

**Table S1: Genes within the association peaks.**

Comparison	Contig	Chromosome	Number of annotations	Genes
SASC vs. Makira	400	11	15	TANGO6, GAS8, DBNDD1, DEF8, TUBB3, MC1R, TCF25, SPIRE2, FANCA, ZNF276, VPS9D1, SPATA2L, CDK10, LOC100232461, LOC115496849
Ugi vs. Makira	1042	Z	1	LOC115491070
Ugi vs. Makira	224	Z	15	SLC44A1, SLC27A6, ISOC1, ADAMTS19, MINAR2, CHSY3, HINT1, LYRM7, CDC42SE2, SPTLC1, ROR2, NFIL3, DIRAS2, GADD45G, LOC116806816
Ugi vs. Makira	5	1	23	EFHC2, FUNDC1, DIPK2B, MIR221, ICOSLG, GATD3A, PWP2, TRAPPC10, AGPAT3, PDXK, RRP1B, HSF2BP, SIK1, CRYAA, U2AF1, LOC116808847, LOC115496971, LOC115496975, LOC115496977, LOC115497347, LOC115496981, LOC115496994, LOC115497018
Ugi vs. Makira	866	20	10	TP53INP2, NCOA6, GGT7, RAB5IF, MYL9, TGIF2, DLGAP4, EPB41L1, SRSF6, L3MBTL1
Ugi vs. Makira	947	20	14	RALY, EIF2S2, ASIP, AHCY, ITCH, DYNLRB1, FSIP2, CEP250, GDF5, FAM83C, EIF6, MMP24, LOC116806597, LOC116809171
Ugi vs. Makira	62	Z	19	PLPP1, MSMP, GALT, SIGMAR1, DCTN3, ENHO, FAM219A, MYORG, NUDT2, UBAP1, DCAF12, NOL6, AQP3, KIAA1328, MAPK6, LOC115491277, LOC105760850, LOC115491042, LOC100226213
Ugi vs. Makira	318	6	1	LOC116808540

**Table S2: Details for the samples used in this study.**

<b>Subspecies</b>	<b>Island</b>	<b>Collection date</b>	<b>Locality</b>	<b>Latitude</b>	<b>Longitude</b>	<b>Ventral Coloration</b>	<b>Sample ID</b>	<b>Sex*</b>
<i>megarhynchus</i>	Makira	5/16/09	Kira Kira	-10.5	161.9	Chestnut	MA132	M
<i>megarhynchus</i>	Makira	5/16/09	Kira Kira	-10.5	161.9	Chestnut	MA133	M
<i>megarhynchus</i>	Makira	5/22/09	Kira Kira	-10.5	161.9	Intermediate	MA166	M
<i>megarhynchus</i>	Makira	6/21/09	Star Harbour	-10.8	162.2	Chestnut	MA175	M
<i>megarhynchus</i>	Makira	6/22/09	Star Harbour	-10.8	162.2	Chestnut	MA180	M
<i>megarhynchus</i>	Makira	6/22/09	Star Harbour	-10.8	162.2	Intermediate	MA182	F
<i>megarhynchus</i>	Makira	6/23/09	Star Harbour	-10.8	162.2	Chestnut	MA183	M
<i>megarhynchus</i>	Makira	6/23/09	Star Harbour	-10.8	162.2	Chestnut	MA184	M
<i>megarhynchus</i>	Makira	6/24/09	Star Harbour	-10.8	162.2	Intermediate	MA185	F
<i>megarhynchus</i>	Makira	6/24/09	Star Harbour	-10.8	162.2	Chestnut	MA187	M
<i>megarhynchus</i>	Makira	6/15/11	Kira Kira	-10.5	161.9	Chestnut	MA193	F
<i>megarhynchus</i>	Makira	3/15/12	Kira Kira	-10.5	161.9	Chestnut	MA230	F
<i>megarhynchus</i>	Makira	3/15/12	Kira Kira	-10.5	161.9	Chestnut	MA231	F
<i>megarhynchus</i>	Makira	6/30/12	Kira Kira	-10.5	161.9	Chestnut	MA250	F
<i>megarhynchus</i>	Makira	3/11/14	Waimasi	-10.4	161.7	Intermediate	MA434	M
<i>megarhynchus</i>	Makira	3/11/14	Waimasi	-10.4	161.7	Intermediate	MA435	F
<i>megarhynchus</i>	Makira	3/12/14	Waimasi	-10.4	161.7	Chestnut	MA440	M
<i>megarhynchus</i>	Makira	3/12/14	Waimasi	-10.4	161.7	Intermediate	MA441	F
<i>megarhynchus</i>	Makira	6/15/18	Waimasi	-10.4	161.7	Chestnut	MA704	M
<i>megarhynchus</i>	Makira	6/15/18	Waimasi	-10.4	161.7	Melanic	MA705	M
<i>megarhynchus</i>	Makira	7/2/18	Waimasi	-10.4	161.7	Chestnut	MA714	M
<i>megarhynchus</i>	Makira	5/15/09	Kira Kira	-10.5	161.9	Chestnut	MA129**	M
<i>ugiensis</i>	Santa Ana	8/8/06	Gupuna	-10.8	162.5	Melanic	SA082	M
<i>ugiensis</i>	Santa Ana	8/10/06	Gupuna	-10.8	162.5	Intermediate	SA085	M
<i>ugiensis</i>	Santa Ana	8/10/06	Gupuna	-10.8	162.5	Melanic	SA087	F

