system.

Admixture and population structure

We found a bimodal distribution of admixture proportions in the hybrid zone. Early generation hybrids are rare and multi-generational backcrosses dominate. This pattern is largely consistent with our previous analyses using a small set of microsatellite markers (Cortés-Ortiz et al. 2019). However, we detected admixture in areas where the parental species are thought to be allopatric. We detected a few admixed individuals east of the contact zone in Campeche and Quintana Roo and west of the contact zone in Veracruz. In addition to autosomal markers, we previously amplified a Y-linked (*SRY*) locus, X-linked loci including the microsatellite locus HAM80, as well as the mitochondrial control region for most individuals sequenced in the study (Table A3.2). Considering the sex-linked genotypes for these individuals together with the admixture proportions calculated in this study, it is clear that these individuals are not F1 hybrids. Due to the apparent absence of non-admixed individuals of the opposite species in these areas, and their distance from the contact zone (~200km or greater), we suspect that the presence of admixed individuals in these regions is likely due to either long distance migration from the contact zone, movement of animals by humans, or to past introgression during a period when the contact zone occurred in a different location than in present day and has since shifted. It will be

possible to test the hypothesis that the hybrid zone has moved by looking at linkage disequilibrium (LD) in a transect across the hybrid zone (e.g., Wang et al. 2011).

Introgression in the hybrid zone

We found evidence for differential introgression in the hybrid zone. The majority of loci exhibited neutral introgression, but a small percentage of markers showed extreme introgression (Figure 3.4). We were particularly interested in loci with reduced introgression ($\beta>0$) as this pattern is expected of loci associated with reproductive isolation. We identified 255 such loci. These loci were distributed on 206 contigs, which may support the hypothesis that reproductive isolation has a genome-wide basis (Parchman et al. 2013, Scordato et al. 2017). However, because the *A. palliata* genome is not assembled to chromosome-level, it is possible that these contigs may be physically linked.

Point estimates of β were much less variable for loci with a pattern of reduced introgression ($\beta>0$) than for loci with a pattern of increased introgression ($\beta<0$), particularly for β outliers (Figure 3.4). These coincident $\beta>0$ clines may be a reflection of the coupling of multiple barrier effects in the hybrid zone (Butlin & Smadja 2018). Recent admixture between divergent populations causes correlations between linked loci that persist over many generations (Stephens et al. 1994, Verardi et al. 2006), so the effects of indirect selection on loci near barrier loci can be strong in hybrid zones. The *Alouatta* hybrid zone is bimodal (Cortés-Ortiz et al. 2019) and differentiation is high between the parental species (Figure 3.5A), so admixture linkage disequilibrium is likely high. Thus, strong barrier effects may influence the whole genome via indirect selection (e.g. Szymura & Barton 1991), making it difficult to identify loci underlying individual barrier effects (Butlin & Smadja 2018).

In a previous analysis, we found complete lack of introgression for a Y-linked marker and limited to no introgression for three X-linked markers (Cortés Ortiz et al. 2019), consistent with other studies suggesting the sex chromosomes may plan an important role in reproductive isolation (Tucker et al. 1992, Masly & Presgraves 2007). We expanded upon these results by identifying putatively X-linked markers in our reduced-representation dataset based on sequence homology to known X-linked human genes. Of 90 putatively X-linked contigs, three had loci that showed reduced introgression and three had loci that showed increased introgression, while the remaining contigs had loci with neutral introgression (Figure A3.3). These results may indicate that few regions of the X chromosome underlie reproductive isolation in this system, although our interpretation may have limitations. First, we may have excluded many loci as putatively X-linked due to divergence between the *Alouatta* assembly and human gene sequences. However, the mammalian X chromosome is known to be highly conserved across mammals (Delgado et al. 2009, Mueller et al. 2013), only 146 of 2,367 human X genes did not have a match in the *Alouatta* genome assembly, and percent identity was generally high across BLASTN hits (mean=89.5%). Second, although the X is highly conserved, New World primates are known to have a high rate of chromosomal rearrangements (de Oliveira et al. 2012) including autosome-to-sex chromosome translocations in *A. pigra* and *A. palliata* (Solari & Rahn 2005, Steinberg et al. 2008, 2014). Thus, many loci we considered to be autosomal may be on regions translocated to the X. Validation of chromosome-linkage for assembly contigs and high density genotype data across the genome will be desirable to overcome these limitations.

Loci with reduced introgression are highly differentiated in allopatry

We found that compared to neutral loci and loci with increased introgression, loci with reduced introgression were more highly differentiated in allopatric parental populations (Figure 3.5B), suggesting a role for selection in driving reproductive isolation as a by-product of divergence in allopatry. As such, it seems likely that in this system, some level of reproductive isolation was already present upon secondary contact.

Because allele frequency differences are a prerequisite for testing introgression, the amount of locus-specific introgression and differentiation may be non-independent. Simulations have shown that when overall differentiation is low ($F_{ST}<0.1$) spurious correlations between F_{ST} and genomic cline parameters can occur in the absence of selection (Gompert et al. 2012b). However, this should not be much of an issue here since mean differentiation is relatively high (Figure 3.5A). Further, such an effect should shape the relationship between F_{ST} and β similarly for loci across β categories and thus would not explain why F_{ST} is greater for loci with non-neutral introgression.

Our results mirror the few studies that have examined the relationship between locus-specific differentiation and introgression in animals. In manakins (Parchman et al. 2013) and lycaenid butterflies (Gompert et al. 2012b), loci with non-neutral introgression also showed elevated differentiation in parental populations compared to neutral markers. In the house mouse hybrid zone, Janoušek et al. (2015) also observed higher differentiation for markers with reduced introgression, but contrary to our findings, loci with increased introgression showed lower differentiation compared to neutral markers. These results are consistent with the hypothesis that reproductive isolation arises as a byproduct of selection in allopatry, although we recognize that locus-specific differentiation and introgression are determined by the complex interaction of

many factors and that disentangling the effects of selection from other processes can be challenging (Beaumont & Balding 2004, Gompert et al. 2012c). For instance, high differentiation can occur in regions of reduced recombination due to low levels of within-species diversity possibly confounding any signals of divergent selection (Cruickshank & Hahn 2014). Therefore, elucidating the cause of elevated differentiation for loci with reduced introgression will provide key insight on the genetics of reproductive isolation.

Although differentiation between allopatric parental populations was greatest for loci with reduced introgression, it was not extremely high for all markers with reduced introgression (Figure 3.5B). Similarly, we observed neutral and increased introgression for markers with very high, moderate, and no differentiation (i.e. from F_{ST}=0 to F_{ST}=1) (Figure 3.5B). These observations are similar to those in the house mouse hybrid zone (Janoušek et al. 2015) and are likely a reflection of the complexity of the interaction between selection, drift and recombination and suggest that these forces vary across the genome. Despite the mechanism of divergence, this also demonstrates that high differentiation in allopatry is not a perfect predictor for reproductive isolation.

With the data presented here, it is not possible to quantify the contribution of drift to the divergence of loci associated with reproductive isolation. However, if the majority of loci associated with reproductive isolation in this system diverged via genetic drift, we might expect to see a similar distribution of locus-specific F_{ST} for loci with reduced introgression and those with neutral introgression. Instead, we observed a significantly greater mean F_{ST} for loci with reduced introgression (Figure 3.5B, Table A3.5). Further, it has been recognized that although it is theoretically possible, drift alone is unlikely to result in reproductive isolation (Turelli et al. 2001, Sobel et al. 2010). Phenotypes with the potential to be associated with reproductive

isolation (e.g., sterility/fertility phenotypes) are likely to be subject to selection within species and are thus not likely to be driven to fixation by drift.

Distinguishing the effects of divergent ecological selection from drift may be useful in understanding the role the environment played in shaping these species' evolutionary history. Some have concluded, however, that ecology is rarely, if ever, divorced from speciation and that multiple mechanisms likely contribute to and interact during the speciation process (Sobel et al. 2010, Templeton 2008), and, consequently, the idea that speciation occurs by either ecological divergence or drift is a false dichotomy (Sobel et al. 2010). Regardless, for this system, it will be necessary to take into consideration the possibility that any potential environmental differences encountered during divergence may differ from those currently encountered considering the estimated divergence time of 3 MA for these species (Cortés-Ortiz et al. 2003).

Evidence for a role of reinforcement

Reinforcement enhances barriers to reproduction between species and can act to complete the speciation process when partially isolated species come into contact after experiencing some divergence in allopatry (Servedio 2004, Butlin & Smadja 2018). We tested for a signature of reinforcement by comparing locus-specific differentiation between allopatric parental populations of *A. palliata* and *A. pigra* to the differentiation between backcrossed hybrids of each parental type (i.e., *A. palliata*-like and *A. pigra*-like backcrosses) in the hybrid zone for loci with reduced introgression. If reinforcing selection shaped loci with reduced introgression and thus contributed to reproductive isolation, we would expect to see greater differentiation in sympatry than in allopatry for these markers. Our results are consistent with this prediction.

There are at least two other mechanisms that may result in greater differentiation between sympatric than allopatric populations of the same species (e.g., Wang et al. 2014). First, strong genetic drift after independent range expansions of the parental species may result in greater differentiation in sympatry than in allopatry. However, effects of drift would be expected to have a genome-wide impact (e.g., Li et al. 2008), and we only observed a pattern of overall elevated differentiation in sympatry for loci with reduced introgression, which are expected to be associated with reproductive isolation. Second, it is plausible that greater differentiation in sympatry could result indirectly from independent local adaptation within each species to sites within their allopatric and sympatric ranges (i.e., mutation-order effects, Schluter 2009). However, the divergent alleles that underlie local adaptation in the hybrid zone may be expected to have neutral or increased introgression because such alleles should be advantageous on either species' genomic background (barring any involvement in hybrid incompatibilities). Thus, it seems likely our results reflect selection to increase reproductive isolation in sympatry under the extended view of reinforcement (Butlin & Smadja 2018). However, we still need to investigate if the loci driving this pattern underlie phenotypes under reinforcing selection.

