Supplemental Information for:

Ancient and recent introgression in hybridizing Mesoamerican crocodiles (*Crocodylus acutus x Crocodylus moreletii*) from Belize

Helen W. Sung[1*], Marisa Tellez[2], Floyd A. Reed[1]

Supplemental Methods S1:

Sample Collection, DNA extraction, library preparation, and sequencing

Crocodiles were captured either by hand or catchpole and morphometric and environmental data were collected at each capture site. Tissue clips were obtained from the lower tail scutes following a unique numerical pattern for each individual and were immediately placed into sterile Eppendorf tubes containing 95% ethanol for subsequent DNA extraction. Before releasing the crocodiles at the capture site, we assessed them for health, visible injuries, and external parasites. We extracted DNA using the NucleoMag™ tissue DNA kit (Macherey-Nagel, Düren, Germany) following manufacturer instructions using a KingFisher® Flex instrument (Thermo Scientific, Wilmington DE, USA). To ensure we recovered high-purity DNA, we reran steps 2-7 from the extraction protocols using the final eluted DNA preserved in 100 µL of Elution Buffer MB6. We quantified DNA concentration using a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and normalized concentrations to 20 ng/µl using sterile water before library preparation. We sent the normalized samples to the BadDNA EHS lab at the University of Georgia to prepare the 3RAD libraries using the high-throughput protocol outlined in Bayona-Vásquez et al. (2019). Briefly, 3RAD is a variant of duel-digest RADseq that uses 3 restriction enzymes and makes use of simultaneous restriction digestion and ligation to efficiently add the adapters onto each piece of DNA with compatible ends. Prior to library preparation, we sent out a subset of 8 samples to optimize which combination of enzymes best fit our samples and selected Adapterama III Design 2. Once optimized, 5.0 µL input genomic DNA of each sample was set up to be digested using restriction enzymes MspI, ClaI, and BamHI-HF (New England Biolabs, Ipswich, MA, USA). Internal indexes were built into adapters to uniquely identify individual samples in order to pool up to 96 samples by volume following ligation. To create each of these pools, 250 µL of ligation products were cleaned with 300 µL of SpeedBeads (Thermo Fisher Scientific, Waltham, MA, USA), purified, resuspended in 30 µL of dH2O, and combined into one tube totaling 60 µL. Then 20 µL of ligated DNA fragments were used for PCR cycled at 98°C for 1 min.; then, 12 cycles of: 98°C for 20 sec., 60°C for 15 sec., 72°C for 30 sec.; 72°C for 5 min. and held at 15°C. To each of the pooled-ligation PCR reactions, 50 µL of dH2O, + 100 µL SpeedBeads were added, purified as normal, and resuspended in 20 µL dH2O. Pooled beads from all 3 PCR replicates (60 µL total) were then placed on a magnet and pulled for all liquid (~55 µL), leaving the beads behind. To ensure each pool worked, 5 µL of the pooled PCR product was run on a 1.5% agarose gel to visually confirm the expected size distribution. The pools of PCR products were quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), combined with SpeedBead at a 1:1.25 DNA to SpeedBead ratio, and size selected for 550 bp on a Pippin Prep (Sage Science, Beverly, MA, USA). The final pool of libraries was quantified on a Qubit 4 Fluorometer, mixed in appropriate proportions with other libraries, and sent to be 150 bp paired-end (PE150) sequencing on an Illumina HiSeq X using TruSeq primers and dual (8nt) index reads.

SNP calling and filtering

In iPyrad, low-quality sites (Q < 20) were trimmed from the 3' ends of reads. Reads containing more than five ambiguous (N) sites or were less than 75 bp in length were also removed. The remaining reads were then mapped to the Crocodylus porosus reference genome (GCA_001723895.1; St John et al., 2012).

Clusters with less than six reads were excluded to ensure accurate base calls, and we set a maximum cluster depth within samples to 10,000. We used the default setting of 0.05 for the maximum fraction of uncalled ('N's') and heterozygous ('H's') bases allowed in the consensus sequences. We set the minimum number of samples per locus for output as 218, the maximum number of SNPs per locus as 0.2, and the maximum number of heterozygous sites per locus as 0.5. Our post-assembly dataset (DS1) produced a total of 89,083 SNPs with 30.03% missing sites.

Parameter file used in the iPyrad analysis of SNP dataset 1 (DS1).

------- ipyrad params file (v.0.7.30)----------	
noreponly_v2	## [0] [assembly_name]: Assembly name. Used to name output directories for assembly steps
./	## [1] [project_dir]: Project dir (made in curdir if not present)
	## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
	## [3] [barcodes_path]: Location of barcodes file
	## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
reference	## [5] [assembly_method]: Assembly method
../reference/reference.fna	## [6] [reference_sequence]: Location of reference sequence file
pair3drad	## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
ATCGG,GATCC	## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
5	## [9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
33	## [10] [phred_Qscore_offset]: phred Q score offset (33 is default and very standard)
6	## [11] [mindepth_statistical]: Min depth for statistical base calling
6	## [12] [mindepth_majrule]: Min depth for majority-rule base calling
10000	## [13] [maxdepth]: Max cluster depth within samples
0.9	## [14] [clust_threshold]: Clustering threshold for de novo assembly
1	## [15] [max_barcode_mismatch]: Max number of allowable mismatches in barcodes
2	## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
75	## [17] [filter_min_trim_len]: Min length of reads after adapter trim
2	## [18] [max_alleles_consens]: Max alleles per site in consensus sequences
0.05	## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus (R1, R2)
0.05	## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1, R2)
218	## [21] [min_samples_locus]: Min # samples per locus for output
0.2	## [22] [max_SNPs_locus]: Max # SNPs per locus (R1, R2)
8	## [23] [max_Indels_locus]: Max # of indels per locus (R1, R2)
0.5	## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus (R1, R2)
0, 0, 0, 0	## [25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see docs)
0, 0, 0, 0	## [26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2)
*	## [27] [output_formats]: Output formats (see docs)
	## [28] [pop_assign_file]: Path to population assignment file

Supplemental Tables:

Supplemental Table S1. Summary of the total sampled American crocodile (*C. acutus*), Morelet's crocodile (*C. moreletii*), and putative hybrids from Belize (2014 - 2021). Individuals are grouped by Morphological Species and by Sampling localities.

		Morphological Species			
Sampling Locality	**Total Sampled**	***C. moreletii* (CM)**	**Mainland *C. acutus* (MCA)**	**Cayes *C. acutus* (CA)**	**Putative Hybrids (HY)**
Belize River Watershed (BRW)	18	8	3	0	7
Cockscomb Basin (CB)	67	17	24	0	26
Coastal Cayes (CC)	39	0	0	39	0
Chiquibul Forest (CF)	5	5	0	0	0
North Belize District Watershed (NBDW)	1	0	1	0	0
Northern Cayes (NC)	38	0	0	38	0
New River Watershed (NRW)	62	46	2	0	14
North/South Lagoon Watershed (NSLW)	6	4	0	0	2
Northern Toledo Watershed (NTW)	7	1	3	0	3
Rio Honda Watershed (RHW)	3	2	0	0	1
Southern Toledo Watershed (STW)	14	0	14	0	0
Unknown Origins	13	4	3	1	5
Total	273	87	50	78	58

Supplemental Table S2. Data table of morphometric and environmental metadata collected from all *C. acutus, C. moreletii,* and their hybrids (n=273). Metadata for the filtered sample data set (n=242) used in the manuscript analyses can be found on DataDryad (https://doi.org/10.5061/dryad.3bk3j9kt9) and GitHub (https://github.com/helenwsung/Croc-ADMIXTURE-scripts.git). To preserve sensitive occurrence data for threatened/at-risk species, we generalized the precision of the geographic coordinates (Lat/Long) by reducing the number of decimal places to 0.1 decimal degrees as recommended by the Guide to Best Practices for Generalising Sensitive Species Occurrence Data [Chapman AD (2020) Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. https://doi.org/10.15468/doc-5jp4-5g10.].

Supplemental Table S3. Summary of the total number of working samples (n = 242) after data filtering. Individuals are grouped by Morphological species and Sampling localities. Individual sample metadata can be found in Supplementary Table S1.

		Morphological Species Group			
Sampling Locality	**Total Sampled**	***C. moreletii* (CM)**	**Mainland *C. acutus* (MCA)**	**Cayes *C. acutus* (CA)**	**Putative Hybrids (HY)**
Belize River Watershed (BRW)	17	7	3	0	7
Cockscomb Basin (CB)	61	15	23	0	23
Coastal Cayes (CC)	36	0	0	36	0
Chiquibul Forest (CF)	5	5	0	0	0
Northern Cayes (NC)	29	0	0	29	0
New River Watershed (NRW)	55	42	2	0	11
North/South Lagoon Watershed (NSLW)	6	4	0	0	2
Northern Toledo Watershed (NTW)	7	1	3	0	3
Rio Honda Watershed (RHW)	2	1	0	0	1
Southern Toledo Watershed (STW)	14	0	14	0	0
Unknown	10	3	2	1	4
Total	242	78	47	66	51

Supplemental Table S4. Output of filtered VCF files used in analyses.

File name	SNP count	# of individuals	% missing data
DS1_unfiltered_vcf	89,083	273	0
filtered.75	37,843	242	10.54
filtered.80	34,482	242	9.522
filtered.85	28,316	242	8.118
filtered.95	5,212	242	3.252
filtered.75_LDpruned	12,866	242	10.19
filtered.85_LDpruned	11,296	242	8.144

Supplemental Table S5. ADMIXTURE groups based on shared genomic ancestry proportions (Q-score) at K=3. We defined 'Hybrid_moreletti' had between 60% < Q < 90% of the *C. moreletii* genetic cluster (Cluster 1), 'Hybrid_acutus_A' for 60% < Q < 90% of the first *C. acutus* cluster (Cluster 2) which predominately was represented by *C. acutus* from the mainland populations, 'Hybrid_acutus_B' for 60% < Q < 90% of the second *C. acutus* cluster (Cluster 3) which predominately was represented by *C. acutus* from the Cayes populations. 'Hybrid_acutus' represents the admixed individuals with Q < 40% from the *C. moreletii* genetic cluster but have Q > 60% between the two *C. acutus* clusters combined. We referred to individuals between 40% < Q < 60% as 'Hybrid' due to the uncertainty of whether they were truly F_1 Hybrids. Output for each individual sample's Q-scores, ADMIXTURE group assignments, and morphological species group assignments can be found in **Supplemental Table S6.**

Group Assignments	**Cluster 1 proportion**	**Cluster 2 proportion**	**Cluster 3 proportion**	**n**
Pure *C. moreletii*	>0.90	<0.10	<0.10	90
Pure *C. acutus lineage A* (Mainland)	<0.10	>0.90	<0.10	17
Pure *C. acutus lineage B* (Cayes)	<0.10	<0.10	>0.90	61
Hybrid_moreletii	0.60 < Q < 0.90	<0.40	<0.40	23
Hybrid_acutus_A	<0.40	0.60 < Q < 0.90	<0.40	22
Hybrid_acutus_B	<0.40	<0.40	0.60 < Q < 0.90	9
Hybrid_acutus	<0.40	<0.60	<0.60	8
Hybrid	0.40 < Q < 0.60	<0.60	<0.60	12

Supplemental Table S6. Individual sample data (n=242) including morphological species group assignments, sampling locality, ADMIXTURE group assignments, cluster assignment, and Q-scores proportions.

Supplemental Table S7. Estimates of admixture age inferred from linkage disequilibrium decay. Recombination fractions (r) at LD half-decay points were calculated using an average recombination rate of 7.53 cM/Mb derived from sex-specific linkage map lengths of *Crocodylus porosus* and converted to physical distance using mapped marker positions. Admixture times were expressed in years assuming a generation time of 8–25 years for *C. acutus.*

Supplemental Table S8. Output of the ancestry painting plot for Figure 6 in the main text. The table includes the sample ids, the number of homozygous sites for *C. moreletii* (n_hom_P1), number of heterozygous sites (n_het), number of homozygous sites for *C. acutus* (n_hom_P2), and the proportion of heterozygosity at the genotypic sites (p_het). Sample names are listed in the order they appear on Figure 6 and are similarly color coordinated for unadmixed *C. moreletii* (red) and unadmixed *C. acutus* (blue), but hybrids are given in black. Unadmixed species are defined as having a Q-score > 98% determined by our ADMIXTURE results, and for *C. acutus* we combined individuals from both Cayes and Mainland lineages.

Supplemental Table S9. Population genetic summary statistics comparing *C. moreletii,* Mainland *C. acutus* lineage (acutus_A), and Cayes *C. acutus* lineage (acutus_B). Population groups only incorporated 'pure' individuals with a Q-score > 90% for respective groups determined via ADMIXTURE results at K=3. For each population, the table shows the number of individuals, observed (H_O) and expected (H_S) heterozygosity, allelic richness, and inbreeding coefficient (F_{IS}). Across all populations observed heterozygosity was lower than expected (t = 241.1, df = 84947, p-value < 2.2e-16).

Population	**Sample Size**	**H_o**	**H_s**	**Allelic Richness**	**F_{IS}**
Acutus_A (Mainland)	17	0.08864163	0.1815246	1.583719	0.4534186
Acutus_B (Cayes)	61	0.09596688	0.2461285	1.624927	0.4455452
C. moreletii	90	0.09569294	0.2553162	1.633896	0.4510475

Supplemental Table S10. Population genetic summary statistics comparing pure *C. acutus* vs pure *C. moreletii*. Population groups only incorporated 'pure' individuals with a Q-score > 90% for respective groups determined via ADMIXTURE results at K=3. *C. acutus* includes all pure individuals from acutus_A (Mainland) and acutus_B (Cayes), in addition to individuals with a combined Q-score > 90% (and <10% *C. moreletii*) when summing the Q-scores between the two *C. acutus* lineages. Across all populations observed heterozygosity was lower than expected (t = 212.32, df = 56631, p-value < 2.2e-16)

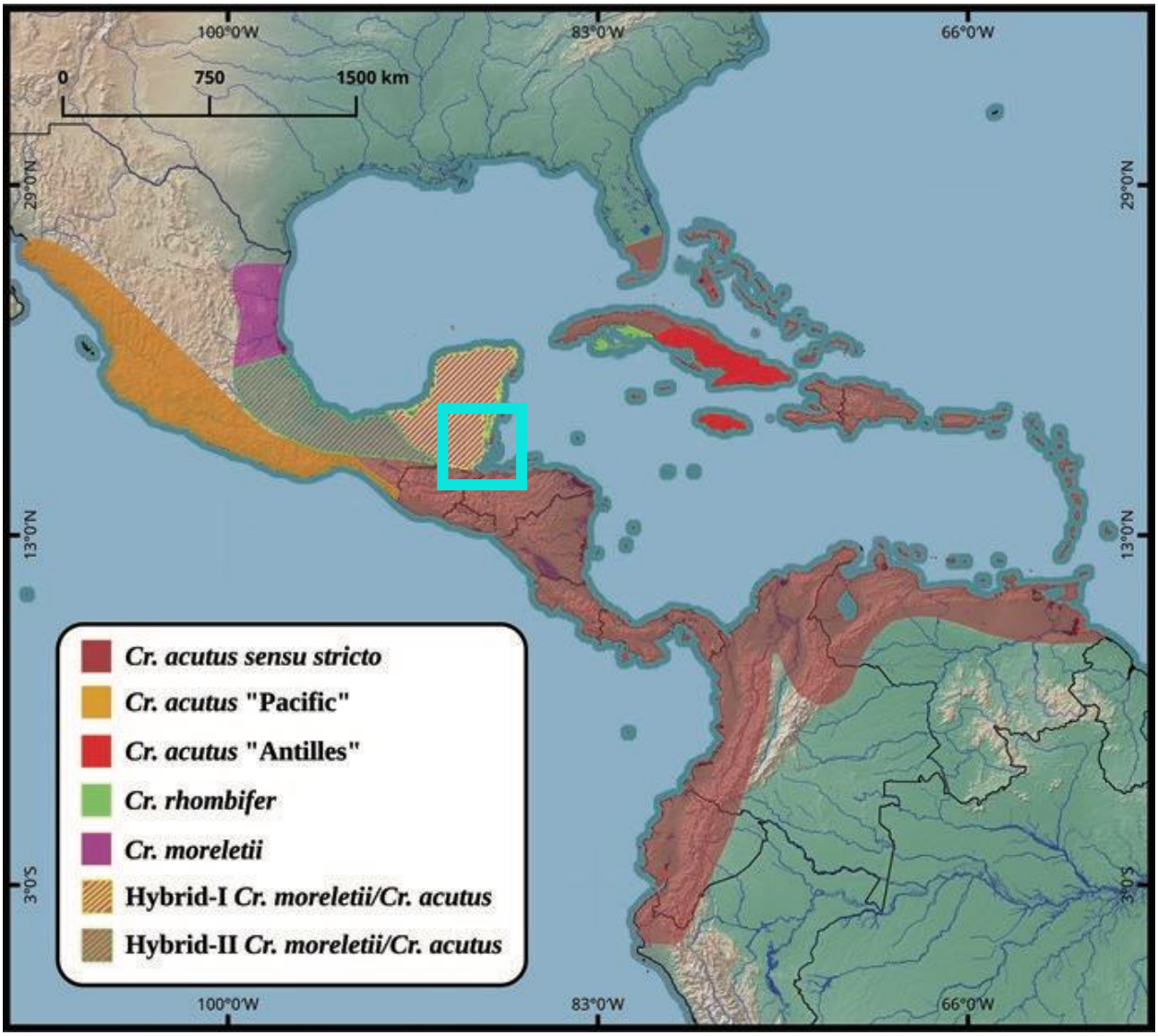
Population	**Sample Size**	**H_o**	**H_s**	**Allelic Richness**	**F_{IS}**
C. acutus	89	0.09425501	0.2456719	1.789168	0.4314963
C. moreletii	90	0.09937433	0.2507669	1.792689	0.4330564

Supplemental Table S11. Mean and weighted F_{ST} estimates comparing *C. moreletii,* Mainland *C. acutus* lineage (acutus_A), and Cayes *C. acutus* lineage (acutus_B). Population groups only incorporated 'pure' individuals with a Q-score > 90% for respective groups determined via ADMIXTURE. 'Combined *C. acutus'* includes all individuals from acutus_A (Mainland) and acutus_B (Cayes), in addition to individuals with a Q-score > 90% (and <10% *C. moreletii*) when summing the Q-scores between the two *C. acutus* lineages. Hybrids are all admixed individuals with a proportion of shared ancestry between species groups.