<i>ugiensis</i>	Santa Ana	8/11/06	Gupuna	-10.8	162.5	Melanic	SA095	M
<i>ugiensis</i>	Santa Ana	6/17/07	Gupuna	-10.8	162.5	Melanic	SA105	M
<i>ugiensis</i>	Santa Ana	6/17/07	Gupuna	-10.8	162.5	Melanic	SA106	M
<i>ugiensis</i>	Santa Ana	6/17/07	Gupuna	-10.8	162.5	Melanic	SA107	M
<i>ugiensis</i>	Santa Ana	5/7/08	Gupuna	-10.8	162.5	Melanic	SA121	M
<i>ugiensis</i>	Santa Ana	5/7/08	Gupuna	-10.8	162.5	Melanic	SA123	M
<i>ugiensis</i>	Santa Ana	5/8/08	Gupuna	-10.8	162.5	Melanic	SA124	M
<i>ugiensis</i>	Santa Ana	5/7/08	Gupuna	-10.8	162.5	Melanic	SA125	F
<i>ugiensis</i>	Santa Ana	3/20/13	Gupuna	-10.8	162.5	Melanic	SA267	M
<i>ugiensis</i>	Santa Catalina	6/22/13	Santa Catalina	-10.9	162.5	Intermediate	SC275	F
<i>ugiensis</i>	Santa Catalina	6/22/13	Santa Catalina	-10.9	162.5	Intermediate	SC277	F
<i>ugiensis</i>	Santa Catalina	6/22/13	Santa Catalina	-10.9	162.5	Intermediate	SC278	M
<i>ugiensis</i>	Santa Catalina	6/22/13	Santa Catalina	-10.9	162.5	Melanic	SC283	M
<i>ugiensis</i>	Santa Catalina	6/24/13	Santa Catalina	-10.9	162.5	Melanic	SC296	M
<i>ugiensis</i>	Santa Catalina	6/24/13	Santa Catalina	-10.9	162.5	Melanic	SC402	M
<i>ugiensis</i>	Santa Catalina	6/24/13	Santa Catalina	-10.9	162.5	Melanic	SC404	M
<i>ugiensis</i>	Ugi	4/26/08	Pawa	-10.3	161.7	Melanic	UG108	M
<i>ugiensis</i>	Ugi	4/27/08	Pawa	-10.3	161.7	Melanic	UG115	M
<i>ugiensis</i>	Ugi	4/27/08	Pawa	-10.3	161.7	Melanic	UG116	F
<i>ugiensis</i>	Ugi	5/18/09	Pawa	-10.3	161.7	Melanic	UG147	M
<i>ugiensis</i>	Ugi	5/18/09	Pawa	-10.3	161.7	Melanic	UG148	M
<i>ugiensis</i>	Ugi	5/19/09	Pawa	-10.3	161.7	Melanic	UG152	M
<i>ugiensis</i>	Ugi	5/19/09	Pawa	-10.3	161.7	Melanic	UG155	M
<i>ugiensis</i>	Ugi	5/19/09	Pawa	-10.3	161.7	Melanic	UG156	M
<i>ugiensis</i>	Ugi	5/20/09	Pawa	-10.3	161.7	Melanic	UG160	F
<i>ugiensis</i>	Ugi	5/20/09	Pawa	-10.3	161.7	Melanic	UG161	M
<i>ugiensis</i>	Ugi	6/8/11	Pawa	-10.3	161.7	Melanic	UG197	M
<i>ugiensis</i>	Ugi	3/13/12	Pawa	-10.3	161.7	Melanic	UG220	M

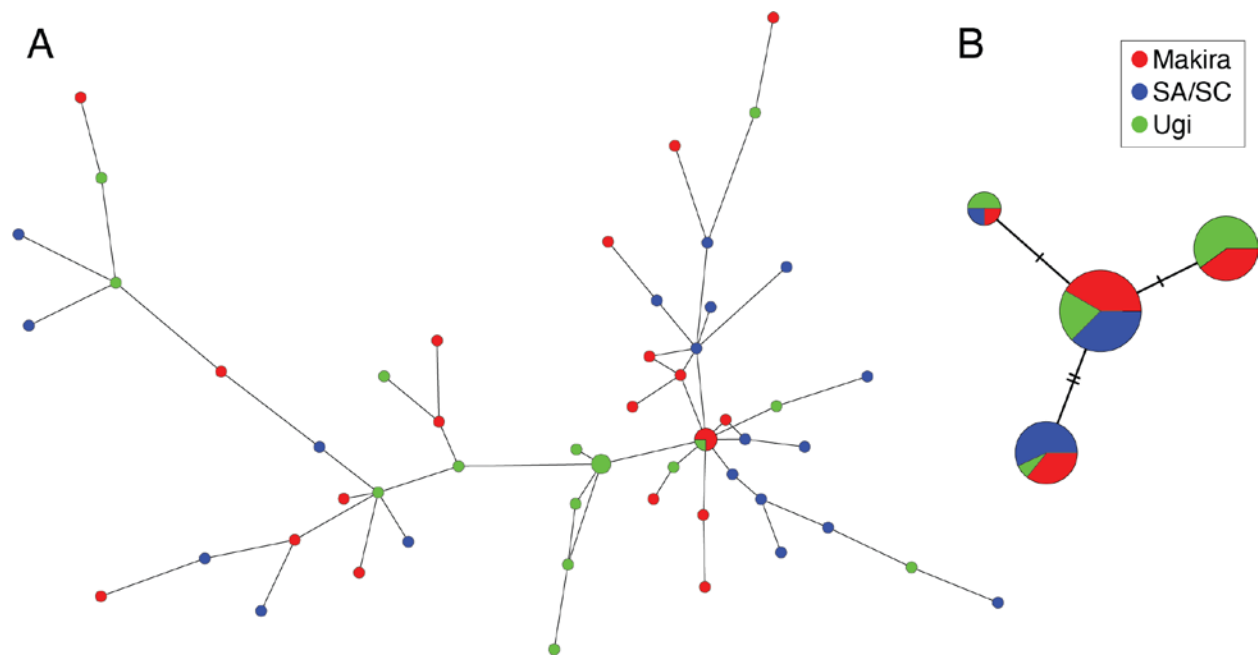
<i>ugiensis</i>	Ugi	7/4/13	Bio	-10.2	161.7	Melanic	UG407	M
<i>ugiensis</i>	Ugi	3/15/14	Pawa	-10.3	161.7	Melanic	UG448	M
<i>ugiensis</i>	Ugi	3/15/14	Pawa	-10.3	161.7	Melanic	UG450	F
<i>ugiensis</i>	Ugi	3/15/14	Pawa	-10.3	161.7	Melanic	UG451	M

\*We determined sex by calculating the average depth of coverage across all the positions in each of six different contigs. Three of these contigs were autosomal and three were part of the Z chromosome. We subsequently averaged the depth of coverage for the three Z-linked contigs and divided it by the average from the three autosomal contigs. This process produced values around 0.5 for heterogametic females, and values close to 1 for males. Additionally, for a subset of 39 individuals we also determined sex through PCR as described in reference (1). Both methods produced congruent results.

