Supplementary Materials for *Recombination suppression and selection affect local ancestries in genomes of a migratory songbird*

Jun Ishigohoka[†,1] Karen Bascón-Cardozo[†,1] Andrea Bours[1] Janina Fuß[2]
Arang Rhie[3] Jacquelyn Mountcastle[4] Bettina Haase[4] William Chow[5]
Joanna Collins[5] Kerstin Howe[5] Marcela Uliano-Silva[5] Olivier Fedrigo[4]
Erich D. Jarvis[4,6,7] Javier Pérez-Tris[8] Juan Carlos Illera[9]
Miriam Liedvogel[1,10,*]

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany

[2]Institute of Clinical Molecular Biology (IKMB), Kiel University, Kiel, Germany

[3]Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[4]The Vertebrate Genome Lab, Rockefeller University, New York, NY, USA

[5]Wellcome Sanger Institute, Cambridge, UK

[6]Laboratory of Neurogenetics of Language, Rockefeller University, New York, NY, USA

[7]The Howards Hughes Medical Institute, Chevy Chase, MD, USA

[8]Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid, Spain

[9]Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, Mieres, Spain

[10]Institute of Avian Research, Wilhelmshaven, Germany

[†] These authors contributed equally to this work.

[*] Correspondence: Miriam Liedvogel <miriam.liedvogel@ifv-vogelwarte.de>

Supplementary tables

Supplementary Table 1: Samples used in this study

Download link to the table

Supplementary Table 2: Mann-Whitney U tests comparing heterozygosity between genotypes of the class-1 genomic islands. For genotypes with sample size smaller than 5, statistical tests were not performed.

chr	pos.from	pos.to	geno1	geno2	n1	n2	p.value	Bonferroni-corrected p.value
chr_6	5684792	6260313	AA	BB	171	1	NA	NA
chr_6	5684792	6260313	AA	AB	171	7	7.63e-06	8.39e-05
chr_6	5684792	6260313	BB	AB	1	7	NA	NA
chr_12	14126710	22227355	AA	BB	121	5	1.77e-04	1.95e-03
chr_12	14126710	22227355	AA	AB	121	53	7.02e-25	7.72e-24
chr_12	14126710	22227355	BB	AB	5	53	2.55e-04	2.81e-03
chr_14	43	259166	AA	BB	159	2	NA	NA
chr_14	43	259166	AA	AB	159	18	3.32e-03	3.65e-02
chr_14	43	259166	BB	AB	2	18	NA	NA
chr_28	917037	1154843	AA	BB	115	9	3.78e-06	4.16e-05
chr_28	917037	1154843	AA	AB	115	55	2.49e-03	2.74e-02
chr_28	917037	1154843	BB	AB	9	55	3.92e-06	4.31e-05
chr_30	105195	1372664	AA	BB	89	11	2.53e-07	2.78e-06
chr_30	105195	1372664	AA	AB	89	79	5.01e-22	5.51e-21
chr_30	105195	1372664	BB	AB	11	79	9.61e-08	1.06e-06

Supplementary Table 3: Permutation tests(n=10,000) for π of genotypes of the class-1 genomic islands by randomising the positions of the windows. P values are based on the rank of the observed value from left to right, asking whether observed value is significantly lower than expected from the null distribution.

chr	pos.from	pos.to	p.val(π_{AA})	p.val(π_{BB})
chr_12	14126710	22227355	1e-04	1e-04
chr_14	43	259166	1e-04	1e-04
chr_28	917037	1154843	0.9921	0.9758
chr_30	105195	1372664	1	6e-04

Supplementary Table 4: Permutation tests(n=10,000) comparing d_{xy} and F_{ST} between genotypes of the class-1 genomic islands by randomising the positions of the windows. P values of d_{xy} and F_{ST} are based on the rank of the observed value from right to left, asking whether observed value is significantly higher than expected from the null distribution.

chr	pos.from	pos.to	p.val(d_{xy})	p.val(F_{ST})
chr_12	14126710	22227355	1e-04	1e-04
chr_14	43	259166	1	1e-04
chr_28	917037	1154843	0.0115	0.1156
chr_30	105195	1372664	1e-04	1e-04

Supplementary Table 5: The effect of genotype of the putative inversion and interval(inside/outside of the class-1 genomic island) on coalescence time by analysis of variance based on generalised linear mixed-effect model with the Poisson error distribution and the log link function.

chr	term	χ^2	df	p.value
chr_6	geno	29.599	2	3.7375e-07
chr_6	interval	1669.102	1	0.0000e+00
chr_6	geno:interval	101.551	2	8.8808e-23
chr_12	geno	310.056	2	4.7020e-68
chr_12	interval	837.357	1	4.0763e-184

chr	term	χ^2	df	p.value
chr_12	geno:interval	5135.296	2	0.0000e+00
chr_14	geno	10.577	2	5.0493e-03
chr_14	interval	424.354	1	2.7531e-94
chr_14	geno:interval	49.984	2	1.3999e-11
chr_28	geno	25.647	2	2.6973e-06
chr_28	interval	13.381	1	2.5421e-04
chr_28	geno:interval	59.538	2	1.1791e-13
chr_30	geno	35.960	2	1.5541e-08
chr_30	interval	45.107	1	1.8655e-11
chr_30	geno:interval	257.081	2	1.4980e-56

Supplementary Table 6: Post hoct tests for the coalescence time in class-1 genomic islands by pairwise comparisons using Z-tests corrected with Bonferroni method following the GLMM.

chr	comparison	Z.value	P.value	
chr_6	AAin-ABin	-10.533	0.0000	*****
chr_6	AAin-BBin	-2.980	0.0432	*
chr_6	AAin-AAout	-31.699	0.0000	*****
chr_6	AAin-ABout	-30.490	0.0000	*****
chr_6	AAin-BBout	-22.109	0.0000	*****
chr_6	ABin-BBin	3.724	0.0029	***
chr_6	ABin-AAout	-20.909	0.0000	*****
chr_6	ABin-ABout	-24.538	0.0000	*****
chr_6	ABin-BBout	-14.512	0.0000	*****
chr_6	BBin-AAout	-14.776	0.0000	*****
chr_6	BBin-ABout	-15.357	0.0000	*****
chr_6	BBin-BBout	-14.138	0.0000	*****
chr_6	AAout-ABout	-1.301	1.0000	n.s.
chr_6	AAout-BBout	0.107	1.0000	n.s.

chr	comparison	Z.value	P.value	
chr_6	ABout-BBout	0.926	1.0000	n.s.
chr_12	AAin-ABin	-10.175	0.0000	*****
chr_12	AAin-BBin	25.657	0.0000	*****
chr_12	AAin-AAout	-10.463	0.0000	*****
chr_12	AAin-ABout	-3.802	0.0021	***
chr_12	AAin-BBout	-0.813	1.0000	n.s.
chr_12	ABin-BBin	35.757	0.0000	*****
chr_12	ABin-AAout	6.914	0.0000	*****
chr_12	ABin-ABout	21.591	0.0000	*****
chr_12	ABin-BBout	9.362	0.0000	*****
chr_12	BBin-AAout	-28.898	0.0000	*****
chr_12	BBin-ABout	-29.435	0.0000	*****
chr_12	BBin-BBout	-72.044	0.0000	*****
chr_12	AAout-ABout	-0.540	1.0000	n.s.
chr_12	AAout-BBout	2.449	0.2149	n.s.
chr_12	ABout-BBout	2.989	0.0420	*
chr_14	AAin-ABin	-0.117	1.0000	n.s.
chr_14	AAin-BBin	7.464	0.0000	*****
chr_14	AAin-AAout	-12.723	0.0000	*****
chr_14	AAin-ABout	-9.577	0.0000	*****
chr_14	AAin-BBout	-9.100	0.0000	*****
chr_14	ABin-BBin	7.547	0.0000	*****
chr_14	ABin-AAout	-11.058	0.0000	*****
chr_14	ABin-ABout	-10.749	0.0000	*****
chr_14	ABin-BBout	-9.000	0.0000	*****
chr_14	BBin-AAout	-14.911	0.0000	*****
chr_14	BBin-ABout	-13.899	0.0000	*****
chr_14	BBin-BBout	-14.827	0.0000	*****

chr	comparison	Z.value	P.value	
chr_14	AAout-ABout	1.655	1.0000	n.s.
chr_14	AAout-BBout	0.349	1.0000	n.s.
chr_14	ABout-BBout	-1.005	1.0000	n.s.
chr_28	AAin-ABin	-1.685	1.0000	n.s.
chr_28	AAin-BBin	7.257	0.0000	*****
chr_28	AAin-AAout	-0.779	1.0000	n.s.
chr_28	AAin-ABout	-1.568	1.0000	n.s.
chr_28	AAin-BBout	0.822	1.0000	n.s.
chr_28	ABin-BBin	8.905	0.0000	*****
chr_28	ABin-AAout	1.085	1.0000	n.s.
chr_28	ABin-ABout	0.154	1.0000	n.s.
chr_28	ABin-BBout	2.500	0.1865	n.s.
chr_28	BBin-AAout	-7.844	0.0000	*****
chr_28	BBin-ABout	-8.788	0.0000	*****
chr_28	BBin-BBout	-7.920	0.0000	*****
chr_28	AAout-ABout	-0.968	1.0000	n.s.
chr_28	AAout-BBout	1.419	1.0000	n.s.
chr_28	ABout-BBout	2.382	0.2584	n.s.
chr_30	AAin-ABin	-1.685	1.0000	n.s.
chr_30	AAin-BBin	7.156	0.0000	*****
chr_30	AAin-AAout	-0.769	1.0000	n.s.
chr_30	AAin-ABout	-0.206	1.0000	n.s.
chr_30	AAin-BBout	0.917	1.0000	n.s.
chr_30	ABin-BBin	8.836	0.0000	*****
chr_30	ABin-AAout	1.406	1.0000	n.s.
chr_30	ABin-ABout	4.438	0.0001	****
chr_30	ABin-BBout	2.576	0.1498	n.s.
chr_30	BBin-AAout	-7.315	0.0000	*****

chr	comparison	Z.value	P.value	
chr_30	BBin-ABout	-7.262	0.0000	*****
chr_30	BBin-BBout	-16.790	0.0000	*****
chr_30	AAout-ABout	0.049	1.0000	n.s.
chr_30	AAout-BBout	1.156	1.0000	n.s.
chr_30	ABout-BBout	1.107	1.0000	n.s.

Supplementary Table 7: 6 Models of recombination suppression at an inversion in SLiM simulations

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Inversion frequency	0.2	0.2	0.2	0.8	0.8	0.8
Recombination suppression N-I	Yes	Yes	Yes	Yes	Yes	Yes
Recombination suppression I-I	No	Yes	Yes	No	Yes	Yes
Recombination suppression N-N	No	No	Yes	No	No	Yes

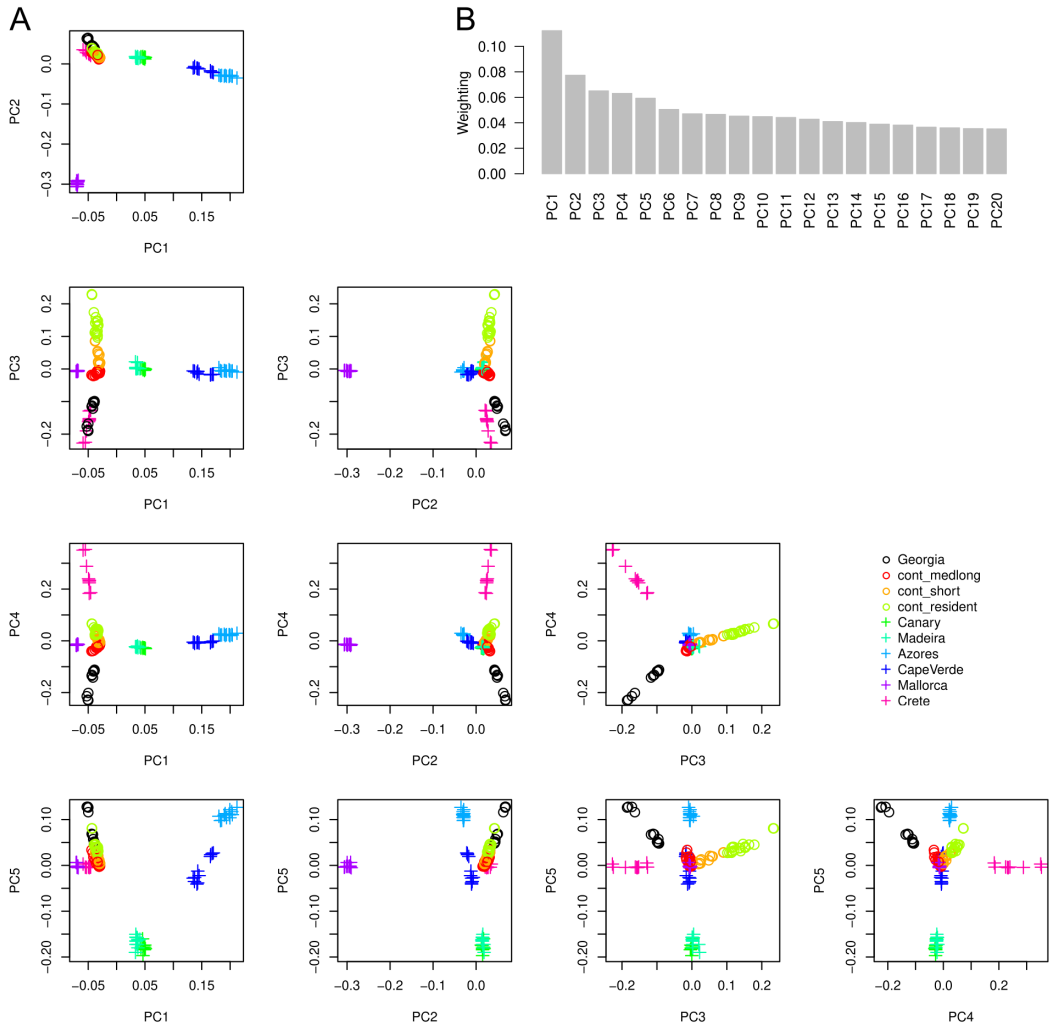
Supplementary Table 8: Permutation tests(n=10,000) for π of four populations at the class-2 genomic island on chromosome 21 in comparison to the chromosomal background by randomising the positions of the windows. P values are based on the rank of the observed value from left to right, asking whether observed value is significantly lower than expected from the null distribution.

