Selection inference in a complex genomic landscape: the role of inversions in the spruce bark beetle

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. Using whole-genome resequencing data from 312 individuals across 23 populations, we applied two complementary selection scan methods to assess how structural genomic variants, such as inversions, affect the detection of adaptive signals. Our findings reveal that large polymorphic inversions are significantly enriched for selection signals, thereby reinforcing their pivotal role in adaptation. However, these same inversions can complicate the interpretation of selection signals within their boundaries, where recombination suppression can obscure the true targets of selection. In contrast, the detection of selection in collinear regions is largely unaffected by the presence of inversions. To overcome the confounding effects within inversions, we demonstrate that focusing on homozygote genotypes enables the identification of candidate regions and haplotype-specific signals. Our findings highlight the contribution of inversions to adaptation, emphasising the importance of understanding species-specific genomic architecture when interpreting selection scan results.

Key words: inversion-rich genomes, spruce bark beetle, selection scans, recombination suppression

Introduction

Elucidating the genetic basis of adaptation is one of the major goals of evolutionary biology (Bomblies & Peichel 2022). 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 (Barghi et al. 2020, Barrett & Schluter 2008, Bomblies & Peichel 2022, Herrmisson & Pennings 2017, Höllinger et al. 2019, Rockman 2012, Sella & Barton 2017, Stephan 2019) has led to a large body of empirical adaptation research (e.g. Bomblies & Peichel 2022, Enbody et al. 2023, Roberts Kingman et al. 2021). The progress in this field has been facilitated by a growing number of increasingly

sophisticated methods for genomic selection scans (Cheng & DeGiorgio 2022, DeGiorgio & Szpiech 2022, Kern & Schrider 2018). 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 (e.g. polymorphic inversions) affect our ability to identify true selection signals and the general importance of such regions in adaptation (Lotterhos 2019, Schaal et al. 2022).

A growing body of theoretical (Guerrero et al. 2012, Fuller et al. 2018, Kirkpatrick & Barton 2006, Schaal et al. 2022, Yeaman 2013, Yeaman et al. 2016) and empirical research identifies polymorphic inversions as key drivers of adaptation (Berdan et al. 2023, Harringmeyer & Hoekstra 2022, Kapun & Flatt 2019, Rane at al. 2015, Thompson & Jiggins 2014). Inversions are chromosomal mutations that reverse the order of the sequence in the affected region and suppress recombination between two inversion haplotypes in heterozygous individuals. The overall magnitude of recombination reduction depends on the frequency of heterozygous individuals within the population/species (Seich Al Basatena et al. 2013) and as suggested by recent studies, the effect of suppressed recombination may extend beyond the inverted region itself (Adrion et al. 2020, Crown et al. 2018, Stevison et al. 2011). Recombination suppression between the inverted and non-inverted arrangement is a critical mechanism that contributes to keeping coadapted genes together, facilitating local adaptation and allowing specific allelic combinations to remain sheltered from the homogenising effects of gene flow (Hoffmann & Rieseberg 2008, Kirkpatrick 2010).

Accumulating evidence suggests that genomes harbouring multiple polymorphic inversions, and thus characterised by complex recombination landscapes, may be more common than previously thought (so called inversion-rich genomes; Harringmeyer & Hoekstra 2022, Reeve et al. 2024). Emerging and still largely unexplored questions include: Does positive selection contribute to shaping patterns of inversion polymorphisms in species with inversion-rich genomes? Are we able to identify causal variants underlying adaptation within inversions using population genomic approaches? Simulation studies have shown that genome-wide studies can distinguish between neutral and adaptive inversions (Berdan et al. 2023, Lotterhos, 2019, Schaal et al. 2022), and many studies use either inversion frequency and/or phenotypic information to associate specific adaptive traits with inversion genotypes/haplotypes (Faria et al. 2019, Harringmeyer & Hoekstra 2022, Merot et al. 2021). However, empirical studies using classical selection scan approaches, that consider the effect of inversion landscapes on the

selection inference are still scarce. Such an approach may provide valuable insights into how inversions contribute to adaptation and how powerful selection scans are in the context of complex genomic recombination landscapes.

