



UNIVERSITY OF
GOTHENBURG

DEPARTMENT OF BIOLOGICAL AND
ENVIRONMENTAL SCIENCES

DIFFERENT PHYLOGENETIC RELATIONSHIPS WITHIN THE GENOME OF EUROPEAN *OXYRIA DIGYNA* (L.) HILL



Sofie Kedbäck

Degree project for Master of Science (120 hec) with a major in Biology

BIO715, Degree project in Evolutionary and behavioral ecology, 30 hec

Second cycle

Semester/year: Spring 2026

Supervisors: Christine Bacon, Department of Biological & Environmental Sciences, GU &
Cassandra Elphinstone, University of Oslo

Examiner: Bengt Oxelman, Department of Biological & Environmental Sciences

Lindman, C. A. M. (1901-1925). *Bilder ur Nordens Flora* (1st ed.). Wahlström & Widstrand.

Index

Abstract	3
Sammanfattning	3
1. Introduction	5
1.1 Phylogenetic trees and networks	5
1.2 Core- and accessory genes	5
1.3 The Arctic and plant dispersal	6
1.5 <i>Oxyria digyna</i> (L.) Hill.....	6
1.6 Aim and hypothesis	7
1.7 Broader impact of the work.....	8
2. Materials and methods	8
2.1 Data collection and sequencing	8
2.2 Extraction of SNPs from the raw data and filter	9
2.3 Subset the data.....	10
2.3.1 <i>Subset the SNPs by chromosome and create a midpoint root</i>	10
2.3.2 <i>Subset the SNPs by core- and accessory gene enriched regions</i>	10
2.4 Create phylogenies based on core- and accessory genes.....	11
2.4.1 <i>Create phylogenetic graphs in TreeMix</i>	11
2.4.2 <i>Create phylogenetic networks in SplitsTree</i>	12
3. Results.....	12
3.1 The phylogenetic trees based on chromosomes one to seven	12
3.2 The phylogenetic graphs based on core- and accessory enriched regions	12
3.1.1 <i>Migration edges needed to explain the data</i>	12
3.1.2 <i>Comparison between the phylogenetic graphs for regions enriched in core- and accessory genes</i>	15
3.3 Phylogenetic network in SplitsTree for regions enriched in core- and accessory genes.....	15
4. Discussion	15
4.1 Phylogenetic relationships within <i>Oxyria digyna</i>	15
4.1.1 <i>The relationships based on the midpoint rooted chromosome one tree</i>	15
4.1.2 <i>Divergent relationships between the seven chromosomes</i>	17
4.2 Phylogenetic differences between core- and accessory genes	18
4.2.1 <i>Phylogenetic difference based on core- and accessory genes</i>	18
4.2.2 <i>Core genes show more migration edges and are more consistent than accessory genes</i>	18
4.2.3 <i>Core genes show more reticulations than accessory genes</i>	18
4.3 Limitations, further studies, and broader implications.....	19

4.3.1 <i>Limitations and further studies</i>	19
4.3.2 <i>Broader implications</i>	20
4.4 Conclusion.....	20
5. Code availability	21
6. Acknowledgements	21
7. References	21
Appendix Popular science summary and Supplementary material	25

Abstract

Innovations in genomics are revolutionizing our understanding of biodiversity. Complex phylogenetic relationships can now be represented through graphs and networks, enabling incorporation of for instance gene flow and recombination. This thesis investigates the phylogenetic relationships between ten European populations of the Arctic-alpine plant species *Oxyria digyna* (L.) Hill (*Polygonaceae*). I explore the phylogenetic differences between core genes (conserved across species) and accessory genes (unique to a species). I hypothesize 1) divergence of the sampled populations will follow the retreat of the glaciers, and 2) core- and accessory genes will show different phylogenies and gene flow, since core genes are under strong purifying selection, accessory genes may be locally adaptive, and with these genes having different locations on the chromosome affecting recombination. Phylogenies for core and accessory genes were compared using two bioinformatic approaches (TreeMix and SplitsTree). Results show 1) southern populations diverge early as expected, but not earlier than populations in Greenland, and 2) there are phylogenetic differences between core- and accessory genes. The phylogeny not only following the glacial retreat may be due to different dispersal ways into and across Europe. Contrary to our expectations, core genes needed more migration events and network reticulations to explain the data perhaps due to these genes occurring in regions of high recombination. Accessory genes showed higher variation and fewer total migration events than core genes, perhaps suggesting local adaptation affect these genes more. Core gene phylogenies were more consistent than those from accessory genes, potentially because of higher purifying selection due to their essential cellular functions. For phylogenetic studies of interbreeding populations, evolutionary differences between regions of the genome should be considered, as treating variable sites equally may not show the diverse evolutionary patterns within the species.

Key words: accessory genes, core genes, gene flow, Oxyria digyna, phylogeny, recombination

Sammanfattning

Innovationer inom genomik har revolutionerat vår förståelse av biologisk mångfald. Komplexa fylogenetiska släktskap kan nu representeras av grafer och nätverk, vilket gör det möjligt att inkludera till exempel genflöde och rekombination. Denna uppsats undersöker de fylogenetiska släktskapen mellan tio europeiska populationer av den arktisk-alpina växtarten *Oxyria digyna* (L.) Hill (*Polygonaceae*). Jag utforskar de fylogenetiska skillnaderna mellan kärngener (bevarade i flera arter) och accessoriska gener (unika för en art). Mina hypoteser är: 1) divergensen hos de studerade populationerna kommer att följa glaciärernas tillbakadragande, och 2) kärn- och accessoriska gener kommer att visa olika fylogener och genflöde, eftersom kärngener är under stark selektiv restriktion, accessoriska gener är lokalt anpassningsbara, och med olika placering på kromosomen för generna vilket påverkar rekombination. Fylogener för kärn- och accessoriska gener jämfördes med två bioinformatiska metoder (TreeMix och SplitsTree). Resultaten visar 1) att sydliga populationer divergerar tidigt som förväntat, men inte tidigare än populationer på Grönland, och 2) att det finns fylogenetiska skillnader mellan kärn- och accessoriska gener. Fylogenin som inte bara följer glaciär tillbakagång kan bero av olika spirdningsvägar in i, och över Europa. I motsats till våra förväntningar behövde kärngener fler migrationshändelser och mer nätverksstruktur för att förklara datan, kanske på grund av att dessa gener förekommer i regioner med hög rekombination. Accessoriska gener visade högre variation och färre totala migrationshändelser än kärngener, vilket kan tyda på att lokal anpassningar påverkar dessa gener mer. Kärngeners fylogener var mer konsekventa än de från accessoriska gener, potentiellt på grund av renande selektion och deras väsentliga cellulära funktioner. För fylogenetiska studier av populationer som korsas med varandra

bör evolutionära skillnader mellan regioner i genomet beaktas, eftersom lika hantering av variabla platser i genomet kanske inte visar mångfalden av evolutionära mönster inom arten.

Nyckelord: accessoriska gener, fylogeni, genflöde, kärngener, Oxyria digyna, rekombination

1. Introduction

1.1 Phylogenetic trees and networks

Phylogenetics can help us infer the evolutionary history of species or populations (Huson & Bryant, 2006; Young & Gillung, 2020). A phylogenetic tree, from morphological or genomic data, can be used to find the source of a pathogen, aid in conservation prioritization, or investigate the evolution of a taxon (Ruiz-García, 2021). As sequencing methods advances, phylogenetic relationships become possible to infer between populations. Low-coverage whole genome sequencing (lcWGS) enables cost-effective sequencing that provides large numbers of single nucleotide polymorphisms (SNPs) across the whole genome, which therefore is a useful tool for both population genetics and phylogenetics (Lou *et al.*, 2021).

A phylogenetic tree assumes bifurcating relationships, even though data often violate these assumptions, for example due to gene flow and recombination (Huson & Bryant, 2006; Huson & Scornavacca, 2011). Software to infer phylogenies are ever increasing and now include sophisticated algorithms to model evolutionary complexity through non-dichotomous branching. For example, TreeMix (Pickrell & Pritchard, 2012) builds phylogenies that model populations as a graph, enabling the explicit incorporation of gene flow. Here, in a set of interbreeding populations, one population can share its evolutionary history with multiple other populations. This is represented in TreeMix as “migration edges”. Edges, or possible branches, with the highest weight becomes the parental population in the graph (Pickrell & Pritchard, 2012), while the other edges will be represented as the migration edges. Weight for a migration edge relates to the fraction of alleles inherited from the parent population, ranging from zero (weak inheritance and low weight) to 0.50 (strong inheritance and high weight). SplitsTree (Huson & Bryant, 2024) is a network approach to account for conflicting signals in population data (Huson & Bryant, 2006). Although internal nodes cannot be interpreted as hypothetical ancestors like in a phylogenetic tree, the reticulations in the network may be explained by gene tree discordance, gene flow, hybridization, recombination, and gene duplication and loss (Huson & Bryant, 2006).

1.2 Core- and accessory genes

Advances in sequencing methods have also led to the realization of reference bias when mapping the sequenced data to a single reference genome (Degner *et al.*, 2009; Loegler *et al.*, 2026). Sequences more similar to the reference genome have higher representation than the alternative allele, making the new genome more similar to the reference (Degner *et al.*, 2009). A possible solution to minimise reference bias is to use a pangenome (Eizenga *et al.*, 2020) which represent more genetic diversity across multiple populations than a single reference genome (Golicz *et al.*, 2020; Loegler *et al.*, 2026; Soltis & Soltis, 2026).

The exact definition of the pangenome varies, but it generally includes all genes in all studied individuals (Tranchant-Dubreuil *et al.*, 2019). A pangenome consists of core genes, which are shared among all studied individuals, and accessory (also called variable- or dispensable) genes which are not shared by all individuals (Golicz *et al.*, 2020; Tellini *et al.*, 2025). Core genes are conserved and are responsible for essential cellular processes such as housekeeping functions, DNA replication, and maintenance of cell-homeostasis (Tranchant-Dubreuil *et al.*, 2019; Tellini *et al.*, 2025). Since the core genes are more conserved, they have stronger selective constraints than accessory genes (Golicz *et al.*, 2020). The accessory genes are instead suggested to be involved in local adaptation and phenotypic trait diversity such as stress response, defence mechanism, organ sizes, and flowering time (Tranchant-Dubreuil *et al.*, 2019; Hu *et al.*, 2022; Tellini *et al.*, 2025; Soltis & Soltis, 2026).

Species level core- and accessory genes are concentrated in different locations of plant chromosomes (e.g. Fig. 1a; Elphinstone et al. in prep.). The core genes are primarily located close to the telomeres where recombination rates are generally higher, and the accessory genes are located near centromeres with lower recombination rates (Fig. 1a; Nachman, 2002; Elphinstone et al. in prep). Recombination is when homologous chromatids exchange DNA segments during gamete formation, and thus shuffles genetic information among the homologous pairs (Hillis *et al.*, 2020).

