

Figure S1. Bayesian information criterion (BIC) as a function of the number of clusters using combined data. The number of clusters that best differentiated genetic groups among all samples was  $K = 2$  clusters. BIC values showed little variation for  $K > 2$  clusters.

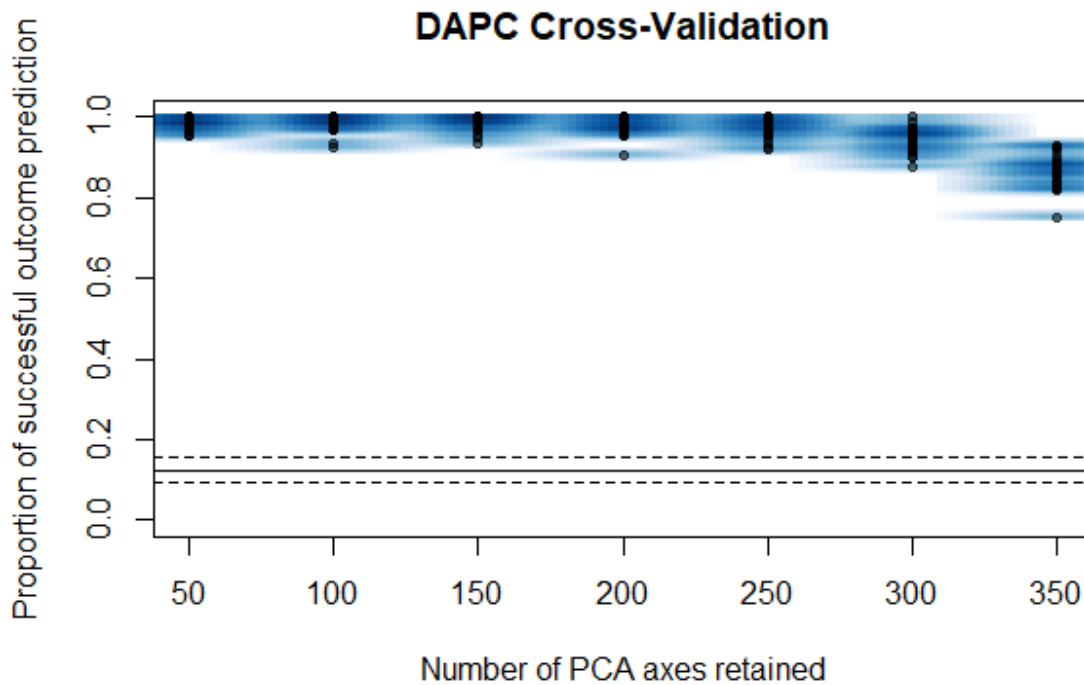


Figure S2. Cross-validation of discriminant analysis of principal components. The x-axis is the number of PCA axes retained for DAPC, and the y-axis is the proportion of successful outcome prediction. Individual replicates appear as points, and the density of those points in different regions of the plot is displayed in blue. The horizontal solid and dashed lines represent the mean and the 95% confidence interval for a random chance. The highest proportion of successful outcome predictions, and the lowest dispersion, was achieved when 50 principal components were retained.