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 (Hlásny et al. in 2021, Seidl et al. in 2016). 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 (Dobor et al. 2024, Gohli et al. 2024, Kautz et al. 2024). 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 (Mykhailenko et al. 2024, Zieliński et al. 2025). Importantly, these studies identified 27 large polymorphic inversions covering approximately 30% of the species genome (Mykhailenko et. al 2024). 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.

Here, we performed genome-wide selection scans in the spruce bark beetle, with a particular focus on the multiple inversion polymorphisms present across its genome. We investigated how inversion-rich genome landscape affects performance of two selection scans methods that differ in their assumptions and explored solutions to minimise the effects of inversions on selection inference. Finally, we revealed candidate loci under selection within the spruce bark beetle genome.

Materials and Methods

Sampling, sequencing, and data preparation

We used data presented in Mykhailenko et al. (2024) and Zieliński et al. (2025). 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 2; Table S1). The details of library preparation, sequencing and processing are given in Mykhailenko et al (2024) and Zieliński et al. (2025). 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 (Andrews 2010). Low quality reads and adaptors were trimmed with Trimmomatic (Bolger et al. 2014) and mapped to the spruce bark beetle genome (Powell et al. 2021) with Bowtie2 (Langmead & Salzberg 2012). 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 (Browning et al. 2021). 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).

The effect of inversions on selection scans - general overview and predictions

There are two important aspects of inversions that can affect selection scans: low recombination rate and genetic subdivision within the inverted regions. Inversion regions are characterised by lower recombination due to suppressed recombination in heterozygous individuals. Thus, 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. Importantly, haplotype-based selection statistics tend to be biased toward low-recombination regions - selection signals are more frequently found in such regions (O'Reilly et al. 2024) and, at least in some cases, this may be a consequence of a higher false-positive rate in such regions (Ferrer-Admetlla et al. 2014). Additionally, 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

141 inversion region - a situation that is akin to a demographic scenario of two isolated populations.
142 In addition, gene flux (double-crossover and conversion events) that can happen between
143 inversion haplotypes mimics a scenario of two isolated populations with some migration
144 between them. This is important since panmixia is a common assumption of selection scan
145 methods. Inference methods are often evaluated for their robustness to population subdivision
146 (or other complex demographic scenarios; DeGiorgio & Szpiech 2022, Ferrer-Admetlla et al.
147 2014) but almost never consider inversions as a source of potential bias (Novo et al. 2023).

148 To test the effect of inversions on selection inference in the spruce bark beetle genome, we used
149 two different selection scan methods: the number of segregating sites by length (nSL, Ferrer-
150 Admetlla et al. 2014) and Λ from the saltiLASSI maximum likelihood approach (DeGiorgio &
151 Szpiech 2022). We focused on methods that do not require recombination rate information, as
152 this is still lacking for many of the species for which genomic data are being generated. Both
153 methods are designed to search for hard and soft sweep signatures of ongoing or recent
154 selection but differ in the way they identify and summarise selection signals and can be
155 differently affected by reduced recombination and genetic subdivision within inversion
156 regions. While nSL has been shown to be quite robust to variation in recombination rate, Λ
157 tends to be negatively correlated with recombination rate, and strong Λ signals in low-
158 recombination regions have to be interpreted with extra caution (DeGiorgio & Szpiech 2022,
159 Ferrer-Admetlla et al. 2014). Due to within inversions genetic subdivision nSL may be affected
160 by substructure created by inversions – the stronger the isolation between isolated haplotypes
161 the stronger the bias (Ferrer-Admetlla et al. 2014 tested several different demographic
162 scenarios including population subdivision). We may expect a similar effect on Λ, but
163 population subdivision has not been evaluated in saltiLASSI. In addition, both methods may
164 be differentially affected by the presence of multiple inversions within the investigated
165 genome. The saltiLASSI method is based on a maximum likelihood framework that
166 incorporates spatial patterns of haplotype frequency distortion into a statistical model of a
167 sweep. This is achieved by defining a genome-wide background haplotype frequency spectrum
168 representing neutral expectations and haplotype frequency spectra representing distorted
169 frequency spectra consistent with a model of one or more sweeping haplotypes. Selection
170 targets are identified by looking for extreme distortions in haplotype frequency spectra relative
171 to genome-wide (neutral) expectations. Since the neutral expectations are based on whole-
172 genome data, in the case of inversion-rich genomes, they will include multiple (and often large)
173 regions that will shape neutral expectation while they can clearly have different haplotype

