Received: 24 October 2015 | Revised: 3 July 2016 | Accepted: 10 July 2016

DOI 10.1002/ajp.22587

RESEARCH ARTICLE

WILEY AMERICAN JOURNAL OF PRIMATOLOGY

Application of PE-RADSeq to the study of genomic diversity and divergence of two Brazilian marmoset species (*Callithrix jacchus* and *C. penicillata*)

Joanna Malukiewicz[1,2*] | Katerina Guschanski[3] | Adriana D. Grativol[4] | Maria Adélia B. Oliveira[5] | Carlos R. Ruiz-Miranda[4] | Anne C. Stone[6]

[1] Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, MG, Brazil

[2] School of Life Sciences, Arizona State University, Tempe, Arizona

[3] Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

[4] Laboratório de Ciências Ambientais, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, RJ, Brazil

[5] Departamento de Morfologia e Fisiologia Animal, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil

[6] School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona

***Correspondence**
Joanna Malukiewicz, Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Avenida PH Rolfs s/n, Viçosa, MG 36570-000, Brazil.
Email: jmalukie@gmail.com

Funding Information
This research was supported by National Science Foundation DDIG proposal 1061508, Fulbright Fellowship, ASU SOLS FIGG Grant, ASU GPSA Jump Start and Research Grants, ASU Chapter of Sigma Xi and International Primatological Society Research Grant.

Callithrix jacchus and *C. penicillata* are among the smallest anthropoid primates, are highly specialized tree gougers, and largely occupy Brazil's most extreme, semi-arid biomes. However, the underlying genomic factors that underpin the evolution of these species and their unique traits are under-investigated. Additionally, exotic populations of these two species are widely established throughout Brazil and hybridize with threatened native congers. Thus, both genomic and conservation factors call for a better understanding of *C. jacchus* and *C. penicillata* evolution. Here, we applied PE-RADseq to characterize genomic variation in these two species, using six *C. jacchus* and seven *C. penicillata* individuals. We identified an average of 7,463 and 5,180 SNPs/individual in *C. penicillata* and *C. jacchus*, respectively, and also found 1,395 variable sites that were represented in both species. *C. penicillata* showed overall higher levels of genetic diversity than *C. jacchus* at the variable sites present in both species. Additionally, among these variable sites, 106 showed relative interspecific divergence levels that were significantly higher than the genome-wide average. We further compared relative and absolute divergence for *C. penicillata* and *C. jacchus* between RAD loci associated with the 106 significantly diverged variable sites and all other RAD loci present in both species. The former RAD loci set showed significantly elevated relative and absolute divergence measures in comparison to the latter set. This convergence suggests that *C. jacchus* and *C. penicillata* may have diverged under a scenario of gene flow under secondary contact. Here, we demonstrate that RADseq is an efficient method to simultaneously discover and genotype a large number of markers and validate the utility of RADseq for examining *Callithrix* evolution.

KEYWORDS
Callithrix, divergence, diversity, genomics, RADseq, speciation

1 | INTRODUCTION

The young Neotropical *Callithrix* genus probably emerged approximately 2.5 million years ago (MYA) (Perelman et al., 2011) and is composed of six species that are endemic to Brazil (Rylands, Coimbra-Filho, & Mittermeier, 2009). *Callithrix* marmosets possess several unique anthropoid primate traits such as cooperative breeding (Digby, Ferrari, & Saltzman, 2007), socially modulated female reproduction (Smith, Schaffner, & French, 1997), and high hemopoietic chimersim levels (Hershkovitz, 1977). The two youngest *Callithrix* species, *C. jacchus* and *C. penicillata*, are thought to have diverged from each other less than a million years ago (Perelman et al., 2011) and still share many behavioral and physical similarities (Digby et al., 2007; Rylands and de Faria, 2003). They are distinct from their congers by (1) being the smallest *Callithrix* marmosets (Fuzessy et al., 2014; unpublished data, Malukiewicz); (2) possessing morphological specializations for facultative tree gouging (Vinyard et al., 2009); and (3) being the only *Callithrix* species distributed largely outside of the Brazilian Atlantic

Forest (Malukiewicz et al., 2014; Rylands et al., 2009). Further, *C. jacchus* and *C. penicillata* are differentiated from each other by (1) facial and ear tuft pelage coloration and patterning (see Malukiewicz et al., 2014); (2) vocalizations (Mendes, Vielliard, & de Marco, 2009); (3) mitochondrial haplotypes (Malukiewicz et al., 2014); and (4) allopatric geographical ranges that largely occur within two distinct semi-arid biomes (Malukiewicz et al., 2014; Rylands et al., 2009). However, the demographic and genomic factors that underpin *C. jacchus* and *C. penicillata* divergence from their congeners and each other are yet to be investigated.

On-going anthropogenic hybridization between *C. jacchus* and *C. penicillata* themselves and their congers signals a further need to understand the genomic diversity and divergence of these two species. Although *Callithrix* species possess historically allopatric geographic ranges (Rylands et al., 2009), exotic *Callithrix* populations, particularly those of *C. jacchus* and *C. penicillata*, are now established throughout Brazil as result of human introductions through the illegal pet-trade (de Carvalho, 2015; Malukiewicz et al., 2014, 2015; Ruiz-Miranda et al., 2000, 2006; Zago, 2012; unpublished data, Silva; personal observations, Malukiewicz). As a result, exotic *C. jacchus* and *C. penicillata* are also found together in artificial sympatry (de Carvalho, 2015; Malukiewicz et al., 2014, 2015; Ruiz-Miranda et al., 2000, 2006; unpublished data, Silva; personal observations, Malukiewicz). Exotic *C. jacchus* and *C. penicillata* hybridize with native species like *C. aurita* (vulnerable under IUCN Red List) and *C. flaviceps* (threatened under IUCN Red List), (de Carvalho, 2015; personal observations, Malukiewicz), but the extent to which hybridization threatens the genetic integrity of endangered *Callithrix* species needs further investigation. Hence, a clearer understanding of the population genomics of *C. jacchus* and *C. penicillata* will elucidate both the evolutionary history of the two species and conservation aspects of anthropogenic *Callithrix* hybridization.

Currently, the available theoretical models explaining genomic divergence during speciation evoke different evolutionary factors. For example, the speciation-with-gene flow model postulates that strong differentiation occurs in genomic regions experiencing reduced levels of inter-population gene flow (Feder et al., 2013). This model puts speciation genes, or those genes "whose divergence made a significant contribution to the evolution of reproductive isolation between populations" (Nosil & Schluter, 2011), within the most extremely divergent genomic regions. Conversely, ecological models minimize the role of gene flow and consider local adaptation as the main driver of population divergence (Schluter, 2009). Species are expected to diverge at genetic loci associated with adaptations to contrasting environments (Schluter, 2009). Genomic features like recombination also greatly influence genomic divergence patterns (Renaut et al., 2013).

In simulations testing the theoretical expectations of several divergence models, Cruickshank and Hahn (2014) showed that under primary and secondary gene flow, high levels of absolute divergence (D_{XY}) and relative divergence (F_{ST}) should coincide together at diverged loci involved in species isolation. On the other hand, high F_{ST}, low intraspecific nucleotide diversity, and unchanged to lowered D_{XY} values are expected at loci that are the targets of linked selection relative to the genome wide average (Cruickshank & Hahn, 2014). Accordingly, an emergent model put forth by Cruickshank and Hahn (2014) explains genomic divergence as interplay between the strength of selection, shared ancestry, and stochastic variation in coalescence, regardless of speciation mode. In this model, extremely divergent loci do not necessarily contain speciation genes, but can instead result from different levels of background selection (Cruickshank & Hahn, 2014).