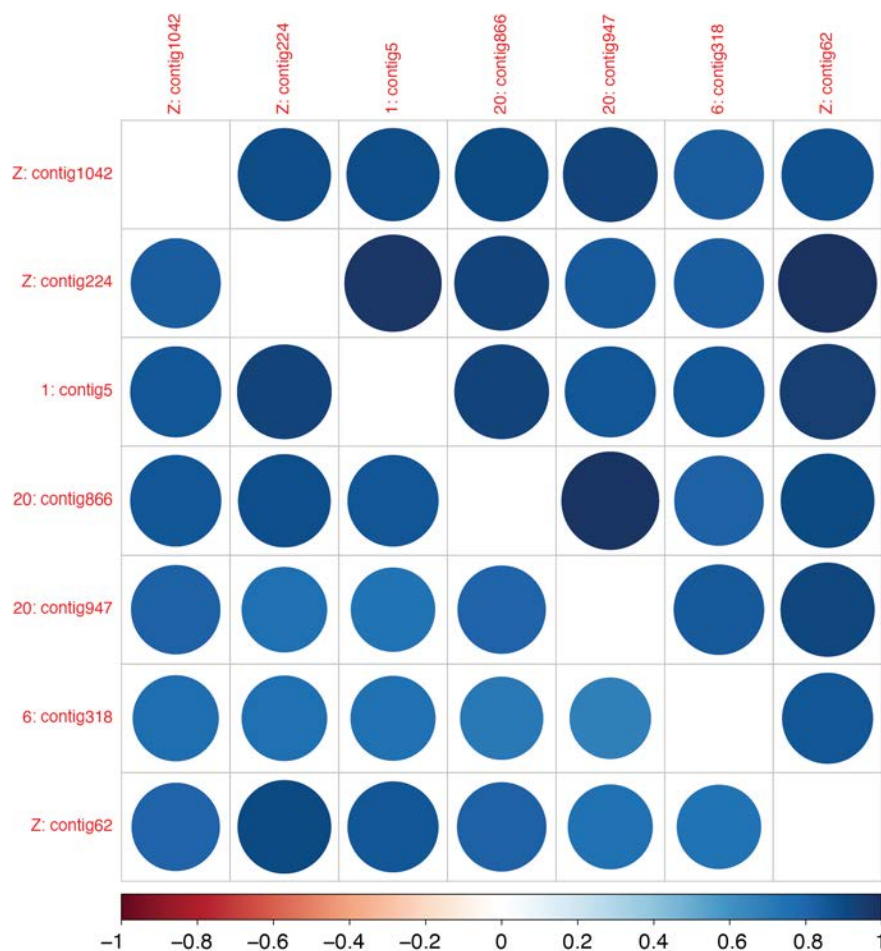
\*\*This sample is labelled as UG129 and not MA129 in the different files from our bioinformatics pipeline (e.g., vcf files), yet was treated correctly as a sample originating from Makira and not from Ugi (as other samples denoted with UG).

## References

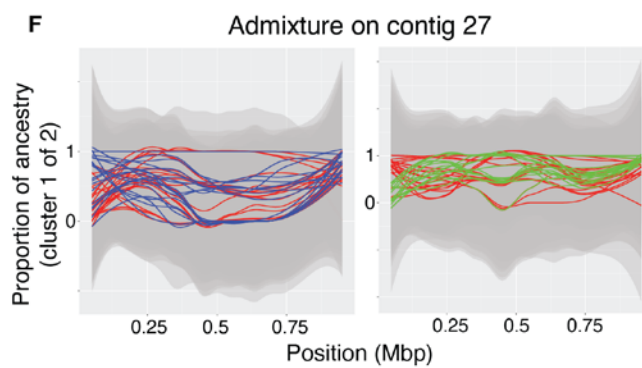
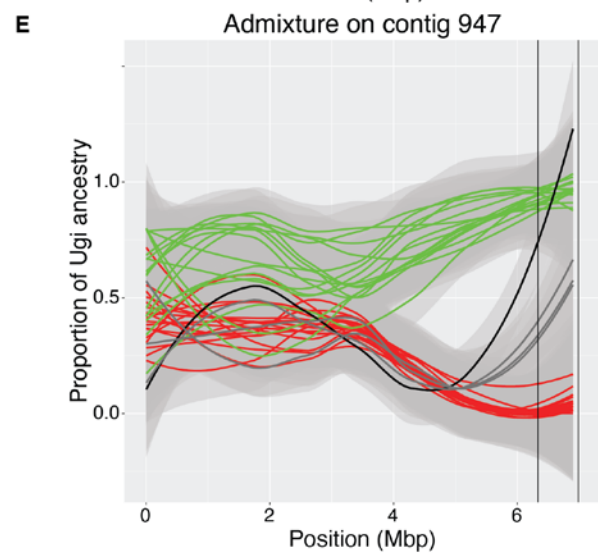
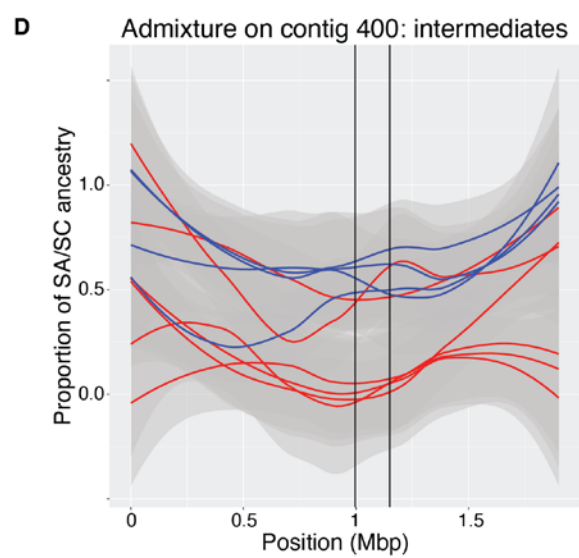
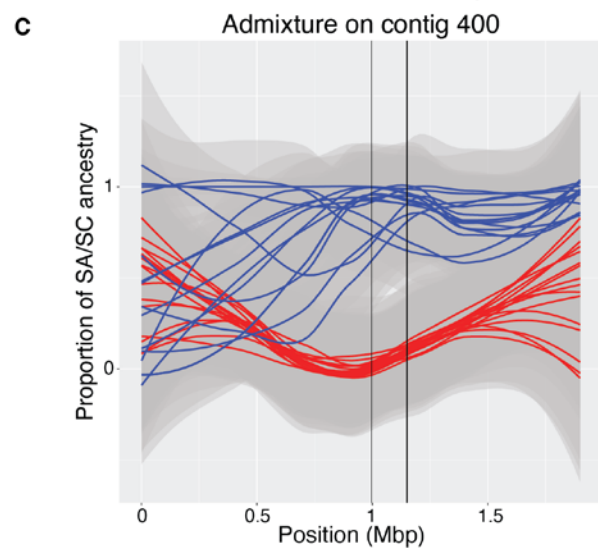
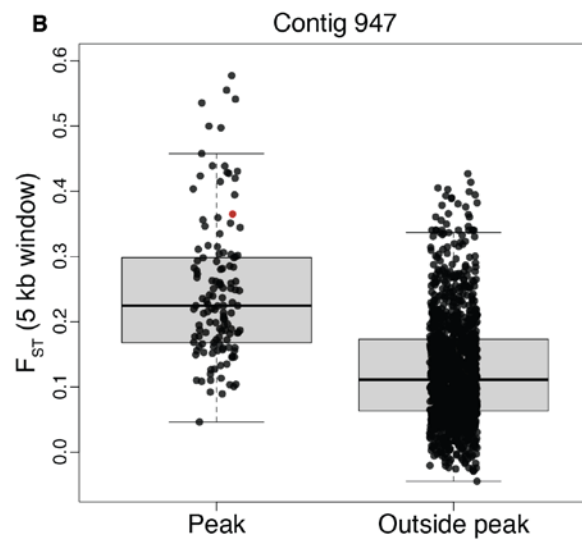
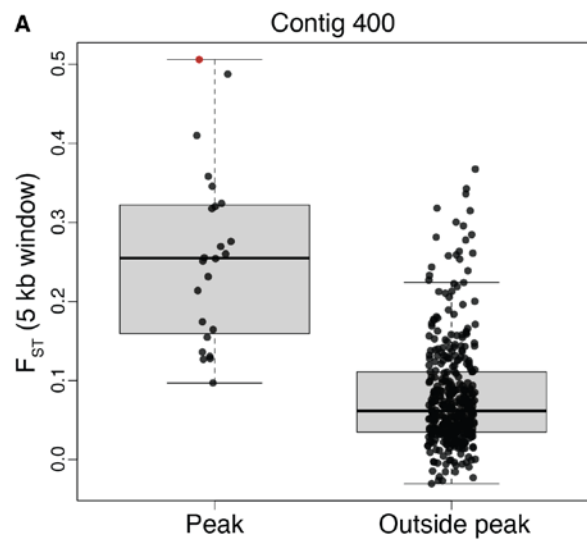
1. N. W. Kahn, J. St. John, T. W. Quinn, Chromosome-specific intron size differences in the avian CHD gene provide an efficient method for sex identification in birds. *The Auk*, 1074–1078 (1998).



