Supplementary material 3. Additional Analyses of Genetic Structure

To investigate fine-scale genetic structure across the Isle of Wight red squirrel population, we first conducted STRUCTURE analyses (Pritchard et al. 2000) on the full dataset using the admixture model. We used a burn-in of 250,000 iterations followed by 750,000 MCMC repetitions, with 10 replicates for each K (number of genetic clusters). The most likely number of clusters was identified using StructureSelector (Li and Liu 2018), which incorporates both the Evanno and Puechmaille methods. Run convergence and consistency across replicates were assessed using CLUMPP (Jakobsson and Rosenberg 2007).

STRUCTURE results indicated that K = 2 was the best-supported model, suggesting the presence of two genetic clusters broadly corresponding to an east–west division across the island. This result is presented in the main manuscript (Fig. 3). Additional results for K = 3 and K = 4 were also investigated but no more clusters were found (Figs. S1 and S2).

Given STRUCTURE's known tendency to favour K = 2 even when true structure is weak or absent (Janes et al. 2017), we performed separate STRUCTURE analyses on the eastern and western subpopulations using the same parameters. Again, STRUCTURE selected K = 2 as the most likely model in both cases (Figs. S3, S4), which likely reflects the algorithm's inherent bias rather than meaningful internal structure within each region.

To complement these results and avoid assumptions of Hardy-Weinberg or linkage equilibrium, we also conducted a Discriminant Analysis of Principal Components (DAPC) using the Adegenet package in R (Jombart 2008). Clustering was performed using the find.clusters function, setting the maximum number of clusters to 10 and using 1,000 starting points (n.start = 1000) across 1,000,000 iterations (n.iter = 1e6). The optimal number of clusters was selected based on the lowest Bayesian Information Criterion (BIC).

Cross-validation (xvalDapc) was used to identify the optimal number of principal components to retain, with a maximum of 35 PCs tested. The final DAPC was run retaining 35 principal components (n.pca = 35) and 4 discriminant functions (n.da = 4). Cluster assignment accuracy and the proportion of variance explained by each discriminant axis were examined to ensure robust group separation. The analysis identified two main clusters corresponding to eastern and western subpopulations (Fig. S5), consistent with STRUCTURE and spatial patterns inferred from the Monmonier algorithm.

Together, these analyses indicate weak but geographically consistent genetic structuring across the Isle of Wight, with an east–west pattern repeatedly supported despite low levels of divergence.

Figure captions:

Fig. S1 Population structure of red squirrels from the Isle of Wight inferred from STRUCTURE analysis at K = 3. Each vertical bar represents an individual, with colours indicating the proportion of their genome assigned to each of the three inferred genetic clusters. Individuals are grouped along the x-axis by sampling region, with '1' representing the western subpopulation and '2' the eastern subpopulation.

Fig. S2 Population structure of red squirrels from the Isle of Wight inferred from STRUCTURE analysis at K = 4. Each vertical bar represents an individual, with colours indicating the proportion of their genome assigned to each of the four inferred genetic clusters. . Individuals are grouped along the

x-axis by sampling region, with '1' representing the western subpopulation and '2' the eastern subpopulation.

Fig. S3 Genetic structure of red squirrels in the western subpopulation of the Isle of Wight. A) shows the Delta K plot from STRUCTURE analysis, identifying K = 2 as the most likely number of clusters using the Evanno method. B) displays the corresponding STRUCTURE bar plot for K = 2, with each vertical bar representing an individual grouped by sampling location. Colours indicate genomic assignment proportions.

Fig. S4 Genetic structure of red squirrels in the eastern subpopulation of the Isle of Wight. A) shows the Delta K plot from STRUCTURE analysis, identifying K = 2 as the most likely number of clusters using the Evanno method. B) displays the STRUCTURE bar plot for K = 2, with each vertical bar representing an individual grouped by sampling location. Colours indicate genomic assignment proportions.