Presently, several next-generation sequencing (NGS) methods are available for making empirical inferences regarding divergence, diversity, and hybridization using a large number of bi-parental nuclear variable sites. Among these NGS methods, restriction-associated DNA sequencing (RADseq) has emerged as a leading technique in the genomics of non-model faunal species (e.g., Bell, Drewes, & Zamudio, 2015; Schield et al., 2015). RADseq is based on the concurrent discovery and genotyping of SNPs contained within short DNA sequences located close to restriction enzyme recognition sites (described in more detail in Baird et al., 2008). Several versions of RADseq have been developed over the last few years (e.g., paired-end [PE] RADseq (Etter & Johnson, 2012), double-digest RADseq (ddRADseq) (Peterson et al., 2012), and ezRAD (Toonen et al., 2013)), and several recent reviews discuss the advantages and disadvantages of these RADseq techniques (e.g., Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Davey et al., 2013; Mastretta-Yanes et al., 2015).

To date, the few available studies on the genetic diversity of *C. jacchus* and *C. penicillata* have been carried out through relatively time-consuming approaches or limited to uni-parental genetic markers. Here, we present an interspecific application of RADseq to study the evolution of these two species with the following objectives: (1) survey intraspecific diversity of each species; (2) characterize genomic divergence between these two species; and (3) determine whether diverged genomic loci are ontologically enriched for either biological processes, molecular functions, or cellular components. These preliminary data will allow us to investigate the evolutionary history of *C. jacchus* and *C. penicillata* as well as lay the groundwork for testing hypotheses about the nature and extent of introgression in *Callithrix* hybrid zones.

2 | METHODS

2.1 | Sample collection and whole genome amplification (WGA)

Samples for this study consisted of six *C. jacchus* and seven *C. penicillata* collected from captive and wild marmoset populations between 2010 and 2011 (Table 1). Permission for capture and tissue collection from wild marmosets was obtained from the Brazilian Ministry for the Environment and Natural Resources (IBAMA, protocol #28075-2). The Arizona State University Institutional Animal Care and Use Committee Animals approved the capture and sampling of wild Brazilian and US captive marmoset populations (ASU IACUC, protocol #11-1150R). All possible steps were taken to minimize animal

TABLE 1 Summary of sampled individuals from captive and wild pure *C. jacchus* and *C. penicillata* populations

Sample	Sex	Species	WGA[a]	Origin	Origin
cja013	F	J[b]	No	CRC[d], Omaha, NE	NA
cja014	M	J	No	CRC[d], Omaha, NE	NA
cja030	F	J	Yes	CETAS[e], Recife, PE, Brazil	NA
cja031	M	J	No	CETAS[e], Recife, PE, Brazil	NA
cja034	M	J	Yes	CETAS[e], Recife, PE, Brazil	NA
cja041	F	J	Yes	CETAS[e], Recife, PE, Brazil	NA
cpe001	F	P[c]	Yes	CRC[d], Omaha, NE	NA
cpe009	M	P	Yes	Muriaé, MG, Brazil	21° 7′ 15.60″S, 42° 22′ 2.50″W
cpe018	M	P	Yes	Brasilía, DF, Brazil	15°45′1.15″S, 47°50′34.10″W
cpe022	F	P	No	Brasilía, DF, Brazil	15°51′56.46″S, 47°58′13.94″W
cpe023	F	P	No	Brasilía, DF, Brazil	15°42′33.45″S, 47°54′44.80″W
cpe026	F	P	Yes	Brasilía, DF, Brazil	15°54′38.07″S, 47°57′10.64″W
cpe041	M	P	Yes	CETAS[e], Goiânia, GO, Brazil	NA

[a]Refers to "whole genome amplified."
[b]Indicates *C. jacchus*.
[c]Indicates *C. penicillata*.
[d]Callitrichid Research Center, University of Nebraska at Omaha.
[e]IBAMA Wild Animal Triage Center, Brazilian Institute of the Environment and Natural Resources.

suffering and maximize their safety by adhering to protocols approved by the ASU IACUC, to legal requirements within Brazil, and to the American Society of Primatologists Principles for the Ethical Treatment of Non Human Primates. Table 1 lists sample information regarding species, sex, origin, and whole genome amplification (WGA) status. More detailed information about collection permits, sample collection, storage, sampling sites/facilities, and DNA extraction from these biological samples is provided in Malukiewicz et al. (2014).

DNA concentrations of all samples were determined with a Qubit 2 Fluorometer and the Qubit dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA). Four *C. jacchus* samples and five *C. penicillata* samples initially showed DNA amounts too low for further processing. DNA amounts of these samples were increased through WGA with the Repli-g Mini Kit (Qiagen, Valencia, CA) using manufacturer's instructions for purified genomic DNA. One *C. jacchus* (cja014) sample was prepared for sequencing with and without WGA to check for amplification bias during WGA. DNA concentrations of WGA samples were rechecked with the Qubit dsDNA BR Assay Kit. An investigation of WGA and non-WGA preparations of cja014 for genotypic differences due to possible WGA bias is described in Supplementary Methods. Hereafter, these samples are respectively referred to as cja014-WGA and cja014-nonWGA. Although some minor WGA bias was detected (see Results), the WGA preparation of cja014 was used in further analyses because it resulted in a higher number of sequencing reads (see Results).

2.2 | RADseq library preparation and sequencing analysis

A single *Sbf*I-digsted RADseq library was prepared following a modified version of the PE-RADseq protocol of Etter and Johnson (2012). As a rare cutter enzyme, *Sbf*I was chosen for library preparation to increase read count of each RADseq locus while still reducing per individual sequencing costs (Guo et al., 2014). A RADseq locus is synonymous with the sequence of DNA that immediately flanks the recognition site of a given restriction enzyme. PE sequence reads are those that are read from the 5′ and 3′ end, respectively, of the same DNA fragment. A unique 5-base pair barcode identified each DNA sample, and barcodes differed from each other by 3 base pairs. We modified the protocol of Etter and Johnson (2012) by (1) chemically shearing DNA samples with NEBNext dsDNA Fragmentase (New England BioLabs, Ipswich, MA) following the manufacturer's instructions, (2) carrying out DNA concentration steps with DNA Clean and Concentrator-5 capped columns (Zymo Research, Irvine, CA), and (3) using AMPure XP beads (Agencourt, Brea, CA) for purification steps. The final step of the library preparation protocol consisted of enriching the RADseq library through the polymerase chain reaction, and then purifying the amplified product on a 1% agarose, 1x TBE gel. A 350–850 bp band was cut from the gel in an inverted trapezoidal shape, following the procedure outlined by Etter and Johnson (2012). The final RADseq library was sequenced on a single lane of an Illumina HiSeq 2500 sequencer at the University of Arizona Genetics Core for a total of 100 cycles to produce PE reads. The library contained a low complexity region where all DNA fragments possessed the *Sbf*I recognition site. To prevent data loss during sequencing due to such low complexity regions (Krueger, Andrews, & Osborne, 2011), the library was spiked with a PhiX control (Illumina, San Diego, CA), with a mix of 60% of library and 40% control loaded for sequencing.

2.3 | Data analysis

PE reads were first demultiplexed by individual barcodes with the *process_radtags* program in the STACKS 1.40 pipeline (Catchen et al., 2013). *Process_radtags* filtered lower quality data by dropping any

reads with a phred score <10, uncalled bases, or ambiguous barcodes. The data were also filtered for Illumina PE adaptor sequences. Then, using both forward and reverse reads from each PE sequence, STACKS *clone_filter* program removed polymerase chain reaction (PCR) duplicates from the data. Beyond this PCR duplicate filtering step, the STACKS user manual (available online at: http://catchenlab.life.illinois.edu/stacks/manual) states that the software does not directly support inclusion of paired-end reads for analysis without anchoring by a second restriction enzyme. As we only used a single restriction enzyme in preparing our RADseq libraries, subsequent analyses were only carried out using the forward end of each PE sequence read.

