

ORIGINAL ARTICLE OPEN ACCESS

Linked Selection and Gene Density Shape Genome-Wide Patterns of Diversification in Peatmosses

Olena Meleshko¹  | Michael D. Martin¹ | Kjell Ivar Flatberg¹ | Hans K. Stenøien¹ | Thorfinn Sand Korneliussen² | Péter Szövényi³  | Kristian Hassel¹

¹Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway | ²Section for GeoGenetics, Globe Institute, University of Copenhagen, Copenhagen, Denmark | ³Department of Systematic and Evolutionary Botany & Zurich-Basel Plant Science Center, University of Zurich, Zurich, Switzerland

Correspondence: Olena Meleshko (olena.meleshko@ntnu.no) | Péter Szövényi (peter.szovenyi@uzh.ch)

Received: 13 November 2023 | **Revised:** 22 July 2024 | **Accepted:** 29 July 2024

Funding: This study was supported by Norges Forskningsråd, 250541/F20, 287327, H2020 Marie Skłodowska-Curie Actions, PlantHUB-No. 722338, Georges and Antoine Claraz Foundation, Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, 131726, 160004, 184826, and Universität Zürich.

Keywords: gene density | genome evolution | introgression | linked selection | peatmoss | speciation

ABSTRACT

Genome evolution under speciation is poorly understood in nonmodel and nonvascular plants, such as bryophytes—the largest group of nonvascular land plants. Their genomes are structurally different from angiosperms and likely subjected to stronger linked selection pressure, which may have profound consequences on genome evolution in diversifying lineages, even more so when their genome architecture is conserved. We use the highly diverse, rapidly radiated group of peatmosses (*Sphagnum*) to characterize the processes affecting genome diversification in bryophytes. Using whole-genome sequencing data from populations of 12 species sampled at different phylogenetic and geographical scales, we describe high correlation of the genomic landscapes of differentiation, divergence, and diversity in *Sphagnum*. Coupled with evidence from the patterns of covariation among different measures of genetic diversity, phylogenetic discordance, and gene density, this provides strong support that peatmoss genome evolution has been shaped by the long-term effects of linked selection, constrained by distribution of selection targets in the genome. Thus, peatmosses join the growing number of animal and plant groups where functional features of the genome, such as gene density, and linked selection drive genome evolution along predetermined and highly similar routes in different species. Our findings demonstrate the great potential of bryophytes for studying the genomics of speciation and highlight the urgent need to expand the genomic resources in this remarkable group of plants.

1 | Introduction

Since the development of next-generation sequencing technologies, the field of speciation genomics has been thriving, and our knowledge about speciation, one of the most fundamental processes in evolution, has widened drastically (Seehausen et al. 2014; Wolf and Ellegren 2017; Jiggins 2019). Deep insights into ongoing and past speciation processes can be achieved by exploring interspecific genetic differentiation and its genome-wide

distribution (Wolf and Ellegren 2017). We now know that patterns of differentiation between species are highly variable across the genome (see Seehausen et al. 2014 for a review). This is interpreted as a result of speciation by gradual establishment of barriers to gene flow due to divergent selection associated with regions of accentuated differentiation, that is, differentiation islands (Wu 2001; Turner, Hahn, and Nuzhdin 2005; Feder and Nosil 2010). However, the heterogeneous differentiation landscape may have little to do with speciation itself,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Evolutionary Applications* published by John Wiley & Sons Ltd.

but rather be derived from variation in recombination rate and distribution of functional features of the genome, that is, genome architecture, that in turn define the effects of selection or gene flow across the genome (Nachman and Payseur 2012; Cruickshank and Hahn 2014; Jiggins 2019). Positive and purifying selection lead to a decrease in genetic diversity and effective population sizes (N_e) at sites linked to the actual target of selection due to the process of genetic hitchhiking (Smith and Haigh 1974; Charlesworth, Morgan, and Charlesworth 1993). These polymorphism-reducing effects are termed linked selection (reviewed by Cutter and Payseur 2013), and result in a local increase in among-species differentiation intensified in regions with low recombination and high density of functional elements (Kaplan, Hudson, and Langley 1989; Begun and Aquadro 1992; Payseur and Nachman 2002). Long-term linked selection will also locally decrease divergence due to repeated reduction in variation at linked sites (Cruickshank and Hahn 2014; Burri et al. 2015).

According to the linked selection model, the distribution of genetic differentiation can be conjectured by variation in genome architecture (Charlesworth 1998; Nachman and Payseur 2012; Cruickshank and Hahn 2014). When genome architecture is highly conserved across species, diversifying lineages would be experiencing similar outcomes of selection, leading to highly similar genome-wide distributions of differentiation in independent species across a speciation continuum (Cutter and Payseur 2013; Burri 2017). Such highly correlated heterogeneous differentiation landscapes have been described in various groups of birds (Burri et al. 2015; Van Doren et al. 2017; Han et al. 2017; Delmore et al. 2018), insects (Martin et al. 2013; Kronforst et al. 2013; Edelman et al. 2019) and vascular plants (Renaut et al. 2013; Stankowski et al. 2019).

Genome evolution has been studied more and understood better in crops or model plants, while nonmodel plants still lack extensive genomic resources and hence understanding of how their genomes evolve in the process of speciation. This topic has hardly been investigated at all in nonvascular plants, such as bryophytes, where genomic resources are nonexistent in the vast majority of species. Bryophytes are the largest group of nonvascular land plants and share a common ancestor with all vascular plants (Puttick et al. 2018). Bryophytes play a crucial role in the global carbon and nitrogen cycling, as well as in ecosystem succession, and provide a wide range of important ecosystem services (reviewed in Turetsky 2003; Alatalo et al. 2020). These early land plants represent hundreds of millions of years of evolutionary history encompassing major evolutionary transitions after land colonization (reviewed by Fernandez-Pozo et al. 2022). The key feature differentiating these plants from angiosperms is that haploid gametophyte dominates their life cycle, which makes them more efficient in purging deleterious mutations, even more so in predominantly selfing species (Szövényi et al. 2014). Bryophytes have high incidences of clonal reproduction (Stenøien and Sastad 2001; Cronberg, Rydgren, and Økland 2006) which should lower their effective recombination rate. In contrast to angiosperm genomes where chromosomes have distinct gene-rich regions at the ends and gene-sparse regions at the center, functional elements are distributed evenly along the chromosomes in genomes of bryophytes (Lang et al. 2018; Diop et al. 2020; Li et al. 2020; Szövényi, Gunadi, and

Li 2021; Healey et al. 2023). These features of bryophyte biology could imply that strong effects of linked selection might be observed in their genomes.

One of the emerging model groups in bryophyte genomics is peatmosses (*Sphagnum* L., Sphagnaceae), a relatively well-studied, compared with the rest of bryophytes, genus comprising five subgenera with numerous species with wide geographical distributions (Michaelis 2019). We know that most peatmoss species are haploid, with a small and surprisingly stable genome size (0.39–0.49 pg DNA, Temsch, Greilhuber, and Krisai 1998), chromosome number ($n=19$) and karyotype (Fritsch 1991) across species. A recent study shows that there is high collinearity between the genomes of two peatmosses *Sphagnum divinum* and *S. angustifolium* (Healey et al. 2023) which represent different subgenera of peatmosses. These traits enable side-by-side comparison of multiple species with a speciation continuum, which allows using peatmosses to study bryophyte genome evolution. They also indicate that genome structure may be conserved in peatmosses which would exacerbate the effects of linked selection. If this is the case, genomic differentiation landscapes are expected to be similar across multiple species (Cutter and Payseur 2013; Burri 2017), and variation in this landscape should be associated with variation in genetic diversity and distribution of selection targets in the genome (Delmore et al. 2018; Rettelbach, Nater, and Ellegren 2019). Peatmosses are also characterized by relatively large N_e (Stenøien and Sastad 1999; Szövényi et al. 2008), ancient and recent interspecific hybridization, and widespread incomplete lineage sorting (ILS) stemming from rapid radiation of the genus (Meleshko et al. 2021) which occurred 7–20 Ma (Shaw et al. 2010). Therefore, the effects of gene flow and ILS might result in patterns of co-distribution of genome-wide landscapes of genetic diversity, differentiation, and divergence that differ from that produced by linked selection (Wolf and Ellegren 2017; Liang et al. 2022).

In addition to being a model of bryophyte genomics, *Sphagnum* is the most important plant genus for global carbon cycling because it is a key component of peatland ecosystems. The amount of carbon stored in peatlands exceeds that stored in all other types of vegetation combined even though peatlands only occupy ca. 3% of terrestrial area worldwide (Yu et al. 2011; IUCN 2021). Peatmoss distribution and species diversity are currently concentrated in the Northern Hemisphere (Laine et al. 2018), which is warming up faster than the global average (reviewed in Rantanen et al. 2022). This is especially concerning given that peatlands turn from carbon sink into source when destabilized or overheated (Frolking et al. 2011; Wilson et al. 2016). Studying speciation processes underlying peatmoss diversity may provide important insights which can help conservation efforts and mitigate the effects of climate change on peatmoss-mediated carbon cycling.

In this study, we use peatmosses to characterize the processes affecting genome diversification in bryophytes. We use low-depth whole-genome sequencing data from 12 peatmoss species (Meleshko et al. 2021) to describe and compare the genomic landscapes of differentiation, divergence, and diversity across the speciation continuum in *Sphagnum*. The selected species are co-occurring haploid species that represent all five subgenera within the genus and were sampled at different

phylogenetic and geographical scales. The species all evolved in the second wave of diversification and are estimated to be younger than 14 Ma (Shaw et al. 2010). Specifically, we aim to (i) estimate genome-wide distributions of genetic differentiation, divergence, and diversity in all species and pairwise species comparisons, (ii) evaluate the degree of their similarity across the species and species comparisons, (iii) explore how these statistics co-vary when taking interspecific variability into account, and (iv) estimate the relationship between these statistics and genome-wide distribution of phylogenetic discordance and density of selection targets. We use these analyses to identify the factors shaping the genomic landscape of diversification in this diverse group of plants, seeking to differentiate between the effects of linked selection, gene flow variation, and ILS.