spectra patterns than collinear parts of the genome. In contrast, nSL does not consider genome-wide expectations but its value depends on allele frequency and nSL values are normalised in frequency bins using whole genome variation information. Inversions, especially if old and maintained as polymorphic by, e.g., balancing selection, may harbour multiple fixed differences (resulting in intermediate allele frequency when both haplotypes are considered together), influencing the assignment of the nSL values into the frequency bins and the normalisation process. Finally, both methods are capable of detecting hard and soft sweeps, and it is possible that multiple high-frequency inversion haplotypes can mimic soft sweep signatures. Given the different approaches and assumptions used by these methods, predicting their behaviour in inversion-rich genomes is challenging, requires empirical assessment and careful interpretation of the results.

nSL was calculated in selscan 2.0.2 (Szpiech 2024) 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 (DeGiorgio & Szpiech 2022). 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 (DeGiorgio & Szpiech 2022). We tested different window sizes and based on our tests and previous studies (DeGiorgio & Szpiech 2022, Yang et al. 2023), we decided to use two window sizes: 52 (sliding window step 12) and 117 SNPs (sliding window step 12). The top 1% of Λ values were considered outliers.

To be able to compare selection signals between statistics, nSL results were translated into SNP-based Λ windows and mean values of nSL were calculated in each window. The top 1% mean (window) nSL absolute values were considered outliers. In order to be able to compare different datasets that contained different individuals and thus different numbers of SNPs and different numbers of SNP-based windows, we further standardised the results by summarising the statistics (both nSL and Λ) in non-overlapping 50 kb genomic windows. For each 50 kb window, we calculated the mean value of each statistic of all SNP-based windows that

overlapped it. This allowed a consistent window-based comparison across datasets and statistics.

Finally, to test if selection signals tend to cluster within inversions, we plotted outlier windows (SNP or 50kb windows according to each case) along the genome, investigated the position of each outlier window in comparison to inversion coordinates and calculated the sizes of clusters composed of adjacent outlier windows. Additionally, to correlate selection statistics with the genetic differentiation between inversion haplotypes we summarised Fst in SNP-based windows (in the same way as we did for nSL results) using vcftools (Danecek et al. 2011).

Selection scans - datasets, tested hypotheses and specific predictions

Multiple datasets used in the study are presented in Figure 1. Individuals were grouped according to the previously described genetic structuring into the southern and northern clade (Mykhailenko et al. 2024, Zielinski et al. 2025; Figure 2, Table S1) creating three datasets: 1) all individuals from southern and northern region (SN, 258 individuals), 2) only individuals from the southern populations (S, 138 individuals), 3) only individuals from the northern populations (N, 120 individuals). These datasets did not include Polish populations that are highly admixed. Each of these datasets was analysed in two versions, differing in the included parts of the genome: inversion regions were either included (SN, S, N) or excluded (SNcol, Scol, Ncol; only included collinear genome). To investigate the effects of selection in inverted homozygotes, all individuals were also divided into datasets based on their inversion genotype, forming three datasets per inversion: 1) individuals carrying major inversion haplotype (e.g. MJAInv2); 2) individuals carrying minor inversion haplotype (e.g. MNAInv2); 3) individuals carrying all possible inversion genotypes (e.g. SNInv2). Polish individuals were included in this analysis to ensure sufficient sample sizes, and because this analysis focused on differences between inversion genotypes rather than population-level structure. Inversion datasets consisted of inversion in question and collinear part of the genome (thus, excluding other inversions apart from the one in question).

