DR. RIMJHIM ROY CHOUDHURY (Orcid ID : 0000-0002-0499-4124)

MISS AUDE ROGIVUE (Orcid ID : 0000-0002-6864-0587)

PROF. CHRISTIAN PARISOD (Orcid ID : 0000-0001-8798-0897)

Impact of polymorphic transposable elements on linkage disequilibrium along chromosomes

Rimjhim Roy Choudhury[1], Aude Rogivue[2], Felix Gugerli[2], Christian Parisod[1,*]

[1] Institute of Plant Sciences, University of Berne, Altenbergrain 21, CH–3013 Bern, Switzerland

[2] WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH–8903 Birmensdorf, Switzerland

*Correspondence: christian.parisod@ips.unibe.ch

Abstract

Recombination and selection drive the extent of linkage disequilibrium (LD) among loci and therefore affect the reshuffling of adaptive genetic variation. However, it is poorly known to what extent the enrichment of transposable elements (TEs) in recombinationally-inert regions reflects their inefficient removal by purifying selection and whether the presence of polymorphic TEs can modify the local recombination rate. In this study, we investigate how TEs and recombination interact at fine scale along chromosomes and possibly support linked selection in natural populations. Whole-genome sequencing data of 304 individuals from nearby alpine populations of

Arabis alpina were used to show that the density of polymorphic TEs is specifically correlated with local LD along chromosomes. Consistent with TEs modifying recombination, the characterization of 28 such LD blocks of up to 5.5Mb in length revealed strong evidence of selective sweeps at a few loci through either site frequency spectrum or haplotype structure. A majority of these blocks were enriched in genes related to ecologically relevant functions such as responses to cold, salt stress or photoperiodism. In particular, the S-locus (i.e. supergene responsible for strict outcrossing) was identified in a LD block with high levels of polymorphic TEs and evidence of selection. Another such LD block was enriched in cold-responding genes and presented evidence of adaptive loci related to photoperiodism and flowering being increasingly linked by polymorphic TEs. These results are consistent with the hypothesis that TEs modify recombination landscapes and thus interact with selection in driving blocks of linked adaptive loci in natural populations.

1. INTRODUCTION

Recurrent environmental changes or shifts in the genetic background promote selection of independent alleles segregating in natural populations (Carja, Liberman, & Feldman, 2014). However, the maintenance of recombining genetic variability may be costly and the evolution of linked loci may be adaptive for traits under consistent selective pressure (Grant & Flake, 1974; Schwander, Libbrecht, & Keller, 2014). Numerous studies have accordingly reported clustered alleles contributing to local adaptation forming "genomic islands of divergence" (Lotterhos, Yeaman, Degner, Aitken, & Hodgins, 2018; Nosil, Funk, & Ortiz-Barrientos, 2009; Via & West, 2008). Factors influencing their origin and persistence remain elusive, but genomic rearrangements promoting tight linkage and increasing the stability of combinations of locally co-adapted alleles have been postulated to play a role (Yeaman & Whitlock, 2011). Such a process may allow several alleles of small effect, which may be hardly determined against the genome-wide background (Pritchard, Pickrell, & Coop, 2010), to effectively function and be selected as a single unit of larger effect (Yeaman, 2013).

New combinations of alleles that are produced through recombination and passed on to offspring possibly support adaptive walks, but can also break apart favourable combinations of alleles with potential reduction in fitness (Orr, 2005). However, molecular mechanisms regulating

recombination rates along chromosomes remain elusive (Dapper & Payseur, 2017). Although the distribution of recombination events is noticeably non-random, commonly restricted to specific regions referred to as hotspots (Mézard, Tagliaro Jahns, & Grelon, 2015), recombination is frequently variable across the genome and among populations (Stapley, Feulner, Johnston, Santure, & Smadja, 2017). Above-average rates of recombination have often been reported in genome regions of high gene density and immediately upstream of genes (Hellsten et al., 2013; Shilo, Melamed-Bessudo, Dorone, Barkai, & Levy, 2015), but variation in recombination rate is observed at all genomic scales: between chromosomes, among regions spanning megabases along chromosomes, and across loci spanning only a few kilobases (Haenel, Laurentino, Roesti, & Berner, 2018).

In most eukaryotes studied to date, meiotic recombination rates appear in strong negative association with transposable elements (TEs), with heterochromatic regions such as pericentromeres showing high TE densities and strongly reduced recombination (Rizzon, Marais, Gouy, & Biémont, 2002; Tian et al., 2009; Willing et al., 2015). Such a pattern seemingly holds at a finer scale with local increase in TEs collocating with reduced rates of recombination (Pan et al., 2011; Rodgers-Melnick et al., 2015). Negative correlation between recombination rates and specific TEs may reflect their biased insertion towards heterochromatic regions (e.g. Gao, Hou, Ebina, Levin, & Voytas, 2008), but available studies rather showed contrasted patterns with preferential insertion in high-recombining, open chromatin regions (e.g. Marand et al., 2017) or lack of specific preferences (e.g. Mirouze et al., 2009; Tsukahara et al., 2009). Accordingly, purifying selection against TEs inserted in or near genes has been argued as a more general and prevalent driver of TE accumulation in regions of reduced recombination (e.g. Wright, Agrawal, & Bureau, 2003). Available data from *Drosophila* and *Arabidopsis* congruently supported greater purifying selection on silenced TEs inserted near genes (Hollister & Gaut, 2009; Lee, 2015). However, this hypothesis remains blind to the actual drivers of recombination rates and to the growing evidence of polymorphism for local recombination hotspots within and between populations (Baudat et al., 2010; Comeron, Ratnappan, & Bailin, 2012). Suppression of recombination events by TEs may originate from their epigenetic silencing leading to repressive chromatin conformations (Zamudio et al., 2015). An alternative hypothesis is that chromosomal restructuring induced by highly dynamic TEs interact with recombination (Bonchev & Parisod, 2013), as suggested by studies in maize showing divergent blocks of TEs around the bz1 locus (He & Dooner, 2009). Accordingly, polymorphic TEs of typically several kilobases would locally reduce recombination even in gene-rich, euchromatic regions of the genome. A better understanding of the role of TEs as modifiers of the local and large-scale recombination landscape is necessary to understand their coevolution with genomes (Kent, Uzunović, & Wright, 2017).