BOWTIE 2.2.4 (Langmead & Salzberg, 2012; Langmead et al., 2009) was used to align reads for each sampled individual to the published marmoset genome (calJac3 build, Worley et al., 2014). BOWTIE was run in "very-sensitive" mode to maximize the alignment accuracy of the program. The k-reporting mode was set to 2 during alignments, which resulted in two distinct alignments, ordered by alignment score. These dual results were later used to filter out non-unique reads that aligned to more than one location within the marmoset genome using a custom Python script (available from JM upon request). Because the majority of downstream evolutionary analyses were carried out in STACKS, which is designed for diploid loci, we did not align reads to the marmoset X and Y chromosomes. We also did not align reads to calJac3 contigs with "chrUN" or "random" designations, as these contigs do not yet have confident placement in the marmoset reference genome. After alignment, samples cja034 and cpe041 were excluded from further analyses due to poor overall alignment to the marmoset reference genome.

Aligned reads for each sampled individual were processed after genome alignment in a series of programs within the STACKS pipeline. To determine the minimum stack size for the pipeline, the *pstacks* −m parameter was systematically varied between 5 and 10, but above a stack size of 6 many samples started to drop out of the analysis. Thus, due to our small sample size, we set −m = 6 and left all other *pstacks*, *sstacks*, and *cstacks* settings at their defaults. The *pstacks* program created a locus catalogue for each sampled individual and identified polymorphic nucleotide sites (Catchen et al., 2013). Also, *pstacks* assigns log likelihoods to each assembled stack, which can be used for further quality filtering during the *rxstacks* step described below. Then the *cstacks* program made locus cross-matches between sampled individuals and *sstacks* determined the allelic state for each RAD locus (Catchen et al., 2013).

Next, genotype calls for individual samples were corrected with *rxstacks* using aggregated sample-wide data where catalogue loci are assigned log likelihoods based on *pstacks* nucleotide calls. Highly negative log likehood values are indicative of loci with low coverage or high sequencing error (available online at: http://creskolab.uoregon.edu/stacks/comp/rxstacks.php) and are not informative in a STACKS analysis. The best log likelihoods are close to zero. *Rxstacks* was run with the lnl_dist parameter to list mean log likelihoods for each catalogue locus. After manually checking this list, the log likelihood limit was chosen as −8.0, which represents the point of a precipitous drop in mean log likelihood values among catalogue loci. All loci below this threshold were blacklisted from the analysis, and *cstacks* and *sstacks* were reran without these loci.

The STACKS *populations* program was used to conduct population genomic analyses based on further filtering of RAD loci. An initial *populations* run was conducted to determine a threshold minor allele frequency (MAF) because low MAF may bias genome scan studies (Roesti, Salzburger, & Berner, 2012). For this and subsequent *populations* analyses, all *C. jacchus* and *C. penicillata* samples were respectively considered as belonging to two different populations, and we filtered our dataset for RAD loci present in at least 55% of individuals within each species (parameter −r = 0.55). Due to possible variance in coverage of RAD loci across samples (Andrews et al., 2016), it was expected that not all RAD loci would be represented equally between *C. jacchus* and *C. penicillata*. The MAF threshold was determined by calculating the correlation of AMOVA F_{ST} (see below) between "sister" SNPs linked together at the same RAD locus within a given group of samples (see Roesti et al., 2012). Thus, we were interested in maximizing the amount of available sister SNPs within each species at this stage of *population* analyses, and we required a RAD locus to be present in only one species (parameter −p = 1). We calculated the Spearman rank correlation between sister SNPs within each species, while systematically eliminating loci with MAFs of 0.10, 0.125, 0.15, 0.2, 0.25, and 0.30. Because the correlation between AMOVA F_{ST} values of sister SNPs was above 0.70 for both species at a MAF of 0.125, this value was used as the MAF threshold frequency.

Subsequently, we carried out our main *populations* analyses, where we required all RAD loci to be present in both *C. jacchus* and *C. penicillata*. Because none of these RAD loci showed a MAF below our threshold value of 0.125, we filtered RAD loci with $H_O > 0.5$ and those with SNPs that had more than two alleles. *Populations* allows for further filtering based on stack log likelihoods generated in *pstacks* and initially filtered by *rxstacks*. For our final *populations* likelihood threshold (parameter −lnl_lim), we chose a more stringent value of −3, as the majority of RAD loci after *rxstacks* filtering possessed log likelihoods between 0 and −2. We employed this strategy to get the log likelihood threshold as close to zero as possible while not filtering out the majority of RAD loci available for analyses through *populations*.

In *populations*, intraspecific diversity was based on both variable and monomorphic sites within RAD loci and measured by major allele frequency (P), observed heterozygosity (H_O), expected heterozygosity (H_E), nucleotide diversity (π), and F_{IS} for *C. jacchus* and *C. penicillata*, respectively. We examined interspecific divergence in *populations* by calculating genome-wide pairwise F_{ST} values between *C. jacchus* and *C. penicillata*. Please note that F_{ST} values are only calculated by *populations* at nucleotide sites that are variable in a pair of populations (or species as in our case) (Catchen et al., 2013).

For point estimates of F_{ST} values at each interspecific variable site, *populations* provided two calculations of F_{ST} measures, that of Weir and Cockerham (1984) and an Analysis of Molecular Variance (AMOVA) alternate of Weir (1996). Both versions of F_{ST} measures were corrected by using Fisher's exact test to see if allele frequencies at each variable site differed significantly from zero. The F_{ST} value was set to zero at each variable site where $P > 0.05$ after application of Fisher's exact test. Variable sites with significant pairwise F_{ST} values, as

calculated by *populations*, were subsequently referred to as "significantly diverged" between *C. jacchus* and *C. penicillata*.

Populations was also used to carry out kernel-smoothing of intraspecific π and interspecific F_{ST} values using the default window size of 150,000 base pairs. Please note that for kernel-smoothing of F_{ST}, *populations* only compares those sites that are variable in the pair of populations in question (Catchen et al., 2013), as explained above. Bootstrapping was used to identify SNPs that possessed kernel-smoothed intraspecific π and variable sites that had interspecific F_{ST} values above the genome-wide average. The number of bootstrapping iterations was set to 1,000,000. For F_{ST} kernel-smoothing, any variable site with a kernel-smoothed F_{ST} value whose $P < 0.05$ after bootstrapping was considered as significantly diverged.

Currently, STACKS does not support the calculation of absolute divergence (D_{XY}) between different species, which unlike relative measures such as F_{ST}, is not biased by intraspecific diversity (Cruickshank & Hahn, 2014). We adapted the custom Python code (available upon request from JM) of Martin et al. (2014) for EGGLIB (De Mita & Siol, 2012) for calculations of D_{XY} for all RAD loci present in both *C. jacchus* and *C. penicillata*. EGGLIB was also used to make F_{ST} calculations for all RAD loci present in *C. jacchus* and *C. penicillata*. Finally, separate *C. jacchus* and *C. penicillata* π calculations were made with EGGLIB for those RAD loci present in both species. The Mann–Whitney U test was used to determine whether F_{ST}, D_{XY}, *C. jacchus* π and *C. penicillata* π values were equal between all non-significantly diverged and significantly diverged RAD loci, respectively. All statistical tests were carried out at an alpha-level of 0.05.

RAD loci were included in gene ontology (GO) enrichment analysis if they possessed variable sites with statistically significant *populations* F_{ST} measures and were found within a gene-coding genomic region. GO analysis was conducted in GORILLA (Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009; Eden, Lipson, Yogev, & Yakhini, 2007) using the mode that compared a target and background gene list under *Homo sapiens* (no *Callithrix* species is currently supported by GORILLA). GORILLA GO analysis settings included process, function, and componente ontologies, and a *P*-value threshold of 10^{-3} with a multiple testing correction using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995).