**S1 Fig.** Mitochondrial minimum spanning networks. **A.** Haplotype network based on a ~17 kbp alignment of the mitochondrial genome. **B.** Haplotype network based on 650 bp of the mitochondrial COI gene, commonly used for species identification. Branch lengths are proportional to the number of nucleotide differences between haplotypes, which are indicated by short lines on each branch (omitted for simplicity in the case of the full mitochondrial network).

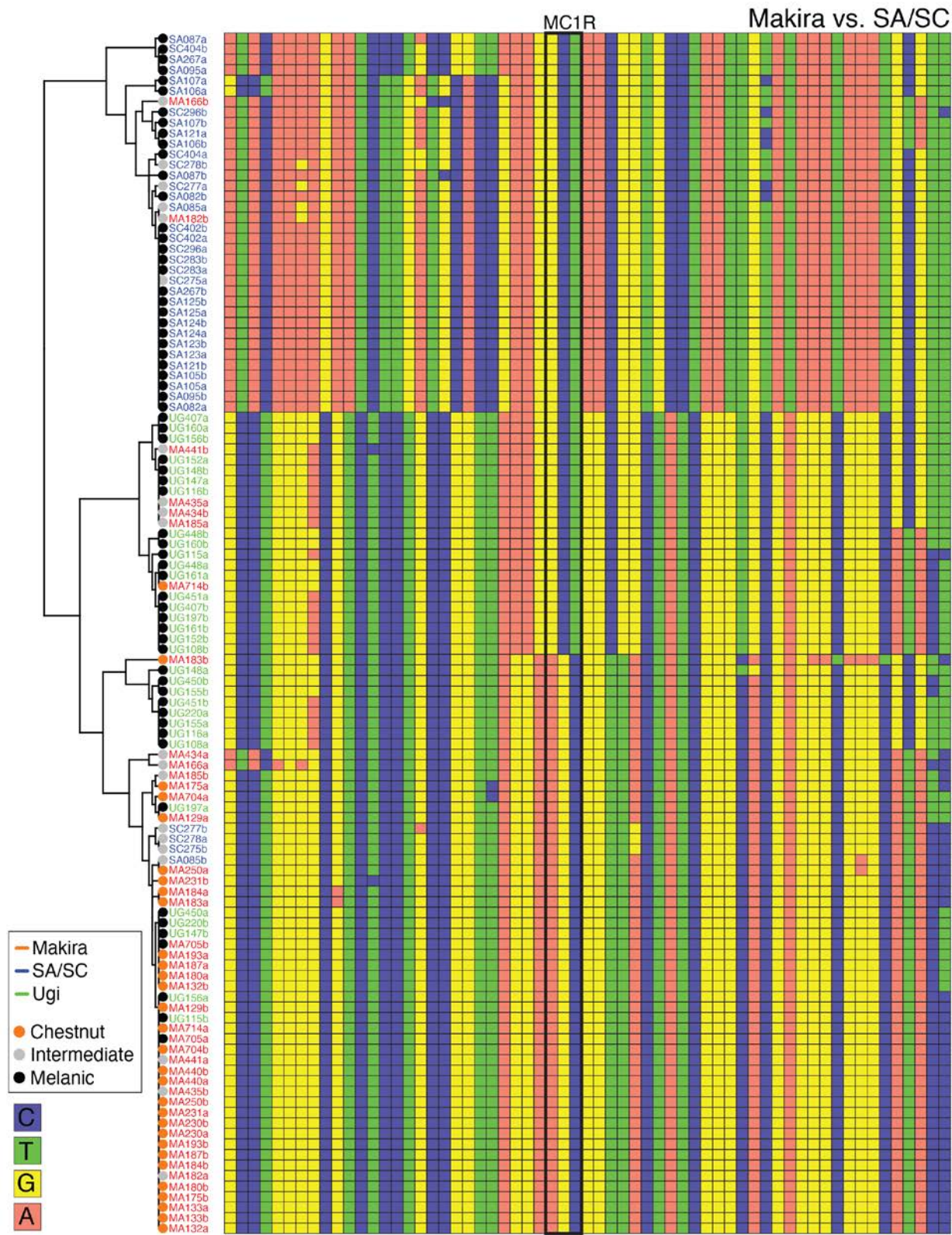


**S2 Fig.** Linkage disequilibrium among association peaks identified in the GWAS conducted with Makira and Ugi individuals. Average (below the diagonal) and maximum (above the diagonal)  $R^2$  values among all the statistical outlier sites in the GWAS from different pairs of association peaks. The chromosome and contig to which each peak belongs is indicated in red, and the size and color of the circles denotes the magnitude of LD.