Variation in recombination along chromosomes can experience indirect selection when altering haplotype frequencies and promote the sorting of linked combinations of alleles by selection (Yeaman, 2013). The impact of indirect selection on recombination depends on the degree and form of non-random associations between alleles at different loci, or linkage disequilibrium (LD). As far as standing genetic variation is concerned, patterns of LD among nearly-neutral sites are determined by recombination, with adjacent loci expectedly showing stronger LD than physically unlinked loci. Selection can also impact LD between nearby loci showing strong epistatic interactions (Schwander, Libbrecht, & Keller, 2014) or when selection on a focal site spreads its action along the chromosome ('hitchhiking') and drives adjacent nearly-neutral sites towards high frequencies (selective sweep; Cutter & Payseur, 2013; Ellegren & Galtier, 2016). When considering mutations, purifying selection that removes such typically deleterious alleles also reduces local variation in the vicinity of selective targets (background selection; Charlesworth, Morgan, & Charlesworth, 1993). Linked selection expectedly results in transient linkage of neutral by-products, whereas selection for tight linkage of co-adapted loci is expected to persist longer, but a deeper understanding is necessary of the processes driving the evolution of reduced recombination and how it promotes adaptation and speciation in the face of gene flow (Levy, 2013).

In this study, we address how polymorphic TEs may interact with the recombination landscape in populations of *Arabis alpina*, a model plant species for ecological genomics (Woetzel et al. in prep.) Although mixed-mating and predominant outcrossing has been reported within the European range of the species (Laenen et al., 2018), self-fertilization appears to be prevalent in the area under scrutiny (Buehler, Graf, Holderegger, & Gugerli, 2012). Surveyed populations in this study belong to a single recolonizing lineage (Rogivue, Graf, Parisod, Holderegger, & Gugerli, 2018) and, as also apparent from low genome-wide differentiation (F_{ST} = 0.13), largely share standing genetic variation (Rogivue, Choudhury, Zoller, Joost, Felber, Kasser, Parisod, & Gugerli, accepted). As indirect selection acting on recombination modifiers is expectedly stronger under such conditions (Roze & Lenormand, 2005), we here assessed the fine-scale association of LD and the density of polymorphic long terminal retrotransposons (hereafter referred as TEs) along chromosomes to (i) highlight the role of polymorphic TEs in shaping recombination and (ii) address how such interactions may support the rise of LD blocks presenting adaptive clusters of genes.

2. METHODS

2.1. Genome-wide variation

Paired-end Illumina sequencing generated whole-genome data for 304 individuals of *Arabis alpina* from four geographically proximate regions in the western Swiss Alps: 69 in La Para, 69 in Pierredar, 70 in Les Essets, and 96 in Les Martinets, sampled between 2016m and 2457m a.s.l. (Rogivue et al., accepted). Belonging to a single recolonization lineage (Rogivue et al., 2018), these populations showed limited substructure based on SNPs (F_{ST} = 0.13; Rogivue et al., accepted). While individuals were sampled 1m apart to avoid clonal replicates as much as possible, the presence of unique multilocus genotypes shared by samples was tested using Poppr (Kamyar et al. 2014) in R v3.34 (R Development Core Team, 2018). Up to 17% of the individuals presented similar multilocus genotyes, but were usually geographically distant and also showed evidence of a high inbreeding coefficient *F*, as expected from the sampling of independent genets in highly inbred populations (Text S1).

As fully described in Rogivue et al. (accepted), raw reads were filtered after quality control and mapped to the reference genome v5.1 (Jiao et al., 2017) with BWA 0.7.12 (Li & Durbin, 2010). Single nucleotide polymorphisms (SNP) were called using FreeBayes v1.0.2 (Garrison & Marth, 2012) to evaluate background genetic variation. Biallelic SNPs with a minimum allele frequency (MAF) of 0.025, a maximum of 10% missing data and that were non-overlapping with annotated TEs were retained as the dataset of 291,396 SNPs. Reference TEs were annotated (Choudhury, Neuhaus, & Parisod, 2017) and polymorphic TEs (i.e. presence/absence of TE insertions at specific loci) were identified using TEPID v0.6 (Stuart et al., 2016) as described in Rogivue et al. (accepted). Loci with <0.025 MAF and > 10% missing data were removed, and zygosity was determined from coverage at the individual level for the remaining set of 20,548 polymorphic TEs. SNPs and polymorphic TEs belonging to the eight assembled chromosomes of *Arabis* were used here.

Various population genetic parameters were estimated in non-overlapping windows of 50kb along each of the eight chromosomes of *A. alpina*. Densities of reference TEs, polymorphic TEs and SNPs were accordingly estimated, and outlier regions presenting a density of polymorphic TEs above the 97.5th percentile of the whole data were identified.

Linkage disequilibrium (LD) was reported to decay by half within 22 to 45kb in these populations (Rogivue et al., accepted). Here, LD was estimated as the averaged correlation coefficient (r^2) between adjacent SNPs less than 1000bp apart within a window of 50kb using Plink (Purcell et al., 2007). Linked variation captured by LD within windows of 50kb is thus expected to

reflect effective recombination rates to a large extent. Accordingly, locally inflated LD can be assumed as indicative of reduced recombination or footprints of a selective sweep.

Nucleotide diversity (π) was computed as the number of nucleotide differences between all possible pairs of sequences within windows of 50kb in the overall population (Nei, 1987). Similarly, divergence among sampling populations (D_{XY}) was estimated as the number of pairwise nucleotide difference between sequences from two populations for all combinations of the four populations (Hudson, Slatkin, & Maddison, 1992; Wakeley, 1996). The 97.5th percentile of all D_{XY} values was used to identify outlier loci across all pairwise population comparisons. These proxies were calculated using the R package PopGenome v2.2.4 (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) with implementation of the 'diversity.stats()' and 'diversity.stats.between()' functions.

2.2. Multiple regression analyses

We used multiple linear regressions to model the influence of polymorphic TE densities (pte) on LD along chromosomes, while taking the nucleotide diversity (π) and the background density of reference TEs (te) into account. As SNPs were largely filtered out from genomic regions containing high densities of reference TEs, proxies of genomic variation were all included as explanatory variables in linear models with all interactions (Full Model: LD ~ pte + te + π + pte * te + pte * π + te * π + pte * te * π). Density estimates were logit transformed (log(y/[1 -y])) to fulfil normality. Exhaustive search accounting for over-fitting was performed using the LEAPS package v.3 (Lumley, 2017), testing all possible combinations of descriptor variables with respect to LD and selecting the best model based on minimized Bayesian Information Criterion (BIC).

Variation in LD was investigated using such models (i) at the chromosome level (i.e. all windows of 50kb of each chromosome) and (ii) along chromosomes, using moving windows of 1.5Mb (i.e. 30 windows of 50kb each) with a step size of 50kb. Models were evaluated when at least 15 windows with all data (i.e. pte, te and π) were available. Independent chromosome blocks showing correlation of LD and polymorphic TEs were identified by merging all overlapping and adjacent windows of 1.5Mb with significant models. Significant *p*-values were adjusted for false discovery rate in multiple testing using the Benjamini-Hochberg procedure within each block. All statistics were performed in R v3.34.