pop	p.val
Azores	0.6331
Belgium	0.998
CapeVerde	0.9583
Spain_Caz	0.9984

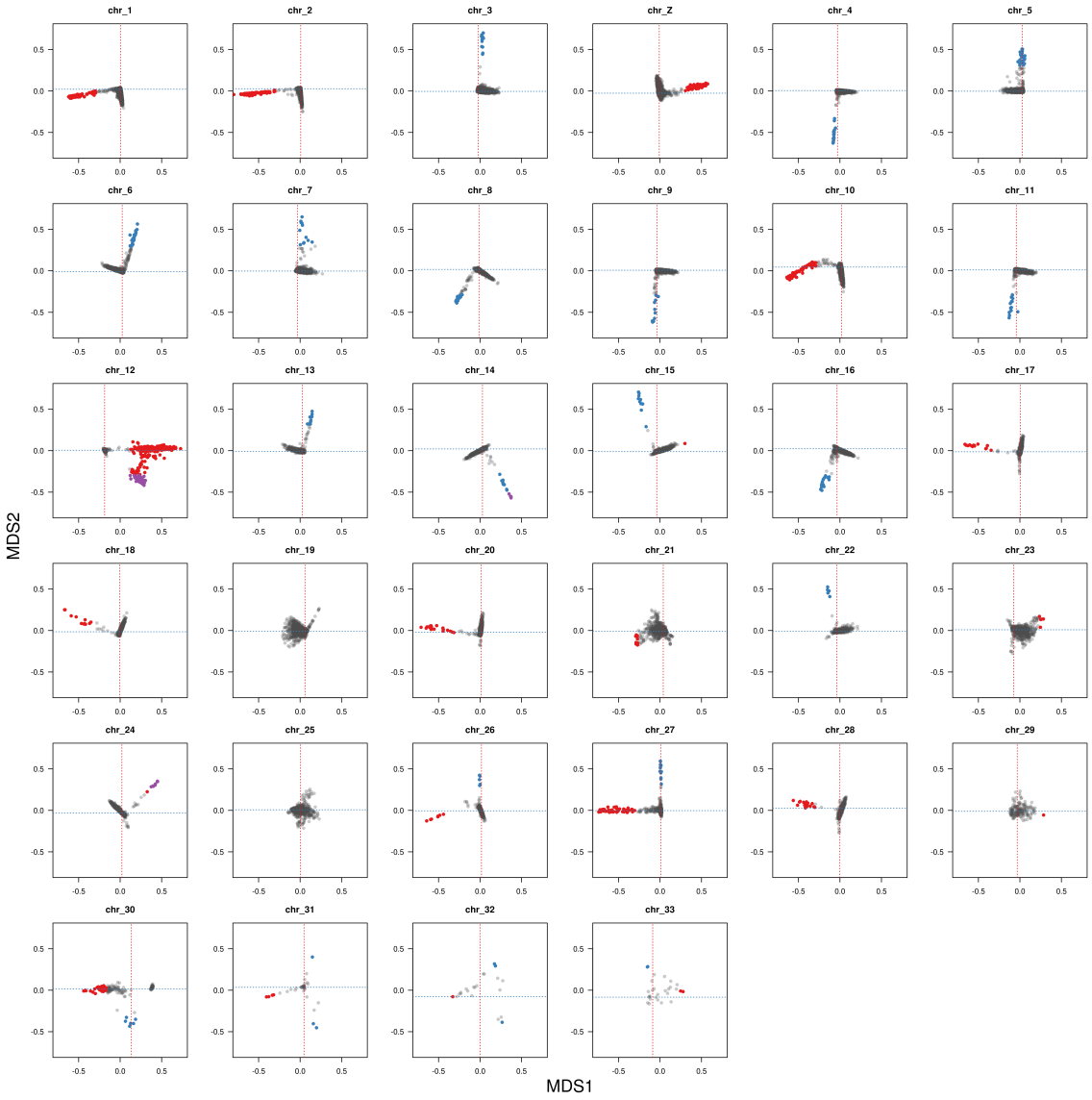
Supplementary Table 9: Permutation tests(n=10,000) d_{xy} and F_{ST} between populations at the class-2 genomic island on chromosome 21 in comparison to the chromosomal background by randomising the positions of the windows. P values of d_{xy} and F_{ST} are based on the rank of the observed value from right to left, asking whether observed value is significantly higher than expected from the null distribution.

pop1	pop2	p.val(d_{xy})	p.val(F_{ST})	p.val($\Delta\pi$)
Azores	Spain_Caz	0.0022	1e-04	2e-04
Azores	Belgium	0.0031	1e-04	2e-04
Belgium	CapeVerde	0.0028	1e-04	2e-04
Belgium	Spain_Caz	0.0023	0.5652	0.9756
Azores	CapeVerde	0.3264	1	2e-04
CapeVerde	Spain_Caz	0.001	2e-04	2e-04

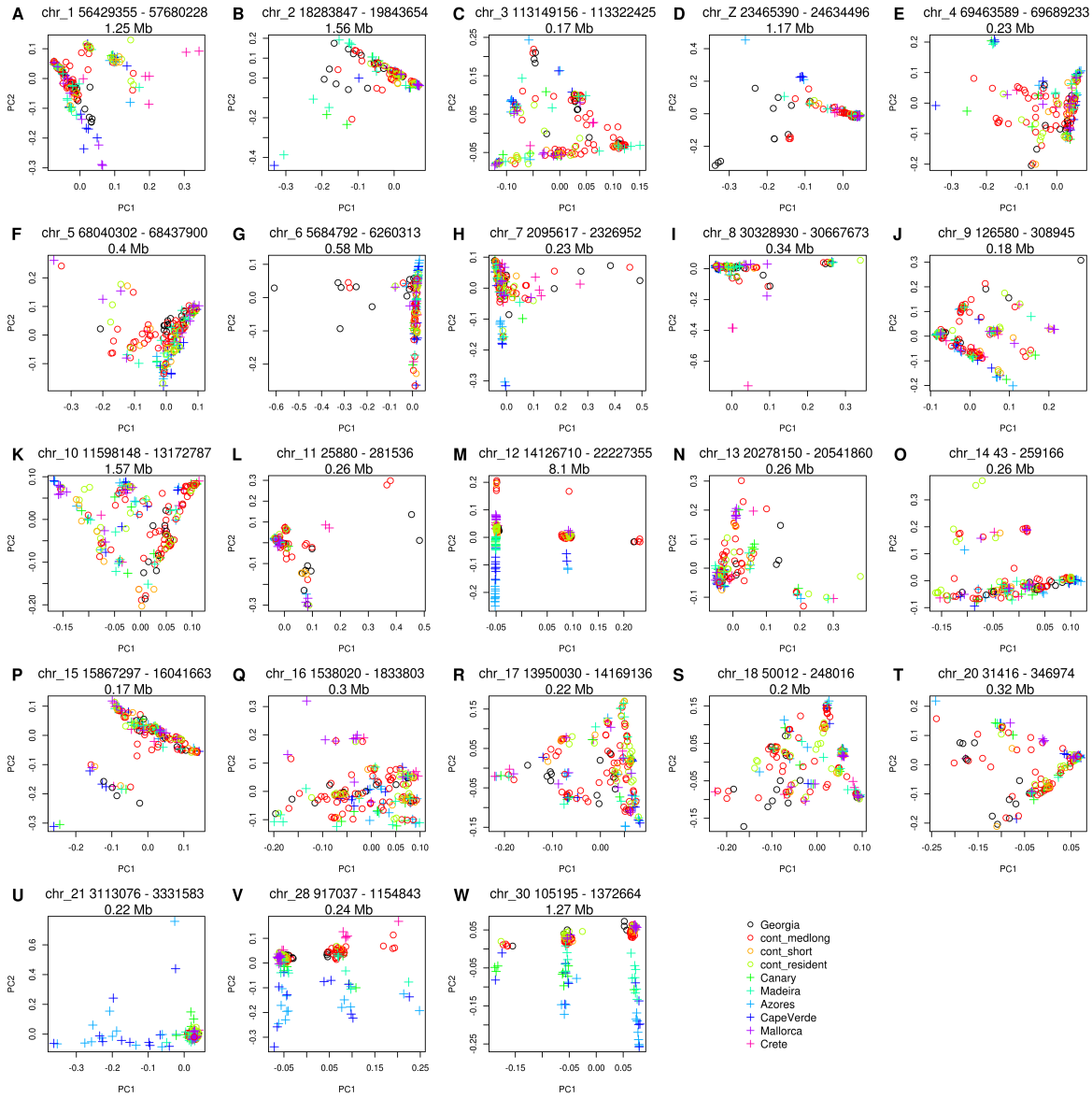
Supplementary figures



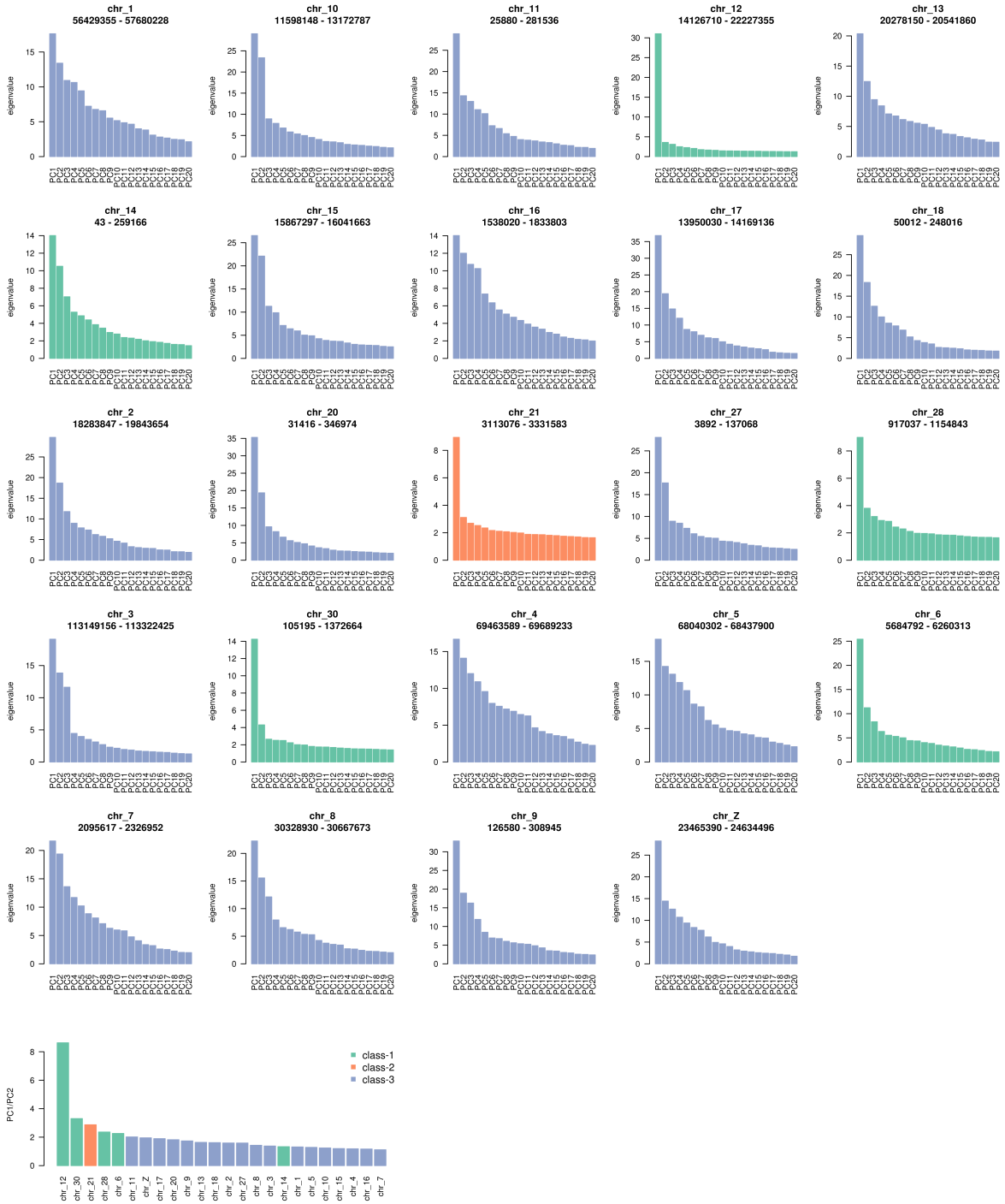
Supplementary Figure 1: Population structure analysed with whole-genome PCA. **A**. Whole-genome PCA. **B**. Scree plot.