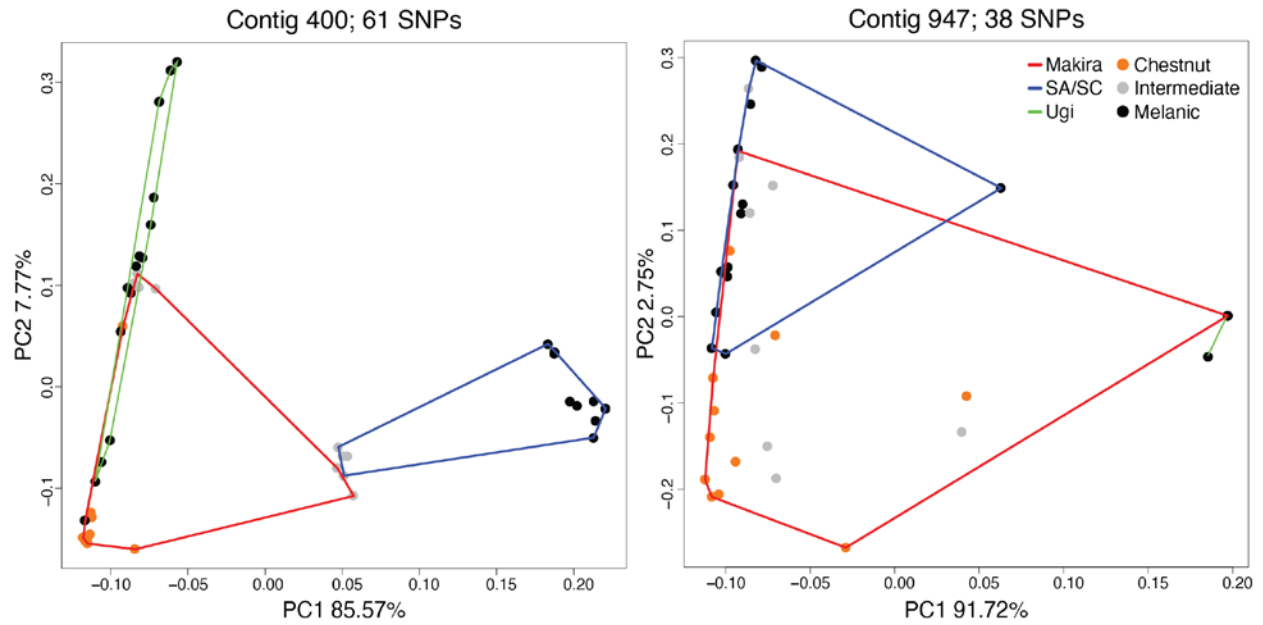


**S3 Fig.** Genetic differentiation within and outside of association peaks. **A and B.**  $F_{ST}$  values calculated for 5 kb windows inside and outside association peaks. The plot for contig 400 compares individuals from SA/SC and Makira, while the plot for contig 947 compares individuals from Ugi and Makira. The red dot denotes the window containing the *MC1R* and *ASIP* genes. **C to F.** Smoothed ancestry values across contigs, with association peaks indicated between vertical lines. Ancestry values were calculated in 100 kb sliding windows in Admixture, and the analysis was restricted to Makira vs. SA/SC for contig 400 and Makira vs. Ugi for contig 947. The plot in (**C**) shows increased separation between individuals from Makira and SA/SC in the peak region on contig 400 and extending approximately 0.5 Mbp in each direction. The plot includes only individuals from Makira and SA/SC that were homozygotes for the *Asp119* or *Asn119* *MC1R* mutation, respectively. The plot in (**D**) shows the six individuals from Makira and the four individuals from SA/SC which were heterozygotes for this mutation and had intermediate plumage, and shows overall more admixture than what is seen in (**C**). **E.** Ancestry values across contig 947 showing increased resolution in the peak region and extending approximately 2 Mb downstream. Three individuals from Makira which were heterozygotes for the *Tre55* *ASIP* mutation are labeled in gray (only one of these individuals had intermediate plumage). The single melanic individual from Makira (MA705), which was homozygous for the derived *Tre55* mutation, is labelled in black. For all four individuals, Ugi ancestry decreases to levels comparable to other Makira individuals about 1 Mb downstream of the peak. **F.** Ancestry across contig 27 shows little resolution compared to the association peaks on contigs 400 and 947. We note that ancestry values range from 0 to 1, but that the plots extend beyond this range because of the smoothing algorithm and particularly the uncertainty shown by the confidence bands. Values beyond the [0,1] interval are therefore meaningless.