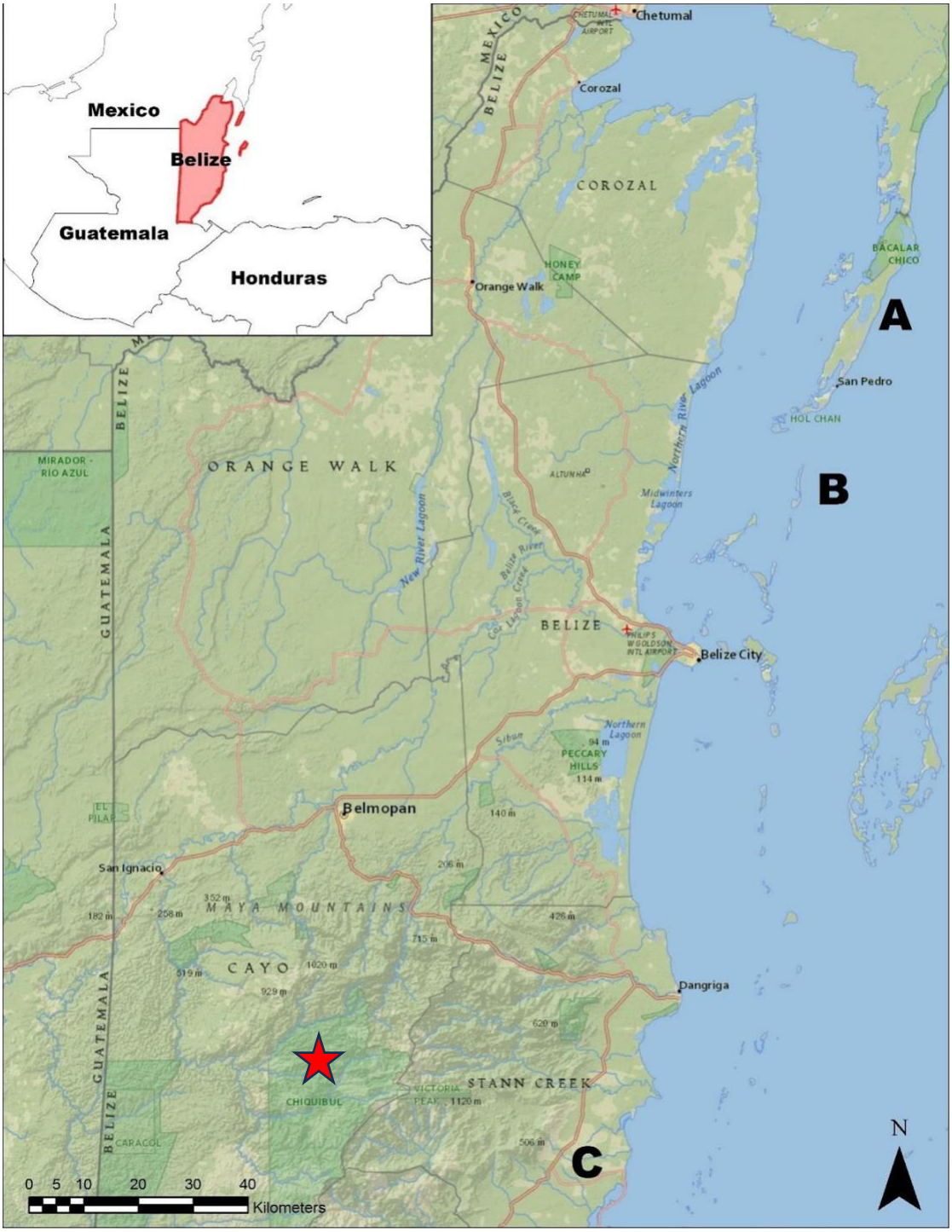
Population 1	**Population 2**	**Weir and Cockerham mean F_{ST} estimates**	**Weir and Cockerham weighted F_{ST} estimate**
combined *C. acutus*	*C. moreletii*	0.42289	0.74569
combined *C. acutus*	Hybrid	0.20178	0.30851
acutus_B (Cayes)	*C. moreletii*	0.45315	0.76006
acutus_B (Cayes)	Hybrid	0.21012	0.31704
acutus_B (Cayes)	acutus_A (Mainland)	0.094842	0.17608
C. moreletii	Hybrid	0.21694	0.35594
C. moreletii	acutus_A (Mainland)	0.49362	0.7593
acutus_A (Mainland)	Hybrid	0.16678	0.25033

Supplemental Figures:

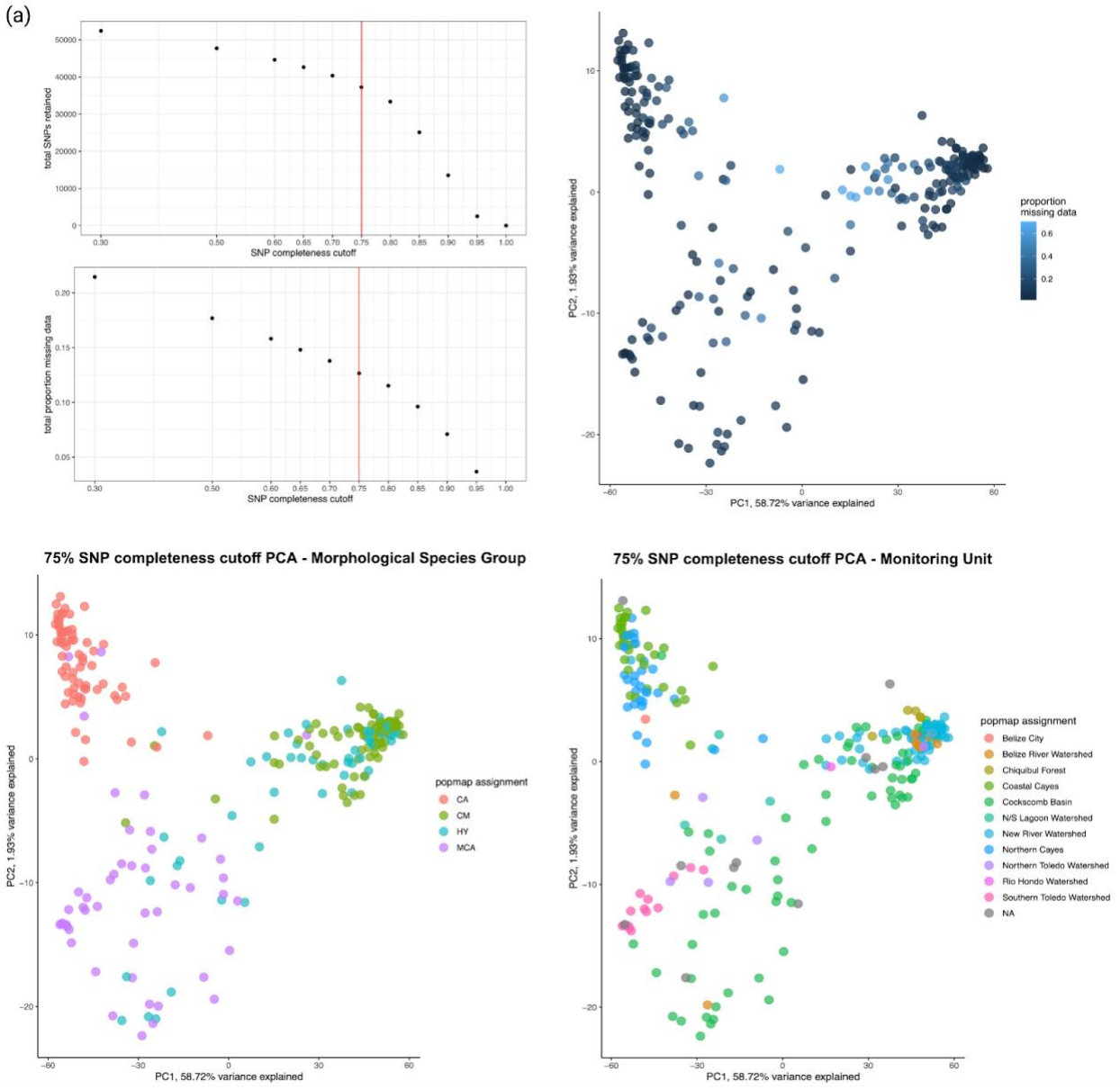
Supplemental Figure S1: Modified from Figure 4.8 taken from Muniz et al. (2021). Geographic distribution of *Crocodylus acutus, C. moreletii, and C. rhombifer*. Map produced with the QGIS software and species distribution shapefile provided by the IUCN Red List. The light blue box indicates samples collected in the present study (Belize, Central America) within the range of 'Hybrid-I *Cr. moreletii/Cr. acutus*'.



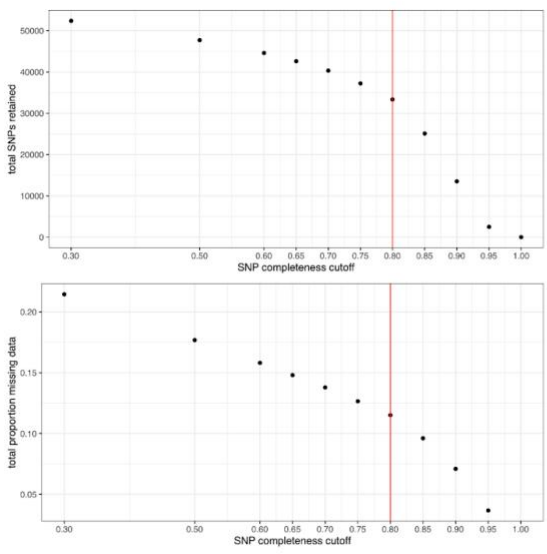
Supplemental Figure S2: Geographical map of Belize modified from Boucher et al. (2020). Letters dictate critical coastal nesting habitat for *C. acutus;* (A) Ambergris Caye, (B) Caye Caulker, and (C) Stann Creek District. A red star highlights sampling locality, Chiquibul Forest (CF) located west of the Maya Mountain range.

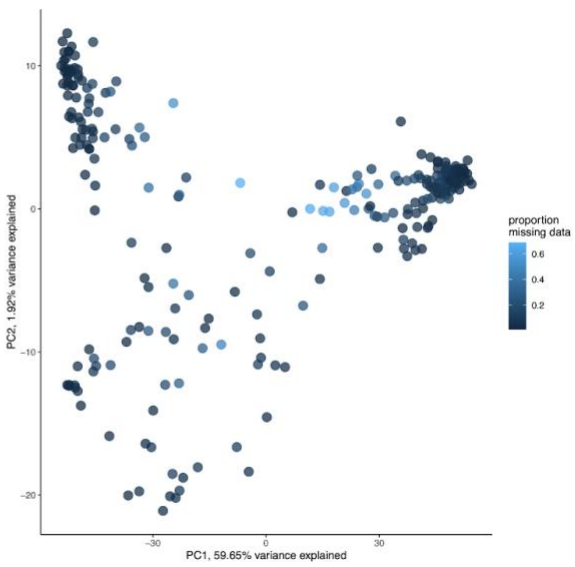


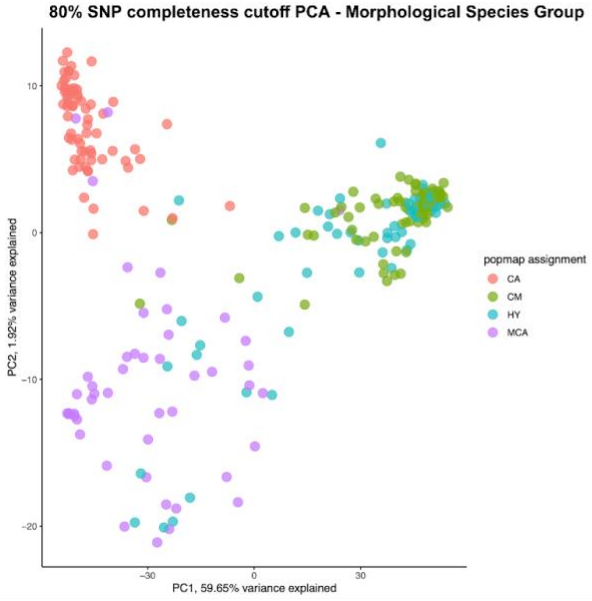
Supplemental Figure S3. Principal component analysis (PCA) plots visualizing whether missing data drove clustering patterns across SNP cutoff thresholds of (a) 75%, (b) 80%, (c) 85%, and (d) 90% across both morphological species groups (left: popmap assignment) and sampling locality (right: popmap assignment).