2 | Results

2.1 | Study Species, Sampling, and Sequencing

We obtained previously published (Meleshko et al. 2021) whole-genome resequencing data which include 190 individuals from sympatric/parapatric and allopatric populations of 12 peatmoss species (Figure 5A, Table S1). Phylogenetic relationships among the species are represented by the species tree (Figure 1A) reconstructed by (Meleshko et al. 2021). For each of the species, at least one population from each of the three geographical locations (Norway, Germany, and Austria) was sampled, with one to four individuals included in each population (Table 1). The average per-sample number of raw sequencing reads after quality filtering was 65 ± 45 M (SD) reads, which were mapped to a reference genome assembly of *Sphagnum angustifolium* v0.5 (Healey et al. 2023). The average per-sample portion of uniquely mapped reads was $38 \pm 16\%$ (SD) and did not differ significantly among the species in line with the previous findings, reflecting the variation in per-sample endogenous DNA content (Meleshko et al. 2021). The sequencing coverage of 6.25 ± 2.6 (SD) was obtained after mapping of reads.

2.2 | Genomic Variation and Divergence Across Peatmosses

To describe genetic variability, among-species divergence and differentiation, we first calculated the genome-wide average of within-population nucleotide diversity (π) and Tajima's D for all species, as well as among-species genetic diversity (d_{XY}) for all species pairs, based on biallelic sites in 1774 100-kbp non-overlapping windows across all genomic scaffolds longer than 2Mbp (49 scaffolds, in total equal to 44.3% of the total length of the reference genome). We also estimated the genome-wide weighted pairwise genetic differentiation (F_{ST}), as well as pairwise F_{ST} in the same 100-kbp windows, for all species pairs. We found that genome-wide estimates of π varied between 0.0008 and 0.005 per species (0.0022 ± 0.0012 SD), indicating moderate within-population diversity (Leffler et al. 2012; Carvalho et al. 2019). In turn, average pairwise d_{XY} ranged from 0.011 to 0.025 (0.021 ± 0.0027 SD) (Figure 1B), corresponding to high levels of sequence divergence among species (Han et al. 2017;

Stankowski et al. 2019), which is in line with the strong genetic differentiation among species inferred by genome-wide F_{ST} (0.75–0.98, 0.91 ± 0.04 SD, Figure 1B). In peatmosses, high levels of F_{ST} are normally observed even among sister species (e.g., Yousefi et al. 2017; Petlund 2021) and among populations of the same species (e.g., Yousefi et al. 2019; Nilsen 2021).

The sliding-window based estimates permitted us to assess the genome-wide distribution of the statistics. We found that per-window estimates of π and Tajima's D in the investigated species varied from 0.00008 to 0.011 (0.0022 ± 0.0017 SD) (Figure 1D) and from -4.71 to 0.96 (-3.19 ± 0.85 SD) (Figure 1E), respectively, whereas in different pairwise comparisons, F_{ST} ranged from 0.17 to 0.99 (0.90 ± 0.01 SD) and d_{XY} ranged from 0.004 to 0.037 ($0.021 \pm 0.0033 \times 10^{-3}$ SD) (Figure 1C). We found that in all pairwise comparisons, F_{ST} -based differentiation was very high along most of the scaffolds interspersed with 2–5 narrow F_{ST} valleys indicating localized differentiation that is lower than the genome-wide average (Figure 2). Within-species nucleotide diversity (π) and Tajima's D , as well as divergence (d_{XY}), followed the opposite pattern, with peaks coinciding with the F_{ST} valleys. This resembles the pattern discovered in studies describing the late stages of speciation, where a small number of less differentiated genomic islands are present in an otherwise highly differentiated landscape (Riesch et al. 2017; Han et al. 2017; Ravinet et al. 2018). This initial visual assessment revealed that variation of F_{ST} , d_{XY} , and π across the genome tends to be correlated among species. The remarkable similarity in the distribution of pairwise F_{ST} values is shown for four randomly selected species pairs in Figure 2. Such a correlation would strongly suggest that evolution of genome diversification in peatmosses is indeed constrained by one or more genomic features conserved across species leading to similar outcomes of selection and/or introgression in different species (Cutter and Payseur 2013; Burri et al. 2015). To further investigate this question, we carried out various analyses to verify and describe the correlated nature of genome-wide genetic diversity and differentiation, which we provide below.

2.3 | Correlated Landscape of Differentiation and Diversity Across the Peatmoss Species

In order to summarize and normalize the level of differentiation and diversity across all species pairs and species in one statistic and to describe the strength of correlation among species, we carried out a principal component analysis (PCA) on all statistics calculated. Using this approach, the correlation among the differentiation landscapes in different species would be reflected by the proportion of variance explained by the main component, and the loadings on the main component would represent a single relative differentiation landscape for the group (Burri 2017). We performed a PCA on the sliding window data for 66 among-species estimates of pairwise F_{ST} and d_{XY} , and for 12 within-species estimates of π and Tajima's D . For each of these statistics, the genomic landscapes were highly correlated across the species or pairwise comparisons, with most of the variation explained by the first principal component (PC1) (73%, 73%, 56%, and 57% for F_{ST} , d_{XY} , π , and Tajima's D , respectively, Figures S1A,B and S2A,B). For each

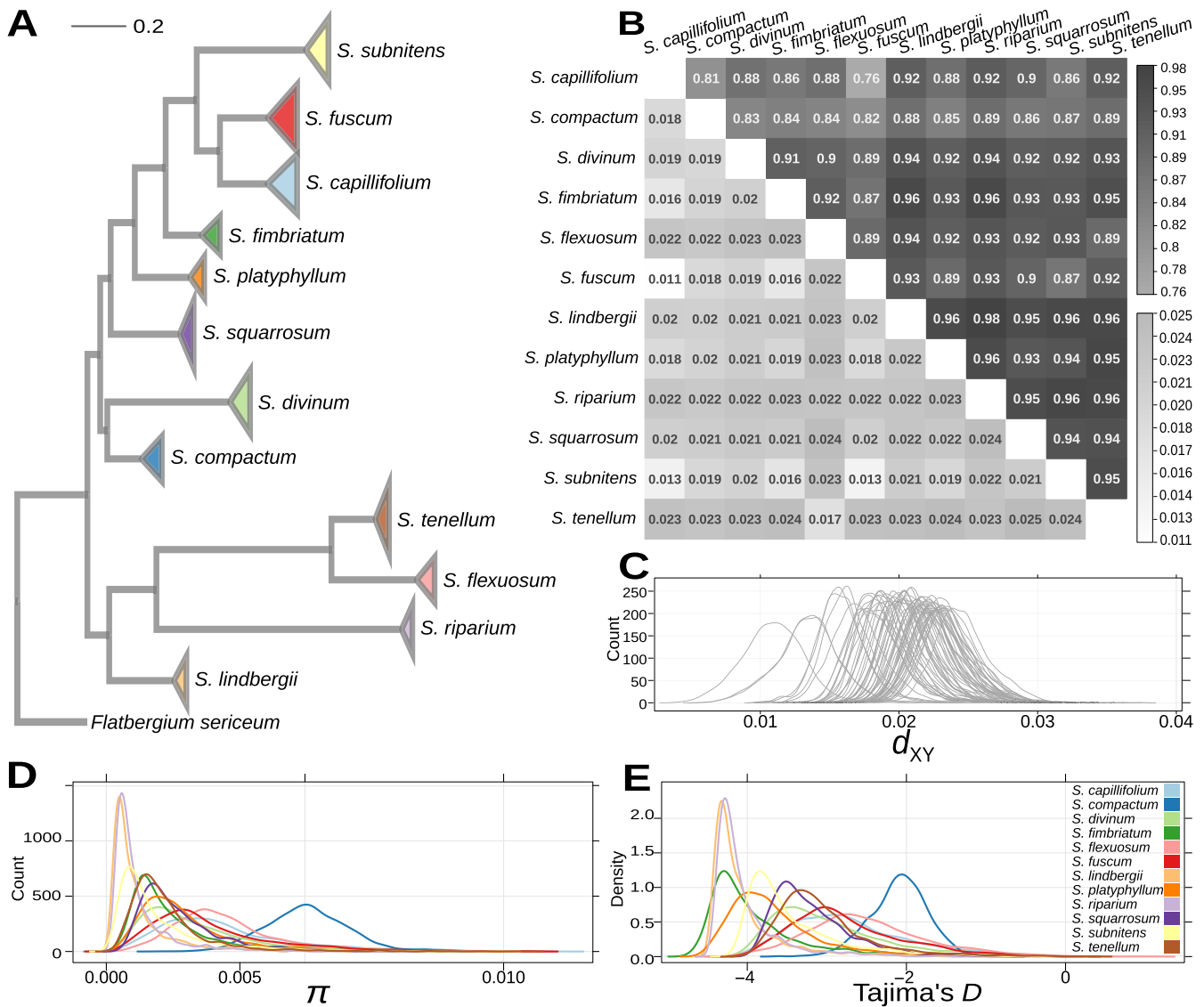


FIGURE 1 | Genetic diversity and differentiation. (A) Phylogenetic relationships among the studied species as reconstructed by Meleshko et al. (2021), with all the branch tips within each species collapsed as shown by triangular symbols that cover the same space as the clade. Colors represent species as shown in (E). (B) Pairwise d_{XY} (lower diagonal) and F_{ST} (upper diagonal) for all species pairs. The colors represent the value of the corresponding statistics as shown on the right: Lower scale bar— d_{XY} , upper scale bar— F_{ST} . (C) Distribution of pairwise d_{XY} in 100-kbp sliding windows for all species pairs. (D) Distribution of nucleotide diversity (π), and (E) Tajima's D in 100-kbp sliding windows for all species, colors represent species as shown on the right in (E).

statistic, the loadings on PC1 were strongly and significantly ($p < 0.0001$) correlated with corresponding window-based values in all individual comparisons (Figure S1C), and with the mean window-based values, that is, per-window value averaged among all species or pairwise comparisons ($r_s > |0.99|$, $p < 0.0001$, Figure S2D,E). In turn, the mean window-based values were highly significantly ($p < 0.0001$, Benjamini and Hochberg correction for multiple tests) correlated with corresponding window-based values in each species or pairwise comparison for each statistic. Since the mean window-based values are easier to interpret than the loadings on PC1, we used them as a representation of the genomic landscape for the corresponding statistic in the subsequent analyses (Table S2). Altogether, our analyses revealed highly correlated landscapes of diversity, divergence, and differentiation among the investigated peatmoss species (Figure 3, Figure S3).

2.4 | The Effects of Recurrent Linked Selection

The high correlations among the landscapes that we observed suggest that the evolution of genome diversification in peatmosses might be constrained by genomic features conserved across the species, which results in similar effects of selection (direct or indirect), introgression, or ILS in different species (Cutter and Payseur 2013; Burri 2017). To further dissect the relative contributions of these processes to the evolution of the diversification landscapes in peatmosses, we assessed the spatial distribution of these statistics and their correlation across the genome while taking among-species variability into account.