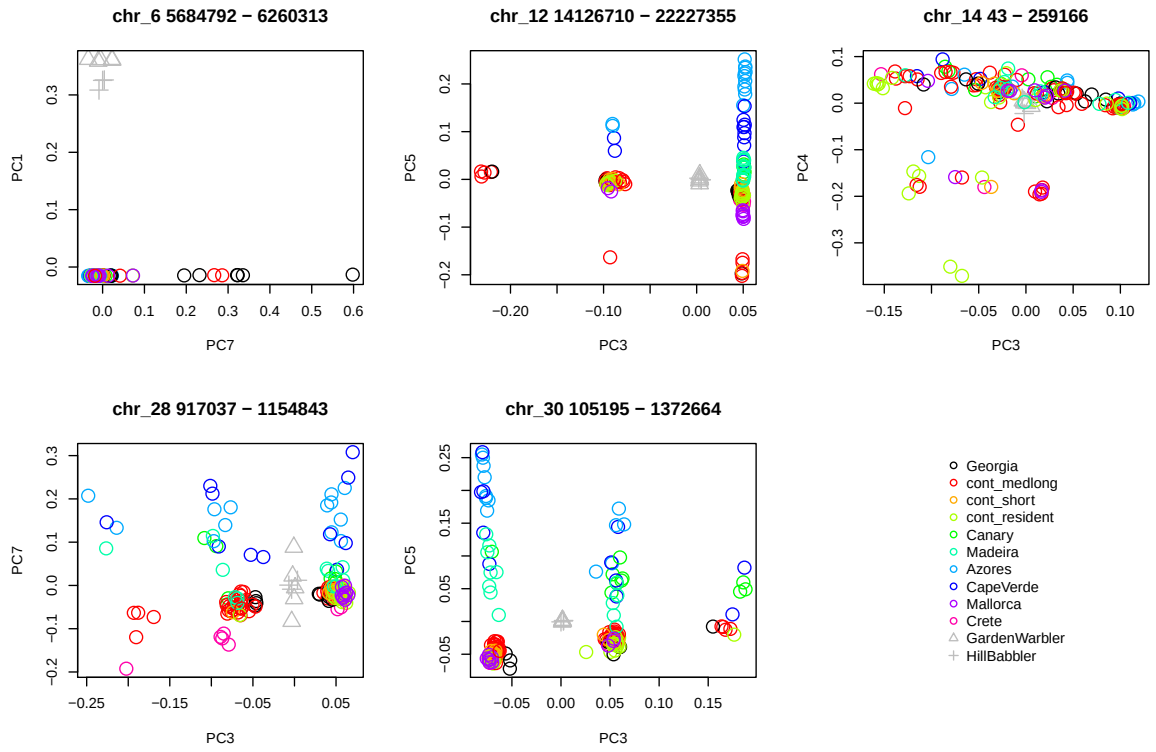
Supplementary Figure 2: MDS1 and MDS2 of local PCA. A data point represents a 1000-SNP genomic window. Red and blue dotted lines show modes of MDS1 and MDS2 values. Red and blue points show outlier windows based on MDS1 and MDS2 values. Purple points show outlier windows based both on MDS1 and MDS2 values.



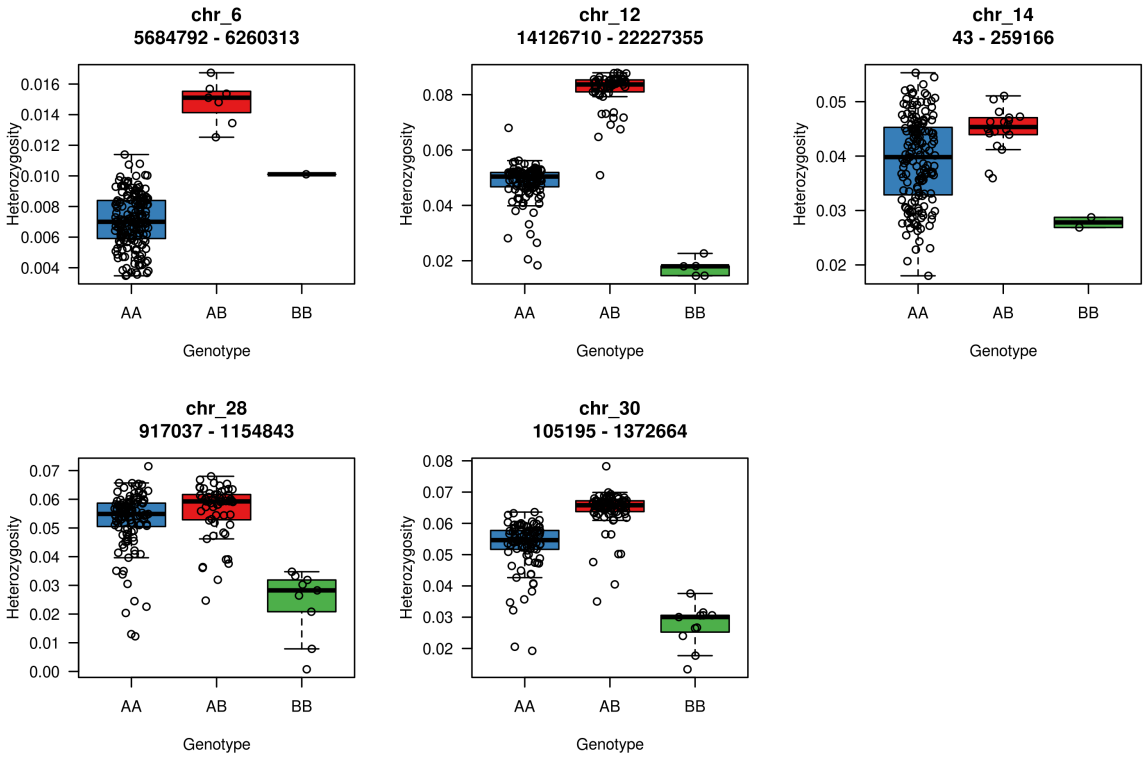
Supplementary Figure 3: PCA for each genomic island of deviated relatedness. A data point represents a blackcap individual. Colours represent populations. Class-1: **G, M, O, V, W**. Class-2: **U**. Class-3: **A-F, H-L, N, P-T**



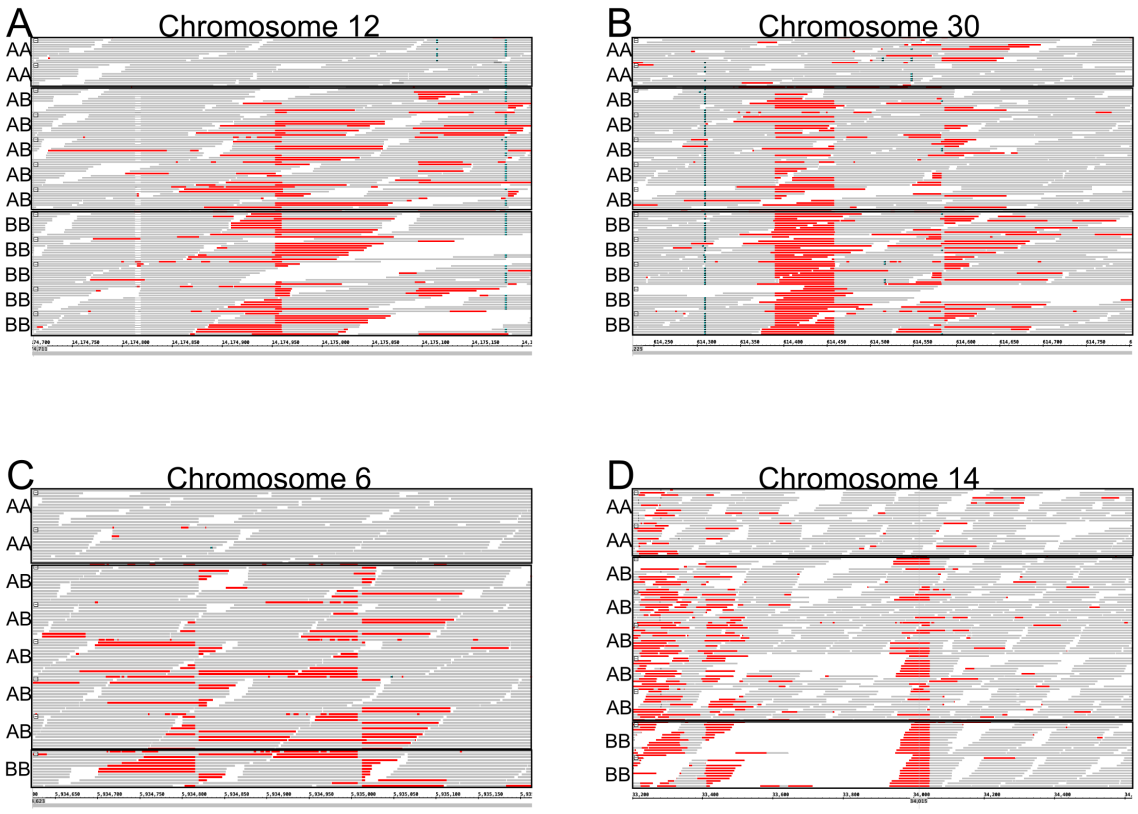
Supplementary Figure 4: Eigenvalues of PCA on genomic islands of deviated relatedness.



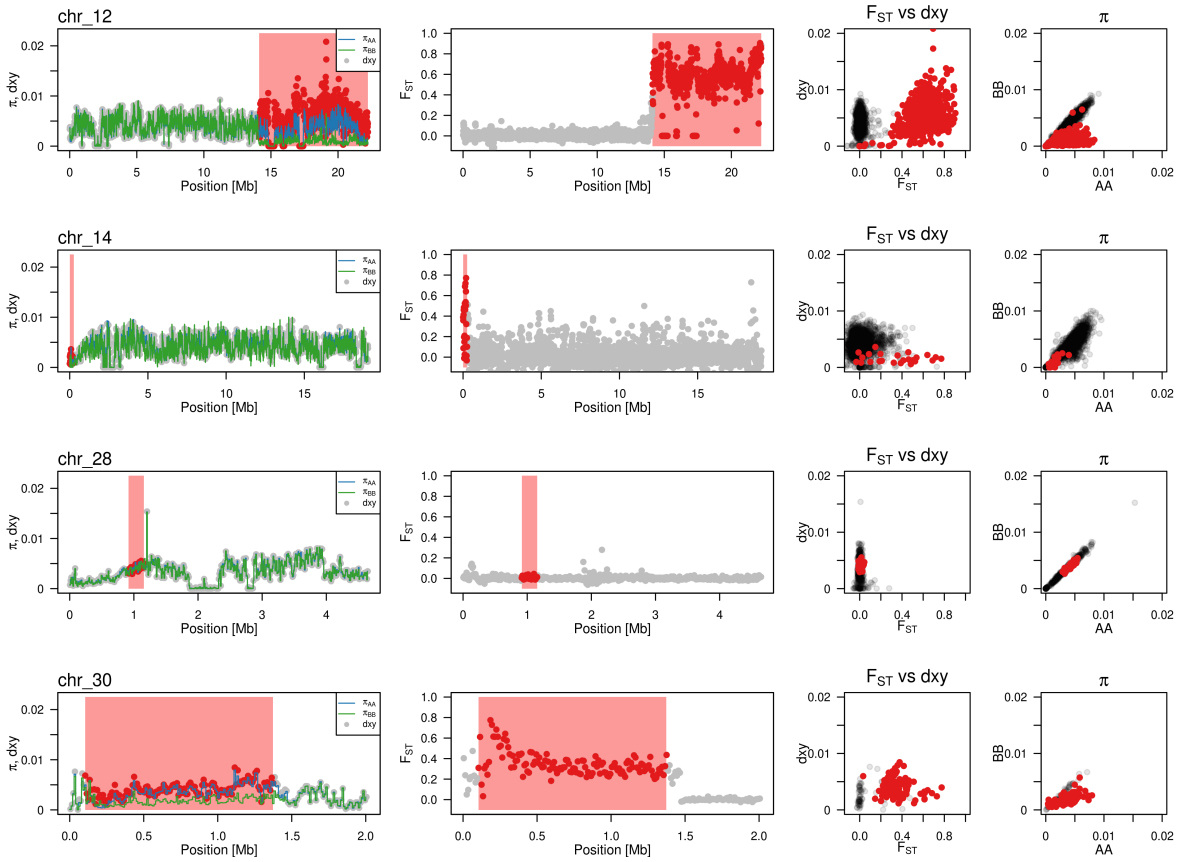
Supplementary Figure 5: PCA for class-1 genomic islands including outgroup samples.



