Received: 2 May 2023 | Revised: 13 July 2023 | Accepted: 28 July 2023

DOI: 10.1111/mec.17100

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Holocentric repeat landscapes: From micro-evolutionary patterns to macro-evolutionary associations with karyotype evolution

Camille Cornet[1] | Pablo Mora[2,3] | Hannah Augustijnen[4] | Petr Nguyen[3] | Marcial Escudero[5] | Kay Lucek[1]

[1]Biodiversity Genomics Laboratory, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[2]Department of Experimental Biology, Genetics Area, University of Jaén, Jaén, Spain

[3]University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic

[4]Department of Environmental Sciences, University of Basel, Basel, Switzerland

[5]Department of Plant Biology and Ecology, University of Seville, Seville, Spain

Correspondence
Camille Cornet and Kay Lucek, Biodiversity Genomics Laboratory, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland.
Email: camille.cornet@unine.ch and kay.lucek@unine.ch

Funding information
Agencia Estatal de Investigación, Grant/Award Number: PID2021-122715NB-I00; Grantová Agentura České Republiky, Grant/Award Number: 23-06455S; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 310030_184934 and PCEFP3_202869; Spanish Ministry of Universities - European Union's NextGenerationEU, Grant/Award Number: UJAR10MS

Handling Editor: Cristina Arias-Sardá

Abstract

Repetitive elements can cause large-scale chromosomal rearrangements, for example through ectopic recombination, potentially promoting reproductive isolation and speciation. Species with holocentric chromosomes, that lack a localized centromere, might be more likely to retain chromosomal rearrangements that lead to karyotype changes such as fusions and fissions. This is because chromosome segregation during cell division should be less affected than in organisms with a localized centromere. The relationships between repetitive elements and chromosomal rearrangements and how they may translate to patterns of speciation in holocentric organisms are though poorly understood. Here, we use a reference-free approach based on low-coverage short-read sequencing data to characterize the repeat landscape of two independently evolved holocentric groups: *Erebia* butterflies and *Carex* sedges. We consider both micro- and macro-evolutionary scales to investigate the repeat landscape differentiation between *Erebia* populations and the association between repeats and karyotype changes in a phylogenetic framework for both *Erebia* and *Carex*. At a micro-evolutionary scale, we found population differentiation in repeat landscape that increases with overall intraspecific genetic differentiation among four *Erebia* species. At a macro-evolutionary scale, we found indications for an association between repetitive elements and karyotype changes along both *Erebia* and *Carex* phylogenies. Altogether, our results suggest that repetitive elements are associated with the level of population differentiation and chromosomal rearrangements in holocentric clades and therefore likely play a role in adaptation and potentially species diversification.

KEYWORDS
Carex, *Erebia*, Lepidoptera, speciation, transposable elements

1 | INTRODUCTION

Repetitive DNA elements (repeats) are major components of most eukaryotic genomes and are increasingly recognized as important drivers of adaptation and speciation (Biémont & Vieira, 2006). Transposable elements (TEs), for example, are mobile DNA sequences propagating within genomes, occasionally generating adaptive variation (Schrader & Schmitz, 2019). For instance, the industrial melanism of the peppered moth *Biston betularia*, a textbook example of adaptation, is caused by the insertion of a TE in the wing patterning gene, which resulted in its altered expression (Van't Hof et al., 2016). Repeats may also contribute to reproductive isolation between populations and even promote speciation. This has been suggested for the adaptive radiation of *Anolis* lizards, where TEs accumulated in clusters of developmental genes may have facilitated morphological adaptations (Feiner, 2016). The insertion density of TEs is similarly correlated with speciation rates in mammals (Ricci et al., 2018).

Repetitive elements are also important drivers of genome evolution through chromosomal rearrangements, such as inversions, fusions and fissions of chromosomes (Fedoroff, 2012; Lönnig & Saedler, 2002). These large-scale rearrangements can be caused by ectopic recombination, that is recombination events between repeat copies from different genomic regions (Cáceres et al., 1999; Delprat et al., 2009). Repeat-mediated chromosomal rearrangements could promote speciation through different processes. First, by reducing fitness of heterokaryotypes, that is hybrids that are heterozygous for chromosomal rearrangements (Faria & Navarro, 2010; Rieseberg, 2001). However, this mode of chromosomal speciation suffers from the so-called '*underdominance paradox*', whereby highly deleterious rearrangements resulting in unfit hybrids are unlikely to become fixed in a population (Spirito, 1998). Second, novel rearrangements could suppress recombination between rearranged parts of the genome (Faria & Navarro, 2010; Rieseberg, 2001) or change gene expression (Li et al., 2023). The suppressed recombination scenario relies on reduced gene flow in rearranged genomic regions and on linkage disequilibrium between genes involved in reproductive isolation (Rieseberg, 2001).

Chromosomal speciation research has primarily focussed on organisms with monocentric chromosomes that have a single localized centromere per chromosome (Coyne & Orr, 2004; White, 1978). However, the centromeric activity can be distributed along a large portion of the so-called holocentric chromosomes (Mandrioli & Manicardi, 2020). Holocentricity has evolved independently at least 19 times in various clades across the tree of life including Lepidoptera (i.e. butterflies and moths), some Angiosperms (e.g. sedges including the genus *Carex*) or nematodes (including the model species *Caenorhabditis elegans*) (Escudero, Márquez-Corro, & Hipp, 2016; Melters et al., 2012). Chromosomal rearrangements such as fusions and fissions are supposedly less deleterious in holocentric than in monocentric organisms (Lucek et al., 2022; Melters et al., 2012) and meiotic adaptations to holocentricity can further mitigate the underdominance of the chromosomal rearrangements (Lukhtanov et al., 2018). Interestingly, karyotype diversity is correlated with diversification rates in angiosperms (Carta & Escudero, 2023) and in butterflies (de Vos et al., 2020), especially in some clades such as *Erebia* (Augustijnen et al., 2023) and Polyommatinae butterflies (Talavera et al., 2013). Chromosomal rearrangements can act as partial reproductive barriers in holocentric taxa as distant as butterflies (Lukhtanov et al., 2018; Mackintosh et al., 2023) and sedges of the genus *Carex* (Escudero, Hahn, et al., 2016). This suggests that chromosomal rearrangements might commonly be involved in speciation of holocentric groups.

The genetic features and molecular mechanisms underlying fusion and fission of holocentric chromosomes are mostly unknown, although different types of repeats have recently been implicated in *Rhynchospora* (Hofstatter et al., 2022) and *Carex* sedges (Escudero et al., 2023), butterflies (Ahola et al., 2014; Höök et al., 2023), aphids (Mathers et al., 2021) and nematodes (Yoshida et al., 2023). Holocentric species constitute some of the most karyotypically diverse taxonomic groups, including several insect clades (e.g. Lepidoptera, $n = 5–223$; de Vos et al., 2020) and plants (e.g. *Carex* sedges, $n = 5–66$; Hipp et al., 2009; Márquez-Corro et al., 2021). However, this diversity in chromosome numbers is not equally represented among holocentric clades, and the reasons for these disparities are still unknown. For example, most butterfly genera have retained the putative ancestral chromosome number of $n = 31$ (de Vos et al., 2020), while others show increased karyotypic diversity (e.g. chromosomes numbers ranging from $n = 7$ to $n = 52$ in the genus *Erebia*; Augustijnen et al., 2023). An explanation to the unequal representation of karyotypic diversity among holocentric clades could be that the karyotypically conserved clades lack the repeats involved in chromosomal rearrangements.

Using a reference-free approach on low-coverage short-read sequencing data, we first aimed to uncover the diversity of repeats and test how these may differ among species with different degrees of intraspecific differentiation. This could inform us on the potential association between repeats and population differentiation. For this, we compared individuals from different populations of four *Erebia* species and tested to which extent divergent populations also differ in their repeat landscape. We expected a positive correlation between the overall genetic differentiation among populations and level of differentiation of their repeat landscape (Bourgeois & Boissinot, 2019). We then employed macro-evolutionary inferences at the genus level to determine whether repeats could be associated with karyotype changes along the phylogeny of *Erebia*, with an emphasis on the most species-rich and karyotypically variable subclade Tyndarus (Augustijnen et al., 2023). Finally, we performed the same phylogenetic analyses for a subset of *Carex* species to test whether similar macro-evolutionary patterns would occur in independently evolved holocentric groups.

2 | MATERIALS AND METHODS

2.1 | Study species and sample collection

The main focus of our study was on the Palearctic genus *Erebia*, which comprises ~100 species of primarily alpine butterflies with closely related species often forming narrow contact zones in the

Alps (Augustijnen et al., 2022; Cupedo, 2014; Peña et al., 2015). For the species-level analyses of *Erebia*, we focussed on four species that fall along a gradient of intraspecific genetic differentiation: (i) *E. cassioides* colonized part of the Western Alps since the last glaciation, likely from a refugia in Southwestern Europe (Lucek et al., 2020), and shows little genetic differentiation to populations from the Pyrenees or the French Massive Central (Schmitt et al., 2016); (ii) *Erebia tyndarus* is endemic to the central Alps and shows a higher population structure than *E. cassioides* in the Alps (Gratton et al., 2016; Schmitt et al., 2016), potentially because *E. tyndarus* is less affected by range expansion. For (iii) *E. nivalis*, we included samples from their disjunct populations in the Alps, that is from Switzerland and Austria, that are considered distinct subspecies (Table S1; Figure S1; Schmitt et al., 2016; Sonderegger, 2005). Finally, for iv) *E. pronoe*, we included individuals of two nominal, ecologically differentiated subspecies, that is *E. pronoe psathura* ($n = 6$) and *E. pronoe vergy* ($n = 5$; Table S1; Figure S1). We wondered whether the levels of population differentiation in those four species are also mirrored in the repeat landscape. *Erebia cassioides*, *E. tyndarus* and *E. nivalis* are sibling species and belong to the karyotypically diverse Tyndarus clade within *Erebia*. For each of these three species, we sampled 2-3 individuals from 4-7 sites (Table S1; Figure S1). *Erebia cassioides* and *E. tyndarus* form secondary contact zones in populations GRW and GRF (Sonderegger, 2005), with very few F1 hybrids and almost no introgression (Augustijnen et al., 2022; Lucek et al., 2020). To ensure that this would not cause a bias in our analyses, we selected individuals as far away as possible from the point of secondary contact in those populations. For the *Erebia* genus-level (i.e. macro-evolutionary) analyses, we took advantage of a short-read re-sequencing data set (NCBI BioProject PRJNA1000734) that was used to establish a dated phylogeny of *Erebia* (Augustijnen et al., 2023). We used species for which chromosome number information was available from the literature. As *Erebia* is one of the best karyotyped butterfly genera (de Vos et al., 2020; Robinson, 1971), this resulted in 47 species in total (Table S2).

Finally, to assess whether there could also be a macro-evolutionary association between repeats and karyotype changes in an independently evolved holocentric clade with high karyotypic diversity, we extended our study to the plant genus *Carex* (Cyperaceae). The genus *Carex* with ca. 2000 species is the largest genus in the family Cyperaceae and one of the largest angiosperm genera, with the highest species richness in the temperate areas of the northern hemisphere (Escudero et al., 2012). For the *Carex* genus-level analyses, we included 25 samples from 14 species, two individuals per species (exceptionally only one individual for three of the species; Table S3). These 14 species are distributed across the whole genus *Carex*, represent all major *Carex* lineages (with the exception of *Carex* subgenus *Siderosticta*; Global Carex Group et al., 2021) and differ in their karyotype (n = 17–42): eight species from *Carex* subgenus *Carex* representing six different sections (*Phacocystis*, *Limosae*, *Spirostachyae*, *Ceratocystis*, *Aulocystis* and *Mitratae*), five species from *Carex* subgenus *Vignea* representing four sections (*Glareosae*, *Stellutae*, *Ovales* and *Foetida*) and one species from *Carex* subgenus *Eutyceras* (sect. *Capituligerae*).

2.2 | DNA extraction and sequencing

All *Erebia* butterflies for the population level analyses were collected between 2006 and 2022 (Table S1) using hand nets and stored at −20°C. We obtained DNA from the thorax of each individual using the standard protocol of the Qiagen Blood & Tissue Kit (Qiagen AG). We outsourced Illumina library preparation, which included a PCR amplification step, to the Department of Biosystems Science and Engineering (DBSSE) of ETH Zürich in Basel, where they carried out the subsequent paired-end whole-genome resequencing on an Illumina NovaSeq 6000 platform. The samples used for *E. tyndarus* and *E. cassioides* were previously used in Augustijnen et al. (2022). *Carex* specimens were collected between 2005 and 2021 and stored in silica gel (Table S3). DNA was extracted using the Qiagen Dneasy Plant Pro Kit following the manufacturer's protocol. Sequencing was carried out at the same time and using the same protocol as the *Erebia* specimens.