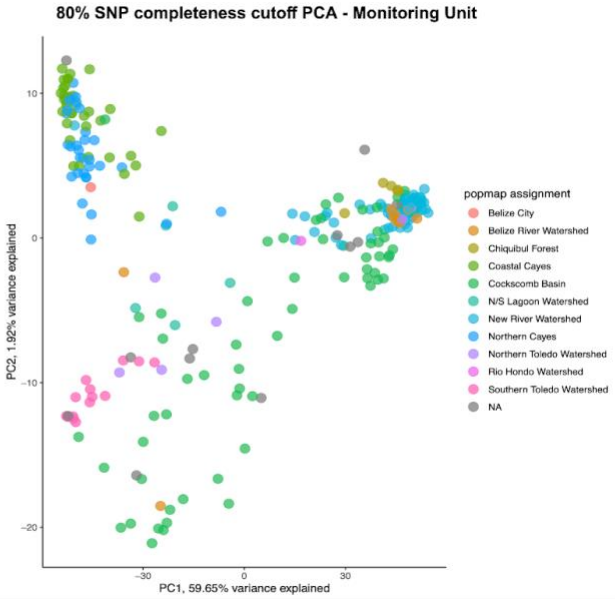


(b)

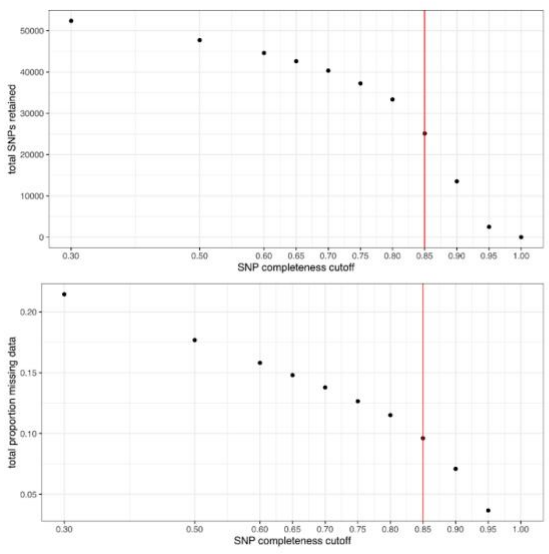

total SNPs retained
SNP completeness cutoff
total proportion missing data
SNP completeness cutoff

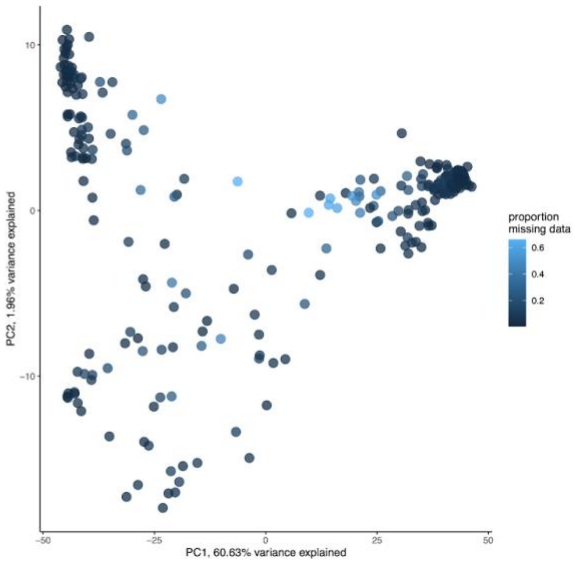

PC2, 1.92% variance explained
PC1, 59.65% variance explained
proportion missing data

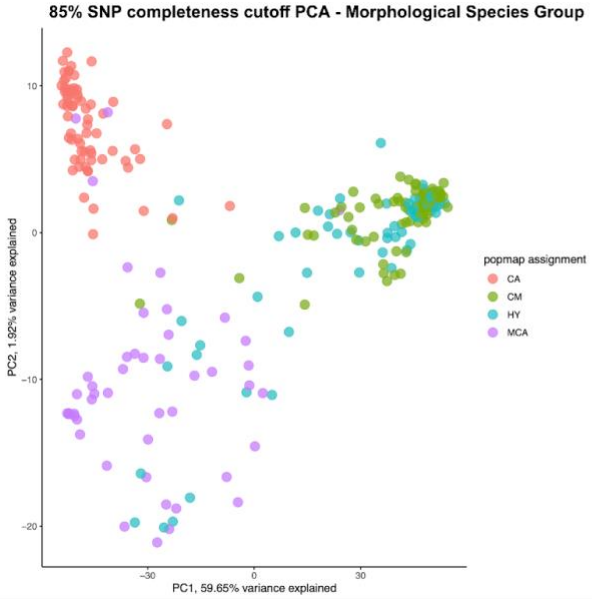

80% SNP completeness cutoff PCA - Morphological Species Group
PC2, 1.92% variance explained
PC1, 59.65% variance explained
popmap assignment
CA
CM
HY
MCA

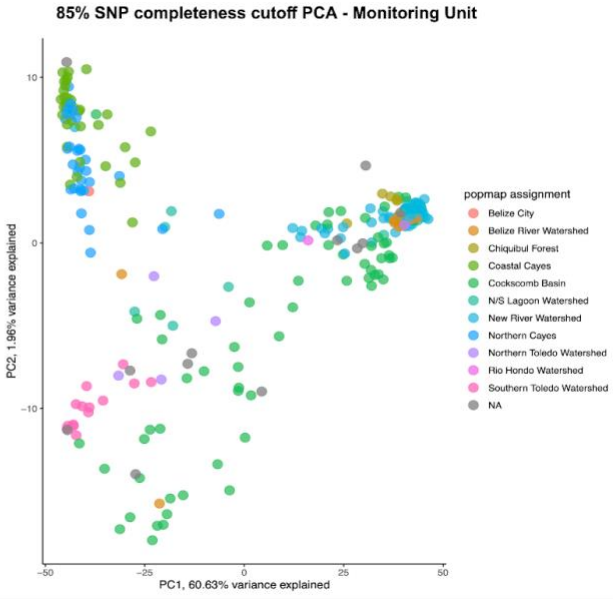

80% SNP completeness cutoff PCA - Monitoring Unit
PC2, 1.92% variance explained
PC1, 59.65% variance explained
popmap assignment
Belize City
Belize River Watershed
Chiquibul Forest
Coastal Cayes
Cockscomb Basin
N/S Lagoon Watershed
New River Watershed
Northern Cayes
Northern Toledo Watershed
Rio Hondo Watershed
Southern Toledo Watershed
NA

(c)

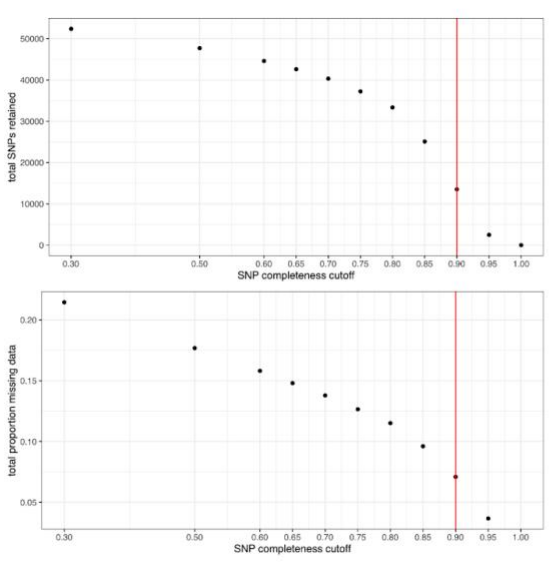

total SNPs retained
SNP completeness cutoff
total proportion missing data
SNP completeness cutoff

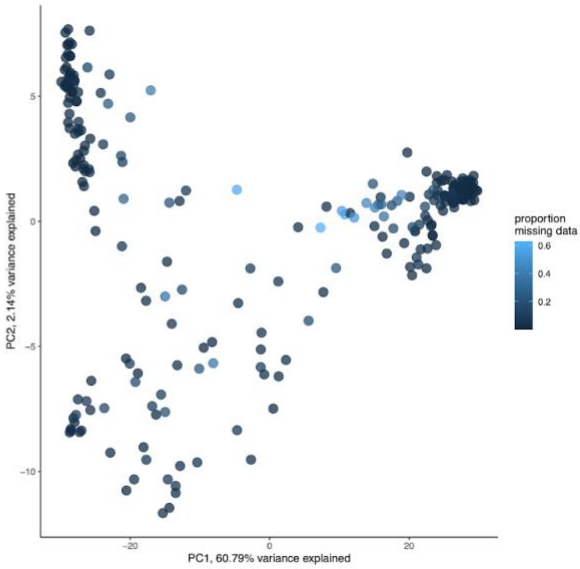

PC2, 1.96% variance explained
PC1, 60.63% variance explained
proportion missing data

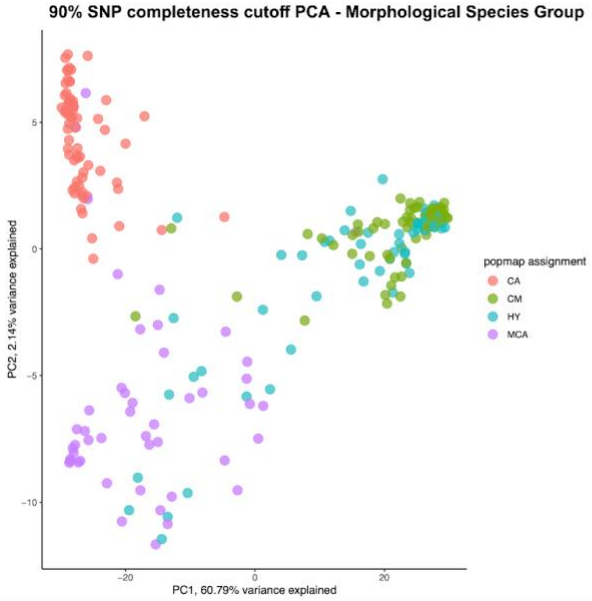

85% SNP completeness cutoff PCA - Morphological Species Group
PC2, 1.92% variance explained
PC1, 59.65% variance explained
popmap assignment
CA
CM
HY
MCA

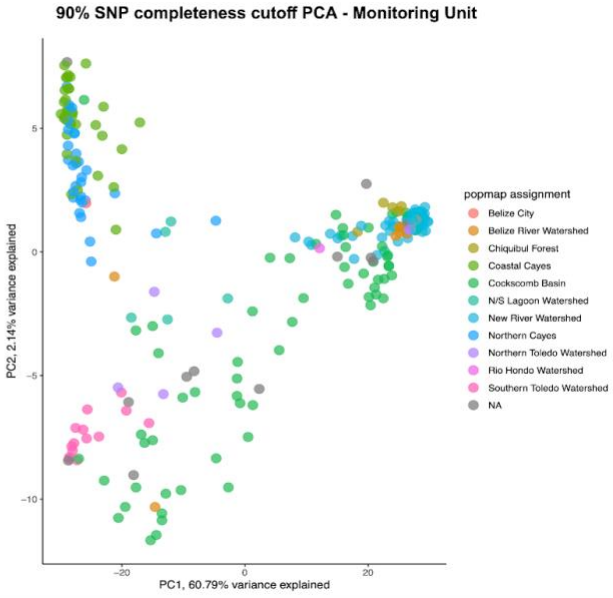

85% SNP completeness cutoff PCA - Monitoring Unit
PC2, 1.96% variance explained
PC1, 60.63% variance explained
popmap assignment
Belize City
Belize River Watershed
Chiquibul Forest
Coastal Cayes
Cockscomb Basin
N/S Lagoon Watershed
New River Watershed
Northern Cayes
Northern Toledo Watershed
Rio Hondo Watershed
Southern Toledo Watershed
NA

(d)

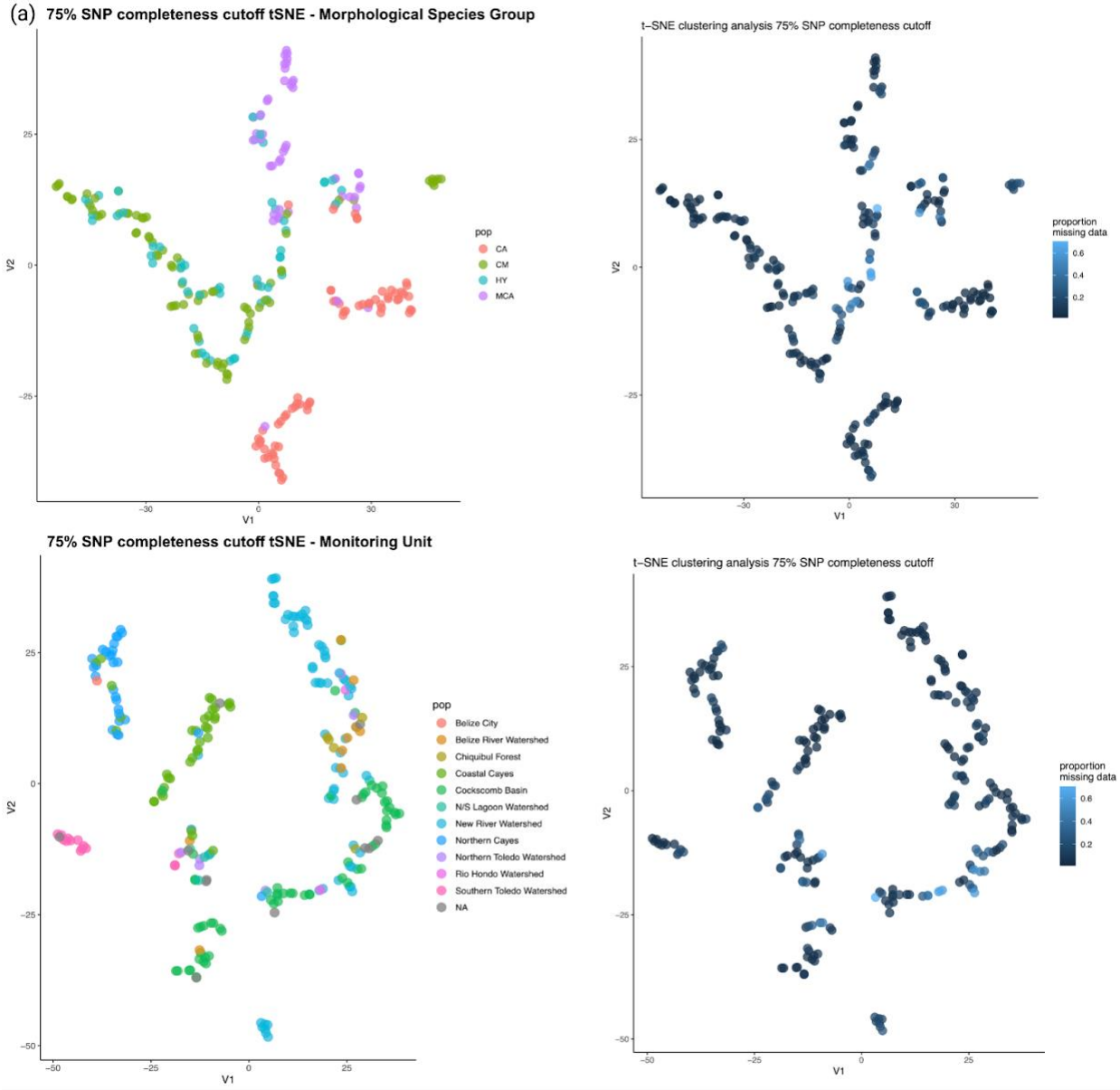

total SNPs retained
SNP completeness cutoff
total proportion missing data
SNP completeness cutoff