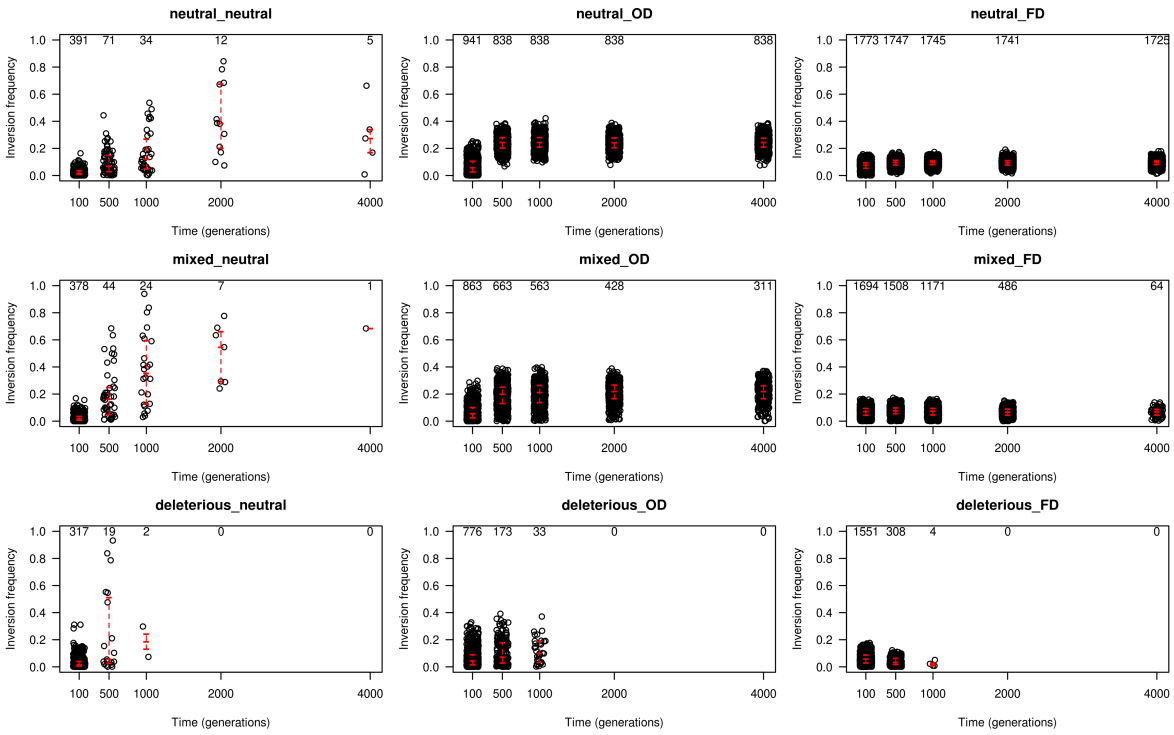
Supplementary Figure 6: Heterozygosity at class-1 genomic islands. A data point represents a blackcap individual with a particular putative genotype (AA, AB, BB) at the class-1 genomic island.



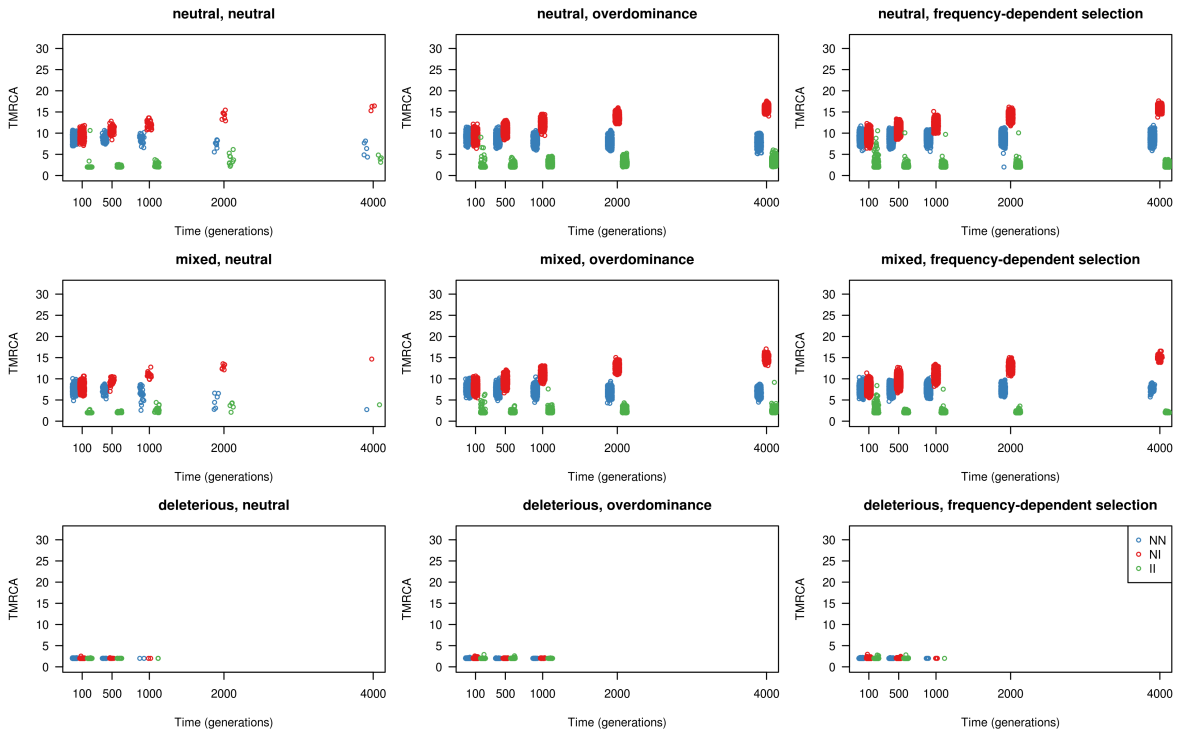
Supplementary Figure 7: Alignment visualised at for class-1 genomic islands with `IGB`. Soft-slipped reads are depicted with red segments. **A**. Chromosome 12 14,174,953 bp had a boundary of soft-clipped reads associated with PCA-based genotype. This position was outside of one of the outermost SNPs with perfect association with PCA-based genotype (14,191,392 bp and 22,245,485 bp), making it a strong candidate for a breakpoint. The soft-clipped reads were mapped to multiple scaffolds, suggesting the other breakpoint is missing from chromosome 12 of the reference assembly. **B**. Chromosome 30 614,390 and 614,583 bp had boundaries of soft-clipped reads associated with PCA-based genotype. This position was outside of the outermost SNPs with perfect association with PCA-based genotype (180,599 and 502,816 bp) and the mapping of the soft-clipped reads indicates B haplotype has a short duplicate. **C**. Chromosome 6 5,934,820 bp and 5,935,008 bp had boundaries of soft-clipped reads associated with PCA-based genotype. This position was between the two outermost SNPs with perfect association with PCA-based genotype (5,802,734 bp and 7,601,966 bp) thus unlikely represents a inversion breakpoint. **D**. Chromosome 14 34,047 bp had a boundary of soft-clipped reads associated with PCA-based genotype, with soft-clipped reads mapped to class-3 genomic islands of chr_8. This position was between the two outermost SNPs with perfect association with PCA-based genotypes (9,016 bp and 167,167 bp) thus unlikely represents a inversion breakpoints No SNPs or soft-clipped reads alignment associated with PCA-based genotype were found on chromosome 28.



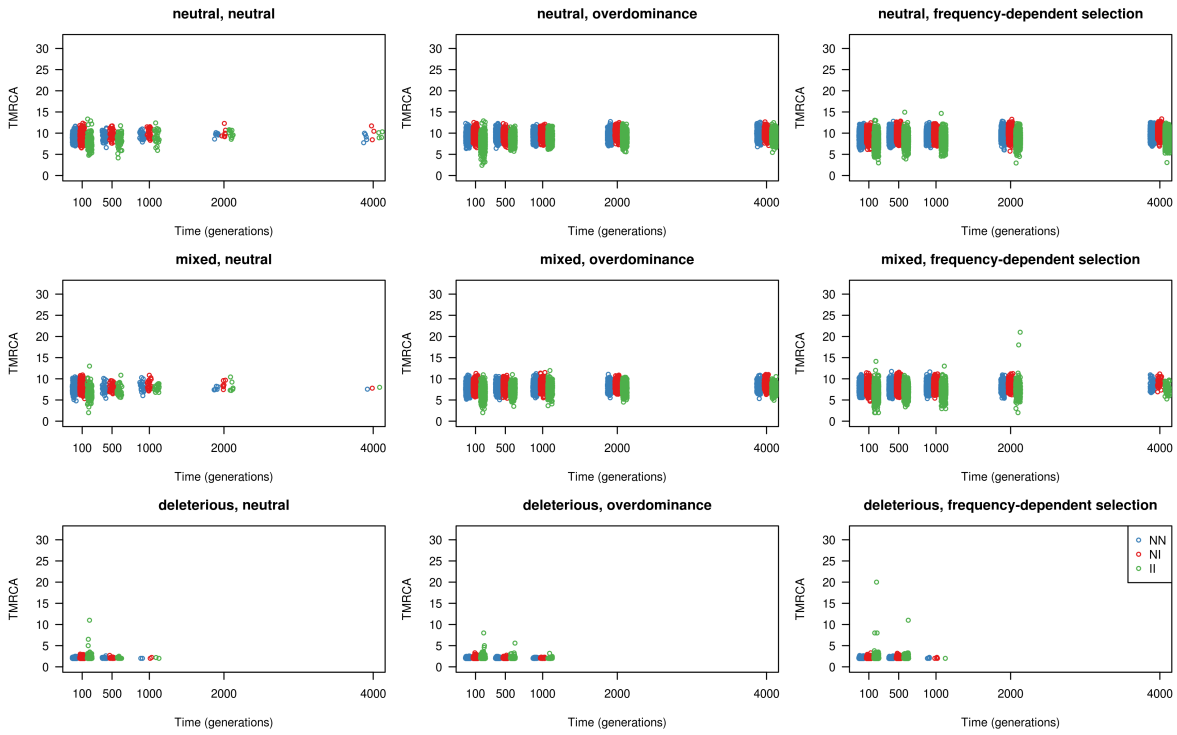
Supplementary Figure 8: π, d_{xy}, and F_{ST} at class-1 genomic islands. π within AA and BB genotypes of the class-1 genomic islands and d_{xy} and F_{ST} between the AA and BB were calculated in a 10-kb sliding window.