1.3 The Arctic and plant dispersal

The Arctic is defined as the area above the Arctic Circle or north of the treeline (Meltofte, 2013). The area has experienced repeated environmental changes because of the Pleistocene (ca. 2.6 million years ago (Ma) to 11,700 years ago; Davinack, 2024) glacial cycles (Hewitt, 2000). During the last glacial maximum (ca. 29,000-19,000 years ago; Clark *et al.*, 2009), large parts of Europe and North America were covered by ice sheets (Ehlers & Gibbard, 2007; Fig. 1b). Landscape and vegetation changed repeatedly during the Pleistocene (Hewitt, 2000) (White *et al.*, 1997; Hoffmann & Röser, 2009). Today the Arctic primarily consists of tundra and open landscapes (Meltofte, 2013). These environmental changes have caused strong selection on contemporary Arctic taxa, forcing them to be highly adapted to their extreme habitats (Meltofte, 2013; Qian *et al.*, 2021).

Although there are species with an Arctic origin (e.g. species in *Cassiope* and *Diapensia*), other Arctic species originated in southern mountain ranges, many in Central Asia (Makhrov *et al.*, 2019). For instance, a molecular phylogeny of the American Arctic lineage of *Artemisia* identified Asian progenitors (Tkach *et al.*, 2008).

Some Arctic species are also distributed in southern mountain areas (in or close to the temperate region; Makhrov *et al.*, 2019). A classical hypothesis posits that contemporary Arctic-alpine species moved into the Arctic multiple times, dispersing to southern mountains in response to advancing ice sheets, and then subsequently reestablishing in the Arctic following the retreating ice sheets (Hultén, 1937; Hultén, 1962; Darwin, 2011). This hypothesis is supported by recent literature reviews on a range of Arctic-alpine species, where many of the species showed an origin in Asian mountains (e.g. the Qinghai-Tibetan Plateau), with repeated colonisation of the Arctic and southern mountain regions in North America and Europe, following environmental change. (Makhrov *et al.*, 2019; Brochmann *et al.*, 2025). Species colonizing the Arctic are suggested to be pre-adapted to open and cold environments because of their previous high mountain- or close lowland habitats (Zhang *et al.*, 2023). This pattern is supported by phylogenetic studies showing that many plant species independently evolved Arctic adaptations with immigration to the Arctic, rather than via in situ diversification (Hoffmann & Röser, 2009; Zhang *et al.*, 2023).

1.5 *Oxyria digyna* (L.) Hill

Oxyria (*Polygonaceae*) is a genus consisting of two species (Sun *et al.*, 2012): *Oxyria digyna* (L.) Hill (Fig. 2A; Artdatabanken, 2026) and *O. sinensis* Hemsl. (WFO, 2026). While *O. sinensis* only occurs in China (WFO, 2026), *O. digyna* has a circumarctic distribution, with parapatric populations in southern mountains of the Northern Hemisphere (such as the Alps in Europe, the Himalayas in Asia, and the Rocky Mountains in North America; Hultén, 1971; Allen *et al.*, 2012; Fig. 2B). Fossil evidence of *O. digyna* dates back to 3.0 Ma (Matthews & Ovenden, 1990), while molecular studies dates some clades within the species back to 3.7 Ma (Allen *et al.*, 2012). A biogeographic analysis of the ancestral area of *O. digyna* pointed to an origin in the high mountains of southeast Asia, spreading both towards the east (into North America) and the west (into Europe) through

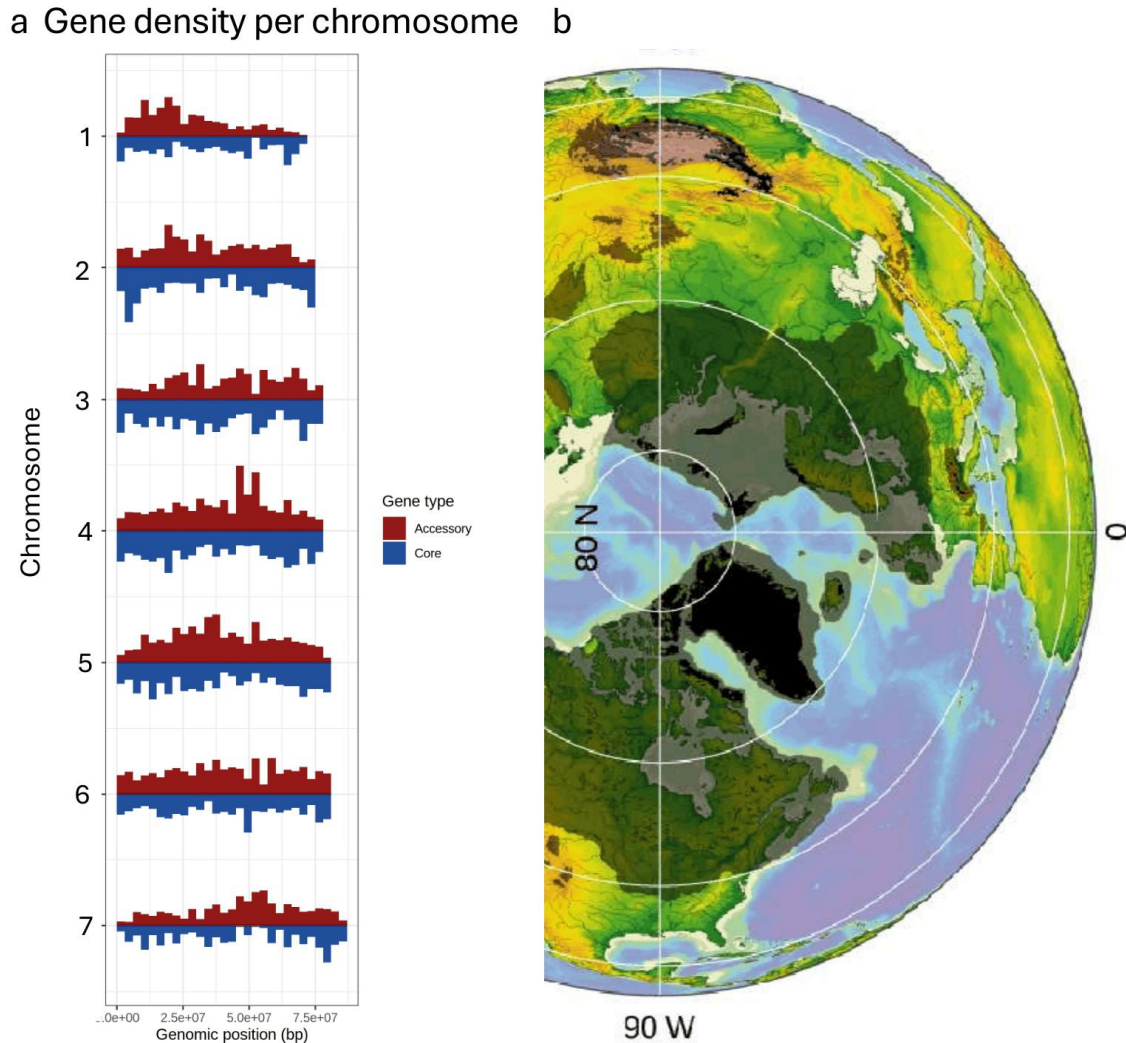


Fig. 1 a, Accessory- (red) and core (blue) gene densities across the seven *Oxyria digyna* (L.) Hill chromosomes. Height of staples indicate density at a genomic position on the chromosome. b, Last Glacial Maximum on the northern hemisphere between 90° W and 90° E longitude, shown as the shadowed area (image source: Grobe, 2007).

Russia, and eventually colonizing Greenland from both the east and west (Wang *et al.*, 2015). The analysis provides evidence for the colonization of Europe thorough Russia at ca. 1.96 Ma.

There are no previous studies on the diversity of European populations of *O. digyna*, but Wang *et al.* and Allen *et al.* present unresolved relationships for Europe populations in global phylogeographic studies of *O. digyna* (Allen *et al.*, 2012; Wang *et al.*, 2015) supplementary)

1.6 Aim and hypothesis

The aims of this study are to first infer a phylogenetic tree of European *Oxyria digyna* (L.) Hill., from an existing whole genome data set, and second to determine how phylogenetic relationships amongst populations vary when comparing regions enriched in core genes and regions enriched in accessory genes.

Hypothesis 1: Since glaciers covered large parts of Europe until relatively recently, I hypothesize that the early diverging lineages of *O. digyna* are from the southern mountain regions (Fig. 3a).

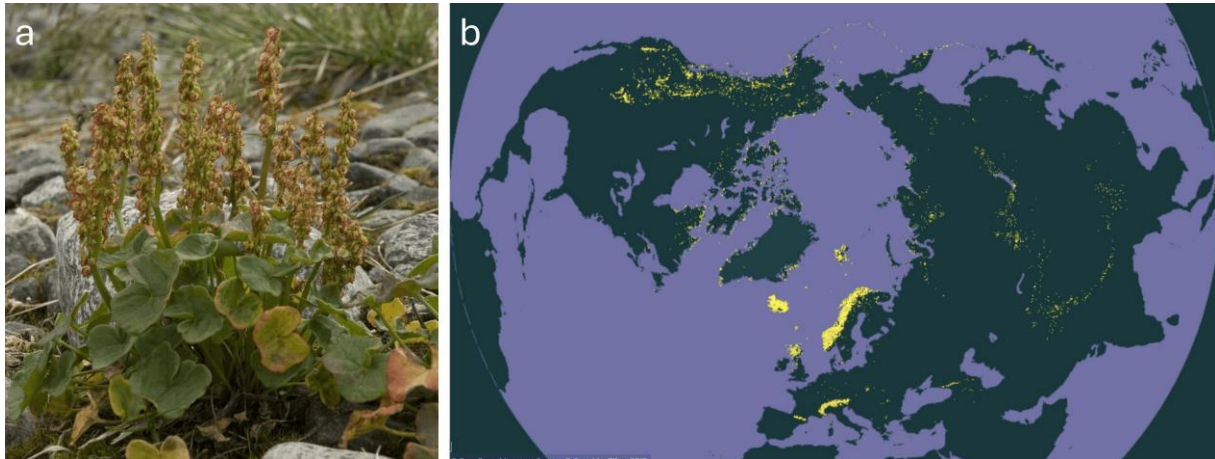


Fig 2. a, *Oxyria digyna* (L.) Hill (Polygonaceae; image source: van Kruijsbergen). b, Global distribution of *Oxyria digyna* (L.) Hill. Yellow dots indicate georeferenced records of the species (image source: Global Biodiversity Information Facility, 2016).

The more recently diverging lineages in the phylogeny are hypothesised to have colonized from the southern mountains to northern Arctic areas, following the retreat of the ice.

Hypothesis 2: Due to different recombination rates and selection pressures, I hypothesize that different phylogenetic relationships between the regions enriched in core- and accessory genes will be inferred.

1.7 Broader impact of the work

Studies of Arctic plant diversification are lacking, preventing a full understanding of their biogeographic history (Winkler *et al.*, 2012). Inferring the phylogeny of European *O. digyna* provides insight into the species evolutionary history and relationships between European populations, possibly as a response to past environmental change (Allen *et al.*, 2012; Wang *et al.*, 2015; Van Rossum *et al.*, 2018; Ruiz-García, 2021). As the Arctic is experiencing almost four times faster warming than global rates (Rantanen, 2022), studies of Arctic species is of high importance to improve our understanding of plant dispersal in relation to climate events.