The analysis of the datasets with and without inversions (SN, S and N, and their respective SNcol, Scol and Ncol) allowed us to investigate whether all or some inversions appear as selection outliers, to test whether the presence of multiple inversions affects the detection of selection signals outside of inversions (in collinear part of the genome), and to compare the behaviour of the two methods in these respects. Specifically, we predicted: 1) to find more

238 selection signals in inversions than in collinear parts due to the effect of reduced recombination
239 and inversions' genetic subdivision, 2) that balanced inversions with intermediate frequency of
240 both inversion haplotypes will show stronger selection signals than inversions with one allele
241 rarer (assuming that rare allele is younger and have low within haplotype polymorphism), 3)
242 that outlier detection outside inversions is affected by inversions present within the genome
243 due to their influence on haplotype frequency spectra (Λ) and the assignment of the statistics
244 values into the frequency bins (nSL) . To test whether selection signals are more often found
245 in inversion regions than in collinear parts of the genome we calculated the proportion of
246 outliers in inversions and performed a permutation test for each of the analysed dataset (1000
247 permutations, permuting SNP-based window labels - inverted and collinear). To check the
248 effect of inversion polymorphism frequency on selection scan performance, we correlated the
249 inversion haplotype frequency and the percentage of outliers within the inversion. To test
250 whether inversions affect selection inference in the collinear part of the genome we compared
251 datasets with and without inversions. We counted the number of outlier windows and
252 investigated whether the same regions were classified as outliers in both datasets and whether
253 new selection signals appear if we excluded inversions from the analysis. We compared both
254 methods by checking the number and distribution along the genome of the outlier windows
255 identified in both methods and across all datasets.

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

262 The inversion datasets (MJAInv, MNAInv and SNInv) allowed us to test whether: 1) selection
263 signals present in inversions when analysing dataset with all individuals (SNInv; all possible
264 inversion genotypes) disappear when the selection scans were performed using only
265 homozygous individuals (MJAInv or MNAInv); 2) There is a positive association between
266 inversion polymorphism and signal disappearance (more false positive selection signals
267 expected in high frequency of heterozygotes); 3) Selection signals disappear only in one
268 homozygotes revealing haplotype which carries alleles under selection. We expect that the
269 behaviour of the scans depends on the frequency of the inversion haplotype (produced by lower
270 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 as response

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

305

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

307 We identified genes harbouring or closest to outlier windows from the top 1% of nSL or Λ
308 values. We used gene annotations from Powell at. al (2021) and tested for overrepresentation
309 of GO using R package topGO (Alexa et al. 2006). We applied Fisher's exact test and
310 "weight01" algorithm (Alexa & Rahnenführer 2009) to deal with the GO graph structure and
311 only GO categories with at least 10 members among the SNP associated genes were considered.
312 GO analysis were performed for datasets including: 1) selection outliers outside inversions and
313 specific for northern group; 2) selection outliers outside inversions and specific for southern
314 group; 3) selection outliers outside inversions common for both groups; 4) selection outliers
315 specific for particular inversion haplotype or present in both inversion haplotypes (the
316 inversion signals that we consider true signals).

317

318 **Results**

319 We used whole-genome sequence data from hundreds of individuals and performed nSL and Λ
320 selection scans across the inversion-rich spruce bark beetle genome. The list of all datasets and
321 their characteristics (e.g. whether they contained inversions or not) are given in Figure 1. In
322 short, the overall results of the selection scans are presented in Figure 3 and Tables 1-2 (nSL
323 and Λ for 52 SNPs window size), Figures S1 and Table S4 (nSL and Λ for 117 SNPs window)
324 and Figure S2 (nSL results before translating to SNP-based windows). The distribution of
325 selection signals in collinear regions is shown in Figure 4 and Table 3 (nSL and Λ for 50kb
326 windows), and the selection scan results in homozygous datasets are shown in Figure 5, Table
327 S5 and Figure S4 (nSL and Λ for 52 SNPs window size).

328

329 *Selection landscape in inversion-rich genome*

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

331 Most of the outliers were identified within inversion regions (Figure 3, Figure S1, Table 1),
332 many more than expected by chance (permutation tests; $p < 0.001$; Supplementary Information

