Selection inference in a complex genomic landscape: the impact of polymorphic inversions

Julia Morales-García[1,2], Wiesław Babik[1], Piotr Zieliński[1], Krystyna Nadachowska-Brzyska[1*]

[1]Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Gronostajowa 7, 30-387, Kraków, Poland

[2]Doctoral School of Exact and Natural Sciences, Jagiellonian University, ul. Łojasiewicza 11, 30-348, Kraków, Poland

[*]Corresponding author: Krystyna Nadachowska-Brzyska, krystyna.nadachowska@uj.edu.pl

Abstract

Understanding the genetic basis of adaptation is a key objective in evolutionary biology. Although advances in genomic selection scans have greatly improved our ability to detect signatures of adaptation, distinguishing true signals from false positives remains challenging. This task is particularly difficult in regions of reduced recombination, such as polymorphic inversions. In this study, we examine the genome-wide landscape of selection in the European spruce bark beetle (*Ips typographus*), Europe's most destructive forest pest, which has one of the most complex inversion-associated recombination landscapes known. Using simulation-based analyses and whole-genome resequencing data from 312 individuals across 23 populations, we applied two complementary selection scan methods (nSL and Λ) to assess how inversions influence the detection of adaptive signals. Simulations revealed that the two selection scan methods differ in their susceptibility to false-positive signals within inversions and that partitioning data by inversion genotype (i.e., homozygote classes) can substantially reduce these errors. Consistent with the results of our simulations, our empirical data showed that inversions are highly enriched for selection signals when all inversion genotypes are analyzed but are depleted when only homozygotes are considered. This demonstrates that focusing on homozygote genotypes allows for a more precise identification of putative selection targets and haplotype-specific signals, as it overcomes the confounding effects of suppressed recombination in heterokaryotypic individuals. In contrast, the detection of selection in collinear regions was largely unaffected by the presence of inversions. Our findings highlight that inversions may play an important role in shaping adaptation, underscoring the need to account for species-specific genomic architecture when interpreting signals from selection scans.

Author summary

Detecting the genetic basis of adaptation is one of the main goals in evolutionary biology. However, it is often difficult to tell whether selection signals identified by genome scans truly reflect natural selection signatures or are simply artefacts of genome structure. In this study, we examined how polymorphic inversions affect the detection of selection in the European spruce bark beetle, one of Europe’s most destructive forest pests. First, we assessed two selection scan methods using simulation approach. Second, we tested the influence of polymorphic inversion on the selection inference using genome-wide data from 23 beetle

populations. We showed that inversions can create false signals of selection when all individuals are analyzed together but by separating individuals according to their inversion genotypes, we were able to distinguish potential adaptive regions from false positives. Our results demonstrate that accounting for genomic architecture is crucial for identifying selection signatures and provide practical guidelines for studying species with inversion-rich genomes.

Introduction

Elucidating the genetic basis of adaptation is one of the major goals of evolutionary biology [1]. The long-standing debate over whether selection acts primarily on many loci of small effect or few loci of large effect, and whether de novo mutations or standing genetic variation are more important in adaptation [1-8] has led to a large body of empirical adaptation research [1,9-10]. The progress in this field has been facilitated by a growing number of increasingly sophisticated methods for genomic selection scans [11-13]. Despite many methodological advances and the fact that selection leaves characteristic signatures around its targets, there are still serious limitations in distinguishing true selection signals from false positives. In particular, we still have a limited understanding of how regions of reduced recombination affect our ability to identify true selection signals and the general importance of such regions in adaptation [14-16].

Local reduction of recombination is one of the main characteristics of polymorphic inversions - chromosomal mutations that reverse the order of the sequence in the affected region and suppress recombination between two inversion haplotypes in heterozygous individuals. The effect of suppressed recombination may extend beyond the inverted region itself [16-20]. Furthermore, the recombination landscape in these regions is shaped by the frequency of inversion haplotypes - the more balanced the frequencies, the higher the number of heterozygous individuals in the population and the lower the population recombination rate in the inverted region [21]. A direct consequence of suppressed recombination between alternative inversion haplotypes is independent accumulation of mutations on both arrangements. This creates a genetic subdivision within the inversion region - a situation that is similar to a demographic scenario of two isolated populations [22-25]. In addition, gene flux (double-crossover and conversion events) that can happen between inversion haplotypes mimics a scenario of two isolated populations with some migration between them [20,23,26]. These inherent features of inversion polymorphism are important since free recombination

(between loci) and panmixia are common assumptions of selection scans and many other inference methods. While new methods are often evaluated for their robustness to recombination rate variation and population subdivision [12,14,27], explicit inversion polymorphic scenarios have rarely been evaluated (but see [28] or [15]).

There are several reasons why including inversion regions in selection inference is important. First, theoretical investigations clearly showed that selection plays a key role in the establishment, spread and the long-term fate of inversions [16,29-35] and unlike other mutations, inversions are rarely selectively neutral [31,36]. Polymorphic inversions can be maintained by divergent selection and migration (e.g. inversion captures locally adapted alleles that are kept together in the face of gene flow [33]), different types of balancing selection [37-38] or, most likely, by a combination of these and other evolutionary processes (for a review, see [36,39]). Importantly, the relative roles of the different selection forces, drift and new mutations that accumulate within different inversion haplotypes (including both advantageous and deleterious ones) may change over time. Second, there is a growing body of empirical research that identifies polymorphic inversions as key drivers of adaptation that capture genes underlying traits involved in local adaptation [40-46]. Finally, accumulating evidence suggests that genomes harboring multiple polymorphic inversions and characterized by complex recombination landscapes may be more common than previously thought (so-called inversion-rich genomes [40,47]). This provides an opportunity to investigate selection signatures among many inversions in a single genome.

Due to the dynamic nature of inversions, it can be difficult to detect and/or distinguish true selection signatures from false positive signals. However, simulation-based studies have demonstrated that adaptive inversions carrying locally adaptive loci can be distinguished from neutral ones [16]. Yet it has also been shown that several selection scan methods can produce strong false positive signals within inversions [15]. Studies have also shown that inversions not only capture beneficial alleles but can also accumulate new advantageous mutations over time [16]. These mutations can undergo selective sweeps within inversion homozygotes and should be detectable in datasets containing the relevant haplotype, unless the underlying phenotypes are polygenic. In this case, the usual challenges of identifying selection signals at loci of small effect apply [2,15-16,48-49].

Here, we performed both simulation-based and empirical assessments of two selection scan methods using the inversion-rich genome of the spruce bark beetle [50] and provide guidance

for methods usage in such genomes. The evaluated methods were the number of segregating sites by length (nSL) [27] and Λ from the saltiLASSI maximum likelihood approach [12]. We focused on methods that do not require recombination rate information, as this is still lacking for many of the species for which genomic data are being generated. Both methods are designed to search for signatures of hard and soft sweeps of ongoing or recent selection, but they differ in how they are affected by reduced recombination. While nSL has been shown to be quite robust to variation in recombination rate, Λ tends to be negatively correlated with recombination rate and strong Λ signals in low-recombination regions must be interpreted with extra caution [12,27]. Guided by simulation results we tested if applying selection scan methods in datasets restricted to inversion homozygotes reveal haplotype specific selection signals. Finally, we identified candidate loci under selection within the spruce bark beetle genome.

Materials and Methods

Study system

The European spruce bark beetle (*Ips typographus*, hereafter the spruce bark beetle) is found in Norway spruce forests throughout Eurasia. It is often considered to be the most destructive forest pest, which under certain conditions can cause mass mortalities of spruce stands [51-52]. In recent decades, exacerbated by climate change, the frequency and severity of outbreaks have dramatically increased, making the spruce bark beetle a major economic and environmental concern in European forests [53-55]. Extensive research on species ecology and biology, as well as, on causes and consequences of the spruce bark beetle outbreaks, has been recently complemented by population genomic studies [50,56]. Importantly, these studies identified 27 large polymorphic inversions covering approximately 30% of the species genome [50]. Most of the identified inversions are polymorphic across species range but can significantly vary in frequency among populations which makes the spruce bark beetle an excellent model to study natural selection signatures in the context of inversion rich genome.

Sampling, sequencing, and data preparation

We used data presented in Mykhailenko et al. (2024) [50] and Zieliński et al. (2025) [56]. The data included 312 individuals from 23 populations (13-14 individuals per population). Samples

covered most of the species' range in Europe and ranged from northern Italy to northern Fennoscandia (Figure 1; Table S1). The details of library preparation, sequencing and processing are given in Mykhailenko et al. (2024) [50] and Zieliński et al. (2025) [56]. In short, DNA was extracted from the whole beetles using the Wizard Genomic DNA Purification Kit (Promega). Genomic libraries were subject to 2×150 bp paired-end sequencing using the NovaSeq technology (Illumina Inc.). Raw data was subject to quality control with FastQC [57]. Low quality reads and adaptors were trimmed with Trimmomatic [58] and mapped to the spruce bark beetle genome [59] with Bowtie2 [60]. Duplicate reads were removed with PicardMarkDulicates (https://broadinstitute.github.io/picard/). Variant calling, genotyping, base quality score recalibration, and filtering were performed with GATK, following the GATK best practice guideline. Filtering included: removing five sites around indels and sites within repeat-masked regions of the genome, and variants exhibiting: low-quality (mapping quality below 30), total depth higher than mean + 1SD, excessive heterozygosity. Individual genotypes with coverage lower than 8 and genotype quality lower than 20 were treated as missing. In downstream analysis we used only biallelic SNPs with more than 50% of genotypes called. The data for all individuals was phased using default parameters of Beagle 5.2 [61]. All analyses were performed using contigs larger than 1 >Mb (IpsContigs 1-36, 78% of assembly length). Moreover, IpsContig33 which had high similarity to mtDNA and potentially sex linked contigs were excluded (IpsContig9, IpsContig27, IpsContig31 and IpsContig35, Nadachowska-Brzyska in preparation).

Selection scans

nSL was calculated in selscan 2.0.2 [62] using default parameters and phased data (it is not possible to use unphased data if it contains missing genotypes - a very common situation with whole-genome resequencing data). The nSL values for each SNP were normalised against genome-wide distribution using 'norm' function and default settings. Because the sign of nSL depends on the unknown ancestral versus derived state of the allele, we followed standard practice and used the absolute values to capture both types of signatures. The top 1% of nSL absolute values were treated as outliers. Λ was inferred using LASSI-Plus1.1.2 [12]. In contrast to the nSL approach, which provides statistics for each analysed SNP, Λ is given for windows consisting of a predefined number of SNPs (hereafter called SNP-based window). In addition, different window sizes may be more suitable for detecting recent or more ancient sweeps [12].