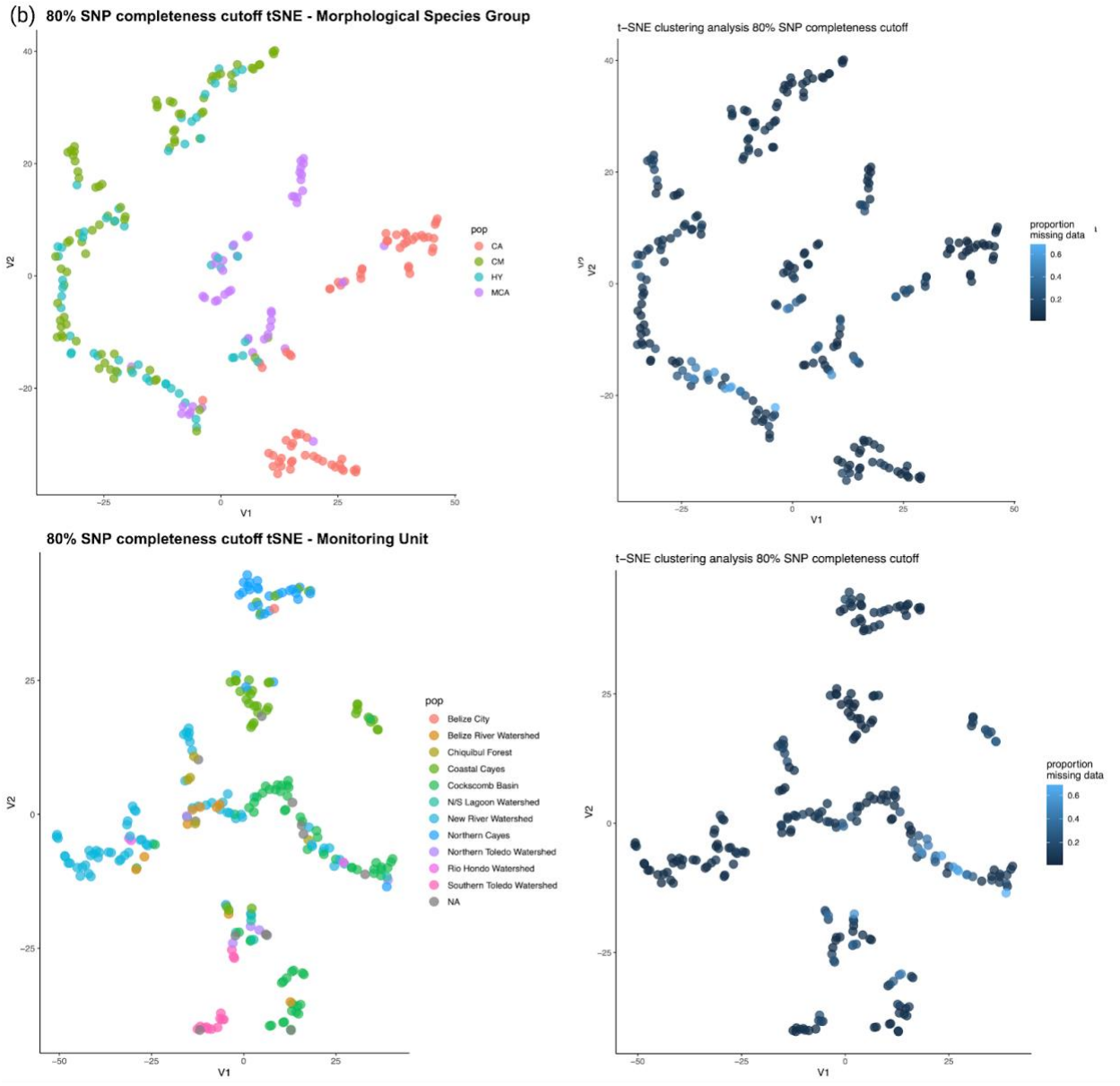

PC2, 2.14% variance explained
PC1, 60.79% variance explained
proportion missing data


90% SNP completeness cutoff PCA - Morphological Species Group
PC2, 2.14% variance explained
PC1, 60.79% variance explained
popmap assignment
CA
CM
HY
MCA

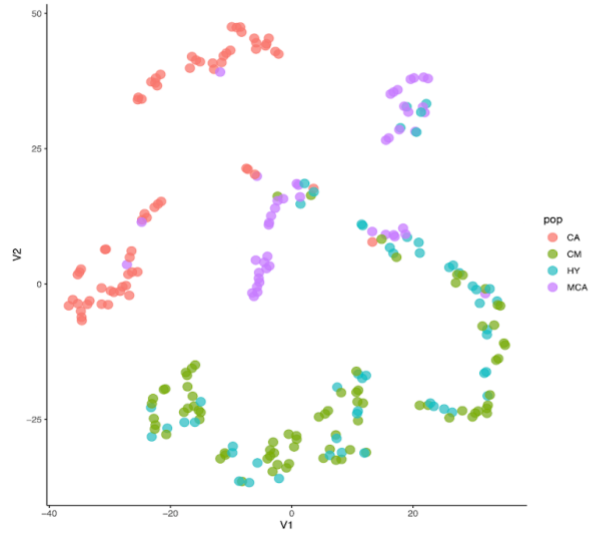

90% SNP completeness cutoff PCA - Monitoring Unit
PC2, 2.14% variance explained
PC1, 60.79% variance explained
popmap assignment
Belize City
Belize River Watershed
Chiquibul Forest
Coastal Cayes
Cockscomb Basin
N/S Lagoon Watershed
New River Watershed
Northern Cayes
Northern Toledo Watershed
Rio Hondo Watershed
Southern Toledo Watershed
NA

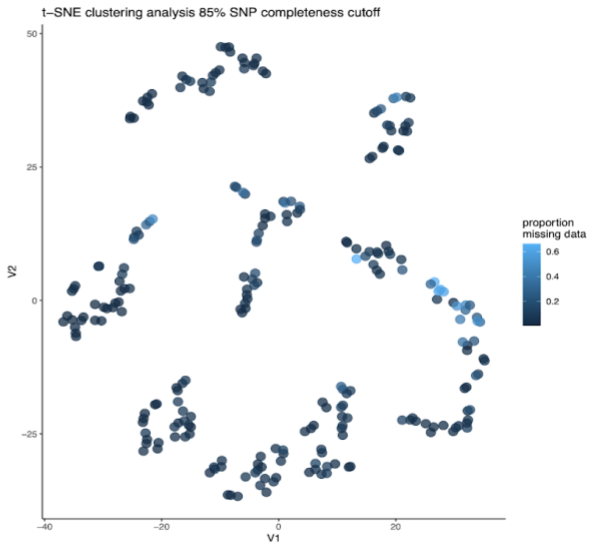
Supplemental Figure S4. T-distributed stochastic neighbor embedding (tSNE) plots visualizing whether missing data drove clustering patterns across SNP cutoff thresholds of (a) 75%, (b) 80%, (c) 85%, and (d) 90% across both morphological species groups (top: pop) and sampling locality (bottom: pop).



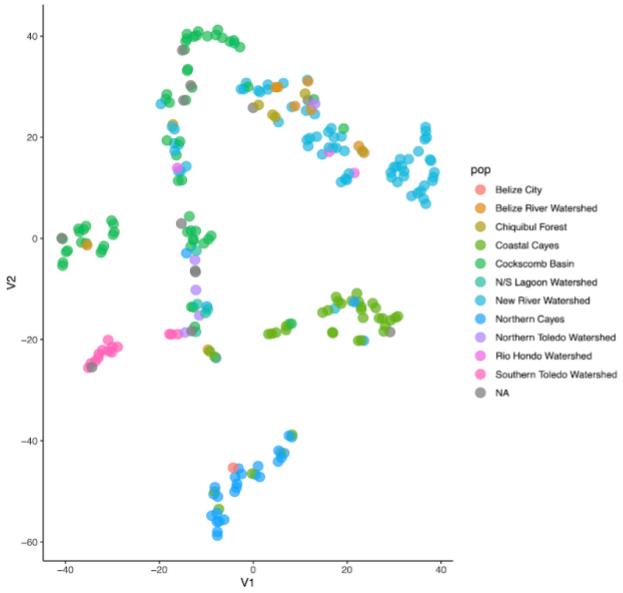

(b) 80% SNP completeness cutoff tSNE - Morphological Species Group
t–SNE clustering analysis 80% SNP completeness cutoff
pop
CA
CM
HY
MCA
proportion missing data
0.6
0.4
0.2
V1
V2
80% SNP completeness cutoff tSNE - Monitoring Unit
t–SNE clustering analysis 80% SNP completeness cutoff
pop
Belize City
Belize River Watershed
Chiquibul Forest
Coastal Cayes
Cockscomb Basin
N/S Lagoon Watershed
New River Watershed
Northern Cayes
Northern Toledo Watershed
Rio Hondo Watershed
Southern Toledo Watershed
NA
proportion missing data
0.6
0.4
0.2
V1
V2

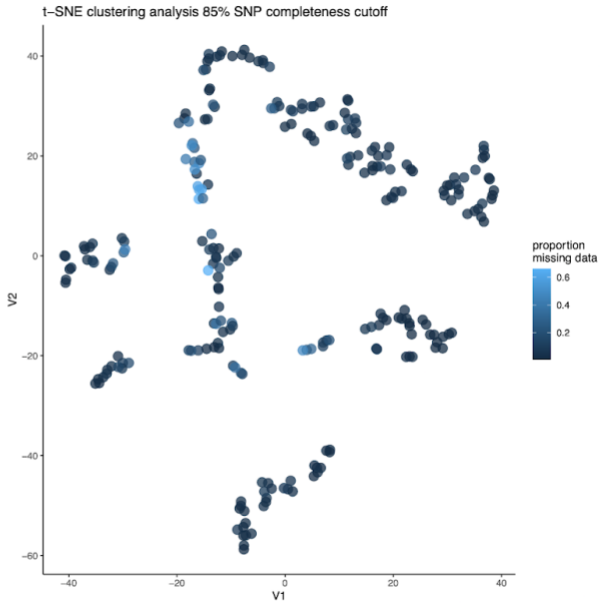
(c) **85% SNP completeness cutoff tSNE - Morphological Species Group**



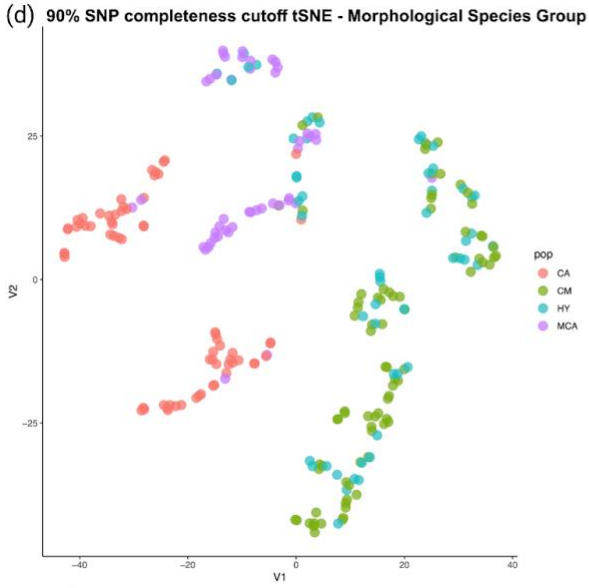


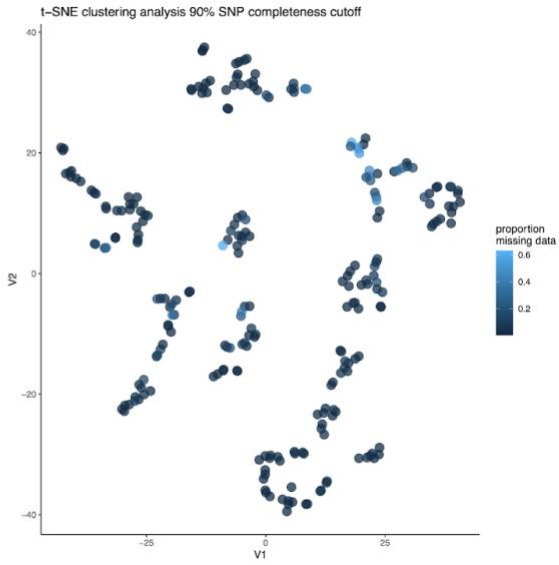
85% SNP completeness cutoff tSNE - Monitoring Unit



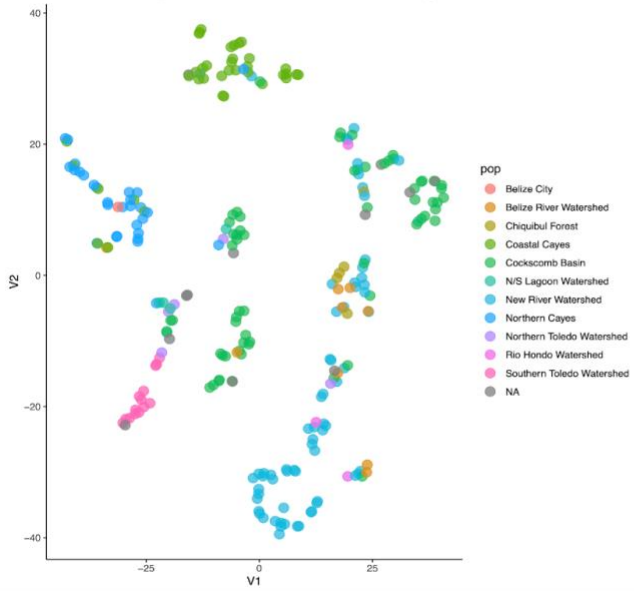


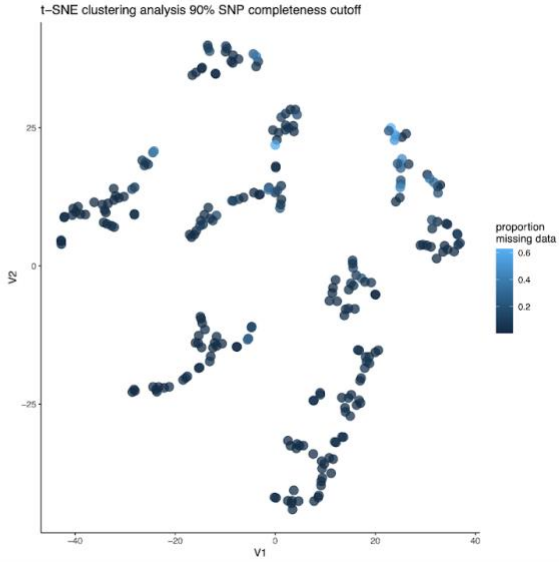
(d) **90% SNP completeness cutoff tSNE - Morphological Species Group**