333 1.1). This pattern was consistent across both statistics, window sizes and all datasets (SN, S,
334 N), except for Λ scan and SN dataset and 52 SNPs window size, where more outliers than
335 expected were found in the collinear part of the genome (Table 1). In the latter case the signal
336 was mainly driven by one region in IpsContig3 (39% of all outliers and 55% of collinear
337 outliers in IpsContig3). Only 7 inversions (out of 24) did not show any selection signals in nSL
338 or Λ analysis (Table S2). Inversions not only included more outliers in total but also harboured
339 longer clusters (in particular in Λ scans; permutation test, $p < 0.001$; Table S3; Supplementary
340 Information 1.8). This was also true for the Λ scan with 52 SNPs window size that had a smaller
341 percentage of outliers in inversions compared to the collinear part of the genome, even
342 considering that inversions represent only ~30% of the genome (Figure 3, Table S3). The
343 largest inversion outlier clusters were found in inversion Inv16 (0.324 Mb, 3% of the inversion,
344 in S dataset) and Inv5 (0.253 Mb, 6% of the inversion, in N dataset, Figure 3, Table S3). The
345 largest cluster outside inversion was found in IpsContig3 (ranging from 0.317 to 0.369 Mb in
346 different datasets). We did not find a significant correlation between frequency of inversion s
347 and amount of selection signals (Supplementary Information 1.3). In summary, we found that
348 inversions are enriched in selection signals and that these signals form large blocks of outlier
349 regions.

350 *2. Selection landscape identified by both methods in collinear parts is minimally affected*
351 *by inversions present in the genome*

352 Both nSL and Λ scans conducted using the dataset with inversions, revealed multiple signals
353 of selection outside inversion regions (Table 1-2, Figure 3-4, Figure S1-S2). The proportion of
354 outliers detected depended on the dataset and selection scan method used. Figure 4 and Table
355 3 compare the selection landscape in the collinear part of the genome in datasets that included
356 inversions when the selection scans were performed (SN, S, N, inversions were kept or
357 removed for the outlier threshold calculation) and datasets in which inversions were excluded
358 before selection scan analysis (SNcol, Ncol, Scol; both in 50kb widows that allowed us to
359 compare datasets with different number of SNP windows). There was a significant correlation
360 between the results obtained in the collinear part of the genome using datasets that included
361 and excluded inversions (r^2 ranged from 0.76-0.95, permutation test; p-value < 0.001;
362 Supplementary Information 1.6). In summary, the general selection landscape in collinear
363 genome was almost the same between datasets that included and excluded inversions indicating
364 minimal influence of inversions on the selection scan results in the collinear part of the genome.

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

366 Both selection scan methods produced a fairly similar selection landscape across all datasets
367 (r^2 ranged from 0.30 to 0.45, permutation test; $p < 0.001$; Supplementary Information 1.4).
368 Only the selection landscape obtained from Λ scans with window size 117 SNPs was visually
369 different and dominated by two large outlier clusters in inversions Inv5 and Inv7 (Figure S1,
370 Table S4). Despite this difference, many regions identified in the analysis that used 52 SNPs
371 window size (both nSL and Λ) were still visible as distinct peaks when using window size of
372 117 SNPs but were not classified as outliers in this dataset due to the extreme values of the two
373 inversions mentioned above. In general, the genome-wide selection landscapes were similar
374 across different statistics (as indicated by a high correlation of values in 50kb widows), but the
375 actual overlap between outlier SNP-based windows in nSL and Λ was rather small due to the
376 difference in values for particular windows - the same regions appear as outliers but particular
377 small SNP windows can differ in the magnitude of selection signals (range from 11% to 20%
378 of shared outliers in Λ and nSL for 52 SNPs window size in different datasets; Table 1), but
379 this is still more overlap then expected by chance (permutation test, $p < 0.001$; Supplementary
380 Information 1.4). In summary, both methods identify similar selection landscapes despite the
381 observed difference in magnitude of selection signal between nSL and Λ.

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

384 Selection landscapes inferred for S and N datasets were highly correlated with each other (r^2
385 0.76 for Λ, 0.81 for nSL; $p < 0.001$; Figure 3, Figures S1–S2, Tables 1-2, Supplementary
386 Information 1.5). This overall similarity was consistent across both nSL and Λ and reflected
387 the largely shared genomic landscape of selection between geographic regions. However, a
388 region-specific outlier windows were also detected (Figure S3, Table 2). In addition, both nSL
389 and Λ values were positively correlated with Fst values calculated across the genome and
390 between the N and S datasets (r^2 0.0004–0.0634 across datasets; $p < 0.001$; Supplementary
391 Information 1.2), indicating that regions of high genetic differentiation may be enriched for
392 signals of selection and at least some of them may be involved in local (regional) adaptation.