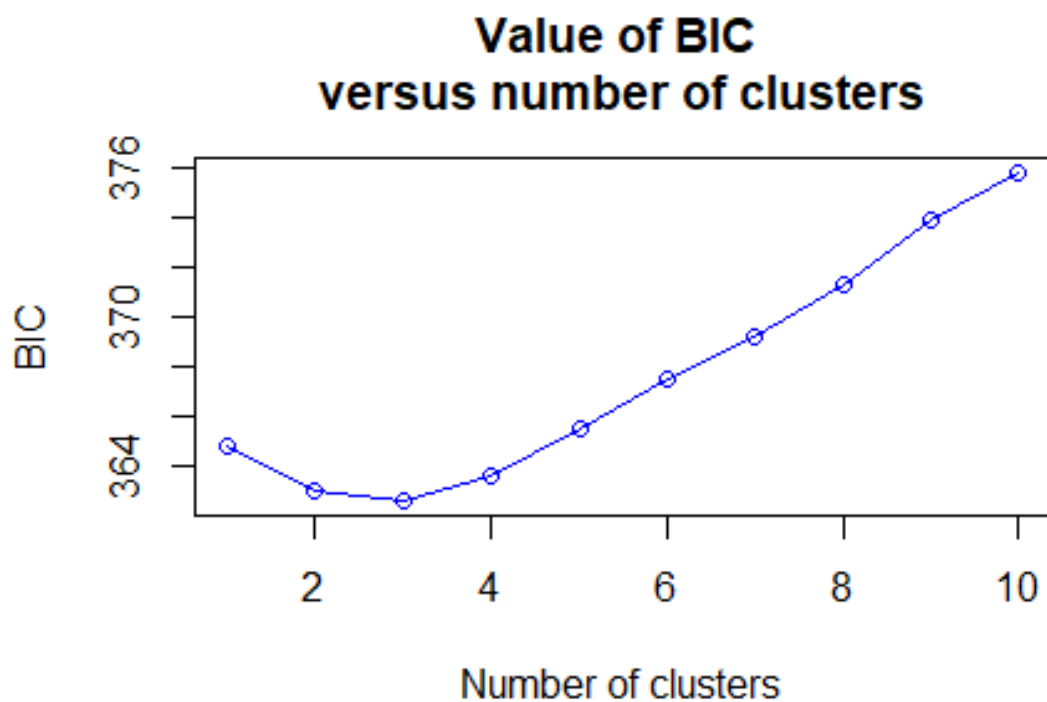


Figure S3. Bayesian information criterion (BIC) for each number of clusters using free-living data.