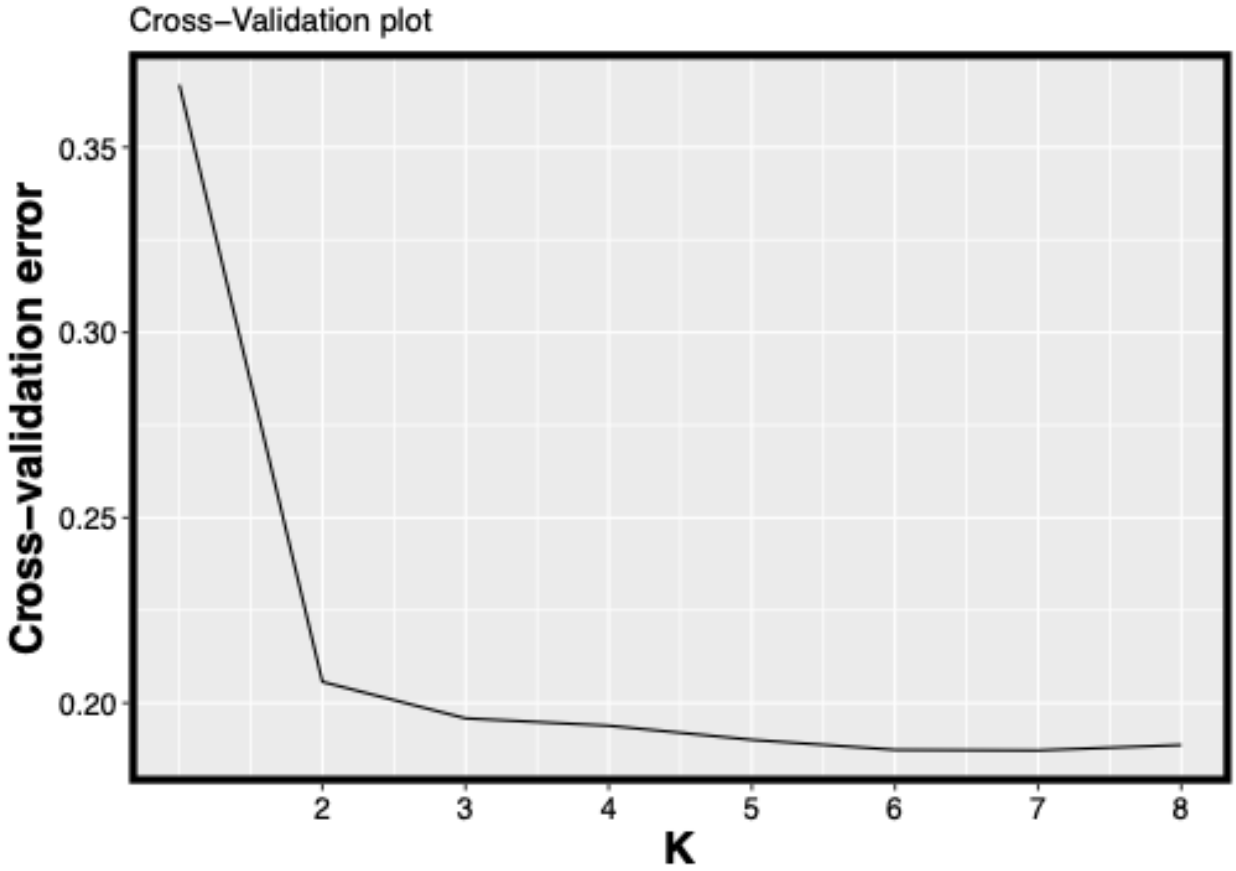


90% SNP completeness cutoff tSNE - Monitoring Unit

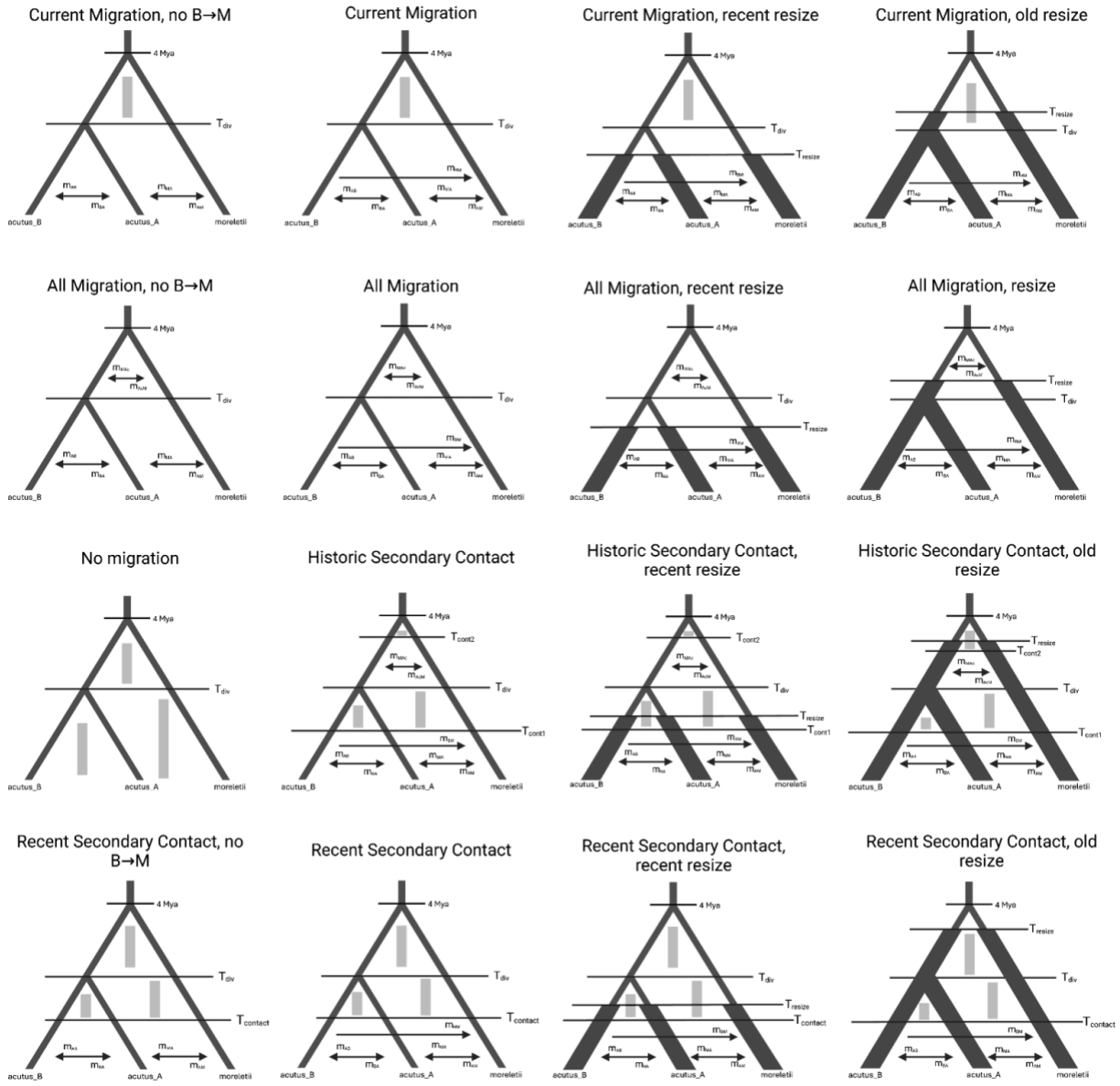




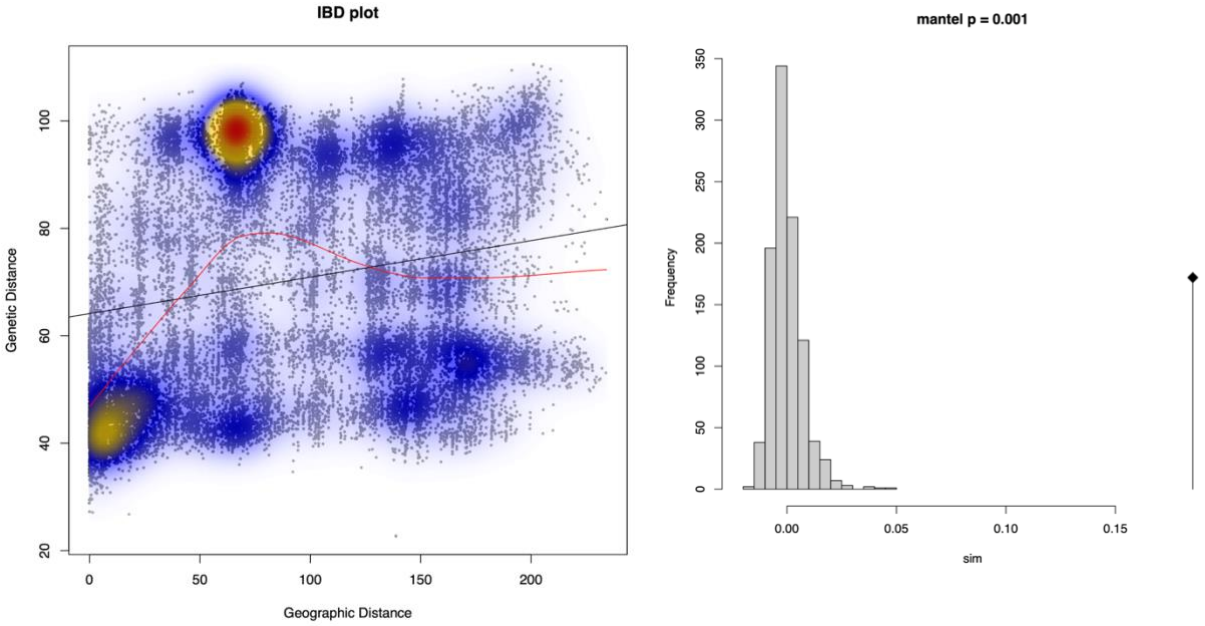
Supplementary Figure S5: Cross-validation error plot obtained from ADMIXTURE run with cluster values (K) evaluated from K=1 to K=8.



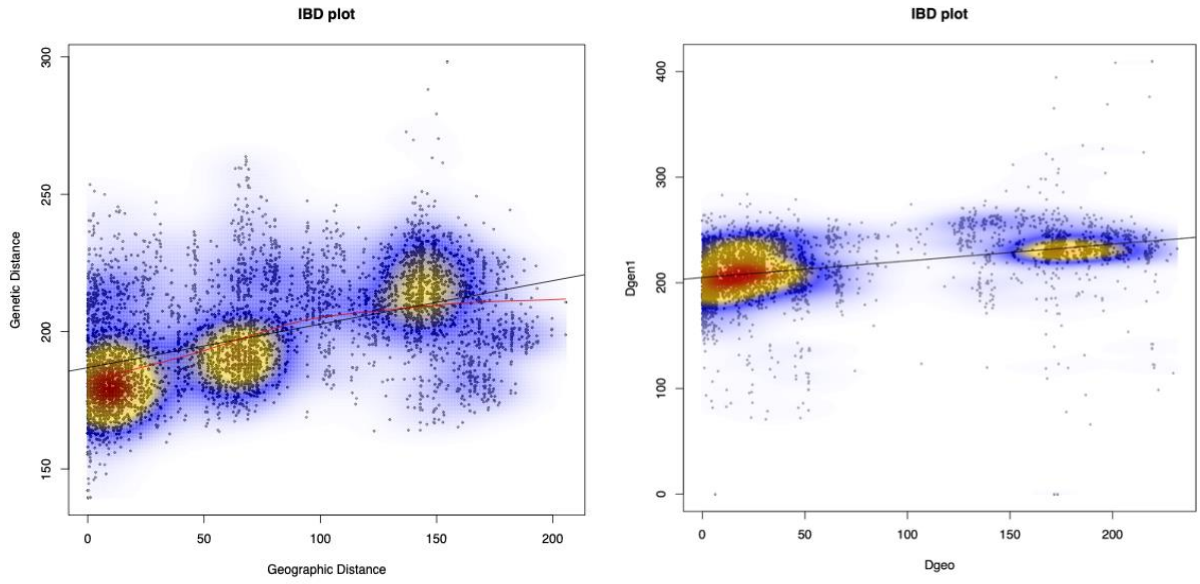
Supplemental Figure S6. Schematic representation of 16 demographic models tested in FastSimCoal2 analyses adopted from Harrington and Burbrink (2023). Abbreviations of parameters in the schematic are m = migration and T = time. Subscripts of m: A = *C. acutus* lineage A (acutus_A), B = *C. acutus* lineage B (acutus_B), M = *C. moreletii,* Ac = ancestral *C. acutus.* Vertical grey bars indicate periods of time with no migration among lineages. In models 'Current Migration, recent resize' and 'All Migration, recent resize' the migration rates are constant from the time of divergence (Tdiv) to the present.



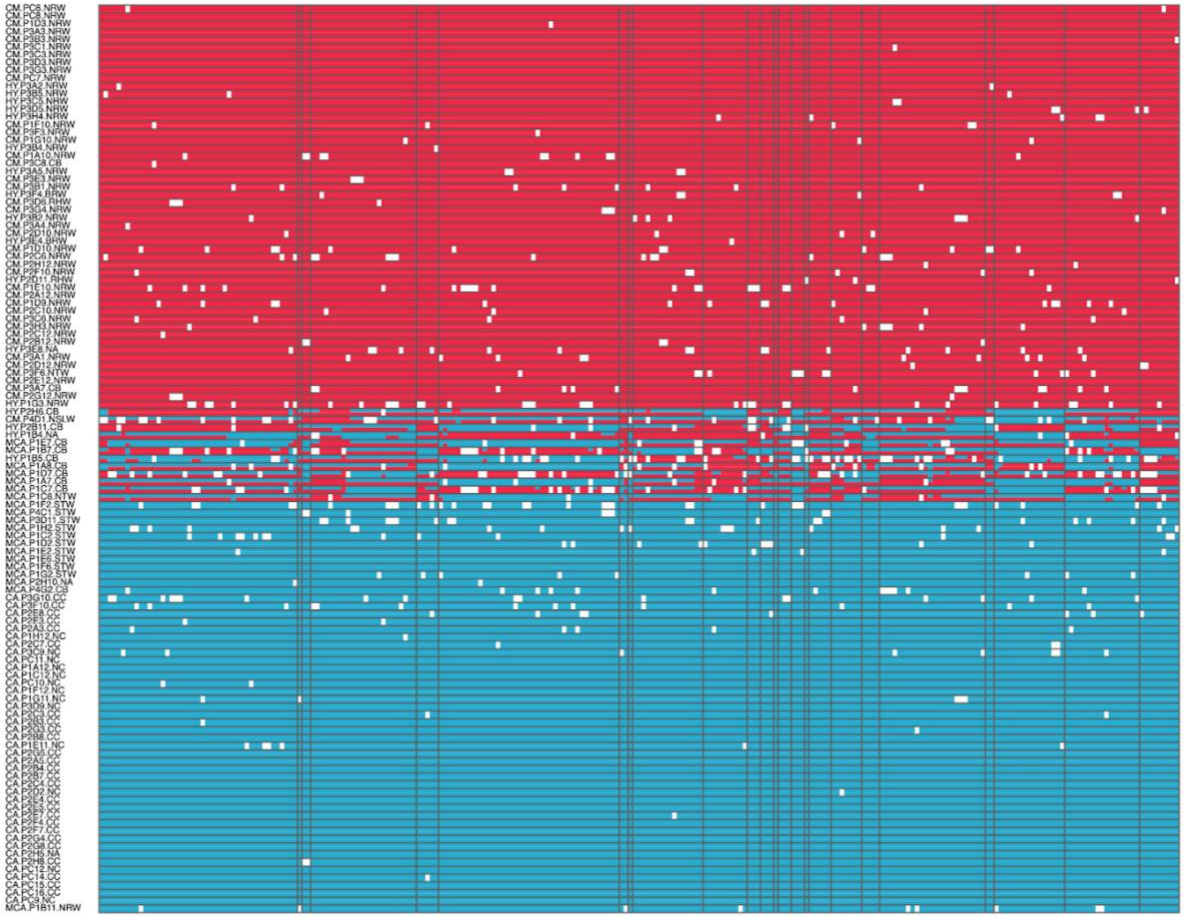
Supplemental Figure S7. Kernel density plot of genetic versus geographical distance for the full sample dataset. Hotter colors correspond to a higher density of points. Geographic distances are in kilometers and genetic distances are uncorrected pairwise distances. We performed a mantel test correlating the genetic and geographic data into pairwise distance matrices and compared results to a permuted null distribution and obtained a highly significant p-value of 0.001. IBD is present in our data yet discrete structure remains present across both species and in *C. acutus* lineages.



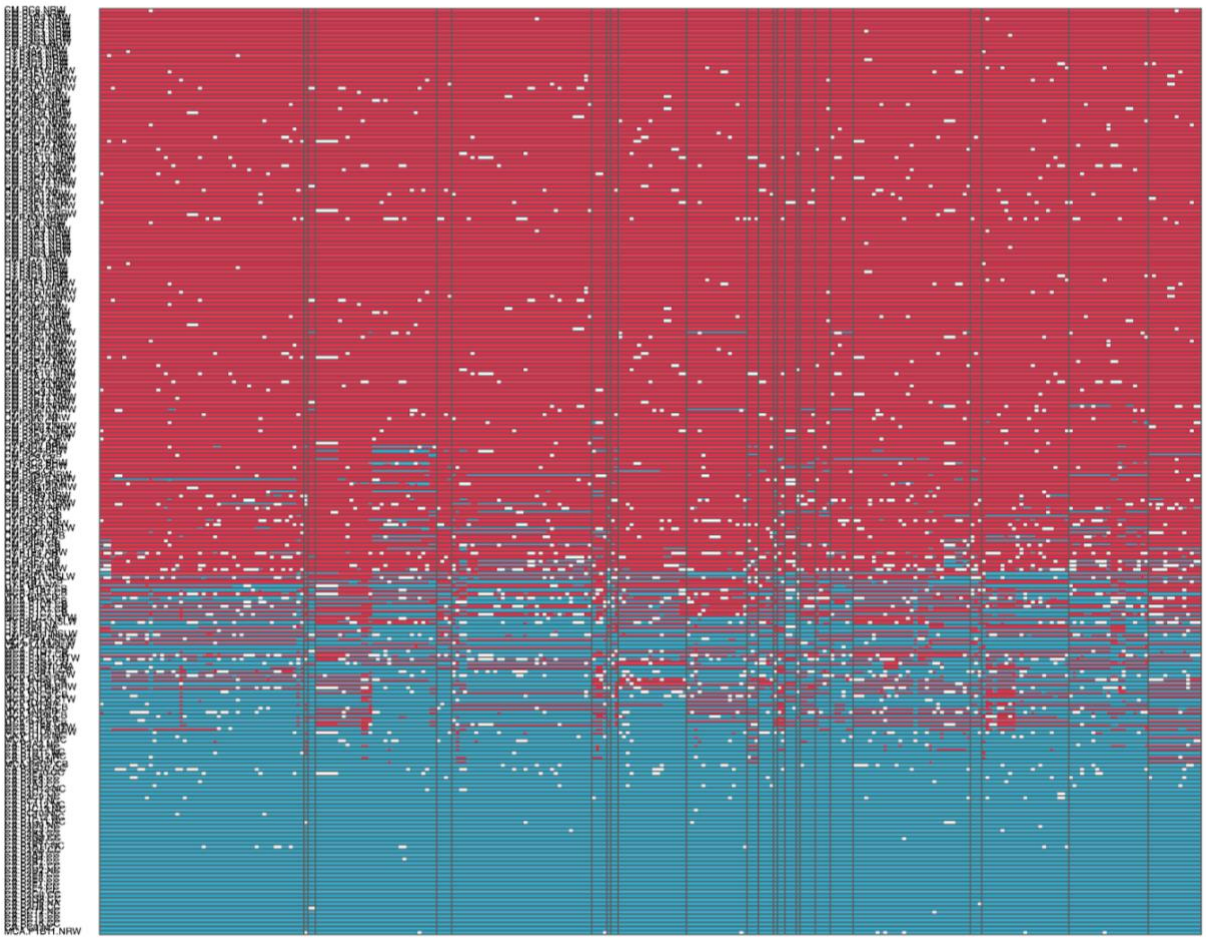
Supplemental Figure S8. Kernel density plot of genetic versus geographical distance for *C. moreletii* (left) and *C. acutus* (right). Hotter colors correspond to a higher density of points. Geographic distances are in kilometers and genetic distances are uncorrected pairwise distances. We performed a mantel test correlating the genetic and geographic data into pairwise distance matrices and compared results to a permuted null distribution and obtained a highly significant p-value of 0.001 for both species.