We investigated mammalian phenotypes associated with genes occurring on contigs of the *Alouatta* genome containing loci with reduced introgression that represented the top 10% of those with the greatest difference between F_{ST} in sympatry and F_{ST} in allopatry. We found that some genes in these regions have been linked to mammalian phenotypes that could conceivably be under selection for prezygotic or postzygotic isolation in the hybrid zone (Table A3.6), thus contributing to the extended view of reinforcement (Butlin & Smadja 2018). For example, several genes are associated with the phenotype "abnormal behavior" (MP:0004924). In mice, one of these genes, the histone methyltransferase *KMT2A*, is known to play a role in complex

behaviors in mice including anxiety, nest-building behavior, spatial working memory, and learning (Gupta et al. 2010, Jakovcevski et al. 2015). In several taxa, learning is known to play a role in mate choice (e.g., sexual imprinting, learned avoidance of heterospecific mates), and thus can potentially be linked to prezygotic isolation (Servedio et al. 2009, Verzijden et al. 2012, Dukas 2013). Learning and memory have also been linked to postzygotic isolation since deficiencies in these traits can be selected against in hybrids (Rice & McQuillan 2018). Recently McQuillan et al. (2018) found that hybrid chickadees scored lower than parental chickadees in associative spatial learning and problem solving tasks. Learning and memory have been implicated in goal-oriented foraging behavior in Neotropical primates (Garber 1989, Janson 1998), traits presumably important for howler monkeys, which maintain a predominantly folivorous-frugivorous diet in highly diverse tropical forests where they adjust their dietary intake on seasonal availability of preferred foods (Raño et al. 2016). Thus, learning and memory deficiencies in hybrids, possibly mediated by *KMT2A*, may hinder foraging efforts potentially contributing to lower viability or fitness of howler monkey hybrids in their environment.

Many genes are also associated with abnormal embryonic/fetal development or lethality phenotypes (e.g., MP:0001672, MP:0011092, MP:0010865, MP:0011101). It is possible that incompatible alleles between the parental species in these genes contribute to postzygotic isolation in this system. The contig with the greatest difference in F_{ST} between sympatry and allopatry annotated using our framework contains *SHROOM3*, a gene that encodes a PDZ domain-containing protein. In mice, *SHROOM3* mutant embryos suffer severe neural tube defects resulting in perinatal death (Hildebrand & Soriano 1999). Although our functional annotation results offer some plausible genetic mechanisms that could be involved in reinforcement, they should be interpreted with caution. First, it is not known how mutations in

any of these genes affect phenotypes in howler monkeys or what role these genes may play in reproductive isolation in the *Alouatta* hybrid zone. Further, it is not known whether incompatibilities associated with the genes identified here are directly driving reduced introgression of our loci, or are physically linked to causal variants not sequenced in our reduced-representation library. Similar analyses using whole genome sequence data (e.g., Rafati et al. 2018) will be a valuable step to better associate loci driving reduced introgression of genomic regions with potential functions and phenotypes under selection for pre- and postzygotic reproductive isolation in this system.

Although we did not measure prezygotic isolation with respect to any phenotype, we suspect that there are many traits beyond the phenotypes identified above that reinforcement could potentially be acting upon in this system. Specifically, traits known to be associated with mating behavior in howler monkeys (and thus may have potential involvement with prezygotic isolation) include olfactory cues in urine and other scent markings that males likely use to detect female sexual receptivity (Glander 1980, Horwich 1983), and behavioral displays of sexual solicitation and mate guarding (Glander 1980, Horwich 1983, Van Belle et al. 2009). Color traits are known to be used in mate discrimination in other animal species (Hill 1991, Seehausen & van Alphen 1998, Jiggins et al. 2001, Waitt et al. 2003) and some have hypothesized that sexual selection on coat color has shaped female choice of mates in howler monkeys, as it may signal a male's competitive ability, health status, maturity, etc. (Crockett 1987, Bicca-Marques & Calegaro-Marques 1998). Similarly, traits that might influence the outcome of competition between males for access to females may shape the dynamics of heterospecific copulation in the hybrid zone (and thus prezygotic isolation). Such traits include body size, canine length, and testis volume (Kelaita et al. 2011), as well as the loud roaring vocalizations for which howler

monkeys are known (Kowalewski & Garber 2010, Holzmann et al. 2012, Van Belle et al. 2014, Kitchen et al. 2015). In order to test any of these hypotheses, it will be necessary to quantify these traits and compare the characteristics of sympatric and allopatric individuals with the prediction that if reinforcement has occurred, trait differences between the species will be more pronounced in sympatry than in allopatry (i.e., reproductive character displacement; although reinforcement does not always produce a signature of reproductive character displacement, Servedio 2004). In this study, we did not directly measure selection against hybridization or prezygotic isolation, let alone any potential phenotypes under reinforcing selection. More research will be necessary in order to connect the pattern we observed here, greater divergence in sympatry than in allopatry for loci associated with reproductive isolation, with conclusive evidence for reinforcing selection on any mating discrimination trait. Regardless, in the genomic era, scans similar to the one employed in this study may enhance efforts to understand the frequency and importance of reinforcement in the speciation process by providing a means to detect the signature of reinforcement in the genome.

Conclusion

We identified a subset of genomic markers with reduced introgression in a natural primate hybrid zone, suggesting an association with reproductive isolation. These markers were more differentiated between allopatric parental populations than neutral loci and loci with increased introgression, consistent with the idea that reproductive isolation is a byproduct of divergence in allopatry. These markers also showed a signature of reinforcement, suggesting that reproductive isolation may have initially been driven by divergence in allopatry, but reinforced

by divergent selection in sympatry. These results reflect the contribution of different selective processes that have shaped the evolution of reproductive isolation in this system.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics,* **16**, 113–148.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Bergman, T. J., Cortés-Ortiz, L., Dias, P. A. D., Ho, L., Adams, D., Canales-Espinosa, D., & Kitchen, D. M. (2016). Striking differences in the loud calls of howler monkey sister species (*Alouatta pigra* and *A. palliata*). *American Journal of Primatology*, **78**, 755–766.

Bicca-Marques, J. C., & Calegaro-Marques, C. (1998). Behavioral thermoregulation in a sexually and developmentally dichromatic neotropical primate, the black-and-gold howling monkey (*Alouatta caraya*). *American Journal of Physical Anthropology*, **106**, 533–546.

Butlin, R. (1987). Speciation by reinforcement. *Trends in Ecology & Evolution*, **2**, 8–13.

Butlin, R. K., & Smadja, C. M. (2018). Coupling, Reinforcement, and Speciation. *The American Naturalist*, **191**, 155–172.

Chapman, C. A., & Balcomb, S. R. (1998). Population characteristics of howlers: ecological conditions or group history. *International Journal of Primatology*, **19**, 385–403.

Cortés-Ortiz, L., Bermingham, E., Rico, C., Rodrıguez-Luna, E., Sampaio, I., & Ruiz-Garcıa, M. (2003). Molecular systematics and biogeography of the Neotropical monkey genus, *Alouatta*. *Molecular Phylogenetics and Evolution*, **26**, 64–81.

Cortés-Ortiz, L., Duda, T. F., Canales-Espinosa, D., García-Orduña, F., Rodríguez-Luna, E., & Bermingham, E. (2007). Hybridization in large-bodied New World primates. *Genetics,* **176**, 2421–2425.

Cortés-Ortiz, L., Agostini, I., Aguiar, L. M., Kelaita, M., Silva, F. E., & Bicca-Marques, J. C. (2015). Hybridization in howler monkeys: current understanding and future directions. In

Kowalewski, M., Garber, P., Cortés-Ortiz, L., Urbani, B., Youlatos, D. (Eds.), *Howler Monkeys: Adaptive Radiation, Systematics, and Morphology* (pp. 107–131). New York, NY: Springer.

Cortés-Ortiz, L., Nidiffer, M. D., Hermida-Lagunes, J., García-Orduña, F., Rangel-Negrín, A., Kitchen, D. M., … Canales-Espinosa, D. (in press). Reduced introgression of sex chromosome markers in the Mexican howler monkey (*Alouatta palliata* x *A. pigra*) hybrid zone. *International Journal of Primatology*.

Coyne, J. A., & Orr, H. A. (1989). Two rules of speciation. In Otte, D., Endler, J. (Eds.), *Speciation and its Consequences* (pp. 180–207). Sunderland, MA: Sinauer Associates.

Crockett, C. M. (1987). Diet, dimorphism and demography: perspectives from howlers to hominids. In Kinzey, W. G. (Ed.), *The Evolution of Human Behavior: Primate Models* (pp. 115–135). Albany, NY: SUNY Press.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 156–158.

de Oliveira, E. H. C., Neusser, M., & Müller, S. (2012). Chromosome evolution in new world monkeys (Platyrrhini). *Cytogenetic and Genome Research*, **137**, 259–272.

Delgado, C. L. R., Waters, P. D., Gilbert, C., Robinson, T. J., & Graves, J. A. M. (2009). Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Research*, **17**, 917–926.

Dodd, D. (1989). Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. *Evolution*, **43**, 1308–1311.

Dukas, R. (2013). Effects of learning on evolution: robustness, innovation and speciation. *Animal Behaviour*, **85**, 1023–1030.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **4**, 1184–1191.

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Ellsworth, J. A., & Hoelzer, G. A. (2006). Genetic evidence on the historical biogeography of Central American howler monkeys. In Lehman, S. M, Fleagle, J. G. (Eds.), *Primate Biogeography* (pp. 81–103). New York, NY: Springer.

Ford, S. M. (2006). The biogeographic history of Mesoamerican primates. In Estrada, A., Garber, P. A., Pavelka, M. S. M., Luecke, L. (Eds.), *New Perspectives in the Study of Mesoamerican Primates* (pp. 81–114). Boston, MA: Springer.

Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National academy of Sciences of the United States of America*, **103**, 3209–3213.

Garber, P. A. (1989). Role of spatial memory in primate foraging patterns: *Saguinus mystax* and *Saguinus fuscicollis*. *American Journal of Primatology*, **19**, 203–216.

Glander, K. E. (1980). Reproduction and population growth in free-ranging mantled howling monkeys. *American Journal of Physical Anthropology*, **53**, 25–36.

Gompert, Z., & Buerkle, C. (2011a). Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert, Z., & Buerkle, C. A. (2011b). A hierarchical Bayesian model for next-generation population genomics. *Genetics*, **187**, 903–917.

Gompert, Z., & Buerkle, C. A. (2012a). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012b). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012c). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 439–450.

Gupta, S., Kim, S. Y., Artis, S., Molfese, D. L., Schumacher, A., Sweatt, J. D., ... & Lubin, F. D. (2010). Histone methylation regulates memory formation. *Journal of Neuroscience*, **30**, 3589–3599.

Hildebrand, J. D., & Soriano, P. (1999). Shroom, a PDZ domain–containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell*, **99**, 485–497.

Hill, G. E. (1991). Plumage coloration is a sexually selected indicator of male quality. *Nature,* **350**, 337–339.

Ho, L., Cortés-Ortiz, L., Dias, P. A. D., Canales-Espinosa, D., Kitchen, D. M., & Bergman, T. J. (2014). Effect of ancestry on behavioral variation in two species of howler monkeys (*Alouatta pigra* and *A. palliata*) and their hybrids. *American Journal of Primatology*, **76**, 855–867.

Holzmann, I., Agostini, I., & Di Bitetti, M. (2012). Roaring behavior of two syntopic howler species (*Alouatta caraya* and *A. guariba clamitans*): evidence supports the mate defense hypothesis. *International Journal of Primatology*, **33**, 338–355.

Horwich, R. H. (1983). Breeding behaviors in the black howler monkey (*Alouatta pigra*) of Belize. *Primates*, **24**, 222–230.

Janson, C. H. (1998). Experimental evidence for spatial memory in foraging wild capuchin monkeys, *Cebus apella*. *Animal Behaviour*, **55**, 1229–1243.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., & Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, **411**, 302–305.

Jakovcevski, M., Ruan, H., Shen, E. Y., Dincer, A., Javidfar, B., Ma, Q., ... Lin, C. L. (2015). Neuronal Kmt2a/Mll1 histone methyltransferase is essential for prefrontal synaptic plasticity and working memory. *Journal of Neuroscience*, **35**, 5097–5108.

Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, **32**, 1208–1220.

Kelaita, M., Dias, P. A. D., Aguilar-Cucurachi, M., Canales-Espinosa, D., & Cortés-Ortiz, L. (2011). Impact of intrasexual selection on sexual dimorphism and testes size in the Mexican howler monkeys *Alouatta palliata* and *A. pigra*. *American Journal of Physical Anthropology*, **146**, 179–187.

Kelaita, M. A., & Cortés-Ortiz, L. (2013). Morphological variation of genetically confirmed *Alouatta pigra* × *A. palliata* hybrids from a natural hybrid zone in Tabasco, Mexico. *American Journal of Physical Anthropology*, **150**, 223–234.

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, **12**, 656–664.

Kilias, G., Alahiotis, S. N., & Pelecanos, M. (1980). A multifactorial genetic investigation of speciation theory using *Drosophila melanogaster*. *Evolution*, **34**, 730–737.

Kitchen, D. M., da Cunha, R. G. T., Holzmann, I., & de Oliveira, D. A. G. (2015). Function of loud calls in howler monkeys. In Kowalewski, M., Garber, P., Cortés-Ortiz, L., Urbani, B., Youlatos, D. (Eds.), *Howler Monkeys: Adaptive Radiation, Systematics, and Morphology* (pp. 369–399). New York, NY: Springer.

Kowalewski, M. M., & Garber, P. A. (2010). Mating promiscuity and reproductive tactics in female black and gold howler monkeys (*Alouatta caraya*) inhabiting an island on the Parana River, Argentina. *American Journal of Primatology*, **72**, 734–748.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997*.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., ... Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **319**, 1100–1104.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS biology*, **5**, e243.

McQuillan, M. A., Roth, T. C., Huynh, A. V., & Rice, A. M. (2018). Hybrid chickadees are deficient in learning and memory. *Evolution*, **72**, 1155–1164.

Mueller, J. L., Skaletsky, H., Brown, L. G., Zaghlul, S., Rock, S., Graves, T., ... Page, D. C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics*, **45**, 1083–1087.

Nosil, P., Parchman, T. L., Feder, J. L., & Gompert, Z. (2012a). Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in *Timema* stick insects. *BMC Evolutionary Biology*, **12**, 164.

Nosil, P., Gompert, Z., Farkas, T. E., Comeault, A. A., Feder, J. L., Buerkle, C. A., & Parchman, T. L. (2012b). Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B,* **279**, 5058–5065.

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., ... Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rafati, N., Blanco-Aguiar, J. A., Rubin, C. J., Sayyab, S., Sabatino, S. J., Afonso, S., ... & Andersson, L. (2018). A genomic map of clinal variation across the European rabbit hybrid zone. *Molecular Ecology*, **27**, 1457–1478.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Raño, M., Kowalewski, M. M., Cerezo, A. M., & Garber, P. A. (2016). Determinants of daily path length in black and gold howler monkeys (*Alouatta caraya*) in northeastern Argentina. *American Journal of Primatology*, **78**, 825–837.

Rice, A. M., & McQuillan, M. A. (2018). Maladaptive learning and memory in hybrids as a reproductive isolating barrier. *Proceedings of the Royal Society B*, **285**, 20180542.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Scordato, E. S., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology,* **26**, 5676–5691.

Seehausen, O., & van Alphen, J. J. (1998). The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioral Ecology and Sociobiology*, **42**, 1–8.

Servedio, M. R., & Noor, M. A. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 339–364.

Servedio, M. R. (2004). The what and why of research on reinforcement. *PLoS Biology*, **2**, e420.

Servedio, M. R., Sæther, S. A., & Sætre, G. P. (2009). Reinforcement and learning. *Evolutionary Ecology*, **23**, 109–123.

Smith, J. D. (1970). The systematic status of the black howler monkey, *Alouatta pigra* Lawrence. *Journal of Mammalogy*, **51**, 358–369.

Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, **64**, 295–315.

Solari, A. J., & Rahn, M. I. (2005). Fine structure and meiotic behaviour of the male multiple sex chromosomes in the genus *Alouatta*. *Cytogenetic and Genome Research*, **108**, 262–267.

Steinberg, E. R., Cortés-Ortiz, L., Nieves, M., Bolzán, A. D., García-Orduña, F., Hermida-Lagunes, J., ... Mudry, M. D. (2008). The karyotype of *Alouatta pigra* (Primates: Platyrrhini): mitotic and meiotic analyses. *Cytogenetic and Genome Research*, **122**, 103–109.

Steinberg, E. R., Nieves, M., & Mudry, M. D. (2014). Multiple sex chromosome systems in howler monkeys (Platyrrhini, *Alouatta*). *Comparative Cytogenetics*, **8**, 43–69.

Stephens, J. C., Briscoe, D., & O'Brien, S. J. (1994). Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *American Journal of Human Genetics*, **55**, 809–824.

Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.

Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. *Bioessays*, **30**, 470–479.

Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C., & Eicher, E. M. (1992). Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution*, **46**, 1146–1163.

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution*, **16**, 330–343.

Van Belle, S., Estrada, A., Ziegler, T. E., & Strier, K. B. (2009). Sexual behavior across ovarian cycles in wild black howler monkeys (*Alouatta pigra*): male mate guarding and female mate choice. *American Journal of Primatology*, **71**, 153–164.

Van Belle, S., Estrada, A., & Garber, P. A. (2014). The function of loud calls in black howler monkeys (*Alouatta pigra*): food, mate, or infant defense? *American Journal of Primatology*, **76**, 1196–1206.

Verardi, A., Lucchini, V., & Randi, E. (2006). Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Molecular Ecology*, **15**, 2845–2855.

Verzijden, M. N., ten Cate, C., Servedio, M. R., Kozak, G. M., Boughman, J. W., & Svensson, E. I. (2012). The impact of learning on sexual selection and speciation. *Trends in Ecology & Evolution*, **27**, 511–519.

Waitt, C., Little, A. C., Wolfensohn, S., Honess, P., Brown, A. P., Buchanan-Smith, H. M., & Perrett, D. I. (2003). Evidence from rhesus macaques suggests that male coloration plays a role

in female primate mate choice. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, S144–S146.

Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkova, P., Vyskočilová, M. M., … Pialek, J. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology*, **20**, 2985–3000.

Wang, J., Abbott, R. J., Ingvarsson, P. K., & Liu, J. (2014). Increased genetic divergence between two closely related fir species in areas of range overlap. *Ecology and Evolution*, **4**, 1019–1029.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., ... Gil, L. (2017). Ensembl 2018. *Nucleic Acids Research*, **46**, D754–D761.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

APPENDIX A3

Table A3.1. Number of loci and individuals retained in each data set after filtering. Apa=allopatric *A. palliata*, HZ=hybrid zone, Api=allopatric *A. pigra*.

