Supplementary Material

Lotterhos, Yeaman, Degner, Aitken, and Hodgins

Contents

List of Figures 1

1. Supplementary Figures 2

List of Figures

1 S1 XTX histogram for outliers 2

2 S2 Enlarged Multi network 3

3 S3 Enlarged Aridity network 4

4 S4 Enlarged Freezing network 5

5 S5 Enlarged Geography network 6

6 S6 Structure-corrected heatmap 7

7 S7 Linkage disequilibrium heatmap 8

8 S8 Recombination heatmap, clustered by recombination rates 9

9 S9 Loadings of environments onto PC axes 10

10 S10 Outliers on PC axes 11

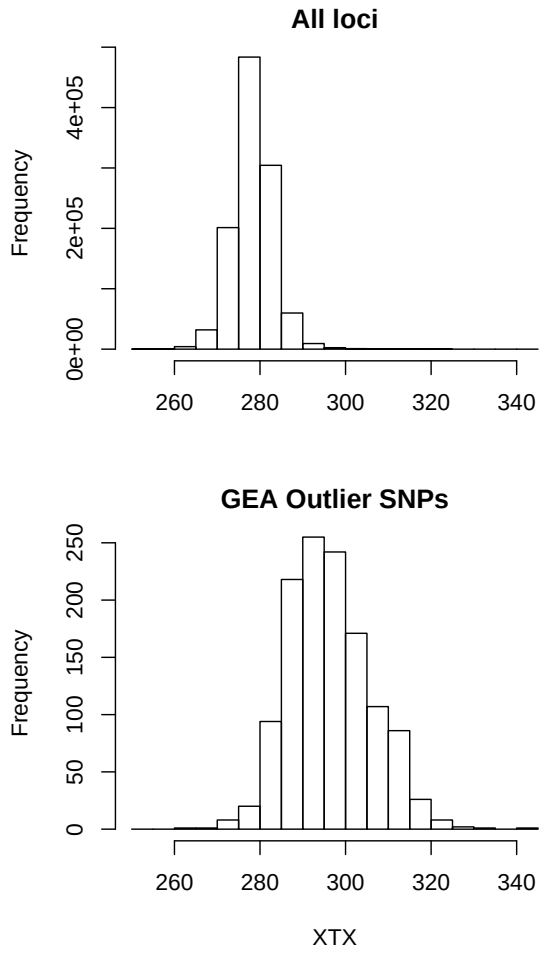
11 S11 SNP annotations and genomic features 12

12 S12 Error rates in simulations 13

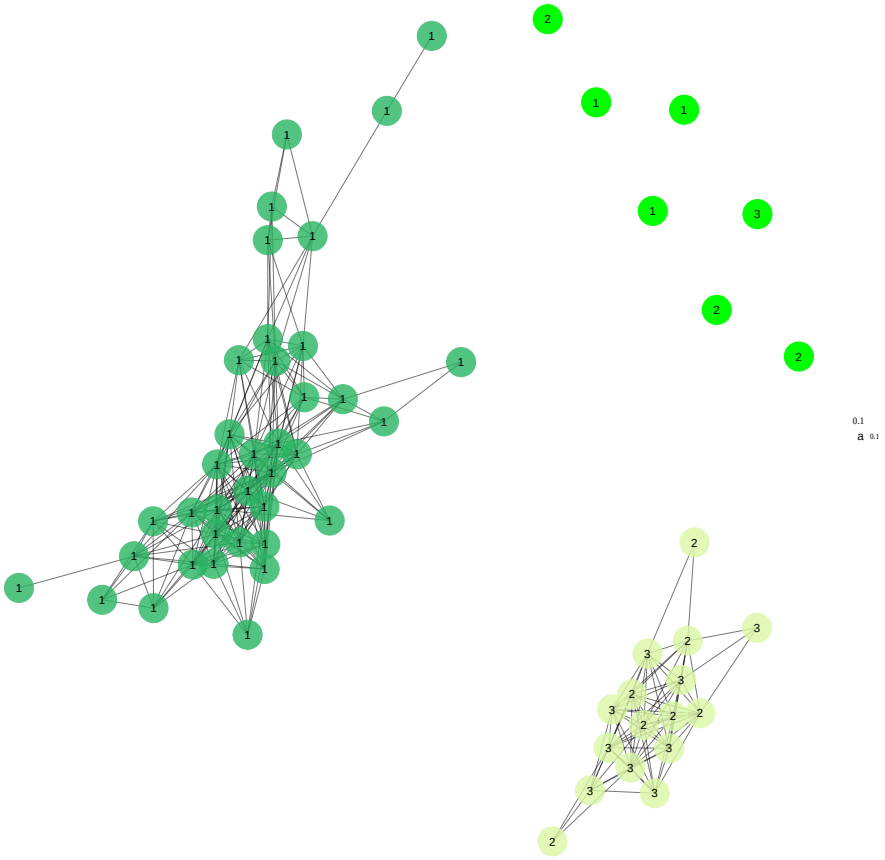
13 S13 Pairwise distances among loci as a function of selection for simulated data 14