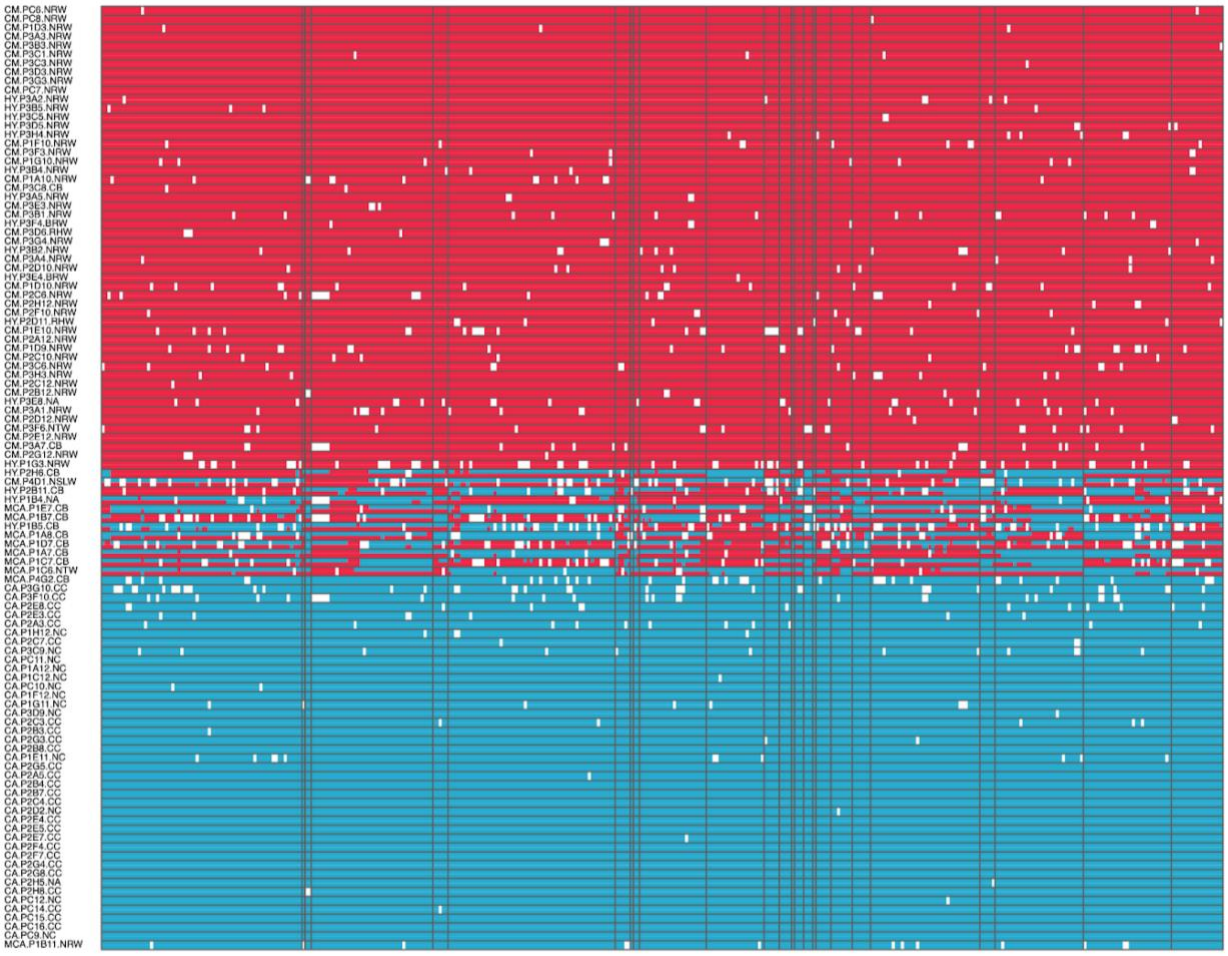
Supplemental Figure S9. Ancestry painting for 52 unadmixed *C. moreletii* (red), 12 Hybrids (40:60 Hybrids), 53 unadmixed *C. acutus* (blue). Dataset included 9899 SNPs and 275 sites with complete fixation of different alleles in the parental population. Unadmixed parental species are defined as having a Q-score > 98% determined by our ADMIXTURE results, and *C. acutus* combined individuals from both lineages. The top and bottom horizontal bars represent 275 sites that are fixed for different alleles between the two species; all other bars indicate the alleles at each of those sites. White blocks indicate missing data. Heterozygous alleles are shown with the top half in each bar matching the second parental species and vice versa.



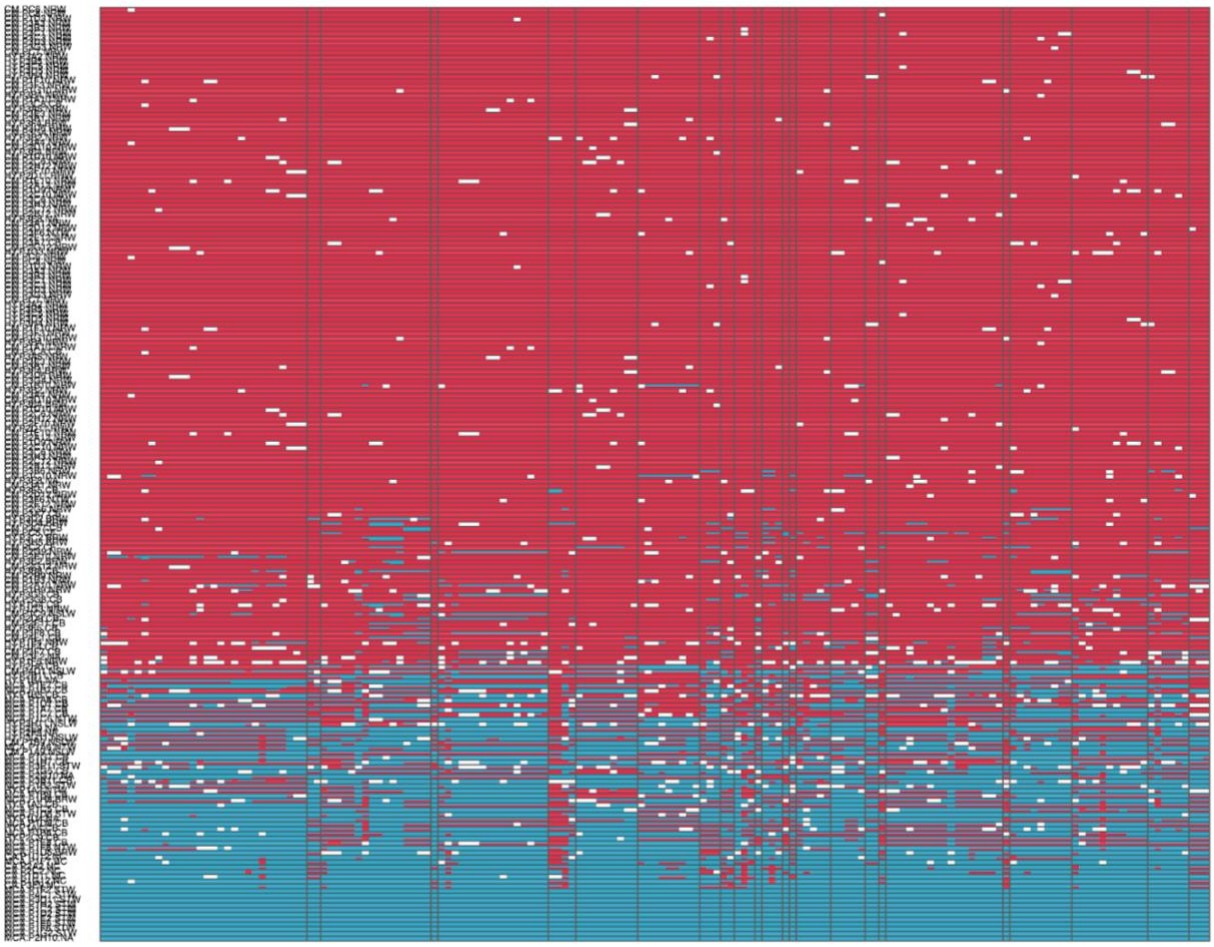
Supplemental Figure S10. Ancestry painting for 52 unadmixed *C. moreletii* (red), 133 Hybrids, 42 unadmixed Cayes lineage *C. acutus* (blue). Dataset included 12383 SNPs and 402 sites with complete fixation of different alleles in the parental population. Unadmixed parental species are defined as having a Q-score > 98% determined by our ADMIXTURE results. The top and bottom horizontal bars represent 275 sites that are fixed for different alleles between the two species; all other bars indicate the alleles at each of those sites. White blocks indicate missing data. Heterozygous alleles are shown with the top half in each bar matching the second parental species and vice versa.



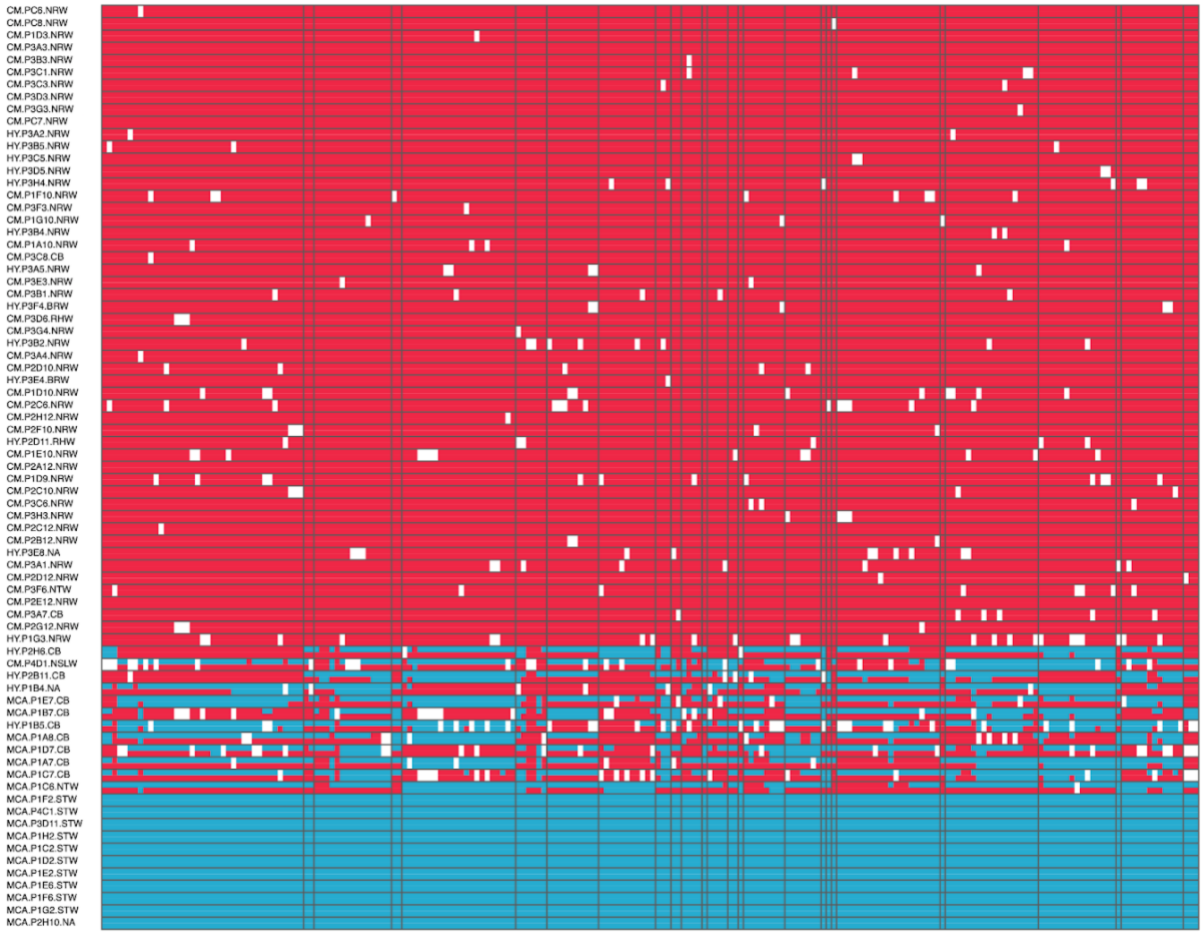
Supplemental Figure S11. Ancestry painting for 52 unadmixed *C. moreletii* (red), 12 Hybrids (40:60 Hybrids), 42 unadmixed Cayes lineage *C. acutus* (blue). Dataset included 12383 SNPs and 402 sites with complete fixation of different alleles in the parental population. Unadmixed parental species are defined as having a Q-score > 98% determined by our ADMIXTURE results. The top and bottom horizontal bars represent 275 sites that are fixed for different alleles between the two species; all other bars indicate the alleles at each of those sites. White blocks indicate missing data. Heterozygous alleles are shown with the top half in each bar matching the second parental species and vice versa.