2.3 | Repetitive elements identification and quantification

To detect, identify and quantify repeats in both *Erebia* and *Carex*, we used the graph-based clustering software RepeatExplorer2 (Novák et al., 2010, 2020). Briefly, RepeatExplorer2 uses low-coverage raw short-read data to cluster sequences based on their similarities (i.e. an all-to-all sequence similarity search). Read clusters correspond mostly to repeat families, which are subsequently identified based on the graph structure and on the consensus sequence of each cluster. The use of low-coverage data (<0.5X) ensures that sequences retained in the clusters are repeats rather than duplicated gene families. This allows for repeat identification and comparison between different individuals of the same species or even across species, without the need for reference genomes. RepeatExplorer2 includes an implementation of TAREAN (TAndem REpeat ANalyzer), a graph-based automated satellite DNA (satDNA) identification pipeline (Novák et al., 2017). This approach has been used extensively to characterize the repeat landscape of various organisms, including plants (e.g. Macas et al., 2015) and insects (e.g. Silva et al., 2019).

We trimmed raw reads based on quality (per base quality score > 30) retaining only reads longer than 120 bp and removed adaptors and poly-G tails using fastp 0.22.0 (Chen et al., 2018). We then subsampled trimmed reads to obtain a coverage of 0.1X using seqtk 1.3 (https://github.com/lh3/seqtk) assuming genome sizes of 500 Mbps for all *Erebia* species (Lohse, Hayward, et al., 2022; Lohse, Lohse, et al., 2022) and using flow cytometry genome size estimates for *Carex* (Pellicer & Leitch, 2020; https://cvalues.science.kew.org/; Table S3). We performed individual-based repeat detection and classification for each individual in the species-level

analyses and for each species in the genus-level analyses, using the command line version of RepeatExplorer2 with default settings and the built-in REXdb databases (METAZOA3.0 for *Erebia* and VIRIDIPLANTAE3.0 for *Carex*). We then used the comparative mode of RepeatExplorer2 to compare the repeat landscape of all individuals for each of the four *Erebia* species in the species-level analyses, and of all species for *Erebia* or *Carex* in the genus-level analyses. We then merged the top repeat clusters by annotation, corresponding to at least 0.01% of the analysed reads, excluding reads corresponding to organellar DNA sequences. We subsequently used the number of reads in each cluster as a proxy for repeat abundance of the respective genomes (Novák et al., 2010). For the comparative genus-level analysis of *Erebia*, comprising 47 species, we were restricted to using 0.04X coverage per species due to computational limitations. Therefore, we ran the pipeline twice, on two different read subsamples, to ensure reliability and repeatability of the results. To ensure that an entirely automated annotation would not bias our results, we also performed manual curation following Goubert et al. (2022) on the 250 most abundant repeat clusters (corresponding to 89% of the analysed reads) for the comparative genus-level analysis of *Erebia*.

2.4 | Species-level analyses

To compare the repeat patterns with individual genetic diversity and differentiation, we generated SNP data sets by aligning all raw reads of *E. cassioides*, *E. tyndarus* and *E. nivalis* against a reference assembly of *E. cassioides* (NCBI BioProject PRJNA941023) and for *E. pronoe* against the more closely related reference genome of *E. ligea* (Lohse, Hayward, et al., 2022). Filtering and SNP calling followed Augustijnen et al. (2023) for each dataset.

To characterize the diversity and differentiation in repeats between individuals and populations, we employed established community ecology approaches implemented in the R package vegan 2.6–4 (Oksanen et al., 2022). For this, we considered distinct repeat annotations as *'species'* and individual genomes as *'sampling sites'* (Haley & Mueller, 2022; Venner et al., 2009). First, to estimate the overall differentiation in repeat landscape between individuals in each of the four *Erebia* species, we conducted principal coordinate analyses (PCoA) using the R package ape 5.7–1 (Paradis & Schliep, 2019). The PCoA was performed on the Bray–Curtis dissimilarity matrix based on the repeat clusters of each individual calculated in vegan. We similarly performed a PCoA based on Euclidean distances among individuals for each SNP dataset, calculated with adegenet 2.1.10 (Jombart, 2008) in R. Second, we performed Permutational Multivariate Analyses of Variance (PERMANOVAs) for each species, with the function *adonis2* implemented in vegan and 9′999 permutations, to test for differences between populations in terms of repeat landscape. While performing these analyses, we noticed the presence of two pairs of putative half-siblings in our data sets (one in *E. tyndarus* and one in *E. nivalis*), confirmed by kinship coefficient analysis (0.29 and 0.31, respectively) in plink 1.9 (Chang et al., 2015; Table S4). Therefore, we randomly removed one individual per pair of half-siblings from all subsequent analyses.

We calculated Simpson's diversity index based on the abundances of each repeat type for each individual with the function *diversity* in vegan, and assessed whether this diversity would be correlated with the level of genetic diversity. For the latter, we used mlRho 2.9 (Haubold et al., 2010) which estimates individual-based expected zygosity (θ), using genome-wide SNP data. For mlRho, we only included sites with a minimum quality of 28, a maximum depth of 80X, and a minimum depth of 4X. We subsequently correlated θ and Simpson's diversity index for each of the four species with a Spearman correlation using the function *cor.test* in R.

For each species, we further compared the Bray–Curtis dissimilarity matrix based on repeat clusters of each individual against the matrix of individual-based pairwise Euclidean distances using the SNP data. The comparison between both distance matrices was performed using a distance-based Redundancy Analysis (db-RDA) using the package vegan following Benestan et al. (2021). We performed all analyses in R 4.2.2 (R Core Team, 2022) using Rstudio 2023.3.0.386 (Posit team, 2023).

2.5 | Genus-level analyses

We performed the genus-level analyses within a phylogenetic framework to test for an association between the evolution of the repeat landscape in *Erebia* and *Carex*, respectively, and changes in karyotypes, used as a proxy for large-scale chromosomal rearrangements (chromosomal fusions and fissions). The phylogeny of *Erebia* was taken from Augustijnen et al. (2023). In short, the *Erebia* species tree was established from 2′920 individual gene trees inferred with IQTREE (Nguyen et al., 2015) using a coalescent model implemented in ASTRAL (Zhang et al., 2018). For *Carex*, we used a dated phylogeny that represents the 70% of extant species and is based on three DNA regions (ITS, ETS and matK) and a HybSeq phylogeny backbone (Martín-Bravo et al., 2019). We pruned both phylogenies to the species used in this study using the function *drop.tip* in the R package ape, resulting in 47 *Erebia* species and 14 *Carex* species.

To determine whether the overall differentiation in repeat landscape between species shows a phylogenetic signal, we constructed a Bray–Curtis dissimilarity matrix based on repeat clusters for each species, for *Erebia* and *Carex* separately in vegan (Oksanen et al., 2022). We then contrasted the dissimilarity matrix with the phylogeny of *Erebia* or *Carex* using the R package dendextend 1.17.1 (Galili, 2015). To formally test for the presence of a phylogenetic signal in particular types of repeats, we used the R package phylosignal 1.3 (Keck et al., 2016) to compute Pagel's λ (Pagel, 1999) along with its statistical significance. If λ equals zero, there is no phylogenetic signal, that is closely related species are not more similar than distant ones. If λ equals one, the phylogenetic signal observed corresponds to Brownian motion (Molina-Venegas & Rodríguez, 2017).

To assess whether there could be a macro-evolutionary impact of repeats on chromosome numbers and therefore on chromosomal

rearrangements, for both *Erebia* and *Carex*, we fitted phylogenetic generalized least square (PGLS) models with diploid chromosome number as a dependent variable and each repeat abundance as explanatory variables, using the *gls* function in NLME 3.1–162 (Pinheiro et al., 2023). We used Pagel's λ coefficient, computed using the function *corPagel* in APE to take phylogenetic correlation into account. Because we had more information available for *Carex*, we fitted additional models including the rate of chromosomal evolution (calculated with CHROMOHISSE, Tribble et al., in preparation) and number of different karyotypes per species (Table S3) as explanatory variables. As some *Carex* species show intraspecific variation in chromosome numbers (Hipp et al., 2009), we used the mean chromosome number for each species. No intraspecific chromosome number variation has been reported in *Erebia* (Augustijnen et al., 2023; Robinson, 1971).

We used a Bayesian reversible-jump multiregime Ornstein–Uhlenbeck (OU) approach as implemented in the R package BAYOU 2.0 (Uyeda & Harmon, 2014) to infer major shifts in chromosome number and abundance of each repeat type in *Erebia*. The OU model has two components: the stochastic and the deterministic components. The stochastic component is a Brownian Motion model with a single parameter, sigma, which quantifies the rate of stochastic evolution of a given trait. The deterministic component has two parameters: theta and alpha. Theta is the optimum towards which the trait evolves, and alpha is the rate of evolution towards the optimum. Using the BAYOU approach, we estimated the overall sigma, overall alpha, a theta for each inferred optimum shift and an additional theta value for the root of the phylogeny for each trait (the diploid chromosome number and abundance of each repeat type). Shifts were considered significant if they showed a posterior probability (PP) higher than 0.30 (Uyeda & Harmon, 2014). We set up the analyses following Larridon et al. (2021). Because BAYOU implements a Bayesian approach, we ran at least two independent Markov Chain Monte Carlo (MCMC) analyses of 1–3 million generations for each of the variables (chromosome number and abundance of each repeat type) to test for convergence, with a burn-in of 25%–30% to consider parameters estimates only after reaching stationarity.

Finally, we used the *slouch.fit* function in the R package SLOUCH 2.1.4 (Kopperud et al., 2020) to test the effect of repeat landscape on chromosome number. In SLOUCH, chromosome number is modelled as evolving towards an optimum following an OU process that is a linear function of the predictors (abundance of each repeat type). The predictors are modelled as evolving on the tree according to a Brownian Motion process (Hansen et al., 2008).

REPEATMASKER identifies and masks repetitive elements in genomic DNA sequences. We ran REPEATMASKER 4.0.7. (Smit et al., 2013) using the Illumina short-reads from all the species with the default settings and the *'-a'* option to keep the alignment in order to quantify the divergence of repeats and to plot the overall repeat landscape. As custom library, we used a custom-made database combining three different datasets, coming from (i) RepBase2018, (ii) the Dfam 3.7 repeat libraries and (iii) REPEATEXPLORER2 contigs using the scripts *id_rmasker_rexp.py* and *annot_to_rexp.py* (https://github.com/fjruizruano/ngs-protocols; accessed 10.04.2023). This combined database was used to improve the detection of novel and divergent repeat elements, and the REPEATEXPLORER2 database was included to improve the detection of lineage-specific and novel repeat elements (including satDNAs) that may not be present in the RepBase2018 and Dfam 3.7 databases. Using the REPEATMASKER output and the Perl script *calcDivergenceFromAlign.pl*, we calculated the divergence for each repeat type. Then, using a custom script in R and GGPLOT2 3.3.6 (Wickham, 2016), we generated a repeat landscape for all the species included in our study. The repeat landscape allowed us to visualize the distribution and abundance of different repeat elements across the genomes of each species and infer the divergence for each repeat type.

3 | RESULTS

3.1 | Species-level analyses

The proportion of repeats in the individual genomes ranged between 26.4% and 28.6% for *E. cassioides*, 24.7% and 26.0% for *E. tyndarus*, and 21.8% and 25.9% for *E. nivalis* (Figure S2). The proportion of repeats in *E. pronoe* was consistently higher in the subspecies *vergy* (25.7%–34.7%) than in the subspecies *psathura* (24.1%–25.3%). The most abundant repeats in all four species were LINE retrotransposons (genome proportion ranging from 1.2% to 3.0%). Among the detected repetitive sequences, the proportion of unidentified repeats ranged between 18.7% and 32.7%. In all four species, some individuals showed an interestingly high amount of satDNA (e.g. PDM_2 in *E. cassioides*, GRO_2 in *E. nivalis*; Figure S2). Satellite DNA can accumulate on a nonrecombining female-specific W sex chromosome (e.g. Cabral-de-Mello et al., 2021), but this is not the cause of the observed differences between individuals, as the only female in the data set is GRI_3 in *E. nivalis* (Table S1).