2.3. Demographic inference

The sequential Markov coalescent implemented in SMC++ (Terhorst et al., 2017) inferred ancestral population sizes of *A. alpina* in the studied area based on unphased SNPs and LD information across the genome. It was run on each population separately as well as on the overall region after masking regions containing TEs and genes in the *A. alpina* reference assembly. Remaining intergenic regions in TE-poor regions should indeed be less affected by linked selection and thus adequate for demographic inferences. A polarization error was set to 0.5 as the identity of the ancestral alleles is unknown. Finally, a mutation rate of 7e−9 per site per generation (estimate for *Arabidopsis thaliana*, Ossowski et al., 2010) was assumed to translate coalescence times into generations using a generation time of 1.5 years (Laenen et al., 2018).

2.4. Detection of selective sweeps

Selection on linked polymorphisms across the regional population was investigated through analyses of both the site frequency spectrum (SFS) and haplotype structure in this dataset. An SFS-based method may have low power when each particular gene shows limited difference with the overall F_{ST}. This is expected in case of selfing species for which the genome-wide efficacy of selection is low, hindering the detection of outliers. In contrast, haplotype-based methods for detecting recent adaptation look for the spread of a beneficial variant that reduces sequence diversity at linked neutral sites. Selfing reduces fixation times of beneficial alleles, as homozygote mutations arise earlier, exposing the mutant to selection (Charlesworth, 1992); reduced effective recombination under selfing means that recombinant haplotypes are less common, making it easier to detect sweeps based on haplotypes. Thus, both SFS and haplotype-based methods were used here to detect signatures of selective sweeps.

Tajima's *D* was here estimated within windows of 50kb using the R package PopGenome v2.2.4. Evaluating deviation of allele frequency at segregating sites from an expected population evolving neutrally, negative values of Tajima's *D* indicate excess of rare alleles (selective sweep or population expansion) and positive values of *D* indicate excess of frequent alleles (balancing selection or population contraction) (Tajima, 1989). Demographic effects on Tajima's *D* were here taken into consideration by focusing on local windows presenting significant deviation from genome-wide estimates to detect footprints of selection.

Haplotype structure in the populations was determined by imputing loci with missing data and phasing SNPs using BEAGLE v5 (Browning & Browning, 2007) with default settings and an overall effective population size of 100 based on SMC++ (see Figure S1). After BEAGLE locally clustered haplotypes based on their similar probabilities for alleles conditional on the genotype of individuals, hard selective sweeps were distinguished from soft sweeps using SelectionHapStats (Garud, Messer, Buzbas, & Petrov, 2015). Defining haplotype homozygosity as $H1 = \sum_{i=1,...n} p_i^2$ (p_i being the frequency of the *i*th most common haplotype and n the number of observed haplotypes), it is expected to be particularly high for hard sweeps. In contrast, haplotype homozygosity after excluding the most abundant haplotype (H2) is expected to be larger for soft sweeps than for hard sweeps. Accordingly, a summary statistics (H12), combining frequencies of the first and the second most common haplotypes as $H12 = (p_1 + p_2)^2 + \sum_{i>2} p_i^2 = H1 + 2p_1 p_2$, is expected to be large following both a hard sweep (i.e. with only a single dominant haplotype) and a soft sweep model (i.e. with at least two frequent haplotypes). Accordingly, high H12 together with low H2/H1 is consistent with a hard sweep fixing a single haplotype, whereas elevated H2/H1 rather suggests a soft sweep arising from standing genetic variation or, although rather unlikely here, recurrent mutations. As the possible range of H2/H1 varies inversely with H12, such a test has high power to differentiate hard and soft sweeps when H12 is large (Garud et al., 2015). Such haplotype homozygosity was here estimated in window sizes of 50 SNPs, with a step size of 10 SNPs between centres of windows. Windows presenting H12 values above the genome-wide median of H12 were grouped into regions considered as having undergone a selective sweep. Using highest H12 values as representative of considered regions, the top 50 regions were selected as candidates showing strongest and/or most recent selective sweeps. Blocks of significant correlations between polymorphic TEs and LD that were flanking such top 50 regions were defined as influenced by such sweeps. Candidate blocks presenting at least three values of H2/H1 greater than 0.05 (included within the block) were considered as influenced by soft sweeps.

2.5. Gene ontology enrichment

Using high-quality annotations of genes from the *A. alpina* reference genome v5.1 (Jiao et al., 2017) and their assigned ontology (Willing et al., 2015), the enrichment of gene ontology (GO) terms in each of the LD blocks was tested using topGO v2.28 (Alexa, Rahnenfuhrer, & Lengauer, 2006). Significance of terms was determined using Fisher's exact tests. We used the "weight01" method in topGO as it was shown to improve the explanatory power of GO groups by taking hierarchical relationships among terms into account (Alexa et al., 2006). Regions were considered significantly

enriched for genes related to a specific biological process when they displayed a corrected *p*-value <0.05 and also harboured at least five genes associated to the focal GO term. GO enrichment was accordingly evaluated for each block as compared to the whole genome to identify closely located genes related to similar functions. The enrichment of genes included in candidate blocks showing positive or negative correlation of polymorphic TEs with LD was similarly compared to the whole genome.

3. RESULTS

3.1. Distribution and association between LD and polymorphic TEs

Along the eight chromosomes of *A. alpina*, we found high LD within windows of 50kb in genomic regions of high TE density such as pericentromeres, although such a pattern appeared disrupted at several locations (Figure S2). The filtering of SNPs from regions with dense reference TEs may have influenced LD to a certain extent, but genome-wide patterns suggest otherwise as LD was significantly and positively associated with pairwise nucleotide diversity in most chromosomes (Table S1).

Looking at the fine scale along the eight chromosomes of *A. alpina*, multiple regressions highlighted 57 blocks showing significant association between polymorphic TE density and LD while taking reference TE density and pairwise nucleotide diversity into account. Such blocks associating LD with polymorphic TEs were detected on all chromosomes, in TE-rich pericentromeric regions as well as chromosome arms of more euchromatic nature (Figures 1, S2). Noticeably, blocks were detected on the arm of chromosome 4 with very high LD following the increasing density of polymorphic TEs. This finding contrasts with the expected high LD observed in regions of high reference TE density but rather low LD along chromosome arms, as seen in case of e.g. chromosome 5.