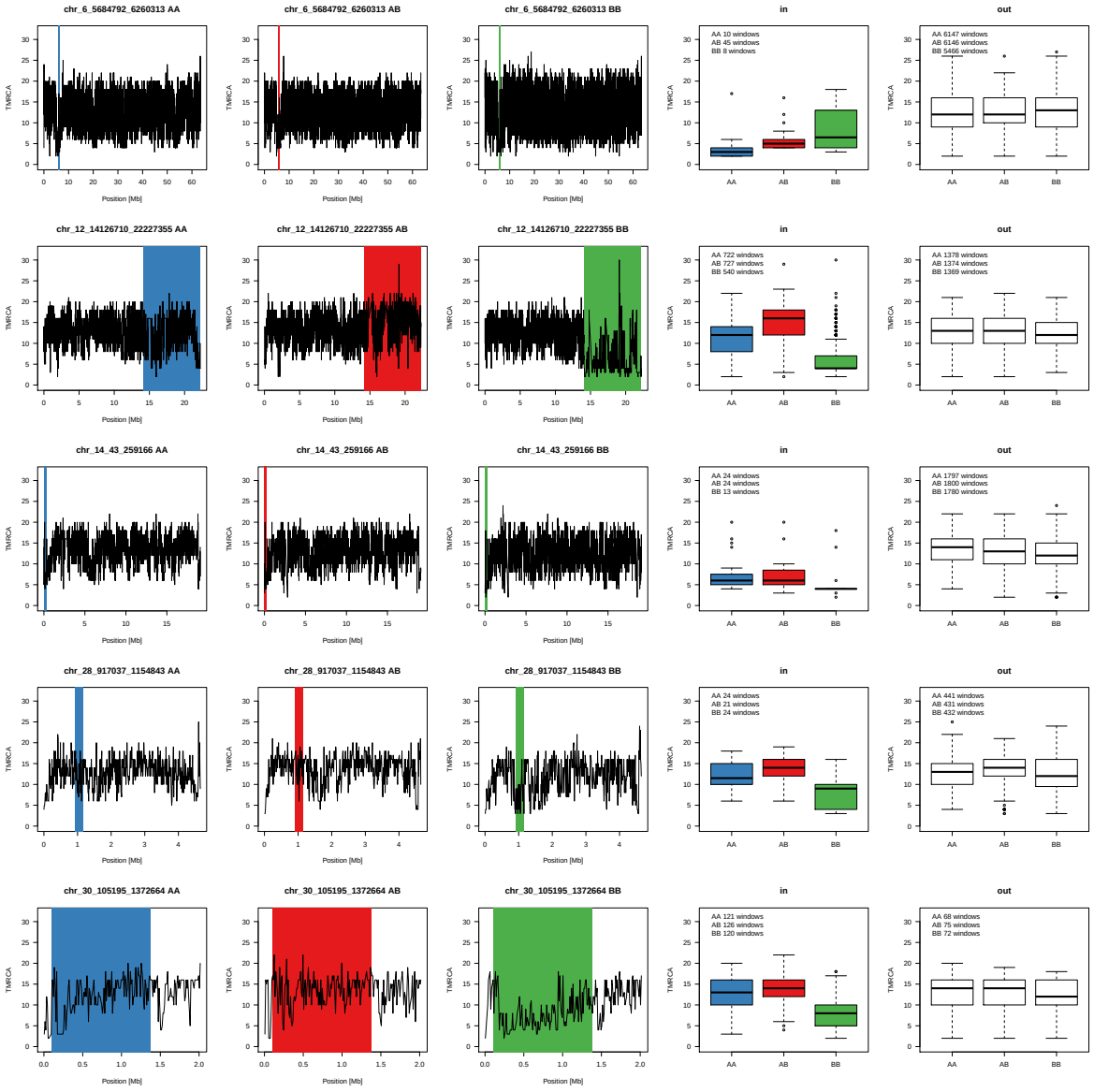
Supplementary Figure 9: Simulated inversion frequency over time. Each data point depicts inversion frequency of a simulation at a time point. Numbers at the top show how many simulations out of 10,000 still have polymorphic inversion in the population at the time point. Ticks of a vertical red dotted line depict 25, 50, and 75 percentiles of inversion frequencies.



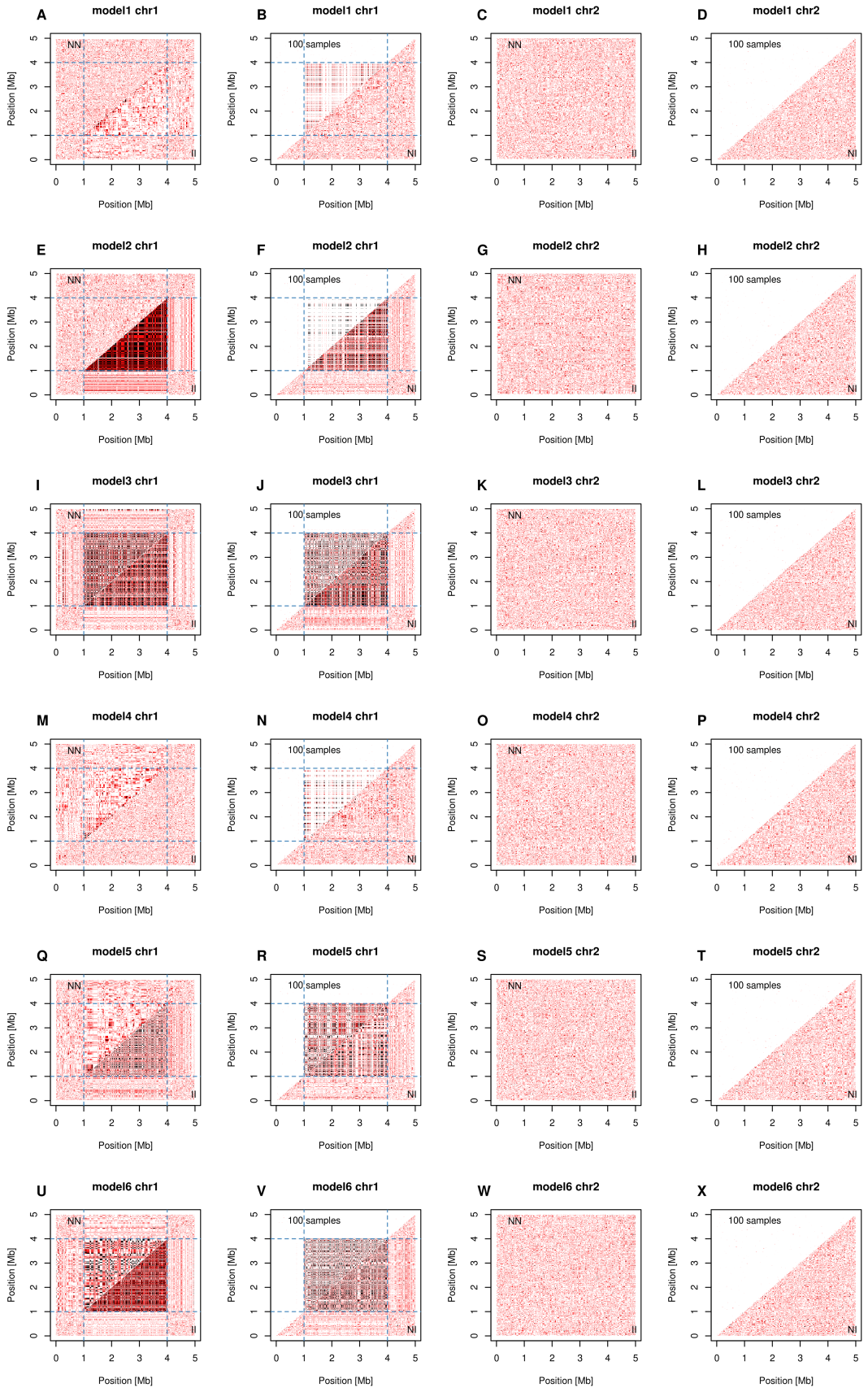
Supplementary Figure 10: Coalescent time within inversions over time simulated by `SLiM` and inferred by `MSMC2-decode`. One data point depicts mean discretised coalescent time within the inversion locus for four pairs of NN (blue), NI (red), or II (green) sequences averages over windows.



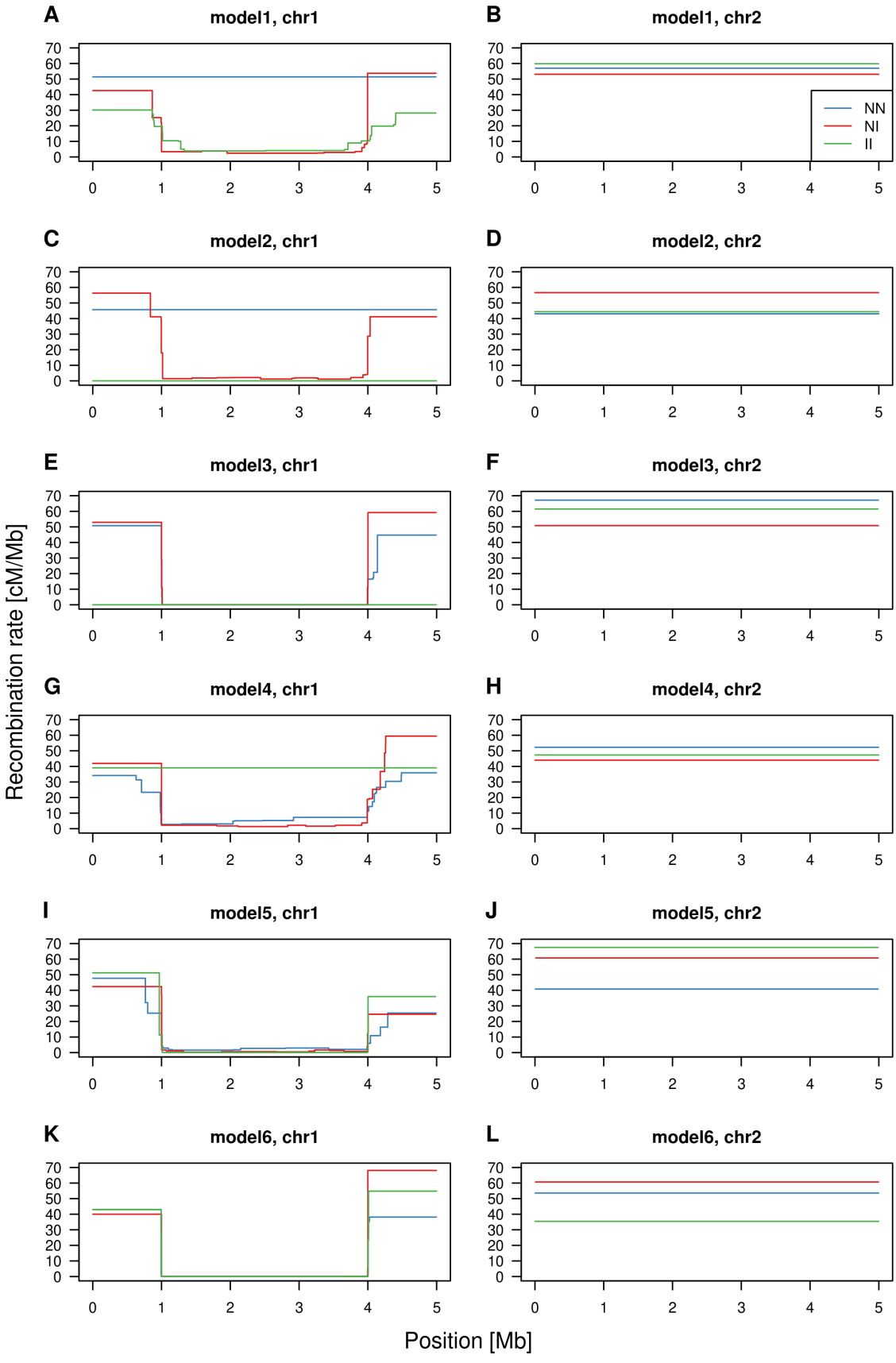
Supplementary Figure 11: Coalescent time outside inversions over time simulated by `SLiM` and inferred by `MSMC2-decode`. One data point depicts mean discretised coalescent time within the inversion locus for four pairs of NN (blue), NI (red), or II (green) sequences averaged over windows.



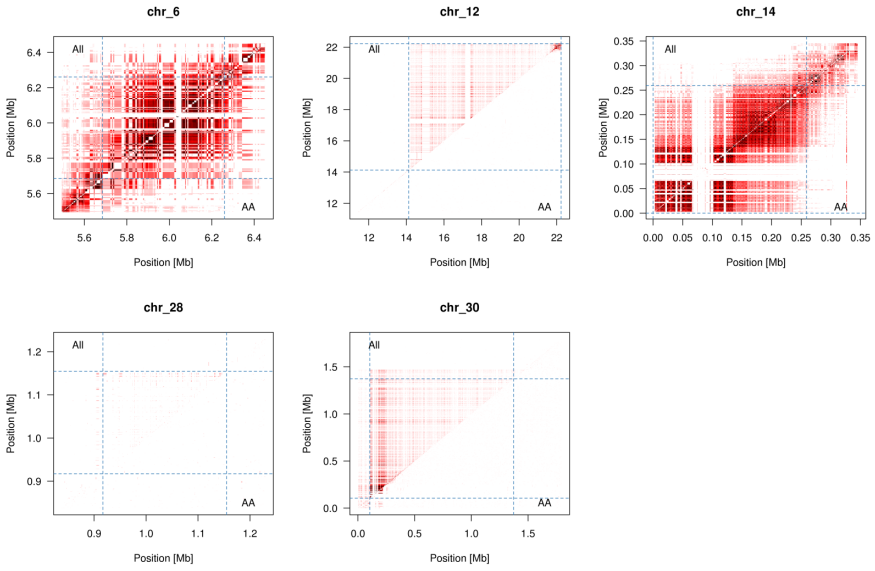
Supplementary Figure 12: Coalescence time at class-1 genomic islands. Coalescence time between A-A, A-B, and B-B haplotypes were estimated with `MSMC2-decode`, using up to 4 samples for each genotype, one sample per run. Discretised coalescence time were averaged over samples with the same genotype at the same genomic windows. Coloured box plots show coalescence time within the class-1 genomic islands, and uncoloured box plots show coalescence time in other part of the same chromosomes.