Data set	Loci	SNPs	Indels	Individuals	Apa	HZ	Api	Filters
Raw data	6,415,368	5,766,502	648,866	181	38	99	44	Removed SNPs within 5bp of an indel and variants with quality score <20
fastStructrue, PCA	74,448	74,448	0	158	36	81	41	Same as raw data set, plus removed non-biallelic loci, indels, loci with minor allele frequency ≤0.01, and sites with minimum mean depth of <10 across individuals, sites out of HWE in either allopatric parental population, and thinned sites within 200bp
bgc, F_{ST}	5,763	5,222	541	147	34	81	32	Same as raw data set, plus removed non-biallelic loci, loci with a minor allele frequency ≤0.05, sites with minimum mean depth of <30 across individuals, and sites that were absent in >20% of individuals in either allopatric parental population

Table A3.2. Details for individuals we determined to be admixed that were sampled outside of the contact zone. Q_I=admixture proportion, *SRY* (Y-linked marker) haplotype, and Ham80 (X-linked marker) microsatellite genotype, and mtDNA haplotype. *SRY*, Ham80, and mtDNA data are from Cortés-Ortiz et al. (2019). Api=*A. pigra* type, Apa=*A. palliata* type.

Locality	Sample ID	Sex	Q_I	*SRY*	Ham80	mtDNA
Ver	S003	F	0.06	na	Apa/Apa	Apa
Ver	S209	F	0.17	na	Apa/Apa	Apa
Cam	S283	F	0.92	na	Api/Api	Api
Cam	S284	F	0.63	na	Api/Api	Api
Cam	S228	F	0.81	na	Api/Api	Api
Cam	S231	F	0.28	na	Api/Api	Api
Cam	S220	F	0.81	na	Api/Api	Api
Cam	S514	F	0.38	na	Api/Api	Api
Cam	S518	M	0.39	Api	Api	Api
Cam	S245	F	0.85	na	Api/Api	Api
QR	S192	M	0.89	Api	Api	Api

Table A3.3. BLAT search results against the human genome (GRCh38/hg38) for the first and last 25kb of sequence in contigs of the *Alouatta* genome assembly (accession ID PVKV00000000) that contain the top 10% of loci with reduced introgression that show the greatest difference in F_{ST} between sympatry and allopatry (F_{ST} diff). N is the number of ddRAD loci per contig, S1 and S2 are matching strands in the human assembly, Length human range is the number of nucleotides between the outermost coordinates of BLAT results for the first and last 25kb, BiomaRt query is the human genomic coordinates used to identify human genes within each range, and N HG is the number of human genes within each range retrieved from BiomaRt.

Alouatta Contig	Contig length (bp)	F_{ST} diff	N	BLAT human chromosome first 25kb	S1	BLAT human chromosome last 25kb	S2	Length human range (bp)	BiomaRt query	N HG
flattened_line_18002	92,090	0.46	1	chr7:3,943,431-3,968,534	+	chr7:4,012,579-4,041,355	+	97,924	7:3943431:4041355	1
flattened_line_5159	203,373	0.45	1	chr8:37,127,197-37,152,308	-	chr8:36,938,986-36,979,670	-	213,322	8:36938986:37152308	0
flattened_line_935	372,272	0.44	1	chr4:76,528,855-76,551,847	-	chr4:76,185,427-76,207,572	-	366,420	4:76185427:76551847	6
flattened_line_3619	239,471	0.42	1	chr3:9,583,310-9,645,137	+	chr3:9,841,400-10,010,823	+	427,513	3:9583310:10010823	17
flattened_line_7652	166,002	0.33	1	chr16:31,362,282-31,384,551	+	chr16:31,506,364-31,556,497	+	194,215	16:31362282:31556497	9
flattened_line_18038	91,979	0.32	1	chr3:94,011,037-94,019,161	-	chr3:93,931,346-93,962,898	-	87,815	3:93931346:94019161	3
flattened_line_7514	167,656	0.27	2	chr4:136,574,471-136,623,075	+	chr4:136,748,530-136,774,476	+	200,005	4:136574471:136774476	0
flattened_line_9274	147,816	0.25	1	chr11:118,438,183-118,461,242	-	chr11:118,316,079-118,335,411	-	145,163	11:118316079:118461242	6

Table A3.4. Mean admixture proportion Q_1 scores across ten replicate *fastStrucutre* runs at K=2. Individuals are arranged by longitude as in Figure 3.2, from East (top) to West (bottom).

ID	Locality	Mean Q_1	SD
S087	Ver	0.000006	0.00E+00
S088	Ver	0.000005	0.00E+00
S089	Ver	0.000004	4.22E-07
S090	Ver	0.000005	8.43E-07
S091	Ver	0.000005	0.00E+00
S092	Ver	0.000005	0.00E+00
S142	Ver	0.000005	1.32E-06
S143	Ver	0.000004	0.00E+00
S001	Ver	0.018777	5.27E-06
S002	Ver	0.010092	4.42E-06
S003	Ver	0.064615	6.29E-06
S004	Ver	0.000006	0.00E+00
S005	Ver	0.000005	1.51E-06
S144	Ver	0.003320	4.18E-06
S145	Ver	0.000006	0.00E+00
S146	Ver	0.000004	3.16E-07
S147	Ver	0.025385	5.53E-06

S204	Ver	0.011731	3.71E-06
S205	Ver	0.000004	0.00E+00
S207	Ver	0.000007	0.00E+00
S208	Ver	0.000538	1.71E-04
S209	Ver	0.167801	8.32E-06
S172	Ver	0.000006	0.00E+00
S173	Ver	0.000007	2.84E-06
S174	Ver	0.000005	3.16E-07
S175	Ver	0.000005	0.00E+00
S176	Ver	0.000005	1.32E-06
S613	Ver	0.002811	4.75E-05
S614	Ver	0.000004	6.99E-07
S615	Ver	0.000006	1.81E-06
S618	Ver	0.000004	0.00E+00
S608	Ver	0.000004	4.22E-07
S609	Ver	0.000006	2.21E-06
S610	Ver	0.015049	4.79E-06
S611	Ver	0.000006	2.21E-06
S612	Ver	0.000004	0.00E+00
S010	Tab	0.017828	4.19E-06
S334	Tab	0.000005	0.00E+00
S335	Tab	0.000005	0.00E+00
S336	Tab	0.000005	0.00E+00
S337	Tab	0.203766	8.97E-06
S339	Tab	0.000005	0.00E+00
S340	Tab	0.000004	0.00E+00
S341	Tab	0.000004	3.16E-07
S164	Tab	0.800458	1.76E-05
S165	Tab	0.926332	2.00E-05
S293	Tab	0.005780	4.10E-06
S296	Tab	0.000006	0.00E+00
S625	Tab	0.000005	0.00E+00
S629	Tab	0.000006	0.00E+00
S630	Tab	0.000006	0.00E+00
S411	Tab	0.052870	5.64E-06
S432	Tab	0.028693	4.63E-06
S434	Tab	0.012284	4.62E-06
S437	Tab	0.000006	0.00E+00
S578	Tab	0.000005	0.00E+00
S586	Tab	0.016930	1.10E-05
S588	Tab	0.003233	7.71E-06
S589	Tab	0.002384	1.41E-05
S590	Tab	0.007959	5.10E-06
S592	Tab	0.428938	1.25E-05
S593	Tab	0.000006	0.00E+00
S544	Tab	0.979716	1.98E-05
S545	Tab	0.970032	1.97E-05
S551	Tab	0.934743	1.82E-05
S553	Tab	0.000006	0.00E+00
S555	Tab	0.004910	6.08E-06
S559	Tab	0.033319	7.60E-06
S561	Tab	0.000006	0.00E+00
S562	Tab	0.000106	6.24E-05

S536	Tab	0.000006	0.00E+00
S538	Tab	0.000006	0.00E+00
S532	Tab	0.959883	1.90E-05
S103	Tab	0.000328	1.80E-04
S158	Tab	0.000005	0.00E+00
S160	Tab	0.000008	0.00E+00
S161	Tab	0.009625	3.75E-06
S162	Tab	0.014718	4.06E-06
S599	Tab	0.948552	1.69E-05
S601	Tab	0.958249	1.75E-05
S407	Tab	0.972485	2.63E-05
S408	Tab	0.999733	1.84E-04
S409	Tab	0.974402	2.25E-05
S410	Tab	0.951865	2.18E-05
S101	Tab	0.059595	6.09E-06
S093	Tab	0.006061	4.02E-06
S094	Tab	0.000005	0.00E+00
S439	Tab	0.880956	1.75E-05
S596	Tab	0.062070	5.51E-06
S163	Tab	0.964544	2.34E-05
S308	Tab	0.992908	3.20E-05
S309	Tab	0.492985	1.35E-05
S326	Tab	0.000005	0.00E+00
S097	Tab	0.605330	1.55E-05
S099	Tab	0.017105	4.23E-06
S310	Tab	0.461529	1.34E-05
S311	Tab	0.000005	0.00E+00
S305	Tab	0.948270	1.88E-05
S307	Tab	0.956388	2.02E-05
S401	Tab	0.430550	1.24E-05
S177	Tab	0.942184	2.02E-05
S301	Tab	0.937896	2.03E-05
S470	Tab	0.435858	1.26E-05
S471	Tab	0.000012	0.00E+00
S472	Tab	0.000005	0.00E+00
S473	Tab	0.080627	6.54E-06
S474	Tab	0.000006	0.00E+00
S476	Tab	0.000005	0.00E+00
S477	Tab	0.048104	5.65E-06
S479	Tab	0.000009	0.00E+00
S313	Tab	0.969246	1.70E-05
S315	Tab	0.960596	1.77E-05
S316	Tab	0.927754	1.76E-05
S465	Tab	0.745545	1.72E-05
S182	Tab	0.841754	1.85E-05
S184	Tab	0.810838	1.86E-05
S015	Tab	0.959593	1.79E-05
S283	Cam	0.919731	1.73E-05
S284	Cam	0.625811	1.54E-05
S274	Cam	0.999995	0.00E+00
S279	Cam	0.999994	0.00E+00
S234	Cam	0.999995	0.00E+00
S289	Cam	0.980676	1.53E-05