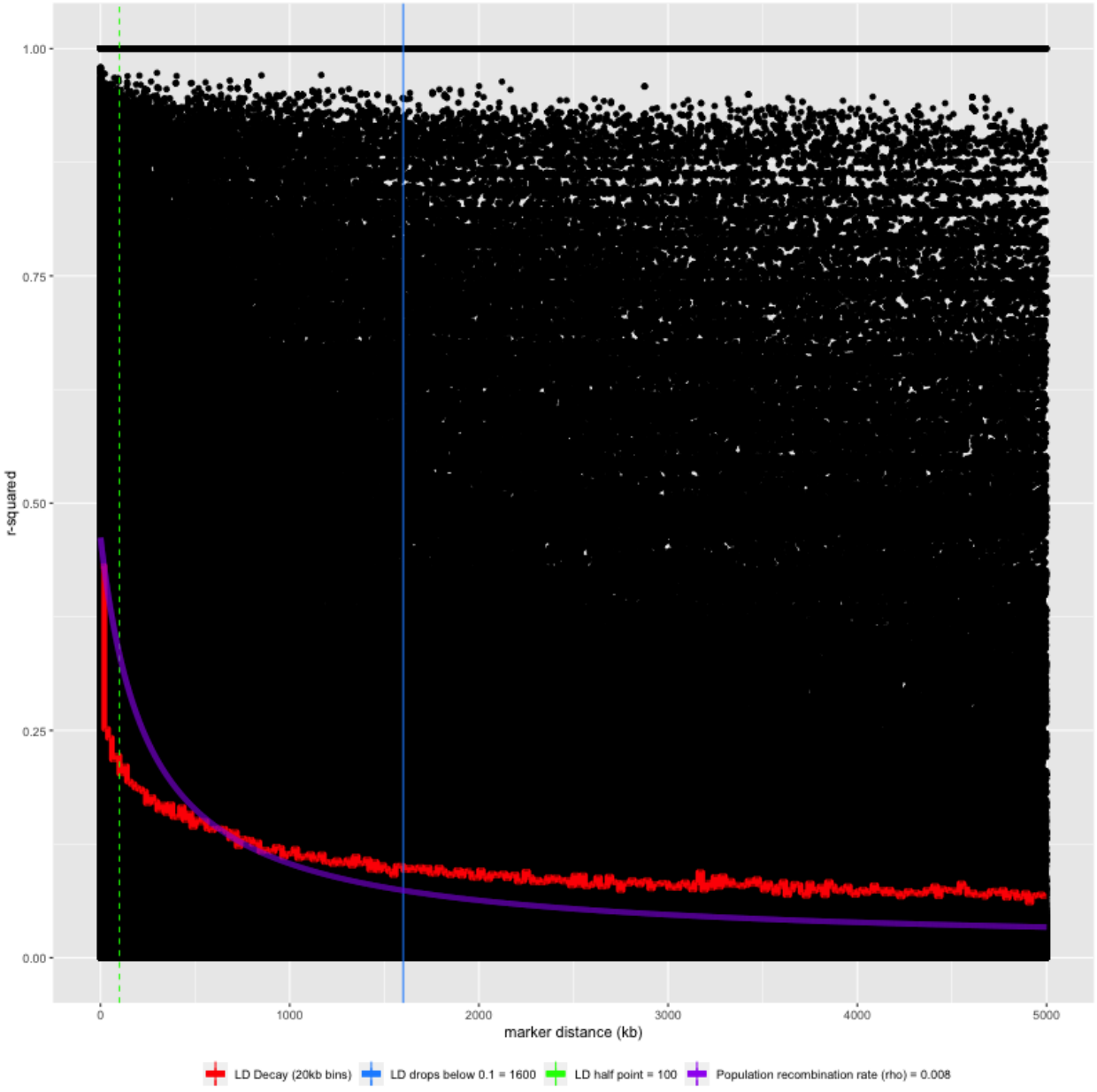
Supplemental Figure S12. Ancestry painting for 52 unadmixed *C. moreletii* (red), 133 Hybrids, 11 unadmixed Mainland lineage *C. acutus* (blue). Dataset included 6865 SNPs and 255 sites with complete fixation of different alleles in the parental population. Unadmixed parental species are defined as having a Q-score > 98% determined by our ADMIXTURE results. The top and bottom horizontal bars represent 275 sites that are fixed for different alleles between the two species; all other bars indicate the alleles at each of those sites. White blocks indicate missing data. Heterozygous alleles are shown with the top half in each bar matching the second parental species and vice versa.



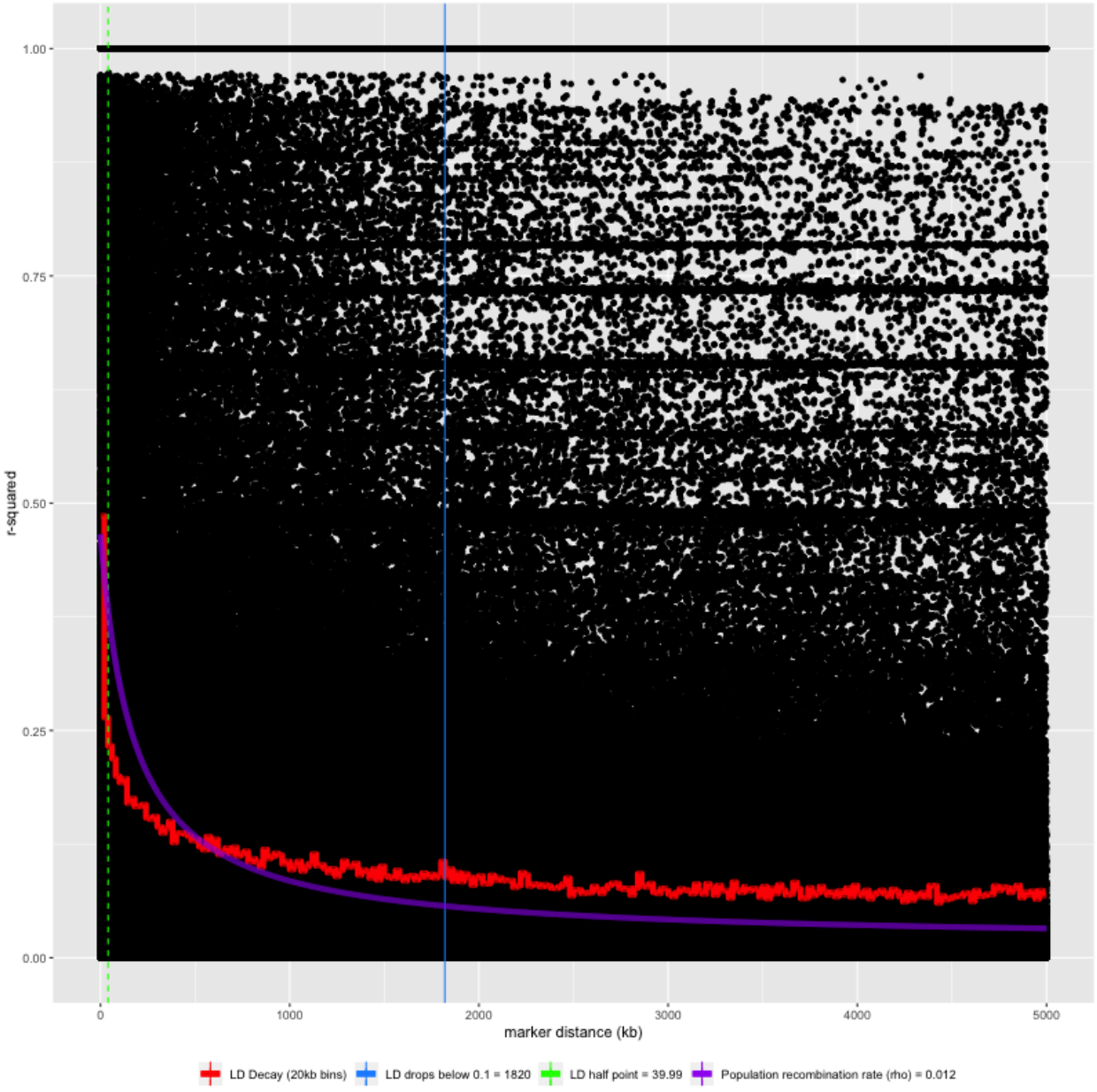
Supplemental Figure S13. Ancestry painting for 52 unadmixed *C. moreletii* (red), 12 Hybrids (40:60 Hybrids), 11 unadmixed Mainland lineage *C. acutus* (blue). Dataset included 6865 SNPs and 255 sites with complete fixation of different alleles in the parental population. Unadmixed parental species are defined as having a Q-score > 98% determined by our ADMIXTURE results. The top and bottom horizontal bars represent 275 sites that are fixed for different alleles between the two species; all other bars indicate the alleles at each of those sites. White blocks indicate missing data. Heterozygous alleles are shown with the top half in each bar matching the second parental species and vice versa.



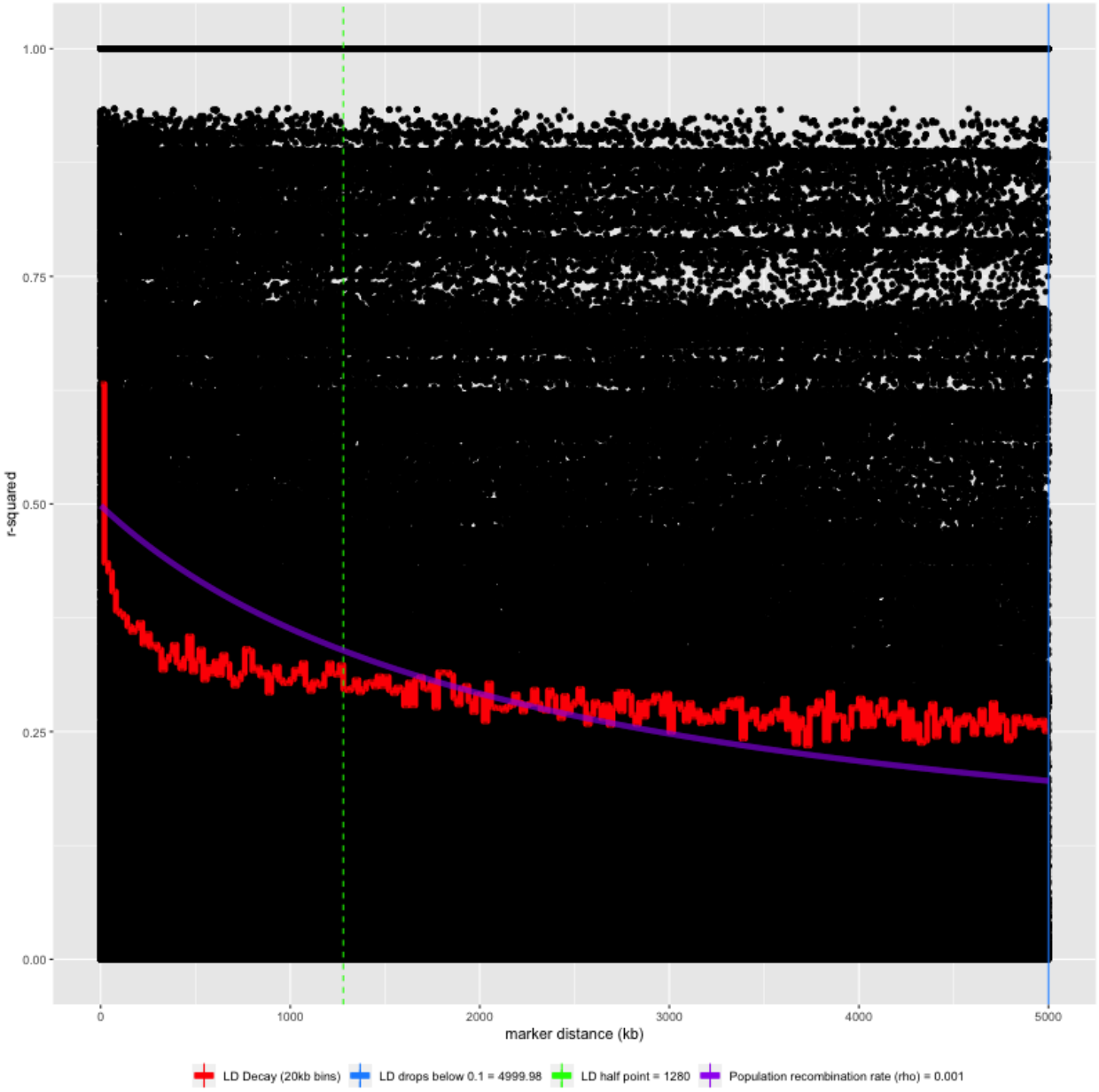
Supplemental Figure S14: LD decay plot calculated for pure *C. moreletii* using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



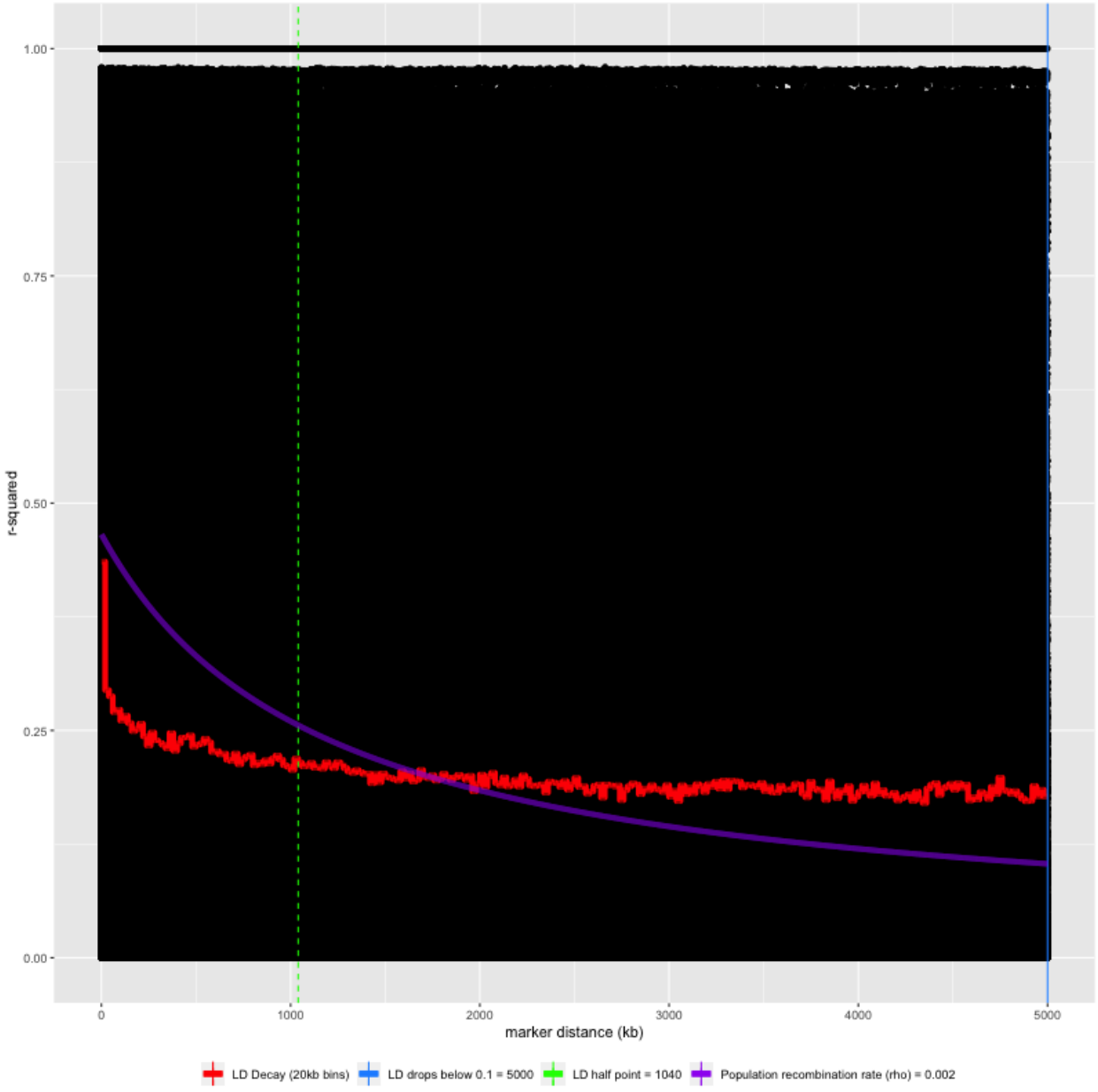
Supplemental Figure S15: LD decay plot for pure Cayes *C. acutus* lineage using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