14 S14 More examples of networks from simulations 15

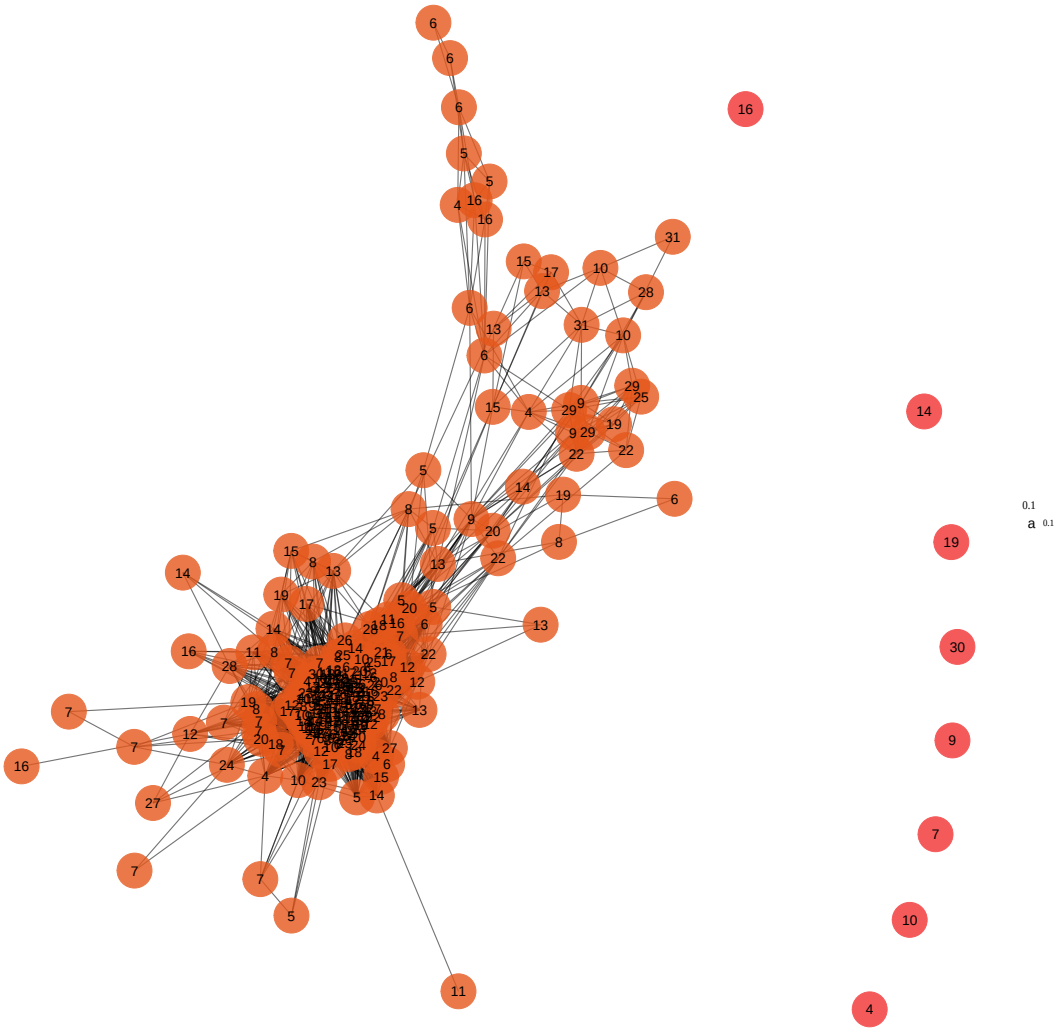
1. Supplementary Figures



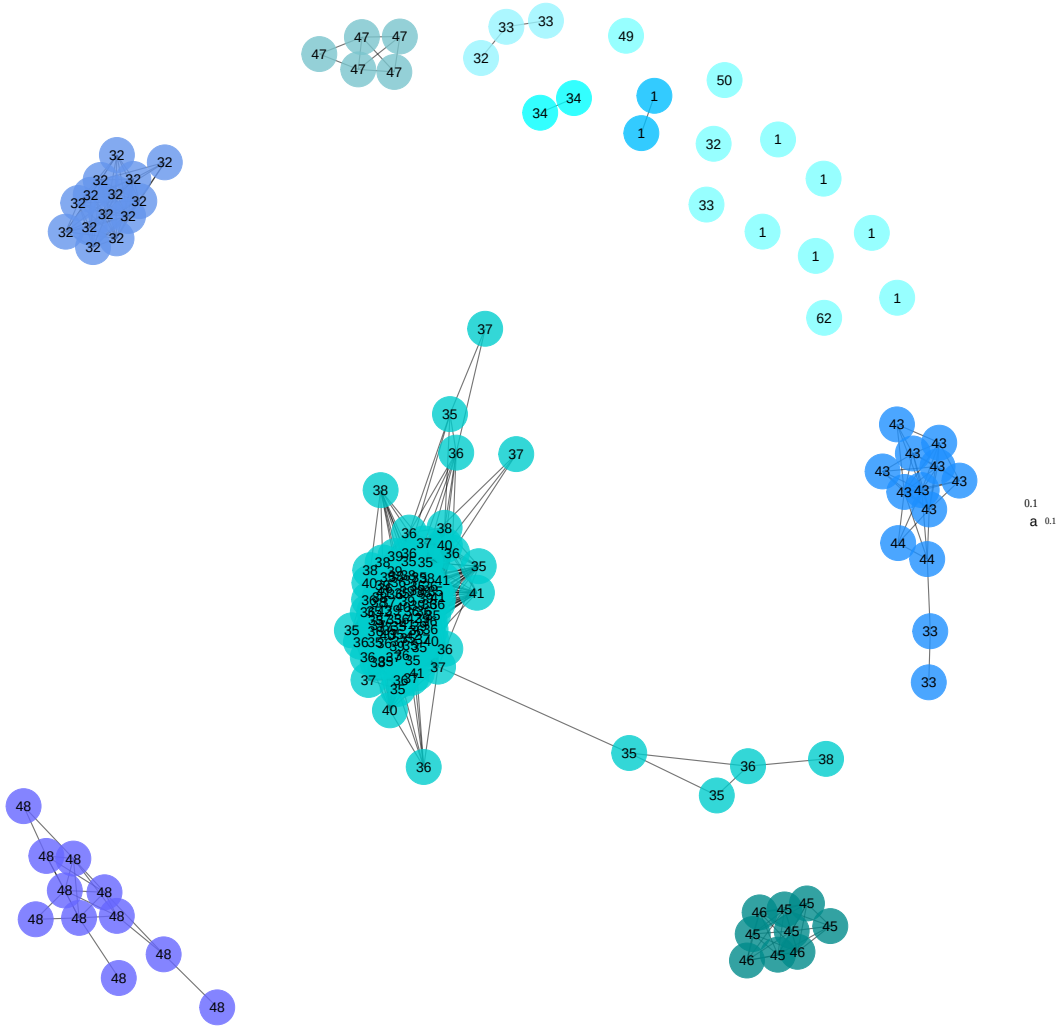
Supplementary Figure 1. Histogram of $X^T X$ estimated from Bayenv2 for all SNPs (top) and for top candidate SNPs (bottom).