Focusing on blocks presenting outlier polymorphic TE densities, 28 candidate blocks made of overlapping and adjacent windows presented significant association between polymorphic TE density and LD (Table 1). The length of these blocks ranged from 1.1 Mb to 5.3 Mb, totaling 74.5Mb (20%) of the genome of *A. alpina* that appeared here with high levels of polymorphic TEs associated to local LD (Figure S2). Across chromosomes, a total of 18 blocks of up to 4.3Mb (called P01 to P18) showed positive association indicative of reduced recombination (high LD) in regions of otherwise high TE density. These blocks included a known supergene (S-locus) that is responsible for self-incompatibility and strict outcrossing on Chromosome 7 (P15). Blocks with low polymorphic TE

density not considered and the remaining 10 blocks up to 5.3Mb (called N01 to N10) showed negative correlations between polymorphic TEs and LD, indicating high recombination in regions of high density of polymorphic TEs. As a whole, these blocks where TE insertions seemingly have an impact on LD and may influence linked selection accounted for 21.4% of polymorphic TEs segregating within the population and 22.5% of genes annotated in *A. alpina*.

3.2. Pattern of selection

Inferences based on unphased SNPs suggested similar demographic histories of the four populations in the last 500-1000 years (Figure S1). Despite fluctuation in the last ca. 50 generations, the impact of demography on inferences of selection is here expectedly minimal and we therefore examined to what extent candidate blocks showing high density of polymorphic TEs associated with high LD were consistent with signatures of selective sweeps. Using Tajima's *D* along chromosomes (Figure S3), six out of the corresponding 18 candidate blocks presented significantly negative *D* values as compared to genome-wide estimates (Table 1). In contrast to such evidence of selective sweeps in regions of high LD, three blocks showed positive Tajima's *D* indicative of linked alleles maintained under balancing selection.

Haplotype homozygosity further supported signals of selective sweeps using the H12 statistic (Figure S3). Grouping consecutive windows with H12 values greater than the median value of H12 (0.35), a total of 986 regions with a maximum length of 1.7Mb were highlighted. Those regions were assigned their highest H12 value and 11 out of the top 50 regions (H12 values between 0.82 and 0.95) showed overlap with nine chromosome blocks where LD was associated with polymorphic TEs (Table 1). As compared to evidence based on Tajima's *D*, this approach identified four additional candidate blocks consistent with selective sweep. In particular, two blocks showing marginally higher Tajima's *D* values (*p*-value >0.01) were characterized here as the result of selective sweeps based on H12 values. It further distinguished soft from hard selective sweeps based on the H2/H1 statistic, and the top 50 loci with maximum H12 values overlapped with several blocks listed in Table 1. Four blocks showing positive correlation between LD and polymorphic TEs contained at least three haplotypes with H2/H1 values >0.05 and were thus considered consistent with soft selective sweeps.

More generally, signatures of adaptive divergence were inspected through the local increase of D_{XY} along chromosomes above the 97.5% percentile of its genome-wide distribution. Although values of pairwise comparisons involving individuals form La Para vs all other populations generally appeared higher, windows of 50kb here showing significantly increased D_{XY} were coinciding with regions under selective sweep as identified by haplotype-based signals.

The S-locus (P15) was highly ranked among the top 50 regions with strong H12 support for selective sweep, although not identified as such by Tajima's D (Table 1). This block is located in the region of high H2/H1 values (Figure S3), supporting at least two frequent haplotypes segregating in the population. Such a signal of soft sweep noticeably coincides with a significant increase in D_{XY} above the 97.5th percentile of all D_{XY} values that indicates putatively adaptive divergence in this region (Figure S4).

3.3. Functional characterization of genes within candidate blocks

Candidate blocks showing significant association of LD with polymorphic TEs contained between 70 and 840 (average 274) annotated genes covering up to 38.8% of the blocks. Among candidate blocks, 19 showed significant enrichment of specific GO terms. Blocks with high density of polymorphic TEs and high LD were enriched in ecologically relevant GO terms related to cold acclimatization (11 significant genes, 28.95% of genes under this term in *Arabis*), response to salt stress (127 genes, 13.57%) and photoperiodic control of flowering time (20 genes, 20%; Table S2). In particular, a block enriched in GO terms related to response to cold showed 21 genes across 2.5Mb of chromosome 8 (P18), whereas others highlighted carbohydrate metabolic process (P04, 20 genes in 1.65Mb) or shoot system development (P05, 18 genes in 2Mb). Moreover, we found genes related to response to bacterium (N09, Figure S5) and response to oxidative stress (N10) in blocks with significantly negative association between polymorphic TEs and LD (Table S3).

The block enriched in cold-responding genes (P18; Figure 2) showed positive correlation of LD with polymorphic TEs and presented high H12 values suggestive of a strong selective sweep. Consistent with a scenario of adaptive divergence, significant peaks of increased D_{XY} were further identified throughout this region, especially among individuals from La Para as compared to all other populations. This LD block presented not only 21 genes enriched in response to cold, but also nine genes related to photoperiodism and seven genes related to repression of flower development (Table S4). In particular, an *Arabis* gene (*Aa_G720620*) is orthologous to the *FLOWERING LOCUS M* (*FLM*) in *Arabidopsis thaliana* and is thus certainly involved in temperature-dependent flowering.

Other genes orthologous to *MADS AFFECTING FLOWERING3* (*MAF3, Aa_G585310*) and *MAF4* (*Aa_G585320*) were identified here. The haplotype-based signature of selective sweep encompasses cold-responding genes as well as genes responding to photoperiodism and flowering present in central position. Both the density of polymorphic TEs and LD, which were already high as compared to whole-genome estimates, further increased in the middle of the block with predominantly heterozygous copies of the abundant and recently active retrotransposon ATLANTYS2 at this position.

4. DISCUSSION

4.1. Pattern of nucleotide diversity and LD in Arabis alpina *populations*

Whole-genome sequencing of 304 individuals of *A. alpina* highlighted limited genomic variation and low genetic substructure among our geographically proximate sampling sites, contrasting with reports from other species (Fischer et al., 2013; Hellsten et al., 2013; Rogivue et al., accepted). These populations indeed have reduced effective population sizes as expected following postglacial expansion under regular selfing (Koch et al., 2006; Buehler et al., 2012; Rogivue et al., 2018). Correspondingly low effective recombination rates are expected to limit both the fixation of adaptive alleles due to positive selection (Hartfield & Glémin, 2014) and the removal of deleterious alleles by purifying selection (Arunkumar, Ness, Wright, & Barrett, 2015). Nearly-neutral alleles are mostly expected under such circumstances (Eyre-Walker, 2006), and selective sweeps were accordingly supported for only a few loci in our study.