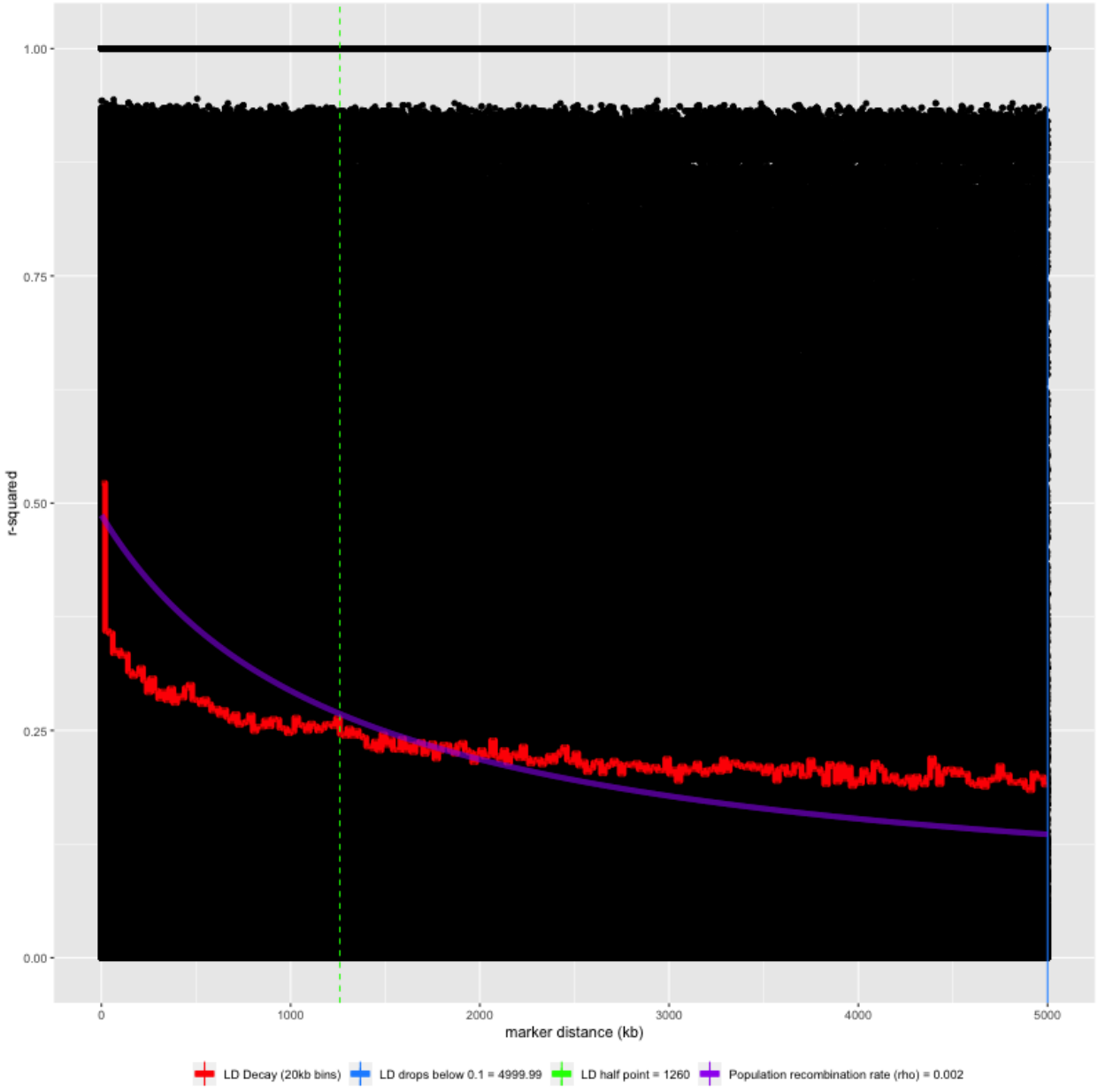
Supplemental Figure S16: LD decay plot for pure Mainland *C. acutus* lineage using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



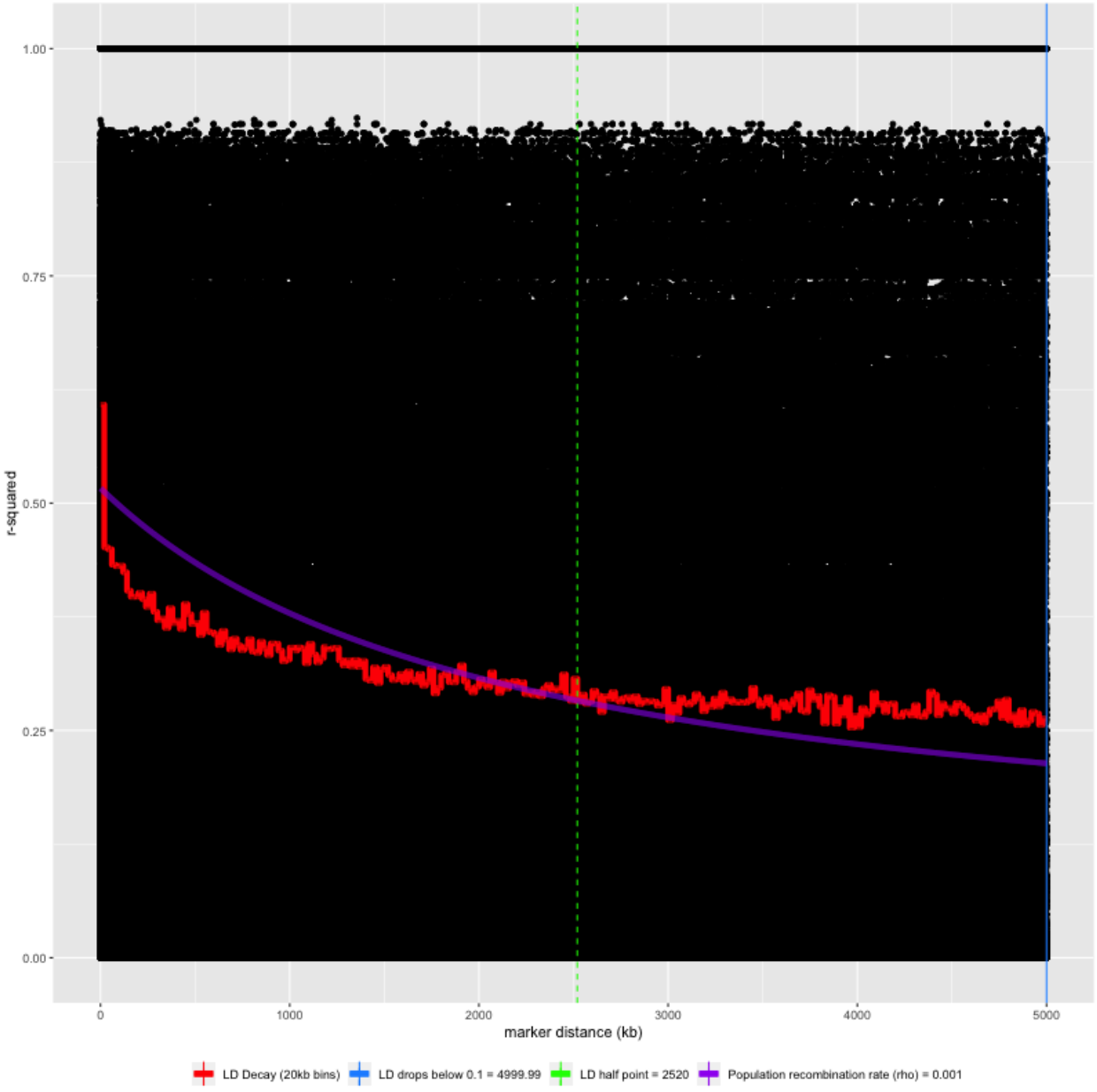
Supplemental Figure S17: LD decay plot for all hybrids using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



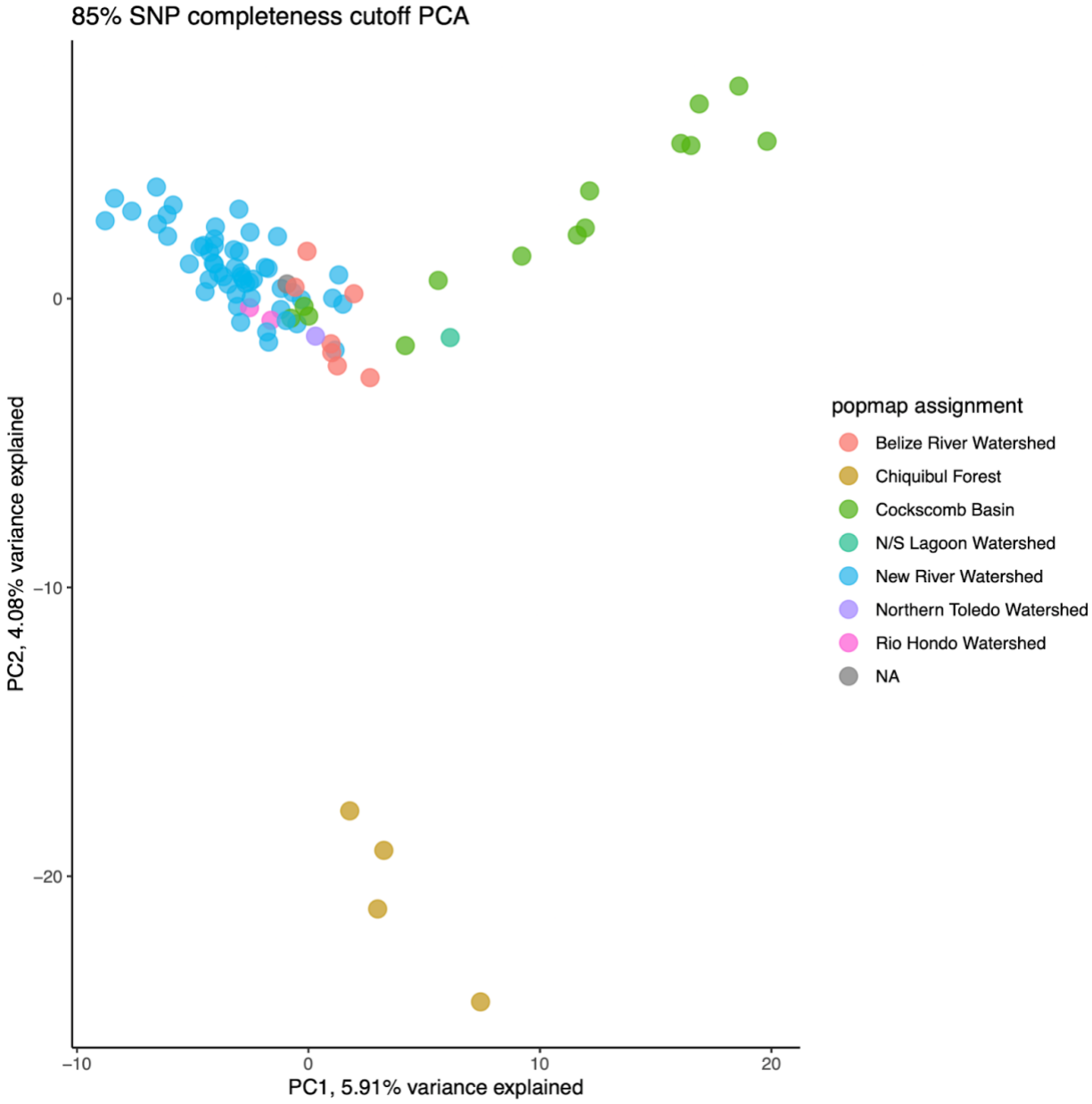
Supplemental Figure S18: LD decay plot for 25%-75% Hybrids using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



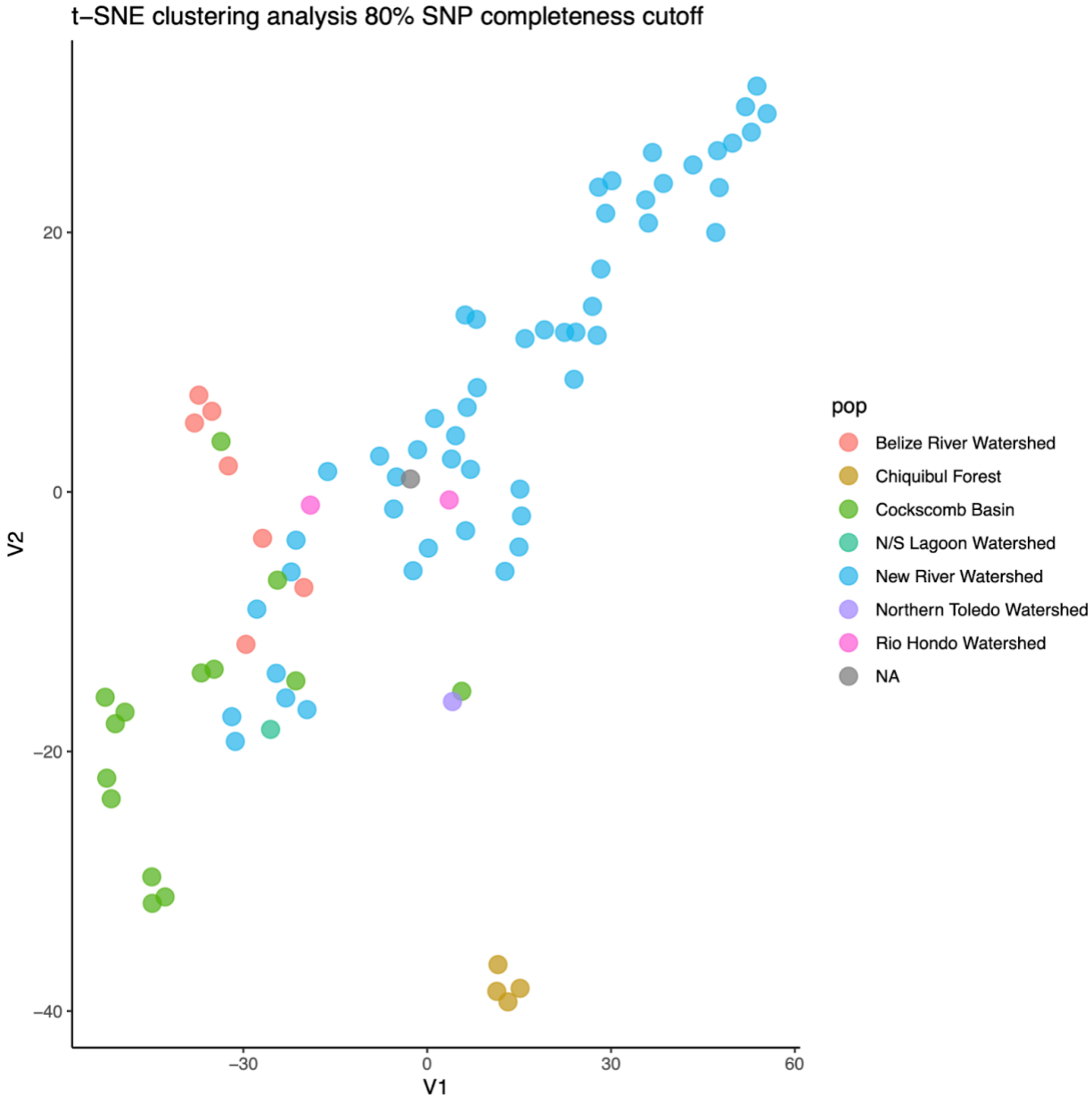
Supplemental Figure S19: LD decay plot for 40%-60% Hybrids. using mean r^2 for all pairwise SNPs along distance intervals of 20 kb up to a maximum distance of 5 Mb (red). Blue line represents the point where LD drops below 0.1, green line indicates points where LD drops by half, and purple line is the estimate for population recombination rate (rho) for the expected R-squared.



Supplemental Figure S20. Principal component analysis (PCA) plot visualizing clustering patterns in pure *C. moreletii* (Q>90%) by Sampling locality (popmap assignment) based on the 85% filtering threshold dataset.



Supplemental Figure S21. T-distributed stochastic neighbor embedding (tSNE) plots visualizing clustering patterns in pure *C. moreletii* (Q>90%) by Sampling locality (pop) based on the 85% filtering threshold dataset.



SI References

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., & Troendle, N. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, *7*, e7724.

Boucher, M., Tellez, M., & Anderson, J. T. (2020). Differences in distress: Variance and production of American Crocodile (*Crocodylus acutus*) distress calls in Belize. *Ecology and Evolution*, *10*(18), 9624–9634. https://doi.org/10.1002/ece3.6556

Harrington, S., & Burbrink, F. T. (2023). Complex cycles of divergence and migration shape lineage structure in the common kingsnake species complex. *Journal of Biogeography*, *50*(2), 341–351. https://doi.org/10.1111/jbi.14536

Muniz, F. de L., Bittencourt, P. S., Hernández-Rangel, S. M., Roberto, I. J., Farias, I. P., & Hrbek, T. (2021). Biogeography and Comparative Phylogeography of New-World Crocodylians. In R. B. Zucoloto, P. S. Amavet, L. M. Verdade, & I. P. Farias (Eds.), *Conservation Genetics of New World Crocodilians* (pp. 95–122). Springer International Publishing. https://doi.org/10.1007/978-3-030-56383-7_4

St John, J. A., Braun, E. L., Isberg, S. R., Miles, L. G., Chong, A. Y., Gongora, J., Dalzell, P., Moran, C., Bed'Hom, B., & Abzhanov, A. (2012). Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biology*, *13*(1), 1–12.