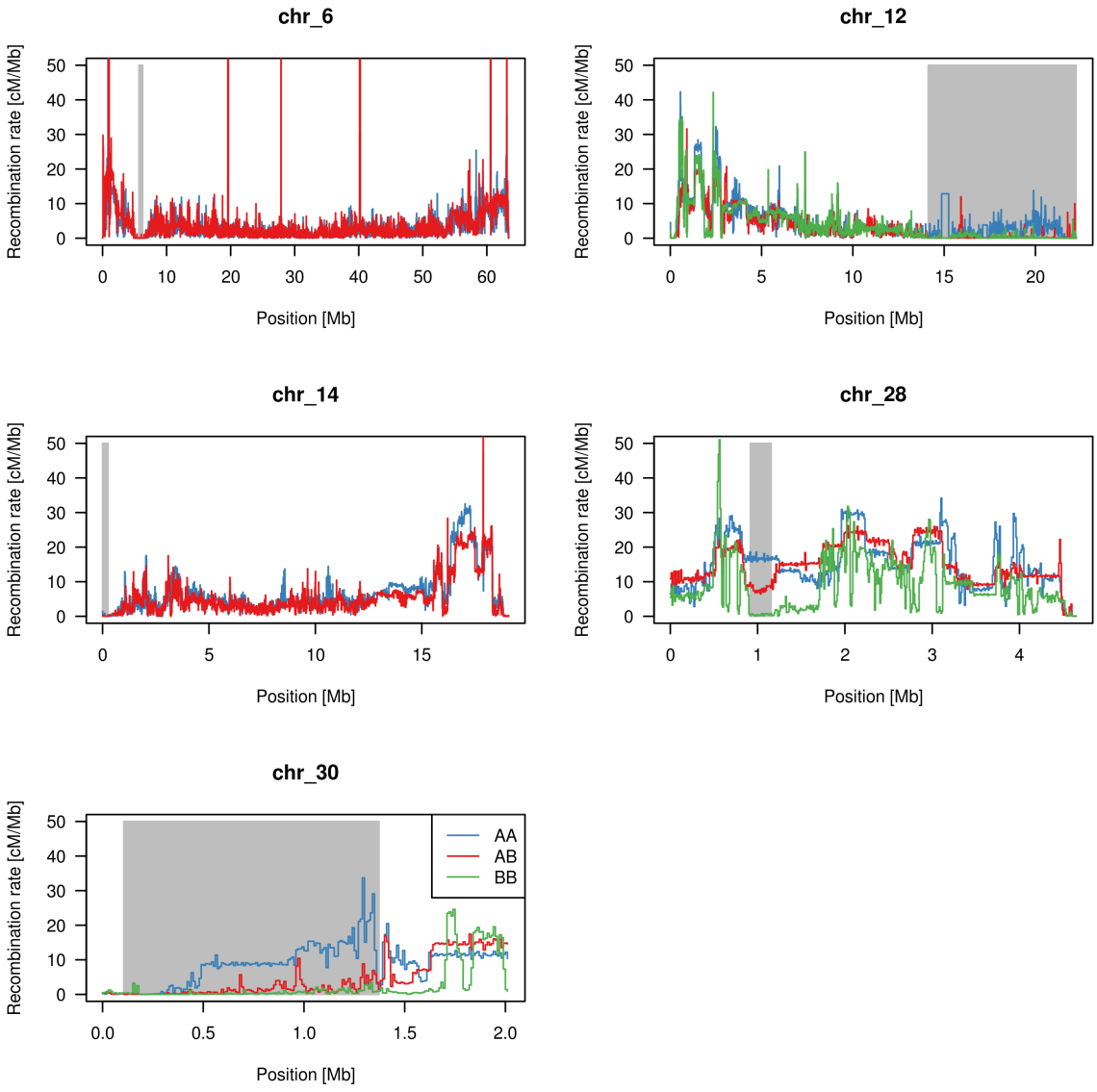
Supplementary Figure 13: Linkage disequelibrium in simulated inversion. Chromosome 1 includes a polymorphic inversion, whose boundaries are depicted by blue dotted lines. Chromosome 2 does not have an inversion.



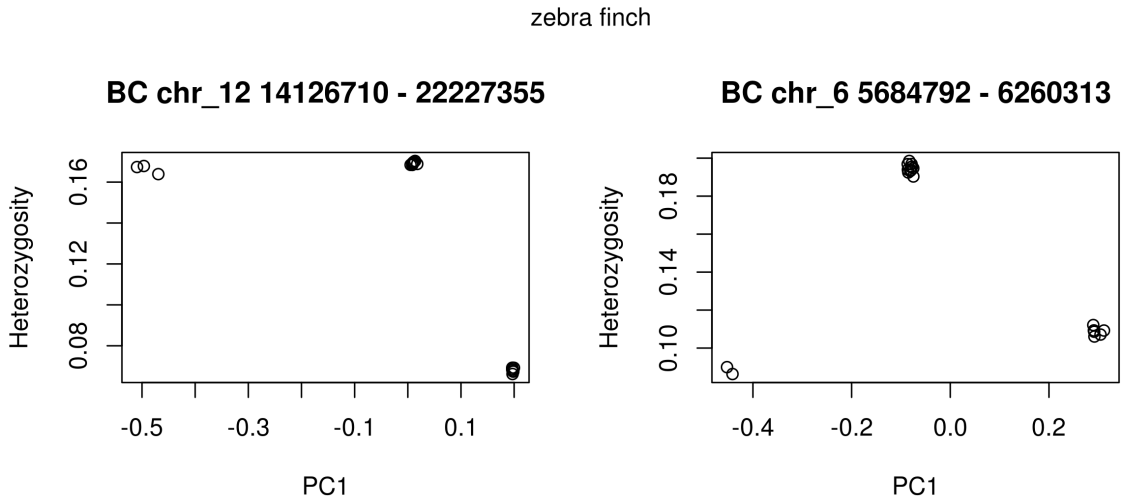
Supplementary Figure 14: Recombination rate inference by `Pyrho` performed on a polymorphic inversion simulated in `SLiM`. Chromosome 1 includes a polymorphic inversion, whose breakpoints are at 1 and 4 Mb. Chromosome 2 does not have an inversion. Three colours depict recombination rate inference using samples with different inversion genotypes.



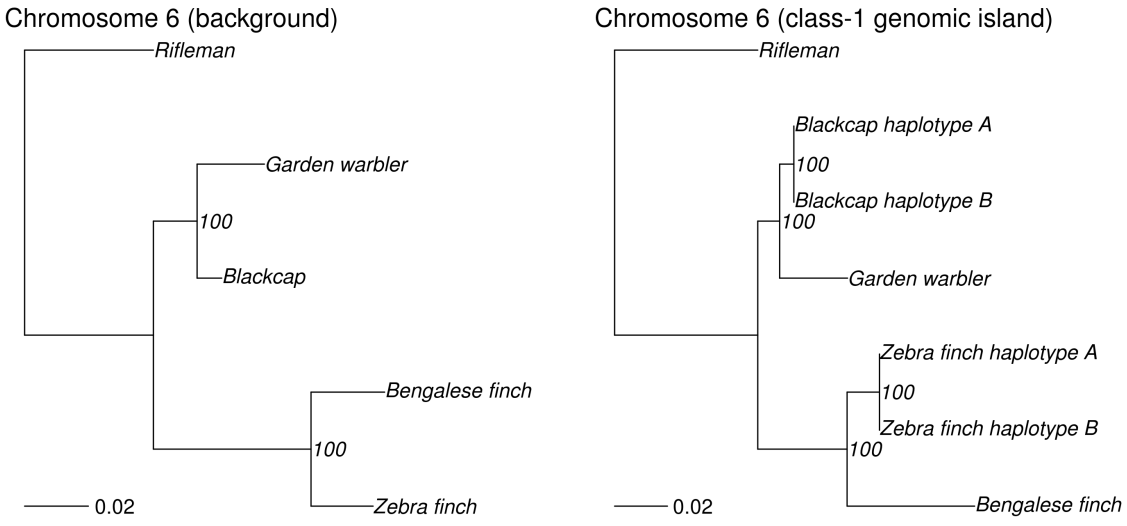
Supplementary Figure 15: Linkage disequelibrium at class-1 genomic islands. The topleft diagonal shows LD calcualted using all blackcap samples. The bottomright diagonal shows LD calculated using blackcap samples homozygous for the major haplotype at the class-1 genomic islands.



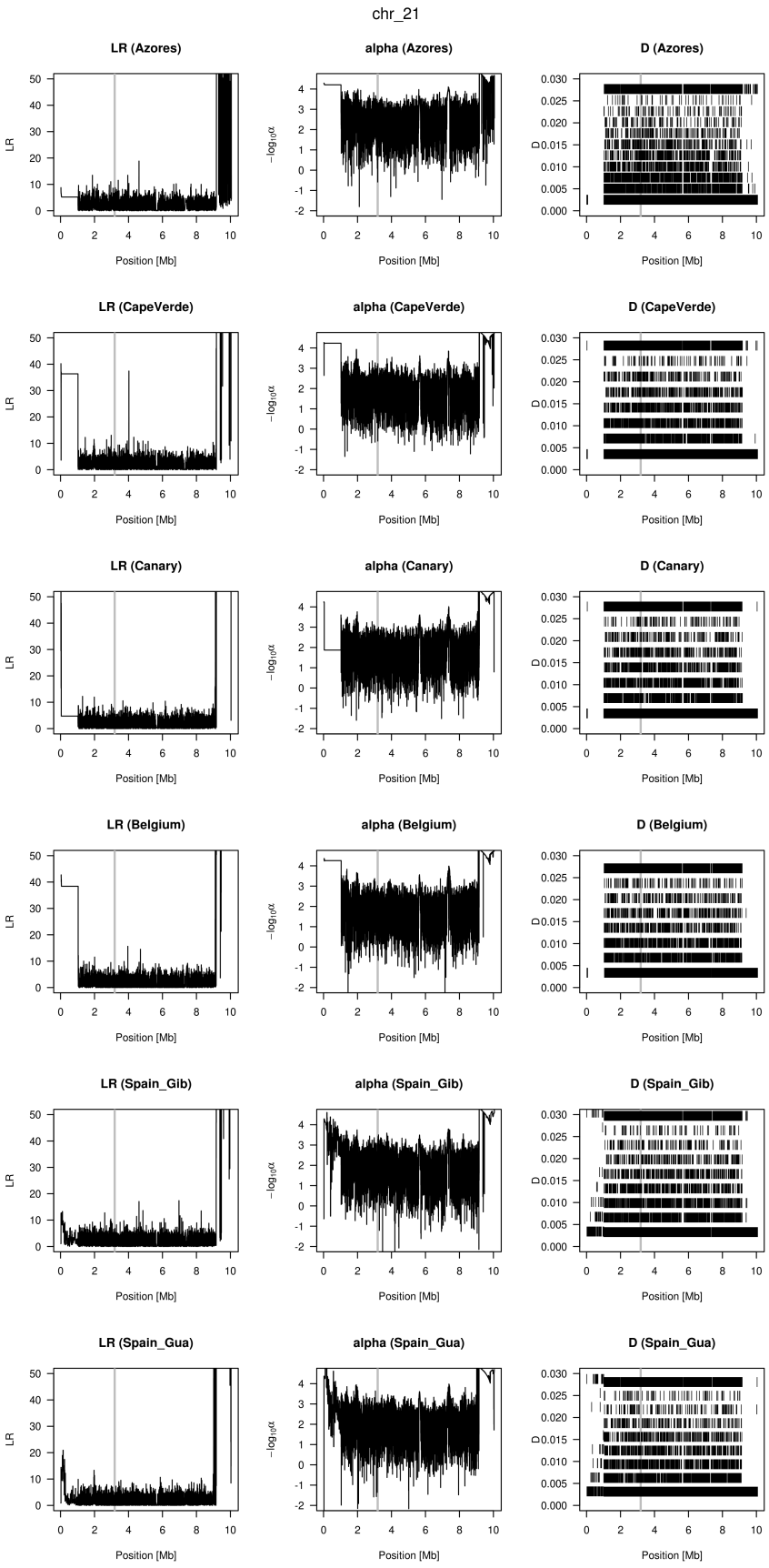
Supplementary Figure 16: Recombination maps inferred with AA, AB, and BB samples in class-1 genomic islands.