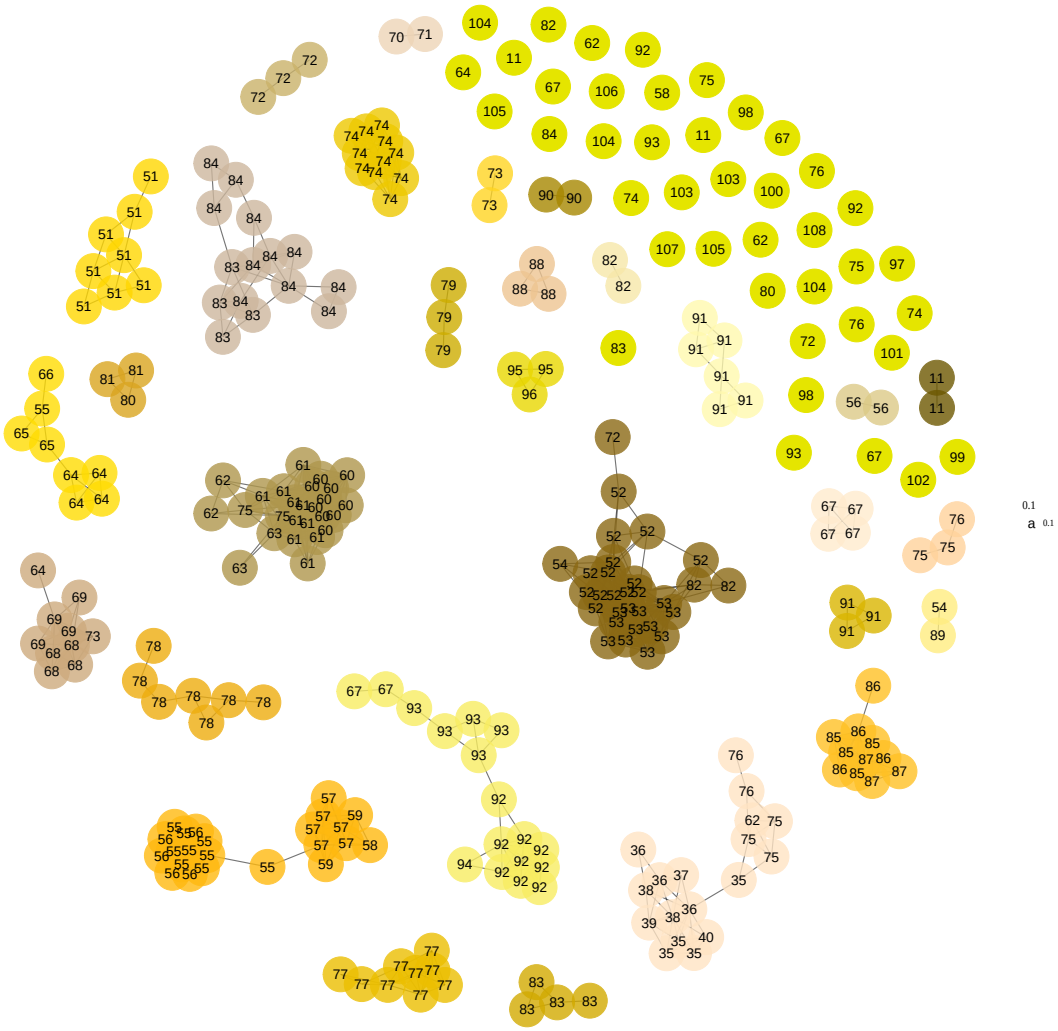
SUPPLEMENTARY FIGURE 2. Undirected graph network for the Multi group (enlarged version of Figure 2C).



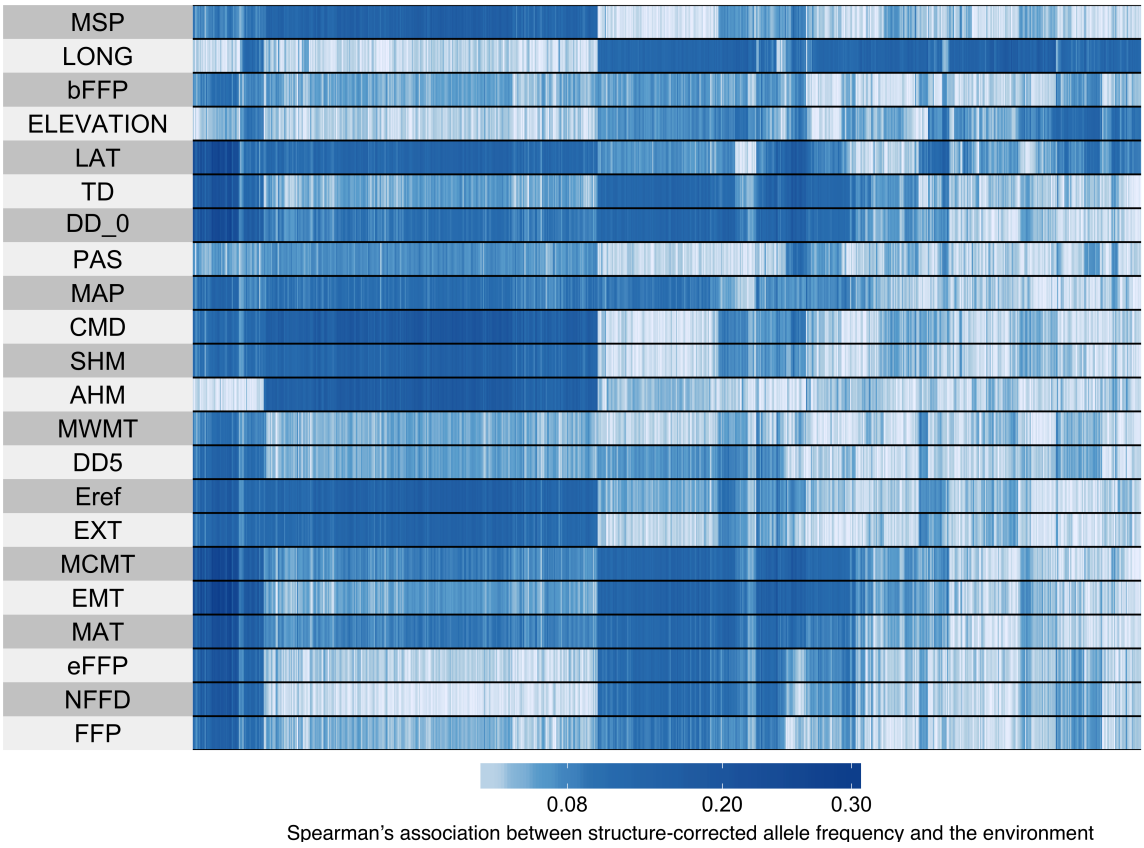
SUPPLEMENTARY FIGURE 3. Undirected graph network for the Aridity group (enlarged version of Figure 2D).