The PCoA revealed increasing levels of differentiation between populations in terms of repeat landscape, corresponding to the increasing genetic differentiation from *E. cassioides* populations to *E. pronoe* subspecies. Indeed, *E. cassioides* populations showed similar repeat landscapes (Figure 1a), although there were significant differences between populations (PERMANOVA: $F_{6,14} = 4.49$, $p < .001$). *Erebia tyndarus* populations were more strongly separated (PERMANOVA: $F_{6,13} = 6.83$, $p < .001$), especially along the first axis (i.e. the axis explaining most of the differences) but also along the second axis for population ARO (Figure 1b). *Erebia nivalis* populations appeared even more distinct (Figure 1c), even though this differentiation between populations in terms of repeat landscape was less significant (PERMANOVA: $F_{3,6} = 3.38$, $p = .036$). The two Austrian populations GRO and SAJ formed a distinct cluster to the Swiss populations, suggesting distinct repeat landscapes between these geographically isolated populations. Finally, *E. pronoe* subspecies differed greatly in repeat landscape (PERMANOVA: $F_{1,9} = 8.75$, $p = .002$; Figure 1d), such a pattern being likely due to the higher overall repeat density in *E. pronoe vergy*, especially in the two individuals VER_2 and VER_4 (Figure S2). This increase in

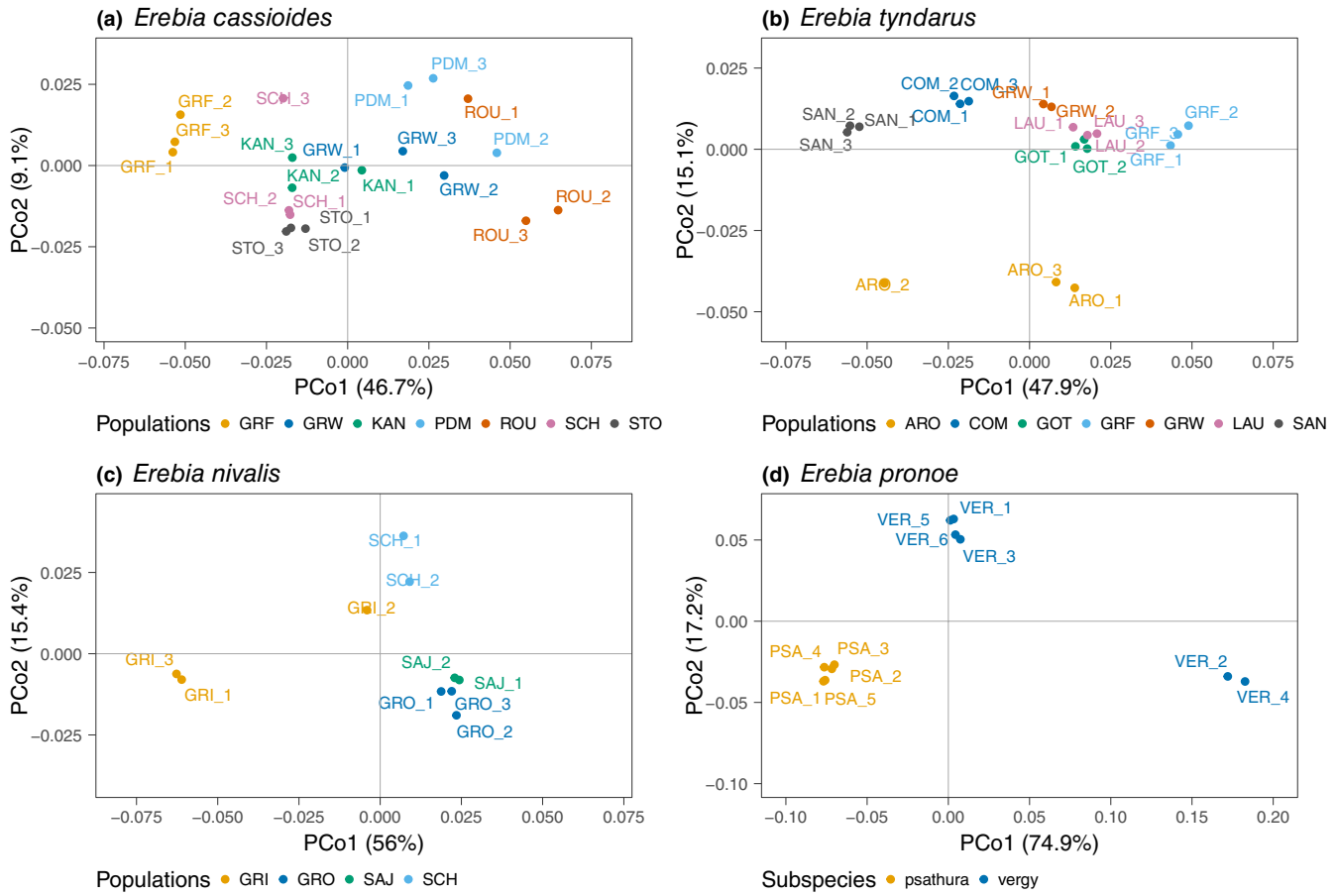

FIGURE 1 Principal Coordinate (PCo) analysis based on Bray–Curtis dissimilarity matrices of repetitive elements abundance and diversity in populations of four *Erebia* species in increasing order of genetic differentiation between populations: (a) *E. cassioides*, (b) *E. tyndarus*, (c) *E. nivalis* and (d) *E. pronoe*.

the levels of differentiation in repeat landscape from *E. cassioides* populations to *E. pronoe* subspecies was also noticeable by the increasing variation in repeat landscape accounted for by the two leading PCoA axes (46.7% and 9.1% for *E. cassioides* to 74.9% and 17.2% for *E. pronoe*; Figure 1). PCoAs on the Euclidean distances between individuals based on the SNP data set recovered similar patterns of differentiation among individuals (Figure S3).

There was no significant correlation between individual-based estimates of genetic diversity θ and Simpson's diversity index in repeat landscape for *E. cassioides*, *E. tyndarus* and *E. nivalis* ($p > .3$; Figure 2a), when it was significant for *E. pronoe* (rho = 0.74, $p = .013$; Figure 2a). The RDA revealed a significant positive association between Euclidean genetic distance between individuals and Bray–Curtis distance in repeat landscape for all four species (*E. cassioides*: $F_{4,16} = 1.19$, $p = .002$; *E. tyndarus*: $F_{3,16} = 1.44$, $p < .001$; *E. nivalis*: $F_{1,8} = 1.99$, $p < .001$; and *E. pronoe*: $F_{1,9} = 2.63$, $p = .015$; Figure 2b).

3.2 | Genus-level analyses

The proportion of repeats in the genomes of *Erebia* species (Figure 3a) ranged between 18.3% (*E. triarius*) and 33.2% (*E. pharte*). Concordantly with the species level analyses of *Erebia*, the most abundant identified repeats were LINE retrotransposons in all species (genome proportion ranging from 0.7% to 5.2%). More generally, retrotransposons (i.e. Class I TEs characterized by a *'copy-paste'* replication mechanism with an RNA intermediate; Wells & Feschotte, 2020) were much more common in *Erebia* (genome proportion ranging from 0.9% to 9.3%) than DNA transposons (i.e. Class II TEs characterized by a *'cut-and-paste'* replication mechanism with a DNA intermediate (Wells & Feschotte, 2020); genome proportion ranging from 0.01% to 0.6%). Among the identified repetitive sequences, the proportion of unidentified repeats ranged between 15.0% and 27.8%. As the genus-level analyses for *Erebia* were performed only on 0.04X genomic coverage, proportions of repeats in the genomes were computed twice on different runs of subsampling. The results were largely congruent, differing by less than 0.1% of the genome for each identified repeat type, and by 0.3% for unidentified repeats (Table S5a). The results were also largely congruent between the automated and manually curated repeat annotation for the 250 most abundant clusters in the *Erebia* dataset (Table S5b), and no wrongly annotated clusters were detected.

The proportion of repeats for the *Carex* species (Figures 3b and S4) was more variable, ranging between 13.7% (*C. capitata*) and

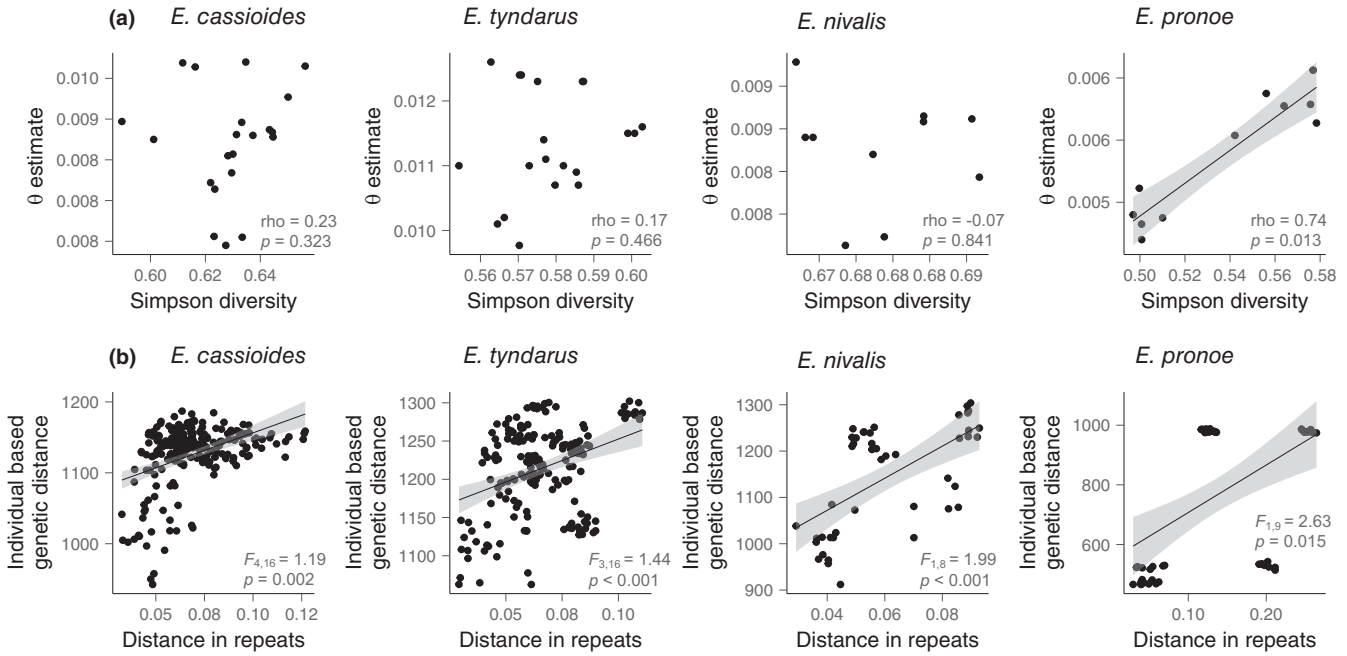

FIGURE 2 Correlations between genome-wide SNP-based estimates and estimates based on repetitive elements abundance and diversity for *Erebia cassioides*, *E. tyndarus*, *E. nivalis* and *E. pronoe*. (a) Correlations between individual-based θ estimates of genetic diversity and Simpson diversity index based on repeat landscape. (b) Correlations between Euclidean genetic distance between individuals and Bray-Curtis distance in repeat landscape. Significance was assessed with a distance-based Redundancy Analysis, even though straight lines and shaded grey areas represent a linear regression and the 95% confidence interval.

48.8% (*C. sempervirens*). The most abundant repeat types were LTR retrotransposons, particularly the Ty1/copia and Ty3/gypsy families (genome proportion ranging from 0.2% to 9.9% and from 0.03% to 20.0%, respectively). Interestingly, DNA transposons (especially the TIR family) were more abundant in *Carex* than in *Erebia* (*t*-test: $t_{1,24.8} = 5.90$, $p < .001$), with values higher than those of retrotransposons in some species (genome proportion ranging from 0.3% to 32.0% for retrotransposons and from 0.2% to 4.7% for DNA transposons). The proportion of identified rDNA was higher for *Carex* than for *Erebia* ($t_{1,13} = 5.88$, $p < .001$), while no statistical difference occurred for satDNA ($t_{1,26} = 0.53$, $p = .599$). Among the identified repetitive sequences, the proportion of unidentified repeats ranged between 10.1% and 24.8%. It was expected that this proportion would be lower for *Carex* than for *Erebia* as the RepeatExplorer2 pipeline was initially designed for plants (Novák et al., 2020). Genome size was positively correlated with total repeats proportion in the *Carex* genomes (PGLS: $t = 11.76$, $p < .001$). For 11 *Carex* species, two individuals were analysed per species, showing some intraspecific variation (Figure S5).

The overall repeat landscape showed a global concordance between the phylogeny and the clustering of the species based on the repeat landscape, both for *Erebia* (Figure 4a) and *Carex* (Figure 4b), although several notable discrepancies occurred. For example, the Tyndarus clade in *Erebia* clustered together based on the repeat landscape, except for *E. cassioides* and *E. rondui*. For *Carex*, the position of *C. sempervirens* (the species with the highest repeat proportion in the genome) was very different in the phylogeny and in the clustering based on the repeat landscape. It is important to note that, both for *Erebia* and *Carex*, the topology of the two trees differed at several scales (i.e. sister species, subclades and deeper nodes), suggesting that the repeat landscape as characterized here is insufficient for phylogenetic inference.