Fig. S5 Density plot showing the distribution of individuals along Discriminant Function 1 from a Discriminant Analysis of Principal Components (DAPC) of the full dataset. Each curve represents one of the two inferred genetic clusters. Clear separation along the discriminant axis supports the presence of two genetically distinct populations, with minimal overlap indicating spatial genetic structure.

Fig. S1-

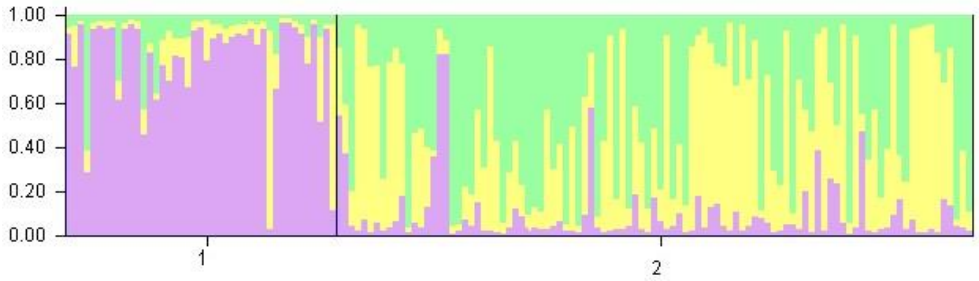


Fig. S2-

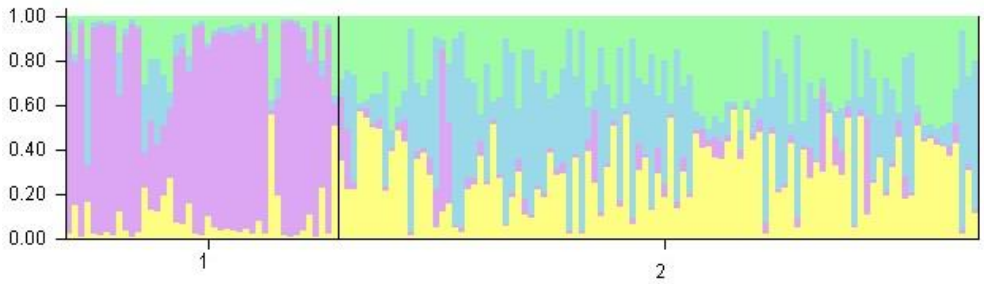


Fig. S3-

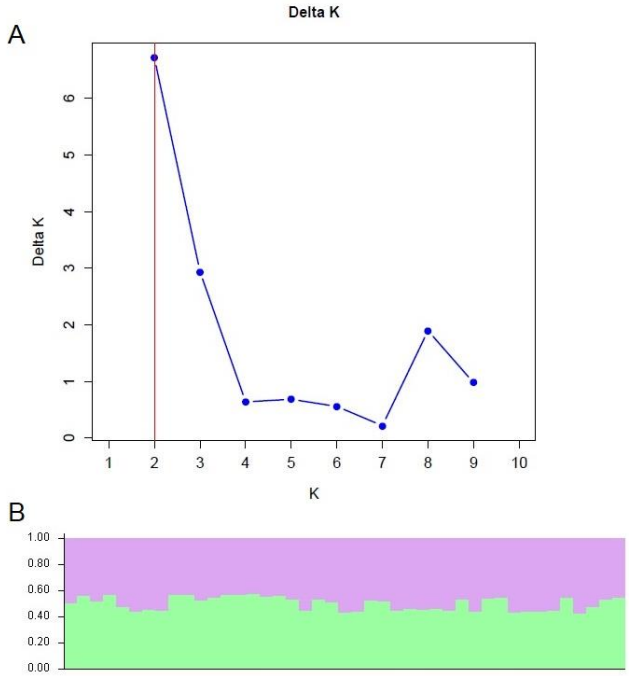


Fig. S4-

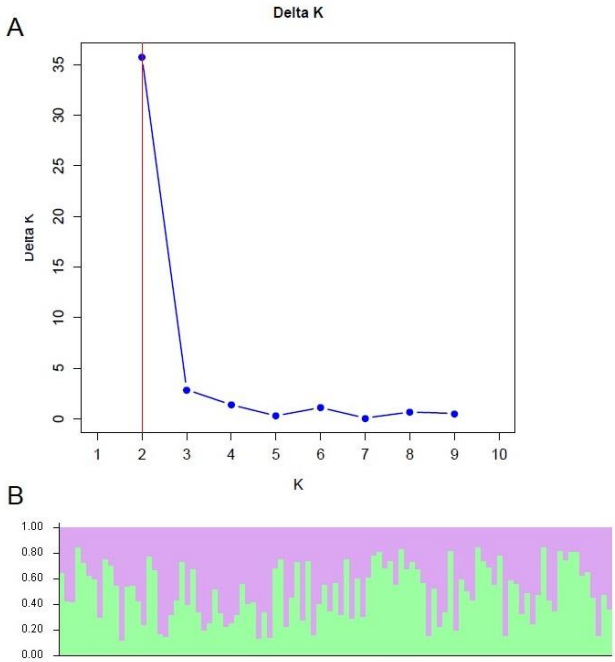