To do so, and to better describe how these population genetic statistics co-vary, we calculated the Spearman-rank correlation

among mean window-based values of gene density, π , Tajima's D , d_{XY} , and F_{ST} . We also investigated if selected genomic features could account for most of the variability in the genomic landscapes we observed. Density of functional elements, such as genes, is one of the main features of genome architecture that modulates the effect of evolutionary forces on the genomic pattern of diversification and divergence (Phung, Huber, and Lohmueller 2016; Schrider 2020). Gene density can predict variation in genetic diversity (Flowers et al. 2012) and has been previously shown to be one of the major drivers of evolution of correlated genomic landscapes via linked selection in other species (Burri et al. 2015; Delmore et al. 2018; Stankowski et al. 2019). Gene density was calculated in 100-kbp sliding windows as the ratio of the total number of bases within gene regions in a sliding window to the total length of the window (Figure S4). We also calculated gene count as the total number of genes per sliding window.

First of all, we found strong negative correlations between F_{ST} and π , and between F_{ST} and Tajima's D ($r_s = -0.97$, $p < 0.0001$, Figure 4). This suggests that F_{ST} valleys are associated with more shared polymorphisms and a higher proportion of intermediate frequency variants, which can potentially be caused by the effect of selection target density on the strengths of gene flow, rate of sorting of ancestral polymorphisms, or positive/negative selection. Assuming recurrent linked selection, gene density is expected to be positively correlated with genetic differentiation and negatively correlated with within-species genetic diversity (Cutter and Payseur 2013). Indeed, we found that gene density was weakly positively correlated with F_{ST} and negatively correlated with π ($r_s = 0.1$ and $r_s = -0.25$, respectively, $p < 0.0001$, Figure 4). By repeatedly reducing nucleotide diversity close to the target regions, recurrent positive or purifying selection should also create local reductions of d_{XY} (Burri et al. 2015; Matthey-Doret and Whitlock 2019), which is the pattern we observe. Indeed, d_{XY} was positively correlated with π and weakly negatively with F_{ST} ($r_s = 0.29$ and $r_s = -0.12$, respectively, $p < 0.0001$, Figure 4). High values of F_{ST} saturated around 1 could explain lower strength of correlations of F_{ST} with other statistics. We found a stronger skew toward an excess of rare alleles (as expressed by negative Tajima's D) in regions with higher gene density ($r_s = -0.16$, $p < 0.0001$, Figure 4). These findings strongly suggest that the differentiation landscape has been shaped by recurrent linked selection.

2.5 | Gene Flow and the Differentiation Landscape

Alternatively, the regions with accentuated differentiation might have occurred when reproductive barriers formed in the process of speciation leading to locally reduced gene flow (Cutter and Payseur 2013; Burri et al. 2015; Ma et al. 2018). In such a case, however, both F_{ST} and d_{XY} are expected to be elevated in these regions relative to the rest of the genome (homogenized by gene flow), which does not correspond to our findings. Despite the extensive interspecific hybridization among co-occurring peatmoss species, postzygotic barriers seem to have been established in the studied species, leading to very low levels of recent, postspeciation interspecific gene flow (Meleshko 2021). Nevertheless, even low levels of gene flow can have a profound effect on genome evolution by reinforcing reproductive isolation

among well-separated species (Hopkins 2013; Twyford, Kidner, and Ennos 2015). Assuming the differentiation landscape is affected by gene flow, one could expect to see a consistent difference between sympatric and allopatric comparisons (Burri et al. 2015; Yamasaki et al. 2020). Therefore, we tested if closely located populations of each species pair are less or more differentiated than their allopatric populations, which would suggest that gene flow is confined to or hampered in sympatric populations (Renaut et al. 2013; Martin et al. 2013). We calculated F_{ST} among sympatric and allopatric populations sampled in Norway and Austria for all species pairs (Figure 5B) and found no significant correlation between genetic differentiation and geographical distance between species (Mantel test, $r = -0.08$, $p = 0.88$). We then calculated the difference between mean F_{ST} in sympatric comparisons and mean F_{ST} in allopatric comparisons in 100-kbp sliding windows for each species pair (Figure 6) and performed a PCA to identify if variation in distribution of this difference along the genome is correlated among the comparisons in the same way as the overall interspecific F_{ST} . We found that only 14.6% of variation was explained by PCI (Figure S5), suggesting that there was no systematic difference between allopatric and sympatric comparisons. These findings rule out recent gene flow as a major factor affecting the differentiation landscape. We also found that d_{XY} was strongly negatively correlated with gene density ($r_s = -0.78$, $p < 0.0001$, Figure 4), which provides further support for our recurrent linked selection scenario (Han et al. 2017). Taken together, the findings suggest reduced impact of linked selection or efficacy of selection acting in the F_{ST} valleys and indicate that recurrent linked selection is one of the main forces that contributes to the correlated pattern of genomic divergence, differentiation, and diversity observed.

2.6 | The Effect of the Stochastic Coalescent Processes

In our previous study, we found that ILS and not introgression was the primary factor causing extensive genome-wide phylogenetic discordance among the investigated peatmoss species (Meleshko et al. 2021). Therefore, one might expect that, besides recurrent selection, ILS might also have influenced the genomic landscape of differentiation. In the presence of positive or purifying selection, the extent of ILS-induced discordance across the genome should be reduced in the regions with high density of selection targets (Hobolth et al. 2011), unless N_e in the extant lineages differs significantly from one of the ancestral populations (Slatkin and Pollack 2006). Thus to estimate the extent to which ILS might have shaped the genome evolution in peatmoss, we investigated how phylogenetic discordance, a measure of ILS, is distributed across the 49 longest scaffolds in 100-kbp sliding windows in relation to gene density, differentiation, divergence, and diversity.

To characterize the distribution of incongruence across the genome, we assessed concordance between individual 100-kbp sliding window trees and the species tree reconstructed by Meleshko et al. (2021). We kept one accession per species and calculated the topological distance score, equal to the number of steps required to interconvert the compared unrooted trees (Penny and Hendy 1985; Rzhetsky and Nei 1992). A high score of this metric represents a high mismatch between the species

TABLE 1 | Sampling summary.

Species	Subgenus	Population							
		Austria			Norway			Germany	Total
		A1	A2	Total	N1	N2	Total		
<i>S. capillifolium</i>	<i>Acutifolia</i>	6	1	7	3	5	8	2	17
<i>S. compactum</i>	<i>Rigida</i>	3	0	3	3	5	8	2	13
<i>S. divinum</i>	<i>Sphagnum</i>	6	3	9	3	6	9	0	18
<i>S. fimbriatum</i>	<i>Acutifolia</i>	0	4	4	4	0	4	3	11
<i>S. flexuosum</i>	<i>Cuspidata</i>	0	3	3	3	6	9	2	14
<i>S. fuscum</i>	<i>Acutifolia</i>	6	3	9	3	5	8	2	19
<i>S. lindbergii</i>	<i>Cuspidata</i>	3	0	3	3	6	9	2	14
<i>S. platyphyllum</i>	<i>Subsecunda</i>	1	3	4	3	5	8	0	12
<i>S. riparium</i>	<i>Cuspidata</i>	3	2	5	3	5	8	1	14
<i>S. squarrosom</i>	<i>Acutifolia</i>	3	5	8	3	6	9	2	19
<i>S. subnitens</i>	<i>Acutifolia</i>	3	6	9	3	5	8	2	19
<i>S. tenellum</i>	<i>Cuspidata</i>	0	6	6	3	9	12	2	20

Note: Number of samples collected from each of the populations for each of the 12 studied *Sphagnum* species: A1—Tamsweg district, Austria; A2—Upper Austria, Austria; N1—Namsos area, Norway; N2—Trondheim area, Norway; G—Germany.

tree and a sliding window tree. The analysis showed that topological discordance was highly variable among and within the genomic scaffolds, ranging from 0 (the same topology as the species tree) to 18 (5.99 ± 2.70 SD, Figure 7B). Most of the scaffolds contain outliers that differ from the genome-wide mean discordance score by >2 SD, showing very low correspondence to the species tree (Figure 7A, Figure S6). There was only a weak negative correlation between topological discordance and the number of parsimony-informative sites in the alignment ($r_s = -0.08$, $p < 0.0001$, Figure 7C). Therefore, variable discordance across the genome is real and not of technical origin.

We found that windows with lower topological discordance have significantly higher gene density ($r_s = -0.31$, $p < 0.0001$, Figures 4 and 7D). Moreover, windows with lower topological discordance have significantly higher differentiation (F_{ST}), and lower π and d_{XY} ($r_s = -0.27$, 0.31, and 0.3, respectively, $p < 0.0001$, Figure 4). Such local increase in F_{ST} and phylogenetic concordance may stem from rapid sorting of ancestral variants due to reduced diversity, resulting from recurrent linked selection (e.g., background selection or selective sweeps) (Pease and Hahn 2013; Stankowski et al. 2019). Directional selection would lead to a wider span of concordance at linked sites (Slatkin and Pollack 2006), thus under directional selection one would expect spatial aggregation of regions with high concordance over longer distances than under the neutral model involving ILS (Slatkin and Pollack 2006; Hobolth et al. 2011). Therefore, we tested if topological incongruence was spatially aggregated across the genome by performing autocorrelation analysis along the scaffolds using a lag size of 3–10 corresponding to 0.3–1 Mbp (Figure S7). We found that 84%–98% of the windows did not demonstrate significant correlation with the neighboring windows ($p < 0.01$, Figure S8)

implying that topological discordance was spatially aggregated at this scale in only a few cases, mostly at the lag size of 0.3 Mbp (see autocorrelation analysis in ‘Section 4’). The window sets with significant and insignificant autocorrelation coefficients did not differ from each other in discordance scores for all lag sizes (Student’s *t*-test, Figure S9).

3 | Discussion

In this study, we found high correlations among genomic landscapes of genetic diversity (π), divergence (d_{XY}), and differentiation (F_{ST}) of 12 peatmoss species which suggests that genome evolution in these species has been constrained by distribution of selection targets in the genome. The analyses of correlation and spatial co-distribution of the statistics across the genome showed that regions with higher gene density have higher genetic differentiation, lower divergence, and nucleotide diversity, as well as higher number of rare alleles and lower phylogenetic discordance. Our analyses of sympatric and allopatric species pairs rule out gene flow as a major factor affecting the diversification landscape. Taken together, our findings indicate that the diversification landscapes in the studied species have been shaped by recurrent linked selection constrained by distribution of functional elements in the genome. Below, we discuss how these and other factors may have influenced genome evolution during speciation in peatmosses.