393

394

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

To minimise the effect of the recombination suppression and genetic subdivision between inversion haplotypes, both selection scans were performed in individuals carrying particular inversions only in homozygous states (Figure 1, MJAInv and MNAInv datasets). Each inversion was evaluated separately for both inversion homozygotes and excluding the genomic regions that contained inversions (e.g. MJAInv2 and MNAInv2; Figure 5) and compared to the results obtained using a dataset that included all genotypes in inversion in question (dataset e.g. SNInv2). We tested several hypothesis: 1) Selection signals present in inversions when analysing a dataset with SNInv individuals disappear when selection scans are performed using only homozygous individuals (MJAInv or MNAInv); 2) Selection signals disappear only in inversions with the highest frequency of heterozygotes (false positive signals would be more often in such cases); 3) Selection signals disappear only in one homozygote, revealing a haplotype that carries alleles under selection.

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

nSL selection signal was detected in ten inversions in both homozygotes, eight inversions in only one homozygote (four of which were minor alleles), and in three inversions selection signals disappeared completely in the homozygous datasets (Table S5). In the Λ scans, three inversions showed selection signals in both homozygotes, and five inversions showed signals in only one homozygote. None of the inversions lost their selection signals in both homozygotes. Figure 5 shows the large variation in the difference between the observed and the expected percentages of outliers among inversions and datasets. For example, Inv2 and Inv16 show a much higher proportion of observed outliers in SNInv and one of the homozygote datasets—dozens observed compared to only a few percent expected. In contrast, Inv13 exhibits similar observed and expected percentages in both SNInv and MJNInv, and notably fewer observed outliers than expected in MJAInv (Figure 5a). Overall, analysis of homozygote datasets revealed that the majority of inversions harbour selection signals in either both or one inversion haplotype (in the latter case revealing haplotype under selection). However, in few cases selection signals in homozygote datasets disappeared completely indicating false positive signals obtained in SNInv datasets.

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

428 While analyses using SN dataset (all individuals and all inversions), showed that inversion
429 regions are overall enriched for selection signals compared to collinear regions, we observed
430 that the strength of these signals tends to be reduced in homozygote datasets. Our tests indicated
431 a depletion of the signal in homozygotes (MJAInv and MNAInv datasets) compared to the
432 SVInv dataset (Wilcoxon tests, $p < 0.02$ for both nSL and Λ scans; see
433 Supplementary_Information_Selection_model.html). However, the magnitude of the selection
434 signal in inversion regions did not correlate with inversion haplotype frequency. Similarly,
435 inversion haplotype frequency did not explain the magnitude of selection signals within
436 inversions or indicate any clear patterns regarding what happens to selection signals in
437 inversions (the disappearance of the signal is not dependent on inversion frequency; see
438 Supplementary_Information_Selection_model.html).

439 In summary, although selection signal tends to be reduced in homozygous datasets compared
440 to SNInv, there is a considerable heterogeneity in this respect among inversions and among
441 datasets (Figure 5, Figure S4, Table S5). The strength of the selection signals and its behaviour
442 in different datasets did not depend on the frequency of inversions.

443

444 *Gene ontology terms enriched in selection targets*

445 To explore the functions associated with regions under selection, we performed Gene Ontology
446 (GO) enrichment analysis using outlier windows that (1) were located outside inversions and
447 specific to N population, (2) were located outside inversions and specific to S population, (3)
448 were shared between S, N, and SN datasets, or (4) remained significant after excluding
449 heterozygotes in the analysis performed with only homozygotes. In total, we identified 24
450 significantly enriched GO terms (Table S6). Of these, four terms were shared between both
451 nSL and Λ statistics in the only homozygote datasets, while five terms were exclusive to Λ and
452 three to nSL.

453 In the parts outside of the inversions from SN, S, and N datasets, two GO terms were shared.
454 Only one term was enriched specifically for S and also only one for N populations. Enriched
455 functions included terms such as "signalling receptor activity" and "response to stimulus,"
456 potentially related to odorant receptor function. A complete list of enriched GO terms, including
457 fold enrichment and significance levels, is provided in Table S6.

