

**An example of how mixing proportions can potentially be determined based on genomic data from Teske et al. (2021)**

**P. Teske<sup>1</sup>**

Previous molecular data generated for South African *Sardinops sagax*, which included a reduced-representation approach (ddRADseq, genomic data) to identify stock structure in a large number of samples, and a more expensive approach (exome data mined from RNA-seq transcriptome data) that was applied to a small subset of the data, showed congruent results: one genetic cluster was dominant on the west coast, another was dominant on the south coast (Teske et al. 2021). However, the level of resolution differed. While the ddRADseq data were unsuitable to assign any individuals to a single cluster, but rather various mixtures of ancestry components (Fig. 1), several individuals that were assigned to one of the two clusters with a probability of 100% were found for the more informative exome data (Fig. 2).

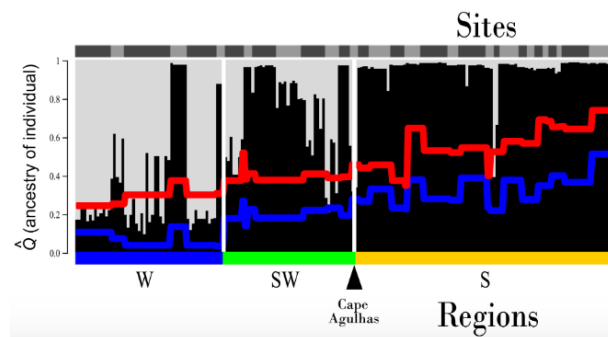


Fig. 1. Simplified STRUCTURE barplot from Teske et al. (2021) assigning sardines to one of two clusters (representing populations or stock components): grey: cool-temperate sardines (CTS), black: warm-temperate sardines (WTS). W = west coast, SW = southwest coast, S = south coast. East coast samples (sardine run participants) are excluded. No individuals were found that were 100% assigned to CTS.

<sup>1</sup> Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, Auckland Park 2006

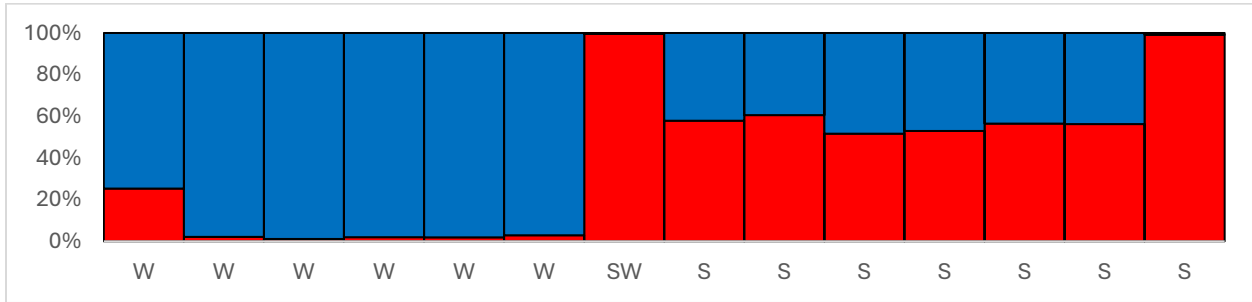
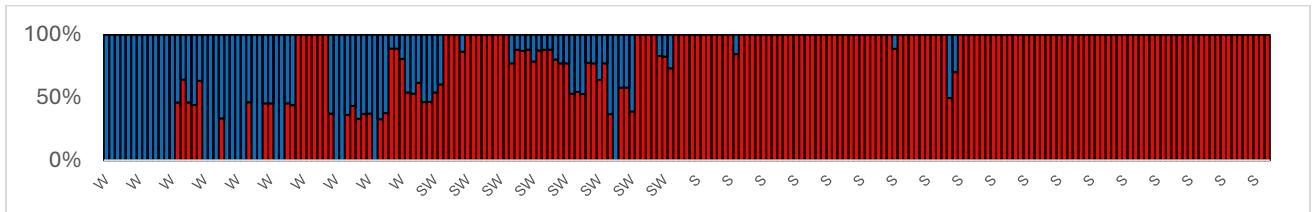


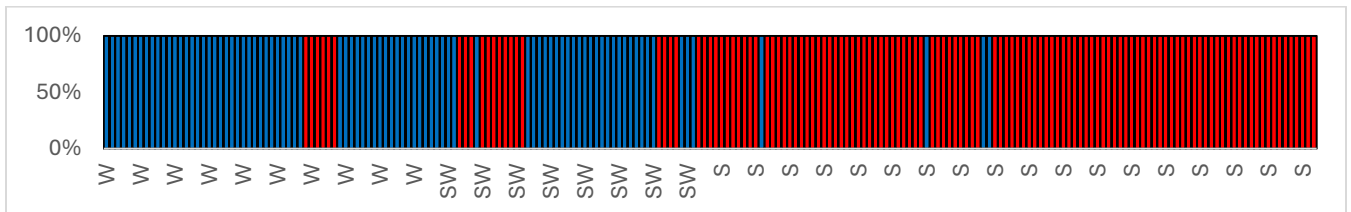
Fig. 2. STRUCTURE barplot based on exome data from RNA-seq assigning sardines to a cool temperature cluster (CTS, blue) and a warm-temperate cluster (WTS, red); W = west coast, SW = southwest coast, S = south coast (east coast samples representing sardine run participants) are excluded.

Examples of how to potentially make the genomic (ddRADseq) barplot more similar to this in order to assess gene flow, are shown below (Fig. 3). In a), portions are removed from the top and bottom of the barplot so that some individuals are 100% assigned to one particular cluster, as was the case for the exome dataset. In b) this is further simplified by assuming that any individual that has about 20% “blue” ancestry (which is common in the W and SW, but rare on the S coast) represents a member of the CTS stock.

a)



b)



Based on Fig. 3b, the percentages of individuals from from a particular stock (CTS and WTS) that occur in a particular region can be worked out by counting “red” and “blue” individuals in each region (Table 1).

Table 1. Counts and percentages  
of CTS and WTS in each region.

	<b>Counts</b>	<b>%</b>
CTS in W:	50	89
WTS in W:	6	11
CTS in SW:	34	69
WTS in SW:	15	31
CTS in S:	4	4
WTS in S:	105	96

Note that these considerations based on the original dataset may no longer be relevant, as the situation has changed somewhat following the reanalysis of the genomic data. Please see MARAM/IWS/2024/Sardine/P7 for details. Moreover, migration rate estimates calculated using BA3-SNPs (MARAM/IWS/2024/Sardine/BG2) are likely more meaningful than the above percentages.