Leveraging a pangenome, core- and accessory genes comparison can contribute to our understanding of multiple evolutionary paths within a genome, and exploring more complex phylogenetic relationships within a species (Huson, 1998; Huson & Bryant, 2006; Tranchant-Dubreuil *et al.*, 2019). The differences between gene types may inform about biologically meaningful diversity within populations (Loegler *et al.*, 2026; Soltis & Soltis, 2026). Further, the differences in the gene types may provide insight into the different evolutionary processes acting on different parts of the genome, such as recombination (Fig. 1a; Nachman, 2002)) and selection pressures (Golicz *et al.*, 2020; Tellini *et al.*, 2025).

2. Materials and methods

2.1 Data collection and sequencing

Leaves, fruits, and seeds from individuals of *Oxyria digyna* were collected from ten Arctic and alpine sites between the latitudes 46.41°N to 83.09°N in Europe (spanning ca. 40°s latitude; Table

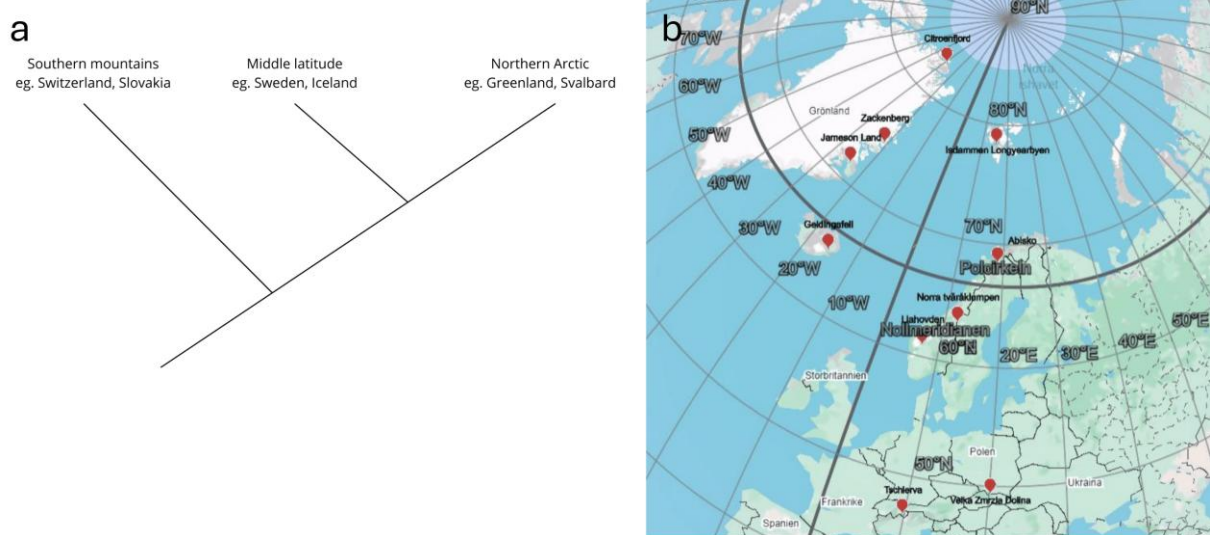


Fig. 3 a, Hypothesised diversification of *Oxyria digyna* (L.) Hill populations in Europe. b, Sample locations (red markers) of *Oxyria digyna* (L.) Hill in Europe, mapped in Google Earth (Google Earth, 2015-2021).

S1). Sampling locations included Slovakia (Veľká Zmrzlá dolina), Switzerland (Tschieriva), Sweden (Norra Tväråklumpen and Abisko), Norway (Isdammen Longyearbyen on Svalbard and Liahovden), Iceland (Geldingafell), and Greenland (Citronen Fjord, Jameson Land and Zackenberg; Fig. 3b; Table S1). In total, 103 individuals were sampled and sequenced, with 10-12 samples per sampling location. Genomic DNA was extracted. Sequencing libraries were prepared, and DNA sequencing was conducted at National Genomics Infrastructure (NGI), Uppsala. For sequencing, NovaSeq X Plus system, 25B flowcell, and XLEAP-SBS sequencing chemistry were used, for low-coverage whole genome sequencing. The sequencing yielded 4385 million (M) and 4255 M paired-end 150 base pairs read pairs in total. For each sample over 28 M read pairs ($\approx 15\times$ coverage) were obtained. The data was stored at National Academic Infrastructure for Super-computing in Sweden (Naiss, 2026).

2.2 Extraction of SNPs from the raw data and filter

The raw FASTQ files, a file format to store nucleotide sequences and the sequences' "probability of error in base calling" (Severin, 2026), were mapped to a reference genome. The reference genome came from one individual of *O. digyna*, collected from Alexandra Fiord on Ellesmere Island, Canada. Binary Alignment MAP (BAM) files were generated, containing the sequence alignments in binary format (Severin, 2026), and described where the short reads (150bp long) mapped to the reference genome. Single nucleotide polymorphisms (SNPs) were called from the BAM files for each individual, using GATK (v. 4.6.1.0; (Broad Institute, 2026)). SNPs were single nucleotide bases which differed between the sequenced individuals and the reference genome (Edwards *et al.*, 2007).

VCFtools (v. 0.1.16; Vcftools, 2015), and PLINK (v. 2.0-20231024-avx2; Purcell *et al.*, 2007; Chang, 2026) were used to filter out SNPs with high linkage. Linkage meant that genes or sequences close to each other on the chromosome were more likely to be inherited together, which could cause bias due to autocorrelation of sites (National Cancer Institute 2012; Wetterstrand, 2026). SNPs were also filtered for quality (GQ30). In the final dataset only biallelic SNPs were kept; indels, SNPs missing in $>10\%$ of the data and SNPs with a minor allele frequency of <0.01

were removed. A Variant Call Format (VCF; Severin, 2026) file contained the final 79,567 SNPs that remained after filtering.

2.3 Subset the data

2.3.1 Subset the SNPs by chromosome and create a midpoint root

As the filtered VCF file contained too much data to work with locally, the SNPs were subset by chromosome. Although large differences between the seven *O. digyna* chromosomes were not expected, subsetting the data by chromosome enabled me to explore any phylogenetic differences between the separate chromosomes. The filtered SNPs were subset by chromosome (one to seven) using bcftools (v.1.22; Danecek *et al.*, 2021), a program that processed and analysed sequence data. Each individual chromosome VCF file was run in TreeMix (v.1.13; Pickrell & Pritchard, 2012) without any migration edges ($m=0$). I inferred a midpoint root (Farris, 1972), for the TreeMix output files in newick format (“treeout”-files), for each of the seven chromosomes, as there was no data from a true outgroup available to root with. Midpoint rooting was used because an outgroup was missing (Hess & De Moraes Russo, 2007; Kinene *et al.*, 2016). The midpoint was placed between the two points furthest apart from each other, i.e. between the two longest branches (EMBL, 2026; Revell, 2026). I performed the midpoint rooting using the R programming language (v. 4.4.0; R Core Team, 2024) in R-studio (v. 2026.01.1-403; Posit team, 2025), using the phylogenetic analysis R packages ape (v. 5.8.1; Paradis & Schliep, 2019) and phytools (v. 2.4.4; Revell, 2024).

2.3.2 Subset the SNPs by core- and accessory gene enriched regions

As core genes occurred across species and accessory genes only within *O. digyna*, the total VCF file was subset into two files, one containing SNPs found in regions of the genome enriched in core genes and one in regions enriched in accessory genes. The regions enriched in core- and accessory genes spanned all seven *O. digyna* chromosomes. As previously mentioned (see 1.2), the pangenome with core- and accessory genes could vary in their definition (Tranchant-Dubreuil *et al.*, 2019). Here, core genes were defined as genes with orthologs in 20 distantly related species, from three different plant families (*Oxyria digyna*, *Polygonum aviculare*, *Rheum tangaticum*, *R. nobile*, *Fagopyrum tataricum*, and *F. esculentum* from Polygonaceae; *Capsella rubella*, *Brassica oleracea*, *Arabidopsis lyrata*, *Arabis alpina*, *Draba nivalis*, *Cochlearia groenlandica* and *Thlaspi arvense* from Brassicaceae; and *Rosa rugosa*, *Prunus persica*, *Malus sylvestris*, *Pyrus bretschneideri*, *Fragaria vesca*, *Dryas octopetala*, and *Argentina anserina* from Rosaceae). Accessory genes were instead genes unique to *O. digyna* (multiple haplotypes), but were not present in the other 19 species.

Core- and accessory genes were identified using the output from Orthofinder (Emms *et al.*, 2025). The core gene enriched regions were defined by comparing the Core:TotalGene counts to the ExpectedCore:Total in each 1 Megabase pair (Mbp) window. The same calculation was made to find the accessory gene enriched regions. The expected counts in each 1Mbp window of the genome were calculated based on the fraction of core- and accessory genes across the whole genome combined with the total number of genes in that window. Regions with more than double the amount of core genes observed than expected were defined as core enriched. In the same way, accessory gene enriched regions were defined as regions with more than double the number of expected accessory genes were observed. All SNPs were extracted from the regions enriched in core genes and the regions enriched in accessory genes. SNPs from core gene enriched regions and SNPs from accessory gene enriched regions were each run in TreeMix (v.1.13; Pickrell & Pritchard, 2012).

2.4 Create phylogenies based on core- and accessory genes

2.4.1 Create phylogenetic graphs in TreeMix

One rooted phylogenetic graph for core gene enriched regions, and one for accessory gene enriched regions were inferred based on the previously created midpoint, including the number of migration edges needed to explain the data, and the most frequently observed pattern of migration edges.

To root the phylogenies for the extracted SNPs with core gene enriched regions and accessory gene enriched regions, the trees were fixed in TreeMix (v.1.13; Pickrell & Pritchard, 2012) with the midpoint rooting for chromosome 1, performed previously (see 2.3.1). The SNPs of the regions enriched with core- and accessory genes was run in TreeMix with the midpoint root for 0 to 9 migration edges, as this interval was sufficient to capture the number of migration edges needed to explain the data in previous trials. TreeMix was run for 51 iterations of each number of migration edges (0-9). Each iteration was based on different subsets of the SNPs to avoid linkage and gave confidence for the inference of the number of migration edges.

To find the number of migration edges needed to explain the data, I calculated the *variance of relatedness between populations explained by the model* (Var(R)_expl) using R (v. 4.4.0; R Core Team, 2024) and following available code (Card, 2015). The 99,8% Var(R)_expl-method was first suggested by the TreeMix authors (Pickrell & Pritchard, 2012), and have come to be used in various TreeMix studies (e.g. Grimm *et al.*, 2007; Chen *et al.*, 2013; Decker *et al.*, 2014; Tremetsberger *et al.*, 2015; Soares & Freitas, 2024). For each number of migration edges, I calculated the mean for the Var(R)_expl, and the standard deviation for Var(R)_expl of the 51 iterations. The number of migration edges needed to explain the data was acquired by stopping to add more migration edges when 99,8% of the data was explained (Fitak, 2025), in this case when the mean Var(R)_expl reached 99.8%. Up until 99.8% Var(R)_expl the model became more accurate with every added migration edge, but after the 99.8% threshold, false migration edges might have been added (Pickrell & Pritchard, 2012).