171 We tested different window sizes and based on our tests and previous studies [12], we decided
172 to use two window sizes: 52 (sliding window step 12) and 117 SNPs (sliding window step 12).
173 The top 1% of Λ values were considered outliers.

174 The nSL results based on the spruce bark beetle data were translated into SNP-based Λ
175 windows and mean values of nSL were calculated in each window. This was done to be able
176 to statistically compare selection signals between statistics. The top 1% mean (window) nSL
177 absolute values were considered outliers. In order to be able to compare different datasets that
178 contained different individuals and thus different numbers of SNPs and different numbers of
179 SNP-based windows, we further standardised the results by summarising the statistics (both
180 nSL and Λ) in non-overlapping 50 kb genomic windows. For each 50 kb window, we calculated
181 the mean value of each statistic of all SNP-based windows that overlapped it. This allowed a
182 consistent window-based comparison across datasets and statistics.

183 Finally, to test if selection signals tend to cluster within inversions, we plotted outlier windows
184 (SNP or 50kb windows according to each case) along the genome, investigated the position of
185 each outlier window in comparison to inversion coordinates and calculated the sizes of clusters
186 composed of adjacent outlier windows. Additionally, to correlate selection statistics with the
187 genetic differentiation between inversion haplotypes we summarised Fst in SNP-based
188 windows (in the same way as we did for nSL results) using vcftools [63].

189

190 *Simulations*

191 To evaluate the effect of inversions on selection scan methods we first investigated selection
192 scans performance using simulations (Figure 2a). We based our test on the existing and publicly
193 available set of simulation performed by Lotterhos (2019) [15] who created a resource to test
194 new methods and their sensitivity to realistic patterns of genomic variation, including
195 polymorphic inversion. In short, the author performed 200 replicate forward-time simulations
196 of a metapopulation adapting to a heterogenous spatial environment. The simulations were
197 performed in SLiM v 3.2 [64]. The simulated data included nine linkage groups (LGs) each
198 evolving independently and under different evolutionary scenarios: two LGs followed neutral
199 evolution (LG1 and LG2), and the rest included multiple casual quantitative trait nucleotides
200 (LG3 and LG4), complete selective sweep (LG5), partial sweep (LG6), inversion maintained
201 polymorphic by weak frequency dependent selection (LG7), low recombination region
202 mimicking centromere (LG8) and randomly varying recombination across the LG (LG9). Each

LG was 50kb and 50cM in length and simulations included 1000 individuals adapting to heterogenous spatial environment. Simulated inversions were 10kb long and were located in the middle of the LG7, their fitness was not related to local adaptation to the environment and all mutations within the inversion were neutral. For our evaluation we used vcf files deposited at Dryad ([15] results from 181 replicate simulations). In each simulation, individuals were classified according to their inversion genotypes using the approach described in Mykhailenko et al. (2024) [50]. Reliable inversion genotyping was possible for 179 of the 181 simulations. Classification accuracy was verified with inversion PCA plots, which showed individuals clustering distinctly into three genotype groups, and were coloured according to the inferred inversion genotypes (Supplementary Information: Simulations Results .zip folder).

Selection scans were conducted using all replicate simulations and all individuals (including all inversion genotypes), as well as datasets consisting of homozygotes of one type only (Figure 2a). Additionally, all simulations were downsampled from 1000 to 200 individuals (similar sample size as the spruce bark beetle sample size) to evaluate the impact of sample size on selection statistics. Since homozygote datasets in the spruce bark beetle data often substantially differ in sample size, in particular if one inversion haplotype is rare, we also downsampled homozygote datasets in simulated data to 10 individuals. We used the selection scan results across these datasets to test whether the two evaluated methods of selection scans produced false positive signals within neutral inversions and whether sample size affected the inferred selection landscape within inversions. To assess false positives, we counted the number of outliers within inversion regions and compared this to the number of outliers detected in neutral linkage groups (LG1 and LG2; chi-squared tests were used to compare the proportion of outliers in inversion versus collinear regions). This test was conducted for datasets including all individuals, homozygote-only datasets, and all replicate simulations. To examine the effect of sample size, we used generalized linear mixed models (GLMM; *glmmTMB* package [65]; Supplementary Information: GLMM.R) with a binomial response representing the number of outlier SNPs and non-outlier SNPs detected in inversion region. The simulation replicate was included as a random effect to account for the variation between simulations. The fixed effects were sample size (200 versus 1000 individuals in datasets including all genotypes, and 200 versus 10 individuals in datasets including only homozygotes). Through this approach, we investigated whether a sample size of approximately 200 individuals (comparable to the spruce bark beetle dataset) would yield selection scan outcomes within the inversion region that mirrored those achieved with a sample size of 1000 individuals. We also tested whether a very

236 small sample size of 10 individuals would produce unbiased results similar to those obtained
237 with a dataset of 1000 individuals divided into homozygotes.

238

239 *Datasets, tested hypotheses and statistical tests*

240 Multiple spruce bark beetle datasets used in the study are presented in Figure 2b and 2c.
241 Individuals were grouped according to the previously described genetic structuring into the
242 southern and northern clade ([50,56]; Figure 1, Table S1) creating three datasets: 1) all
243 individuals from southern and northern region (SN, 258 individuals), 2) only individuals from
244 the southern populations (S, 138 individuals), 3) only individuals from the northern populations
245 (N, 120 individuals). These datasets did not include Polish populations that are highly admixed.
246 Each of these datasets was analysed in two versions, differing in the included parts of the
247 genome: inversion regions were either included (SN, S, N) or excluded (SNcol, Scol, Ncol;
248 only included collinear genome).

249 To investigate the effects of selection in inverted homozygotes, all individuals were also
250 divided into datasets based on their inversion genotype, forming three datasets per inversion:
251 1) individuals carrying major inversion haplotype (e.g. MJAInv2); 2) individuals carrying
252 minor inversion haplotype (e.g. MNAInv2); 3) individuals carrying all possible inversion
253 genotypes (e.g. SNInv2). Polish individuals were included in this analysis to ensure sufficient
254 sample sizes, and because this analysis focused on differences between inversion genotypes
255 rather than population-level structure. Inversion datasets consisted of inversion in question and
256 collinear part of the genome (thus, excluding other inversions apart from the one in question).

257

258 *Datasets with and without inversions*

259 The analysis of the datasets with and without inversions (SN, S and N, and their respective
260 SNcol, Scol and Ncol, Figure 2b) allowed us to investigate whether all or some inversions
261 appear as selection outliers, to test whether the presence of multiple inversions affects the
262 detection of selection signals outside of inversions (in collinear part of the genome), and to
263 compare the behaviour of the two methods in these respects. Specifically, we predicted: 1) to
264 find more selection signals in inversions than in collinear parts due to the effect of reduced
265 recombination and inversions' genetic subdivision, 2) that balanced inversions with
266 intermediate frequency of both inversion haplotypes will show stronger selection signals than

267 inversions with one allele rarer (assuming that rare allele is younger and have low within
268 haplotype polymorphism), 3) that outlier detection outside inversions is affected by inversions
269 present within the genome due to their influence on haplotype frequency spectra (Λ) and the
270 assignment of the statistics values into the frequency bins (nSL). To test whether selection
271 signals are more often found in inversion regions than in collinear parts of the genome we
272 calculated the proportion of outliers in inversions and performed a permutation test for each of
273 the analysed dataset (1000 permutations, permuting SNP-based window labels - inverted and
274 collinear). To check the effect of inversion polymorphism frequency on selection scan
275 performance, we correlated the inversion haplotype frequency and the percentage of outliers
276 within the inversion. To test whether inversions affect selection inference in the collinear part
277 of the genome we compared datasets with and without inversions. We counted the number of
278 outlier windows and investigated whether the same regions were classified as outliers in both
279 datasets and whether new selection signals appear if we excluded inversions from the analysis.
280 We compared both methods by checking the number and distribution along the genome of the
281 outlier windows identified in both methods and across all datasets.

282 To be able to identify selection signals specific for the genetically differentiated groups
283 identified across the species range, selection scans described above were performed separately
284 for southern and northern regions/clades (S and N; Scol and Ncol; Figure 1, Figure 2b). To test
285 if selection signals correlate with Fst between the southern and northern group we correlated
286 Fst with nSL and Λ values (all calculated in SNP-based windows, significance of the
287 correlation was calculated using permutations).

288

289 *Datasets consisted of different inversion genotypes*

290 The inversion datasets (MJAInv, MNAInv and SNInv, Figure 2c) allowed us to test whether:
291 1) selection signals present in inversions when analysing dataset with all individuals (SNInv;
292 all possible inversion genotypes) disappear when the selection scans were performed using
293 only homozygous individuals (MJAInv or MNAInv); 2) There is a positive association between
294 inversion polymorphism and signal disappearance (more false positive selection signals
295 expected in high frequency of heterozygotes); 3) Selection signals disappear only in one
296 homozygotes revealing haplotype which carries alleles under selection. We expect that the
297 behaviour of the scans depends on the frequency of the inversion haplotype (produced by lower
298 recombination within haplotypes and higher genetic subdivision between them) and that the

most pronounced changes in selection signal occur when the inversion frequency is the highest (around 0.5). We also predict that in some inversions false positive signals would disappear, leaving only inversion haplotype-specific selection signals (if they exist).

In total, twenty-four inversions were analysed. In each of the inversion dataset all the other inversions (apart from the inversion in question) were excluded before performing selection scans. We calculated the percentage of outliers in each inversion in all datasets and in both of the selection scans. Since 1% of the selection scans results are always treated as outliers (even if the selection landscape is flat) we corrected the observed percentage of outliers in each inversion for the expected number of outliers given the inversion length and analysed genome size (given in number of SNP windows). Since several observed percentages equalled zero, we did not normalize them by simply dividing observed by expected values. Instead, we transformed the observed data as (observed percentage - expected percentage)/(observed percentage + expected percentage)/2. This produced values that were easy to interpret: positive values indicated regions enriched in the selection signal and negative values regions that were depleted in the selection signals. These transformed percentages were used in all the downstream tests.