When statistically investigating the phylogenetic signal of specific repeat types, contrasting results were observed for *Erebia* (Table 1a) and *Carex* (Table 1b). For *Erebia*, LTR retrotransposons that could be identified down to repeat family level (i.e. Bel-Pao, Ty1/copia and Ty3/gypsy), Penelope retrotransposons and Helitron DNA transposons showed phylogenetic signals corresponding to Brownian motion (i.e. Pagel's λ close to one) that remained significant following a False Discovery Rate (FDR) correction ($p < .05$). On the contrary, retrotransposons of the families DIRS, LINE and LTR that could not be identified at a finer scale, and Maverick DNA transposons did not show a significant phylogenetic signal (i.e. closely related species were not significantly more similar in their repeat abundance than distantly related species; Pagel's λ close to zero; $p > .6$). For *Carex*, LINE, Ty3/gypsy, LTR retrotransposons that could not be identified to repeat family level showed a significant phylogenetic signal (Pagel's λ close to one; $p < .05$ following FDR). The same relationship for TIR DNA transposons was not significant following the FDR. Interestingly, both for *Erebia* and *Carex*, rDNA and satDNA did not show an overall phylogenetic signal (Pagel's λ close to zero; $p > .05$), suggesting that rDNA and satDNA evolved independently to the phylogeny and that other factors such as selection could be involved in the evolution of these two types of repetitive elements.

None of the associations between specific repeat types and chromosome number were significant following FDR (Table 1;

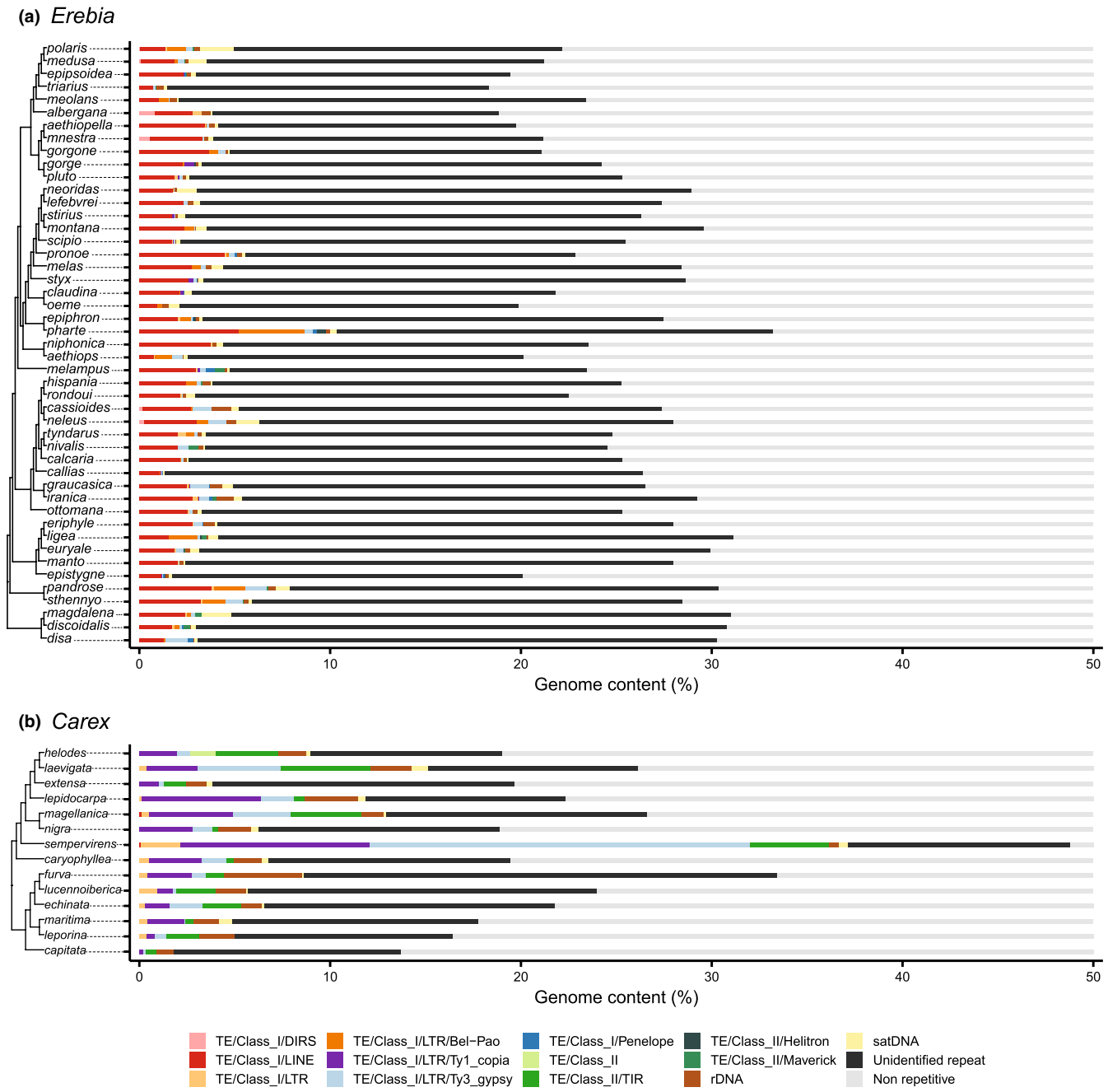

FIGURE 3 Proportion of repetitive elements in the genomes of (a) *Erebia* and (b) *Carex* species, calculated by the proportion of reads in clusters corresponding to repetitive elements in RepeatExplorer2. Unrooted phylogenies are represented on the left. TE—Transposable Element, rDNA—ribosomal DNA, satDNA—satellite DNA.

Figure S6), which could also indicate limited statistical power given our sample size. We summarize the relationships that were significant before applying the FDR as they could hint towards potential associations that would be ideally tested at a larger scale in future. First, there could be an association between the abundance of LTR retrotransposons and chromosome number for both *Erebia* and *Carex* (Table 1). The association was positive for *Erebia*, suggesting that a higher LTR abundance in the genome could corresponds to a higher chromosome number, contrarily to *Carex* for which this association was negative, suggesting that a higher LTR abundance in the genome corresponds to a lower chromosome number. Second, in *Erebia*, DNA transposons were negatively associated with chromosome number for Helitrons and positively for Mavericks. Lastly, in *Carex*, Ty3/Gypsy retrotransposons were negatively associated with chromosome number. When adding other variables in the model for *Carex* (Table S6), rates of chromosome evolution were positively associated with the abundance of LINEs ($t = 5.68$, $p < .001$), LTRs ($t = 2.99$, $p = .014$) and Helitrons ($t = 3.60$, $p = .005$) following an FDR, when the same relationship for Ty1/copia and Ty3/gypsy did not remain significant following the FDR. Similarly, the negative

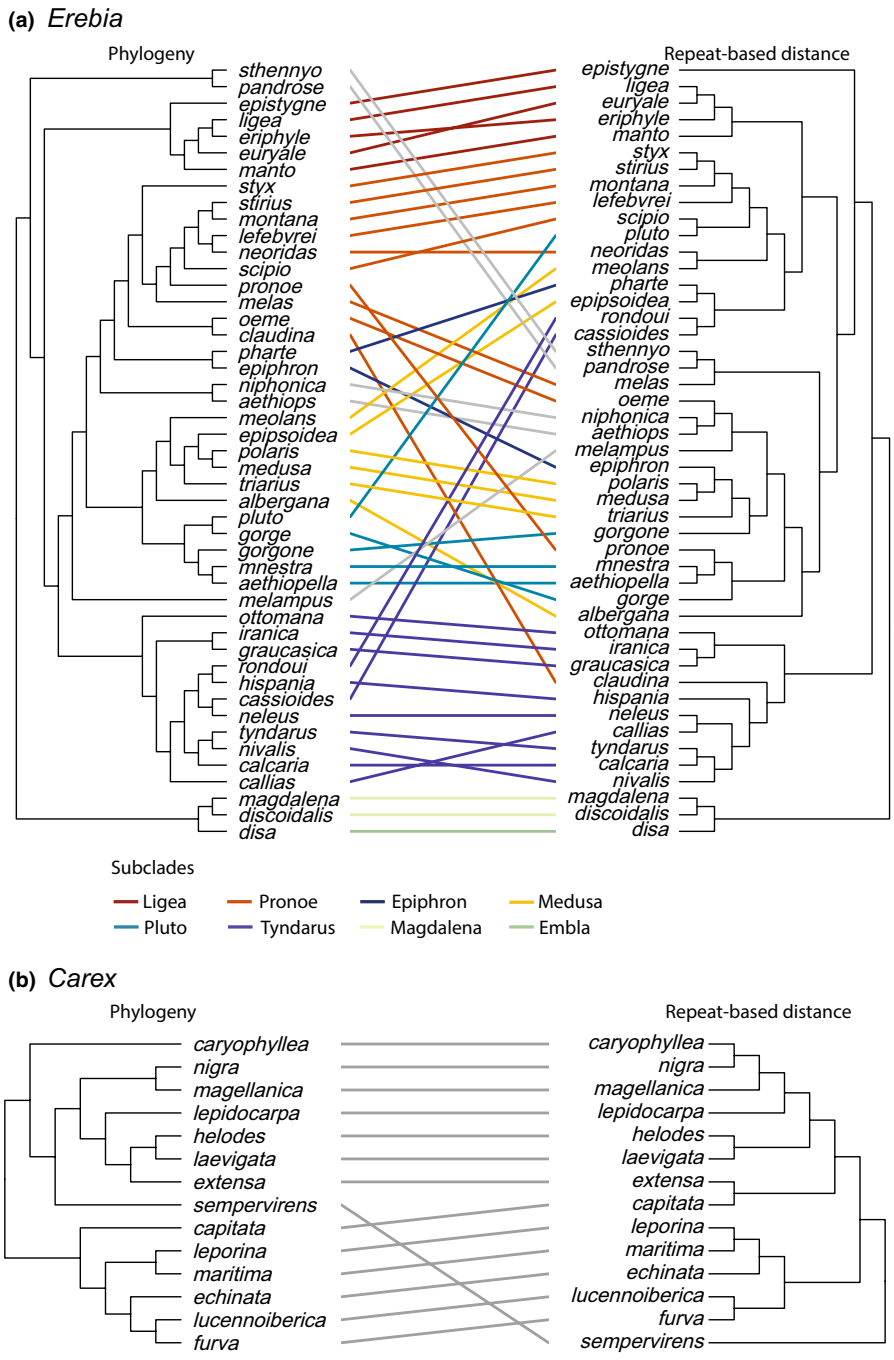

FIGURE 4 Correspondence between the unrooted phylogeny (left side) and Bray–Curtis distance between species based on the repeat landscape (right side), for (a) *Erebia* and (b) *Carex*. Lines for *Erebia* are coloured according to the different subgenera defined in Augustijnen et al. (2023). Tyndarus is the most species-rich and karyotypically variable clade.

association between chromosome number with the abundance of LTR and Ty3/gypsy were nonsignificant following the FDR. Number of different karyotypes per species was never significantly associated with repeats (all $p > .074$, Table S6).

The estimated sample sizes (ESSs) for all parameters estimated in BAYOU were > 200, which indicate that the Bayesian analyses reached convergence and stationarity. The BAYOU results displayed no significant shift in optima for some of the repeat landscape variables (LINEs, Ty3/gypsy, Penelope, and Maverick; Table S7, Figure S7). However, chromosome number and most of the repeat variables (DIRS, Bel-Pao, LTRs, Ty1/copia, Helitrons, 45S rDNA and satDNA) showed significant shifts in optima. Chromosome number showed seven shifts (eight optima), three of the four with the highest PPs occurred within the karyotypically diverse and species rich Tyndarus group (Table S7, Figure 5a). DIRS, Bel-Pao, LTRs, Ty1/copia and satDNA showed 1–4 shifts but related to terminal branches or small clades (2–3 species; Table S7, Figure S7). Exceptionally, one of the shifts inferred for satDNA corresponded to a bigger clade, the Pronoe group (Table S7; Figure 5d). Helitrons (Table S7; Figure 5b) and 45S rDNA (Table S7; Figure 5c) showed four shifts each, 3 and 4 of them, respectively, related to the Tyndarus group.

TABLE 1 Transposable element families identified in (a) *Erebia* and (b) *Carex*, the phylogenetic signal in their genomic abundance (Pagel's λ), the *p*-value associated with λ (testing if λ is significantly different from 1) and their association with diploid (2*n*) chromosome number changes (tested by phylogenetic generalized least square regression, PGLS).

	Phylogenetic signal		Chromosome 2*n*	
	Pagel's λ	*p*	*t*	*p*
(a) *Erebia*				
TE/Class_I/DIRS	0.00	1.000	−1.70	.097
TE/Class_I/LINE	0.00	1.000	0.96	.341
TE/Class_I/LTR	0.00	1.000	2.15	.037
TE/Class_I/LTR/Bel-Pao	1.01	.008*	−0.78	.437
TE/Class_I/LTR/Ty1_copia	1.14	<.001***	−1.45	.154
TE/Class_I/LTR/Ty3_gypsy	0.85	.003**	1.55	.128
TE/Class_I/Penelope	1.16	<.001***	0.83	.409
TE/Class_II/Helitron	0.57	.003**	−2.66	.011
TE/Class_II/Maverick	0.24	.617	2.10	.041
rDNA	0.37	.217	1.69	.098
satDNA	0.00	1.000	1.20	.235
(b) *Carex*				
TE/Class_I/LINE	1.07	<.001***	−1.52	.153
TE/Class_I/LTR	1.07	.003*	−2.59	.024
TE/Class_I/LTR/Ty1_copia	1.05	.058	−2.02	.066
TE/Class_I/LTR/Ty3_gypsy	1.07	.007*	−2.59	.024
TE/Class_II/TIR	1.06	.026	−0.95	.361
TE/Class_II/Helitron	0.92	.134	−0.42	.681
rDNA	0.00	1.000	1.55	.147
satDNA	0.30	.384	−0.70	.500

Note: A negative *t*-value represent a negative association between repeat abundance and chromosome number, and a positive *t*-value a positive association. Asterisks indicate significance after correcting for multiple testing (FDR) (*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$).