To generate GO target and background gene lists, we first determined whether RAD loci possessing significantly diverged variable sites were located within a gene. For this step, genomic regions consisting of the first 100 nucleotides up and downsteam of each target variable site were uploaded into the UCSC Genome Table Browser (Karolchik et al., 2004). Then an ensGene table of Ensembl gene and gene predictions (Yates et al., 2016) was obtained for each region using the calJac3 *Callithrix jacchus* reference genome. Ensembl IDs were extracted from this table, uploaded to the UCSC Genome Table Browser and an ensembltoGeneName table was obtained whose output contained gene symbols of associated target RAD loci. In the case that the UCSC Genome Table Browser did not recognize an Ensembl gene identifier of a target RAD locus, that locus was excluded. Because the resulting target gene list only contained protein-coding genes (see Results), we then used the Biomart portal (Smedley et al., 2015) to query the Ensembl Genes 84 database under the caljac3 resference genome for a list of gene symbols for all available protein coding genes.

3 | RESULTS

3.1 | Genomic coverage of RAD loci, sequencing depth, and WGA bias

The Illumina HiSeq2500 run produced a total of 111,557,292 single reads of which 71,692,114 were retained after demultiplexing and initial filtering. Table S1 shows the total number of single reads obtained for each individual along with the number of PE read pairs retained after demultiplexing. The average number of single reads obtained per individual was 6,429,482, with *C. jacchus* producing a higher number of reads than *C. penicillata*. After demultiplexing, the average number of retained PE read pairs per individual was 2,286,336, and further PCR duplicate filtering resulted in an average of 967,019 PE read pairs per individual. Only the forward member of each PE read pair was used for alignment to the calJac3 *C. jacchus* reference genome and Table S2 shows individual alignment results. The average number of aligned forward reads per individual was 542,090 and a similar average number of *C. jacchus* and *C. penicillata* reads aligned to the calJac3 reference genome. Table S2 also shows individual results for the total number of *cstacks* catalogued RAD loci, total number of SNPs contained within RAD loci, and average coverage per *cstacks* catalogued RAD locus. On average, there were 6,322 SNPs recovered per individual and each cataloged RAD locus had average coverage of 10.5x. Cpe018 dropped out of the analysis at this point because it did not pass STACKS filters for minimum stack size per RAD locus.

The comparison of genotype calls between cja014-WGA and cja014-nonWGA showed differences in genotyping between certain RAD loci present in both samples. For the STACKS analysis that excluded the *rxstacks* step, there was a total of 10,238 RAD loci processed for both cja014-WGA and cja014-nonWGA. Of these RAD loci, 246 (2.4%) had haplotypes that differed between the non-WGA and WGA samples of individual cja014. The majority of RAD loci that differed between cja014-WGA and cja014-non-WGA had unfavorable log likelihoods, below our set threshold of −8.0 for *rxstacks* filtering. In 62 cases, the WGA sample possessed heterozygous RAD locus allele while the non-WGA sample was homozygous for the locus. For 116 cases, cja014-nonWGA was heterozygous but cja014-WGA was homozygous. At the remaining loci, one or both samples possessed more than two alleles for a RAD locus.

For the STACKS analysis that included the *rxstacks* step, there was a total of 8,068 loci processed for cja014-WGA and cja014-nonWGA. A total of 113 (1.4%) RAD loci had haplotypes that differed between the WGA and non-WGA samples. Among these, 57 were cases of cja014-non-WGA being homozygous at a RAD loci and cja014-WGA being heterozygous. The opposite situation existed at 18 loci, and only two RAD loci possessed more than two alleles for either sample.

3.2 | Genomic diversity and population differentiation in *C. jacchus* and *C. penicillata*

After additional filtering, *populations* identified a total of 1,395 autosomal sites in RAD loci that were variable within or between species and that were represented in both *C. jacchus* and *C. penicillata*. Among these RAD loci, the two species were fixed for different alleles at 127 sites, but shared polymorphisms at 107 sites. For 868 sites, *C. jacchus* was fixed for a single allele and *C. penicillata* was polymorphic. On the other hand, *C. jacchus* was polymorphic and *C. penicillata* was fixed for a single allele at 293 sites.

Populations averages of genetic indices for *C. jacchus* and *C. penicillata* autosomes are shown in Table 2. *C. jacchus* and *C. penicillata* were similar to each other in their average values of P. However, *C. penicillata* showed greater variability than *C. jacchus*, as H_O and π were higher in the former species than in the latter. The two study species showed strong overall differentiation with a F_{ST} value of 0.266. (Point estimates of F_{ST} values for individual variable sites present in both *C. jacchus* and *C. penicillata* are available from JM).

Genome-wide smoothed intraspecific π and AMOVA interspecific F_{ST} values are shown in Figures 1 and 2, respectively. The mean smoothed π for *C. jacchus* was 0.001 with a range of 0–0.0091, and for *C. penicillata* the smoothed π average value was 0.0027 with a range of 0.0–0.0168. Values of π varied widely in both species. Kernel-smoothing of AMOVA F_{ST} values showed an average value of 0.266, with 42.4% of all variable sites present in both *C. jacchus* and *C. penicillata* showing no interspecific differentiation.

A total of 106 variable sites (7.59% of all variable sites) between *C. jacchus* and *C. penicillata* differed significantly from the genome-wide average, as their smoothed AMOVA F_{ST} $P < 0.05$ after bootstrapping. Of these significantly differentiated sites, smoothed AMOVA F_{ST} values ranged from 0.589 to 1. The majority of variable sites that showed significant differentiation between *C. jacchus* and *C. penicillata* also showed smoothed π values that were significantly different from the genome-wide average. For these sites, smoothed π either equaled 0 in both species or was lower than species-specific smoothed π averages. Additionally, observed heterozygosity values at significantly differentiated variable sites tended to be below 0.33 in both species.

Locus-specific and average values of absolute divergence (D_{XY}), relative divergence (F_{ST}) between *C. jacchus* and *C. penicillata* across the full set of RAD loci present in both species are listed in Table S3. Genome-wide D_{XY} values for these RAD loci are shown in Figure 3 and genome-wide adjusted F_{ST} values are shown in Figure S1. Mean D_{XY} for RAD loci with significantly diverged sites was 0.012 and the mean D_{XY} value for RAD loci without significantly diverged sites was 0.006. The former set of sites was significantly different from the latter set ($W = 11{,}673$, $P < 2.2 \times 10^{-16}$, Mann–Whitney two-tailed test). Mean adjusted F_{ST} for RAD loci with significantly diverged sites was 0.943 and the mean adjusted F_{ST} value for RAD loci without significantly diverged sites was 0.224. There was a significant difference between F_{ST} values for RAD loci with and without significantly diverged sites ($W = 6{,}902$, $P < 2.2 \times 10^{-16}$, Mann–Whitney two-tailed test).

Table S4 lists individual and average intraspecific π values for each RAD locus, and average π values for RAD loci with and without significantly diverged sites. Mean intraspecific π for both *C. jacchus* and *C. penicillata* for RAD loci without significantly diverged sites was 0.006. The *C. jacchus* intraspecific π at RAD loci with significantly diverged sites was 0.000 and for *C. penicillata* this value was 0.002. Nucleotide diversity between the two classes of RAD loci differed significantly for *C. penicillata* ($W = 295{,}640$, $P < 2.2 \times 10^{-16}$, Mann–Whitney two-tailed test) and the difference was also significant for *C. jacchus* ($W = 14{,}217$, $P < 2.2 \times 10^{-16}$, Mann–Whitney two-tailed test).

3.3 | GO analyses of significantly divergent loci

The initial list of target RAD loci associated with the 106 significantly diverged SNPs was used to query the USCS Genome Table Browser for ensGene table of Ensembl gene and gene predictions. The final ensembltoGeneName table contained a list of 38 gene symbols (Table S3), which composed our final target gene list. The background gene target list contained a total of 14,761 gene symbols (list available from JM). No GO enrichment was found for process, function, or component between RAD loci significantly diverged between *C. jacchus* and *C. penicillata*.