S242	Cam	0.999995	0.00E+00
S228	Cam	0.814565	1.83E-05
S229	Cam	0.964763	1.67E-05
S231	Cam	0.283298	9.56E-06
S216	Cam	0.990169	1.56E-05
S219	Cam	0.999993	0.00E+00
S220	Cam	0.805036	1.73E-05
S511	Cam	0.999995	0.00E+00
S514	Cam	0.382904	1.14E-05
S518	Cam	0.387363	1.13E-05
S526	Cam	0.999995	0.00E+00
S149	Cam	0.999995	0.00E+00
S245	Cam	0.850562	1.79E-05
S151	Cam	0.999995	0.00E+00
S153	Cam	0.999995	0.00E+00
S248	Cam	0.999995	0.00E+00
S504	DG	0.999993	0.00E+00
S505	DG	0.961795	1.57E-05
S506	DG	0.999984	0.00E+00
S507	DG	0.999994	0.00E+00
S508	DG	0.999995	0.00E+00
S509	DG	0.999993	0.00E+00
S510	DG	0.999993	0.00E+00
S185	QR	0.999992	5.16E-07
S186	QR	0.999995	0.00E+00
S187	QR	0.999995	0.00E+00
S188	QR	0.998290	3.39E-04
S189	QR	0.999995	0.00E+00
S190	QR	0.999995	0.00E+00
S191	QR	0.999993	0.00E+00
S192	QR	0.885564	1.69E-05
S193	QR	0.999993	0.00E+00
S194	QR	0.999994	0.00E+00
S195	QR	0.989286	1.84E-05
S196	QR	0.999131	2.80E-04

Table A3.5. Results of the Tukey HSD post hoc test, showing that F_{ST} between allopatric parental populations is significantly different for loci in each beta category.

Comparison	**Difference**	**Lower**	**Upper**	**Adjusted P**
$\beta>0 - \beta<0$	0.07	0.01	0.14	0.016
$\beta=0 - \beta<0$	-0.15	-0.19	-0.10	<0.001
$\beta=0 - \beta>0$	-0.22	-0.27	-0.17	<0.001

Table A3.6. Mammalian phenotypes associated with the top 10% of loci with the greatest difference in F_{ST} between sympatry and allopatric parental populations. Contig=*Alouatta* assembly contig, FST_{sym}-FST_{allo}=difference between F_{ST} in sympatry and F_{ST} in allopatry, Gene=HGNC symbol, MGI Gene/Marker ID=Mouse Genome Informatics identifier, Name=gene name, MP ID=mouse phenotype identifier, Term=MP definition, RI=tentative type of selection in hybrid zone (pre=prezygotic, post=postzygotic).

Contig	FST_{sym}-FST_{allo}	Gene	MGI Gene/Marker ID	Name	MP ID	Term	RI
18002	0.46	SDK1	MGI:2444413	sidekick cell adhesion molecule 1			
935	0.44	CCDC158	MGI:2444555	coiled-coil domain containing 158			
		FAM47E	MGI:2686227	family with sequence similarity 47, member E	MP:0012768	decreased KLRG1-positive NK cell number	Post
		FAM47E-STBD1	No associated gene				
		SCARB2	MGI:1196458	scavenger receptor class B, member 2	MP:0004738	abnormal auditory brainstem response	Pre/post
					MP:0004924	abnormal behavior	
					MP:0004736	abnormal distortion product otoacoustic emission	
					MP:0005584	abnormal enzyme/coenzyme activity	
					MP:0003632	abnormal nervous system morphology	
					MP:0002882	abnormal neuron morphology	
					MP:0002895	abnormal otolithic membrane morphology	
					MP:0001078	abnormal phrenic nerve morphology	
					MP:0001106	abnormal Schwann cell morphology	
					MP:0004366	abnormal strial marginal cell morphology	
					MP:0000048	abnormal stria vascularis morphology	
					MP:0004368	abnormal stria vascularis vasculature morphology	
					MP:0000534	abnormal ureter morphology	
					MP:0011487	abnormal ureteropelvic junction morphology	
					MP:0011426	abnormal ureter smooth muscle morphology	
					MP:0003630	abnormal urothelium morphology	

			MP:0004438	abnormal vestibular hair cell physiology	
			MP:0004410	absent endocochlear potential	
			MP:0002978	absent otoliths	
			MP:0006359	absent startle reflex	
			MP:0002871	albuminuria	
			MP:0003354	astrocytosis	
			MP:0002857	cochlear ganglion degeneration	
			MP:0004398	cochlear inner hair cell degeneration	
			MP:0004404	cochlear outer hair cell degeneration	
			MP:0001967	deafness	
			MP:0001262	decreased body weight	
			MP:0004411	decreased endocochlear potential	
			MP:0003870	decreased urine glucose level	
			MP:0002988	decreased urine osmolality	
			MP:0000921	demyelination	
			MP:0011305	dilated kidney calyx	
			MP:0001270	distended abdomen	
			MP:0000519	hydronephrosis	
			MP:0011418	leukocyturia	
			MP:0001513	limb grasping	
			MP:0008918	microgliosis	
			MP:0002229	neurodegeneration	
			MP:0000753	paralysis	
			MP:0001426	polydipsia	
			MP:0001762	polyuria	
			MP:0002083	premature death	
			MP:0004740	sensorineural hearing loss	
			MP:0004363	stria vascularis degeneration	
			MP:0000745	tremors	
			MP:0011491	ureteropelvic junction obstruction	
SHROOM3	MGI:1351655	shroom family member 3	MP:0003385	abnormal body wall morphology	Post
			MP:0001672	abnormal embryo development	
			MP:0000075	absent neurocranium	
			MP:0000914	exencephaly	
			MP:0008797	facial cleft	
			MP:0005155	herniated intestine	
			MP:0010912	herniated liver	

					MP:0000108	midline facial cleft	
					MP:0000929	open neural tube	
					MP:0011089	perinatal lethality, complete penetrance	
					MP:0000930	wavy neural tube	
		STBD1	MGI:1261768	starch binding domain 1			
3619	0.42	ARPC4	MGI:1915339	actin related protein 2/3 complex, subunit 4	MP:0002074	abnormal hair texture	Pre/post
					MP:0010088	decreased circulating fructosamine level	
					MP:0005567	decreased circulating total protein level	
					MP:0000218	increased leukocyte cell number	
					MP:0002590	increased mean corpuscular volume	
					MP:0002169	no abnormal phenotype detected	
		ARPC4-TTLL3	No associated gene				
		BRPF1	MGI:1926033	bromodomain and PHD finger containing, 1	MP:0004924	abnormal behavior	Pre/post
					MP:0004045	abnormal cell cycle checkpoint function	
					MP:0005621	abnormal cell physiology	
					MP:0000780	abnormal corpus callosum morphology	
					MP:0000812	abnormal dentate gyrus morphology	
					MP:0014134	abnormal embryo morphology	
					MP:0002761	abnormal hippocampal mossy fiber morphology	
					MP:0008263	abnormal hippocampus CA1 region morphology	
					MP:0000808	abnormal hippocampus development	
					MP:0000807	abnormal hippocampus morphology	
					MP:0009939	abnormal hippocampus neuron morphology	
					MP:0008284	abnormal hippocampus pyramidal cell layer	
					MP:0008262	abnormal hippocampus region morphology	
					MP:0008547	abnormal neocortex morphology	
					MP:0001447	abnormal nest building behavior	
					MP:0006009	abnormal neuronal migration	
					MP:0012203	abnormal neuronal stem cell morphology	

			MP:0003648	abnormal radial glial cell morphology	
			MP:0001265	decreased body size	
			MP:0001262	decreased body weight	
			MP:0000063	decreased bone mineral density	
			MP:0000781	decreased corpus callosum size	
			MP:0012460	decreased dentate gyrus size	
			MP:0011940	decreased food intake	
			MP:0020069	decreased neocortex size	
			MP:0004981	decreased neuronal precursor cell number	
			MP:0020394	decreased neuronal precursor proliferation	
			MP:0020388	decreased radial glial cell number	
			MP:0005018	decreased T cell number	
			MP:0001730	embryonic growth arrest	
			MP:0011092	embryonic lethality, complete penetrance	
			MP:0011098	embryonic lethality during organogenesis, complete penetrance	
			MP:0011108	embryonic lethality during organogenesis, incomplete penetrance	
			MP:0000220	increased monocyte cell number	
			MP:0020536	increased subiculum size	
			MP:0001513	limb grasping	
			MP:0003631	nervous system phenotype	
			MP:0002169	no abnormal phenotype detected	
			MP:0011086	postnatal lethality, incomplete penetrance	
			MP:0010865	prenatal growth retardation	
			MP:0011101	prenatal lethality, incomplete penetrance	
			MP:0008441	thin cortical plate	
CAMK1	MGI:1098535	calcium/calmodulin-dependent protein kinase I	MP:0003049	abnormal lumbar vertebrae morphology	Post
			MP:0013278	decreased fasted circulating glucose level	
			MP:0003019	increased circulating chloride level	
			MP:0002965	increased circulating serum albumin level	
CIDEC	MGI:95585	cell death-inducing	MP:0009642	abnormal blood homeostasis	Pre/post

	DFFA-like effector c		
			abnormal brown adipose tissue
		MP:0002971	morphology
		MP:0009115	abnormal fat cell morphology
		MP:0005334	abnormal fat pad morphology
		MP:0005335	abnormal gonadal fat pad morphology
		MP:0002970	abnormal white adipose tissue morphology
		MP:0002702	decreased circulating free fatty acid level
		MP:0005560	decreased circulating glucose level
		MP:0003458	decreased circulating ketone body level
		MP:0005668	decreased circulating leptin level
		MP:0002644	decreased circulating triglyceride level
		MP:0009283	decreased gonadal fat pad weight
		MP:0009293	decreased inguinal fat pad weight
		MP:0009299	decreased mesenteric fat pad weight
		MP:0009305	decreased retroperitoneal fat pad weight
		MP:0008844	decreased subcutaneous adipose tissue amount
		MP:0003213	decreased susceptibility to age related obesity
		MP:0005659	decreased susceptibility to diet-induced obesity
		MP:0002310	decreased susceptibility to hepatic steatosis
		MP:0010025	decreased total body fat amount
		MP:0001783	decreased white adipose tissue amount
		MP:0008034	enhanced lipolysis
		MP:0011049	impaired adaptive thermogenesis
		MP:0005292	improved glucose tolerance
		MP:0005533	increased body temperature
		MP:0004889	increased energy expenditure
		MP:0011939	increased food intake
		MP:0002891	increased insulin sensitivity
		MP:0005289	increased oxygen consumption
		MP:0010378	increased respiratory quotient
		MP:0010771	integument phenotype