Abbreviations: TE, Transposable Element; rDNA, ribosomal DNA, satDNA, satellite DNA.

The null model (a flat relationship between 2n and any variable) in the SLOUCH analyses (2n ~1) showed a corrected Akaike Information Criterion (AICc) value of 399. The R^2 of all fitted models in slouch was small (<5%) and AICc was higher than the AICc of the null model, with only two exceptions (Table S8). Helitrons explained up to $R^2 = 11.60\%$ of chromosome number variation and the model showed the lowest AICc value (396). The model with the second highest R^2 value was the one that included DIRS as predictor ($R^2 = 6.44\%$, AICc = 398).

The REPEATMASKER analysis revealed contrasting repeat landscapes between *Erebia* species (Figure 6 and Data S1). Species belonging the karyotypically diverse and species-rich Tyndarus group (e.g. *E. cassioides*; Figure 6a) tended to show lower Kimura substitution values than species belonging to more karyotypically conserved groups (e.g. *E. epistygne*; Figure 6b), suggesting lower sequence divergence between copies and thus younger and more active repeats. Moreover, certain repeat types showed low Kimura values in most *Erebia* species, namely Helitrons, Maverick, satDNA and rDNA. On the contrary, the abundant LINEs consistently showed higher Kimura values. In *Carex*, the repeats consistently showing the lowest Kimura values were Ty1/Copia, Ty3/Gypsy, rDNA and satDNA (Data S2).

4 | DISCUSSION

Repetitive elements can drive chromosome evolution and have been suggested to be associated with chromosomal rearrangements through ectopic recombination (Fedoroff, 2012). Species with holocentric chromosomes, that is without a defined centromere, are thought to be more prone to retain chromosomal rearrangements than monocentric species (Melters et al., 2012). However, the impact of repeats on chromosomal rearrangements and chromosomal speciation in holocentric species is poorly understood. Here, we explored repeat landscapes across different levels, from intraspecific population structure to their association with karyotype changes in the holocentric clades *Erebia* and *Carex*. We used a reference-free low coverage (~0.1X) approach (Novák et al., 2010, 2020). This enabled us to increase the taxonomic scope of our analyses, as still very few reference genomes are available.

At a micro-evolutionary level, we found evidence for an increased differentiation in the repeat landscapes that scales with the level of overall genetic differentiation. Furthermore, our results suggest that repeats are associated with large-scale chromosomal rearrangements leading to karyotype changes in both *Erebia* and in *Carex*. We discuss these findings in the context of chromosomal speciation in

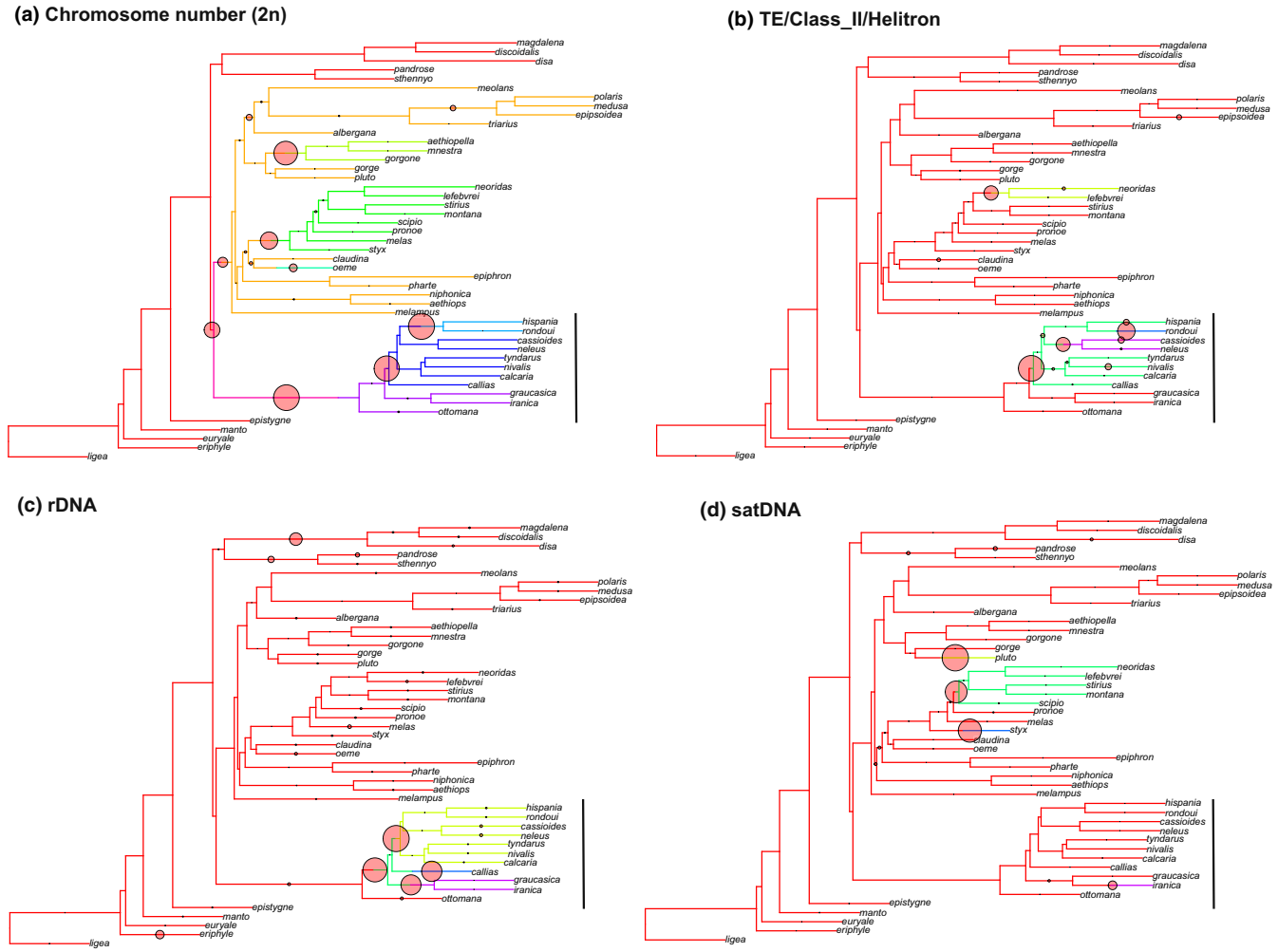

FIGURE 5 Bayou results showing shifts in optima along the *Erebia* phylogeny for chromosome number (a), and genomic abundance of Helitrons (b), rDNA (c) and satDNA (d). Significant shifts (PP>.3) in optima are represented by different branch colours, and red circles represent the PP of the shift. The analyses were done on an unrooted phylogeny of *Erebia*, while the trees were artificially rooted on *E. ligea*. The black vertical lines highlight the species-rich and karyotypically variable Tyndarus clade.

holocentric species, aiming to shed light on how repeats might be implicated in population differentiation at micro-evolutionary levels and chromosomal rearrangements and potentially speciation at macro-evolutionary scale.

4.1 | Micro-evolutionary repeat patterns

Previous studies describing differences in repeats between populations found population specific repeat expansions or losses (e.g. Feliciello et al., 2015; Oggenfuss et al., 2021). Such differences have been suggested to be a result of different mutational processes, genetic drift, natural selection and/or local adaptation (Bourgeois & Boissinot, 2019; Charlesworth & Charlesworth, 1983). In small populations, drift might result in the random loss of rare repeats, or in repeat expansion due to reduced efficacy of selection against deleterious TE insertions (Charlesworth & Charlesworth, 1983). Selection, on the contrary, might favour different repeat landscapes in different environments, because of their potential phenotypic effects (e.g. Huang et al., 2018; Schrader et al., 2014). Comparing four different *Erebia* species, we found that population specific changes in the repeat landscape are common and increase with the level of intraspecific differentiation (Figure 1). Similarly, individual-based genetic distances correlated with distances in repeat landscape in all species (Figure 2b), suggesting that similar evolutionary processes affected the repeat landscape and the genome as a whole. Incongruence between overall genetic distance and distance in repeat landscape potentially represent different selective regimes acting on repeats and on the rest of the genome (Oggenfuss et al., 2021).

Repeat landscape variation between populations can lead to the build-up of genetic incompatibilities, eventually leading to reproductive isolation (Serrato-Capuchina & Matute, 2018). Differences in the repeat landscape may for example lead to differences in gene expression (Rebollo et al., 2012) or differences in chromatin structure (Feliciello et al., 2015). Variation between populations in their repeats could also lead to variation in ectopic recombination

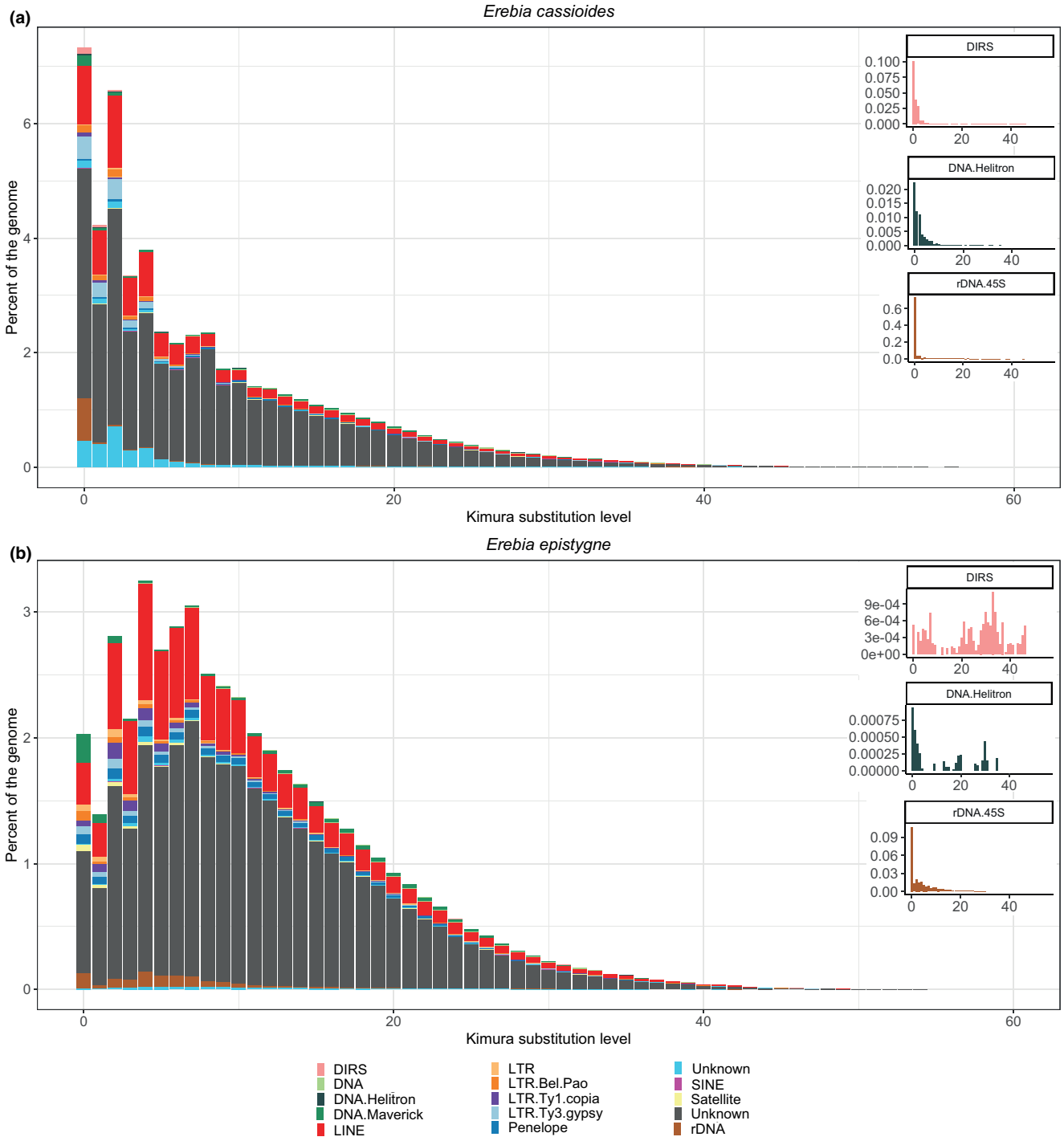

FIGURE 6 Repeat landscape of *Erebia cassioides* (a) and *E. epistygne* (b) determined using RepeatMasker, showing genome repeat proportion in function of Kimura substitution levels. Inserts show specifically the values for DIRS, Helitrons and rDNAs.

and chromosomal rearrangements between populations (Serrato-Capuchina & Matute, 2018). The observed increase in differentiation in the repeat landscape with increasing population divergence could thus reflect increasing levels of reproductive isolation because of drift or local adaptation. This may especially be true for *E. nivalis*, where the Swiss population is geographically strongly restricted and isolated from its broader occurrence in the Eastern Alps (Gratton et al., 2016). Similarly, the strong difference in the repeat landscape for the two subspecies of *E. pronoe*, which even occur on different substrates, may indicate some level of reproductive isolation (Sonderegger, 2005).