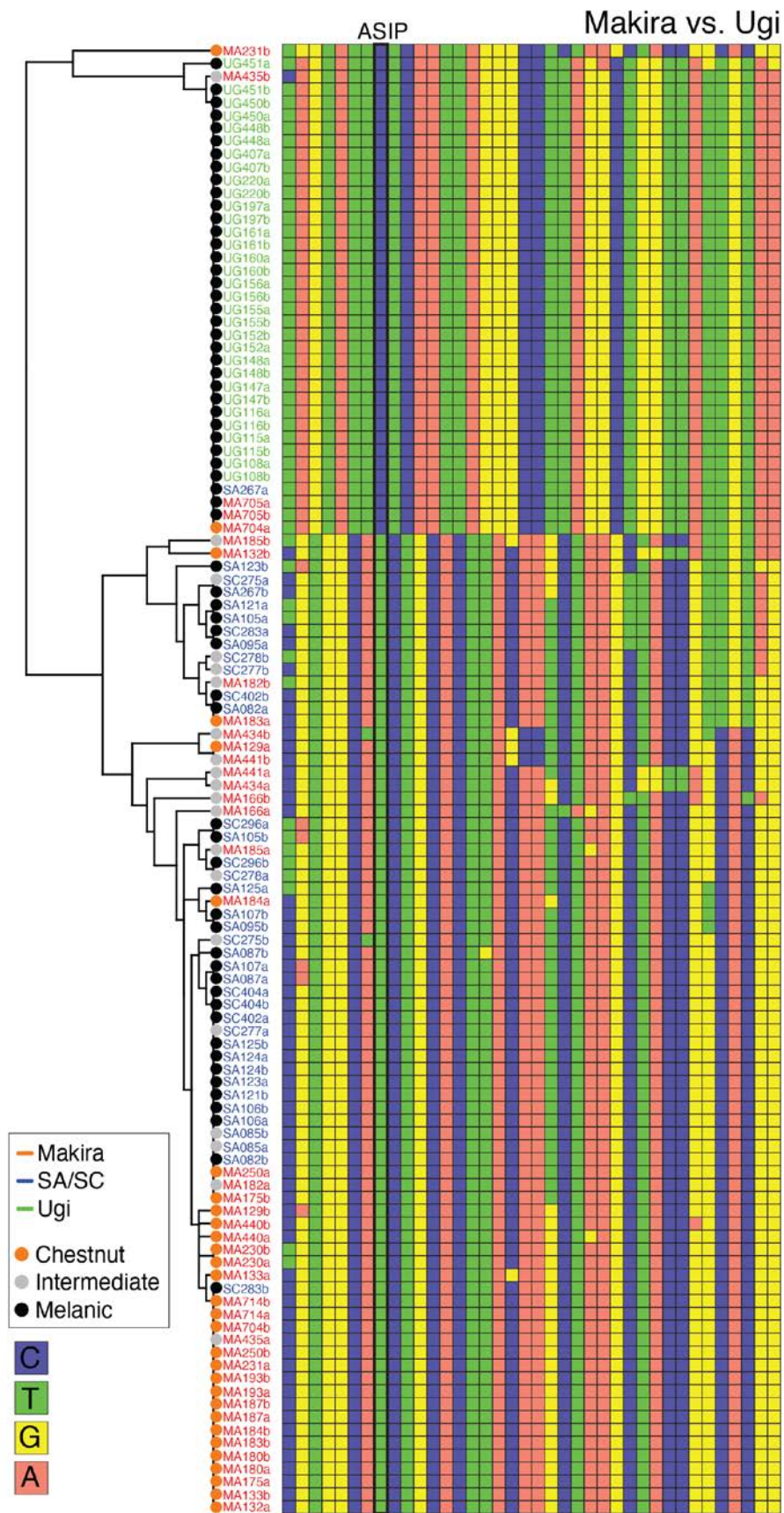


**S4 Fig.** Clustering of haplotypes obtained from the association peak on contig 400. Phased genotypes for the 61 SNPs located in the association peak on contig 400. Rows represent single chromosomes, therefore individuals are represented twice in the clustering tree on the left. The four nucleotides, the collection locality and the coloration phenotype are color-coded as indicated at left. The three SNPs within the *MC1R* coding region are indicated with a black rectangle. All individuals from SA/SC contained at least one haplotype in the region delimited by the SNPs with significant association scores around *MC1R* that differed from the one present in most individuals from Makira and Ugi. All the melanic individuals from SA/SC possessed two copies of this haplotype, while the four individuals with intermediate coloration possessed one of each, as was the case for two of the six individuals with intermediate coloration from Makira. All melanic individuals from SA/SC carried two copies of the derived *Asn119* mutation, while all but one of the chestnut-bellied individuals from Makira had two copies of *Asp119*. The individuals with intermediate coloration (from either Makira or SA/SC) were heterozygotes for this coding mutation. The exception to this pattern was a single chestnut-bellied individual from Makira (MA714), which was a heterozygote yet was scored in the field under heavy molt and may have been incorrectly classified as having a chestnut belly. Finally, the derived *Asn199* mutation existed primarily on the haplotype background found on SA/SC, but also to a lesser extent on the haplotype background found on Makira and Ugi. We note that the chestnut-bellied MA714 bird carried the *Asn119* mutation on the most common haplotype background observed on Makira. The birds from Ugi, which are all melanic, were sometimes homozygous for either allele or heterozygotes.

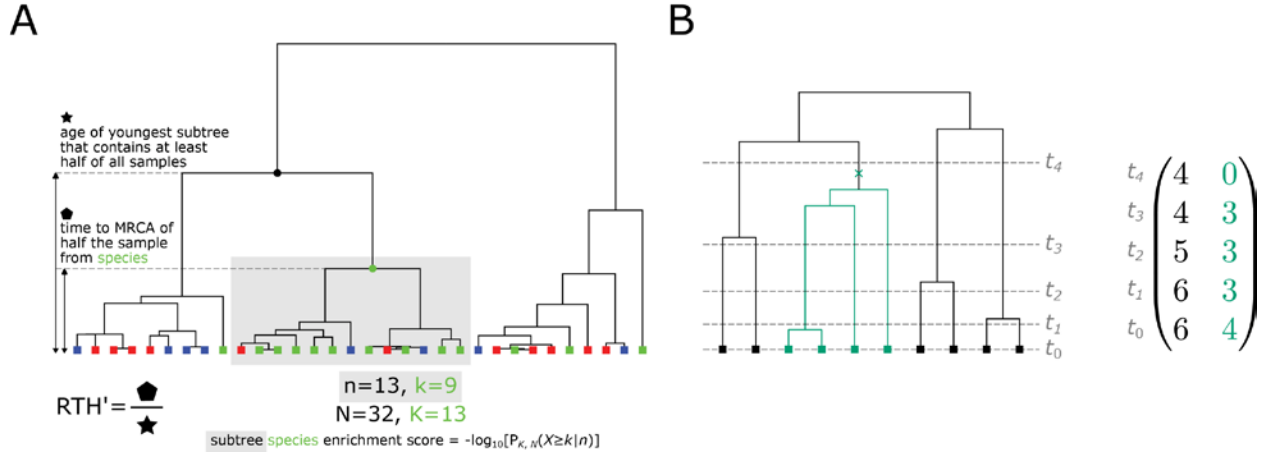


**S5 Fig.** Principal component analyses derived from the SNPs within the association peaks. PCAs from the variants within the association peaks on contig 400 and contig 947. Coloration phenotype and the island where individuals were sampled are color-coded.