3.1 | Correlated Landscapes of Differentiation Suggest Constrained Genome Evolution

Multiple studies report conserved diversification landscapes in various species across the speciation continuum. This is

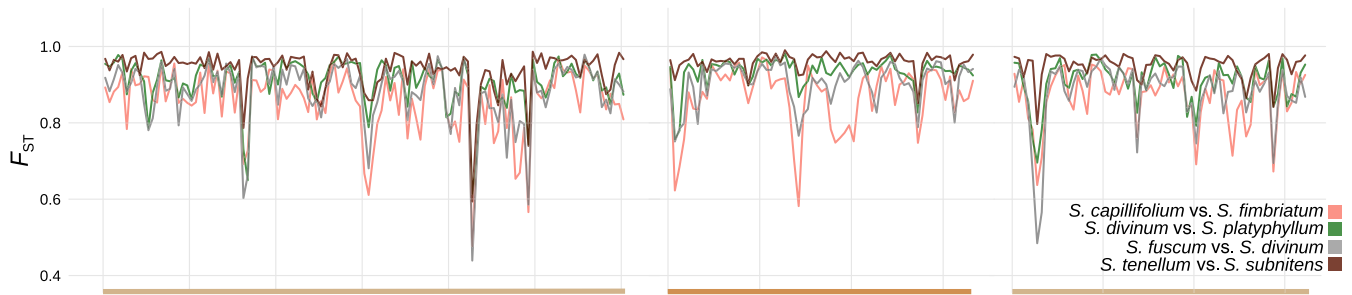


FIGURE 2 | Pairwise differentiation (F_{ST}) in 100-bp sliding windows across the three longest scaffolds (x axis, designated by colors) in four arbitrary species pairs involving seven species. Colors correspond to pairwise comparisons as shown on the right.

attributed to conserved distribution of functional features in the genome, such as gene density and/or recombination rate, which imposes restrictions onto the magnitude and genomic consequences of evolutionary forces acting on the genomes of diversifying lineages, resulting in evolution of similar genomic landscapes of divergence and diversity in them (Martin et al. 2013; Irwin et al. 2016; Vijay et al. 2016; Delmore et al. 2018; Stankowski et al. 2019; Chase, Ellegren, and Mugal 2021; Liang et al. 2022). Our findings of highly correlated landscapes of differentiation described here suggest that peatmosses join the growing number of animal and plant groups where conserved distribution of functional genomic features condition genome evolution under speciation.

3.2 | The Role of Linked Selection in Shaping the Genomic Landscape of Peatmoss Diversification

Haploid nature and even distribution of selection targets across the genome suggest that direct and linked effect of positive and/or purifying selection is expected to be strong in peatmoss genomes. Indeed, we argue that the diversification landscape observed in the studied species can be best explained by the effect of recurrent linked selection. In particular, the covariation among measures of genetic diversity, phylogenetic discordance, and gene density, as demonstrated here, is a predicted consequence of long-term purifying selection and/or background selection (BGS) (Burri 2017). Cumulative indirect effects of recurrent BGS largely explain the genome-wide variation in genetic diversity and correlated landscapes of differentiation in many organisms, especially at late stages of differentiation (Burri 2017; Rettelbach, Nater, and Ellegren 2019). BGS also better explains the negative correlation between divergence and differentiation since, in outcrossing sexual organisms, it affects d_{XY} more strongly than F_{ST} (Matthey-Doret and Whitlock 2019). Similarly, the relationships observed between genetic diversity, gene density, and Tajima's D can be explained by the effect of recurrent BGS/purifying selection. Low Tajima's D could also indicate recent population expansion leading to excess of rare variants (Tajima 1989), but it would not be expected to be associated with regions of higher gene density (Hahn, Rausher, and Cunningham 2002; Stajich and Hahn 2005), in contrast to our observations. Finally, the genomic landscapes of Tajima's D and diversity strongly resemble those produced in simulations under the BGS scenario in *Drosophila* whose genome has a similar gene-dense structure (Schridder 2020), which further substantiates the predominant effect of BGS.

3.3 | Incomplete Lineage Sorting Is Widespread in the Genome but Is Modulated by Selection Targets Density Variation

Peatmosses are characterized by extensive ILS (Meleshko et al. 2021). Recurrently, acting selection reduces neutral genetic variation and N_e , and thereby facilitates lineage sorting at linked sites (Pollard et al. 2006; Hobolth et al. 2011; Pease and Hahn 2013; Li et al. 2019). This effect depends on the strength of linkage and is more pronounced in the regions with higher density of selection targets (Kaplan, Hudson, and Langley 1989; Hudson and Kaplan 1995; Cutter and Payseur 2013; Schridder 2020). Accordingly, we find that phylogenetic discordance is heterogeneously distributed across the genome and negatively, albeit not strongly, associated with gene density. Under recurrent linked selection at late stages of speciation, however, stronger correlation is expected, unless the species evolved considerable differences in recombination rate or gene conversion occurred in the regions of low recombination (Pollard et al. 2006; Burri 2017). The latter is not the case since we do observe correlated diversity, divergence, and differentiation landscapes among the species, which would only be possible with conserved recombination landscape (reviewed in Burri 2017). It has been suggested that under the neutral model with ILS, regions harboring loci responsible for establishment of reproductive isolation between species via hybrid incompatibilities, that is, speciation genes, are expected to produce evolutionary relationships discordant with the species trees (Wang and Hahn 2018). This could explain why discordance was quite strong even in regions with high gene density. The genetic mechanisms and the extent of reproductive isolation in peatmosses are unknown; therefore, this should be interpreted with caution. In addition, peatmosses are characterized by large ancestral N_e (Stenøien and Sæstad 1999; Szövényi et al. 2008; Yousefi et al. 2017) and long generation times, which are expected to increase ILS even further (Copetti et al. 2017).

3.4 | Effects of Demographic History

Natural selection is expected to affect much smaller, localized regions of the genome in comparison with the more broad, genome-wide effects of demographic events (Stajich and Hahn 2005; Schridder 2020). The combination of recurrent selection coupled with a recent bottleneck has been shown to have strong effects on the genome in many organisms, including humans (Stajich and Hahn 2005), crucifers (Slotte et al. 2010), monkeyflowers (Stankowski et al. 2019), and

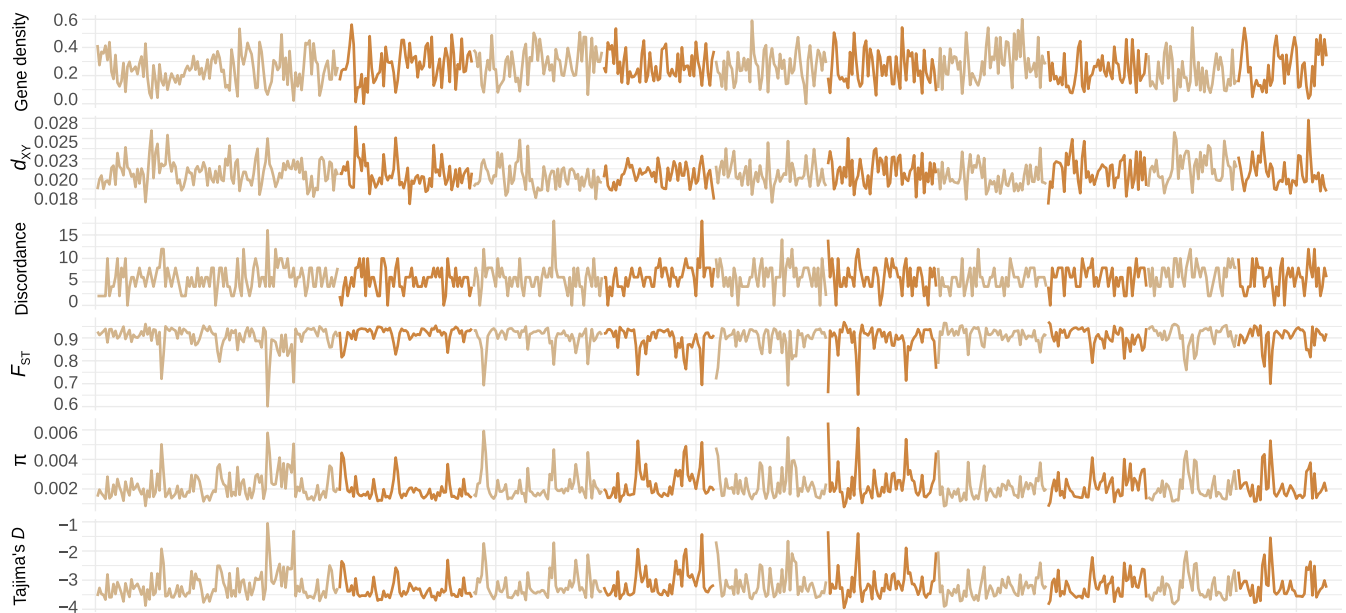


FIGURE 3 | Correlated genomic landscapes of summary statistics (the mean window-based values) in 100-kbp sliding windows across the 10 longest scaffolds (x axis). Alternating colors represent different scaffolds.

poplar species (Ma et al. 2018), among others. The Northern Hemisphere peatmoss species have experienced massive bottlenecks during the Last Glacial Maximum (reviewed in Kyrkjeide et al. 2014). This entails numerous consequences for the subsequent evolutionary history of the studied species and complicates the interpretation of the results of our analyses. First, contemporary N_e may differ markedly from ancestral population sizes, resulting in deviation of the expected relationships between phylogenetic discordance and functional features distribution (Slatkin and Pollack 2006). Secondly, bottleneck events followed by population expansion can lead to false signatures of selection in the genome, and, in particular, in patterns resembling selective sweeps instead of correctly inferring BGS (reviewed in Schrider 2020). Nevertheless, fluctuations in effective population size do not change the fact that recurrent linked selection is expected to produce patterns of genetic diversity and differentiation deviant from that expected under the neutral model (Zeng 2013).

4 | Conclusions

In this study, we demonstrate that genome evolution in the studied species has mainly been shaped by long-term effects of linked, most likely purifying, selection, constrained by distribution of selection targets in the genome. Future perspectives could involve comparing the differentiation landscapes in multiple species according to the time since their split, while taking into account the demographic history of each species.

Our findings are in line with studies in vascular plants, for example, oaks (Liang et al. 2022), sunflowers (Renaut et al. 2013), poplar (Wang et al. 2020), and *Primulina* (Ke et al. 2022) and therefore suggest that genome evolution is guided by similar processes in bryophytes and vascular land plants. Speciation on genomic level is primarily studied in angiosperms, whereas our study demonstrates that bryophytes hold a great potential for

this line of research, calling for facilitating the expansion of the genomic resources in this remarkable group of plants.

