**Methods**

**Sequence Processing and Quality Control**

Sequencing reads were first processed using the nf-core sarek pipeline (Ewels et al., 2020, Di Tommaso et al., 2017), which included steps for trimming, mapping, and marking duplicates. Two individuals were excluded from further analysis due to not meeting our quality control standards: number and quality of sequencing reads. The assembly of the genome of *Melitaea athalia* from the Darwin Tree of Life project was utilized as a reference.

**Quality Control and Phylogenetic Analysis**

Variant calling was performed using SAMtools mpileup (Li et al., 2009), following the settings recommended by the GenErode pipeline. The output VCF file was used for perform PCA analysis for initial quality control and to extract COI genes.To ensure species-specific analysis, we constructed a phylogenetic tree using the COI gene sequences. Sequences for *M. athalia*, *M. diamina* and *M.cinxia* were obtained from GenBank. The tree was constructed using MEGA software (Kumar et al., 2018) and visualized with ETE Toolkit (Huerta-Cepas et al., 2016). Individuals grouping with any of the outgroup species were excluded from the further analysis.

**Genomic Analysis Using ANGSD**

Due to the low and uneven coverage in remaining data, we switched to the ANGSD toolkit (Korneliussen et al., 2014) tailored for such scenarios. ANGSD was employed to directly estimate genotype likelihoods from two datasets: the full dataset of *Melitaea britomartis* and a subset including only individuals from Sweden. We applied following filters to both sample sets : MAF > 0.05. minimum mapping quality = 20, base quality score = 20, 5 < Depth < 30. We selectd sites covered in at least 20 and 10 individuals for full and Swedish datasets respectively. Beagle-format files generated by ANGSD were used as input for pcangsd (Meisner & Albrechtsen, 2018), which produced a covariance matrix. Principal Component Analysis (PCA) was then performed again to explore population structure using Python 3.7 with the scikit-learn library (Pedregosa et al., 2011). Additionally, the site frequency spectrum (SFS) was derived from ANGSD output and used to estimate nucleotide diversity (theta) for the population using the realSFS module. All visualizations were created using custom scripts, which are available on our GitHub repository.

**References**

1. Ewels, P. et al. "The nf-core framework for community-curated bioinformatics pipelines." Nature Biotechnology 38, 276–278 (2020). [Link to Paper](https://www.nature.com/articles/s41587-020-0439-x)
2. Di Tommaso, P. et al. "Nextflow enables reproducible computational workflows." Nature Biotechnology 35, 316–319 (2017). [Link to Paper](https://www.nature.com/articles/nbt.3820)
3. Li, H. et al. "The Sequence Alignment/Map format and SAMtools." Bioinformatics, 25(16), 2078–2079 (2009). [Link to Paper](https://academic.oup.com/bioinformatics/article/25/16/2078/204688)
4. Kumar, S. et al. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." Molecular Biology and Evolution, 35(6), 1547–1549 (2018). [Link to Paper](https://academic.oup.com/mbe/article/35/6/1547/4990888)
5. Huerta-Cepas, J. et al. "ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data." Molecular Biology and Evolution, 33(6), 1635–1638 (2016). [Link to Paper](https://academic.oup.com/mbe/article/33/6/1635/2579822)
6. Korneliussen, T. S. et al. "ANGSD: Analysis of Next Generation Sequencing Data." BMC Bioinformatics, 15, 356 (2014). [Link to Paper](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-014-0356-4)
7. Meisner, J. & Albrechtsen, A. "pcangsd: principal components analysis on next generation sequencing data." bioRxiv (2018). [Link to Preprint](https://www.biorxiv.org/content/10.1101/311951v1)
8. Pedregosa, F. et al. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825