4 | DISCUSSION

4.1 | Challenges of RADseq and WGA in evolutionary studies

C. jacchus and *C. penicillata* are thought to be the most recently diverged of the *Callithrix* marmosets (Perelman et al., 2011; unpublished data, Malukiewicz), and we used PE-RADseq to simultaneously

TABLE 2 Summary of *population* averages of autosomal genetic indices for *C. jacchus* and *C. penicillata*

Species	Average number of individuals (SE)[a]	P[b] (SE)	H_O[c] (SE)	H_E[d] (SE)	π[e] (SE)	F_{IS} (SE)
C. jacchus	3.38 (0.0005)	0.9998 (0.0000)	0.0002 (0.0000)	0.0003 (0.0000)	0.0003 (0.0000)	0.0002 (0.0008)
C. penicillata	3.33 (0.0004)	0.9995 (0.0000)	0.0004 (0.0000)	0.0007 (0.0000)	0.0008 (0.0000)	0.0009 (0.0007)

[a]Standard error.
[b]Average frequency of major allele.
[c]Average observed heterozygosity.
[d]Average expected heterozygosity.
[e]Average nucleotide diversity.

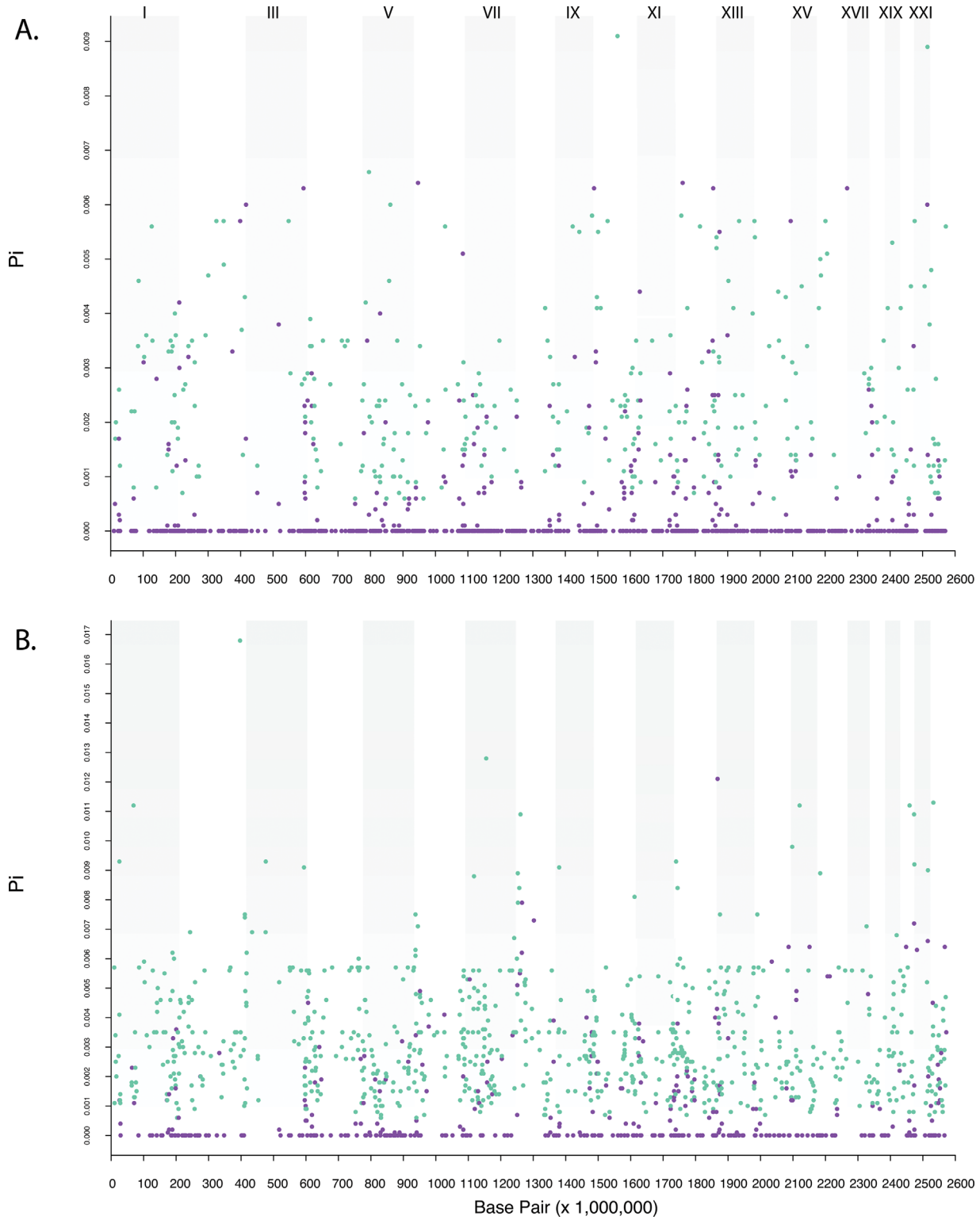

FIGURE 1 Genome-wide SNP-based smoothed π values for (A) *Callithrix jacchus* and (B) *C. penicillata*. In both panels, green indicates SNPs whose values were not statistically significant and purple indicates SNPs whose values were statistically significant after bootstrapping. Roman numerals indicate autosomes and the *x*-axis indicates base pairs along the genome for both panels A and B

discover and genotype 1000+ variable sites present in both species in order to examine genetic diversity and divergence. Prior genetic studies of *C. jacchus* and *C. penicillata* (e.g., de Carvalho, 2015; Malukiewicz et al., 2014, 2015) based on autosomal microsatellites, mtDNA, and the Y chromosome have been limited by their utilization of a smaller number of genetic markers in comparison to this RADseq analysis. Additionally, genetic markers like the non-recombining portion of the Y-chromosome and mtDNA are limited to genetic information from male and female lineages, respectively. Sampling a larger number of bi-parental genetic markers is advantageous in that it