5 | Materials and Methods

5.1 | Dataset

We used whole-genome resequencing data obtained in the previous study (Meleshko et al. 2021) comprising 190 individuals from 12 haploid co-occurring peatmoss species (Table S1) and sampled from sympatric/parapatric and allopatric populations at three geographical scales in Europe (Table 1). Sequencing data processing, read mapping, and filtering was performed using a reference genome assembly of *Sphagnum angustifolium* v0.5 (Healey et al. 2023) and the Paleomix pipeline v1.2.13.4 (Schubert et al. 2014) as described in (Meleshko et al. 2021).

5.2 | Population Genomic Analyses

We used ANGSD v0.931 (Korneliussen, Albrechtsen, and Nielsen 2014) to calculate nucleotide diversity and neutrality statistics within species, as well as species differentiation based on site frequency spectrum (SFS), without calling individual genotypes. First, we performed quality filtering of mapped reads, which we hereafter refer to as “reads quality filtering.” We carried out per-base alignment quality (BAQ) computation (Li 2011) implemented in ANGSD, and adjusted mapping quality score (MAPQ) to 50 for reads containing excessive mismatches. We discarded secondary alignments and reads with unmapped mate, poor quality ($\text{flag} \geq 256$), low MAPQ score (≤ 30), or low base quality score (≤ 20). We discarded individuals from a site if individual filtered read depth at that site differed from 2 to 100.

We further used ANGSD to estimate SFS based on genotype likelihoods (GL) computed using the SAMTools method (Li

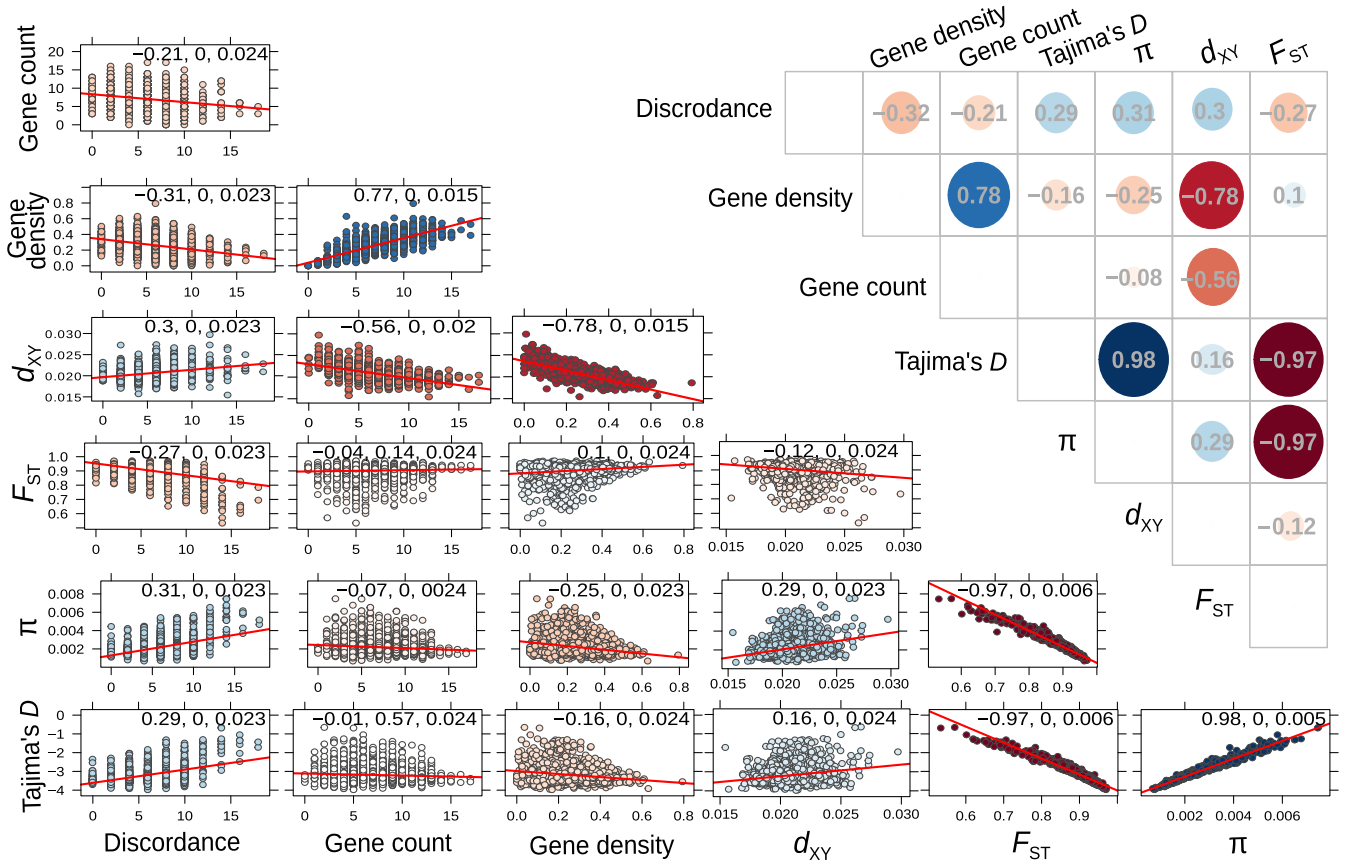


FIGURE 4 | Correlation structure of the summary statistics investigated. The numbers on the scatterplots correspond to Spearman's rank correlation coefficients followed by probability values adjusted for multiple tests and standard error. Tajima's D , π , d_{XY} , and F_{ST} refer to mean window-based Tajima's D , π , d_{XY} , and F_{ST} , respectively. Solid red lines represent the linear regression between the corresponding statistics. The color and size of circles in the upper panel represent the strength of correlations, with blue color representing positive correlation, and red color—negative correlation.

et al. 2009) setting ploidy level to 1 (*-isHap 1*). Based on GLs in biallelic sites, we calculated allele frequencies and inferred minor alleles using ML approach (Skotte, Korneliussen, and Albrechtsen 2013). We discarded sites missing in more than $\frac{1}{3}$ of individuals. Then, site allele frequency likelihood (SAF) was estimated jointly for all individuals for each species. We also estimated the SAF for each population. Using the expectation maximization (EM) algorithm, we optimized and polarized the SAF to obtain an ML estimate of the SFS for each species (Nielsen et al. 2012) and an ML estimate of the pairwise (2D) SFS for each species pair.

We used the Empirical Bayes method implemented in ANGSD to perform neutrality test statistics and to calculate F_{ST} in sliding windows across the genome, using the inferred SFS to take into account genotyping uncertainty (Korneliussen et al. 2013; Fumagalli et al. 2013). Instead of performing computationally intensive estimation of an ML SFS for each sliding window, the method calculates posterior probabilities for the SFS at each site using a prior. For neutrality test statistics, the prior is the joint ML estimate of the SFS for the whole genome generated at the previous step (Korneliussen et al. 2013). Without inferring derived alleles, the posterior estimates of θ_w and θ_π are obtained for each window as linear functions of the folded ML estimate of the SFS assuming

the infinite size model, and Tajima's D is calculated for each window assuming neutral model without recombination as $T = (\theta_\pi - \theta_w) / \sqrt{\text{var}(\theta_\pi - \theta_w)}$ (Korneliussen et al. 2013) SFS for each species pair (Fumagalli et al. 2013). Using the folded estimate of 2D SFS, we calculated whole-genome weighted F_{ST} for each species pair, as well as in sliding windows, in ANGSD (Fumagalli et al. 2013) using an extended version of the method-of-moments estimator (Reynolds, Weir, and Cockerham 1983). We used 100-kbp sliding nonoverlapping windows and kept the scaffolds longer than 2 M and 1 M bases that equal to 44.3% (175.6 M bases) and to 70.3% (278.6) of the total length of the reference for the sliding-window based estimates and the whole-genome F_{ST} calculation, respectively. The cutoff of the scaffolds length was chosen to maximize the number of windows per scaffold and exclude scaffolds with low quality, that is, with high repeat content and contamination, and low number of gene models. We calculated correlation between pairwise F_{ST} matrix and geographical distance matrix using a Mantel test (Mantel 1967) within the package "ape" (Paradis and Schliep 2019).

To estimate divergence (d_{XY}) among the species, we used the collection of scripts for genomic data analysis by Martin (https://github.com/simonhmartin/genomics_general). We calculated d_{XY} among all species in 100-kbp consecutive

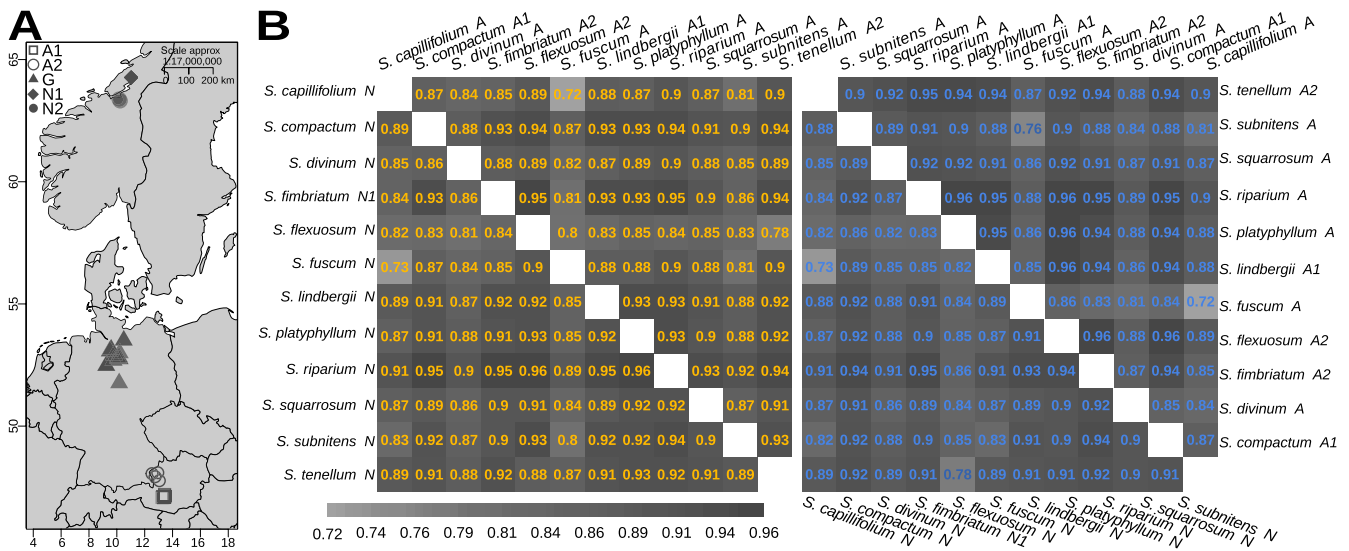


FIGURE 5 | Genetic differentiation in sympatric and allopatric comparisons. (A) Geographic locations of the sampled populations: A1—Tamsweg district, Austria; A2—Upper Austria, Austria; N1—Namsos area, Norway; N2—Trondheim area, Norway; G—Germany. The symbols represent sampling locations as shown on the left. (B) Pairwise F_{ST} among sympatric (in blue) and allopatric (in orange) populations in all species pairs. The plot on the right shows sympatric populations from Norway (lower diagonal) and Austria (upper diagonal). Population codes: A1—Tamsweg district, Austria; A2—Upper Austria; A—all from Austria combined; N1—Namsos area, Norway; N2—Trondheim area, Norway; N—all from Norway combined.