Reduced diversity in regions of low recombination is well supported by both theory and empirical data (Cutter & Payseur, 2013; Nachman, 2002). Findings of the present study using local variation in LD to reflect recombination rates along chromosomes, however, hardly support these predictions and rather highlight pairwise nucleotide diversity being positively associated with LD. Loss of variation in typically euchromatic regions of chromosome arms rather than in heterochromatic pericentromeres thus appears coherent with nucleotide diversity shaped by purifying selection that would efficiently remove deleterious mutations in highly recombining regions with low linkage (Arunkumar et al., 2015; Slotte, 2014). Reduced effective recombination rates would also increase the effect of selection on linked sites belonging to several alleles of small effect over larger genomic distances (Charlesworth et al., 1993). A similar distribution of nucleotide diversity and negative correlation with recombination rate was also reported in rice (Flowers et al., 2012), although this study showed reduced variation at the ends of chromosomes, whereas here it

appeared interspersed along chromosomes in *A. alpina*. Polymorphisms still segregating within and among surveyed populations appear mostly as nearly-neutral variation.

4.2. Transposable elements and recombination along chromosomes

The presence of high TE loads in chromosomal regions with high LD is consistent with either the accumulation of their copies where recombination is low or with a direct role played by TEs in locally reducing recombination in *A. alpina*. The genome-wide pattern detected here hardly rejects the former hypothesis assuming background selection to preferentially remove TEs across highly recombining regions of chromosome arms in such a selfing species. However, it would require a strong pressure of deleterious TE-induced mutations that the relatively ancient TE dynamics of *A. alpina* may hardly offer (Choudhury et al, 2017; Willing et al., 2015). Polymorphic TEs surveyed here are indeed mostly representative of standing genetic variation (Rogivue et al., accepted) and thus seem consistent with TE copies driving the recombination landscape. In addition to avoid recombination modifiers along chromosome arms to be treated as a black box, such a hypothesis is empirically supported by TE insertions causing major reduction in recombination rate at a hotspot in budding yeast (Ben-Aroya, Mieczkowski, Petes, & Kupiec, 2004) as well as studies that showed polymorphic arrangements of TEs affecting recombination in maize (He & Dooner, 2009). Accordingly, chromosomal blocks showing correlation of high LD and high density of polymorphic TEs across euchromatic regions of *A. alpina* suggest that TEs may modify recombination even in gene-rich regions.

Molecular mechanisms linking TEs and recombination remain elusive, but epigenetic silencing of TEs is expected to promote compaction of chromatin that might interfere with recombination initiation typically occurring in regions of open chromatin (Kent et al., 2017). Such an effect is expected to act similarly in collinear (i.e. conserved) and polymorphic TEs and may thus contribute to not only reduce recombination in pericentromeres, but also to locally decrease recombination along chromosome arms. However, effective silencing of TEs likely spreads towards adjacent loci and, in gene-rich regions, may suppress expression with chiefly detrimental impact on phenotypes (Cridland, Thornton, & Long, 2015; Hollister & Gaut, 2009; Stuart et al., 2016). Accordingly, Castanera et al. (2016) reported blocks of TEs likely inherited as supergenes that localized together with strongly repressed genes. Suppression of recombination through silencing of polymorphic TEs cannot be ruled out, but heterochromatinization of gene-rich blocks in *A. alpina* seems unlikely as such loci should have been purged from recombining chromosome arms. Noticeably, a direct impact of TE-related rearrangements promoting chromosomal mispairing and

reduction in efficient resolution of cross-overs may also contribute to reduction in recombination (Schubert & Vu, 2016). Local as well as global correlations between polymorphic TEs and LD along chromosomes of *A. alpina* match predictions of such a structural hypothesis. Selfing and low effective population size have likely supported the initial rise of such chromosomal rearrangements (Hoffmann & Rieseberg, 2008), promoting the segregation of polymorphic TEs here under scrutiny at relatively high frequencies. However, molecular underpinnings of interactions between TEs and recombination however deserve further investigation.

4.3. LD blocks and selection

Chromosomal blocks with higher LD indicative of reduced recombination in regions with high density of polymorphic TEs included the S-locus supergene (P15) encoding self-incompatibility to ensure outcrossing (Vekemans, Poux, Goubet, & Castric, 2014). This block of 2.2 Mb presented genomic variation indicative of a soft selective sweep in the surveyed population. Although consistent with positive rather than the expected frequency-dependent selection, such a soft sweep pattern may reflect recent balancing selection initiated 0.02–0.4 Ne generations ago (Fijarczyk & Babik, 2015). Populations under scrutiny appear mostly selfing (e.g. Rovigue, et al., accepted) and their low Ne makes a recent transition to selfing unlikely. The pattern highlighted here at the S-locus may have been generated by selective sweeps of non-functional alleles conferring advantageous selfing in alpine populations of *A. alpina*. Extended LD at multiple frequent haplotypes showing reduced differentiation among populations is an expectedly long-lasting pattern (Fijarczyk & Babik, 2015), and further characterization of such a locus will have to disentangle how TEs interact with recombination and selection to shape complex signals of genomic variation.

Another LD block enriched in cold response genes (P18) was identified here as having undergone a hard selective sweep as indicated by the haplotype-based H12 statistics. This 2.5Mb locus presented several genes orthologous to loci contributing to vernalisation response in *A. thaliana*. In particular, the centre of the block shows orthologue of *FLM* that appears to be key to the repression of flowering at low ambient temperatures (Lutz et al., 2015). Moreover, expression of the nearby *MAF4* during cold conditions ensures that vernalisation is not achieved prematurely (Kim & Sung, 2013). This block further presented genes related to the regulation of transcription and may trigger chromatin modifications and repression of *FLM* expression as is the case for *PEP1* in *A. alpina* (Wang et al., 2009). Cold-induced freezing tolerance enhanced by short photoperiods highlights systemic responses to the environment mediated by the circadian clock in *A. thaliana* (Franklin, Toledo-Ortiz, Pyott, & Halliday, 2014) and species such as the perennial ryegrass present

intraspecific variation in interactions between the circadian network and freezing tolerance (Paina, Byrne, Domnisoru, & Asp, 2014). Accordingly, such a block enriched in cold-responding as well as photoperiodic and flowering genes may be of functional significance in natural populations of *A. alpina*. In view of increasing evidence that genes are not distributed randomly across chromosomes in eukaryotes (Boycheva et al., 2014; Pavlidis et al., 2012), the enrichment of similar GO terms within blocks of *A. alpina* suggests a role of TEs in modifying linkage in such regions. The pattern of increased LD being modulated by higher polymorphic TE density in the middle of the locus P18 particularly supports the hypothesis that recently inserted TEs increased linkage among adjacent loci, although uncoupling cause and effect is challenging (Kent et al., 2017). Low-recombining regions were suggested to shield sets of adaptive variants and offer examples of 'selection at linked sites' in threespine sticklebacks (Samuk et al. 2017), whereas González et al. (2008) evidenced selection in the regions flanking two TE insertions and later linked them to fitness effects related to insecticide resistance in *Drosophila melanogaster* (Guio, Barrón, & González, 2014; Mateo, Ullastres, & González, 2014). It remains to be demonstrated to what extent polymorphic TEs in the P18 block are here coupling cold-responding genes with genes related to photoperiodism and flowering, supporting the emergence of linked loci under selection.