For confidence, the most common graph variant within the 51 iterations was identified. This included the tree with: I, the most common topology, and II, containing the most common pattern of migration edges. Pattern of migration edges meant the direction and placement of all migration edges in a graph. The TreeMix output file containing the nexus tree ("treeout"-file) was used to compare the topologies with Robinson-Foulds distance (Robinson & Foulds, 1981) using R (R Core Team, 2024). Robinson-Foulds distance calculated the pairwise topological distance between trees, to find the number of nodes not shared (Robinson & Foulds, 1981; Schliep, 2011; Llabrés *et al.*, 2021). By using the TreeMix output file containing only the migration edge information ("edges"-file), the pattern of migration edges (placement and direction) was compared between the 51 iterations using R. The most frequently observed pattern of migration edges for core- respectively accessory enriched regions was calculated.

After having calculated the number of migration edges needed to explain the data, and identifying the most common topology and pattern of migration edges, one graph each for core- and accessory enriched regions was plotted. These two remaining graphs were compared to find the difference in topology (excluding migration edges). A normalized Robinson-Foulds distance (nRF) was calculated to find the number of splits which differed between the two trees, giving a value between 0 (the trees are identical) and 1 (the trees share no nodes; Schliep, 2011). nRF was performed in R, using the packages ape (Paradis & Schliep, 2019) and phangorn (Schliep, 2011). To find if the difference between core- and accessory enriched regions was greater than among the core- respectively accessory enriched regions, all 51 iterations of the number of migration edges needed to

explain the data was calculated for normalized Robinson-Foulds. The calculations were summarised as a mean of the iterations.

The two remaining graphs were also compared between migration edges (excluding topology). The difference in migration edges between the core- and accessory enriched regions was calculated in R using nRF.

2.4.2 Create phylogenetic networks in SplitsTree

I ran the SNP regions enriched in core- and accessory genes separately in the SplitsTree App (v.6.7.5 ;Huson & Bryant, 2024). The default workflow was used. First, P distances (Hamming, 1950) were calculated to create a distance matrix. P distances calculate distances based on normalized Hamming distances, which was based on character-state differences (Huson & Bryant, 2026). The distance matrix was made into an unrooted phylogenetic tree using Neighbour-joining (Saitou & Nei, 1987; Huson & Bryant, 2026). Lastly, a Neighbour net was created based on weighted splits, where splits of weight below zero were removed (Huson & Bryant, 2026).

3. Results

Identification of core and accessory genes on the seven chromosomes of *O. digyna* resulted in 41 regions enriched in core genes with 117 SNPs per Mbp (4836 SNPs / 41 Mbp), and 48 regions enriched in accessory genes with 153 SNPs per Mbp (7360 SNPs / 48 Mbp).

3.1 The phylogenetic trees based on chromosomes one to seven

The midpoint root was used on all seven chromosomes of *O. digyna*, and presented two different topologies. The phylogenetic trees of chromosomes one, three, six, and seven show the same topology, with populations from West Greenland (Citronen Fjord) and southern Europe (Tschierva, Switzerland and Veľká Zmrzlá dolina, Slovakia) separated from a clade comprised of the populations from East Greenland (Jameson Land and Zackenberg), Norway (Liahovden and Isdammen Longyearbyen, Svalbard), Sweden (Norra Tväråklumpen and Abisko), and Iceland (Geldingafell; Fig. 4a). The phylogenetic trees for chromosomes two, four, and five inferred populations from West Greenland, central Europe, and Svalbard as one clade, with the rest of the populations in a separate clade (Fig. 4b). Apart from the position of the Svalbard population, a similar topology was seen in all seven trees after the first diverging lineages. East Greenland populations diverge first, with Zackenberg then Jameson Land. Also, Sweden, and mainland Norway (Liahovden), and Iceland were clustered in a clade. Within the topology of the Sweden-, Norway-, and Iceland populations, Iceland was inferred as the early divergent population in all seven trees. For the other three populations, trees showed varying topologies between the chromosomes.

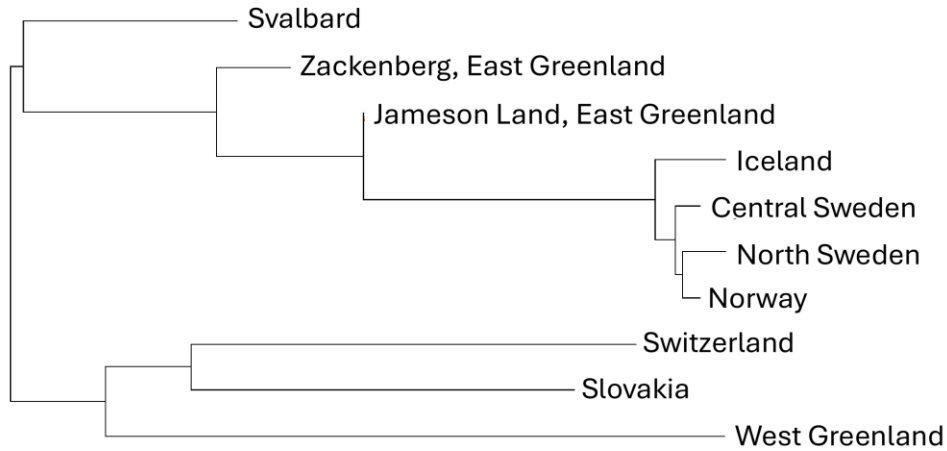
3.2 The phylogenetic graphs based on core- and accessory enriched regions

3.1.1 Migration edges needed to explain the data

The core enriched graphs needed four migration edges to explain the data (mean Var(R)_expl: 0.9984, standard deviation (SD): 0), and the accessory enriched trees needed three migration edges to explain the data (mean Var(R)_expl: 0.9989, SD: 0.0009).

Between the 51 iterations, the pattern (direction and placement) and migration weight of the migration edges differ. The most frequently observed migration edge pattern for the core enriched regions occurred 12 times out of the 51 iterations (24%), with four migration edges (m=4). All four

a Midpoint root of chromosome 1



b Midpoint root of chromosome 2

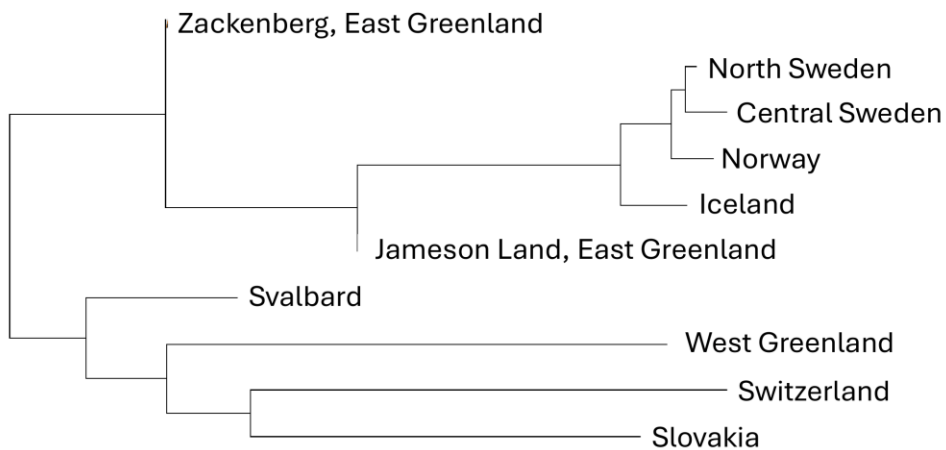
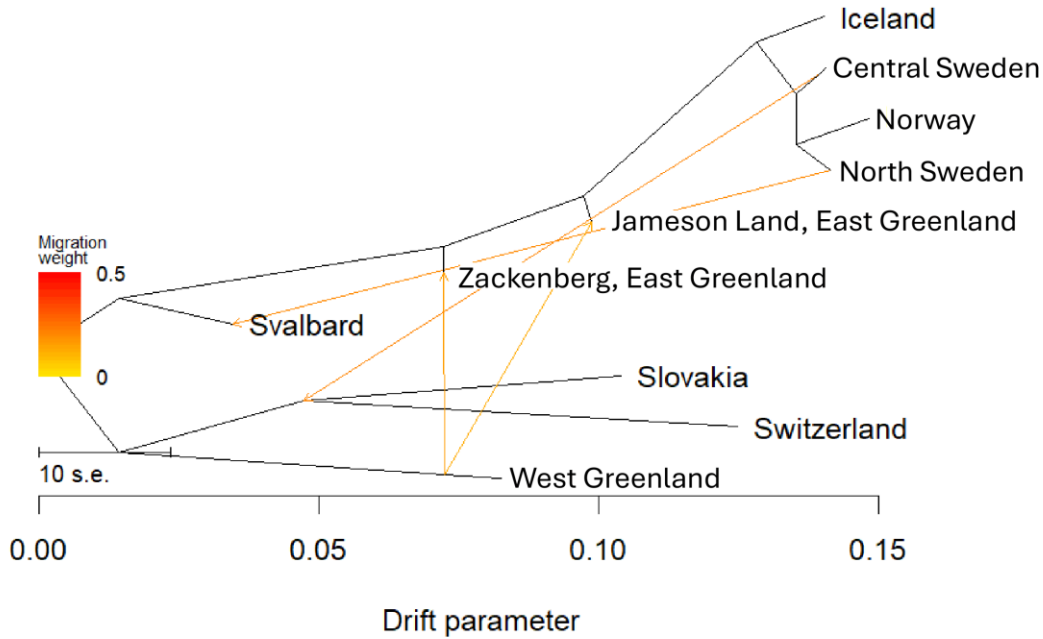


Fig. 4 Phylogenetic trees made in TreeMix (Pickrell & Pritchard, 2012), and visualised in R (R Core Team, 2024) of the topology and phylogenetic relationships among 10 European populations of *Oxyria digyna* (L.) Hill. a, Midpoint rooted phylogenetic tree based on chromosome 1 in *O. digyna*. b, Midpoint rooted phylogenetic tree based on chromosome 2 in *O. digyna*.

migration edges in the most frequently observed pattern had low to intermediate migration weight (0-0.25). Two migration edges pointed from the West Greenland population to each East Greenland populations, one pointed from the central Sweden population (Norra Tväråklumpen) to the node dividing Switzerland and Slovakia, and one was pointing from Northern Sweden (Abisko) to Svalbard, Norway (Fig 5a). Between the iterations, three different less frequently observed patterns occurred six times (12%) each, with most patterns occurring one to three times (2-6%) each (Fig S1).