windows using the script `popgenWindows.py` with command line argument “`popDist`” based on nuclear genome consensus fasta alignments for each sliding window generated using ANGSD. To obtain the alignments, we performed read quality filtering on mapped reads and used the filtered reads to generate fasta files for each sample in ANGSD. At each position, we sampled the base with the highest effective base depth (EBD), which is a product of mapping quality and base quality scores for each base. This method offers better precision in base calling from low-coverage sequencing data (Wang et al. 2013). We then used a custom bash script and `SeqKit` (Shen et al. 2016) to slice the individual alignments into 100-kbp consecutive windows and convert them into multiple sequence alignment fasta files for each sliding window. We discarded one sample (UH58) based on the high number of missing bases. To verify that the exclusion of UH58 did not affect the statistics estimates, and to compare the statistics calculated based on GLs and on the alignments, we further used `popgenWindows.py` to calculate π and F_{ST} . The mean window-based values for each statistic were significantly highly correlated to the ones obtained based on GLs in ANGSD ($r_s = 0.98$ for π , $r_s = 0.96$ for F_{ST} , $p < 0.0001$). We used GLs-based π and F_{ST} for all subsequent analyses described below.

We visually inspected the genomic distribution of these sliding-window statistics for all species or pairwise comparisons concluding that the genomic landscapes of within-species diversity (θ_π , hereafter referred to as π) and Tajima’s D , and between-species differentiation (F_{ST}) and divergence (d_{XY}) were correlated across all species. Thus, we performed PCA on the sliding-window data to summarize the across-species variation in these statistics. The variation explained by the principal component can be used as a proxy for the degree of the correlation among the genomic landscapes across the species (Burri 2017). We extracted the loadings of each sliding window onto PC1 and

calculated their correlation with the mean interspecific window-based values for each statistic. We also obtained Spearman’s correlation coefficient between PC1 for a statistic and the corresponding window-based values for all the pairwise comparisons. False discovery rate correction was performed for these and described below correlation analyses using Benjamini and Hochberg correction (Benjamini and Hochberg 1995).

5.3 | Phylogenetic Discordance

To estimate the distribution of phylogenetic concordance across the genome, we used the consensus alignments prepared as described above to generate phylogenetic trees in 100-kbp consecutive sliding windows. We used `IQ-TREE v1.6.12` (Nguyen et al. 2015) to infer the best ML tree for each window using `GTRGAMMA+I` model and 1000 ultrafast bootstrap replicates (Hoang et al. 2018). We then calculated a measure of phylogenetic discordance between the estimated consensus sliding window trees and the species tree reconstructed from the same dataset by Meleshko et al. (2021) using a topological distance score in the package “ape.” The score is the partition metric defined as the number of steps required to transform the compared unrooted trees into one another (Penny and Hendy 1985; Rzhetsky and Nei 1992). We kept one sample per species in the sliding window trees and the species tree as branching order within each species was poorly resolved in the species tree, whereas all species were well-resolved as monophyletic clades.

5.4 | Correlation Analysis

We calculated Spearman’s correlation coefficient between the phylogenetic discordance score and gene density, gene count, and population genetic parameters such as π , Tajima’s D , d_{XY} ,

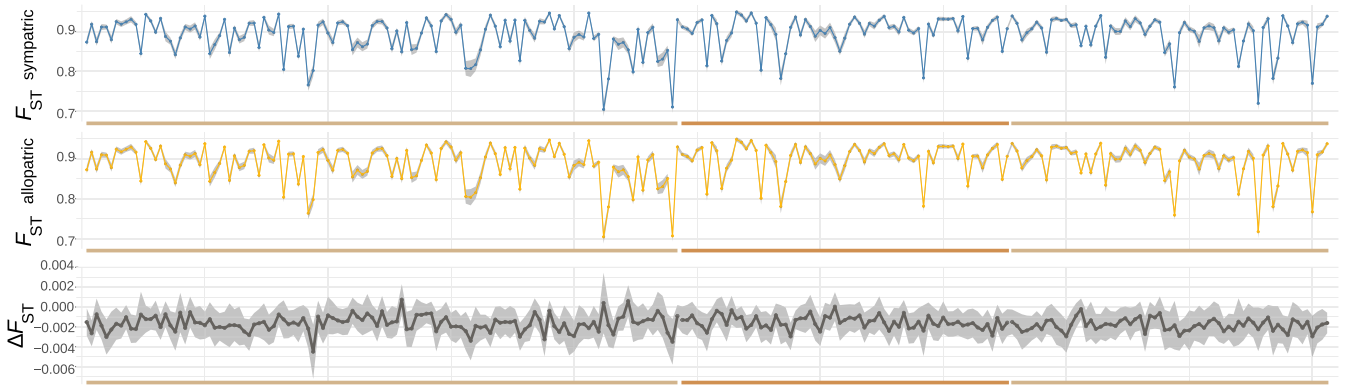


FIGURE 6 | Mean per-window pairwise F_{ST} in sympatric (in blue) and allopatric (in orange) populations across the three longest scaffolds (x axis, designated by colors, same order as on Figure 2). The bottom panel shows the mean per-window difference between allopatric and pairwise sympatric F_{ST} for each pair; light-gray polygons correspond to confidence intervals.

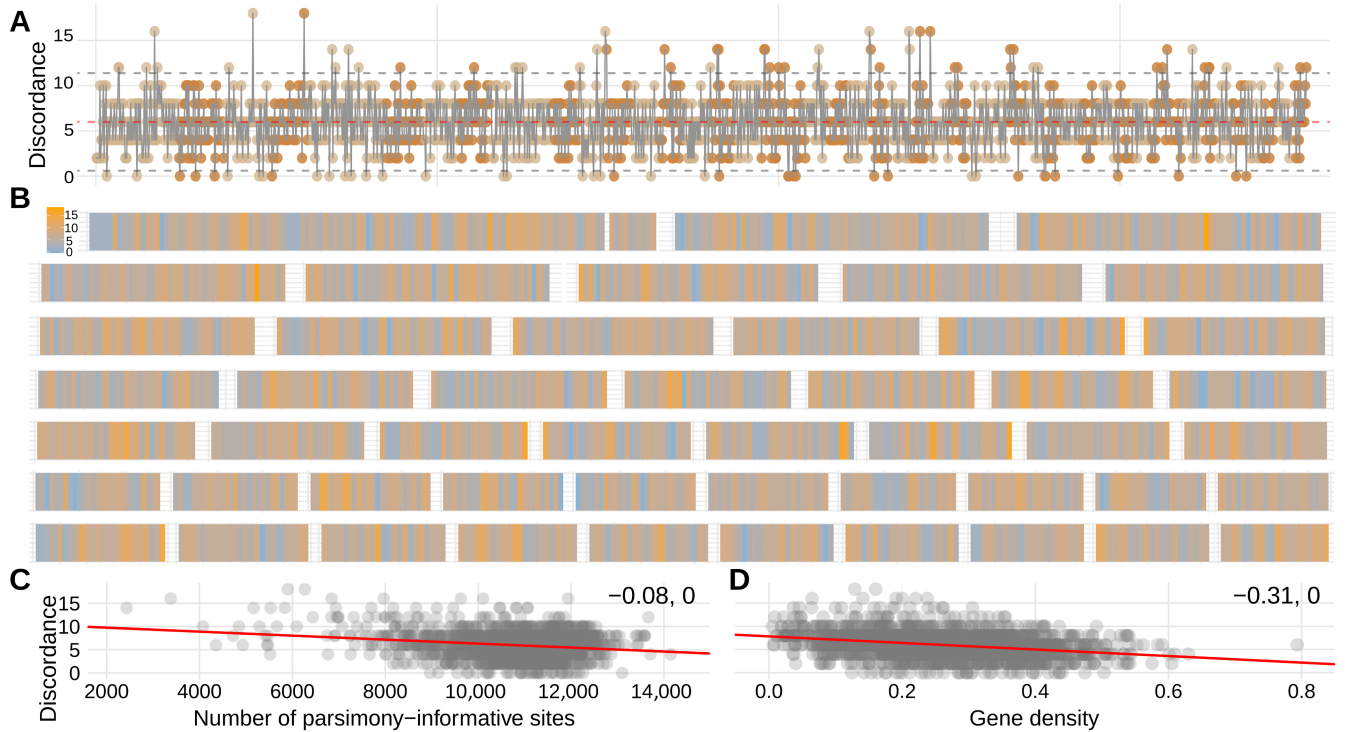


FIGURE 7 | Topological discordance for sliding window trees across the 49 longest scaffolds. (A) Distribution of topological discordance score across the genome, the x axis corresponds to 100-kbp sliding windows, colors refer to the scaffold the sliding window is located on. (B) Distribution of topological discordance score, colors refer to the value of topological discordance score in the corresponding sliding window. Relationships between the topological discordance score and (C) the number of parsimony-informative sites, and (D) gene density per window, the numbers in the top right corner are Spearman's rank correlation coefficient followed by P value.

and F_{ST} . Gene density was estimated as a ratio of the total number of bases within all gene regions located in a given sliding window to the total length of that window, whereas gene count corresponds to the number of genes located in the sliding window. We also calculated spatial autocorrelation coefficients along the scaffolds for phylogenetic discordance score, π , Tajima's D , d_{XY} , and F_{ST} . For each statistic, we calculated Spearman's correlation between the adjacent window sets using lag size of 3–10 windows corresponding to 0.3–1 M bases. The significance of autocorrelation coefficients was determined from a null distribution of the autocorrelation values obtained from 1000 random permutations of the genome-wide data. The results of this

and other analyses were visualized using the package “lattice” (Sarkar 2008), “corrplot” (Wei and Simko 2017), and “ggplot2” (Wickham 2016) in the R statistical environment v3.6.3 (R Core Team 2020).