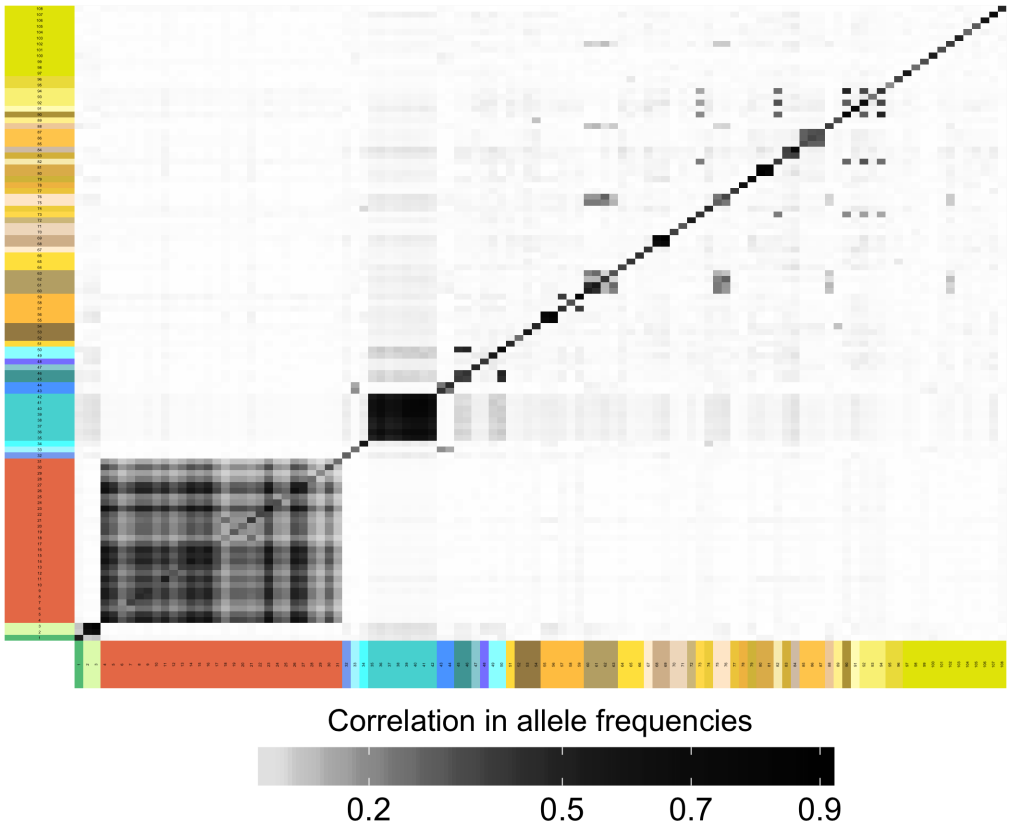
SUPPLEMENTARY FIGURE 4. Undirected graph network for the Freezing group (enlarged version of Figure 2E).