			MP:0002169	no abnormal phenotype detected	
			MP:0005391	vision/eye phenotype	
CPNE9	MGI:2443052	copine family member IX			
CRELD1	MGI:2152539	cysteine-rich with EGF-like domains 1	MP:0000428	abnormal craniofacial morphology	Post
			MP:0003232	abnormal forebrain development	
			MP:0006108	abnormal hindbrain development	
			MP:0003864	abnormal midbrain development	
			MP:0000263	absent organized vascular network	
			MP:0000301	decreased atrioventricular cushion size	
			MP:0011098	embryonic lethality during organogenesis, complete penetrance	
			MP:0013504	increased embryonic tissue cell apoptosis	
EMC3	MGI:1913337	ER membrane protein complex subunit 3			
IL17RC	MGI:2159336	interleukin 17 receptor C	MP:0008721	abnormal chemokine level	Post
			MP:0001876	decreased inflammatory response	
			MP:0004800	decreased susceptibility to experimental autoimmune encephalomyelitis	
			MP:0008681	increased interleukin-17 secretion	
			MP:0002169	no abnormal phenotype detected	
			MP:0008568	abnormal interleukin secretion	
			MP:0003449	abnormal intestinal goblet cell morphology	
			MP:0003453	abnormal keratinocyte physiology	
			MP:0004842	abnormal large intestine crypts of Lieberkuhn morphology	
			MP:0008682	decreased interleukin-17 secretion	
			MP:0002412	increased susceptibility to bacterial infection	
			MP:0009788	increased susceptibility to bacterial infection induced morbidity/mortality	
			MP:0004799	increased susceptibility to experimental autoimmune encephalomyelitis	
			MP:0008537	increased susceptibility to induced colitis	
			MP:0010180	increased susceptibility to weight loss	

			MP:0001193	psoriasis	
JAGN1	MGI:1915017	jagunal homolog 1	MP:0005065	abnormal neutrophil morphology	Post
			MP:0002463	abnormal neutrophil physiology	
			MP:0008720	impaired neutrophil chemotaxis	
			MP:0008719	impaired neutrophil recruitment	
			MP:0005399	increased susceptibility to fungal infection	
MTMR14	MGI:1916075	myotubularin related protein 14	MP:0002972	abnormal cardiac muscle contractility	Post
			MP:0000266	abnormal heart morphology	
			MP:0005620	abnormal muscle contractility	
			MP:0002106	abnormal muscle physiology	
			MP:0004036	abnormal muscle relaxation	
			MP:0004089	dilated sarcoplasmic reticulum	
			MP:0001405	impaired coordination	
			MP:0012106	impaired exercise endurance	
			MP:0002841	impaired skeletal muscle contractility	
			MP:0003646	muscle fatigue	
			MP:0002269	muscular atrophy	
OGG1	MGI:1097693	8-oxoguanine DNA-glycosylase 1	MP:0005621	abnormal cell physiology	Pre/post
			MP:0008058	abnormal DNA repair	
			MP:0002757	decreased vertical activity	
			MP:0001402	hypoactivity	
			MP:0002048	increased lung adenoma incidence	
			MP:0008714	increased lung carcinoma incidence	
			MP:0008235	increased susceptibility to neuronal excitotoxicity	
			MP:0005370	liver/biliary system phenotype	
			MP:0008918	microgliosis	
			MP:0002229	neurodegeneration	
			MP:0002169	no abnormal phenotype detected	
PRRT3	MGI:2444810	proline-rich transmembrane protein 3			
RPUSD3	MGI:2141440	RNA pseudouridylate synthase			

				domain containing 3			
		TADA3	MGI:1915724	transcriptional adaptor 3	MP:0004046	abnormal mitosis	Post
					MP:0010039	abnormal trophoblast giant cell proliferation	
					MP:0004964	absent inner cell mass	
					MP:0011704	decreased fibroblast proliferation	
					MP:0011104	embryonic lethality before implantation, incomplete penetrance	
					MP:0011095	embryonic lethality between implantation and placentation, complete penetrance	
					MP:0002169	no abnormal phenotype detected	
		TTLL3	MGI:2141418	tubulin tyrosine ligase-like family, member 3	MP:0000495	abnormal colon morphology	Post
					MP:0000490	abnormal crypts of Lieberkuhn morphology	
					MP:0000488	abnormal intestinal epithelium morphology	
					MP:0010155	abnormal intestine physiology	
					MP:0005628	decreased circulating potassium level	
					MP:0004499	increased incidence of tumors by chemical induction	
					MP:0002169	no abnormal phenotype detected	
7652	0.33	AHSP	MGI:2158492	alpha hemoglobin stabilizing protein	MP:0002123	abnormal definitive hematopoiesis	Post
					MP:0002447	abnormal erythrocyte morphology	
					MP:0000245	abnormal erythropoiesis	
					MP:0001588	abnormal hemoglobin	
					MP:0005563	abnormal hemoglobin content	
					MP:0010177	acanthocytosis	
					MP:0001577	anemia	
					MP:0002642	anisocytosis	
					MP:0002641	anisopoikilocytosis	
					MP:0000208	decreased hematocrit	
					MP:0005562	decreased mean corpuscular hemoglobin	
					MP:0002591	decreased mean corpuscular volume	
					MP:0000691	enlarged spleen	
					MP:0008412	increased cellular sensitivity to oxidative stress	

			MP:0003135	increased erythroid progenitor cell number	
			MP:0008850	increased hemoglobin concentration distribution width	
			MP:0011171	increased number of Heinz bodies	
			MP:0010067	increased red blood cell distribution width	
			MP:0010175	leptocytosis	
			MP:0003674	oxidative stress	
			MP:0005097	polychromatophilia	
			MP:0002640	reticulocytosis	
ARMC5	MGI:2384586	armadillo repeat containing 5			
C16orf58	No associated gene				
COX6A2	MGI:104649	cytochrome c oxidase subunit VIa polypeptide 2	MP:0001544	abnormal cardiovascular system physiology	Post
			MP:0005385	cardiovascular system phenotype	
COX6A2	MGI:104649	cytochrome c oxidase subunit VIa polypeptide 2	MP:0004011	decreased diastolic filling velocity	Post
ITGAD	MGI:3578624	integrin, alpha D	MP:0001790	abnormal immune system physiology	Post
			MP:0000717	abnormal lymphocyte cell number	
			MP:0002451	abnormal macrophage physiology	
			MP:0001828	abnormal T cell activation	
			MP:0003944	abnormal T cell subpopulation ratio	
			MP:0001793	altered susceptibility to infection	
			MP:0008075	decreased CD4-positive, alpha beta T cell number	
			MP:0005026	decreased susceptibility to parasitic infection	
			MP:0005095	decreased T cell proliferation	
			MP:0008078	increased CD8-positive, alpha-beta T cell number	
			MP:0008617	increased circulating interleukin-12 level	
			MP:0004952	increased spleen weight	
			MP:0002412	increased susceptibility to bacterial infection	
ITGAX	MGI:96609	integrin alpha X	MP:0005025	abnormal response to infection	Post
			MP:0008127	decreased dendritic cell number	

			MP:0008567	decreased interferon-gamma secretion	
			MP:0008682	decreased interleukin-17 secretion	
			MP:0004800	decreased susceptibility to experimental autoimmune encephalomyelitis	
			MP:0008561	decreased tumor necrosis factor secretion	
			MP:0001606	impaired hematopoiesis	
			MP:0008663	increased interleukin-12 secretion	
			MP:0008687	increased interleukin-2 secretion	
			MP:0008524	increased plasmacytoid dendritic cell number	
			MP:0002412	increased susceptibility to bacterial infection	
			MP:0005348	increased T cell proliferation	
			MP:0002169	no abnormal phenotype detected	
SLC5A2	MGI:2181411	solute carrier family 5 (sodium/glucose cotransporter), member 2	MP:0011923	abnormal bladder urine volume	Pre/post
			MP:0009642	abnormal blood homeostasis	
			MP:0000188	abnormal circulating glucose level	
			MP:0005556	abnormal renal filtration rate	
			MP:0011447	abnormal renal glucose reabsorption	
			MP:0009643	abnormal urine homeostasis	
			MP:0005375	adipose tissue phenotype	
			MP:0005386	behavior/neurological phenotype	
			MP:0001262	decreased body weight	
			MP:0002667	decreased circulating aldosterone level	
			MP:0005560	decreased circulating glucose level	
			MP:0002727	decreased circulating insulin level	
			MP:0000198	decreased circulating phosphate level	
			MP:0004804	decreased susceptibility to autoimmune diabetes	
			MP:0010025	decreased total body fat amount	
			MP:0002988	decreased urine osmolality	
			MP:0005378	growth/size/body region phenotype	
			MP:0005292	improved glucose tolerance	