To test for a difference between the observed and the expected percentage of outliers within inversions and to test whether inversions are enriched in (Observed – Expected > 0) or depleted (Observed – Expected < 0) from selection signals we used Wilcoxon signed-rank test. To test whether the magnitude of selection signals depends on the frequency of the inversion haplotype, we fitted logistic regression models for each dataset (SNInv, MJAInv, MNAInv). We transformed the percentage of outliers into binary categories—"Depleted" (negative values) and "Enriched" (positive values). These categories were used as response variables, and the frequency of the major allele and its square (to capture a potential nonlinear relationships) were included as explanatory variables. Finally, to test if selection signals disappear only in one homozygote (major or minor) revealing haplotype which carries loci under selection we classified selection signals into two signal categories: 1) selection signal gets much stronger in one of the homozygous dataset (percentage of outliers in one MJAInv (or MNAInv) dataset is higher than 50% of the percentage observed in the SNInv dataset (all genotypes) and the percentage of outliers in the other homozygote dataset is lower than 50% of the percentage observed in the SNInv dataset); 2) percentage of outliers is different from the first category. After categorization we used binomial regression model with signal behaviour

331 as response variable and frequency of the inversion major haplotype as explanatory variable.
332 The significance of the variables was tested using a Wald test.

333

334 *Candidate genes under selection and gene ontology enrichment analysis*

335 We identified genes harbouring or closest to outlier windows from the top 1% of nSL or Λ
336 values. We used gene annotations from Powell at. al (2021) [59] and tested for
337 overrepresentation of GO using R package topGO [66]. We applied Fisher's exact test and
338 "weight01" algorithm [67] to deal with the GO graph structure and only GO categories with at
339 least 10 members among the SNP associated genes were considered. GO analysis were
340 performed for datasets including: 1) selection outliers outside inversions and specific for
341 northern group; 2) selection outliers outside inversions and specific for southern group; 3)
342 selection outliers outside inversions common for both groups; 4) selection outliers specific for
343 particular inversion haplotype or present in both inversion haplotypes (the inversion signals
344 that we consider true signals).

345

346 **Results**

347 First, we used simulations to assess the performance of the nSL and Λ selection scans and to
348 establish a theoretical baseline for the patterns expected in neutral inversions across the
349 genome. Second, we used whole-genome sequence data from hundreds of individuals and
350 performed both selection scans across the inversion-rich spruce bark beetle genome. The list
351 of all datasets and their characteristics (e.g. whether they contained inversions or not) are given
352 in Figure 2 and the results are presented below.

353

354 *Selection landscape in neutral inversions - lessons from simulation-based datasets*

355 1. *False positive signals in neutral inversions*

356 Neutral inversions contained selection outliers in the majority of simulations (Figure 3,
357 Supplementary Information: Simulations Results .zip folder). This pattern was consistent
358 across datasets with 1000 and 200 individuals and for both selection scans (nSL and Λ, Table
359 1).

360 The two selection scan methods revealed qualitatively different patterns of selection signals.
361 When present, Λ outliers produced distinct and pronounced peaks, with no outliers detected in
362 the two linkage groups that evolved under neutrality in the simulations (Figure 3 and
363 Supplementary Information: Simulation Results .zip folder). In contrast, nSL scans typically
364 identified only a small number of outlier SNPs within inversions (Figure 3 and Supplementary
365 Information: Simulation Results .zip folder). This pattern was very similar to patterns observed
366 in neutral LGs (LG1 and LG2) where a small number of outliers were also detected. In most
367 simulations, the proportion of outliers in inversions and neutral linkage groups did not differ
368 significantly (χ^2 test, $p > 0.05$ after multiple testing correction). In the dataset with 1000
369 individuals, 33 simulations showed significantly more outliers in inversions compared to
370 neutral linkage groups, while in the dataset with 200 individuals, this was the case for 31
371 simulations.

372 *2. What happens to section signals if we divide simulations into homozygotes datasets?*

373 A striking difference emerged between the two selection scan methods in how selection signals
374 changed when datasets were divided into major and minor inversion homozygotes and
375 compared to the simulations that showed selection signal in datasets that included all inversion
376 genotypes. In Λ scans, nearly all previously detected signals disappeared in the homozygote
377 subsets (Figure 3 and Supplementary Information: Simulations Results .zip folder, Table 1). In
378 contrast, nSL scans showed a more stochastic pattern of signal loss and appearance (Table 1).
379 As in the previous analyses, only a small number of simulations showed a significant excess
380 of outliers in inversions compared to neutral linkage groups (χ^2 test, $p < 0.05$ after multiple
381 testing correction), with between 2 and 10 out of 181 simulations showing a significant
382 difference.

383 3. *The effect of dataset sample size on the detection of selection signals*

384 We found no significant difference in the proportion of outliers in inversions between datasets
385 of 200 and 1000 individuals for either nSL or Λ selection scans (GLMM, $p > 0.05$). In these
386 comparisons, sample sizes were fixed at 200 and 1000. When datasets were divided into
387 homozygote subsets, the sample size depended on inversion frequency, and even in the 1000
388 individual dataset, minor haplotype subsets could be as small as 10 individuals (3 simulations
389 were excluded since had only 1 or 2 homozygotes; in total there were 38 simulations had fewer
390 than 100 individuals). For nSL scans, no significant difference was observed between major

391 inversion haplotype datasets from 1000 individuals (smallest n = 208) and those with only 10
392 individuals. For minor inversion homozygotes, slightly fewer outliers were detected in the
393 smallest datasets (n = 10) compared to larger subsets (GLMM, p = 0.01).

394

395 *Selection landscape in inversion-rich genome – lessons from empirical datasets*

396 1. *Inversion regions are enriched in selection signals*

397 Most of the selection outliers were identified within inversion regions (Figure 4, Figure S1,
398 Table 2), many more than expected by chance (permutation tests; $p < 0.001$; Supplementary
399 Information 1.1). This pattern was consistent across both statistics, window sizes and all
400 datasets (SN, S, N), except for Λ scan and SN dataset and 52 SNPs window size, where more
401 outliers than expected were found in the collinear part of the genome (Table 2). In the latter
402 case the signal was mainly driven by one region in IpsContig3 (39% of all outliers and 55% of
403 collinear outliers in IpsContig3). Only 7 inversions (out of 24) did not show any selection
404 signals in nSL or Λ analysis (Table S2). Inversions not only included more outliers in total but
405 also harboured longer selection signal clusters (in particular in Λ scans; permutation test, $p <$
406 0.001; Table S3; Supplementary Information 1.8). This was also true for the Λ scan with 52
407 SNPs window size that had a smaller percentage of outliers in inversions compared to the
408 collinear part of the genome, even considering that inversions represent only ~30% of the
409 genome (Figure 4, Table S3). The largest inversion outlier clusters were found in inversion
410 Inv16 (0.324 Mb, 3% of the inversion, in S dataset) and Inv5 (0.253 Mb, 6% of the inversion,
411 in N dataset, Figure 4, Table S3). The largest cluster outside inversion was found in IpsContig3
412 (ranging from 0.317 to 0.369 Mb in different datasets). We did not find a significant correlation
413 between frequency of inversions and strength of selection signals (Supplementary Information
414 1.3). In summary, we found that inversions are enriched in selection signals and that these
415 signals form large blocks of outlier regions.

416 2. *Selection landscape identified in collinear parts is minimally affected by inversions*
417 *present in the genome*

418 Both nSL and Λ scans conducted using the dataset with inversions, revealed multiple signals
419 of selection outside inversion regions (Table 2-3, Figure 4-5, Figure S1-S2). The proportion of
420 outliers detected depended on the dataset and selection scan method used. Figure 5 and Table
421 4 compare the selection landscape in the collinear part of the genome in datasets that included

422 inversions when the selection scans were performed (SN, S, N) and datasets in which
423 inversions were excluded before selection scan analysis (SNcol, Ncol, Scol). There was a
424 significant correlation between the results obtained in the collinear part of the genome using
425 datasets that included and excluded inversions (r^2 ranged from 0.76-0.95, permutation test; p-
426 value < 0.001; Supplementary Information 1.6). In summary, the general selection landscape
427 in collinear genome was almost the same between datasets that included and excluded
428 inversions indicating minimal influence of inversions on the selection scan results in the
429 collinear part of the genome (Figure 5).

430 *3. Both methods identify a similar selection landscape*

431 Both selection scan methods produced a fairly similar selection landscape across all datasets
432 (r^2 ranged from 0.09 to 0.21, permutation test; $p < 0.001$; Supplementary Information 1.4).
433 Only the selection landscape obtained from Λ scans with window size 117 SNPs was visually
434 different and dominated by two large outlier clusters in inversions Inv5 and Inv7 (Figure S1,
435 Table S4). Despite this difference, many regions identified in the analysis that used 52 SNPs
436 window size (both nSL and Λ) were still visible as distinct peaks when using window size of
437 117 SNPs but were not classified as outliers in this dataset due to the extreme values of the two
438 inversions mentioned above. In general, the genome-wide selection landscapes were similar
439 across different statistics (as indicated by a high correlation of values in 50kb widows), but the
440 actual overlap between outlier SNP-based windows in nSL and Λ was rather small due to the
441 difference in values for particular windows - the same regions appear as outliers but particular
442 small SNP windows can differ in the magnitude of selection signals (range from 11% to 20%
443 of shared outliers in Λ and nSL for 52 SNPs window size in different datasets; Table 1), but
444 this is still more overlap then expected by chance (permutation test, $p < 0.001$; Supplementary
445 Information 1.4). In summary, both methods identify similar selection landscapes despite the
446 observed difference in magnitude of selection signal between nSL and Λ.

447 *4. Northern and southern regions have similar selection landscapes with few region-*
448 *specific selection signals*

449 Selection landscapes inferred for S and N datasets were highly correlated with each other (r^2
450 equalled 0.58 for Λ, 0.66 for nSL; $p < 0.001$; Figure 4, Figures S1–S2, Tables 2-3,
451 Supplementary Information 1.5). This overall similarity was consistent across both nSL and Λ
452 and reflected the largely shared genomic landscape of selection between geographic regions
453 (including similar impact of inversions on selection landscape in both regions). However, a

454 region-specific outlier windows were also detected (Figure S3, Table 3). In addition, both nSL
455 and Λ values were very weakly positively correlated with Fst values calculated across the
456 genome and between the N and S datasets (r^2 ranged from 0.0004 to 0.0634 across datasets; p
457 < 0.001; Supplementary Information 1.2).