4.4. TEs, selection and the recombinational landscape

The evolutionary arms race of TEs and host genomes can create complex recombination landscape across chromosomes (Kent et al., 2017), as also highlighted here through large blocks of chromosome arms presenting high LD going along with high density of polymorphic TEs. To what extent and how polymorphic TEs could act as recombination modifiers through either heterochromatinization or chromosomal restructuring remains elusive. Although TE silencing in gene-dense regions is still largely unknown (Takuno, Seymour, & Gaut, 2017), the expected suppression of expression along flanking sites should mostly yield deleterious variation removed by background selection.

The present study on the impact of polymorphic TEs on LD in *A. alpina* suggests some interactions between selection and specific LD blocks. Several loci highlighted here with indirect evidence of selective sweeps suggest that TEs may be central players of adaptive processes. The majority of LD blocks highlighted here were consistent with polymorphic TEs promoting long-range linked selection. However, a few blocks along euchromatic regions of *A. alpina* presented negative correlation between LD and polymorphic TE density, as expected if high density of polymorphic TEs had promoted elevated recombination (McVean, 2010). Such loci in particularly weak LD were

generally enriched in functions requiring high allelic diversity such as plant resistance genes (e.g. N09, Figure S5). Retrotransposons conferring partial resistance to bacterial pathogens have already been reported (Wang et al., 1998) and, as postulated for the hot pepper genome, high TE dynamics may foster crucial variation proving vital to the evolution of plant disease resistance (Kim et al., 2017). The significance of similar blocks related to other factors such as oxidative stress in heterogeneous alpine environment remains to be determined, but chromosome blocks identified here were generally consistent with the predictions that different genome fractions evolve at variable rates under different constraints (Choi et al., 2016; Dong, Raffaele, & Kamoun, 2015). Such blocks identified in high-resolution datasets from natural populations should be fully characterized regarding not only their functionality, but also their fitness effects (Barrett & Hoekstra, 2011), and thereby contribute to decipher the surmised role of TEs in genome restructuring and adaptive processes (Bonchev & Parisod, 2013).

ACKNOWLEDGEMENTS

REFERENCES

Alexa, A., Rahnenfuhrer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*, 1600–1607.

Arunkumar, R., Ness, R. W., Wright, S. I., & Barrett, S. C. H. (2015). The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics*, *199*, 817–829.

Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, *12*, 767–780.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., … de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, *327*, 836–840.

Ben-Aroya, S., Mieczkowski, P. A., Petes, T. D., & Kupiec, M. (2004). The compact chromatin structure of a Ty repeated sequence suppresses recombination hotspot activity in *Saccharomyces cerevisiae*. *Molecular Cell*, *15*, 221–231.

Bonchev, G., & Parisod, C. (2013). Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources*, *13*, 765–775.

Boycheva, S., Daviet, L., Wolfender, J.-L., & Fitzpatrick, T. B. (2014). The rise of operon-like gene clusters in plants. *Trends in Plant Science*, *19*, 447–459.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, *81*, 1084–1097.

Buehler, D., Graf, R., Holderegger, R., & Gugerli, F. (2012). Contemporary gene flow and mating system of *Arabis alpina* in a central European alpine landscape. *Annals of Botany*, *109*, 1359–1367.

Carja, O., Liberman, U., & Feldman, M. W. (2014). Evolution in changing environments: modifiers of mutation, recombination, and migration. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 17935–17940.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., … Ramírez, L. (2016). Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics*, *12*, e1006108.

Charlesworth, B. (1992). Evolutionary rates in partially self-fertilizing species. *American Naturalist*, *140*, 126–148.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*, 1289–1303.

Choi, K., Reinhard, C., Serra, H., Ziolkowski, P. A., Underwood, C. J., Zhao, X., … Henderson, I. R. (2016). Recombination rate heterogeneity within *Arabidopsis* disease resistance genes. *PLoS Genetics*, *12*, e1006179.

Choudhury, R. R., Neuhaus, J.-M., & Parisod, C. (2017). Resolving fine-grained dynamics of retrotransposons: comparative analysis of inferential methods and genomic resources. *Plant Journal*, *90*, 979–993.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, *8*, e1002905.

Cridland, J. M., Thornton, K. R., & Long, A. D. (2015). Gene expression variation in *Drosophila melanogaster* due to rare transposable element insertion alleles of large effect. *Genetics*, *199*, 85–93.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, *14*, 262–274.

Dapper, A. L., & Payseur, B. A. (2017). Connecting theory and data to understand recombination rate evolution. *Philosophical Transactions of the Royal Society, Biological Sciences*, *372*, 20160469.

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, *35*, 57–65.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*, 422–433.

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, *21*, 569–575.

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, *24*, 3529–3545.

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., … Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, *22*, 5594–5607.

Flowers, J. M., Molina, J., Rubinstein, S., Huang, P., Schaal, B. A., & Purugganan, M. D. (2012). Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Molecular Biology and Evolution*, *29*, 675–687.

Franklin, K. A., Toledo-Ortiz, G., Pyott, D. E., & Halliday, K. J. (2014). Interaction of light and temperature signalling. *Journal of Experimental Botany*, *65*, 2859–2871.

Gao, X., Hou, Y., Ebina, H., Levin, H. L., & Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Research*, *18*, 359–369.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv: 1207.3907v2*.

Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genetics*, *11*, e1005004.

González, J., Lenkov, K., Lipatov, M., Macpherson, J. M., & Petrov, D. a. (2008). High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biology*, *6*, e251.

Grant, V., & Flake, R. H. (1974). Solutions to the cost-of-selection dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, *71*, 3863–3865.

Guio, L., Barrón, M. G., & González, J. (2014). The transposable element Bari-Jheh mediates oxidative stress response in *Drosophila*. *Molecular Ecology*, *23*, 2020–2030.

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, *27*, 2477–2497.

Hartfield, M., & Glémin, S. (2014). Hitchhiking of deleterious alleles and the cost of adaptation in partially selfing species. *Genetics*, *196*, 281–93.

He, L., & Dooner, H. K. (2009). Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8410–8416.

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., … Rokhsar, D. S. (2013). Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 19478–19482.

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, *39*, 21–42.

Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, *19*, 1419–1428.

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*, 583–589.

Jiao, W.-B., Accinelli, G. G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., … Schneeberger, K. (2017). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Research*, *27*, 778–786.

Kent, T. V, Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society, Biological Sciences*, *372*, 20160458.

Kim, D.-H., & Sung, S. (2013). Coordination of the vernalization response through a *VIN3* and *FLC* gene family regulatory network in *Arabidopsis*. *Plant Cell*, *25*, 454–469.