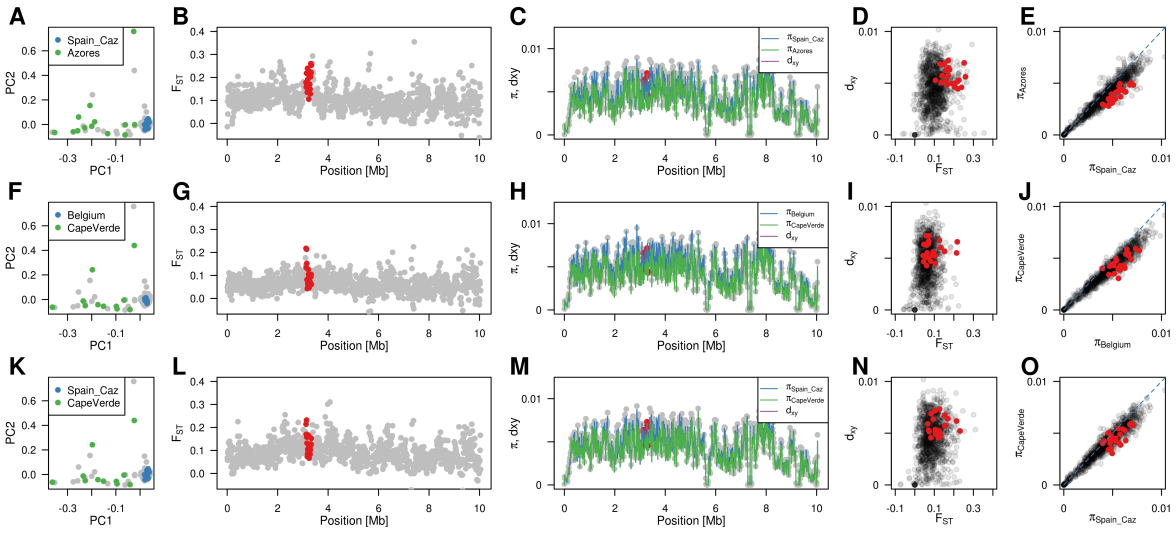
Supplementary Figure 17: PCA and heterozygosity of the zebra finch genomic regions syntenic to the blackcap class-1 genomic islands. Regions syntenic to class-1 genomic islands on the blackcap chromosomes 12 and 6 show patterns supporting inversion polymorphism in the samples.



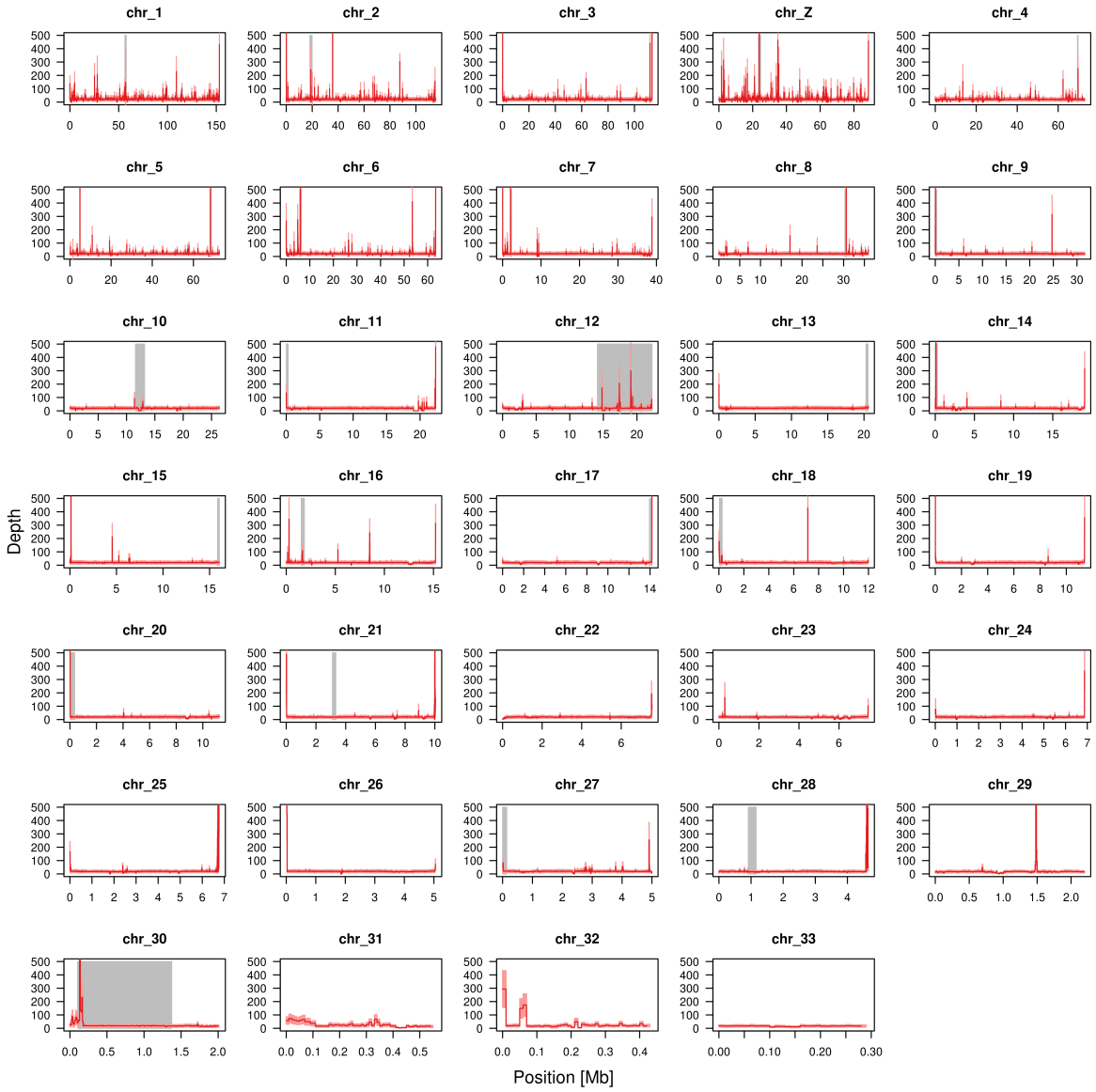
Supplementary Figure 18: Phylogeny at the class-1 genomic island on blackcap chromosome 6.



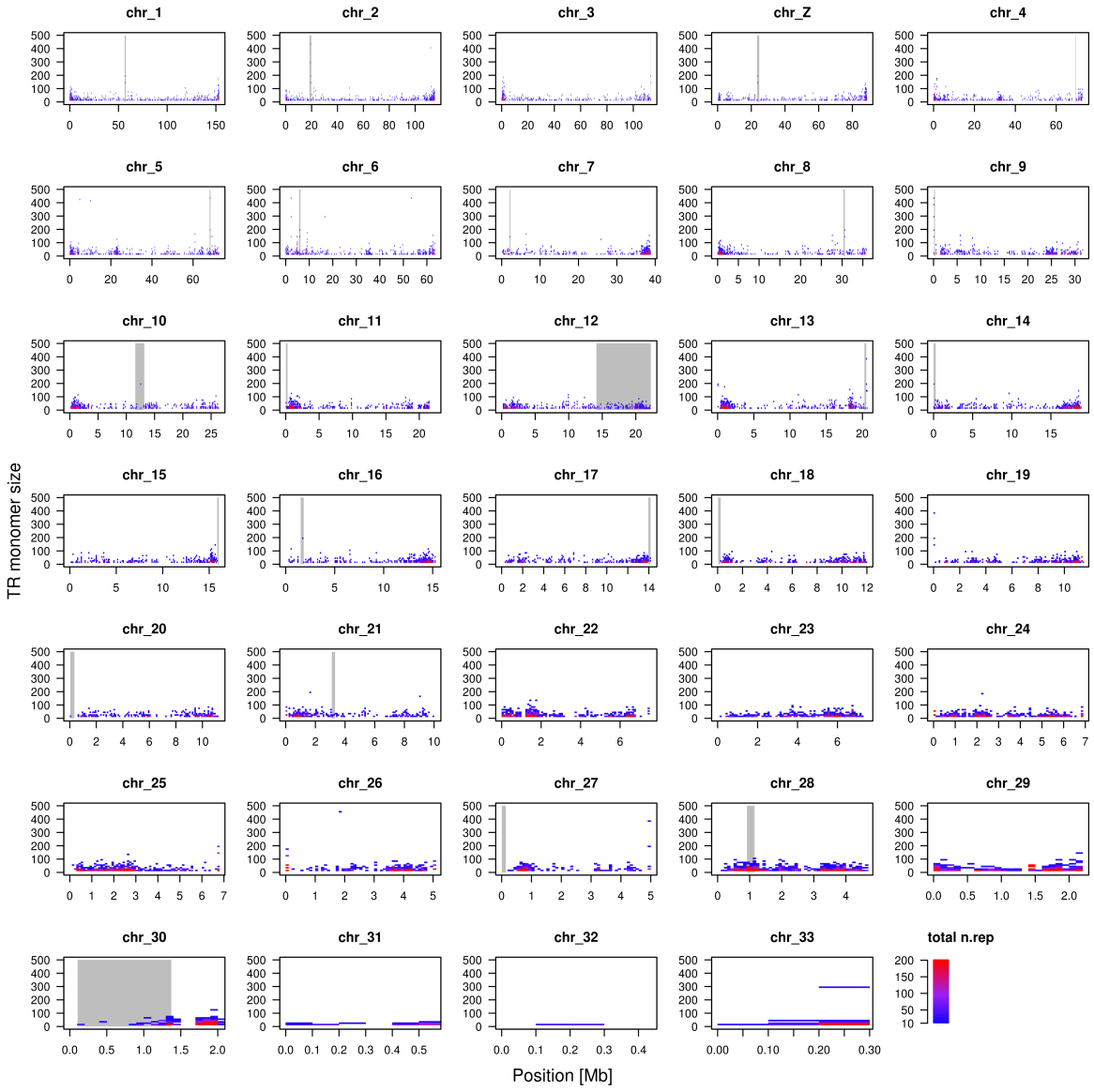
Supplementary Figure 19: `VolcanoFinder` on chromosome 21, testing for introgression into Azores and Cape Verde populations. Gray shade show the position of the class-2 genomic island on chromosome 21. Left column: log likelihood ratio. Middle column: $-ln(\alpha)$. Right column: inferred divergence D between donor and recipient populations.



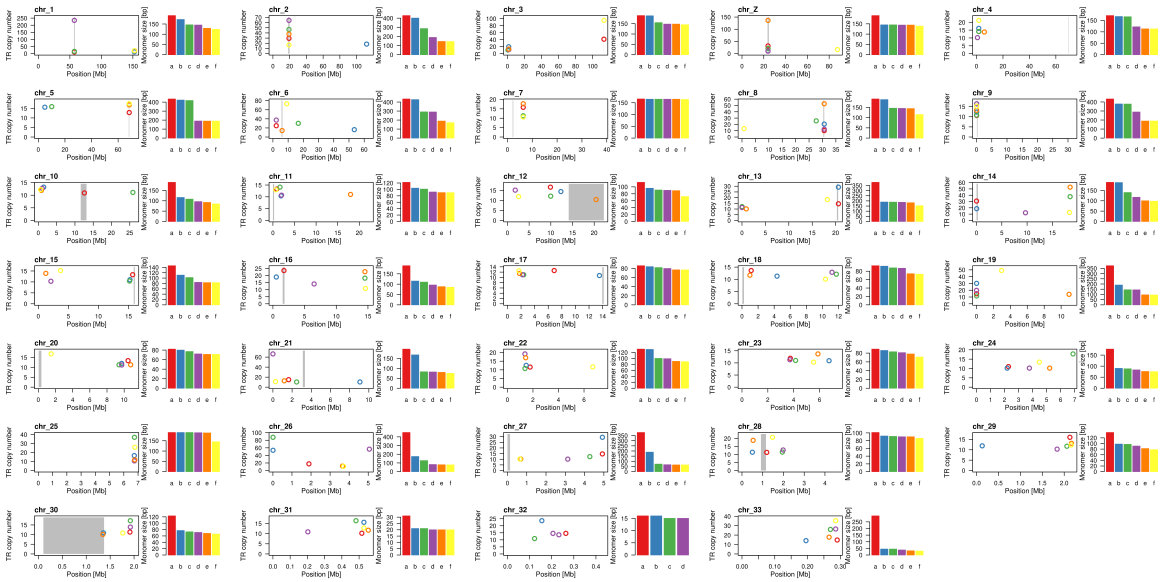
Supplementary Figure 20: F_{ST}, d_{xy}, and π in the class-2 genomic islands. Results of permutation tests of π, F_{ST}, d_{XY}, and $\Delta\pi$ comparing within and outside the class-2 genomic island are in Supplementary Tables. 8, 9.