458

459 *Selection scans in inversion homozygotes - detecting false positive and haplotype-specific*
460 *signals*

461 5. *Selection scans in homozygote datasets reveal haplotype-specific and false positive*
462 *signals*

463 nSL selection signal was detected in ten inversions in both homozygotes, eight inversions in
464 only one homozygote (four of which were minor alleles), and in three inversions selection
465 signals disappeared completely in the homozygous datasets (Table S5). In the Λ scans, three
466 inversions showed selection signals in both homozygotes, and five inversions showed signals
467 in only one homozygote. None of the inversions lost their selection signals in both
468 homozygotes. Figure 6 shows the large variation in the difference between the observed and
469 the expected percentages of outliers among inversions and datasets. For example, Inv2 and
470 Inv16 show a much higher proportion of observed outliers in SNInv and one of the homozygote
471 datasets — dozens observed compared to only a few percent expected. In contrast, Inv13
472 exhibits similar observed and expected percentages in both SNInv and MJNInv, and notably
473 fewer observed outliers than expected in MJAInv (Figure 6a). Overall, analysis of homozygote
474 datasets revealed that the majority of inversions harbour selection signals in either both or one
475 inversion haplotype. However, in few cases selection signals in homozygote datasets
476 disappeared completely indicating false positive signals obtained in SNInv datasets.

477 6. *Selection signal tends to be reduced in homozygote datasets*

478 While analyses using SN dataset (all individuals and all inversions), showed that inversion
479 regions are overall enriched for selection signals compared to collinear regions, we observed
480 that the strength of these signals tends to be reduced in homozygote datasets. Our tests indicated
481 a depletion of the signal in homozygotes (MJAInv and MNAInv datasets) compared to the
482 SVInv dataset (Wilcoxon tests, $p < 0.02$ for both nSL and Λ scans; see
483 Supplementary_Information_Selection_model.html). However, the magnitude of the selection
484 signal in inversion regions did not correlate with inversion haplotype frequency. Similarly,

485 inversion haplotype frequency did not explain the magnitude of selection signals within
486 inversions or indicate any clear patterns regarding what happens to selection signals in
487 inversions (the disappearance of the signal is not dependent on inversion frequency; see
488 Supplementary_Information_Selection_model.html).

489 In summary, although selection signal tends to be reduced in homozygous datasets compared
490 to SNInv, there is a considerable heterogeneity in this respect among inversions and among
491 datasets (Figure 6, Figure S4, Table S5). The strength of the selection signals and its behaviour
492 in different datasets did not depend on the frequency of inversions.

493

494 *Gene ontology terms enriched in selection targets*

495 To explore the functions associated with regions under selection, we performed Gene Ontology
496 (GO) enrichment analysis using outlier windows that (1) were located outside inversions and
497 specific to N population, (2) were located outside inversions and specific to S population, (3)
498 were shared between S, N, and SN datasets, or (4) remained significant after excluding
499 heterozygous individuals in the analysis performed with only homozygotes. In total, we
500 identified 24 significantly enriched GO terms (Table S6). Enriched functions included terms
501 such as "signalling receptor activity" and "response to stimulus," potentially related to odorant
502 receptor function. A complete list of enriched GO terms, including fold enrichment and
503 significance levels, is provided in Table S6.

504

505 **Discussion**

506 In this study we evaluate performance of two selection scan methods using simulated datasets
507 and used whole-genome resequencing data from 312 individuals of the spruce bark beetle to
508 investigate how selection signals are distributed throughout its genome and how chromosomal
509 inversions influence these patterns and our ability to detect them.

510 Selection scan results based on the simulated datasets indicated that the two evaluated statistics
511 differed in their sensitivity to suppressed recombination in the inversion regions. As expected,
512 Λ was highly sensitive to low recombination [12] and produced distinct outliers in the majority
513 of simulation-based selection scans (Figure 3 and Supplementary Materials: Simulation Results
514 .zip folder). In contrast, nSL, consistent with its known robustness to recombination variation

515 [27], yielded much less pronounced outlier peaks in inversions, though a low level of
516 background noise remained, with some outliers detected even in neutrally evolving regions.
517 Importantly, these nSL outliers did not form visible clusters and usually appeared as single-
518 locus signals. Additionally, simulations showed that analysing homozygote-only datasets can
519 substantially reduce false-positive signals indicating that this approach can be valuable in study
520 systems where inversion genotyping is possible.

521 As predicted, we observed a significant enrichment of selection signals within inversion
522 regions using both nSL and Λ statistics in the spruce bark beetle data. Notably, even in the
523 homozygotes, most inversions exhibit some selection signals in at least one homozygote
524 dataset. These findings reinforce the view that most inversions contribute to adaptation [42,68-
525 69] and are unlikely to be selectively neutral [31], particularly in the case of large polymorphic
526 inversions [36] — the kind identified in the spruce bark beetle genome [50].

527 On the other hand, we observed substantial variation in both the magnitude and behaviour
528 (enrichment or depletion) of selection signals across inversions and datasets, in particular when
529 all genotypes were contrasted with homozygotes. In general, in the concordance with
530 simulation-based results, selection signals were weaker in the homozygote datasets, suggesting
531 that some signals detected in the entire dataset may be significantly inflated by the presence of
532 heterozygous individuals. By restricting analyses to individuals carrying only one inversion
533 haplotype, we reduced the likelihood of false positives and were able to detect potential
534 haplotype-specific selection signals. However, the small sample size of some homozygote
535 datasets (directly reflecting inversion frequency) could limit the ability to detect selection
536 signals. Our simulation-based tests indicated that small sample sizes may lead to
537 underestimation of outliers detected in inversions. Overall, the lack of a clear relationship
538 between the extent of inversion polymorphism and the behaviour of selection signals across
539 the spruce bark beetle datasets underscores a key challenge: although inversions are enriched
540 for selection signals and homozygote datasets can help in reducing false positive signals,
541 identifying the specific targets of selection within these regions remains difficult.

542 Contrary to our expectations, we found that the presence of inversions had only a minimal
543 impact on the shape of the selection landscape and our ability of detect selection signals in the
544 collinear genome. Importantly, this funding suggests that identifying loci under selection
545 outside of inversions remains feasible, even in genomes characterized by extensive inversion
546 polymorphism. This result aligns with our earlier findings on demographic inference [56],

547 where we showed that overlooking chromosomal inversions had no effect on model selection
548 and introduced only minor biases in the estimation of demographic parameters.

549 Additionally, the spruce bark beetle selection scans supported the necessity of using multiple
550 complementary approaches to selection scans. While the overall patterns of selection inferred
551 by Λ and nSL were similar, differences were observed at a local scale. There was a high
552 correlation of selection signals at 50 kb windows, but only a limited overlap among outlier
553 windows at 52 SNPs, which demonstrates the general difficulty of selection scans in
554 pinpointing the focal gene under selection [70]. Moreover, the selection of window sizes in the
555 Λ scans can influence which regions are considered candidates for selection. The differences
556 observed in the spruce bark beetle between 52 and 117 SNP windows can, to some extent,
557 reflect differences in the ability to detect more recent and more ancient sweep signatures:
558 smaller window sizes are more effective at detecting ancient sweeps [12].

559 Our study revealed a broadly similar selection landscape between the southern and northern
560 populations of the spruce bark beetle, suggesting that these populations are likely subject to
561 comparable selective pressures and/or that their divergence is relatively recent, with the
562 observed signals possibly reflecting selection in a shared ancestral population. According to
563 our recent demographic modelling, the divergence between the southern and northern clades
564 occurred approximately 80,000 years ago [56]. Signals of selection in the site frequency
565 spectrum are expected to persist for roughly 0.1Ne generations, while haplotype-based signals
566 fade more rapidly, typically within ~0.01Ne generations [71-72]. Assuming an ancestral Ne of
567 250,000 [56], haplotype-based signatures could persist for around 2,500 generations—well
568 after the estimated divergence time—supporting the idea that shared selection pressures may
569 be acting across both populations. The inferred Ne for the northern and southern populations
570 after the split was 180,000 and 830,000, respectively, and even under the higher Ne scenario,
571 the timescale of detectable selection signals remains post-divergence. These findings suggest
572 that spruce bark beetle populations across Europe may be subject to similar selective pressure.
573 Nonetheless, we also detected evidence of regional differentiation: one of the strongest
574 selection signals outside of inversions (on IpsContig3) overlapped with a pronounced Fst peak
575 between the northern and southern groups, indicating that divergent selection may also
576 contribute to shaping the selection landscape in this species.

577 Finally, our GO enrichment analysis identified terms such as "signalling receptor activity" and
578 "response to stimulus," which are potentially linked to odorant receptor function. These

579 findings are of particular interest considering recent studies showing that inversions are
580 enriched in genes encoding odorant receptors [50], and that these inversions may contain
581 receptor variants that respond differently to different components of pheromones [73].

582

583 **Conclusions and future directions**

584 Our study provides new insights into the selection landscape across the spruce bark beetle
585 genome, identifying candidate genes that may guide future research into the genomic basis of
586 adaptation in this ecologically and economically important species. We show that most
587 polymorphic inversions in the genome are enriched for selection signals, supporting their role
588 as key contributors to adaptation, while also highlighting the importance of accounting for
589 structural variation — particularly inversions — when performing selection inference. As
590 illustrated by the changing selection patterns in the homozygote datasets in simulations as well
591 as empirical datasets, inversions can affect the detection of selection within their regions.
592 Nevertheless, their presence does not obscure the identification of selection signals outside
593 inversions, emphasizing the feasibility of conducting genome-wide selection scans even in
594 inversion-rich genomes. More broadly, our results underscore the need to systematically
595 evaluate the performance and robustness of evolutionary inference methods in the presence of
596 structural variation [15,28]. However, such efforts remain rare, and further research is needed
597 to understand how inversions influence evolutionary inference, particularly given that their
598 effects likely vary across species with differing numbers and evolutionary histories of large
599 polymorphic inversions.

600

601 **Acknowledgments**

602 We thank members of the Genomics and Experimental Evolution Group at Jagiellonian
603 University for their help in improving this manuscript. This work was funded by a Polish
604 National Science Center 2018/30/E/NZ8/00105 grant to K.N.B. We gratefully acknowledge
605 Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet
606 AGH) for providing computer facilities and support within computational grant no.
607 PLG/2024/017278 to K.N.B.

608

Data Availability

All DNA sequences have been deposited in the National Centre for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA1013983. Scripts used for data processing, generation of intermediate files, and analyses, have been archived and are publicly available in the Dryad Digital Repository entry [will be added as soon as possible].

Author contributions

J.M.G., K.N.B. conceived the study and wrote the first draft of the manuscript. J.M.G, performed all the main analyses. P.Z processed and supervised the raw data processing. W.B. performed the GO analysis. W.B. and P.Z helped in data interpretation and provided feedback on all manuscript versions. All authors read and approved the final manuscript.