					MP:0000062	increased bone mineral density	
					MP:0010868	increased bone trabecula number	
					MP:0003019	increased circulating chloride level	
					MP:0005633	increased circulating sodium level	
					MP:0011941	increased fluid intake	
					MP:0011939	increased food intake	
					MP:0005582	increased renin activity	
					MP:0011459	increased urine chloride ion level	
					MP:0011470	increased urine creatinine level	
					MP:0001759	increased urine glucose level	
					MP:0011435	increased urine magnesium level	
					MP:0002901	increased urine phosphate level	
					MP:0005619	increased urine potassium level	
					MP:0006316	increased urine sodium level	
					MP:0011466	increased urine urea nitrogen level	
					MP:0001762	polyuria	
		TGFB1I1	MGI:102784	transforming growth factor beta 1 induced transcript 1	MP:0009503	abnormal mammary gland duct morphology	Post
					MP:0005595	abnormal vascular smooth muscle physiology	
					MP:0004883	abnormal vascular wound healing	
					MP:0005385	cardiovascular system phenotype	
					MP:0005292	improved glucose tolerance	
					MP:0005178	increased circulating cholesterol level	
					MP:0003443	increased circulating glycerol level	
					MP:0001556	increased circulating HDL cholesterol level	
					MP:0002169	no abnormal phenotype detected	
		ZNF843	No associated gene				
18038	0.32	ARL13B	MGI:1915396	ADP-ribosylation factor-like 13B	MP:0001706	abnormal left-right axis patterning	Post
					MP:0000926	absent floor plate	
					MP:0020384	absent kidney epithelial cell primary cilium	
					MP:0001265	decreased body size	

			MP:0004860	dilated kidney collecting duct	
			MP:0011098	embryonic lethality during organogenesis, complete penetrance	
			MP:0003068	enlarged kidney	
			MP:0005565	increased blood urea nitrogen level	
			MP:0011440	increased kidney cell proliferation	
			MP:0003917	increased kidney weight	
			MP:0000522	kidney cortex cysts	
			MP:0003675	kidney cysts	
			MP:0003606	kidney failure	
			MP:0011307	kidney medulla cysts	
			MP:0002169	no abnormal phenotype detected	
			MP:0011090	perinatal lethality, incomplete penetrance	
			MP:0008528	polycystic kidney	
			MP:0011085	postnatal lethality, complete penetrance	
			MP:0002083	premature death	
			MP:0003985	renal fibrosis	
			MP:0003215	renal interstitial fibrosis	
			MP:0001954	respiratory distress	
			MP:0003054	spina bifida	
PROS1	MGI:1095733	protein S (alpha)	MP:0002551	abnormal blood coagulation	Post
			MP:0001614	abnormal blood vessel morphology	
			MP:0004950	abnormal brain vasculature morphology	
			MP:0008547	abnormal neocortex morphology	
			MP:0003638	abnormal response/metabolism to endogenous compounds	
			MP:0005048	abnormal thrombosis	
			MP:0003229	abnormal vitelline vasculature morphology	
			MP:0010022	brain vascular congestion	
			MP:0009275	bruising	
			MP:0005385	cardiovascular system phenotype	
			MP:0005607	decreased bleeding time	
			MP:0000606	decreased hepatocyte number	
			MP:0011380	enlarged brain ventricles	
			MP:0001914	hemorrhage	
			MP:0001654	hepatic necrosis	

					MP:0008254	increased megakaryocyte cell number	
					MP:0009763	increased sensitivity to induced morbidity/mortality	
					MP:0003070	increased vascular permeability	
					MP:0001915	intracranial hemorrhage	
					MP:0011099	lethality throughout fetal growth and development, complete penetrance	
					MP:0010768	mortality/aging	
					MP:0002169	no abnormal phenotype detected	
					MP:0003717	pallor	
					MP:0011089	perinatal lethality, complete penetrance	
					MP:0011101	prenatal lethality, incomplete penetrance	
					MP:0003179	thrombocytopenia	
					MP:0003814	vascular smooth muscle hypoplasia	
					MP:0000255	vasculature congestion	
		STX19	MGI:1915409	syntaxin 19	MP:0003020	decreased circulating chloride level	Post
					MP:0005560	decreased circulating glucose level	
					MP:0002727	decreased circulating insulin level	
9274	0.25	ATP5MG	No associated gene				
		CD3D	MGI:88331	CD3 antigen, delta polypeptide	MP:0001825	arrested T cell differentiation	Post
					MP:0008075	decreased CD4-positive, alpha beta T cell number	
					MP:0008079	decreased CD8-positive, alpha-beta T cell number	
		CD3E	MGI:88332	CD3 antigen, epsilon polypeptide	MP:0004939	abnormal B cell morphology	Post
					MP:0004816	abnormal class switch recombination	
					MP:0002407	abnormal double-negative T cell morphology	
					MP:0002408	abnormal double-positive T cell morphology	
					MP:0004946	abnormal regulatory T cell physiology	
					MP:0002145	abnormal T cell differentiation	
					MP:0008037	abnormal T cell morphology	
					MP:0006387	abnormal T cell number	
					MP:0000703	abnormal thymus morphology	

			MP:0003790	absent CD4-positive, alpha beta T cells	
			MP:0000727	absent CD8-positive, alpha-beta T cells	
			MP:0008070	absent T cells	
			MP:0001825	arrested T cell differentiation	
			MP:0005092	decreased double-positive T cell number	
			MP:0010133	increased DN3 thymocyte number	
			MP:0005090	increased double-negative T cell number	
			MP:0009763	increased sensitivity to induced morbidity/mortality	
			MP:0006413	increased T cell apoptosis	
			MP:0002064	seizures	
			MP:0000706	small thymus	
			MP:0001823	thymus hypoplasia	
CD3G	MGI:88333	CD3 antigen, gamma polypeptide	MP:0004392	abnormal CD8-positive, alpha-beta T cell physiology	Post
			MP:0005025	abnormal response to infection	
			MP:0001828	abnormal T cell activation	
			MP:0002145	abnormal T cell differentiation	
			MP:0002444	abnormal T cell physiology	
			MP:0005094	abnormal T cell proliferation	
			MP:0008079	decreased CD8-positive, alpha-beta T cell number	
			MP:0005079	decreased cytotoxic T cell cytolysis	
			MP:0000715	decreased thymocyte number	
			MP:0005387	immune system phenotype	
			MP:0006413	increased T cell apoptosis	
			MP:0008101	lymph node hypoplasia	
			MP:0000694	spleen hypoplasia	
			MP:0001823	thymus hypoplasia	
KMT2A	MGI:96995	lysine (K)-specific methyltransferase 2A	MP:0009879	abnormal arcus anterior morphology	Pre/post
			MP:0002182	abnormal astrocyte morphology	
			MP:0004924	abnormal behavior	
			MP:0002429	abnormal blood cell morphology/development	

	abnormal bone marrow cell morphology/developme
MP:0002398	nt
MP:0009278	abnormal bone marrow cell physiology
MP:0000913	abnormal brain development
MP:0004607	abnormal cervical atlas morphology
MP:0004608	abnormal cervical axis morphology
MP:0003048	abnormal cervical vertebrae morphology
MP:0008249	abnormal common lymphocyte progenitor cell morphology
MP:0006410	abnormal common myeloid progenitor cell morphology
MP:0001081	abnormal cranial ganglia morphology
MP:0002123	abnormal definitive hematopoiesis
MP:0000812	abnormal dentate gyrus morphology
MP:0000961	abnormal dorsal root ganglion morphology
MP:0003743	abnormal facial morphology
MP:0004808	abnormal hematopoietic stem cell morphology
MP:0010763	abnormal hematopoietic stem cell physiology
	abnormal hematopoietic system morphology/developme
MP:0002396	nt
MP:0000596	abnormal liver development
MP:0002619	abnormal lymphocyte morphology
MP:0001601	abnormal myelopoiesis
MP:0011808	abnormal myoblast differentiation
MP:0006009	abnormal neuronal migration
MP:0009937	abnormal neuron differentiation
MP:0008029	abnormal paraxial mesoderm morphology
MP:0004275	abnormal postnatal subventricular zone morphology
MP:0002416	abnormal proerythroblast morphology
MP:0002823	abnormal rib development
MP:0005221	abnormal rostral-caudal axis patterning

MP:0000689	abnormal spleen morphology
MP:0000157	abnormal sternum morphology
MP:0013176	abnormal tail position or orientation
MP:0004599	abnormal vertebral arch morphology
MP:0001577	anemia
MP:0001393	ataxia
MP:0000017	big ears
MP:0004620	cervical vertebral fusion
MP:0004615	cervical vertebral transformation
MP:0005017	decreased B cell number
MP:0009703	decreased birth body size
MP:0001265	decreased body size
MP:0001262	decreased body weight
MP:0000333	decreased bone marrow cell number
MP:0008813	decreased common myeloid progenitor cell number
MP:0002875	decreased erythrocyte cell number
MP:0008973	decreased erythroid progenitor cell number
MP:0011704	decreased fibroblast proliferation
MP:0004810	decreased hematopoietic stem cell number
MP:0002874	decreased hemoglobin content
MP:0004981	decreased neuronal precursor cell number
MP:0008209	decreased pre-B cell number
MP:0003345	decreased rib number
MP:0008008	early cellular replicative senescence
MP:0011094	embryonic lethality before implantation, complete penetrance
MP:0011092	embryonic lethality, complete penetrance
MP:0011098	embryonic lethality during organogenesis, complete penetrance
MP:0000691	enlarged spleen
MP:0004569	glossopharyngeal nerve hypoplasia
MP:0001914	hemorrhage
MP:0001505	hunched posture
MP:0005387	immune system phenotype