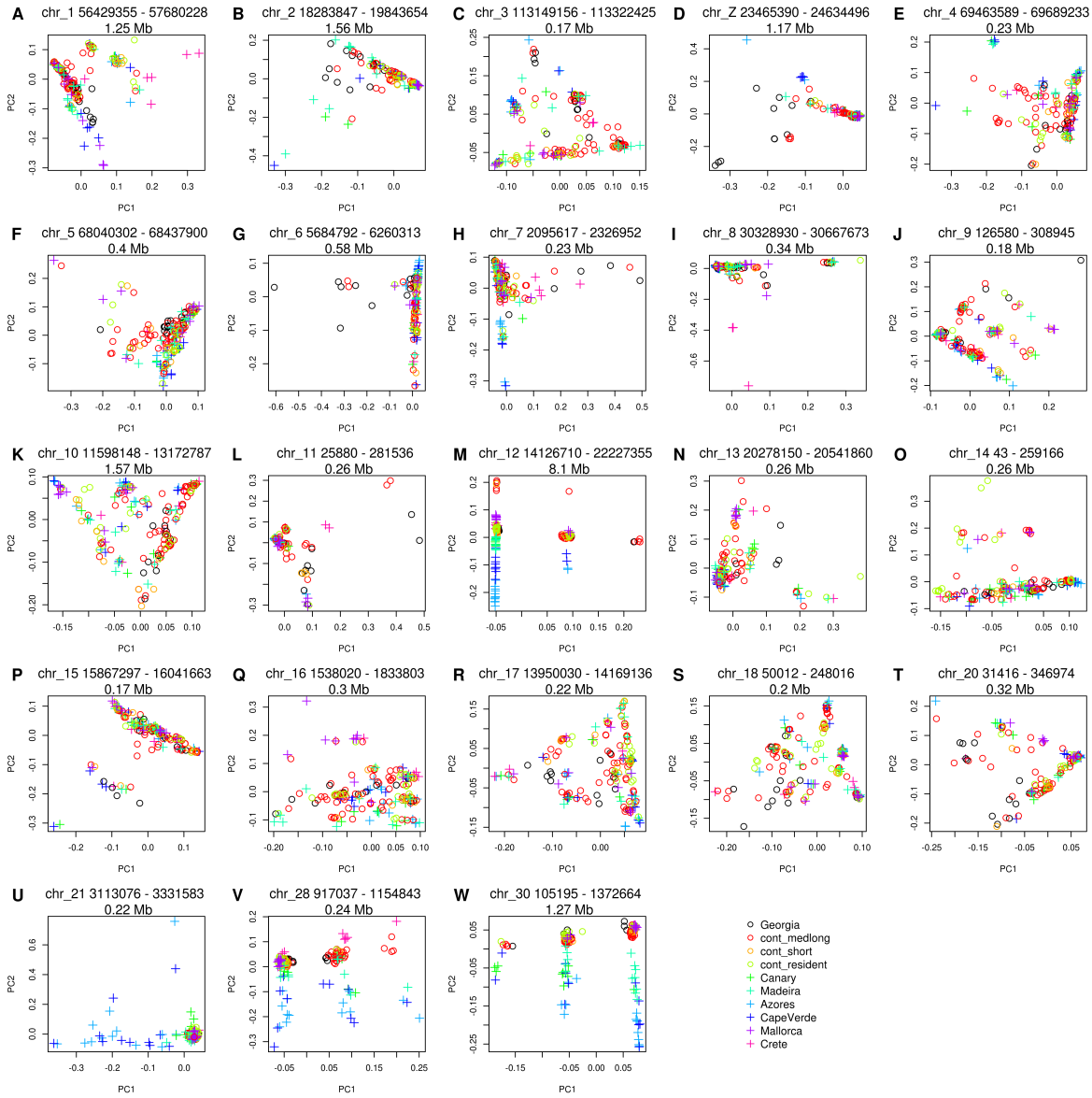
Supplementary Figure 21: Read depth in whole-genome resequencing of blackcaps. Red line: read depth averaged over all blackcap samples in 10-kb genomic sliding windows. Pink shade: mean read depth ± standard deviation. Gray shade: genomic islands of deviated relatedness patterns.



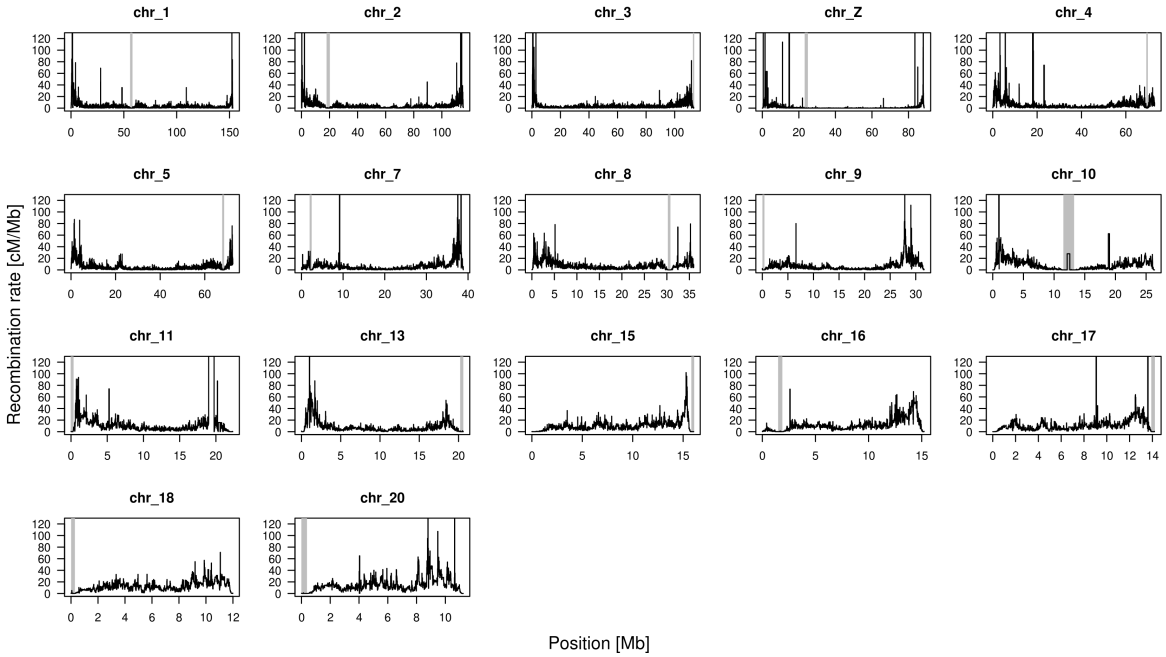
Supplementary Figure 22: Genomic distribution of tandem repeats. Gray shade: genomic islands of deviated relatedness patterns.



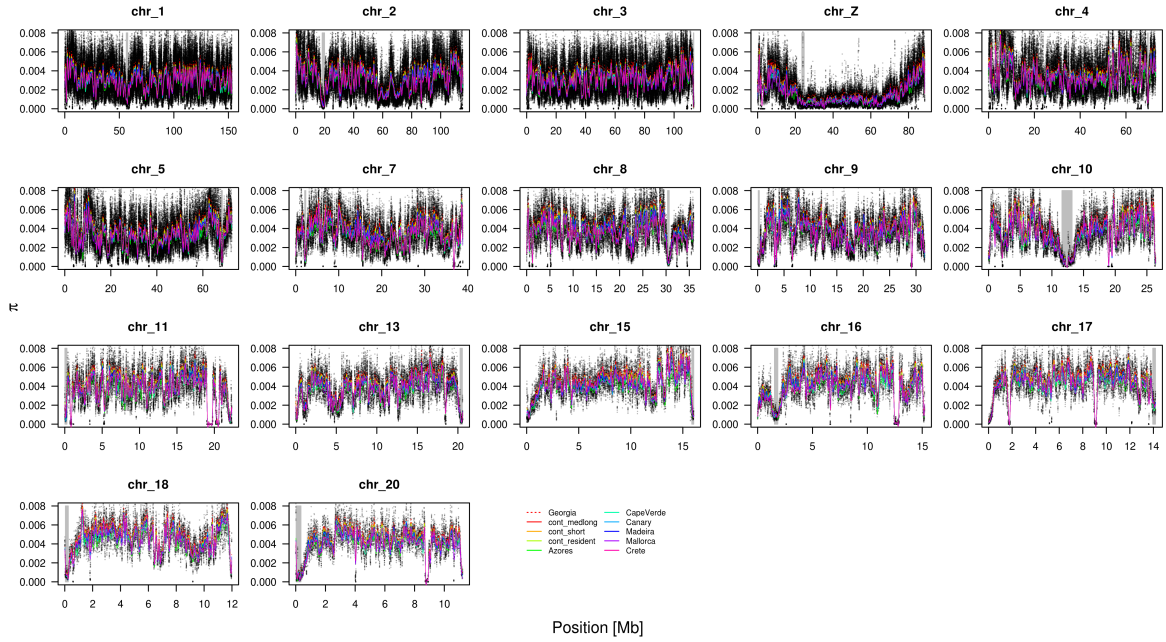
Supplementary Figure 23: Genomic distribution of tandem repeats with long repeat unit. In each chromosome, 6 tandem repeats with the longest repeat units are shown. Gray shade: genomic islands of deviated relatedness patterns.



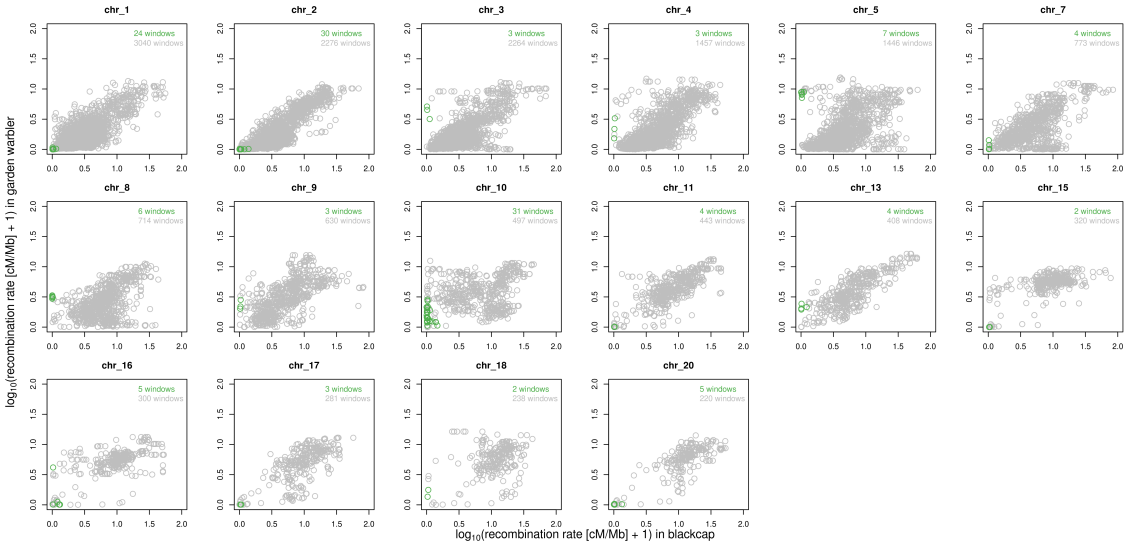
Supplementary Figure 24: PCA for each genomic island of deviated relatedness patterns with TRs masked. A data point represents a blackcap individual. Colours represent populations. Class-1: **G, M, O, V, W**. Class-2: **U**. Class-3: **A-F, H-L, N, P-T**



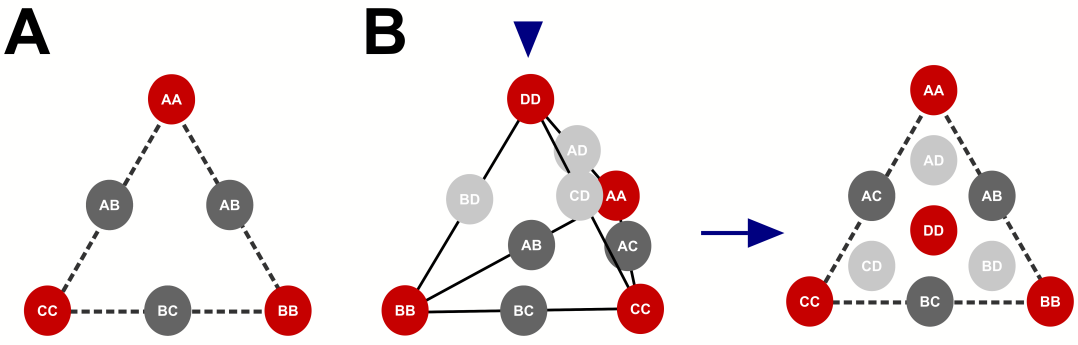
Supplementary Figure 25: Recombination rate along chromosomes with class-3 genomic islands.



Supplementary Figure 26: π along chromosomes with class-3 genomic islands.



Supplementary Figure 27: Recombination rate in blackcaps and garden warblers on chromosomes with class-3 genomic islands.



Supplementary Figure 28: Triangular spread of samples at a multi-allelic site.