Kim, S., Park, J., Yeom, S.-I., Kim, Y.-M., Seo, E., Kim, K.-T., … Choi, D. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biology*, *18*, 210.

Koch, M. A., Kiefer, C., Eheich, D., Vogel, J., Brochmann, C., & Mummenhoff, K. (2006). Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology*, *15*, 825–839.

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., … Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 816–821.

Lee, Y. C. G. (2015). The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. *PLoS Genetics*, *11*, e1005269.

Levy, A. A. (2013). Transposons in Plant Speciation. In N. V. Federoff (Ed.) *Plant Transposons and Genome Dynamics in Evolution* (pp. 165–179). Oxford, UK: Wiley.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595.

Lotterhos, K., Yeaman, S., Degner, J., Aitken, S., & Hodgins, K. (2018). Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biology, 19*, 157.

Lumley, T. (2017). Leaps: regression subset selection. Retrieved from https://cran.r-project.org/web/packages/leaps/leaps.pdf

Lutz, U., Posé, D., Pfeifer, M., Gundlach, H., Hagmann, J., Wang, C., … Schwechheimer, C. (2015). Modulation of ambient temperature-dependent flowering in *Arabidopsis thaliana* by natural variation of *FLOWERING LOCUS M*. *PLoS Genetics*, *11*, e1005588.

Marand, A. P., Jansky, S. H., Zhao, H., Leisner, C. P., Zhu, X., Zeng, Z., … Jiang, J. (2017). Meiotic crossovers are associated with open chromatin and enriched with Stowaway transposons in potato. *Genome Biology*, *18*, 203.

Mateo, L., Ullastres, A., & González, J. (2014). A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLoS Genetics*, *10*, e1004560.

McVean, G. (2010). What drives recombination hotspots to repeat DNA in humans? *Philosophical Transactions of the Royal Society, Biological Sciences*, *365*, 1213–1218.

Mézard, C., Tagliaro Jahns, M., & Grelon, M. (2015). Where to cross? New insights into the location of meiotic crossovers. *Trends in Genetics*, *31*, 393–401.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., … Mathieu, O. (2009). Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature*, *461*, 427–

430.

Nachman, M. W. (2002). Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics & Development*, *12*, 657–663.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*, 1566–1575.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402.

Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, *6*, 119–127.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., … Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, *327*, 92–94.

Paina, C., Byrne, S. L., Domnisoru, C., & Asp, T. (2014). Vernalization mediated changes in the *Lolium perenne* transcriptome. *PLoS ONE*, *9*, e107365.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., … Keeney, S. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell*, *144*, 719–731.

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, *29*, 3237–3248.

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, *30*, 2224–2234.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*, 1929–1936.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208–R215.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

Rizzon, C., Marais, G., Gouy, M., & Biémont, C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Research*, *12*, 400–407.

R Development Core Team (2018) *R: a language and environment for statistical computing*.

http://www.R-project.org

Rodgers-Melnick, E., Bradbury, P. J., Elshire, R. J., Glaubitz, J. C., Acharya, C. B., Mitchell, S. E., … Buckler, E. S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 3823–3828.

Rogivue, A., Choudhury, R. R., Zoller, S., Joost, S., Felber, F, Kasser, M., Parisod, C., & Gugerli. F. (2018). Genome-wide variation in nucleotides and retrotransposons in alpine populations of *Arabis alpina* (Brassicaceae). *Molecular Ecology Resources*, Manuscript accepted.

Rogivue, A., Graf, R., Parisod, C., Holderegger, R., & Gugerli, F. (2018). The phylogeographic structure of *Arabis alpina* in the Alps shows consistent patterns across different types of molecular markers and geographic scales. *Alpine Botany*, *128*, 35–45.

Roze, D., & Lenormand, T. (2005). Self-fertilization and the evolution of recombination. *Genetics*, *170*, 841–857.

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*, 4378–4390.

Schubert, I., & Vu, G. T. H. (2016). Genome stability and evolution: attempting a holistic view. *Trends in Plant Science*, *21*, 749–757.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*, R288-R294.

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., & Levy, A. A. (2015). DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. *Plant Cell*, *27*, 2427–2436.

Slotte, T. (2014). The impact of linked selection on plant genomic variation. *Briefings in Functional Genomics*, *13*, 268–275.

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society, Biological Sciences*, *372*, 20160455.

Stuart, T., Eichten, S. R., Cahn, J., Karpievitch, Y. V, Borevitz, J. O., & Lister, R. (2016). Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *eLife*, *5*, e20777.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Takuno, F., Seymour, D. K., & Gaut, B. S. (2017) The evolutionary dynamics of orthologs that shift in gene body methylation between *Arabidopsis* species. *Molecular Biology and Evolution*, *34*, 1479–1491.

Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., … Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Research*, *19*, 2221–2230.

Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A., & Kakutani, T. (2009). Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature*, *461*, 423–426.

Vekemans, X., Poux, C., Goubet, P. M., & Castric, V. (2014). The evolution of selfing from outcrossing ancestors in Brassicaceae: what have we learned from variation at the *S-* locus? *Journal of Evolutionary Biology*, *27*, 1372–1385.

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334–4345.

Wakeley, J. (1996). The variance of pairwise nucleotide differences in two populations with migration. *Theoretical Population Biology*, *49*, 39–57.

Wang, G. L., Ruan, D. L., Song, W. Y., Sideris, S., Chen, L., Pi, L. Y., … Ronald, P. C. (1998). Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *Plant Cell*, *10*, 765–779.

Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., … Albani, M. C. (2009). *PEP1* regulates perennial flowering in *Arabis alpina*. *Nature*, *459*, 423–427.

Willing, E.-M., Rawat, V., Mandáková, T., Maumus, F., James, G. V., Nordström, K. J. V., … Schneeberger, K. (2015). Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*, *1*, 14023.

Wright, S. I., Agrawal, N., & Bureau, T. E. (2003). Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research*, *13*, 1897–1903.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, E1743-E1751.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*, 1897–1911.

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*, 1256–1270.

SUPPLEMENTARY INFORMATION

Additional Supplementary Information may be found in the online version of this article.

Text S1: Analysis of multilocus genotypes

Table S1: Genome-wide correlation of LD with nucleotide diversity, reference TE density and polymorphic TE density

Table S2: Gene Ontology analysis for genes in positively correlated blocks

Table S3: Gene Ontology analysis for genes in negatively correlated blocks

Table S4: Gene Ontology analysis for the block P18

Table S5: Differences in Dxy per population pair

Figure S1: Changes in effective population sizes over time and split times of populations

Figure S2: Genomic features along chromosomes of *Arabis alpina* in windows of 50kb

Figure S3: Signals of selective sweep along chromosomes of *Arabis alpina*

Figure S4: Representation of block P15 containing the S-locus

Figure S5: Representation of block N09 containing genes responding to bacterium

Figure S6: Frequency distribution of polymorphic TEs per block

Data Accessibility

SNPs and polymorphic TEs used in this manuscript are submitted on Dryad doi:10.5061/dryad.58g217k

Authors' contributions

Designed research: RRC, FG, CP

Analyzed datasets: RRC, AR, CP

Wrote manuscript: RRC, CP

All authors read, commented and approved the final manuscript.