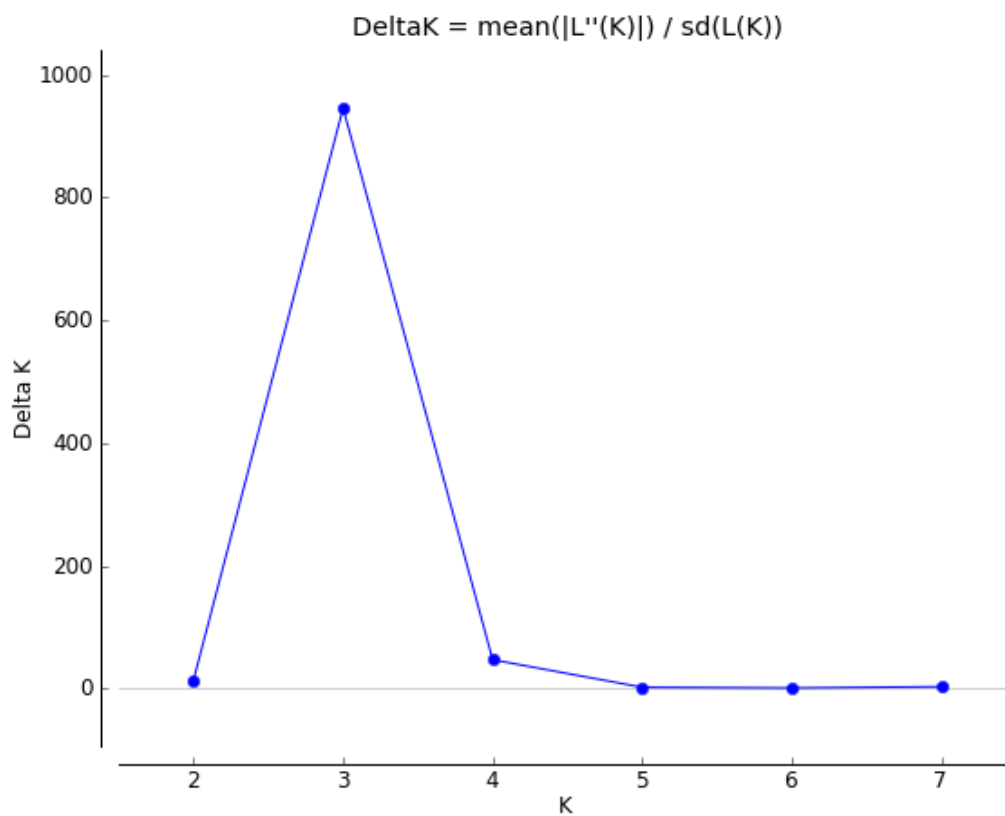


Figure S4. Estimated delta K following Evanno's method for  $n = 80$  free-living Atlantic salmon. The inferred number of genetic clusters that best fit the data was  $K = 3$ .

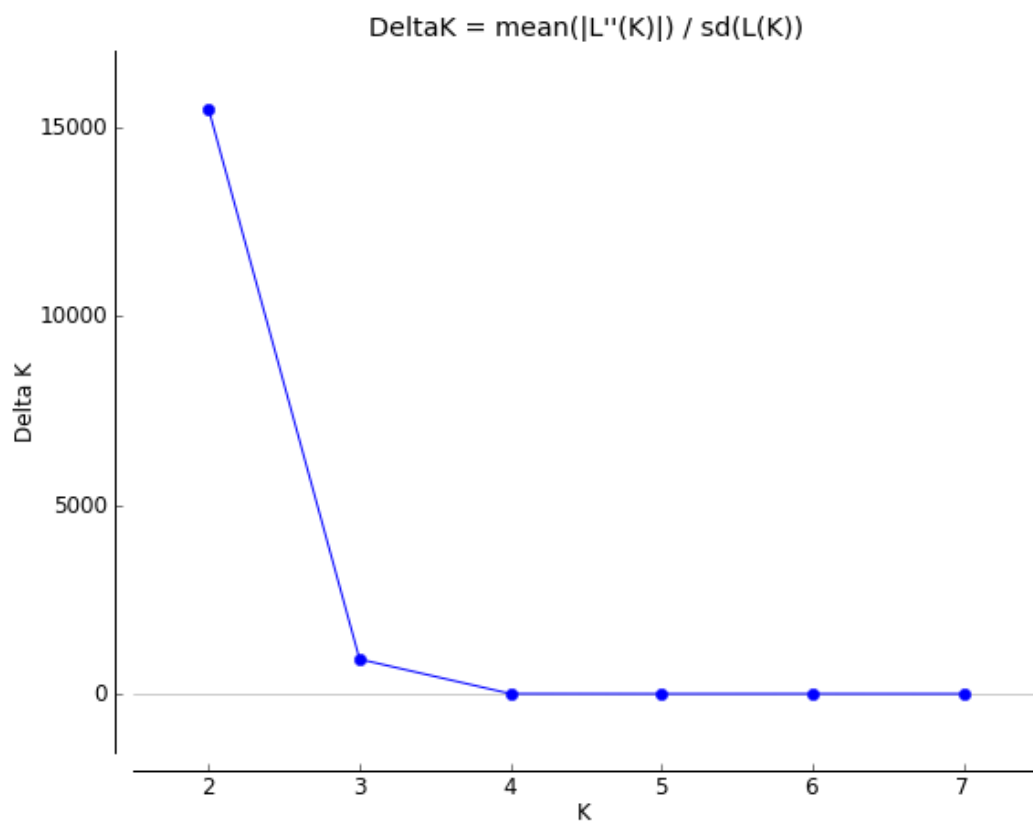


Figure S5. Estimated delta K following Evanno's method for reference collections and free-living Atlantic salmon combined. The number of genetic clusters that best describes the samples is  $K = 2$ , followed by  $K = 3$ .