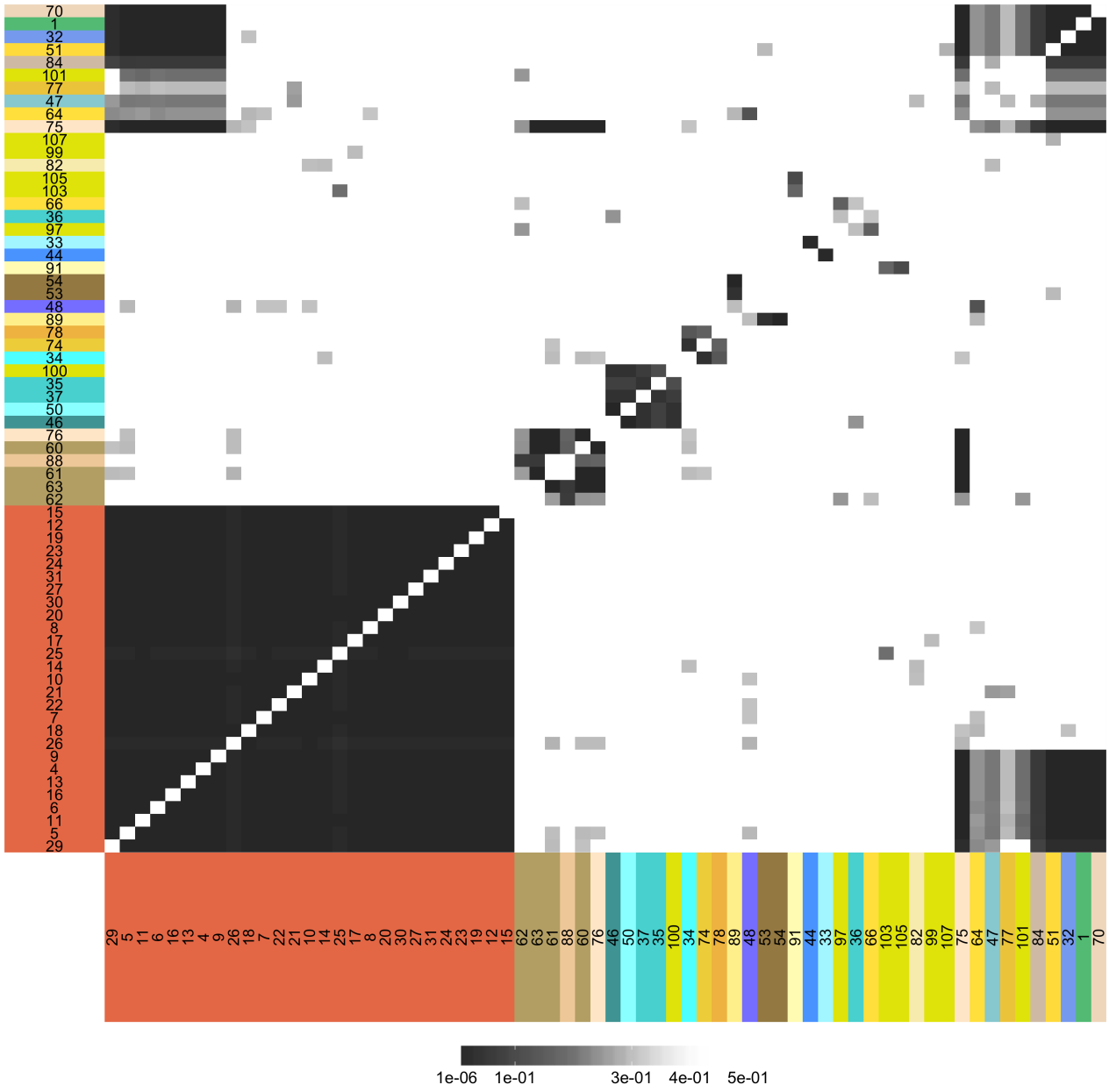
SUPPLEMENTARY FIGURE 5. Undirected graph network for the Geography group (enlarged version of Figure 2F).



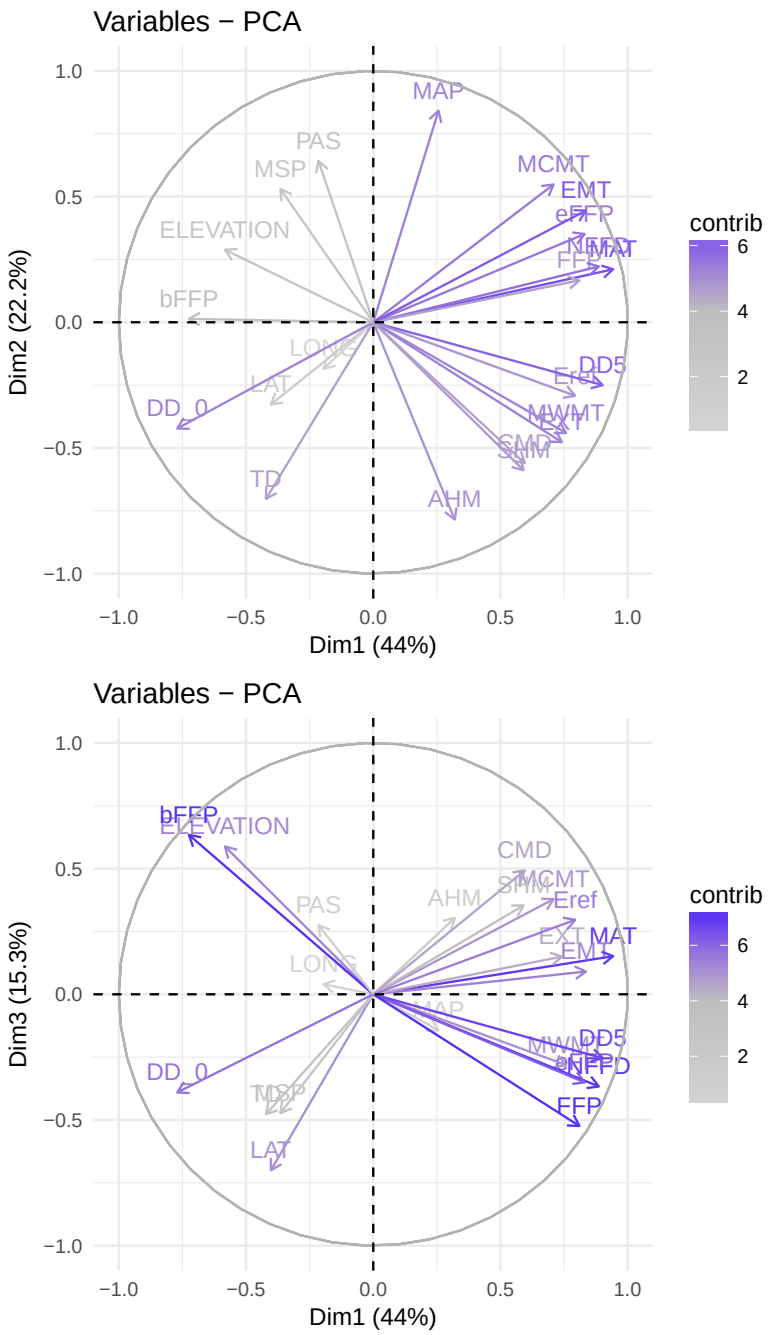
SUPPLEMENTARY FIGURE 6. Heatmap of structure-corrected allele associations with the environment, analogous to Figure 2B in the main paper. Note that although the pattern is very similar, the magnitude of allele correlations is smaller in the structure-corrected data.