622 **References**

623 1. Bomblies K, Peichel CL. Genetics of adaptation. Proc Natl Acad Sci U S A.
624 2022;119(30):1-0. https://doi.org/10.1073/pnas.2122152119.
625 2. Barghi N, Hermisson J, Schlötterer C. Polygenic adaptation: a unifying framework to
626 understand positive selection. Nat Rev Genet. 2020;21(12):769-781.
627 https://doi.org/10.1038/s41576-020-0250-z
628 3. Barrett R, Schluter D. Adaptation from standing genetic variation. Trends Ecol Evol. 2008;
629 23(1):38–44. https://doi.org/10.1016/j.tree.2007.09.008.
630 4. Hermisson J, Pennings PS. Soft sweeps and beyond: understanding the patterns and
631 probabilities of selection footprints under rapid adaptation. Methods Ecol Evol. 2017;8(6),
632 700-716. https://doi.org/10.1111/2041-210X.12808
633 5. Höllinger I, Pennings PS, Hermisson J. Polygenic adaptation: from sweeps to subtle
634 frequency shifts. PLoS Genet. 2019;15(3):1-26.
635 https://doi.org/10.1371/journal.pgen.1008035.
636 6. Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold
637 does not glitter. Evol. 2012;66(1):1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x.
638 7. Sella G, Barton NH. Thinking about the evolution of complex traits in the era of genome-
639 wide association studies. Annu Rev Genomics Hum Genet. 2019;20(1):461-493.
640 https://doi.org/10.1146/annurev-genom-083115-022316.
641 8. Stephan W. Selective sweeps. Genetics. 2019;211(1):5–13.
642 https://doi.org/10.1534/genetics.118.301319
643 9. Enbody ED, Sendell-Price AT, Sprehn CG, Rubin CJ, Visscher PM, Grant BR, et al.
644 Community-wide genome sequencing reveals 30 years of Darwin's finch evolution. Science.
645 2023;381(6665):1-9. https://doi.org/10.1126/science.adf6218
646 10. Roberts Kingman GA, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, et al. Predicting

647 future from past: the genomic basis of recurrent and rapid stickleback evolution. Sci. Adv.
648 2012;7(25):1-12. https://doi.org/10.1126/sciadv.abg5285
649 11. Cheng X, DeGiorgio M. BalLeRMix +: mixture model approaches for robust joint
650 identification of both positive selection and long-term balancing selection. Bioinform.
651 2022;38(3):861–63. https://doi.org/10.1093/bioinformatics/btab720.
652 12. DeGiorgio M, Szpiech ZA. A spatially aware likelihood test to detect sweeps from
653 haplotype distributions. PLoS Genet. 2022;18(4):1-37
654 https://doi.org/10.1371/journal.pgen.1010134.
655 13. Kern AD, Schrider DR. diploS/HIC: an updated approach to classifying selective sweeps.
656 G3. 2018;8(6):1959–1970. https://doi.org/10.1534/g3.118.200262.
657 14. O'Reilly PF, Birney E, Balding DJ. Counfounding between recombination and selection,
658 and the Ped/Pop method for detecting selection. Genome Res. 2024;18:1304-1313.
659 https://doi.org/ 10.1101/gr.067181.107
660 15. Lottheros KE. The effect of neutral recombination variation on genome scans for
661 selection. G3. 2019;9(6):1851–1867. https://doi.org/10.1534/g3.119.400088.
662 16. Schaal SM, Haller BC, Lottheros KE. Inversion invasions: when the genetic basis of local
663 adaptation is concentrated within inversions in the face of gene flow. Philos Trans of R Soc
664 Lond B Biol Sci. 2022;377(1856):1-14. https://doi.org/10.1098/rstb.2021.0200
665 17. Adrion JR, Galloway JG, Kern AD. Predicting the landscape of recombination using deep
666 learning. Mol Biol Evol.2020;37(6):1790–808. https://doi.org/10.1093/molbev/msaa038
667 18. Crown KN, Miller DE, Sekelsky J, Hawley RS. Local inversion heterozygosity alters
668 recombination throughout the genome. Curr Biol. 2018;28(18):2984-2990.
669 https://doi.org/10.1016/j.cub.2018.07.004
670 19. Koury SA. Predicting recombination suppression outside chromosomal inversions in
671 Drosophila melanogaster using crossover interference theory. Heredity. 2023;130(4):196-

672 208. https://doi.org/10.1038/s41437-023-00593-x
673 20. Li H, Berent E, Hadjipanteli S, Galey M, Muhammad-Lahbabi N, Miller DE, et al.
674 Heterozygous inversion breakpoints suppress meiotic crossovers by altering recombination
675 repair outcomes. PLoS Genet. 2023;19(4):e1010702.
676 https://doi.org/10.1371/journal.pgen.1010702
677 21. Seich Al Basatena NK, Hoggart CJ, Coin LJ, O'Reilly PF. The effect of genomic
678 inversions on estimation of population genetic parameters from SNP data. Genetics.
679 2013;193(1):243-253. https://doi.org/10.1534/genetics.112.145599
680 22. Ishii K, Charlesworth B. Associations between allozyme loci and gene arrangements due
681 to hitch-hiking effects of new inversions. Genetics Res. 1977;30(2):93-106.
682 doi:10.1017/S0016672300017511
683 23. Navarro A, Betrán E, Barbadilla A, Ruiz A. Recombination and gene flux caused by gene
684 conversion and crossing over in inversion heterokaryotypes. Genetics. 1997;146(2):695-709.
685 https://doi.org/10.1093/genetics/146.2.695
686 24. Navarro A, Barbadilla A, Ruiz A. Effect of inversion polymorphism on the neutral
687 nucleotide variability of linked chromosomal regions in
688 Drosophila. Genetics. 2000;155(2):685-698. https://doi.org/10.1093/genetics/155.2.685
689 25. Guerrero RF, Rousset F, Kirkpatrick M. Coalescent patterns for chromosomal inversions
690 in divergent populations. Philos Trans of R Soc Lond B Biol Sci. 2012;367(1587):430–438.
691 https://doi.org/10.1098/rstb.2011.0246.
692 26. Korunes KL, Noor MA. Pervasive gene conversion in chromosomal inversion
693 heterozygotes. Mol Ecol. 2019;28(6):1302-1315. https://doi.org/10.1111/mec.14921
694 27. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or
695 hard selective sweeps using haplotype structure. Mol Biol Evol. 2014; 31(5):1275–1291.
696 https://doi.org/10.1093/molbev/msu077

697 28. Novo I, Ordás P, Moraga N, Santiago E, Quesada H, Caballero A. Impact of population
698 structure in the estimation of recent historical effective population size by the software
699 GONE. Genet Sel Evol. 2023;55(1):1-15. https://doi.org/10.1186/s12711-023-00859-2.
700 29. Charlesworth B. The effects of inversion polymorphisms on patterns of neutral genetic
701 diversity. Genetics. 2023;224(4):1-21. https://doi.org/10.1093/genetics/iyad116
702 30. Charlesworth B, Barton NH. The spread of an inversion with migration and
703 selection. Genetics. 2018;208(1):377-382. https://doi.org/10.1534/genetics.117.300426
704 31. Berdan EL, Barton NH, Butlin R, Charlesworth B, Faria R, Fragata I, et al. How
705 chromosomal inversions reorient the evolutionary process. J Evol Biol. 2023;36(12):1761–
706 1782. https://doi.org/10.1111/jeb.14242
707 32. Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N. Ancestral polymorphisms
708 explain the role of chromosomal inversions in speciation. PLoS Genet. 2018;14(7):1-26.
709 https://doi.org/10.1371/journal.pgen.1007526
710 33. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation.
711 Genetics. 2006;173(1):419–434. https://doi.org/10.1534/genetics.105.047985.
712 34. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci.
713 Proc Natl Acad Sci U S A. 2013;110(19):1743-1751.
714 https://doi.org/10.1073/pnas.1219381110.
715 35. Yeaman S, Aeschbacher S, Bürger R. The evolution of genomic islands by increased
716 extablishment probability of linked alleles. Mol Ecol. 2016;25(11): 2542-2558.
717 https://doi.org/10.1111/mec.13611
718 36. Connallon T, Olito C. Natural selection and the distribution of chromosomal inversion
719 lengths. Mol Ecol. 2022;31(13):3627–3641. https://doi.org/10.1111/mec.16091.
720 37. Charlesworth B, Charlesworth D. Selection of new inversions in multi-locus genetic
721 systems. Genetics Res. 1973;21(2):167-183. doi:10.1017/S0016672300013343

722 38. Ohta T. Associative overdominance caused by linked detrimental mutations. Genetics
723 Res. 1971;18(3):277-286. doi:10.1017/S0016672300012684
724 39. Faria R, Johannesson K, Butlin RK, Westram AM. Evolving inversions. Trends Ecol
725 Evol. 2019;34(3):239–48. https://doi.org/10.1016/j.tree.2018.12.005
726 40. Harringmeyer OS, Hoekstra HE. Chromosomal inversion polymorphisms shape the
727 genomic landscape of deer mice. Nat Ecol Evol. 2022;6(12):1965–1979.
728 https://doi.org/10.1038/s41559-022-01890-0.
729 41. Kapun M, Flat T. The adaptive significance of chromosomal inversion polymorphisms in
730 Drosophila melanogaster. Mol Ecol. 2019;28(6):1263-1282.
731 https://doi.org/10.1111/mec.14871
732 42. Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA. Genomic evidence for role of
733 inversion 3RP of Drosophila Melanogaster in facilitating climate change adaptation. Mol
734 Ecol. 2015;24(10):2423–2432. https://doi.org/10.1111/mec.13161
735 43. Thompson MJ, Jiggins CD. Supergenes and their role in evolution. Heredity.
736 2014;113(1):1–8. https://doi.org/10.1038/hdy.2014.20.
737 44. Coughlan JM, Willis JH. Dissecting the role of a large chromosomal inversion in life
738 history divergence throughout the Mimulus guttatus species complex. Mol
739 Ecol. 2019;28(6):1343-1357. https://doi.org/10.1111/mec.14804
740 45. Wallberg A, Schoening C, Webster MT, Hasselmann M. Two extended haplotype blocks
741 are associated with adaptation to high altitude habitats in East African honeybees. PLoS
742 Genet. 2017:13(5):e1006792. https://doi.org/10.1371/journal.pgen.1006792
743 46. Sinclair-Waters M, Bradbury IR, Morris CJ, Lien S, Kent MP, Bentzen P. Ancient
744 chromosomal rearrangement associated with local adaptation of a postglacially colonized
745 population of Atlantic Cod in the northwest Atlantic. Mol Ecol. 2018;27(2):339-351.
746 https://doi.org/10.1111/mec.14442

47. Reeve J, Butlin RK, Koch EL, Stankowski S, Faria R. Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (Littorina Saxatilis and L. arcana). Mol Ecol. 2024;33(24):1-18. https://doi.org/10.1111/mec.17160

48. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20(4):208-215. doi: 10.1016/j.cub.2009.11.055

49. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. BMC Genomics. 2014;15(1):246. https://doi.org/10.1186/1471-2164-15-246

50. Mykhailenko A, Zieliński P, Bednarz A, Schlyter F, Andersson MN, Antunes B, et al. Complex genomic landscape of inversion polymorphism in Europe's most destructive forest pest. Genome Biol Evol. 2024;16(12):1-23. https://doi.org/10.1093/gbe/evae263.