The most frequently observed pattern of migration edges for the accessory enriched regions occurred five times (10%), with three migration edges ($m=3$). The migration edge with the highest weight (ca. 0.5 migration weight) pointed from the node splitting the Switzerland- and Slovakia populations to the West Greenland population (Fig. 5b). The two other migration edges had a lower migration weight (ca. 0.25), and pointed from an early point of the West Greenland population

a Phylogenetic graph of core enriched regions



b Phylogenetic graph of accessory enriched regions

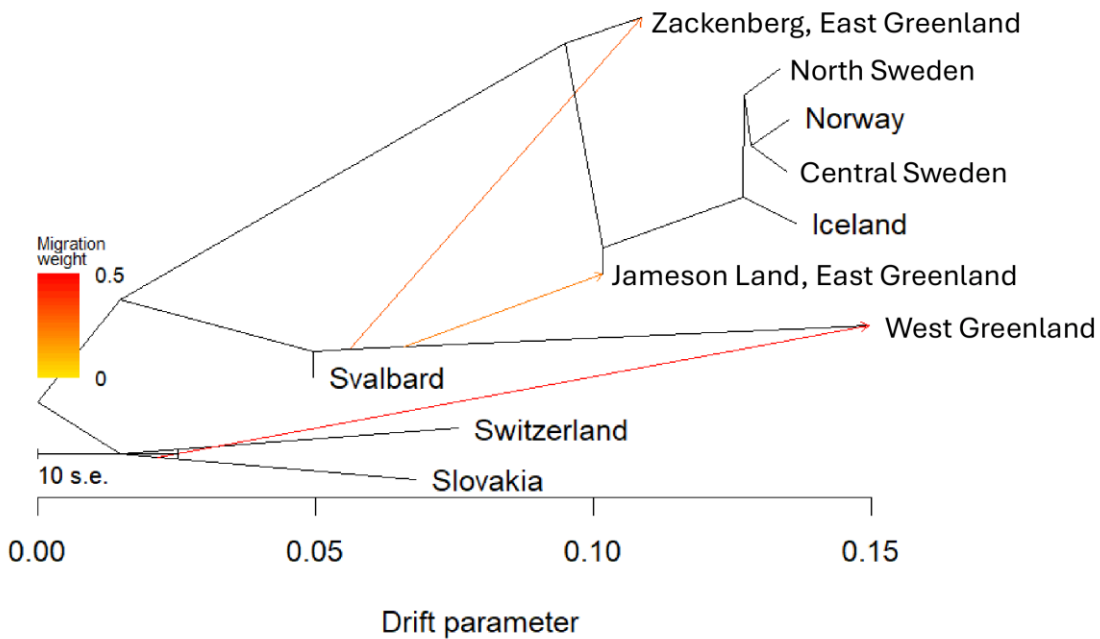


Fig. 5 Midpoint rooted phylogenetic graphs with migration edges (coloured arrows) for European *Oxyria digyna* (L.) Hill., created in TreeMix (Pickrell & Pritchard, 2012) and visualised in R (R Core Team, 2024). Migration edges are alternative parental branches, with migration weight indicating how strong the relationship is between the points of the arrow (Pickrell & Pritchard, 2012). Red indicates a stronger relationship, and yellow a weaker relationship. The phylogenetic trees each have the amount of migration edges needed to explain the data, and the most common migration edge pattern (across 51 iterations). Drift parameter refer to the rate of random drift along the phylogeny (Catalán et al., 2019). a, Graph based on core enriched regions on the chromosomes, with 4 migration edges ($m=4$). b, Graph based on accessory enriched regions of the chromosome, with 3 migration edges ($m=3$).

branch to each East Greenland population (Fig. 5b). Less frequently observed patterns occurred 1 to 2 times (2-4%) each (Fig. S1).

3.1.2 Comparison between the phylogenetic graphs for regions enriched in core- and accessory genes

The difference between graphs with core- and accessory enriched regions were calculated using nRF, and is presented as a number between 0 (the splits are the same) and 1 (the splits are completely different; (Schliep, 2011)). The topology of core- and accessory enriched graphs (excluding migration edges) shared 62% of their splits (nRF=0.375). The nRB showed a greater difference between the best core- and accessory enriched trees, than between the respective tree iterations (Core enriched region nRFmean=0, Accessory enriched region nRFmean=0.15). For the comparison of migration edges, core enriched regions shared no migration edges with accessory enriched regions (nRF=1).

3.3 Phylogenetic network in SplitsTree for regions enriched in core- and accessory genes

The core enriched network inferred more reticulations than the accessory enriched network (Fig. 6 a and b), indicating more conflicts in the data of the core enriched regions (Chen *et al.*, 2013). For instance, in the accessory enriched network, a single path is seen from the centre of the network (East Greenland populations) to the populations in Sweden, mainland Norway, and Iceland (Fig. 6 b). For the same region in the core enriched network, there were many reticulations (Fig. 6 a). The reticulations were both within said populations, but also the Iceland population showed reticulations to the Jameson Land population.

4. Discussion

Despite previous phylogenetic studies on *Oxyria digyna* (L.) Hill, understanding of relationships within and amongst European populations are limited (but see Wang *et al.*, 2015; Allen *et al.*, 2012). Classic theory and studies on dispersal patterns of Arctic-Alpine species hypothesise plants migrated from southern- to northern parts of Europe following retreating glaciers after the Last Glacial Maximum (Hultén, 1937; Hultén, 1962; Darwin, 2011; Makhrov *et al.*, 2019). Here, the phylogeny of European *O. digyna* was inferred, and determined that southern European populations diverged earlier than the more northern populations (with the West Greenland population as an exception). Additionally, the relationships of core- and accessory genes were compared. With advances in sequencing methods, it have been established that different regions of the genome, for instance core- (conserved across a range of species) and accessory (unique to *O. digyna*) genes, were hypothesized to infer different phylogenetic relationships. Here, the difference is hypothesised to be a result of different selection pressures (Tranchant-Dubreuil *et al.*, 2019; Golicz *et al.*, 2020; Hu *et al.*, 2022; Tellini *et al.*, 2025; Soltis & Soltis, 2026) and position on the genome (Fig. 1a). Core and accessory genes differed from each other in their phylogenetic relationships, in terms of topology, migration events, and reticulations.

4.1 Phylogenetic relationships within *Oxyria digyna*

4.1.1 The relationships based on the midpoint rooted chromosome one tree

The midpoint rooted, chromosome one phylogenetic tree is here used to represent the “base” relationships of European *O. digyna*, without including gene flow in the inference (Fig. 4a). This base

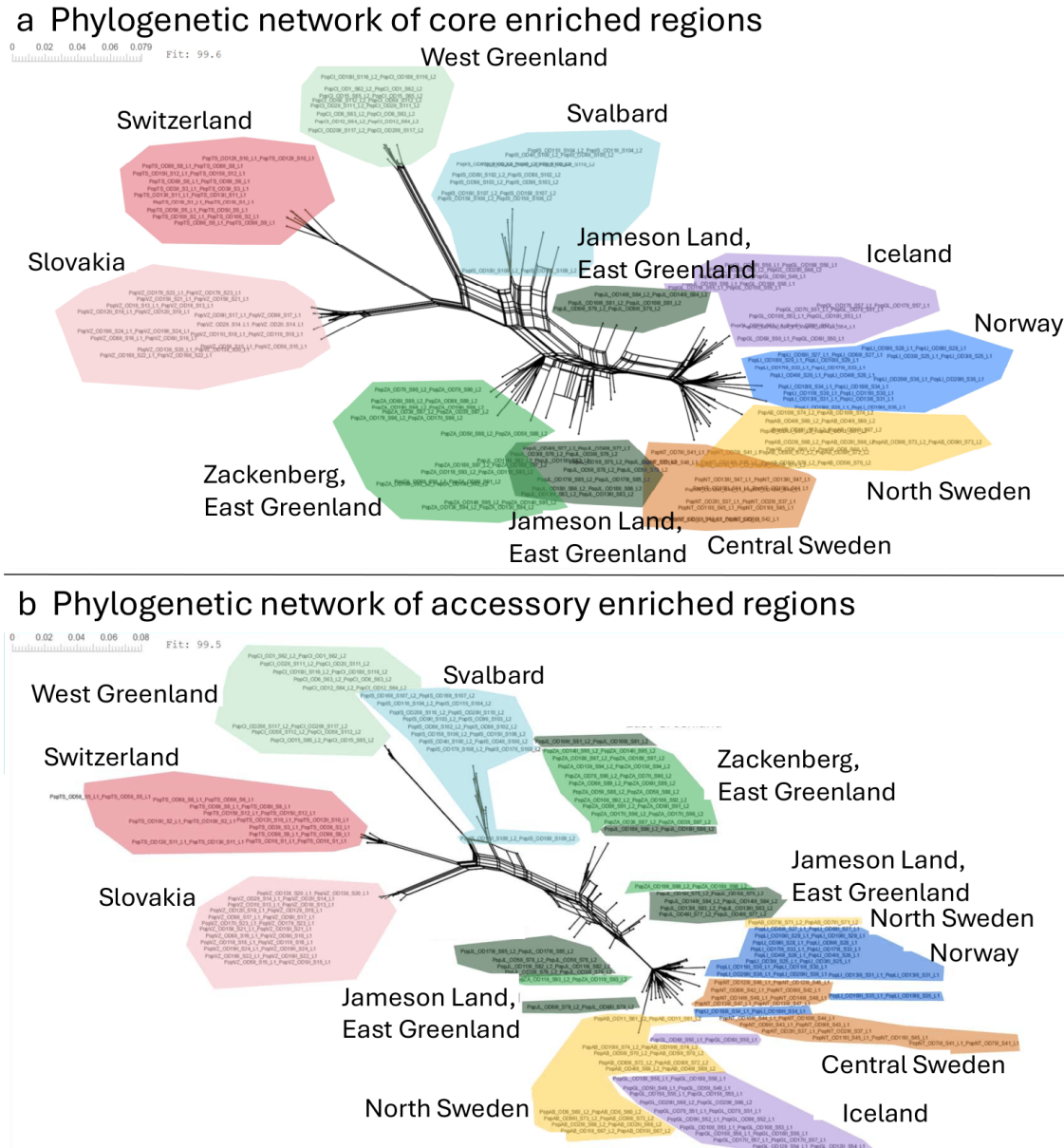


Fig 6 Split network of European *Oxyria digyna* (L.) Hill individuals relationship, created in SplitsTree (Huson & Bryant, 2024). Reticulations indicate conflicts in the data. Fit is how similar the input distance is to the network distance, given in percent (%; Bryant & Huson, 2023). The network is coloured by population. *a*, Split network based on the core gene enriched regions on the chromosomes. *b*, Split network based on the accessory gene enriched regions of the chromosomes.

topology is consistent with the hypothesis, with earlier diverging lineages in southern populations. The West Greenland population is an exception from the hypothesis as it diverges early but is in the north. Later diverging are the more northern populations in the base topology.