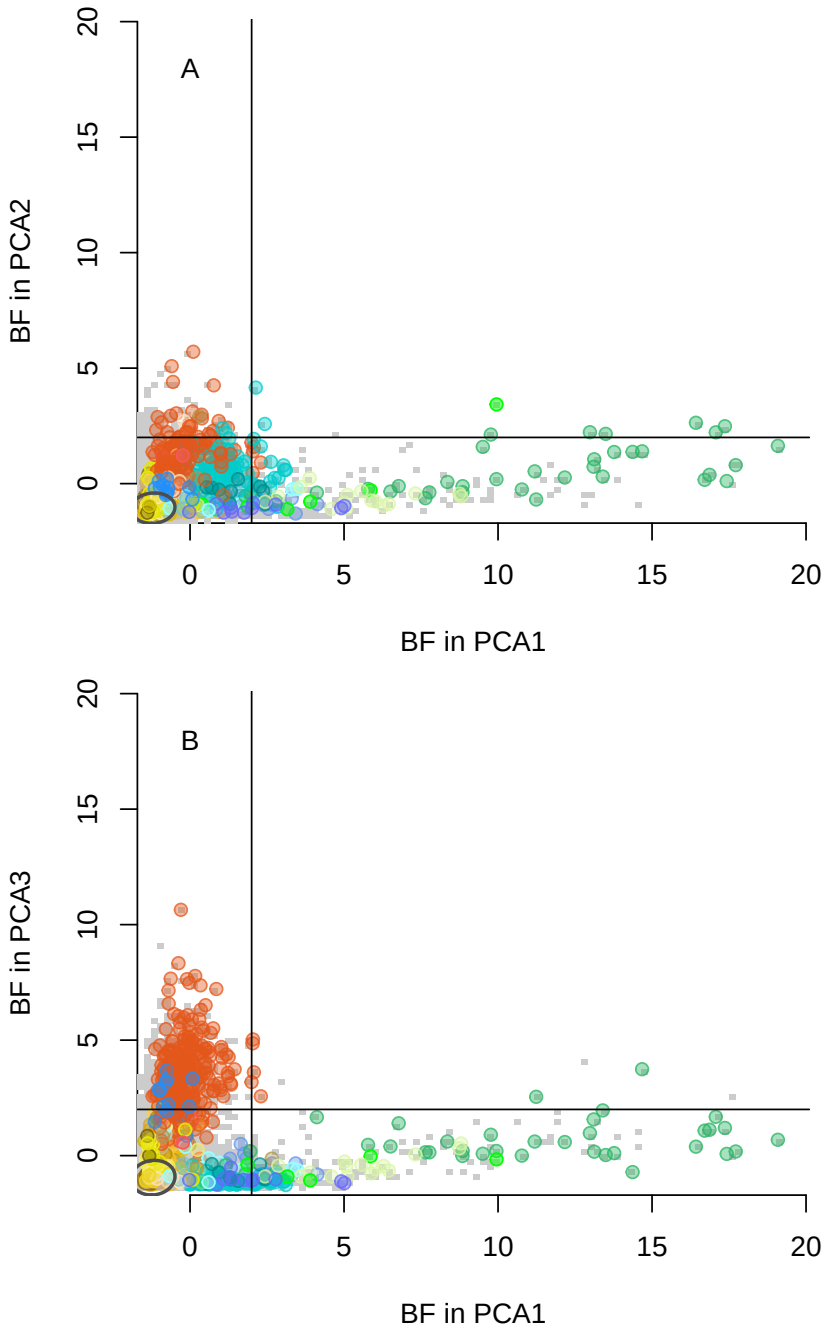
Supplementary Figure 7. Linkage disequilibrium heatmap. Mean correlation among allele frequencies between all 108 top candidate contigs. Contigs are ordered the same as Figure 2G in the main paper.



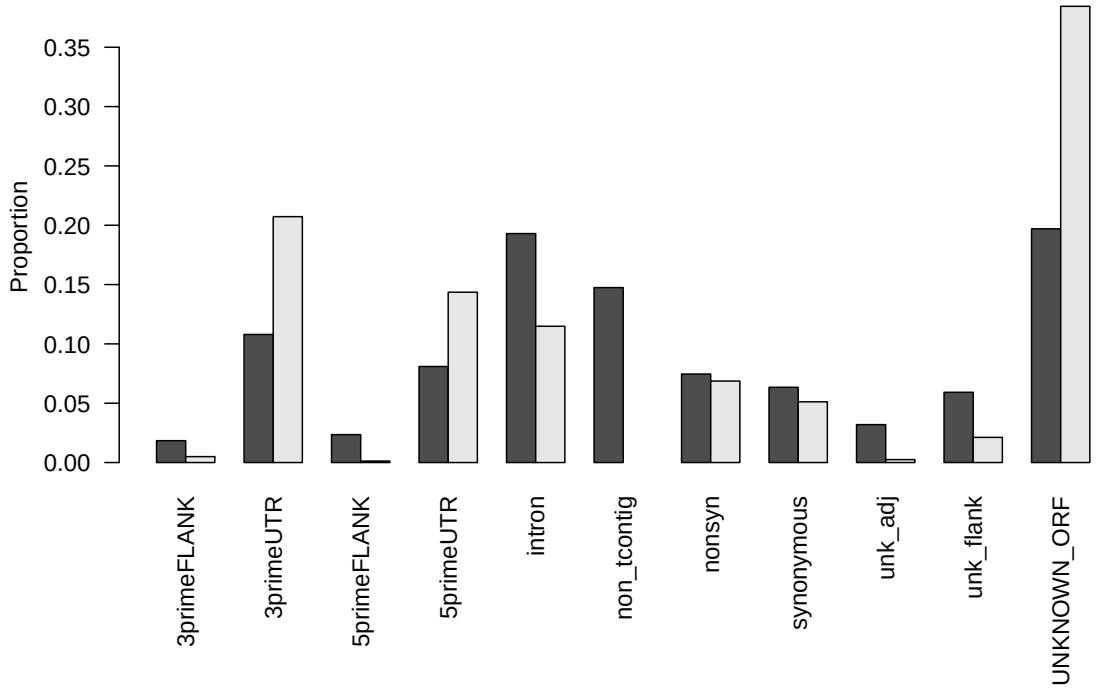
Supplementary Figure 8. Recombination heatmap, clustered by recombination rates. The same data as is shown in Figure 3, except re-clustered by recombination rates to more easily see the patterns of physical linkage.