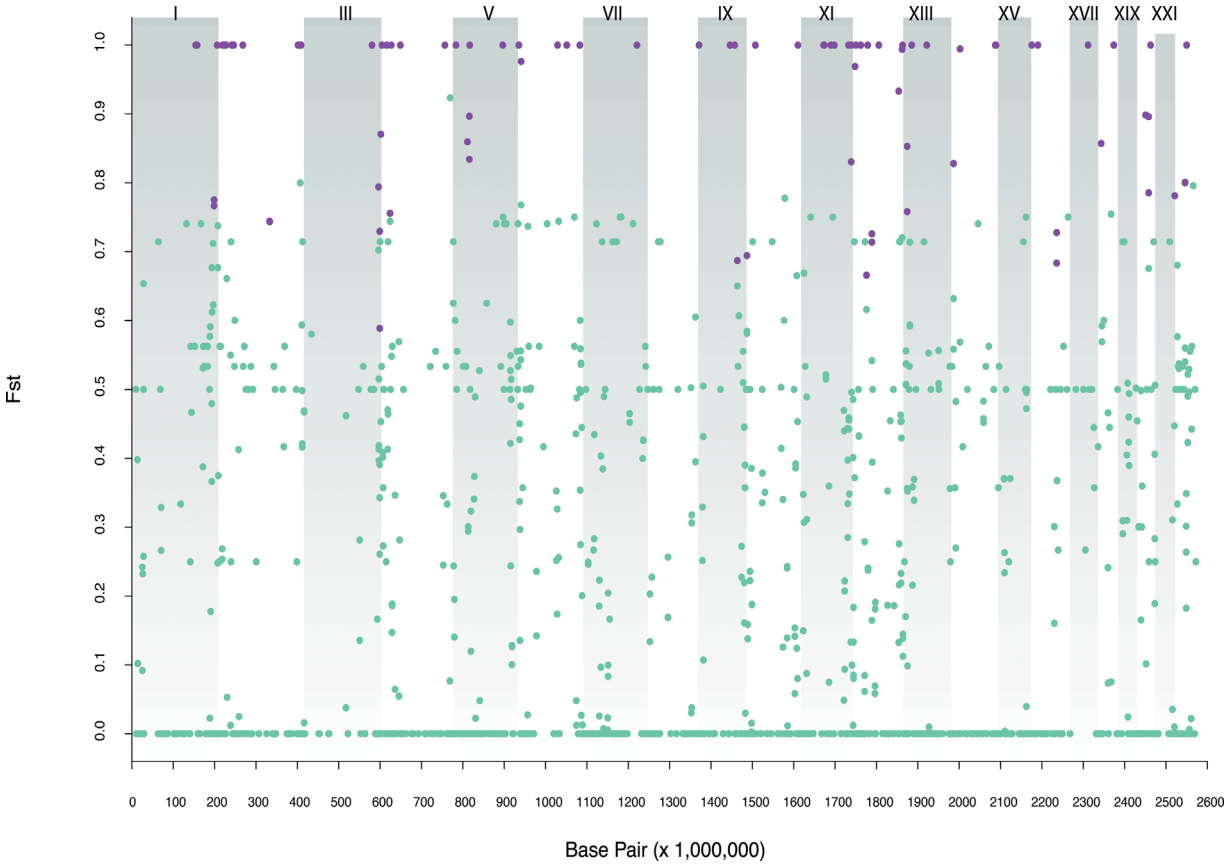

FIGURE 2 Variable site genome-wide kernel-smoothed relative divergence (AMOVA F_{ST}) for *C. jacchus* and *C. penicillata*. Green indicates sites whose values were not statistically significant and purple indicates sites whose values were statistically significant after bootstrapping. Roman numerals indicate autosomes. The *x*-axis indicates base pairs along the genome

gives a more accurate representation of genomic diversity for taxa of interest. Even the use of large microsatellite panels for divergence studies is arguably less time-efficient and garners less genetic information than NGS approaches such as RADseq. For example, genetic marker development by Malukiewicz et al. (2015) required separate testing of genetic marker amplification in *C. jacchus* and *C. penicillata*. Thus, RADseq may be more appropriate for bi-parental marker development in cases of evolutionary inference when a more time-efficient method and/or a large number of genetic markers are needed.

However, there are some challenges to RADseq that introduce inherent biases toward allele frequency estimates, which could affect downstream evolutionary analyses (see Table 1 in Mastretta-Yanes et al., 2015). For PE-RADseq, there is a significant positive correlation between read depth and restriction fragment length for RAD loci (Davey et al., 2013). This bias could result in underestimation of taxon diversity if short fragment RAD loci are filtered out during bioinformatics analysis due to shallow read depth. A mechanical DNA shearing step normally used during library construction may cause this RAD fragment length bias, but using an enzymatic shearing step, as we did, mitigates this issue (Davey et al., 2013).

Formation of PCR duplicates during enrichment of sequencing libraries is another common RADseq bias. Read counts form the basis of RAD locus genotyping and PCR duplicates may bias base calls as non-independent repeats of the same region (Hohenlohe et al., 2013). In turn, not filtering PCR duplicates from RADseq data will produce genotypes calls based on inflated confidence (Andrews et al., 2014). Hence, we took advantage of PE-RADseq to remove PCR duplicates from our dataset by filtering reads from the same RAD locus with the same length. Such reads are likely duplicates because random shearing is otherwise unlikely to cause equal-length fragments at a given locus (Andrews et al., 2014).

Finally, in mapping *C. penicillata* differences to the *C. jacchus* genome during our reference-genome guide STACKS analysis, there is a possibility of generating biased measures of genetic diversity for the former species. However, our mapping efficiency of the two species to the *C. jacchus* reference genome and intraspecific values of genetic diversity indices do not suggest mapping bias of one species over the other. For example, sequence reads from *C. jacchus* and *C. penicillata* showed highly similar mapping percentages to the *C. jacchus* genome (Table S2). Additionally, the average alignment rate of *C. penicillata* reads to the *C. jacchus* reference was slightly higher than that of *C. jacchus* reads. With a strong intraspecific mapping bias in favor of *C. jacchus* over *C. penicillata* reads, we would expect a lower mapping rate for *C. penicillata* to the calJac3 genome. Second, our RADseq-derived genetic diversity indices were very similar between the two species. If indeed there was a strong bias in favor of *C. jacchus* mapping to the *Callithrix* reference genome over *C. penicillata*, we would expect

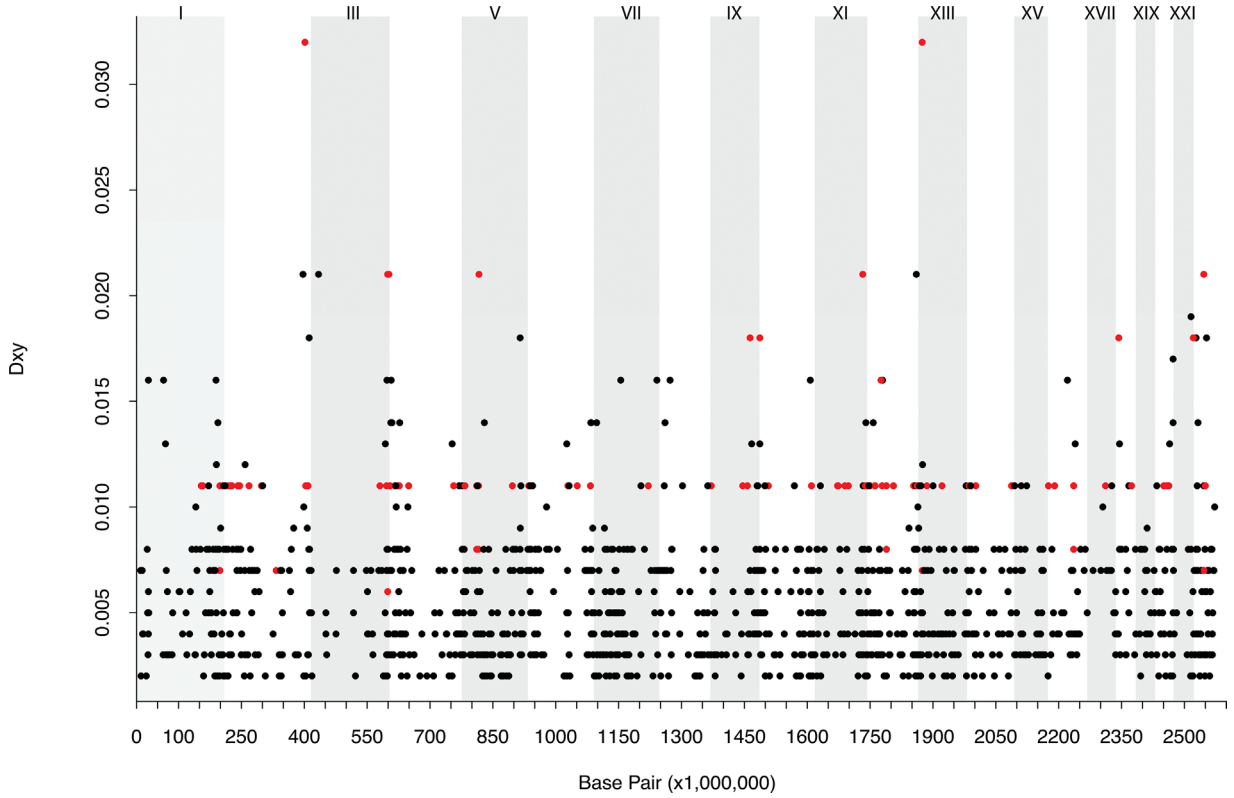

FIGURE 3 Genomic absolute divergence values (D_{XY}) between *C. jacchus* and *C. penicillata* for RAD loci. Red indicates RAD loci associated with significantly diverged sites and black indicates RAD loci associated with otherwise non-significant sites. Roman numerals indicate autosomes and the *x*-axis indicates base pairs along the genome

a tendency of the data to show lower diversity for *C. penicillata* in intraspecific genetic diversity index values.