The southern lineages diverging early agrees with literature and the classic theory behind Arctic-alpine establishment (Hultén, 1937; Hultén, 1962; Darwin, 2011; Makhrov *et al.*, 2019). As explained in recent literature reviews, the more recent divergence of the more northern populations can likely be explained by the retreating ice of the relatively recent glaciations, enabling species to establish in previously glaciated areas (Makhrov *et al.*, 2019; Brochmann *et al.*, 2025). Previous studies on *O. digyna* have not considered population level relationships among Europe. Both Wang *et al.* (2015) and Allen *et al.* (2012) included European populations (for instance Greenland,

Iceland, Norway, Switzerland, and Slovakia), in their studies on global phylogeography of *O. digyna*, but the relationships between European populations are unresolved. Thus, this thesis contributes with an increased understanding of relationships within European *O. digyna*.

The West Greenland population is not consistent with the hypothesis of southern lineages colonising the northern regions, as it is early diverging and clustered with the southern European populations (Fig. 4a). The deviation from the hypothesis could be a consequence of the colonization of Greenland, with establishment both from the west (North America), and from the east (Europe), as suggested by (Wang *et al.*, 2015). The West Greenland population may therefore, in contrast to the East Greenland populations, have its ancestry in North America, although that cannot be identified here as only European data have been considered. If the West Greenland population has its ancestry in North America, it could be the reason the population is early diverging in the phylogeny. Samples from North American populations are needed to give a better understanding of the relationship between the West Greenland population to the other European populations.

The topology of the more northern populations (East Greenland, Sweden, Norway, and Iceland) was also hypothesized to follow a latitudinal gradient, but this was not found. Instead, the East Greenland populations (latitude 71,397°N and 74,465°N) diverged before the populations from Sweden, mainland Norway, and Iceland (latitude 63,208°N to 68,360°N; Table S1). In a study on phylogeography of primarily North American *O. digyna*, Allen *et al.* (2012) suggests that dispersal to new areas was rapid after glacial retreat, as *O. digyna* have been shown to have moderately good dispersal potential (Tackenberg & Stöcklin, 2008). This could have enabled fast wind dispersal to Greenland from continental Europe (Allen *et al.*, 2012). Allen *et al.* (2012) also suggest alternative ways of dispersal after the retreat of the glaciers, from unglaciated areas in the north. Area of dispersal after glaciation could cause a different phylogenetic relationship between the European *O. digyna* populations than hypothesised.

4.1.2 Divergent relationships between the seven chromosomes

The phylogenies of the seven chromosomes showed two alternative midpoints. The alternative midpoints may be explained by incomplete lineage sorting, meaning that alleles did not diverge at the same time as speciation occurred (Nichols, 2001). The different time of diversification in alleles and species can give different branching patterns for phylogenetic trees (Nichols, 2001), making it more difficult to create an accurate species tree (Korstian *et al.*, 2022). In species with fast diversification, incomplete lineage sorting is expected, as there have been less time for complete sorting of ancestral polymorphisms (Korstian *et al.*, 2022; Yisilam *et al.*, 2026). *Oxyria digyna* may have experienced rapid diversification after establishment in the Arctic, following the retreat of the glaciers after the relatively recent ice cover, which would agree with the hypothesis. Similar results can also come from ancient hybridisation or introgression (Yisilam *et al.*, 2026), and testing between these drivers would require analysis of the rate of diversification.

The different midpoint roots between the chromosomes were not considered for further analysis of the relationships among European *O. digyna*, as all chromosomes initially were expected to have the same midpoint. The relationships within the midpoint rooted trees should therefore be interpreted with caution. Further studies on chromosomes of *O. digyna* should consider the different midpoints when creating a root.

4.2 Phylogenetic differences between core- and accessory genes

4.2.1 Phylogenetic difference based on core- and accessory genes

The difference between the phylogenetic graphs of core- and accessory enriched regions testifies for different evolutionary relationships of the core- and accessory genes, as hypothesised. Both the topology and migration edge pattern differed. This main difference is probably a result of the different types of selection pressure of the gene types. The core genes are more conserved and thus likely involved in essential cellular functions, which is likely under purifying selection (Golicz *et al.*, 2020). Accessory genes are, in contrast, suggested to be involved in traits related to local adaptation (Tranchant-Dubreuil *et al.*, 2019; Tellini *et al.*, 2025; Soltis & Soltis, 2026) and under positive selection.

The migration edge patterns (placement and direction) were different for the graphs of core- and accessory genes. No comparison was made to see if the core- and accessory genes migration edges differed more than expected, as in the case of the topology (excluding migration edges; see 3.1.2). Additionally, there was a lot of variation observed between iterations, with a range of alternative migration edge patterns (fig S1), also making the difference between core- and accessory genes less confident. Despite this, here I present a suggested difference between migration edges in core- and accessory enriched regions that could be studied further.

4.2.2 Core genes show more migration edges and are more consistent than accessory genes

The graph for regions enriched in core genes require more migration edges to explain the data ($m=4$; Fig. 5a) than the graph for accessory enriched regions ($m=3$; Fig. 5b), although they only differ with one migration edge. Initially we hypothesized that core genes would follow a more bifurcating topology because of their essential cellular functions making them more constrained with less geneflow (Tranchant-Dubreuil *et al.*, 2019; Golicz *et al.*, 2020; Tellini *et al.*, 2025), and accessory enriched regions, being more locally adaptive, would follow more of a network-like relationship to account for adaptive gene flow (Tranchant-Dubreuil *et al.*, 2019; Hu *et al.*, 2022; Tellini *et al.*, 2025; Soltis & Soltis, 2026). However, more migration edges means that more additional parental lineages are needed to explain the data, and can suggest more gene flow (Pickrell & Pritchard, 2012).

The patterns of migration edges in the two phylogenetic graphs are not representative of all 51 iterations within each gene enriched region, but only the most common migration pattern found across each 51 iterations. Although the core enriched regions need more migration edges to explain the data, the pattern is more consistent across the 51 iterations of the core enriched regions. Comparing the core enriched regions migration edge patterns to the accessory enriched regions, there are fewer different migration edge patterns for the core enriched regions. This agrees with the hypothesis of core genes being more consistent because of their essential cellular functions (Tranchant-Dubreuil *et al.*, 2019; Golicz *et al.*, 2020; Tellini *et al.*, 2025). The consistent patterns in core genes may mean that the migration edges observed are true historical migration events that have been conserved thorough time. More variation in the accessory enriched regions also agrees with accessory genes being involved in local adaptation (Tranchant-Dubreuil *et al.*, 2019; Hu *et al.*, 2022; Tellini *et al.*, 2025; Soltis & Soltis, 2026).

4.2.3 Core genes show more reticulations than accessory genes

The phylogenetic networks are not consistent with the hypothesis, expecting population relationships in core enriched regions to follow the structure of a bifurcating tree, and relationships in the accessory enriched regions to follow the structure of a network. Instead core enriched regions had

more reticulations (Fig. 6a) than the accessory enriched regions (Fig 6b). The reticulations in a SplitsTree network can be caused by recombination (Huson & Bryant, 2006). This is likely the cause for the network like structure of the core enriched regions, as these regions were more often found close to the telomeres on the chromosome (Fig. 1a) with a lot of recombination (Nachman, 2002). The accessory genes could instead be found close to the centromere (Fig. 1a), with little recombination and a lot of linkage (Nachman, 2002), which can explain less of a network-like relationship in accessory enriched regions.

Because of a high recombination rate around the telomeres (Nachman, 2002), the core enriched regions may have experienced recombination before establishing in Sweden, mainland Norway, and Iceland, creating variants. As the core genes are under strong purifying selection (Golicz *et al.*, 2020), the variant will not change a lot, thus being maintained and therefore conserved across the populations. Conserved core genes in the populations can be the reason behind a network reticulations connecting Sweden, mainland Norway, and Iceland to the rest of the network. Tellini *et al.* (2025) hypothesize accessory genes to be “evolutionary accelerators” for adaptive radiation. Adaptive radiation is when an individual establishes in a new environment, and diversify rapidly (Tellini *et al.*, 2025). Because of the variation and adaptability in accessory genes, they can easily colonise and adapt in new places. Once established, they can, unlike the core genes, freely diversify and expand the individual/population traits, which could be why we for the accessory gene network instead see a single path connecting the populations. This may also be related to the retreating glaciers after Pleistocene ice ages, causing rapid colonisation of species into uninhabited areas (Hultén, 1937; Hultén, 1962; Darwin, 2011; Makhrov *et al.*, 2019). Alternatively, as the accessory genes are more adaptive (Tranchant-Dubreuil *et al.*, 2019; Hu *et al.*, 2022; Tellini *et al.*, 2025; Soltis & Soltis, 2026), so many variants likely exist that it is not recorded in the SplitsTree network.

Selection may favour core genes closer to the telomeres, as recombination enables purifying selection to get rid of deleterious mutations (Aeschbacher & Bürger, 2014). Due to the proximity to the centromere, accessory genes have less recombination which possibly slows down removal of maladaptive alleles (Nachman, 2002). Despite this, the location of accessory genes on the chromosome may facilitate the individual’s survival in the local environment. With gene flow and high recombination, alleles beneficial in other local environments may be mixed into the accessory genes, making the genes less adapted to their own local environment (Felsenstein, 1981; Aeschbacher & Bürger, 2014). Previous phylogenetic studies have not considered the placement of the core- and accessory genes along the chromosome, despite the different phylogenetic relationships the placement can result in, as suggested in this thesis. However, in a book chapter on genome reticulations, Edwards *et al.* (2017) suggest local adapted alleles to accumulate in low recombination regions, for instance near centromeres, referring to a previous review on evolution on chromosome inversion by Kirkpatrick (2010). This agrees with the accessory genes (which are suggested to be locally adaptive) location close to the centromere described in this thesis. Further, a computational study on recombination rates in several species of bacteria have shown core genes to have more recombination compared with accessory genes (Preska Steinberg *et al.*, 2022), again agreeing with the location of core genes close to the telomeres as found here, but it may be inappropriate to draw parallels between plant and bacteria, as bacterial recombination includes incorporating genes from outside the cell (Segerman, 2012).

4.3 Limitations, further studies, and broader implications

4.3.1 Limitations and further studies

Despite my results pointing towards a difference between core- and accessory genes, for the bifurcating topology, the migration edges, and number of reticulations, no tests for significance have been made. Therefore, all results should be taken with caution. Despite no test for significance, this

thesis presents new ideas that can be further explored. Following are some methods that could be tested to get more reliable results: I, Topologies (excluding gene flow) from the phylogenetic graphs created in TreeMix were compared within the 51 iterations to see if core- and accessory enriched regions differed more between each other than expected. The migration edges can be compared in a similar way to see if they differ more than expected. The difference within and between should also have been compared for significance, with H_0 = no difference between the differences within iterations and between core- and accessory enriched regions. II, The phylogenetic networks created in SplitsTree should also be compared statistically, not just visually, by calculating the number of reticulations in each graph. III, Although probably difficult, all iterations for the migration edges (core $m=4$, accessory $m=3$) in the phylogenetic graphs can be included in a comparison between core- and accessory enriched regions, since those include many alternative topologies (Fig. S1).

