**Introduction**

Based on the VCF data, 16 individuals were sequenced. There were two different chromosomes (chr), which were chr5 and chrZ. The VCF file contained 3,816,977 single nucleotide polymorphisms (SNPs) and the outgroup (Naxos2) was first removed before data filtering.

**Methods**

Preceding the analysis,the VCF file was then filtered using VCFtools (0.1.16) based on the listed filtration criteria as followed, the minimum depth < 3, and the maximum depth > 15 (Fig. 1A). The dataset was further filtered out missing data. Filtered left 3,664,421 SNPs. vcftools divided the genome into bins of 1,000,000 base pairs each to calculate the number and density of SNPs within bins of a specified size. In this case, 1,000,000 is the size of each bin, measured in base pairs. PLINK v1.90b6.21 64-bit was used to prune and create principal component analysis.

B

A

A graph of a number of blue bars

Description automatically generated with medium confidenceA graph of a number of columns

Description automatically generated with medium confidence

C

A graph showing a number of different sizes of lines

Description automatically generated with medium confidence

Fig. 1. Investigation of data. A) Histogram of mean depth per site averaged across all individuals before filtering. B, C) Bar plots represent a genomic bin, and the height of the bar represents the SNP count within that bin for B) chr5 and C) chrZ.

**Result and Conclusion**

The multivariate relationships between the taxa were visualized using a PCA that decomposes data of high dimensionality (PC1 – PC15) (Fig 2A). PC1 explained the highest percentages of variance, followed by the second component explaining the second most (Fig 2B). PC1 disclosed three major clusters corresponding to 8N, K0 and Lesina.

B

A

A graph of a graph showing the principal component

Description automatically generatedA graph showing different colored dots

Description automatically generated

Fig. 2. The three PCA scatter diagram was made by the first and second principal components.

A Structure-like approach, ADMIXTURE Version 1.3.0 was used to estimate the proportions of admixtures among the taxa by assuming that the number of ancestries (k) was in the range of 2–4. The optimal k value was confirmed based on the minimum Cross-validation (CV) error estimated by ADMIXTURE, k=2 (0.57402), k=3 (0.73722), k=4 (0.89689), k=5 (1.00592). Fst divergences between estimated populations is 0.361 (k=2). Admixture findings

A

B

A graph showing a number of samples

Description automatically generated with medium confidenceA diagram of a sample

Description automatically generated with medium confidence

Fig. 3. Bar plot of individual for the genetic clusters inferred using ADMIXITURE A) (K = 2), B) (K = 2, 3, 4).

found support for two clusters corresponding to the parent species and intermediate admixture proportions, with K0 having the highest probabilities of clustering with the 8N and very small probabilities of clustering with the Lesina (Fig. 3A).

A

B

A red and blue chart

Description automatically generatedA red and blue chart

Description automatically generatedA red and blue chart

Description automatically generated

C

Fig. 4. Manhattan plot illustrating genetic differentiation (Fst) across chromosomes

Differentiation coefficients (Fst) was used as a measure of population structure and bcftools 1.19 and vcftools were applied. The Fst in the range of 0.05–0.15 indicates moderate divergence. Fst estimate between the 8N and K0 was 0.054429, K0 and Lesina was 0.13074, indicating a moderate level of genetic differentiation being studied. 8N and Lesina was 0.20142 indicated that relatively high genetic differentiation between taxa. This level of divergence might imply substantial barriers to gene flow or significant historical isolation between the taxa under study.