Table 1 Blocks along eight chromosomes (Chr) of *Arabis alpina* showing significant correlation (+, positive; -, negative) between linkage disequilibrium (LD) and density of polymorphic transposable elements (PolyTE; Figure S2). Signatures of selective sweep are based on Tajima's *D* and haplotype-based scores (H12; Figure S3). Enriched GO term within each block with maximum number of genes (-, no significant term; Tables S2, S3).

Block No.	Chr	Start (Mb)	Length (Mb)	PolyTE correlation with LD (Count)	*D*[a]	H12 (max H12, Soft Sweep)[b]	GO enrichment (with number of genes)
P01	1	7.2	1.6	+ (123)	- *	NS	Protein autophosphorylation:7
P02	1	24.9	1.8	+ (104)	NS	NS	-
P03	2	21.8	1.6	+ (86)	+ ***	NS	-
P04	3	3.6	1.6	+ (117)	- ***	Sig (0.87,+)	Carbohydrate metabolic process:20
P05	4	0	2.0	+ (102)	NS	Sig (0.95)	Shoot system development:18
P06	4	10.1	2.5	+ (144)	- ***	NS	-
P07	6	6.9	2.1	+ (168)	+ ***	NS	Ubiquitin-dependent catabolic process:8
P08	6	29.7	1.6	+ (94)	+ *	Sig (0.88)	Response to salt stress:11
P09	7	1.4	2.8	+ (149)	- **	NS	Defense response to bacterium:26
P10	7	6.1	1.6	+ (96)	NS	Sig (0.83)	Cellular amino acid metabolic process:9
P11	7	17.3	1.5	+ (99)	NS	NS	-
P12	7	24.2	1.5	+ (88)	NS	NS	-
P13	7	29.9	1.1	+ (57)	- ***	NS	-
P14	7	38.5	1.5	+ (117)	NS	NS	Response to abscisic acid:9
P15	7	43.0	2.2	+ (125)	NS	Sig (0.87, +)	Response to abscisic acid:13

P16	8	0	2.0	+ (147)	NS	NS	Abscisic acid-activated signaling:7
P17	8	25.7	1.6	+ (106)	- *	Sig (0.85, +)	-
P18	8	47.0	2.5	+ (119)	NS	Sig (0.86)	Response to cold:21
N01	2	30.5	1.7	- (140)	NS	NS	Lipid metabolic process:10
N02	2	34.6	5.3	- (290)	- ***	NS	Response to cadmium ion:38
N03	3	8.4	3.4	- (192)	+ *	Sig (0.86,+)	Response to salt stress:20
N04	4	33.7	2.0	- (99)	+ ***	NS	-
N05	5	2.6	2.1	- (148)	- ***	NS	-
N06	5	6.1	3.0	- (224)	- ***	Sig (0.91)	Chromosome organization:21
N07	5	26.7	1.7	- (118)	- ***	NS	Cellular localization:9
N08	6	22.5	4.5	- (393)	+ ***	NS	Ion transport:15
N09	7	33.8	3.7	- (172)	- ***	NS	Response to bacterium:7
N10	7	40.0	2.6	- (222)	+ ***	NS	Response to oxidative stress:15

[a] Tajima's D, with – for excess of rare alleles (selective sweep) and + for excess of frequent alleles (balancing selection), p-values: ***≤0.001, **≤0.01, *≤0.05
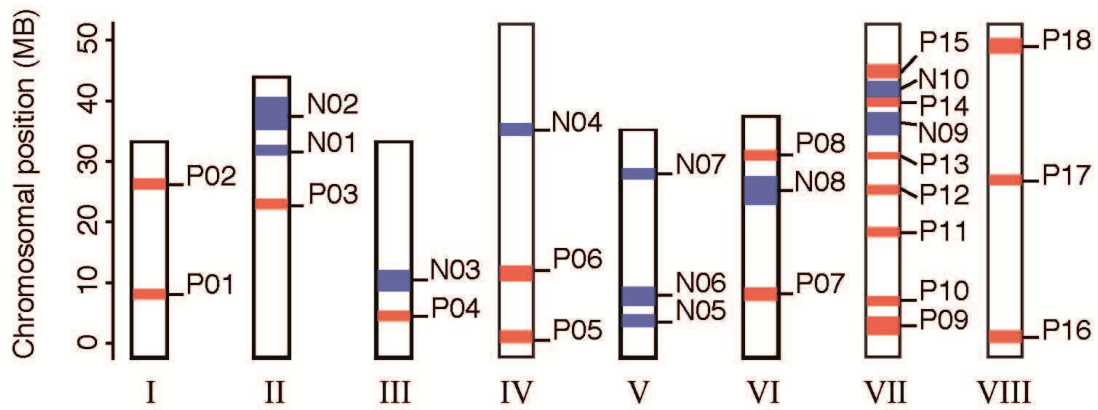
[b] H12 test where blocks containing loci ranked among the top 50 peaks from genome-wide estimates are represented as significant selective sweep (Sig), with the maximum H12 value reported in the blocks as well as evidence of soft sweep (+) in between parentheses; NS, non-significant.

Figure legends

Figure 1 Correlation of polymorphic LTR retrotransposons (TEs) with linkage disequilibrium (LD) in *Arabis alpina*. a) Ideogram of the eight chromosomes showing eighteen blocks with high LD positively correlated with polymorphic TEs (red) as well as ten blocks with negative correlation (blue; numbered according to Table 1). b) Pattern of high LD in windows of 50kb of high-density polymorphic TE density is shown at the start of chromosome IV, contrasting to the pattern along chromosome V, where the overall TE density increases with LD. Each solid line represents a LOESS smooth of the corresponding feature along each chromosome. Genome-wide patterns are presented in Figures S2 and S3.

Figure 2 Representation of chromosome block P18 enriched in cold-responding genes in *Arabis alpina*. a) D_{XY} among all pairs of four studied alpine populations, with dotted line representing the 97.5 percentile across all pairs of populations (symbols according to panel with La Para: Pa, Essets: Es, Pierredar: Pi, Martinets: Ma). b) Haplotype-based H12 statistic for selective sweep with dotted line representing median value of genome-wide H12 estimates, c) pattern of linkage disequilibrium (LD), d) polymorphic LTR retrotransposon (TE) density, and e) location of genes related to cold, photoperiodism and flowering (coloured accordingly).

a)

b)