Repetitive elements diversity has previously been suggested to be associated with stochastic demographic processes (Bourgeois & Boissinot, 2019). Repeat diversity can be reduced due to genetic drift, along with genome-wide diversity (Stritt et al., 2018). However, our results indicate that this pattern is not ubiquitous and

that other evolutionary processes such as selection might blur the correlation between repeat diversity and genome-wide diversity (Figure 2a). Overall, our findings imply that repeats hold valuable information to characterize population differentiation, despite being often ignored in population genetic analyses (Slotkin, 2018). In addition, these results were obtained using low coverage (~0.1X), and without the need for a reference genome (Novák et al., 2010, 2020), highlighting the relevance of repeats for studies with limited funds and genomic resources.

4.2 | A macro-evolutionary view on karyotype evolution and repeat landscape

The most abundant repeats in the genomes of all *Erebia* species were LINE retrotransposons (Figure 3a). This is consistent with previous studies of the repeat landscape of Lepidoptera (Höök et al., 2023; Sproul et al., 2022). The abundance of LINEs probably results from phylogenetic constraints and could have phenotypic consequences due to the high association between LINEs and protein-coding genes (Sproul et al., 2022). For *Carex*, the most abundant repeats differed between species (Figure 3b), particularly for *C. sempervirens* which had by far the highest repeat content among all *Carex* species. Such a variable repeat abundance between species was also found in *Rhynchospora*, another holocentric Cyperaceae genus (Costa et al., 2021). Although full manual curation of our repeat annotations would likely allow for more detailed repeat identification (Table S5b), a fully automated repeat annotation makes our results rather conservative. Indeed, the high unidentified repeat proportion (Figures 3 and S2) may indicate an underestimation of associations between repeats and population differentiation and karyotype changes, but unlikely to false associations.

We found significant phylogenetic signals for some repeats (Table 1), and an overall correspondence between the phylogeny of *Erebia* and *Carex* species and their clustering based on their respective repeat landscape (Figure 4). However, this correspondence is not sufficient for a reliable phylogenetic inference using only repeats, contrarily to former suggestions (Dodsworth et al., 2015; Vitales et al., 2020). This does not seem surprising, as closely related species can differ substantially in their repeat landscapes, which can be attributed to differences in demographic history (e.g. Alioto et al., 2020). For example, satDNA abundance is very variable between species for both *Erebia* and *Carex* (Figure 3). This is expected under the so-called '*library hypothesis*' (Fry & Salser, 1977; Salser et al., 1976), which highlights that different satDNA families are randomly amplified in different lineages, resulting in different satDNA catalogues even between closely related species. Moreover, TEs can shift between host genomes via horizontal gene transfer events (Bartolomé et al., 2009), which can also blur the phylogenetic patterns of repeat landscapes.

Repetitive elements have often been suggested to be associated with karyotypic variation because they can cause chromosomal rearrangements through ectopic recombination (Fedoroff, 2012). Indeed, comparisons between one or a few, often closely related species, indicated that TEs could be involved in fusions of chromosomes (e.g. Ahola et al., 2014; Höök et al., 2023). Here, we took a broader, genus-wide approach to evaluate the association between repeats and karyotype changes in a phylogenetic context. We used chromosome number changes as a proxy for chromosomal fusions and fissions, therefore likely underestimating the number of rearrangements (Mackintosh et al., 2023) and their association with repeats. We found indications that certain types of repeats could be associated with karyotype changes at a macro-evolutionary level, although none of the associations were significant following a FDR correction in both *Erebia* and *Carex* (Table 1), suggesting that future macro-evolutionary investigations should increase taxon sampling. The relationships may nevertheless hint towards potential mechanisms. For instance, Helitrons were negatively associated with chromosome number in *Erebia* (Table 1a) and moreover explained the highest amount of variation in chromosome number (Table S8) with significant optima shifts in the Tyndarus clade (Figure 5b). This is especially interesting given that Tyndarus is the most karyotypically diverse, species-rich, and youngest clade within *Erebia* (Augustijnen et al., 2023), which also hints towards an association of repeats with karyotype changes and associated speciation. Lineage-specific repeat expansion could explain why some *Erebia* subclades are more karyotypically diverse than others (Augustijnen et al., 2023), a pattern also observed between different Lepidoptera genera (de Vos et al., 2020; Robinson, 1971).

Both Maverick repeats and unidentified LTRs showed potential indications for a positive association with chromosome number in *Erebia* (Table 1a), which would differ from another study on wood-white butterflies (*Leptidea* sp.; Höök et al., 2023), where Helitrons were generally depleted from fusion and fission sites, and LINEs and LTRs were enriched in fusion, but not fission, breakpoints. However, the taxonomic resolution of the aforementioned study was limited and unlike *Erebia*, intraspecific karyotypic variation is more common in *Leptidea* (Höök et al., 2023) and the impact of repeats on reproductive isolation may thus differ. In *Erebia*, chromosomal fusions are more common than fissions, but fissions are more often involved in speciation events (Augustijnen et al., 2023). Our findings implicate that fusion and fission events are likely associated with different repetitive elements, where Helitrons might cause frequent fusions but therefore could be less important for speciation than Maverick and LTRs, involved in fissions events in *Erebia*. However, by using chromosome numbers as proxy for rearrangements, we underestimate the impact of repeats that could be involved in both fusions and fissions of chromosomes. Interestingly, an enrichment in Helitrons, other DNA transposons, LINEs and LTRs was observed in rearrangement breakpoints of holocentric aphids (Mathers et al., 2021), but no distinction between fusions and fissions of chromosomes was made. In *Carex*, LINEs, Ty1/Copia and Helitrons were associated with rates of chromosomal evolution (Table S6). We detected associations between unidentified LTRs and Ty3/Gypsies and chromosome number, that were not significant after the FDR (Table 1b). These findings are in accordance with a study based on comparative genomics among

three *Carex* species which found enrichment of similar repeats in rearranged parts of the genome (Escudero et al., 2023), but suggest that the relative impact of specific repeats on karyotypic changes differs among independently evolved holocentric groups.

One of the goals of this study was to determine which repetitive elements could be associated with karyotype changes in *Erebia* and *Carex*. Abundant repeats should be more likely to result in ectopic recombination (Kent et al., 2017). Surprisingly, LINEs, which were by far the most abundant repeats in *Erebia* genomes, showed no association with karyotype changes (Table 1a). Helitrons, on the contrary, were the rarest repeats but showed some indication for an association with karyotype changes (Figure 5). This is possibly because very abundant TEs could be more often silenced through methylation by the host to suppress ectopic recombination (Zamudio et al., 2015), or because retrotransposons such as LINEs might be controlled by piRNAs (Shoji et al., 2023). Differences in repeat abundance between *Erebia* and *Carex* (Figure 3) could partly explain why different repeats were associated with rearrangements in these two genera. Furthermore, *Erebia* and *Carex* likely differ in their holocentromere structure. Indeed, sedges, possibly including *Carex*, have repeat-based holocentromeres (satDNA) that can be involved in chromosomal rearrangements (Hofstatter et al., 2022). Centromeric satDNA can also be involved in chromosomal rearrangements in monocentric organisms (e.g. Antonarakis, 2022). In contrast, the holocentromeres of Lepidoptera are not resolved but do not seem to be repeat-based (Senaratne et al., 2021).

Recently active repeats that show little sequence divergence are the most likely to be involved in ectopic recombination (Li et al., 2006; Renkawitz et al., 2014). Indeed, the karyotypically diverse Tyndarus group in *Erebia* tended to show lower Kimura substitution values suggesting younger and more active repeats than more karyotypically conserved groups (Figure 6 and Data S1). Moreover, repeat types associated with karyotype changes had generally low Kimura substitution levels, both in *Erebia* (e.g. Helitrons, DIRS and rDNA; Data S1) and in *Carex* (e.g. Ty1/Copia and Ty3/Gypsy; Data S2). Differences between TE classes in their importance for ectopic recombination could also be expected given their differences in transposition mechanisms. Indeed, retrotransposons, which do not excise (i.e. replicate via copy-paste mechanism, including LINES and LTRs), might show a stronger negative correlation with recombination rates than DNA transposons which cause double-strand breaks during excision, e.g. Helitrons (Kent et al., 2017). Because ectopic recombination rates are correlated with meiotic recombination rates (Goldman & Lichten, 1996), this suggests that DNA transposons could be more likely involved in chromosomal rearrangements, consistent with our findings for *Erebia*.

Another factor influencing the impact of repeats in rearrangements is their genomic localisation. Repeats physically close to each other, close to telomeres, or close to recombination hotspots might be more likely to undergo ectopic recombination (Goldman & Lichten, 1996; Kent et al., 2017). The repeat-rich sex chromosomes might also be more likely to undergo chromosomal rearrangements than autosomes (Nguyen & Carabajal Paladino, 2016; Wright et al., 2023), but chromosome specific analyses would require chromosome-scale reference genomes for all species. Longer repeats are also more likely to cause ectopic recombination than shorter repeats (Petrov et al., 2003), and repeat orientation could also play a role (Delprat et al., 2009). Because we used a reference-free repeat detection method based on short-read data and in the absence of reference genomes, we cannot investigate these factors here. Also, because we used only a subset of the phylogeny of *Erebia* and *Carex*, evolutionary contingencies might have masked the pattern of repeats involved in rearrangements in the time that passed since the rearrangement evolved. Our results on the association between repeats and chromosomal rearrangements are therefore rather conservative. Further studies could investigate sister species pairs or closely related lineages to reduce the impact of evolutionary contingencies and further precise the role of repetitive elements in chromosomal rearrangements.

5 | CONCLUSION AND FUTURE DIRECTIONS

This study is, to our knowledge, the first to explore repeat diversity in holocentric species at both micro- and macro-evolutionary scales. We highlight intraspecific divergence in repeat landscapes between populations, which scales with levels of population divergence. Consequently, repeats might be associated with population differentiation. Furthermore, we provide evidence for a macro-evolutionary association between repeats and chromosomal rearrangements that lead to karyotype changes in two independently evolved holocentric clades. Our study thus sheds some light on the potential role of repeats on the process of species diversification (Kulmuni et al., 2020) and potentially chromosomal speciation. Future studies comparing reference assemblies between sister species, along with novel pipelines for discovery and annotation of repeats (e.g. Baril et al., 2022), are likely to precise the role of repeats in chromosomal rearrangements, even in non-model organisms, and may help to bridge the gap between micro- and macro-evolutionary patterns. Similarly, understanding the association with the epigenetic landscape could provide crucial insights into why only some clades show high karyotypic variation (Zamudio et al., 2015). A link between repeat-induced chromosomal rearrangements and speciation has long been suggested (Serrato-Capuchina & Matute, 2018), but empirical evidence is still scarce. Here, we highlight variations in repeat landscape at both micro- and macro-evolutionary scales that have the potential to be involved in speciation. Further studies could make use of advances in genome editing technologies to cause chromosomal rearrangements (Ansai & Kitano, 2022) and infer a causal relationship between repeats, chromosomal rearrangements, and potentially speciation.

AUTHOR CONTRIBUTIONS

CC, KL, PN and ME designed the study. HA and KL generated the genomic data. CC, ME and PM analysed the data. CC wrote the manuscript with inputs from all co-authors.