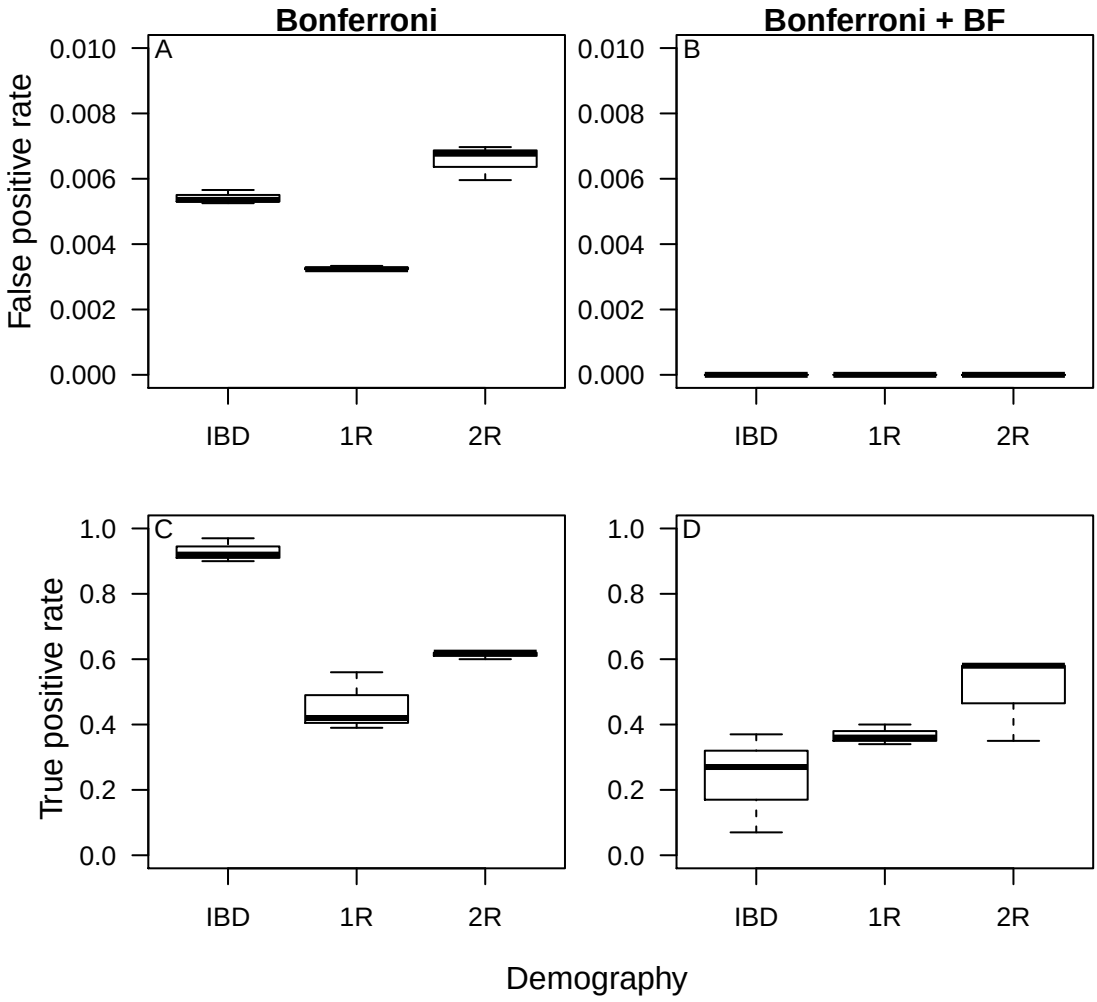
Supplementary Figure 9. The length and direction of each vector represents the scaled loading of that environmental variable onto the PC axis. The color of each vector represents the mean proportion of variance explained by that environment in the two axes plotted.