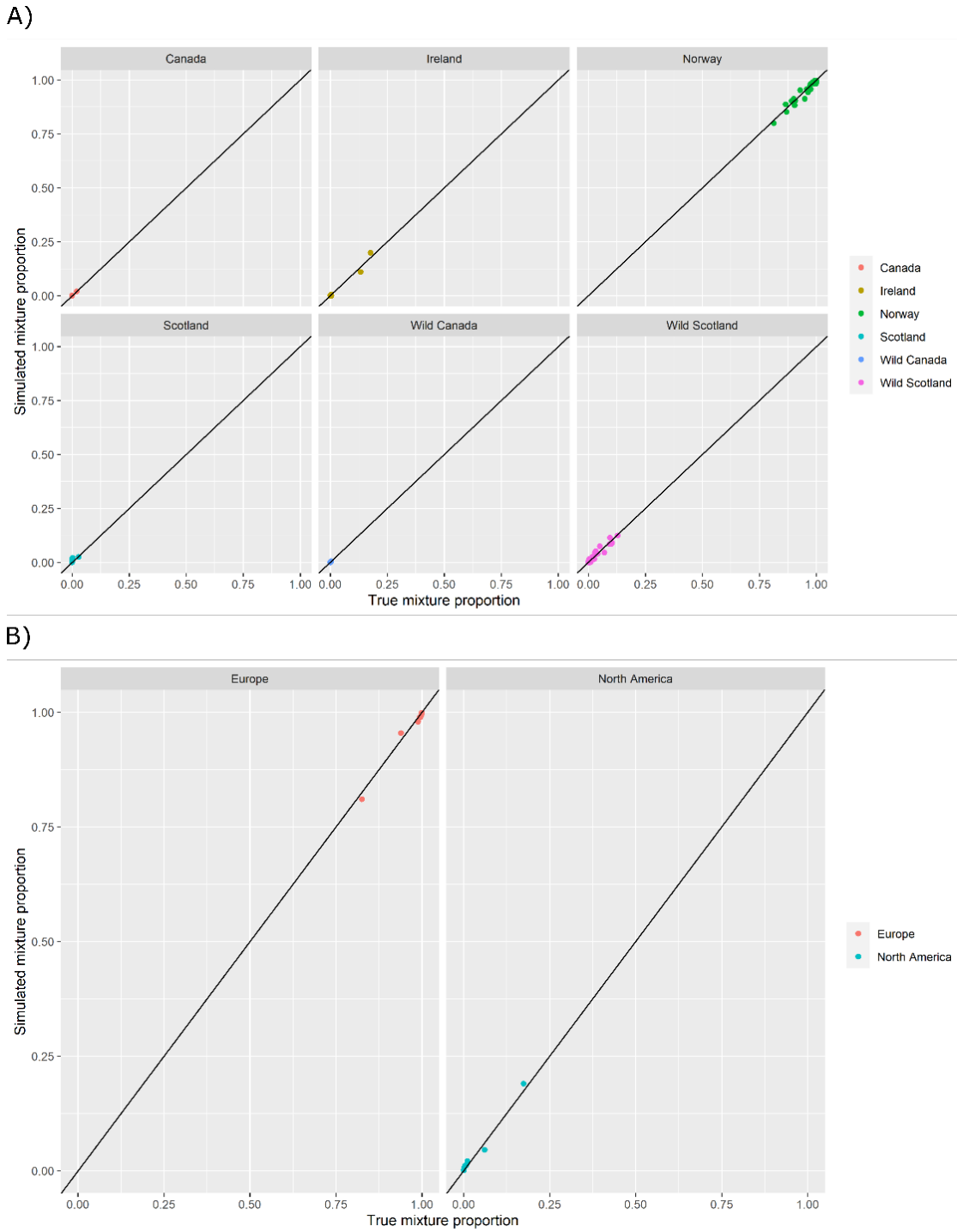


Figure S6. Comparison between true- and simulated-mixture proportions performed by RUBIAS, when using A) six reference collections and B) two reference groups. Points above or below the line represent upward and downward bias respectively.