51. Hlásny T, König L, Krokene P, Lindner M, Montagné-Huck C, Müller J, et al. Bark beetle outbreaks in Europe: state of knowledge and ways forward for management. Curr For Rep. 2021;7(3):138–165. https://doi.org/10.1007/s40725-021-00142-x

52. Seidl R, Müller J, Hothorn T, Bässler C, Heurich M, Kautz M. Small beetle, large-scale drivers: how regional and landscape factors affect outbreaks of the European spruce bark beetle. J Appl Ecol. 2016;53(2):530-540. https://doi.org/10.1111/1365-2664.12540

53. Dobor L, Baldo M, Bílek L, Barka I, Máliš F, Štěpánek P, et al. The interacting effect of climate change and herbivory can trigger large-scale transformations of European temperate forests. Glob Chang Biol. 2024;30(2):1-20. https://doi.org/10.1111/gcb.17194

54. Gohli J, Krokene P, Heggem ESF, Øklandet B al. Climatic and management-related drivers of endemic European spruce bark beetle populations in Boreal forests. J Appl Ecol. 2024;61(4):809–820. https://doi.org/10.1111/1365-2664.14606

55. Kautz M, Feurer J, Adler P. Early detection of bark beetle (Ips typographus) infestations

by remote sensing – a critical review of recent research. For Ecol Manage. 2024;556:1-17. https://doi.org/10.1016/j.foreco.2023.121595.

56. Zieliński P, Morales-García J, Schebeck M, Duduman ML, Nadachowska-Brzyska K. Demographic history inferred from an inversion-rich spruce bark beetle genome. bioRxiv. 2025. https://doi.org/10.1101/2025.07.17.665273

57. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: https://www.bioinformatics.babraham.ac. uk/projects/fastqc/.

58. Bolger AM. Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform. 2014; 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

59. Powell D, Groβe-Wilde E, Krokene P, Roy A, Chakraborty A, Löfstedt C, et al. A highly contiguous genome assembly of the Eurasian spruce bark beetle, Ips typographus, provides insight into a major forest pest. Commun Biol. 2021;4(1):1-9. https://doi.org/10.1038/s42003-021-02602-3.

60. Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359. https://doi.org/10.1038/nmeth.1923.

61. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108(10):1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

62. Szpiech ZA. Selscan 2.0: scanning for sweeps in unphased data. Bioinform. 2024;40(1):1-6. https://doi.org/10.1093/bioinformatics/btae006.

63. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The Variant Call Format and VCFtools'. Bioinform. 2011;27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

64. Haller BC, Messer PW. SLiM 2: flexible, interactive forward genetic simulations. Mol Biol Evol. 2017;34(1), 230-240. doi:10.1093/molbev/msw211

797 65. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al.
798 glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized
799 Linear Mixed Modeling. R J. 2017;9(2):378–400. doi:10.32614/RJ-2017-066.
800 66. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene
801 expression data by decorrelating GO graph structure. Bioinform. 2006;22(13):1600–16007.
802 https://doi.org/10.1093/bioinformatics/btl140
803 67. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. Bioconductor
804 Improv. 2009;27(1-26): 776.
805 68. Christmas MJ, Wallberg A, Bunikis I, Olsson A, Wallerman O, Webster MT.
806 Chromosomal inversions associated with environmental adaptation in honeybees. Mol Ecol.
807 2019;28(6):1358–1374. https://doi.org/10.1111/mec.14944
808 69. Pegueroles C, Ordóñez V, Pascual M. Recombination and selection in the maintenance of
809 the adaptive value of inversions. J Evol Biol. 2010;23(12):2709-2717.
810 https://doi.org/10.1111/j.1420-9101.2010.02136.x
811 70. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev
812 Genet. 2013;47(1):97-120. https://doi.org/10.1146/annurev-genet-111212-133526.
813 71. Pennings PS, Hermisson J. Soft sweeps II—molecular population genetics of adaptation
814 from recurrent mutation or migration. Mol Biol Evol 2006;23(5):1076-1084.
815 https://doi.org/10.1093/molbev/msj117.
816 72. Przeworski M. The signature of positive selection at randomly chosen loci. Genetics.
817 2002;160(3):1179–1189. https://doi.org/10.1093/genetics/160.3.1179.
818 73. Johny J, Diallo S, Nadachowska-Brzyska K, Moliterno AAC, Roy A, Kalinová B, et al.
819 Not all bark beetles smell the same: population-level functional olfactory polymorphisms in
820 Ips Typographus pheromone receptor ItypOR33. Mol Ecol. 2025;34(6):1-19.
821 https://doi.org/10.1111/mec.17693.

Figures

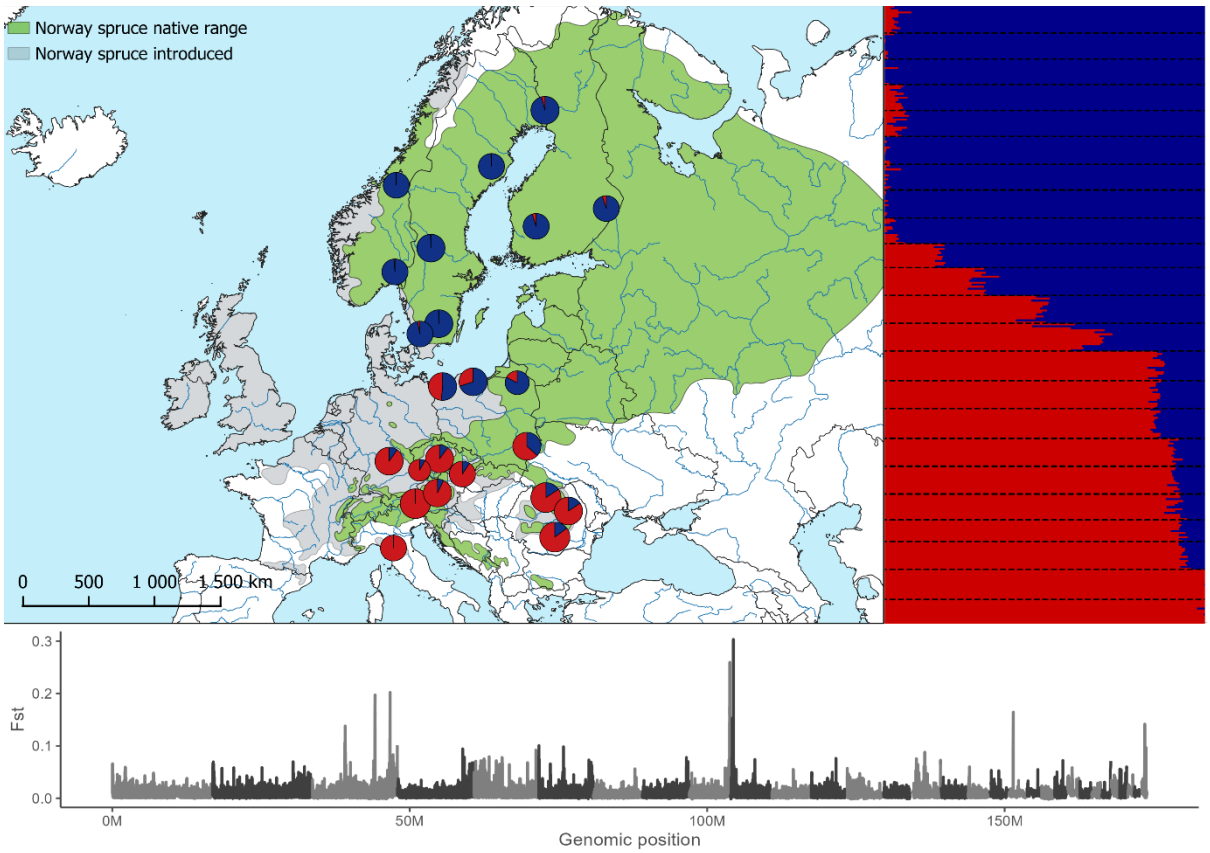

Figure 1. The geographic distribution, admixture, and genetic differentiation of *Ips typographus* populations as inferred in Zielinski et al. 2025. Top left: map of sampled populations, with admixture proportions illustrated by coloured circles. The colours represent the admixture proportions that were identified (top right). Admixture plot showing individuals admixture proportions for K=2. The populations are separated by dashed lines and ordered according to longitude. Bottom plot: pairwise Fst calculated between the two genetic clusters across the genome. Grey and black represent different contigs.