4.2 | Intraspecific diversity in *C. jacchus* and *C. penicillata*

Despite small sample sizes of *C. jacchus* and *C. penicillata*, we were able to observe some general trends of intraspecific diversity for the two species. Notably, PE-RADseq analysis indicated higher levels of intraspecific genetic variation in *C. penicillata* than in *C. jacchus*. In comparing point estimates and kernel-smoothed estimates of nucleotide diversity in *C. penicillata* (average point estimate $\pi = 0.0008$ vs. kernel-smoothed average $\pi = 0.0027$) and *C. jacchus* (average point estimate $\pi = 0.0003$ vs. kernel-smoothed $\pi = 0.001$), the former estimates were lower than those of the latter. Catchen et al. (2013) favor kernel-smoothed estimates of diversity, as these are less susceptible to random biological or sequencing error. Thus, kernel-smoothed π estimates are more likely to reflect true genomic-wide intraspecific diversity in our *Callithrix* study species, and our kernel-smoothed results indicated that *C. penicillata* has over double the levels of genetic diversity than *C. jacchus*.

Overall, our results were close to the *C. jacchus* nucleotide levels ($\pi = 0.00123$) reported by Leffler et al. (2012). *C. jacchus* and *C. penicillata* show higher nucleotide diversity than *Homo sapiens* ($\pi = 0.0007–0.0012$), but marmosets and humans are on the lower end of the scale of nucleotide diversity values in primates (Leffler et al., 2012; Meyer et al., 2015). Leffler et al. (2012) pointed out that the evolutionary forces that maintain genetic diversity in species are still not well understood but genomic surveys of diversity, like those we report here for *Callithrix*, are an important first step to resolving this question. Our RADseq-derived results on intraspecific diversity also corroborate diversity patterns for *C. jacchus* and *C. penicillata* (i.e., lower genetic diversity in the former than the latter species) in the work of Malukiewicz et al. (2014), who used mitochondrial data from a much large set of samples for both *C. jacchus* and *C. penicillata*. The use of a 40+ panel of autosomal microsatellites by Malukiewicz et al. (2015), using the same sample set as Malukiewicz et al. (2014), also indicated a similar pattern of intraspecific diversity within the two species as observed with our RADseq dataset.

4.3 | Divergence between *C. jacchus* and *C. penicillata*

For our analysis of divergence, filtering our dataset for RAD loci available in both species resulted in a set of 1,395 variable sites. Among the variable sites represented in both species, we observed sites that were either fixed in both species, polymorphic in one species or the other, or polymorphic in both species. In future RADseq studies, increasing the sample size per species will be important to determine whether the fixed sites represented here are not an artifact of the small interspecific sample sizes of our study. Nonetheless, our current dataset contained 106 sites that showed statistically significant levels of relative divergence between *C. jacchus* and *C. penicillata*.

Our kernel-smoothed genome-wide relative divergence average (AMOVA F_{ST} = 0.266) was within range of that reported by Wall et al. (2016) for genome-wide differentiation between *Papio cynocephalus* and *Papio anubis* (0.23 < F_{ST} < 0.33). Additionally, Pastorini et al. (2009) also reported similar relative divergence levels (F_{ST} = 0.23) for *Eulemur mongoz* and *E. fulvus* as those for marmosets and baboons, though the *Eulemur* estimate was microsatellite-based. Finally, our estimates of genome-wide relative divergence between *C. jacchus* and *C. penicillata* were within the range of estimates reported by Cruickshank and Hahn (2014) for a diverse array of diverged species. However, relative measures of species divergence should be interpreted with care as F_{ST} measures, which are dependent on within species diversity, can be influenced by a number of different factors such as gene flow, recombination rates, and background selection (Cruickshank & Hahn, 2014). Instead, relative divergence should be considered in concert with absolute divergence, which is not based on within-species polymorphisms (Cruickshank & Hahn, 2014).

Our comparison of relative and absolute divergence for *C. jacchus* and *C. penicillata* is in line for theoretical expectations under speciation with gene flow. Our data showed the co-occurrence of high absolute divergence and high relative divergence between *C. jacchus* and *C. penicillata* within the same genomic regions. However, many of these regions showed low values of within-species polymorphism, a pattern expected under the effects of linked selection via hitchhiking or background selection (Cruickshank & Hahn, 2014). Similar patterns between D_{XY}, $F_{ST,}$ and within species diversity were observed for two flycatcher species (*Ficedula*) (Ellegren et al., 2012), which Cruickshank and Hahn (2014) interpreted as a signal of selection acting prior to species divergence.

Whether speciation with gene flow is driving *C. jacchus* and *C. penicillata* divergence at RAD loci associated with significantly diverged variable sites, needs to be further investigated. In particular, sampling variation due to small sample size may have affected our results by increasing the variance of measured genetic diversity and divergence indices in comparison to measures from larger intraspecific sample sizes. Second, recombination levels show a negative relationship with relative divergence, and high F_{ST} measures maybe the result of low levels of recombination (Burri et al., 2015; Cruickshank & Hahn, 2014). Recombination rates can be estimated either through the use of a high-density linkage map (e.g., Malinsky et al., 2015) or by using resequenced whole genomes (e.g., Feulner et al., 2015), but neither of these resources is currently available for *Callithrix*. Thus, due to the limitations of our study in sample size and knowledge of recombination rates, more data are needed to determine whether *C. jacchus* and *C. penicillata* diverged under speciation-with-gene-flow and whether our significantly diverged sites are located within candidate speciation genes.

4.4 | Implications for primate speciation genomics and future research

A considerable amount of scientific work has already examined genome-level barrier evolution with several key studies from a variety of organisms reporting heterogeneous genome-wide differentiation levels (e.g., Ellegren et al., 2012; Renaut et al., 2013). However, primates remain severely underrepresented in the burgeoning field of speciation genomics, with the very recent work of Wall et al. (2016) representing one of the few examples of genomic work on primate divergence. Further, the evolutionary drivers of genomic divergence during speciation and genomic signature of speciation genes remain contested.

To fill the above gaps in understanding speciation as a genome-level process ideally takes an unbiased genomic account of divergence stages of the same speciation process—an unrealistic task for most primates (Seehausen et al., 2014). However, one way to overcome this challenge for primates is by studying a speciation continuum (SC) of closely related species differing in their extent of divergence. This approach makes it possible to reconstruct the sequential origin of isolating barriers and identify those barriers that are the cause rather than a consequence of speciation (Seehausen et al., 2014). Additionally, including lineages in early-stage divergence is crucial for the SC approach. Young lineages are useful for differentiation between primary and secondary (post-speciation) isolating barriers because divergent loci in these lineages likely arose before the completion of speciation (Schield et al., 2015).

Our RADseq data represent the first crucial step toward better integrating primates, especially New World taxa, into the field of speciation genomics. First, *C. jacchus* and *C. penicillata* are sister species and the youngest members of this genus. Indeed, the discovery of fixed and variable sites between these species the results of this study show that these species are a natural starting point for primate speciation genomics via the SC approach. Older *Callithrix* species possess historically allopatric ranges, are adapted to different environments, and show increasing divergence (Coimbra-Filho, Pissinatti, & Rylands, 1993; Rylands et al., 2009). With the development of additional genomic resources and expanded intraspecific and geographical samplings for the other *Callithrix* species, this genus stands to become a strong primate model for SC speciation genomics.

ACKNOWLEDGMENTS

Thank you to Leanne Nash, Phil Hedrick, and Michael Rosenberg for comments on earlier drafts of this work. We are indebted to Dr. Jeffrey French for providing captive *C. penicillata* samples and IBAMA CETAS for access to their premises and collections. Drs. Vanner Boere and Ita de Oliveira e Silva gave immense support in obtaining wild *C. penicillata* samples. This research was supported by National Science Foundation DDIG proposal 1061508 to ACS and JM, a Fulbright Fellowship to JM, ASU SOLS FIGG grant, ASU GPSA Jump Start and Research Grants, ASU Chapter of Sigma Xi, and an International Primatological Society Research Grant to JM.

REFERENCES

Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G. (2014). Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. 2014. *Molecular Ecology*, *23*, 5943–5946.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376.