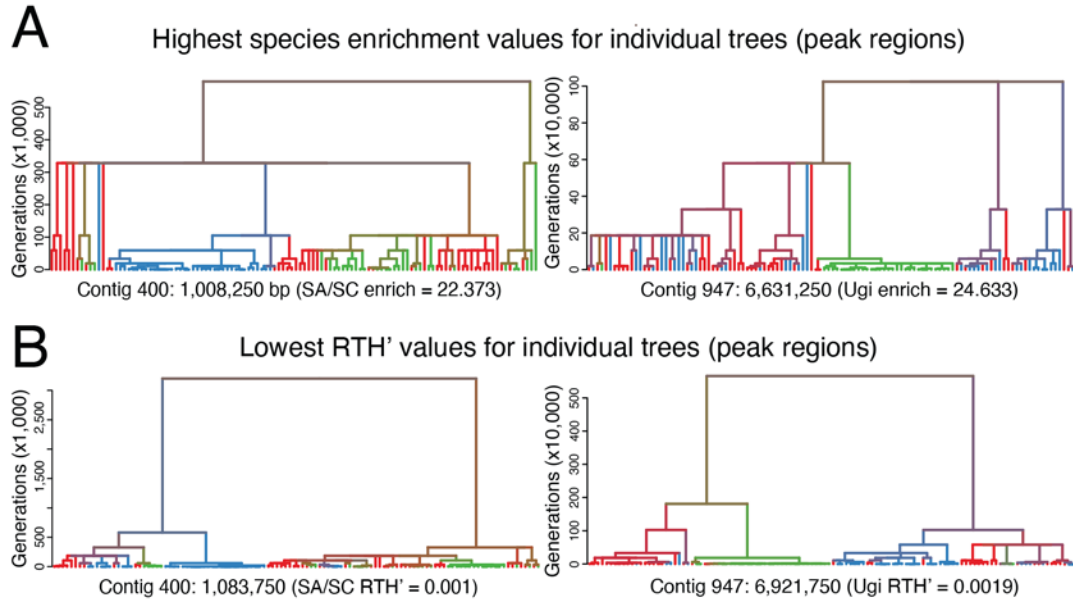




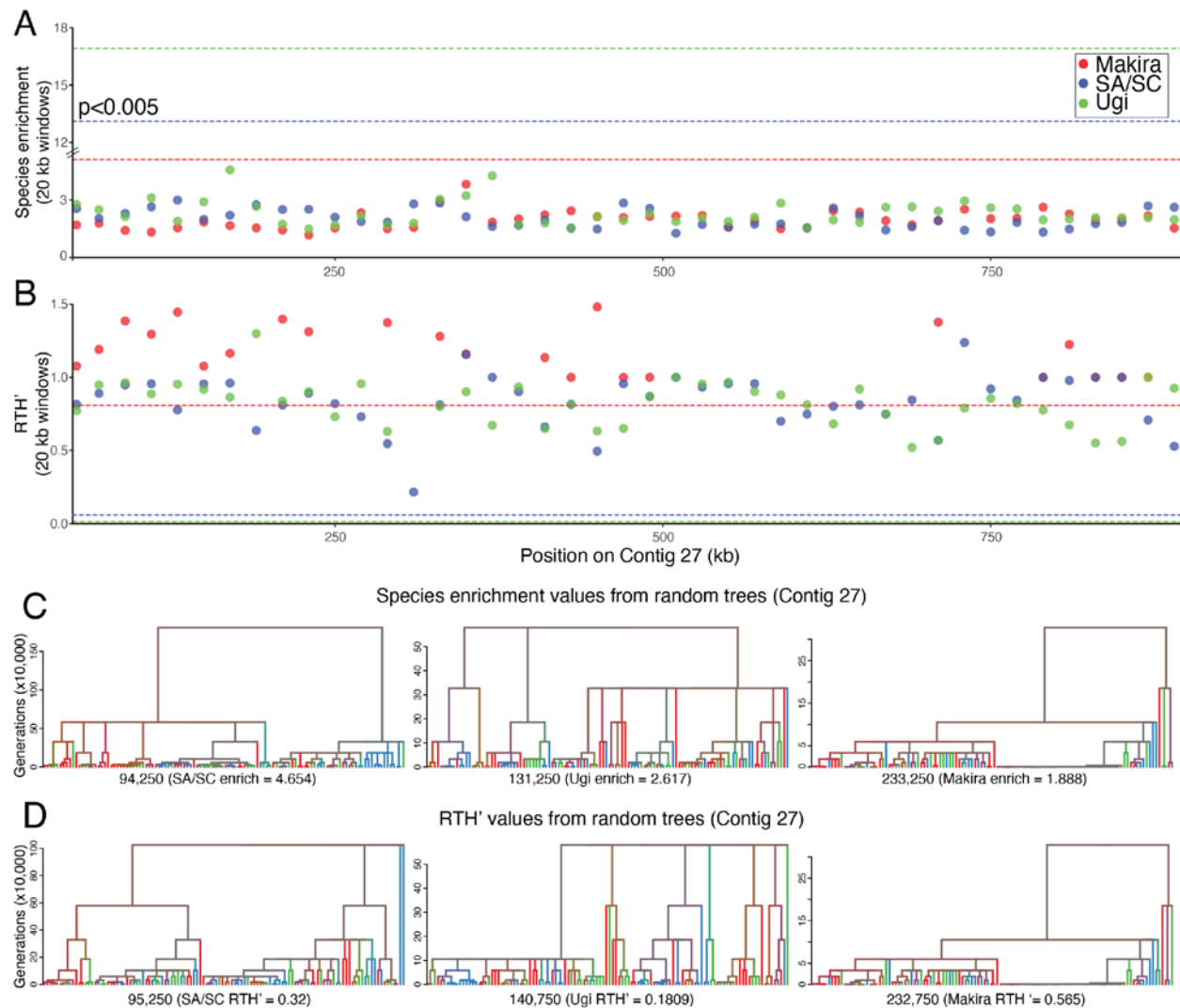
**S6 Fig.** Clustering of haplotypes obtained from the association peak on contig 947. Details as in **S4 Fig**. Ugi individuals had two copies of a haplotype that was different from the one present in SA/SC and Makira individuals. The only melanic individual from Makira (MA705) also carried two copies of the derived Ugi haplotype. We found a few heterozygote individuals but there was no obvious pattern with respect to their coloration.



**S7 Fig.** Statistics and feature encoding of the genealogy. **A.** RTH' tests for the reduction in within-species TMRCA and is defined as the ratio between the TMRCA of half of the samples from a *given species* and the age of the youngest subtree that contains at least half of *all* samples. The species enrichment score tests for species differentiation in local trees and is defined as the maximum score associated with a given species in a subtree of the full genealogy. For a subtree, the species enrichment score is the probability of observing the number of samples of a particular species in that subtree under a hypergeometric distribution. Here the coloring of the leaves indicates hypothetical species and the example illustrates the statistics with respect to the green species. **B.** Genealogical features for the SIA model consist of the number of lineages in the genealogy at a set of discrete time points ( $t_0, t_1, \dots$ ). The time points are chosen in an approximately log-uniform manner resulting in finer discretization of more recent time scales. In addition, when encoding the genealogical features at a particular site of interest, we encode separately the counts of ancestral (shown in black in the example) and derived (shown in aquamarine) lineages. The aquamarine cross indicates the branch where the mutation occurred.