4.3.2 Broader implications

Studying the phylogenetic relationship of *O. digyna* using high resolution SNPs from different regions of the genome gives insight into variation in the pattern of divergence in the species. Despite the rapid warming of the Arctic (Rantanen, 2022), studies on Arctic plant diversification have been lacking (Winkler *et al.*, 2012). This thesis gives insight into the phylogenetic history of *O. digyna*, possibly affected by past climate events (Allen *et al.*, 2012; Wang *et al.*, 2015; Van Rossum *et al.*, 2018). It gives a deepened understanding of *O. digyna*'s within-species relationship and may suggest how other Arctic-alpine plants are related within Europe.

By also looking at the phylogenetic relationships for different genes, I present further understanding of how the genome can show distinct evolutionary relationships depending on what part of it is studied. When working with genome-wide data, core- and accessory genes should be taken into consideration because of the suggested differences between conserved and variable sites. The differences between the core- and accessory genes have also provided insight into possible evolutionary forces acting on different parts of the genome, driving the differentiation in functions of the genes. For instance, recombination and selection pressure seem to affect the different phylogenetic histories in the genes. A broader knowledge of how different evolutionary forces contribute to diversification gives us better understanding of how genomic evolution works. Understanding the genomics behind diversification is of importance for evolutionary biology (Tellini *et al.*, 2025).

4.4 Conclusion

Using whole genome sequence data from 11 populations, I suggest a higher resolution phylogeny, and possible migration routes of the European *O. digyna* populations. My results also point towards a difference between core- and accessory gene phylogenetic relationships. This is possibly explained by core genes occurring in regions of high recombination and being conserved while accessory genes occurred in low recombination regions and may be more locally adaptive. Understanding the genomics behind diversification is of importance for evolutionary biology. When working with genomic data, differing recombination and selection pressures in genes should be taken into consideration, to get a more complete view on genetic variation across the genome. The field of core- and accessory genes needs to be studied further to see if these phylogenetic differences are shared among more species, and what evolutionary factors are driving them.

5. Code availability

R code used by me to visualise TreeMix phylogenies, midpoint root, find number of TreeMix migration edges needed to explain the data, find most common topology and pattern (placement and direction) of migration edges, and compare topologies and migration edges with normalised Robinson-Foulds distance is available at https://github.com/sofiekedback/Masterthesis_Oxyria_phylogeny.git.

6. Acknowledgements

I want to thank my supervisor, Christine Bacon, for helping me develop this project, and for her support with the workflow, analysis of the results, and improvements of this thesis, throughout the project; Cassandra Elpinstone for helping with the SNP-data, running software to generate files for analysis, support with knowledge in the field, and providing feedback on this thesis; Elizabeth Stunz and everyone involved in collecting and sequencing the data, as well as giving helpful input on the project.

7. References

- Aeschbacher S, Bürger R. 2014. The Effect of Linkage on Establishment and Survival of Locally Beneficial Mutations. *Genetics* **197**(1): 317-336.
- Allen GA, Marr KL, McCormick LJ, Hebda RJ. 2012. The impact of Pleistocene climate change on an ancient arctic-alpine plant: multiple lineages of disparate history in *Oxyria digyna*. *Ecology and Evolution* **2**(3): 649-665.
- Artdatabanken S 2026. Artfakta: fjällsyra (*Oxyria digyna*).
- Broad Institute. 2026. GATK Genome Analysis Toolkit. gatk.broadinstitute.org.
- Brochmann C, Elphinstone C, Birkeland S, Ikeda H, Eidesen PB, Alsos IG, Westergaard KB. 2025. Phylogeography of Arctic plants: where are we after 35 years, and where to go? *Plant Ecology & Diversity* **18**(3-4): 1-25.
- Bryant D, Huson DH. 2023. NeighborNet: improved algorithms and implementation. *Front Bioinform* **3**: 1178600.
- Card D. 2015. RADpipe/treemixVarianceExplained.R. *GitHub*.
- Catalán A, Briscoe AD, Höhna S. 2019. Drift and Directional Selection Are the Evolutionary Forces Driving Gene Expression Divergence in Eye and Brain Tissue of *Heliconius* Butterflies. *Genetics* **213**(2): 581-594.
- Chang C. 2026. PLINK 1.9 beta. www.cog-genomics.org.
- Chen S, Kim D-K, Chase MW, Kim J-H. 2013. Networks in a Large-Scale Phylogenetic Analysis: Reconstructing Evolutionary History of Asparagales (Lilianaee) Based on Four Plastid Genes. *PLOS ONE* **8**(3): e59472.
- Clark P, Dyke A, Shakun J, Carlson A, Clark J, Wohlfarth B, Mitrovica J, Hostetler S, McCabe A. 2009. The Last Glacial Maximum. *Science (New York, N.Y.)* **325**: 710-714.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2).
- Darwin C. 2011. *On the Origin of Species [The Origin of Species]*: HarperCollins publishers Ltd.
- Davinack D. 2024. The Pleistocene Epoch. [Wheatoncollege.edu](https://wheatoncollege.edu).
- Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, et al. 2014. Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLOS Genetics* **10**(3): e1004254.

- Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009.** Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**(24): 3207-3212.
- Edwards D, Forster JW, Chagné D, Batley J 2007.** What Are SNPs? In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN eds. *Association Mapping in Plants*. New York, NY: Springer New York, 41-52.
- Edwards SV, Potter S, Schmitt CJ, Bragg JG, Moritz C 2017.** Chapter: 9 Reticulation, Divergence, and the PhylogeographyPhylogenetics Continuum. In: Avise JC, Ayala FJ eds. *In the Light of Evolution*: National Academies Press, 171-190.
- Ehlers J, Gibbard PL. 2007.** The extent and chronology of Cenozoic Global Glaciation. *Quaternary International* **164-165**: 6-20.
- Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffaari A, Hickey G, Chang X, Seaman JD, Rounthwaite R, Ebler J, et al. 2020.** Pangenome Graphs. *Annu Rev Genomics Hum Genet* **21**: 139-162.
- EMBL. 2026.** Root | Phylogenetics. *EMBL-EBI*.
- Emms DM, Liu Y, Belcher L, Holmes J, Kelly S. 2025.** OrthoFinder: scalable phylogenetic orthology inference for comparative genomics. *bioRxiv*.
- Farris JS. 1972.** Estimating Phylogenetic Trees from Distance Matrices. *The American Naturalist* **106**(951): 645-668.
- Felsenstein J. 1981.** Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution* **35**(1): 124-138.
- Fitak R. 2025.** OptM: Estimating the Optimal Number of Migration Edges from 'Treemix'. *rdrr.io*.
- Global Biodiversity Information Facility. 2016.** *Oxyria digyna* (L.) Hill. *Gbif.org*.
- Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. 2020.** Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. *Trends Genet* **36**(2): 132-145.
- Google Earth 2015-2021.** 70°50'24.70"N 12°32'23.66"E.
- Grimm GW, Renner SS, Stamatakis A, Hemleben V. 2007.** A nuclear ribosomal DNA phylogeny of acer inferred with maximum likelihood, splits graphs, and motif analysis of 606 sequences. *Evol Bioinform Online* **2**: 7-22.
- Grobe H 2007.** Recent (black) and maximum (grey) glaciation of the northern hemisphere during the Quaternary climatic cycles Ice coverage. Wikipedia commons.
- Hamming RW. 1950.** Error Detecting and Error Correcting Codes. *Bell System Technical Journal* **29**(2): 147-160.
- Hess PN, De Moraes Russo CAA. 2007.** An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* **92**(4): 669-674.
- Hewitt G. 2000.** The genetic legacy of the Quaternary ice ages. *Nature* **405**(6789): 907-913.
- Hillis DM, Heller HC, Hacker SD, Hall DW, Laskowski MJ, Sadava DE. 2020.** *Life: The Science of Biology*: Macmillan Higher Education.
- Hoffmann MH, Röser M. 2009.** Taxon recruitment of the arctic flora: an analysis of phylogenies. *New Phytologist* **182**(3): 774-780.
- Hu H, Scheben A, Verpaalen B, Tirnaz S, Bayer PE, Hodel RGJ, Batley J, Soltis DE, Soltis PS, Edwards D. 2022.** Amborella gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation. *New Phytol* **233**(4): 1548-1555.
- Hultén E. 1937.** *Outline of the History of Arctic and Boreal Biota During the Quarternary Period*: Bokförlags Aktiebolaget Thule.
- Hultén E. 1962.** *The Circumpolar Plants. I Vascular Cryptograms, Conifers, Monocotyledons*: Almquist & Wiksell.
- Hultén E. 1971.** *The Circumpolar Plants. II Dicotyledons*: Akmqvist & Wiksell.
- Huson D, Bryant D. 2024.** The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. *Nature Methods* **21**.
- Huson DH. 1998.** SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**(1): 68-73.

- Huson DH, Bryant D. 2006.** Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23**(2): 254-267.
- Huson DH, Bryant D. 2026.** SplitsTree App User Manual SplitsTree App (version 6.7.4, built 18 Feb 2026).
- Huson DH, Scornavacca C. 2011.** A Survey of Combinatorial Methods for Phylogenetic Networks. *Genome Biology and Evolution* **3**: 23-35.
- Kinene T, Wainaina J, Maina S, Boykin LM. 2016.** Rooting Trees, Methods for. *Encyclopedia of Evolutionary Biology*: 489–493.
- Kirkpatrick M. 2010.** How and Why Chromosome Inversions Evolve. *PLOS Biology* **8**(9): e1000501.
- Korstian JM, Paulat NS, Platt RN, 2nd, Stevens RD, Ray DA. 2022.** SINE-Based Phylogenomics Reveal Extensive Introgression and Incomplete Lineage Sorting in *Myotis*. *Genes (Basel)* **13**(3).
- Llabrés M, Rosselló F, Valiente G. 2021.** The Generalized Robinson-Foulds Distance for Phylogenetic Trees. *J Comput Biol* **28**(12): 1181-1195.
- Loegler V, Friedrich A, Schacherer J. 2026.** Dynamics of genome evolution in the era of pangenome analysis. *Cell Genom* **6**(1): 101067.
- Lou RN, Jacobs A, Wilder AP, Therkildsen NO. 2021.** A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology* **30**(23): 5966-5993.
- Makhrov AA, Bolotov IN, Artamonova VS, Borovikova EA. 2019.** Mountain Regions are the Putative Place of Origin of Many Arctic Animal and Plant Forms. *Зоологический журнал* **98**(11): 1291-1303.
- Matthews JJV, Ovenden LE. 1990.** Late Tertiary Plant Macrofossils from Localities in Arctic/Subarctic North America: A Review of the Data. *ARCTIC* **43**(4): 384-392.
- Meltofte H. 2013.** *Arctic Biodiversity Assessment. Status and Trends in Arctic Biodiversity: Synthesis.*
- Nachman MW. 2002.** Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics & Development* **12**(6): 657-663.
- Naiss. 2026.** NAISS | The National Academic Infrastructure for Supercomputing in Sweden. *Naiss.se*.
- National Cancer Institute 2012.** <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/linkage>. *NIH National cancer institute*.
- Nichols R. 2001.** Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16**(7): 358-364.
- Paradis E, Schliep K. 2019.** ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics* **35**: 526-528.
- Pickrell JK, Pritchard JK. 2012.** Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics* **8**(11): e1002967.
- Posit team 2025.** RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA.
- Preska Steinberg A, Lin M, Kussell E. 2022.** Core genes can have higher recombination rates than accessory genes within global microbial populations. *Elife* **11**.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007.** PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**(3): 559-575.
- Qian H, Deng T, Ricklefs RE. 2021.** Evolutionary assembly of the Arctic flora. *Global Ecology and Biogeography* **31**(3): 396-404.
- R Core Team 2024.** R: A language and environment for statistical computing.
- Rantanen M. 2022.** The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment* **3**(1): 1-10.
- Revell L. 2026.** R: Midpoint root a phylogeny. *Phytools.org*.
- Revell LJ. 2024.** phytools 2.0: an updated {R} ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**: e16505.
- Robinson DF, Foulds LR. 1981.** Comparison of phylogenetic trees. *Mathematical Biosciences* **53**(1): 131-147.