ACKNOWLEDGEMENTS

We are grateful to Daniel Croll for allowing us access to his server to perform data analysis. We thank Anna Voleníková for her precious help in data analysis and interpretation, and comments on the manuscript, and Sandra Le Bissonnais for helpful comments on a previous version of the manuscript. We are grateful to Thomas Schmitt, Reinhold Öhrlein, Thomas Kissling, Jürgen Hensle, Bene Jost for contributing to sample collection. Finally, we thank the associate editor and three anonymous reviewers that provided comments improving significantly previous versions of this study. CC was supported by the Swiss National Science Foundation (SNSF) Eccellenza Project *The evolution of strong reproductive barriers towards the completion of speciation* (PCEFP3_202869) and HA by the SNSF grant (310030_184934) *Genomic rearrangements and the origin of species* both awarded to KL. ME was supported by FEDER/MICINN – AEI (PID2021-122715NB-I00). PN was supported by the 23-06455S grant of the Czech Science Foundation. PM thanks the University of Jaén for its "Convocatoria de Recualificación del Sistema Universitario Español-Margarita Salas" postdoctoral grant under the "Plan de Recuperación Transformación" program funded by the Spanish Ministry of Universities with European Union's NextGenerationEU funds (grant no. UJAR10MS). Computational resources for RepeatMasker analyses were supplied by the project 'e-Infrastruktura CZ' (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures and the ELIXIR-CZ project (LM2018131), part of the international ELIXIR infrastructure. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Open access funding provided by Universite de Neuchatel.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All *Erebia* genomic sequence data (for the species-level analyses) is deposited at NCBI BioProject PRJNA1001152 and *Carex* at NCBI BioProject PRJNA1000636. Final repeat count tables and repeat genomic proportions are archived on Zenodo at: https://doi.org/10.5281/zenodo.8199371. All code is uploaded on GitHub at: https://github.com/camille-cornet/HolocentricRepeatLandscapes.

ORCID

Camille Cornet https://orcid.org/0000-0002-4545-5485
Pablo Mora https://orcid.org/0000-0001-7967-3379
Hannah Augustijnen https://orcid.org/0000-0002-6605-9748
Petr Nguyen https://orcid.org/0000-0003-1395-4287
Marcial Escudero https://orcid.org/0000-0002-2541-5427
Kay Lucek https://orcid.org/0000-0002-2253-2556

REFERENCES

Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., Välimäki, N., Paulin, L., Kvist, J., Wahlberg, N., Tanskanen, J., Hornett, E. A., Ferguson, L. C., Luo, S., Cao, Z., de Jong, M. A., Duplouy, A., Smolander, O.-P., Vogel, H., … Hanski, I. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature. Communications*, *5*(1), 4737. https://doi.org/10.1038/ncomms5737

Alioto, T., Alexiou, K. G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández I Martí, Á., Frias, L., Galán, B., García, J. L., Howad, W., Gómez-Garrido, J., Gut, M., Julca, I., Morata, J., Puigdomènech, P., Ribeca, P., … Arús, P. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: Results from the almond genome sequence. *The Plant Journal*, *101*(2), 455–472. https://doi.org/10.1111/tpj.14538

Ansai, S., & Kitano, J. (2022). Speciation and adaptation research meets genome editing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1855), 20200516. https://doi.org/10.1098/rstb.2020.0516

Antonarakis, S. E. (2022). Short arms of human acrocentric chromosomes and the completion of the human genome sequence. *Genome Research*, *32*, 599–607. https://doi.org/10.1101/gr.275350.121

Augustijnen, H., Baetscher, L., Cesanek, M., Chkhartishvili, T., Dinca, V., Iankoshvili, G., Ogawa, K., Vila, R., Klopfstein, S., de Vos, J., & Lucek, K. (2023). *A macro-evolutionary role for chromosomal fusion and fission in* Erebia *butterflies* (p. 2023.01.16.524200). bioRxiv. https://doi.org/10.1101/2023.01.16.524200

Augustijnen, H., Patsiou, T., & Lucek, K. (2022). Secondary contact rather than coexistence—*Erebia* butterflies in the Alps. *Evolution*, *76*(11), 2669–2686. https://doi.org/10.1111/evo.14615

Baril, T., Imrie, R. M., & Hayward, A. (2022). Earl Grey: A fully automated user-friendly transposable element annotation and analysis pipeline. *bioRxiv*. https://doi.org/10.1101/2022.06.30.498289

Bartolomé, C., Bello, X., & Maside, X. (2009). Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biology*, *10*(2), R22. https://doi.org/10.1186/gb-2009-10-2-r22

Benestan, L. M., Rougemont, Q., Senay, C., Normandeau, E., Parent, E., Rideout, R., Bernatchez, L., Lambert, Y., Audet, C., & Parent, G. J. (2021). Population genomics and history of speciation reveal fishery management gaps in two related redfish species (*Sebastes mentella* and *Sebastes fasciatus*). *Evolutionary Applications*, *14*(2), 588–606. https://doi.org/10.1111/eva.13143

Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, *443*(7111), 521–524. https://doi.org/10.1038/443521a

Bourgeois, Y., & Boissinot, S. (2019). On the population dynamics of junk: A review on the population genomics of transposable elements. *Genes*, *10*(6), 419. https://doi.org/10.3390/genes10060419

Cabral-de-Mello, D. C., Zrzavá, M., Kubíčková, S., Rendón, P., & Marec, F. (2021). The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Frontiers in Genetics*, *12*, 661417. https://doi.org/10.3389/fgene.2021.661417

Cáceres, M., Ranz, J. M., Barbadilla, A., Long, M., & Ruiz, A. (1999). Generation of a widespread *Drosophila* inversion by a transposable element. *Science (New York, N.Y.)*, *285*(5426), 415–418. https://doi.org/10.1126/science.285.5426.415

Carta, A., & Escudero, M. (2023). Karyotypic diversity: A neglected trait to explain angiosperm diversification? *Evolution*, *77*(4), 1158–1164. https://doi.org/10.1093/evolut/qpad014

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Charlesworth, B., & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetics Research*, *42*(1), 1–27. https://doi.org/10.1017/S0016672300021455

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Costa, L., Marques, A., Buddenhagen, C., Thomas, W. W., Huettel, B., Schubert, V., Dodsworth, S., Houben, A., Souza, G., & Pedrosa-Harand, A. (2021). Aiming off the target: Recycling target capture sequencing reads for investigating repetitive DNA. *Annals of Botany*, *128*(7), 835–848. https://doi.org/10.1093/aob/mcab063

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Oxford University Press.

Cupedo, F. (2014). Reproductive isolation and intraspecific structure in Alpine populations of *Erebia euryale* (Esper, 1805) (Lepidoptera, Nymphalidae, Satyrinae). *Nota Lepidopterologica*, *37*(1), 19. https://doi.org/10.3897/nl.37.7960

de Vos, J. M., Augustijnen, H., Bätscher, L., & Lucek, K. (2020). Speciation through chromosomal fusion and fission in Lepidoptera. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806), 20190539. https://doi.org/10.1098/rstb.2019.0539

Delprat, A., Negre, B., Puig, M., & Ruiz, A. (2009). The transposon Galileo generates natural chromosomal inversions in *drosophila* by ectopic recombination. *PLoS One*, *4*(11), e7883. https://doi.org/10.1371/journal.pone.0007883

Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novák, P., Piednoël, M., Weiss-Schneeweiss, H., & Leitch, A. R. (2015). Genomic repeat abundances contain phylogenetic signal. *Systematic Biology*, *64*(1), 112–126. https://doi.org/10.1093/sysbio/syu080

Escudero, M., Hahn, M., Brown, B. H., Lueders, K., & Hipp, A. L. (2016). Chromosomal rearrangements in holocentric organisms lead to reproductive isolation by hybrid dysfunction: The correlation between karyotype rearrangements and germination rates in sedges. *American Journal of Botany*, *103*(8), 1529–1536. https://doi.org/10.3732/ajb.1600051

Escudero, M., Hipp, A. L., Waterway, M. J., & Valente, L. M. (2012). Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (*Carex*, Cyperaceae). *Molecular Phylogenetics and Evolution*, *63*(3), 650–655. https://doi.org/10.1016/j.ympev.2012.02.005

Escudero, M., Marques, A., Lucek, K., & Hipp, A. (2023). *Highly conserved synteny despite massive chromosome fusion and fission suggest fragile sites in holocentric plants*. *Molecular Ecology* in press.

Escudero, M., Márquez-Corro, J. I., & Hipp, A. L. (2016). The phylogenetic origins and evolutionary history of holocentric chromosomes. *Systematic Botany*, *41*(3), 580–585. https://doi.org/10.1600/036364416X692442

Faria, R., & Navarro, A. (2010). Chromosomal speciation revisited: Rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, *25*(11), 660–669. https://doi.org/10.1016/j.tree.2010.07.008

Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science*, *338*(6108), 758–767. https://doi.org/10.1126/science.338.6108.758

Feiner, N. (2016). Accumulation of transposable elements in Hox gene clusters during adaptive radiation of *Anolis* lizards. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1840), 20161555. https://doi.org/10.1098/rspb.2016.1555

Feliciello, I., Akrap, I., Brajković, J., Zlatar, I., & Ugarković, Đ. (2015). Satellite DNA as a driver of population divergence in the red flour beetle *Tribolium castaneum*. *Genome Biology and Evolution*, *7*(1), 228–239. https://doi.org/10.1093/gbe/evu280

Fry, K., & Salser, W. (1977). Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. *Cell*, *12*(4), 1069–1084. https://doi.org/10.1016/0092-8674(77)90170-2

Galili, T. (2015). dendextend: An R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics (Oxford, England)*, *31*, 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Global Carex Group, Roalson, E. H., Jiménez-Mejías, P., Hipp, A. L., Benítez-Benítez, C., Bruederle, L. P., Chung, K.-S., Escudero, M., Ford, B. A., Ford, K., Gebauer, S., Gehrke, B., Hahn, M., Hayat, M. Q., Hoffmann, M. H., Jin, X.-F., Kim, S., Larridon, I., Léveillé-Bourret, É., … Zhang, S.-R. (2021). A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles. *Journal of Systematics and Evolution*, *59*(4), 726–762. https://doi.org/10.1111/jse.12722

Goldman, A. S. H., & Lichten, M. (1996). The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics*, *144*(1), 43–55.

Goubert, C., Craig, R. J., Bilat, A. F., Peona, V., Vogan, A. A., & Protasio, A. V. (2022). A beginner's guide to manual curation of transposable elements. *Mobile DNA*, *13*(1), 7. https://doi.org/10.1186/s13100-021-00259-7

Gratton, P., Trucchi, E., Trasatti, A., Riccarducci, G., Marta, S., Allegrucci, G., Cesaroni, D., & Sbordoni, V. (2016). Testing classical species properties with contemporary data: How "bad species" in the brassy ringlets (*Erebia tyndarus* complex, Lepidoptera) turned good. *Systematic Biology*, *65*(2), 292–303. https://doi.org/10.1093/sysbio/syv087

Haley, A. L., & Mueller, R. L. (2022). Transposable element diversity remains high in gigantic genomes. *Journal of Molecular Evolution*, *90*(5), 332–341. https://doi.org/10.1007/s00239-022-10063-3

Hansen, T. F., Pienaar, J., & Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, *62*(8), 1965–1977. https://doi.org/10.1111/j.1558-5646.2008.00412.x

Haubold, B., Pfaffelhuber, P., & Lynch, M. (2010). MlRho – A program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Molecular Ecology*, *19*(s1), 277–284. https://doi.org/10.1111/j.1365-294X.2009.04482.x

Hipp, A. L., Rothrock, P. E., & Roalson, E. H. (2009). The evolution of chromosome arrangements in *Carex* (Cyperaceae). *The Botanical Review*, *75*(1), 96–109. https://doi.org/10.1007/s12229-008-9022-8

Hofstatter, P. G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., Zhang, M., Costa, L., Castellani, M., Scott, A., Toegelová, H., Fuchs, J., Mata-Sucre, Y., Dias, Y., Vanzela, A. L. L., Huettel, B., Almeida, C. C. S., Šimková, H., Souza, G., … Marques, A. (2022). Repeat-based holocentromeres influence genome architecture and karyotype evolution. *Cell*, *185*(17), 3153–3168.e18. https://doi.org/10.1016/j.cell.2022.06.045

Höök, L., Näsvall, K., Vila, R., Wiklund, C., & Backström, N. (2023). High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (*Leptidea* spp.). *Chromosome Research*, *31*(1), 2. https://doi.org/10.1007/s10577-023-09713-z

Huang, C., Sun, H., Xu, D., Chen, Q., Liang, Y., Wang, X., Xu, G., Tian, J., Wang, C., Li, D., Wu, L., Yang, X., Jin, W., Doebley, J. F., & Tian, F. (2018). ZmCCT9 enhances maize adaptation to higher latitudes. *Proceedings of the National Academy of Sciences*, *115*(2), E334–E341. https://doi.org/10.1073/pnas.1718058115

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). phylosignal: An R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, *6*(9), 2774–2780. https://doi.org/10.1002/ece3.2051

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160458. https://doi.org/10.1098/rstb.2016.0458

Kopperud, B. T., Pienaar, J., Voje, K. L., Orzack, S. H., Hansen, T. F., & Grabowski, M. (2020). *slouch: Stochastic linear Ornstein-Uhlenbeck comparative hypotheses* (2.1.4). https://CRAN.R-project.org/package=slouch

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., & Westram, A. M. (2020). Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806), 20190528. https://doi.org/10.1098/rstb.2019.0528

Larridon, I., Galán Díaz, J., Bauters, K., & Escudero, M. (2021). What drives diversification in a pantropical plant lineage with extraordinary capacity for long-distance dispersal and colonization? *Journal of Biogeography*, *48*(1), 64–77. https://doi.org/10.1111/jbi.13982