MP:0002023	increased B cell derived lymphoma incidence
MP:0003135	increased erythroid progenitor cell number
MP:0002021	increased incidence of induced tumors
MP:0002026	increased leukemia incidence
MP:0000218	increased leukocyte cell number
MP:0002018	increased malignant tumor incidence
MP:0000220	increased monocyte cell number
MP:0001658	increased mortality induced by gamma-irradiation
MP:0000219	increased neutrophil cell number
MP:0008186	increased pro-B cell number
MP:0005649	increased spleen neoplasm incidence
MP:0008476	increased spleen red pulp amount
MP:0012284	increased sternebra number
MP:0011099	lethality throughout fetal growth and development, complete penetrance
MP:0000600	liver hypoplasia
MP:0004616	lumbar vertebral transformation
MP:0000688	lymphoid hyperplasia
MP:0010373	myeloid hyperplasia
MP:0002006	neoplasm
MP:0002169	no abnormal phenotype detected
MP:0011089	perinatal lethality, complete penetrance
MP:0011090	perinatal lethality, incomplete penetrance
MP:0001732	postnatal growth retardation
MP:0002083	premature death
MP:0001923	reduced female fertility
MP:0000154	rib fusion
MP:0010179	rough coat
MP:0004617	sacral vertebral transformation
MP:0000445	short snout
MP:0001786	skin edema
MP:0002741	small olfactory bulb
MP:0010082	sternebra fusion
MP:0006099	thin cerebellar granule layer
MP:0004618	thoracic vertebral transformation

			MP:0003179	thrombocytopenia
			MP:0003036	vertebral transformation
UBE4A	MGI:2154580	ubiquitination factor E4A		

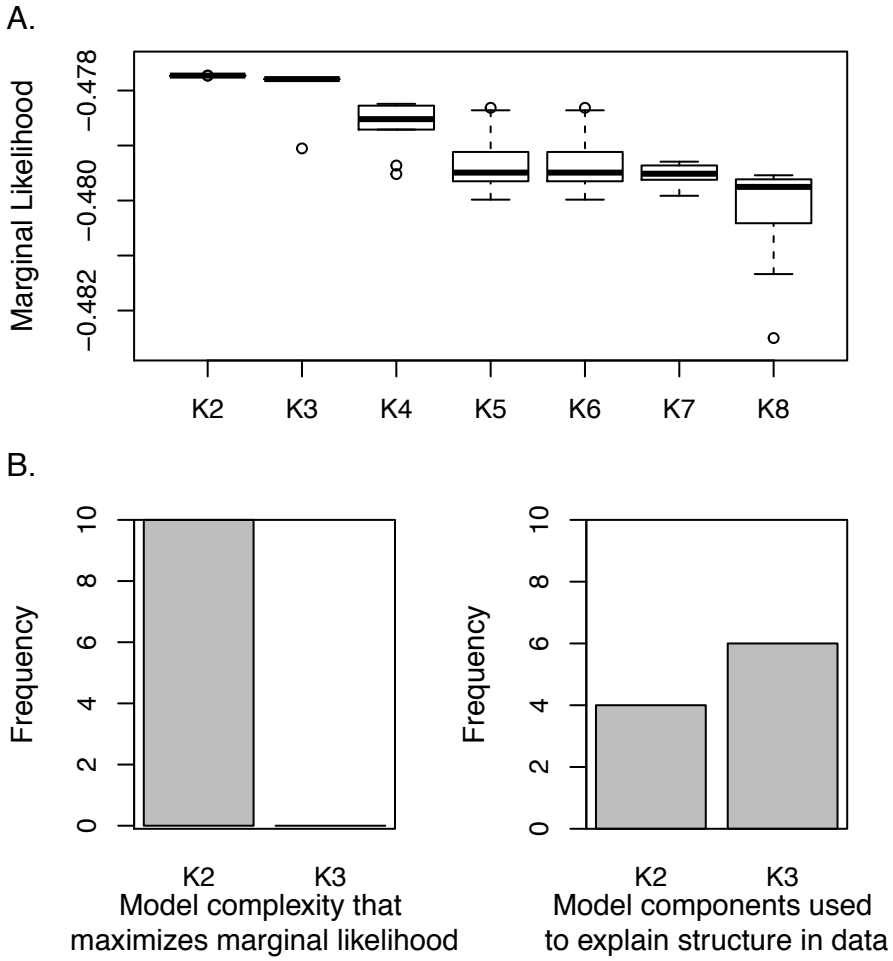

Figure A3.1. Model fitting summary of ten replicate *fastStructure* runs each for K=2-8. A) Boxplot of marginal likelihood scores, and B) number of replicates supporting K=2 and K=3 as the appropriate number of clusters.

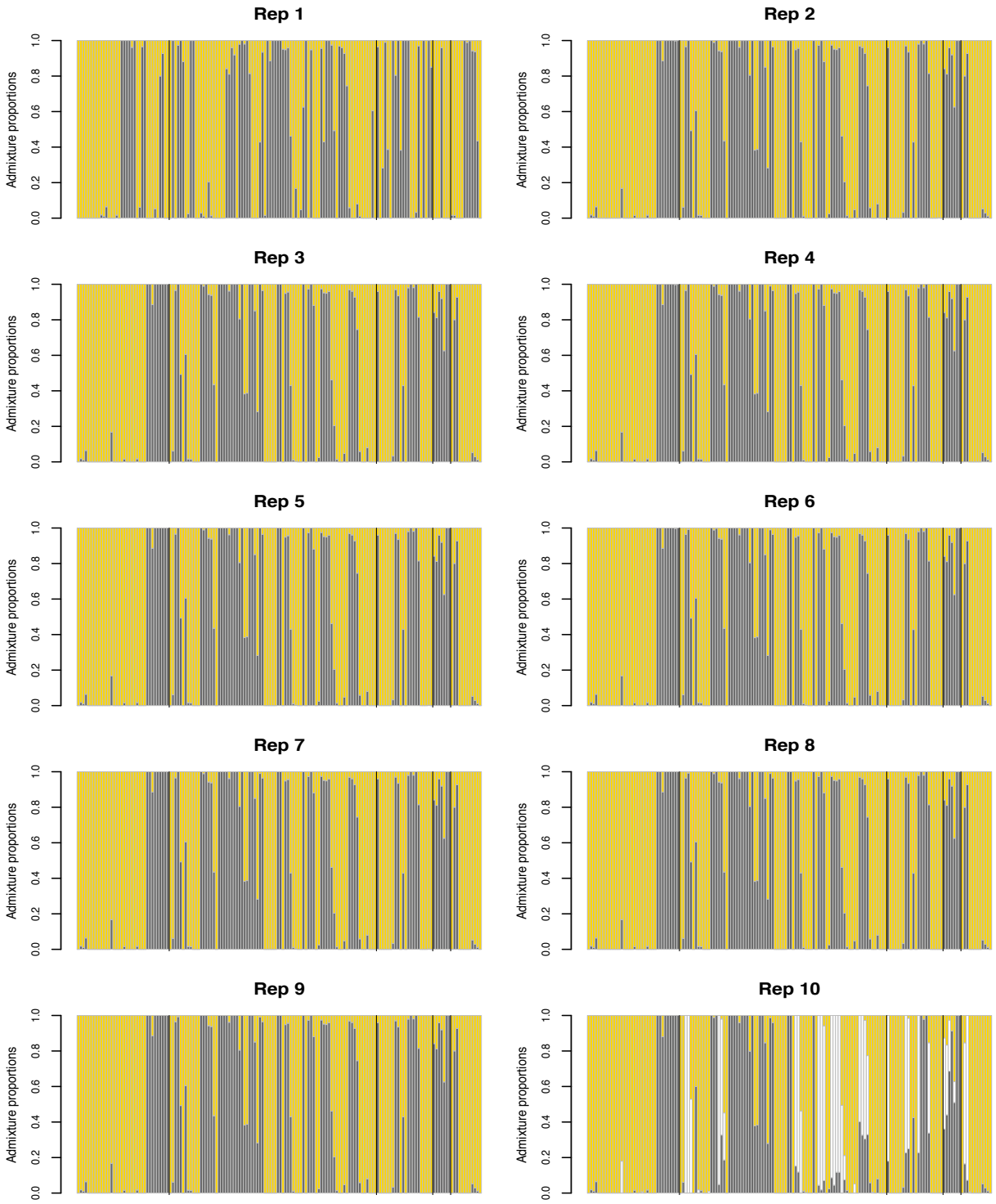

Figure A3.2. Admixture proportions for each of ten replicate *fastStructure* runs at K=3. Individual admixture proportions are very similar to admixture proportions obtained using K=2 due to very low ancestry proportions assigned to the third cluster in nine of ten runs (reps 1–9).

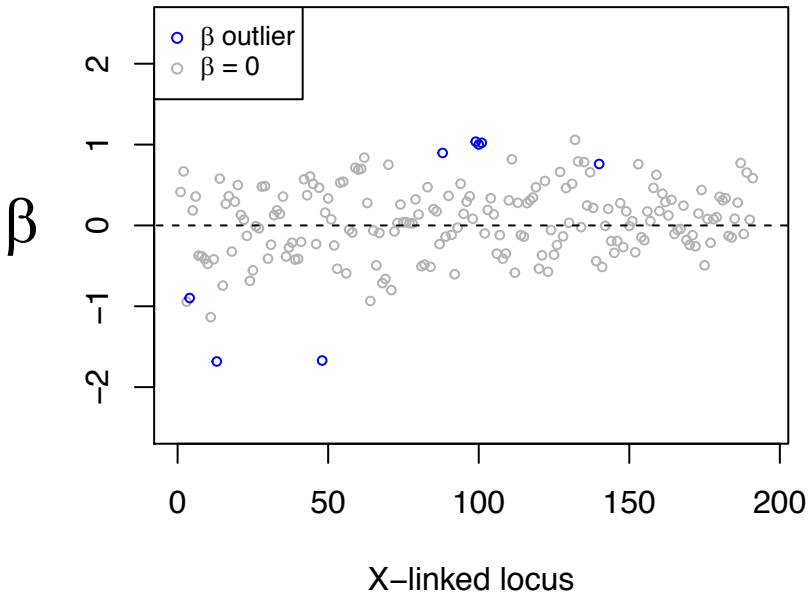

Figure A3.3. Amount of introgression (β) for the subset of loci (N=191) on contigs designated as X-linked based on sequence similarity to genes known to be X-linked in humans.