Bell, R. C., Drewes, R. C., & Zamudio, K. R. (2015). Reed frog diversification in the Gulf of Guinea: Overseas dispersal, the progression rule, and in situ speciation. *Evolution*, *69*, 904–915.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, *25*, 1656–1665.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140.

Coimbra-Filho, A. F., Pissinatti, A., & Rylands, A. B. (1993). Experimental multiple hybridism and natural hybrids among *Callithrix* species from eastern Brazil. In A. B. Rylands (Ed.), *Marmosets and tamarins: Systematics, behaviour, and ecology* (pp. 95–120). New York: Oxford University 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.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology*, *22*, 3151–3164.

de Carvalho, R. D. (2015). Conservação do saguis-da-serra-escuro (*Callithrix aurita* (Primates))—Análise molecular e colormétrica de populações do gênero *Callithrix* e seus híbridos. PhD Dissertation: Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro State.

De Mita, S., & Siol, M. (2012). EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, *13*, 27.

Digby, L. J., Ferrari, S. F., & Saltzman, W. (2007). Callitrichines: The role of competition in cooperatively breeding species. In C. J. Campbell, A. Fuentes, K. C. MacKinnon, M. Panger, & S. K. Bearder (Eds.), *Primates in perspective* (pp. 85–105). New York: Oxford University Press.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*, 48.

Eden, E., Lipson, D., Yogev, S., & Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology*, *3*, e39.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backstrom, N., Kawakami, T., … Wolf J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*, 756–760.

Etter, P. D., & Johnson, E. (2012). RAD paired-end sequencing for local de novo assembly and SNP discovery in non-model organisms. *Methods in Molecular Biology*, *888*, 135–151.

Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 73–97.

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., … Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet*, *11*, e1004966.

Fuzessy, L. F., Silva, I. O., Malukiewicz, J., Silva, F. F., Ponzio, M. C., Boere, V., & Ackermann, R. R. (2014). Morphological variation in wild marmosets (*Callithrix penicillata* and *C. geoffroyi*) and their hybrids. *Evolutionary Biology*, *41*, 480–493.

Guo, Y., Yuan, H., Fang, D., Song, L., Liu, Y., Liu, Y., … Zhang, H. (2014). An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (*Oryza sativa* L.) F2 population. *BMC Genomics*, *15*, 956.

Hershkovitz P. (1977). *Living new world monkeys (Platyrrhini): With an introduction to primates*. Chicago: University of Chicago Press.

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., … Luikart G. (2013). Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecuarl Ecology*, *22*, 3002–3013.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K., Sugnet, C., Haussler D., & Kent, W. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*, D493–D496.

Krueger, F., Andrews, S. R., & Osborne, C. S. (2011). Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE*, *6*, e16607.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, R25.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., … Przeworski, M. (2012). Revisiting an old riddler: What determines genetic diversity levels within species. *PLoS Biology*, *9*, e1001388.

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., … Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*, 1493–1498.

Malukiewicz, J., Boere, V., Fuzessy, L. F., Grativol, A. D., Silva, I. O., Pereira, L. C. M., … Stone, A. C. (2014). Hybridization effects and genetic diversity of the common and black-tufted marmoset (*Callithrix jacchus* and *Callithrix penicillata*) mitochondrial control region. *American Journal of Physical Anthropology*, *155*, 522–536.

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

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2014). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*, 244–257.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Pinero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*, 28–41.

Mendes, S. L., Vielliard, J. M. E., & de Marco, M. Jr. (2009). Vocal identify of the *Callithrix* species (Primates, Callitrichidae). In S. M. Ford, L. M. Ford, & L. C. Davis (Eds.), *The smallest anthropoids: The Marmoset/Callimico radiation* (pp. 63–84). New York: Springer.

Meyer, W. K., Venkat, A., Kermany, A. R., Geijn, B., Zhang, S., & Przeworski, M. (2015). Evolutionary history inferred from the de novo assembly of a nonmodel organism, the blue-eyed black lemur. *Molecular Ecology*, *24*, 4392–4405.

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

Pastorini, J., Zaramody, A., Curtis, D. J., Nievergelt, C. M., & Mundy, N. I. (2009). Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evolutionary Biology*, *9*, 32.

Perelman, P., Johnson, W. E., Roos, C., Seuanez, H. N., Horvath, J. E., Moreira, M. A. M., ... Pecon-Slattery, J. (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.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., ... Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827.

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*, 94.

Ruiz-Miranda, C. R., Affonso, A. G., Martins, A., & Beck, B. (2000). Distribuição do sagui (*Callithrix jacchus*) nas áreas de ocorrência do mico-leão-dourado (*Leontopithecus rosalia*) no estado do Rio de Janeiro. *Neotropical Primates*, *8*, 98–101.

Ruiz-Miranda, C., Affonso, A., Morais, M., Verona, C., Martins, A., & Beck, B. (2006). Behavioral and ecological interactions between reintroduced golden lion tamarins (*Leontopithecus rosalia Linnaeus*, 1766) and introduced marmosets (*Callithrix* spp, Linnaeus, 1758) in Brazil's Atlantic coast forest fragments. *Brazilian Archives of Biology and Technology*, *49*, 99–109.

Rylands, A. B., & de Faria, D. S. (2003). Habitats, feeding ecology, and home range size in the genus *Callithrix*. In A. B. Rylands (Ed.), *Marmosets and tamarins: Systematics, behaviour, and ecology* (pp. 262–272). New York: Oxford University Press.

Rylands, A. B., Coimbra-Filho, A. F., & Mittermeier, R. A. (2009). The systematics and distributions of the marmosets (*Callithrix*, *Callibella*, *Cebuella*, and *Mico)* and *Callimico* (*Callimico*) (Callitrichidae, Primates). In S. M. Ford, L. M. Ford, & L. C. Davis (Eds.), *The smallest anthropoids: The Marmoset/Callimico radiation* (pp. 25–61). New York: Springer.

Schield, D. R., Card, D. C., Adams, R. H., Jezkova, T., Reyes-Velasco, J., Proctor, F. N., ... Castoe T. A. (2015). Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, *83*, 213–223.

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

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Review Genetics*, *15*, 176–192.

Smedley, D., Haider, S., Durink, S., Pandini, L., Provero, P., Allen, J., ... Kasprzyk, A. (2015). The Biomart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, *43*, W589–W598.

Smith, T. E., Schaffner, C. M., & French, J. A. (1997). Social and developmental influences on reproduction in female Wied's black tufted-ear marmosets (*Callithrix kuhli*). *Hormones and Behavior*, *31*, 159–168.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). EzRAD: A simplified method for genomic genotyping in non-model organisms. *Peerj*, *1*, e203.

Vinyard, C. J., Wall, C. E., Williams, S. H. et al. (2009). The evolutionary morphology of tree gouging in marmosets. In S. M. Ford, L. M. Ford, & L. C. Davis (Eds.), *The smallest anthropoids: The Marmoset/Callimico radiation* (pp. 395–408). New York: Springer.

Weir, B. S. (1996). *Genetic analyis II*. Sunderland: Sinauer. 376 p.

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

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

Worley, K. C., Warren, W. C., Rogers, J., Locke, D., Muzny, D. M., Mardis, E. R., ... Wilson, R. K. (2014). The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics*, *46*, 850–857.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., ... Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, *44*, D710–D716.

Zago, L. S. 2012. Fatores determinantes no uso do espaço por *Callithrix* penicillata (E. Geoffroy, 1812) introduzidos em fragmento urbano. Masters Thesis: Universidade Federal do Paraná. Curitiba, Paraná.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Malukiewicz J, Guschanski K, Grativol AD, Oliveira MAB, Ruiz-Miranda CR, and Stone AC. Application of PE-RADSeq to the study of genomic diversity and divergence of two Brazilian marmoset species (*Callithrix jacchus* and *C. penicillata*). *Am J Primatol*. 2017;79:e22587. doi:10.1002/ajp.22587.