Acknowledgements

This work was supported by a Research Council of Norway award (project no. 250541/F20) to K.H.; Swiss National Science Foundation grants 160004, 131726, and 184826; the EU's Horizon 2020 Research and Innovation Programme (PlantHUB-No. 722338); the Georges and Antoine Claraz Foundation; the Forschungskredit and the University

Research Priority Program “Evolution in Action” of the University of Zurich to P.S.; and a Research Council of Norway award (project no. 287327) to M.D.M. Some computational analyses were performed on resources provided by the National Infrastructure for High Performance Computing and Data Storage in Norway (UNINETT Sigma2).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The sequence data used in this study are available in The European Nucleotide Archive, and the accession numbers are provided in the Table S1.

References

Alatalo, J. M., A. K. Jägerbrand, M. B. Erfanian, S. Chen, S.-Q. Sun, and U. Molau. 2020. “Bryophyte Cover and Richness Decline After 18 Years of Experimental Warming in Alpine Sweden.” *AoB Plants* 12: plaa061.

Begun, D. J., and C. F. Aquadro. 1992. “Levels of Naturally Occurring DNA Polymorphism Correlate With Recombination Rates in *D. melanogaster*.” *Nature* 356: 519–520.

Benjamini, Y., and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B: Methodological* 57: 289–300.

Burri, R. 2017. “Interpreting Differentiation Landscapes in the Light of Long-Term Linked Selection.” *Evolution Letters* 1: 118–131.

Burri, R., A. Nater, T. Kawakami, et al. 2015. “Linked Selection and Recombination Rate Variation Drive the Evolution of the Genomic Landscape of Differentiation Across the Speciation Continuum of *Ficedula* Flycatchers.” *Genome Research* 25: 1656–1665.

Carvalho, Y. G. S., L. C. Vitorino, U. J. B. de Souza, and L. A. Bessa. 2019. “Recent Trends in Research on the Genetic Diversity of Plants: Implications for Conservation.” *Diversity* 11: 62.

Charlesworth, B. 1998. “Measures of Divergence Between Populations and the Effect of Forces That Reduce Variability.” *Molecular Biology and Evolution* 15: 538–543.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. “The Effect of Deleterious Mutations on Neutral Molecular Variation.” *Genetics* 134: 1289–1303.

Chase, M. A., H. Ellegren, and C. F. Mugal. 2021. “Positive Selection Plays a Major Role in Shaping Signatures of Differentiation Across the Genomic Landscape of Two Independent *Ficedula* Flycatcher Species Pairs.” *Evolution* 75: 2179–2196.

Copetti, D., A. Búrquez, E. Bustamante, et al. 2017. “Extensive Gene Tree Discordance and Hemiplasy Shaped the Genomes of North American Columnar Cacti.” *Proceedings of the National Academy of Sciences, USA* 114: 12003–12008.

Cronberg, N., K. Rydgren, and R. H. Økland. 2006. “Clonal Structure and Genet-Level Sex Ratios Suggest Different Roles of Vegetative and Sexual Reproduction in the Clonal Moss *Hylocomium splendens*.” *Ecography* 29: 95–103.

Cruickshank, T. E., and M. W. Hahn. 2014. “Reanalysis Suggests That Genomic Islands of Speciation Are due to Reduced Diversity, Not Reduced Gene Flow.” *Molecular Ecology* 23: 3133–3157.

Cutter, A. D., and B. A. Payseur. 2013. “Genomic Signatures of Selection at Linked Sites: Unifying the Disparity Among Species.” *Nature Reviews Genetics* 14: 262–274.

Delmore, K. E., J. S. L. Ramos, B. M. V. Doren, et al. 2018. “Comparative Analysis Examining Patterns of Genomic Differentiation Across Multiple Episodes of Population Divergence in Birds.” *Evolution Letters* 2: 76–87.

Diop, S. I., O. Subotic, A. Giraldo-Fonseca, et al. 2020. “A Pseudomolecule-Scale Genome Assembly of the Liverwort *Marchantia polymorpha*.” *Plant Journal* 101: 1378–1396.

Edelman, N. B., P. B. Frandsen, M. Miyagi, et al. 2019. “Genomic Architecture and Introgression Shape a Butterfly Radiation.” *Science* 366: 594–599.

Feder, J. L., and P. Nosil. 2010. “The Efficacy of Divergence Hitchhiking in Generating Genomic Islands During Ecological Speciation.” *Evolution* 64: 1729–1747.

Fernandez-Pozo, N., F. B. Haas, S. B. Gould, and S. A. Rensing. 2022. “An Overview of Bioinformatics, Genomics, and Transcriptomics Resources for Bryophytes.” *Journal of Experimental Botany* 73: 4291–4305.

Flowers, J. M., J. Molina, S. Rubinstein, P. Huang, B. A. Schaal, and M. D. Purugganan. 2012. “Natural Selection in Gene-Dense Regions Shapes the Genomic Pattern of Polymorphism in Wild and Domesticated Rice.” *Molecular Biology and Evolution* 29: 675–687.

Fritsch, R. 1991. *Index to Bryophyte Chromosome Counts*. Stuttgart, Germany: Schweizerbart'sche Verlagsbuchhandlung.

Frolking, S., J. Talbot, M. C. Jones, et al. 2011. “Peatlands in the Earth's 21st Century Climate System.” *Environmental Reviews* 19: 371–396.

Fumagalli, M., F. G. Vieira, T. S. Korneliussen, et al. 2013. “Quantifying Population Genetic Differentiation From Next-Generation Sequencing Data.” *Genetics* 195: 979–992.

Hahn, M. W., M. D. Rausher, and C. W. Cunningham. 2002. “Distinguishing Between Selection and Population Expansion in an Experimental Lineage of Bacteriophage T7.” *Genetics* 161: 11–20.

Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. “Gene Flow, Ancient Polymorphism, and Ecological Adaptation Shape the Genomic Landscape of Divergence Among Darwin's Finches.” *Genome Research* 27: 1004–1015.

Healey, A. L., B. Piatkowski, J. T. Lovell, et al. 2023. “Newly Identified Sex Chromosomes in the *Sphagnum* (Peat Moss) genome Alter Carbon Sequestration and Ecosystem Dynamics.” *Nature Plants* 9: 1–17.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. “UFBoot2: Improving the Ultrafast Bootstrap Approximation.” *Molecular Biology and Evolution* 35: 518–522.

Hobolth, A., J. Y. Dutheil, J. Hawks, M. H. Schierup, and T. Mailund. 2011. “Incomplete Lineage Sorting Patterns Among Human, Chimpanzee, and Orangutan Suggest Recent Orangutan Speciation and Widespread Selection.” *Genome Research* 21: 349–356.

Hopkins, R. 2013. “Reinforcement in Plants.” *New Phytologist* 197: 1095–1103.

Hudson, R. R., and N. L. Kaplan. 1995. “Deleterious Background Selection With Recombination.” *Genetics* 141: 1605–1617.

Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens. 2016. “Recurrent Selection Explains Parallel Evolution of Genomic Regions of High Relative but Low Absolute Differentiation in a Ring Species.” *Molecular Ecology* 25: 4488–4507.

IUCN. 2021. “IUCN Issues Peatlands and Climate Change.”

Jiggins, C. D. 2019. “Can Genomics Shed Light on the Origin of Species?” *PLoS Biology* 17: e3000394.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. “The ‘Hitchhiking Effect’ Revisited.” *Genetics* 123: 887–899.

Ke, F., L. Vasseur, H. Yi, et al. 2022. “Gene Flow, Linked Selection, and Divergent Sorting of Ancient Polymorphism Shape Genomic