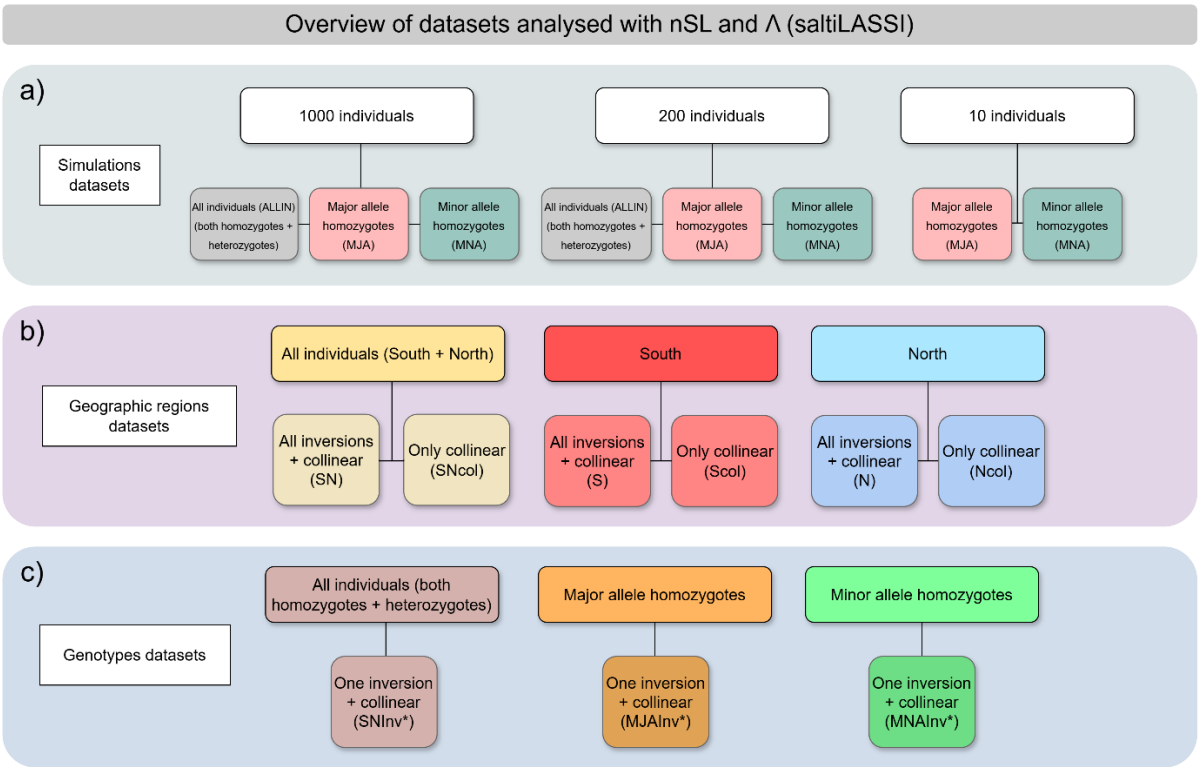

Figure 2. Graphical summary of the datasets analysed with nSL and Λ selection scans. A: Simulation datasets prepared based on the publicly available simulation results from Lotterhos (2019). The datasets differ in samples size (1000, 200, and 10 individuals) and included all individuals (all possible inversion genotypes, ALLIN, grey), or only individuals homozygous for a particular inversion allele (major allele; MJA, pink and minor allele, MNA, turquoise). B: Geographic region datasets: datasets generated from the spruce bark beetle genomic data and including southern populations (S, red), northern populations (N, blue) or both combined (SN, yellow). Each group was analysed including all genomic regions (inversions + collinear) or restricted to analysing collinear regions only (Scol, Ncol and SNcol). C: Genotype-based datasets: generated from the spruce bark beetle genomic data and analysed either including all individuals, regardless of their inversion genotype (both homozygotes and heterozygotes; SNInv*, brown), or including only individuals homozygous for a particular inversion allele, either for the major (MJAInv*, red) or minor (MNAInv*, turquoise) allele. Each inversion was examined separately, and generated dataset consisted of the inversion in question and collinear part of the genome. Inv* indicates 24 inversions examined.

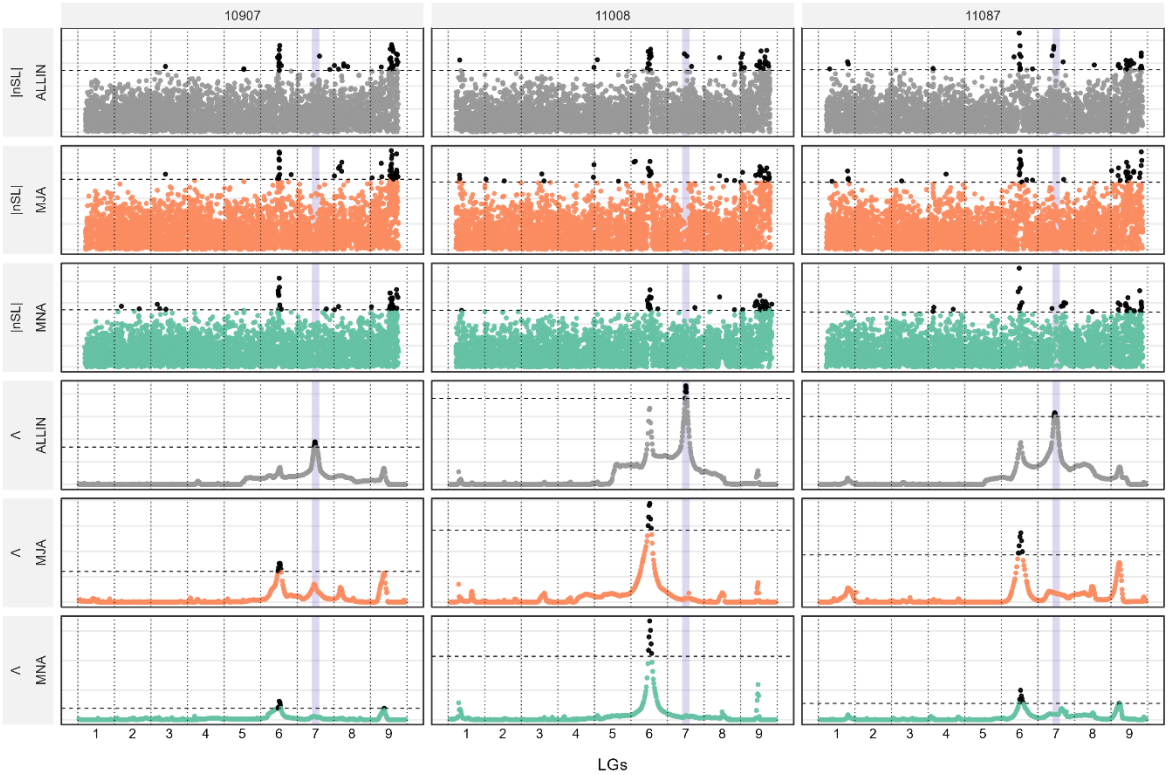

Figure 3. Comparison of selection scan results in three example simulations. Each column represents an independent simulation (numbered according to Lotterhos 2019), and each row corresponds to a dataset and selection statistic. The rows show either the nSL or the Λ statistic, which were computed for three datasets based on the inversion genotype of individuals: ALLIN (both homozygotes and heterozygotes, grey); MJA (major allele homozygotes, orange); and MNA (minor allele homozygotes, green). The dashed horizontal line indicates the 99th percentile threshold and the black dots above this line represent selection outliers. The vertical dotted lines mark the boundaries between linkage groups (LGs), which are numbered along the x-axis. The purple shaded region highlights the inversion region. LG1 and LG2 evolve neutrally and the selection signals within them were compared to selection signals detected within inversion. The most pronounced selection outliers were detected in LG6 included partial selective sweep.

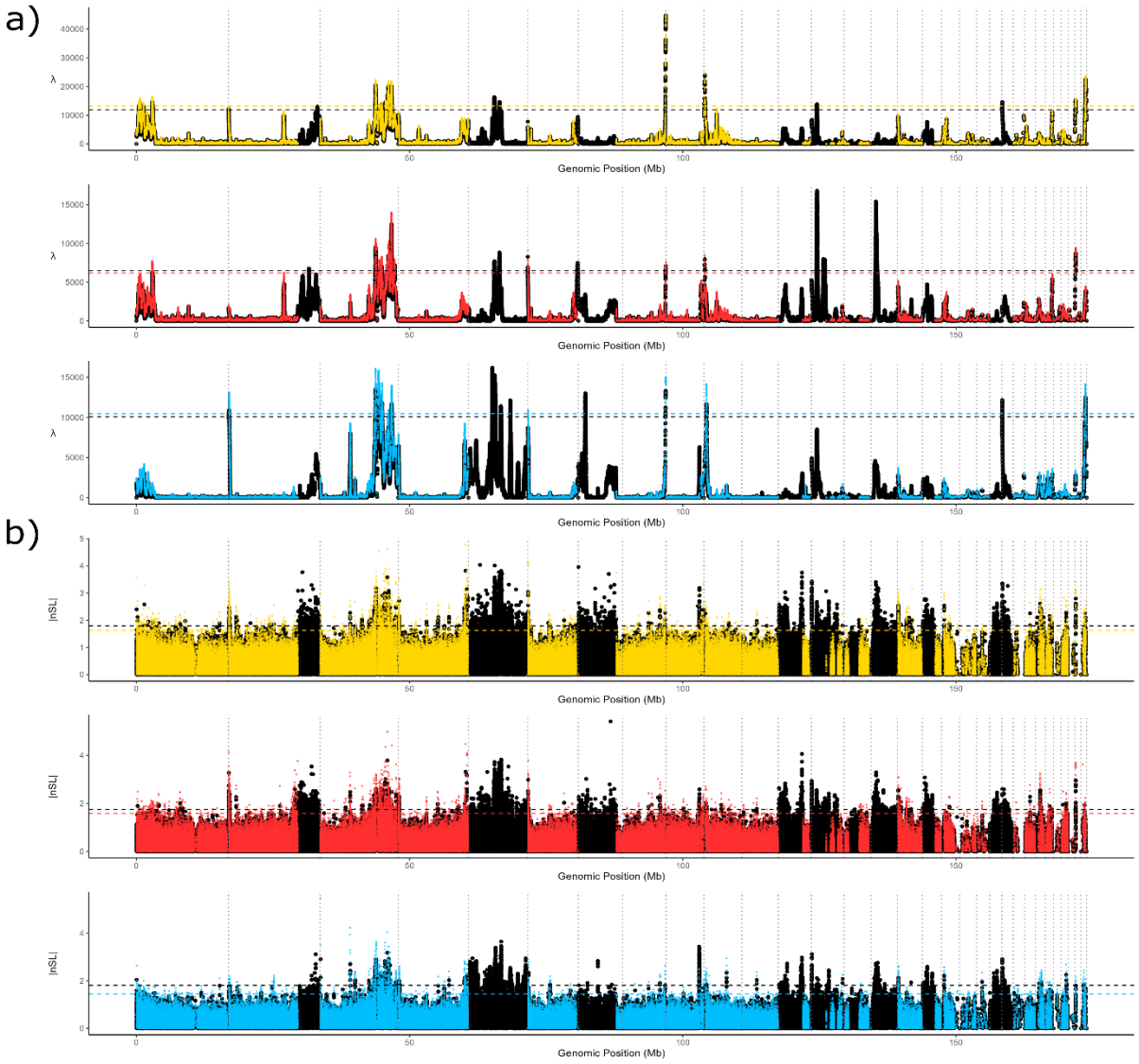

Figure 4. Results of the selection scans for the 52 SNP window sizes and SN, S, and N datasets for both selection scans (a) Λ and (b) nSL (absolute values). Results using both collinear and inversion regions are shown in black and the results of the same analysis performed excluding inversion regions are overlaid (SNcol – yellow, Scol – red, and Ncol – blue). The dashed lines indicate the outlier threshold (1% of all values). In all plots, grey dotted vertical lines represent the end of each contig.