- Ruiz-García M. 2021.** An Overview of Phylogeny and its Applications. *Journal of Phylogenetics & Evolutionary Biology* **9**.
- Saitou N, Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**(4): 406-425.
- Schliep KP. 2011.** phangorn: phylogenetic analysis in R. *phangorn: phylogenetic analysis in R*. **27**(4): 592-593.
- Segerman B. 2012.** The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front Cell Infect Microbiol* **2**: 116.
- Severin A. 2026.** Terminology. *Bioinformatics Workbook*.
- Soares LS, Freitas LB. 2024.** The phylogeographic journey of a plant species from lowland to highlands during the Pleistocene. *Scientific Reports* **14**(1): 3825.
- Soltis PS, Soltis DE. 2026.** A panorama of plant pangenomes. *Science* **391**(6785).
- Sun Y, Wang A, Wan D, Wang Q, Liu J. 2012.** Rapid radiation of Rheum (Polygonaceae) and parallel evolution of morphological traits. *Molecular Phylogenetics and Evolution* **63**(1): 150-158.
- Tackenberg O, Stöcklin J. 2008.** Wind dispersal of alpine plant species:: A comparison with lowland species. *JOURNAL OF VEGETATION SCIENCE* **19**(1): 109-118.
- Tellini N, Tørresen OK, Edwards D, Rieseberg LH, Jakobsen KS, Cerca J. 2025.** Pangenomes as a framework for adaptive radiation, speciation, and adaptation. *American Journal of Botany* **112**(12).
- Tkach NV, Hoffmann MH, Röser M, von Hagen KB. 2008.** Temporal patterns of evolution in the Arctic explored in Artemisia L. (Asteraceae) lineages of different age. *Plant Ecology & Diversity* **1**(2): 161-169.
- Tranchant-Dubreuil C, Rouard M, Sabot F 2019.** Plant Pangenome: Impacts on Phenotypes and Evolution. *Annual Plant Reviews online*, 453-478.
- Tremetsberger K, Ortiz M, Terrab A, Balao F, Casimiro-Soriguer R, Talavera M, Talavera S. 2015.** Phylogeography above the species level for perennial species in a composite genus. *AoB Plants* **8**.
- van Kruijsbergen W** Oxyria digyna 5. Saxifraga.
- Van Rossum F, Martin H, Le Cadre S, Brachi B, Christenhusz MJM, Touzet P. 2018.** Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspectives in Plant Ecology, Evolution and Systematics* **35**: 44-51.
- Vcftools. 2015.** Welcome to VCFtools. *Github.io*.
- Wang Q, Liu J, Allen GA, Ma Y, Yue W, Marr KL, Abbott RJ. 2015.** Arctic plant origins and early formation of circumarctic distributions: a case study of the mountain sorrel, *Oxyria digyna*. *New Phytologist* **209**(1): 343-353.
- Wetterstrand K. 2026.** Linkage. *National Human Genome Research Institute*.
- WFO. 2026.** Oxyria sinensis Hemsl. *Worldfloraonline.org*.
- White JM, Ager TA, Adam DP, Leopold EB, Liu G, Jette H, Schweger CE. 1997.** An 18 million year record of vegetation and climate change in northwestern Canada and Alaska: Tectonic and global climatic correlates. *Palaeogeography Palaeoclimatology Palaeoecology* **130**(1-4): 293-306.
- Winkler M, Tribsch A, Schneeweiss GM, Brodbeck S, Gugerli F, Holderegger R, Abbott RJ, Schönswetter P. 2012.** Tales of the unexpected: Phylogeography of the arctic-alpine model plant Saxifraga oppositifolia (Saxifragaceae) revisited. *Molecular Ecology* **21**(18): 4618-4630.
- Yisilam G, Gao Y-Z, Wei Z, Comes HP, Li Z, Li P, Tian X-M. 2026.** Phylogenomic conflict in Lycium (Solanaceae): Revealing the roles of hybridization, incomplete lineage sorting, and biogeographic history using nuclear and mitochondrial genomes. *Plant Diversity*.
- Young AD, Gillung JP. 2020.** Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology* **45**(2): 225-247.
- Zhang J, Li X-Q, Peng H-W, Hai L, Erst AS, Jabbour F, Ortiz RdC, Xia F-C, Soltis PS, Soltis DE, et al. 2023.** Evolutionary history of the Arctic flora. *Nature Communications* **14**(1): 4021.

Appendix 1 Popular science summary

History of the Mountain sorrel - When the genes are not agreeing

Observing the genetic material of a species is often expected to give a clear history, but what happens when different genes show different histories?

When people distributed over Europe went out to gather leaves from the Arctic and Alpine living Mountain sorrel, they could probably not imagine that the different kinds of genes, or heritable material, extracted from the leaves would show different histories. Was maybe Darwin wrong in thinking species have a history like a tree?

Already in 2015 could Qian Wang and colleagues present evidence of the Mountain sorrels' origin to be in central Asian mountains. Despite this, the history of Mountain sorrel in Europe has been unclear until now.

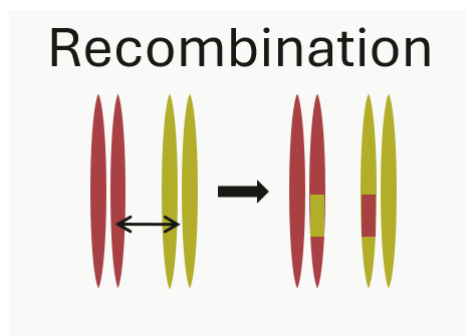
If we could travel through time, we would probably see Mountain sorrel finally being able to expand its European range to the Arctic from the older southern populations, following the retreat of the last ice age which covered the north. But it turns out that the pattern is not the same for all genes.



The Arctic-alpine Mountain sorrel (*Oxyria digyna* (L.) Hill)
Image: Saxifraga, Jan van der Straaten

Some genes are found across many species, while others only are found in one

There has recently been an increased interest in exploring variation within a group or species. The variation can be observed in the genes, the parts of the genome responsible for specific functions and expressions in an individual.



Recombination shuffles genetic information. This happens between chromosomes, which consist of DNA

For instance, there are *core genes*, which as the name suggests are in charge of essential plant functions. The *core genes* can therefore be found across a range of species. *Accessory genes* on the other hand can only be found in a few species or individuals, and are suggested to help the plant survive in the local environment.

How can different genes have different relationships between populations?

As you might understand by now, I have found evidence for core and accessory genes showing different histories for Mountain sorrel.

Despite the general pattern pointing towards movement from the south to the north, the populations of Mountain sorrel in Europe show different relationships when only observing the *core-* or *accessory genes*. For instance, a population in

Greenland is most closely related to populations in Central Europe when only considering the *core genes*. But if we instead consider *accessory genes*, the population in Greenland is closest related to a Svalbard population.

But how can this happen? Probably because the *core genes* are responsible for more essential function which makes them stay the same over long time. The *accessory genes* may instead change over time as they help the plant adapt to the local environment.

But there is also something called recombination, which shuffles genetic information. The *core genes* experience more of this shuffling than the *accessory genes*. Together, these processes are inferred as different histories between the gene types.

We must reconsider how the history of a species is decided

If the different genes can show different histories for Mountain sorrel, then imagine how many different histories and relationships there are to explore in all plants, or even in all life forms! This shows that the history of a species is not following the treelike structure Darwin proposed, once we considering the different relationships of the genes.

Instead, it's a network, created over many, many years, thanks to functional differences and recombination, eventually ending up as us today. But this network does not end with us. It will continue to form over generations, because of the differences in genes relationships.

Appendix 2 Supplementary material

Table S1 Sampled populations, with site and country, the sample locations placement, and the number of individuals sampled.

Sample site	Country	Latitude	Longitude	Sampled individuals
Tschierva	Switzerland	46,406	9,891	15
Velka Zmrzla Dolina	Slovakia	49,202	20,205	20
Liahovden	Norway	60,861	7,188	20
Norra tväråklumpen	Sweden	63,208	12,350	20
Geldingafell	Iceland	64,700	-15,351	20
Abisko	Sweden	68,360	18,772	11
Jameson Land	Greenland	71,397	-23,974	20
Zackenbergl	Greenland	74,465	-20,578	20
Isdammen Longyearbyen	Norway	78,214	15,702	20
Citronen Fjord	Greenland	83,086	-28,274	20

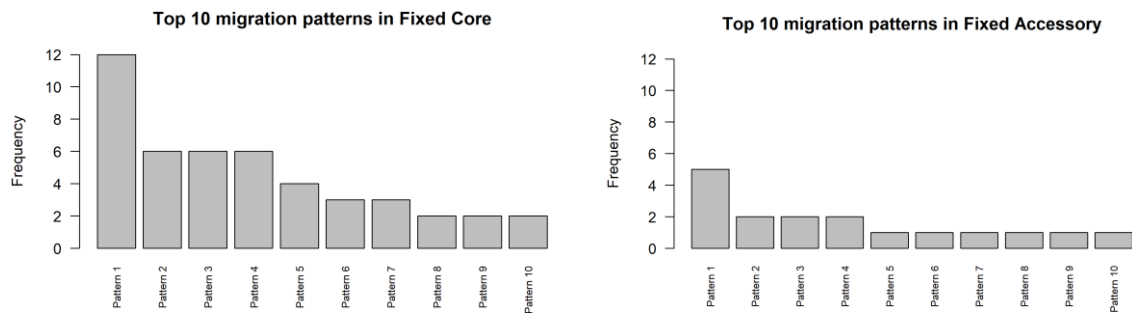


Fig S1 Frequency of migration edge patterns (position and direction of all migration edges in a graph) occurring in the 51 iterations of the phylogenetic graphs created in TreeMix (Pickrell & Pritchard, 2012). The bar plots only show the 10 most common patterns. a, Frequency of migration edge patterns of phylogenetic graphs based on core enriched regions of the chromosomes. b, Frequency of migration edge patterns of phylogenetic graphs based on accessory enriched regions of the chromosomes.