- Divergence Landscape in a Group of Edaphic Specialists." *Molecular Ecology* 31: 104–118.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15: 356.
- Korneliussen, T. S., I. Moltke, A. Albrechtsen, and R. Nielsen. 2013. "Calculation of Tajima's *D* and Other Neutrality Test Statistics From Low Depth Next-Generation Sequencing Data." *BMC Bioinformatics* 14: 289.
- Kronforst, M. R., M. E. B. Hansen, N. G. Crawford, et al. 2013. "Hybridization Reveals the Evolving Genomic Architecture of Speciation." *Cell Reports* 5: 666–677.
- Kyrkjeeide, M. O., H. K. Stenøien, K. I. Flatberg, and K. Hassel. 2014. "Glacial Refugia and Post-Glacial Colonization Patterns in European Bryophytes." *Lindbergia* 37: 47–59.
- Laine, J., K. I. Flatberg, P. Harju, et al. 2018. *Sphagnum Mosses: The Stars of European Mires*. Helsinki, Finland: Sphagna Ky.
- Lang, D., K. K. Ullrich, F. Murat, et al. 2018. "The *Physcomitrella patens* Chromosome-Scale Assembly Reveals Moss Genome Structure and Evolution." *Plant Journal* 93: 515–533.
- Leffler, E. M., K. Bullaughey, D. R. Matute, et al. 2012. "Revisiting an Old Riddle: What Determines Genetic Diversity Levels Within Species?" *PLoS Biology* 10: e1001388.
- Li, F.-W., T. Nishiyama, M. Waller, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants and the Unique Biology of Hornworts." *Nature Plants* 6: 259–272.
- Li, G., H. V. Figueiró, E. Eizirik, and W. J. Murphy. 2019. "Recombination-Aware Phylogenomics Reveals the Structured Genomic Landscape of Hybridizing cat Species." *Molecular Biology and Evolution* 36: 2111–2126.
- Li, H. 2011. "Improving SNP Discovery by Base Alignment Quality." *Bioinformatics* 27: 1157–1158.
- Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25: 2078–2079.
- Liang, Y.-Y., Y. Shi, S. Yuan, et al. 2022. "Linked Selection Shapes the Landscape of Genomic Variation in Three Oak Species." *New Phytologist* 233: 555–568.
- Ma, T., K. Wang, Q. Hu, et al. 2018. "Ancient Polymorphisms and Divergence Hitchhiking Contribute to Genomic Islands of Divergence Within a Poplar Species Complex." *Proceedings of the National Academy of Sciences* 115: E236–E243.
- Mantel, N. 1967. "The Detection of Disease Clustering and a Generalized Regression Approach." *Cancer Research* 27: 209–220.
- Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, et al. 2013. "Genome-Wide Evidence for Speciation With Gene Flow in *Heliconius* Butterflies." *Genome Research* 23: 1817–1828.
- Matthey-Doret, R., and M. C. Whitlock. 2019. "Background Selection and F_{ST} : Consequences for Detecting Local Adaptation." *Molecular Ecology* 28: 3902–3914.
- Meleshko, O. 2021. *Gene Flow and Genome Evolution in Peatmosses (Sphagnum)*. Trondheim, Norway: NTNU.
- Meleshko, O., M. D. Martin, T. S. Korneliussen, et al. 2021. "Extensive Genome-Wide Phylogenetic Discordance Is due to Incomplete Lineage Sorting and Not Ongoing Introgression in a Rapidly Radiated Bryophyte Genus." *Molecular Biology and Evolution* 38: 2750–2766.
- Michaelis, D. 2019. *The Sphagnum Species of the World*. Stuttgart, Germany: Schweizerbart Science Publishers.
- Nachman, M. W., and B. A. Payseur. 2012. "Recombination Rate Variation and Speciation: Theoretical Predictions and Empirical Results From Rabbits and Mice." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367: 409–421.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32: 268–274.
- Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang. 2012. "SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation From New-Generation Sequencing Data." *PLoS One* 7: e37558.
- Nilsen, N. H. 2021. *Spatial Genetic Structure in Northern European Sphagnum Squarrosum Crome in Relation to Its Arctic Morph*. Trondheim, Norway: NTNU.
- Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35: 526–528.
- Payseur, B. A., and M. W. Nachman. 2002. "Gene Density and Human Nucleotide Polymorphism." *Molecular Biology and Evolution* 19: 336–340.
- Pease, J. B., and M. W. Hahn. 2013. "More Accurate Phylogenies Inferred From Low-Recombination Regions in the Presence of Incomplete Lineage Sorting." *Evolution* 67: 2376–2384.
- Penny, D., and M. D. Hendy. 1985. "The Use of Tree Comparison Metrics." *Systematic Zoology* 34: 75–82.
- Petlund, H. T. 2021. "A Genomic Delimitation of European *Sphagnum* Subgenus *Subsecunda* With Particular Emphasis on the Taxonomically Difficult Allopolyploids *Sphagnum auriculatum* Schimp. and *Sphagnum inundatum* Russow." Thesis, NTNU.
- Phung, T. N., C. D. Huber, and K. E. Lohmueller. 2016. "Determining the Effect of Natural Selection on Linked Neutral Divergence Across Species." *PLoS Genetics* 12: e1006199.
- Pollard, D. A., V. N. Iyer, A. M. Moses, and M. B. Eisen. 2006. "Widespread Discordance of Gene Trees With Species Tree in *Drosophila*: Evidence for Incomplete Lineage Sorting." *PLoS Genetics* 2: e173.
- Puttick, M. N., J. L. Morris, T. A. Williams, et al. 2018. "The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte." *Current Biology* 28: 733–745.e2.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>
- Rantanen, M., A. Y. Karpechko, A. Lipponen, et al. 2022. "The Arctic Has Warmed Nearly Four Times Faster Than the Globe Since 1979." *Communications Earth & Environment* 3: 1–10.
- Ravinet, M., K. Yoshida, S. Shigenobu, A. Toyoda, A. Fujiyama, and J. Kitano. 2018. "The Genomic Landscape at a Late Stage of Stickleback Speciation: High Genomic Divergence Interspersed by Small Localized Regions of Introgression." *PLoS Genetics* 14: e1007358.
- Renaut, S., C. J. Grassa, S. Yeaman, et al. 2013. "Genomic Islands of Divergence Are Not Affected by Geography of Speciation in Sunflowers." *Nature Communications* 4: 1–8.
- Rettelbach, A., A. Nater, and H. Ellegren. 2019. "How Linked Selection Shapes the Diversity Landscape in *Ficedula* Flycatchers." *Genetics* 212: 277–285.
- Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. "Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance." *Genetics* 105: 767–779.
- Riesch, R., M. Muschick, D. Lindtke, et al. 2017. "Transitions Between Phases of Genomic Differentiation During Stick-Insect Speciation." *Nature Ecology & Evolution* 1: 1–13.
- Rzhetsky, A., and M. Nei. 1992. "Statistical Properties of the Ordinary Least-Squares, Generalized Least-Squares, and Minimum-Evolution Methods of Phylogenetic Inference." *Journal of Molecular Evolution* 35: 367–375.

- Sarkar, D. 2008. *Lattice: Multivariate Data Visualization With R*. New York, NY, USA: Springer-Verlag.
- Schrider, D. R. 2020. "Background Selection Does Not Mimic the Patterns of Genetic Diversity Produced by Selective Sweeps." *Genetics* 216: 499–519.
- Schubert, M., L. Ermini, C. D. Sarkissian, et al. 2014. "Characterization of Ancient and Modern Genomes by SNP Detection and Phylogenomic and Metagenomic Analysis Using PALEOMIX." *Nature Protocols* 9: 1056–1082.
- Seehausen, O., R. K. Butlin, I. Keller, et al. 2014. "Genomics and the Origin of Species." *Nature Reviews Genetics* 15: 176–192.
- Shaw, A. J., N. Devos, C. J. Cox, et al. 2010. "Peatmoss (*Sphagnum*) Diversification Associated With Miocene Northern Hemisphere Climatic Cooling?" *Molecular Phylogenetics and Evolution* 55: 1139–1145.
- Shen, W., S. Le, Y. Li, and F. Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *PLoS One* 11: e0163962.
- Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. "Estimating Individual Admixture Proportions From Next Generation Sequencing Data." *Genetics* 195: 693–702.
- Slatkin, M., and J. L. Pollack. 2006. "The Concordance of Gene Trees and Species Trees at Two Linked Loci." *Genetics* 172: 1979–1984.
- Slotte, T., J. P. Foxe, K. M. Hazzouri, and S. I. Wright. 2010. "Genome-Wide Evidence for Efficient Positive and Purifying Selection in *Capsella grandiflora*, a Plant Species With a Large Effective Population Size." *Molecular Biology and Evolution* 27: 1813–1821.
- Smith, J. M., and J. Haigh. 1974. "The Hitch-Hiking Effect of a Favourable Gene." *Genetical Research* 23: 23–35.
- Stajich, J. E., and M. W. Hahn. 2005. "Disentangling the Effects of Demography and Selection in Human History." *Molecular Biology and Evolution* 22: 63–73.
- Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. "Widespread Selection and Gene Flow Shape the Genomic Landscape During a Radiation of Monkeyflowers." *PLoS Biology* 17: e3000391.
- Stenøien, H. K., and S. M. Sæstad. 1999. "Genetic Structure in Three Haploid Peat Mosses (*Sphagnum*)." *Heredity* 82: 391–400.
- Stenøien, H. K., and S. M. Sæstad. 2001. "Genetic Variability in Bryophytes: Does Mating System Really Matter?" *Journal of Bryology* 23: 313–318.
- Szövényi, P., N. Devos, D. J. Weston, et al. 2014. "Efficient Purging of Deleterious Mutations in Plants With Haploid Selfing." *Genome Biology and Evolution* 6: 1238–1252.
- Szövényi, P., A. Gunadi, and F.-W. Li. 2021. "Charting the Genomic Landscape of Seed-Free Plants." *Nature Plants* 7: 554–565.
- Szövényi, P., S. Terracciano, M. Ricca, S. Giordano, and A. J. Shaw. 2008. "Recent Divergence, Intercontinental Dispersal and Shared Polymorphism Are Shaping the Genetic Structure of Amphiatlantic Peatmoss Populations." *Molecular Ecology* 17: 5364–5377.
- Tajima, F. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123: 585–595.
- Temsch, E. M., J. Greilhuber, and R. Krisai. 1998. "Genome Size in *Sphagnum* (Peat Moss)." *Botanica Acta* 111: 325–330.
- Turetsky, M. R. 2003. "The Role of Bryophytes in Carbon and Nitrogen Cycling." *Bryologist* 106: 395–409.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. "Genomic Islands of Speciation in *Anopheles gambiae*." *PLoS Biology* 3: e285.
- Twyford, A. D., C. A. Kidner, and R. A. Ennos. 2015. "Maintenance of Species Boundaries in a Neotropical Radiation of *Begonia*." *Molecular Ecology* 24: 4982–4993.
- Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. "Correlated Patterns of Genetic Diversity and Differentiation Across an Avian Family." *Molecular Ecology* 26: 3982–3997.
- Vijay, N., C. M. Bossu, J. W. Poelstra, et al. 2016. "Evolution of Heterogeneous Genome Differentiation Across Multiple Contact Zones in a Crow Species Complex." *Nature Communications* 7: 13195.
- Wang, J., N. R. Street, E.-J. Park, J. Liu, and P. K. Ingvarsson. 2020. "Evidence for Widespread Selection in Shaping the Genomic Landscape During Speciation of *Populus*." *Molecular Ecology* 29: 1120–1136.
- Wang, R. J., and M. W. Hahn. 2018. "Speciation Genes Are More Likely to Have Discordant Gene Trees." *Evolution Letters* 2: 281–296.
- Wang, Y., J. Lu, J. Yu, R. A. Gibbs, and F. Yu. 2013. "An Integrative Variant Analysis Pipeline for Accurate Genotype/Haplotype Inference in Population NGS Data." *Genome Research* 23: 833–842.
- Wei, T., and V. Simko. 2017. "R Package 'Corrplot': Visualization of a Correlation Matrix."
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing.
- Wilson, R. M., A. M. Hopple, M. M. Tfaily, et al. 2016. "Stability of Peatland Carbon to Rising Temperatures." *Nature Communications* 7: 13723.
- Wolf, J. B. W., and H. Ellegren. 2017. "Making Sense of Genomic Islands of Differentiation in Light of Speciation." *Nature Reviews Genetics* 18: 87–100.
- Wu, C.-I. 2001. "The Genic View of the Process of Speciation." *Journal of Evolutionary Biology* 14: 851–865.
- Yamasaki, Y. Y., R. Kakioka, H. Takahashi, et al. 2020. "Genome-Wide Patterns of Divergence and Introgression After Secondary Contact Between *Pungitius* Sticklebacks." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 375: 20190548.
- Yousefi, N., K. Hassel, K. I. Flatberg, et al. 2017. "Divergent Evolution and Niche Differentiation Within the Common Peatmoss *Sphagnum magellanicum*." *American Journal of Botany* 104: 1060–1072.
- Yousefi, N., E. Mikulášková, H. K. Stenøien, et al. 2019. "Genetic and Morphological Variation in the Circumpolar Distribution Range of *Sphagnum warnstorffii*: Indications of Vicariant Divergence in a Common Peatmoss." *Botanical Journal of the Linnean Society* 189: 408–423.
- Yu, Z., D. W. Beilman, S. Frolking, et al. 2011. "Peatlands and Their Role in the Global Carbon Cycle." *Eos, Transactions American Geophysical Union* 92: 97–98.
- Zeng, K. 2013. "A Coalescent Model of Background Selection With Recombination, Demography and Variation in Selection Coefficients." *Heredity* 110: 363–371.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.