Li, L., Jean, M., & Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. *The Plant Journal*, *45*(6), 908–916. https://doi.org/10.1111/j.1365-313X.2006.02657.x

Li, Y., Wang, S., Zhang, Z., Luo, J., Lin, G. L., Deng, W.-D., Guo, Z., Han, F. M., Wang, L.-L., Li, J., Wu, S.-F., Liu, H.-Q., He, S., Murphy, R. W., Zhang, Z.-J., Cooper, D. N., Wu, D.-D., & Zhang, Y.-P. (2023). Large-scale chromosomal changes Lead to genome-level expression alterations, environmental adaptation, and speciation in the Gayal (*Bos frontalis*). *Molecular Biology and Evolution*, *40*(1), msad006. https://doi.org/10.1093/molbev/msad006

Lohse, K., Hayward, A., Laetsch, D. R., Vila, R., Lucek, K., & Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, & Darwin Tree of Life Consortium. (2022). The genome sequence of the arran brown, *Erebia ligea* (Linnaeus, 1758). *Wellcome Open Research*, *7*, 259. https://doi.org/10.12688/wellcomeopenres.18115.1

Lohse, O., Lohse, K., Augustijnen, H., & Lucek, K. (2022). The genome sequence of the scotch argus butterfly, *Erebia aethiops* (Esper, 1777). *Wellcome Open Research*, *7*, 217. https://doi.org/10.12688/wellcomeopenres.17927.1

Lönnig, W.-E., & Saedler, H. (2002). Chromosome rearrangements and transposable elements. *Annual Review of Genetics*, *36*(1), 389–410. https://doi.org/10.1146/annurev.genet.36.040202.092802

Lucek, K., Augustijnen, H., & Escudero, M. (2022). A holocentric twist to chromosomal speciation? *Trends in Ecology & Evolution*, *37*(8), 655–662. https://doi.org/10.1016/j.tree.2022.04.002

Lucek, K., Butlin, R. K., & Patsiou, T. (2020). Secondary contact zones of closely-related *Erebia* butterflies overlap with narrow phenotypic and parasitic clines. *Journal of Evolutionary Biology*, *33*(9), 1152–1163. https://doi.org/10.1111/jeb.13669

Lukhtanov, V. A., Dincă, V., Friberg, M., Šíchová, J., Olofsson, M., Vila, R., Marec, F., & Wiklund, C. (2018). Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *Proceedings of the National Academy of Sciences*, *115*(41), E9610–E9619. https://doi.org/10.1073/pnas.1802610115

Macas, J., Novák, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P., Fuková, I., Doležel, J., Kelly, L. J., & Leitch, I. J. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PLoS One*, *10*(11), e0143424. https://doi.org/10.1371/journal.pone.0143424

Mackintosh, A., Vila, R., Laetsch, D. R., Hayward, A., Martin, S. H., & Lohse, K. (2023). Chromosome fissions and fusions act as barriers to gene flow between *Brenthis* fritillary butterflies. *Molecular Biology and Evolution*, *40*(3), msad043. https://doi.org/10.1093/molbev/msad043

Mandrioli, M., & Manicardi, G. C. (2020). Holocentric chromosomes. *PLOS Genetics*, *16*(7), e1008918. https://doi.org/10.1371/journal.pgen.1008918

Márquez-Corro, J. I., Martín-Bravo, S., Jiménez-Mejías, P., Hipp, A. L., Spalink, D., Naczi, R. F. C., Roalson, E. H., Luceño, M., & Escudero, M. (2021). Macro-evolutionary insights into sedges (*Carex*: Cyperaceae): The effects of rapid chromosome number evolution on lineage diversification. *Journal of Systematics and Evolution*, *59*(4), 776–790. https://doi.org/10.1111/jse.12730

Martín-Bravo, S., Jiménez-Mejías, P., Villaverde, T., Escudero, M., Hahn, M., Spalink, D., Roalson, E. H., Hipp, A. L., Group, the G. C, Benítez-Benítez, C. P., Bruederle, L., Fitzek, E. A., Ford, B. A., Ford, K., Garner, M., Gebauer, S. H., Hoffmann, M., Jin, X.-F., Larridon, I., … Starr, J. R. (2019). A tale of worldwide success: Behind the scenes of *Carex* (Cyperaceae) biogeography and diversification. *Journal of Systematics and Evolution*, *57*(6), 695–718. https://doi.org/10.1111/jse.12549

Mathers, T. C., Wouters, R. H. M., Mugford, S. T., Swarbreck, D., van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, *38*(3), 856–875. https://doi.org/10.1093/molbev/msaa246

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. L. (2012). Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research*, *20*(5), 579–593. https://doi.org/10.1007/s10577-012-9292-1

Molina-Venegas, R., & Rodríguez, M. Á. (2017). Revisiting phylogenetic signal; strong or negligible impacts of polytomies and branch length information? *BMC Evolutionary Biology*, *17*, 53. https://doi.org/10.1186/s12862-017-0898-y

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nguyen, P., & Carabajal Paladino, L. (2016). On the neo-sex chromosomes of Lepidoptera. In P. Pontarotti (Ed.), *Evolutionary biology: Convergent evolution, evolution of complex traits, concepts and methods* (pp. 171–185). Springer International Publishing. https://doi.org/10.1007/978-3-319-41324-2_11

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Research*, *45*(12), e111. https://doi.org/10.1093/nar/gkx257

Novák, P., Neumann, P., & Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, *11*(1), 378. https://doi.org/10.1186/1471-2105-11-378

Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745–3776. https://doi.org/10.1038/s41596-020-0400-y

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, L., Karisto, P., Vonlanthen, T., Mundt, C., McDonald, B. A., & Croll, D. (2021). A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *eLife*, *10*, e69249. https://doi.org/10.7554/eLife.69249

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2022). *Vegan: Community ecology package* (2.6-4). https://CRAN.R-project.org/package=vegan

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*

(Oxford, England), 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pellicer, J., & Leitch, I. J. (2020). The plant DNA C-values database (release 7.1): An updated online repository of plant genome size data for comparative studies. *New Phytologist*, *226*(2), 301–305. https://doi.org/10.1111/nph.16261

Peña, C., Witthauer, H., Klečková, I., Fric, Z., & Wahlberg, N. (2015). Adaptive radiations in butterflies: Evolutionary history of the genus *Erebia* (Nymphalidae: Satyrinae). *Biological Journal of the Linnean Society*, *116*(2), 449–467. https://doi.org/10.1111/bij.12597

Petrov, D. A., Aminetzach, Y. T., Davis, J. C., Bensasson, D., & Hirsh, A. E. (2003). Size matters: Non-LTR Retrotransposable elements and ectopic recombination in *drosophila*. *Molecular Biology and Evolution*, *20*(6), 880–892. https://doi.org/10.1093/molbev/msg102

Pinheiro, J., Bates, D., & R Core Team. (2023). *Nlme: Linear and nonlinear mixed effects models*. R package version 3.1-162. https://CRAN.R-project.org/package=nlme

Posit team. (2023). *RStudio: Integrated development environment for R* [Manual]. http://www.posit.co/

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebollo, R., Romanish, M. T., & Mager, D. L. (2012). Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics*, *46*(1), 21–42. https://doi.org/10.1146/annurev-genet-110711-155621

Renkawitz, J., Lademann, C. A., & Jentsch, S. (2014). Mechanisms and principles of homology search during recombination. *Nature Reviews Molecular Cell Biology*, *15*(6), 369–383. https://doi.org/10.1038/nrm3805

Ricci, M., Peona, V., Guichard, E., Taccioli, C., & Boattini, A. (2018). Transposable elements activity is positively related to rate of speciation in mammals. *Journal of Molecular Evolution*, *86*(5), 303–310. https://doi.org/10.1007/s00239-018-9847-7

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*(7), 351–358. https://doi.org/10.1016/S0169-5347(01)02187-5

Robinson, R. (1971). *Lepidoptera Genetics*. Pergamon Press.

Salser, W., Bowen, S., Browne, D., El-Adli, F., Fedoroff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B., & Whitcome, P. (1976). Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Federation Proceedings*, *35*(1), 23–35.

Schmitt, T., Louy, D., Zimmermann, E., & Habel, J. C. (2016). Species radiation in the Alps: Multiple range shifts caused diversification in ringlet butterflies in the European high mountains. *Organisms Diversity and Evolution*, *16*(4), 791–808. https://doi.org/10.1007/s13127-016-0282-6

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, *5*(1), 6495. https://doi.org/10.1038/ncomms6495

Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, *28*(6), 1537–1549. https://doi.org/10.1111/mec.14794

Senaratne, A. P., Muller, H., Fryer, K. A., Kawamoto, M., Katsuma, S., & Drinnenberg, I. A. (2021). Formation of the CenH3-deficient Holocentromere in Lepidoptera avoids active chromatin. *Current Biology*, *31*(1), 173–181.e7. https://doi.org/10.1016/j.cub.2020.09.078

Serrato-Capuchina, A., & Matute, D. R. (2018). The role of transposable elements in speciation. *Genes*, *9*(5), 254. https://doi.org/10.3390/genes9050254

Shoji, K., Umemura, Y., Katsuma, S., & Tomari, Y. (2023). The piRNA cluster torimochi is an expanding transposon in cultured silkworm cells. *PLoS Genetics*, *19*(2), e1010632. https://doi.org/10.1371/journal.pgen.1010632

Silva, B. S. M. L., Heringer, P., Dias, G. B., Svartman, M., & Kuhn, G. C. S. (2019). De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS One*, *14*(12), e0223466. https://doi.org/10.1371/journal.pone.0223466

Slotkin, R. K. (2018). The case for not masking away repetitive DNA. *Mobile DNA*, *9*(1), 15. https://doi.org/10.1186/s13100-018-0120-9

Smit, A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. http://www.repeatmasker.org

Sonderegger, P. (2005). Die Erebien der Schweiz: (Lepidoptera: Satyrinae, Genus Erebia). P. Sonderegger.

Spirito, F. (1998). The role of chromosomal change in speciation. In *Endless forms: Species and speciation*. Oxford University Press.

Sproul, J. S., Hotaling, S., Heckenhauer, J., Powell, A., Larracuente, A. M., Kelley, J. L., Pauls, S. U., & Frandsen, P. B. (2022). *Repetitive elements in the era of biodiversity genomics: Insights from 600+ insect genomes* (p. 2022.06.02.494618). *bioRxiv*. https://doi.org/10.1101/2022.06.02.494618

Stritt, C., Gordon, S. P., Wicker, T., Vogel, J. P., & Roulin, A. C. (2018). Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of the Mediterranean grass *Brachypodium distachyon*. *Genome Biology and Evolution*, *10*(1), 304–318. https://doi.org/10.1093/gbe/evx276

Talavera, G., Lukhtanov, V. A., Rieppel, L., Pierce, N. E., & Vila, R. (2013). In the shadow of phylogenetic uncertainty: The recent diversification of *Lysandra* butterflies through chromosomal change. *Molecular Phylogenetics and Evolution*, *69*(3), 469–478. https://doi.org/10.1016/j.ympev.2013.08.004

Uyeda, J. C., & Harmon, L. J. (2014). A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology*, *63*(6), 902–918. https://doi.org/10.1093/sysbio/syu057

Van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., & Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, *534*(7605), 102–105. https://doi.org/10.1038/nature17951

Venner, S., Feschotte, C., & Biémont, C. (2009). Dynamics of transposable elements: Towards a community ecology of the genome. *Trends in Genetics*, *25*(7), 317–323. https://doi.org/10.1016/j.tig.2009.05.003

Vitales, D., Garcia, S., & Dodsworth, S. (2020). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Molecular Phylogenetics and Evolution*, *147*, 106766. https://doi.org/10.1016/j.ympev.2020.106766

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, *54*, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

White, M. J. D. (1978). *Modes of speciation* (First ed.). W. H. Freeman.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Wright, C. J., Stevens, L., Mackintosh, A., Lawniczak, M., & Blaxter, M. (2023). *Chromosome evolution in Lepidoptera* (p. 2023.05.12.540473). bioRxiv. https://doi.org/10.1101/2023.05.12.540473

Yoshida, K., Rödelsperger, C., Röseler, W., Riebesell, M., Sun, S., Kikuchi, T., & Sommer, R. J. (2023). Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nature Ecology & Evolution*, *7*(3), 424–439. https://doi.org/10.1038/s41559-022-01980-z

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., & Bourc'his, D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic

recombination. *Genes & Development*, *29*(12), 1256–1270. https://doi.org/10.1101/gad.257840.114

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153. https://doi.org/10.1186/s12859-018-2129-y

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cornet, C., Mora, P., Augustijnen, H., Nguyen, P., Escudero, M., & Lucek, K. (2023). Holocentric repeat landscapes: From micro-evolutionary patterns to macro-evolutionary associations with karyotype evolution. *Molecular Ecology*, *00*, 1–19. https://doi.org/10.1111/mec.17100