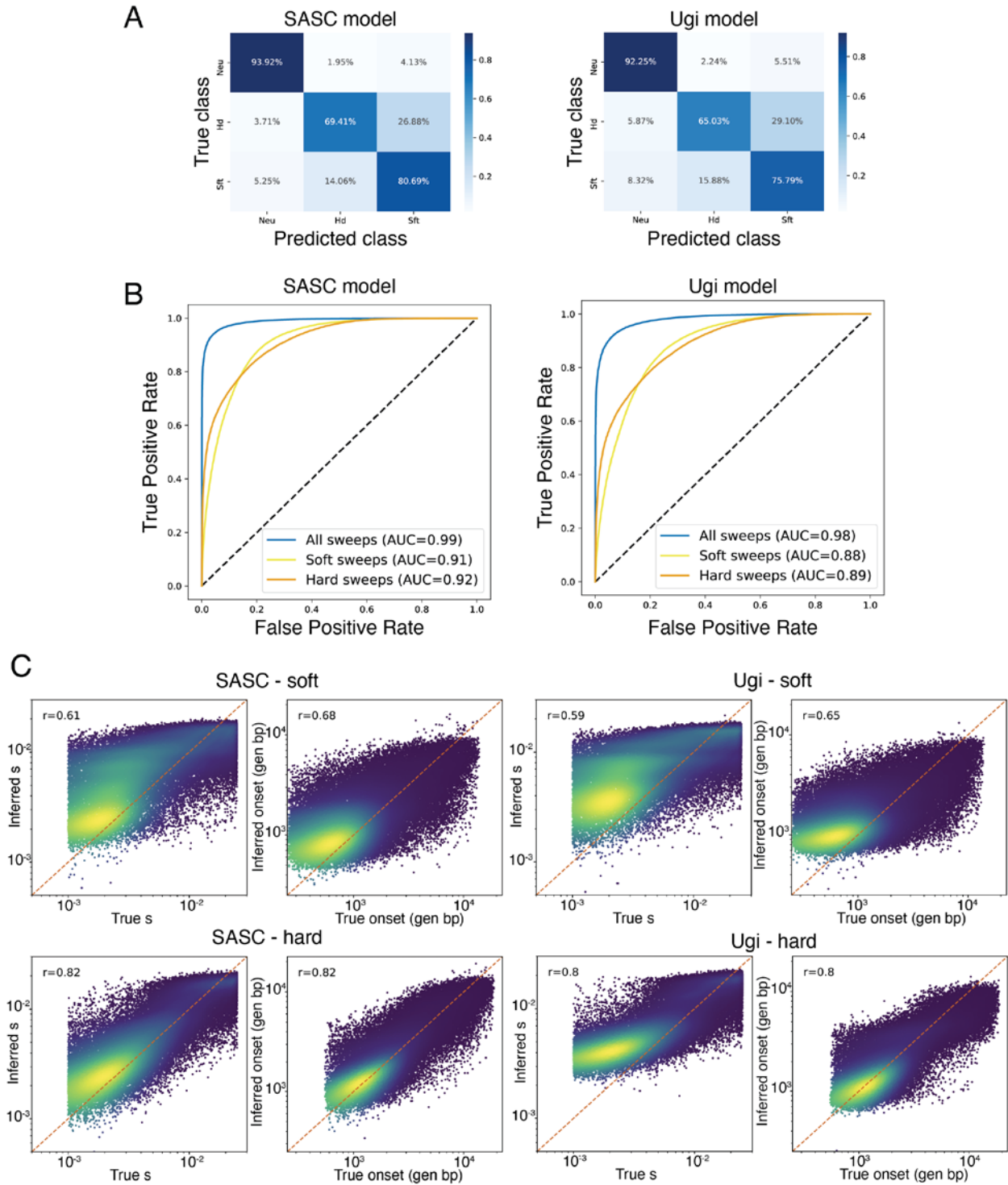


**S8 Fig.** Highest species enrichment scores and lowest RTH' values for peak regions on contigs 400 and 947. Representative trees from the peak regions which show extreme values of species-enrichment (**A**) and RTH' (**B**) for SA/SC and Ugi.



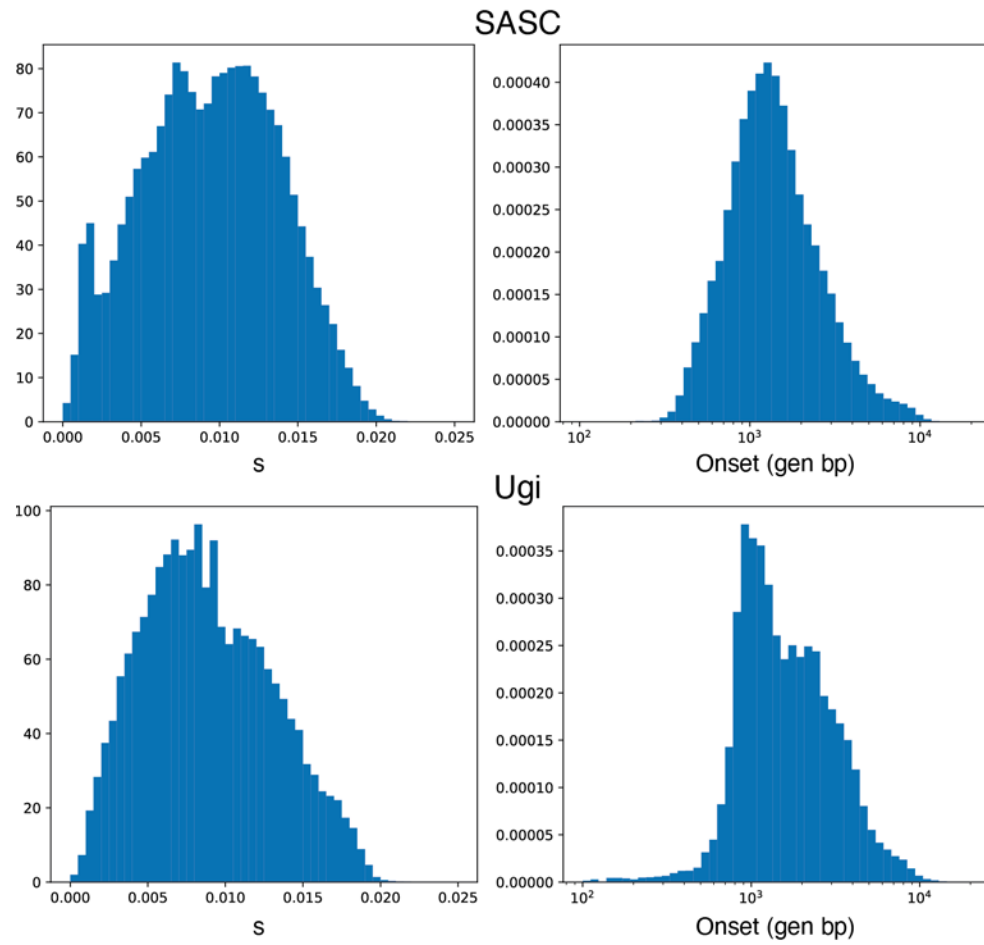
**S9 Fig.** Species enrichment scores and RTH' values for a control contig. Plots showing species enrichment (**A**) and RTH' (**B**) in 20 kb windows along contig 27. Horizontal lines show species-specific levels of statistical significance ( $p < 0.005$ ). Trees obtained from random positions on contig 27 (each position is shown under the tree) indicating species enrichment scores (**C**) and RTH' (**D**) values for each population.





**S10 Fig.** Benchmarking of the SIA models. **A.** Confusion matrices generated by applying the SA/SC and Ugi models to simulated data. Instead of applying a specific probability threshold, the predicted class was identified as the one with the highest probability according to the model. Under this maximum likelihood classification scheme, both models perform best distinguishing neutral from selected sites, and moderately when identifying soft or hard sweeps. **B.** One-versus-rest (OvR) receiver

operating characteristic (ROC) curves of the model classification performance on simulated data. The models perform very well on distinguishing sweeps, and less so on precisely identifying hard or soft sweeps. **C.** Comparisons between true and inferred selection coefficients and time of selection onset (expressed in generations before present) for SA/SC and Ugi models trained to detect soft or hard sweeps. Models trained to predict hard sweeps generally perform better than those trained to predict soft sweeps, which tend to overestimate small selection coefficients and recent selection onset times.



**S11 Fig.** Genome-wide estimates of selection coefficients and time of selection onset. Predictions are based on all the variants from the 190 scaffolds longer than 100kb.