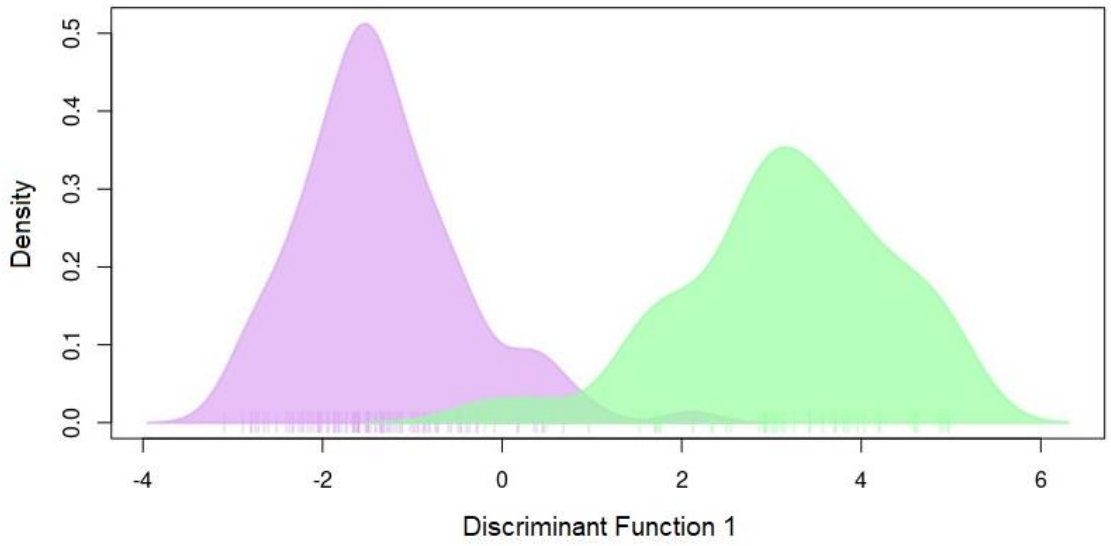


Fig. S5-



References

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Janes JK, Miller JM, Dupuis JR, et al (2017) The *K* = 2 conundrum. Mol Ecol 26:3594–3602. https://doi.org/10.1111/mec.14187

Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Li Y, Liu J (2018) StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18:176–177. https://doi.org/10.1111/1755-0998.12719

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

Corresponding Author

Alyson E. V. Buchanan, Bournemouth University, Christchurch House, Talbot Campus, Poole, Dorset BH12 5BB, U.K

E-mail abuchanan@bournemouth.ac.uk