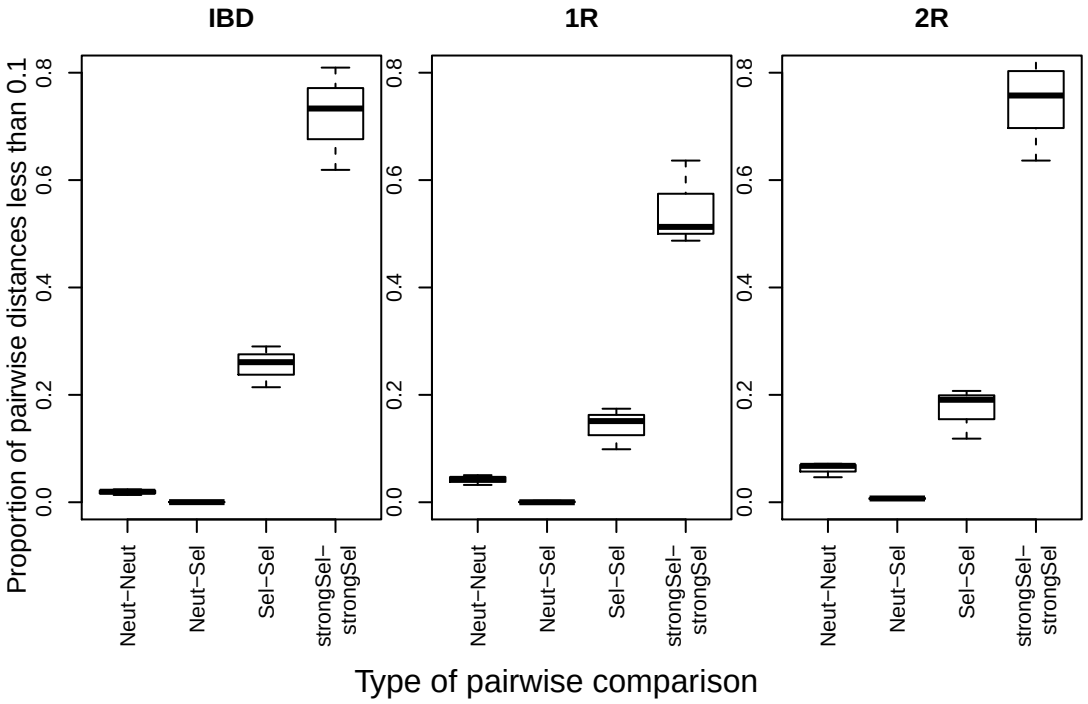
Supplementary Figure 10. The distribution of log-10 Bayes Factors for the association between a SNP and a PC axis. Each point is a SNP colored according to its co-association module in Figure 2C-F. Vertical and horizontal lines represent criteria for significance, and the black ovals represent the 95% prediction ellipse. Note that candidate SNPs all had BF > 2 with at least one univariate environmental variable.



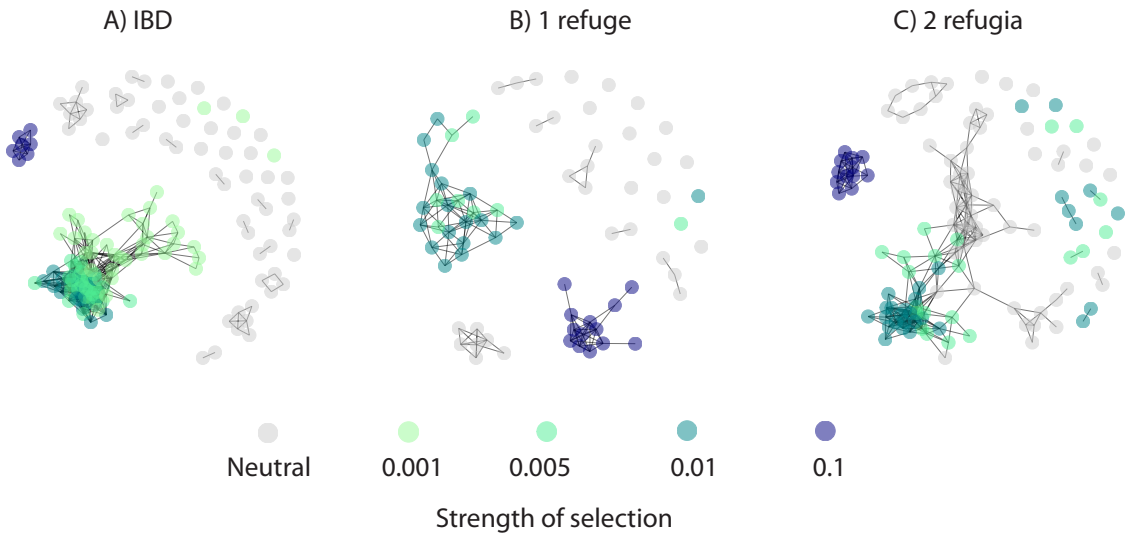
SUPPLEMENTARY FIGURE 11. Proportion of exome SNPs falling into various categories for genomic features compared to in the top candidate list. 3prime-FLANK: 3' flanking region; 3primeUTR: 3' untranslated region; 5prime-FLANK: 5' flanking region; 5primeUTR: 5' untranslated region; non-tcontig: not located in a transcriptomic contig (intergenic); nonsyn: non-synonymous substitution; unk-adj: unknown adjacent region; unk-flank: unknown flanking region; UNKNOWN-ORF: unknown open reading frame.



Supplementary Figure 12. Error rates from the simulations given a less stringent criteria (Bonferroni, left) and a more stringent criteria (Bonferroni and Bayes Factors from bayenv2, right). The less stringent criteria was used for the simulations because it had some false positives (A), while the more stringent criteria was used for the empirical data because it didn't have any false positives (B). The three demographies are isolation by distance (IBD), range expansion from one refuge (1R), and range expansion from two refugia (2R). While using the more stringent criteria resulted in no false positives, it also reduced the number of true positives (compare C and D), with the most severe reduction under isolation by distance.



Supplementary Figure 13. Evaluation of 0.1 as a distance threshold for creating a co-association module. The three demographies are isolation by distance (IBD), range expansion from one refuge (1R), and range expansion from two refugia (2R). For the simulated data, top candidates were chosen as described in the methods. Multivariate euclidean distance was calculated among the loci based on their associations with environments, and the proportion of pairwise distances above the distance threshold of 0.1 (used for the empirical data) was calculated for each type of comparison. We evaluated four types of pairwise comparisions: neutral loci with each other ("Neut-Neut"), neutral loci with selected loci ("Neut-Sel"), all selected loci with each other ("Sel-Sel"), and only loci under strong selection with each other ($s > 0.1$, "strongSel-strongSel"). A higher proportion of pairwise distances above the threshold indicates that these loci would be more connected to each other in the co-association network.



Supplementary Figure 14. The simulated datasets were nested within randomly generated selective environments, such that different demographic histories were simulated on the same environmental landscape. For this randomly generated environment, loci simulated under stronger selection had a propensity to cluster differently than loci simulated under weaker selection. To be clear, they still show the same patterns of associations, but the absolute value of the associations was just larger for the loci under strong selection and this caused the creation of a second cluster.