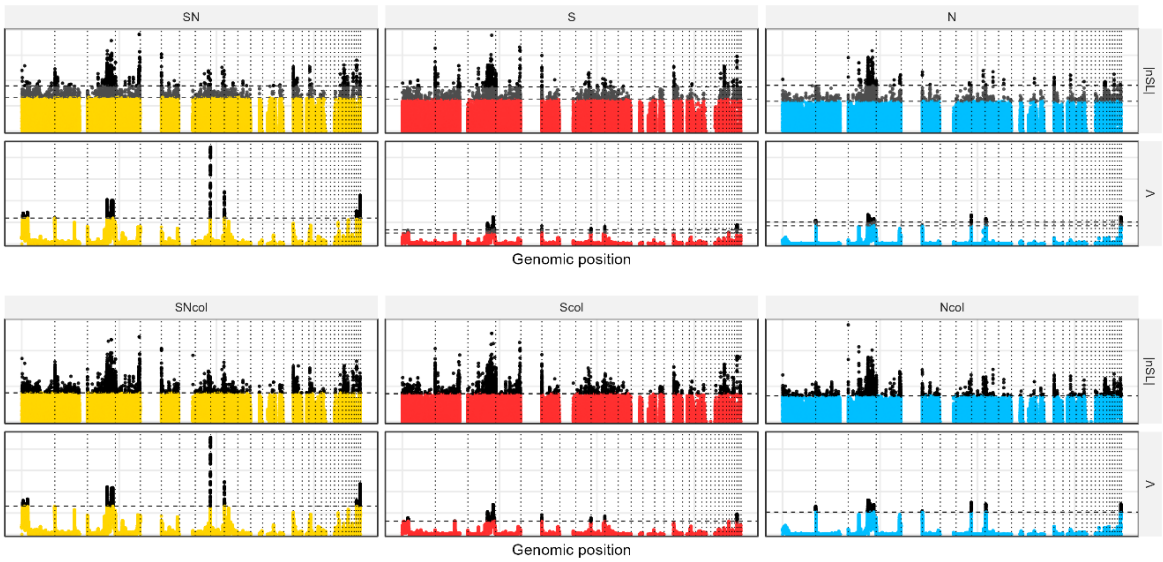

Figure 5: Results of the selection scans for 52 SNP window sizes and the SN, SNcol, S and Scol, N and Ncol datasets for Λ and nSL selection scans. Each column represents the results of the datasets relating to the geographic populations of the spruce bark beetle: SN (all individuals, south and north, yellow), S (south, red) and N (north, blue). Each row shows the results of the nSL or Λ analyses. The upper raw of plots shows the results of analyses performed with both collinear and inverted regions included, while the lower raw shows the results of analyses performed with only collinear regions. Every vertical dashed line differentiates a different contig. The horizontal dashed lines set the threshold at the 99th percentile, with outliers above shown in black and grey (depending on the threshold used). In the upper raw of plots, two horizontal dashed lines indicate the 99th percentile threshold: one was calculated including inversions (black outliers) and the other was calculated excluding them, with the corresponding outliers shown in dark grey and black. By comparing upper and lower raw one can see the inversions inclusion of inversions in the analyses does not affect the selection landscape in collinear genome.

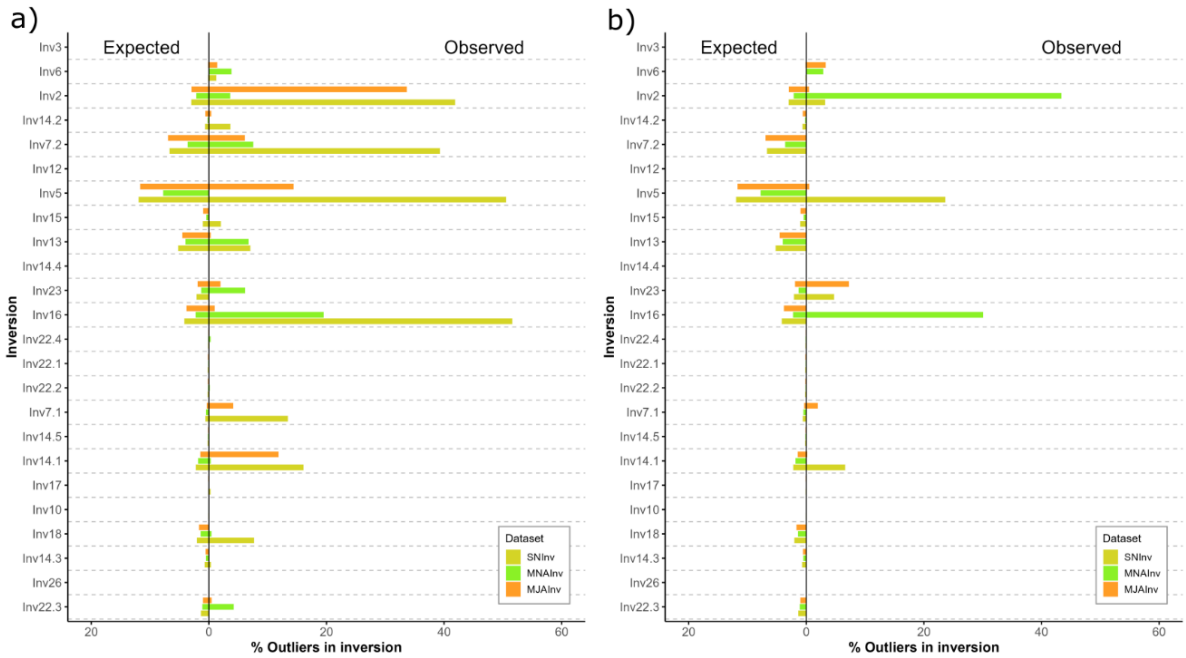

Figure 6. Comparison of the expected and the observed percentages of outliers detected in inversions. The expected values were calculated based on the length of the inversion and a 1% outlier threshold. The results are shown for nSL (a) and Λ (b). All selection scans were performed using 52 SNP windows and three datasets: datasets including individuals carrying any of the possible inversion genotypes (SNInv, yellow); datasets including only individuals that were homozygous for the major allele for the inversion in question (MJAInv, orange); and datasets including only individuals that were homozygous for the minor allele for the inversion in question (MNAInv, green). Each dataset included one analysed inversion and a collinear genome. The inversions are ordered according to the major allele frequency (from lowest to highest, top to bottom).

895 **Tables**

896 Table 1. Percentage of simulated datasets in which at least one outlier was detected within the
897 inversion. Calculations were performed for datasets with 1000 and 200 individuals and for both
898 Λ and nSL selection scans. Within each dataset, individuals were grouped according to their
899 inversion genotype into three categories: ALLIN: all individuals, including both heterozygotes
900 and homozygotes; MJA: including only individuals homozygous for the major inversion
901 haplotype; MNA: including only individuals homozygous for the minor inversion haplotype.

		1000 individuals		200 individuals	
		nSL	Λ	nSL	Λ
% of simulations with at least one outlier in the inversion	ALLIN	84%	59%	83%	60%
	MJA	32%	0%	32%	2%
	MNA	54%	1%	41%	1%

902

903 Table 2**.** The comparison between Λ and nSL selection scans using different datasets and a 52
904 SNP window size. Calculations from nSL analysis are based on the absolute values of the
905 statistic. Genomic region: part of the genome that is included in each dataset. Inversion-
906 associated outliers: the percentage of outlier windows located within inversion regions,
907 calculated from the total number of outliers. Collinear-region outliers: percentage of outliers
908 located outside inversion regions calculated from the total number of outliers. Shared outliers:
909 percentage of outliers that are shared between Λ and nSL results. Shared outliers in inversions:
910 the percentage of outliers shared between Λ and nSL results that are in inversion regions.
911 Shared outliers in collinear: the percentage of outliers shared between Λ and nSL results that
912 are in the collinear genome.

Dataset	Genomic region	Statistic	Inversion associated outliers	Collinear region outliers	Shared outliers	Shared outliers in inversions	Shared outliers in collinear
SN	Inversions + collinear	Λ nSL	29% 86%	71% 14%	15%	12%	3%
SNcol	Only collinear	Λ nSL		100% 100%	11%		11%
S	Inversions + collinear	Λ nSL	70% 87%	30% 13%	20%	18%	2%
Scol	Only collinear	Λ nSL		100% 100%	16%		16%
N	Inversions + collinear	Λ nSL	56% 91%	44% 9%	17%	15%	2%
Ncol	Only collinear	Λ nSL		100% 100%	19%		19%

913

914

Table 3. Comparison of the results of Λ and nSL selection scans obtained using different datasets. Calculations from nSL analysis are based on the absolute values of the statistic. Genomic region: part of the genome that is included in each dataset. Statistic: selection scan which comparisons were done. Shared outliers: percentage of outliers that are shared between datasets. Shared outliers in inversions: the percentage of outliers shared between datasets that are in inversion regions. Shared outliers in collinear: the percentage of outliers shared between datasets that are in the collinear genome

Compared datasets	Genomic region	Statistic	Shared outliers	Shared outliers in inversions	Shared outliers in collinear
SN, N	Inversions + collinear	Λ	23%	5%	18%
		nSL	33%	31%	2%
SNcol, Ncol	Only collinear	Λ	19%		19%
		nSL	32%		32%
SN, S	Inversions + collinear	Λ	30%	6%	24%
		nSL	48%	39%	9%
SNcol, Scol	Only collinear	Λ	44%		44%
		nSL	44%		44%
S, N	Inversions + collinear	Λ	9%	2%	7%
		nSL	20%	18%	2%
Scol, Ncol	Only collinear	Λ	24%		24%
		nSL	20%		20%

Table 4. Changes in percentage of outliers found in collinear parts of the genome when including or excluding inversions. Results are shown for three datasets (SN, N, and S) analysed using both selection scan methods (Λ and nSL) with results summarised in 50kb windows. Calculations from nSL analysis are based on the absolute values of the statistic. Outliers in collinear regions before excluding inversions: the percentage of outliers in collinear regions identified based on datasets that included inversions, inversions were not removed before identifying the outliers. Lost outliers in collinear parts after excluding inversions: the percentage of outlier windows in collinear parts that disappear after excluding inversions and recalculating the outlier threshold. New outliers in collinear parts after excluding inversions: the percentage of outlier windows in collinear parts that appear after excluding inversions and recalculating the outlier threshold.

Dataset	Statistic	Outliers in collinear regions before excluding inversions	Lost outliers in collinear parts after excluding inversions	New outliers in collinear parts after excluding inversions
SN	Λ	83%	9%	0%
	nSL	30%	0%	42%
S	Λ	48%	0%	26%
	nSL	36%	0%	36%
N	Λ	60%	0%	14%
	nSL	21%	0%	53%