Discussion

In this study we used whole-genome resequencing data from 312 individuals of the spruce bark beetle to investigate how selection signals are distributed throughout its genome and how chromosomal inversions influence these patterns and our ability to detect them.

As predicted, we observed a significant enrichment of selection signals within inversion regions using both nSL and Λ statistics. Notably, even in the homozygotes, most inversions exhibit some selection signals in at least one homozygote dataset. These findings reinforce the view that most inversions contribute to adaptation (Christmas et al. 2019, Nadachowska-Brzyska et al. 2019, Pegueroles et al. 2010, Rane et al. 2015) and are unlikely to be selectively neutral (Berden et al. 2023), particularly in the case of large polymorphic inversions (Connallon & Olito, 2022) — the kind identified in the spruce bark beetle genome (Mykhailenko et al. 2024).

On the other hand, we observed substantial variation in both the magnitude and behaviour (enrichment or depletion) of selection signals across inversions and datasets, in particular when all genotypes were contrasted with homozygotes. In general, selection signals were weaker in the homozygote datasets, suggesting that some signals detected in the entire dataset may be inflated by the presence of heterozygous individuals. By restricting analyses to individuals carrying only one inversion haplotype, we reduced the likelihood of false positives and were able to detect haplotype-specific selection signals. This approach also helped refine candidate selection regions and distinguish genuine signals from spurious ones. Importantly, the lack of a clear relationship between the extent of inversion polymorphism and the behaviour of selection signals across datasets underscores a key challenge: although inversions are enriched for selection signals, identifying the specific targets of selection within these regions remains difficult.

Contrary to our expectations, we found that the presence of inversions had only a minimal impact on the shape of the selection landscape and our ability of detect selection signals in the collinear genome. Importantly, this funding suggests that identifying loci under selection outside of inversions remains feasible, even in genomes characterized by extensive inversion polymorphism. This result aligns with our earlier findings on demographic inference (Zieliński et al. 2025), where we showed that overlooking chromosomal inversions had little effect on model selection and introduced only minor biases in the estimation of demographic parameters.

Additionally, the spruce bark beetle selection scans supported the necessity of using multiple complementary approaches to selection scans. While the overall patterns of selection inferred by Λ and nSL were similar, differences were observed at a local scale. There was a high correlation of selection signals at 50 kb windows, but only a small overlap among outlier windows at 52 SNPs, which demonstrates the general difficulty of selection scans in pinpointing the focal gene under selection (Vitti et al. 2013). Moreover, the selection of window sizes in the Λ scans can influence which regions are considered candidates for selection. The differences observed in the spruce bark beetle between 52 and 117 SNP windows can, to some extent, reflect differences in the ability to detect more recent and more ancient sweep signatures: smaller window sizes are more effective at detecting ancient sweeps (DeGiorgio & Szpiech 2022).

Our study revealed a broadly similar selection landscape between the southern and northern populations of the spruce bark beetle, suggesting that these populations are likely subject to comparable selective pressures and/or that their divergence is relatively recent, with the observed signals possibly reflecting selection in a shared ancestral population. According to our recent demographic modelling, the divergence between the southern and northern clades occurred approximately 80,000 years ago (Zieliński et al. 2025). Signals of selection in the site frequency spectrum are expected to persist for roughly 0.1Ne generations, while haplotype-based signals fade more rapidly, typically within ~0.01Ne generations (Pennings & Hermisson 2006, Przeworski 2002). Assuming an ancestral Ne of 250,000 (Zieliński et al. 2025), haplotype-based signatures could persist for around 2,500 generations—well after the estimated divergence time—supporting the idea that shared selection pressures may be acting across both populations. The inferred Ne for the northern and southern populations after the split was 180,000 and 830,000, respectively, and even under the higher Ne scenario, the timescale of detectable selection signals remains post-divergence. These findings suggest that spruce bark beetle populations across Europe may be subject to similar selective pressure. Nonetheless, we also detected evidence of regional differentiation: one of the strongest selection signals outside of inversions (on IpsContig3) overlapped with a pronounced Fst peak between the northern and southern groups, indicating that divergent selection may also contribute to shaping the selection landscape in this species.

Finally, our GO enrichment analysis identified terms such as "signalling receptor activity" and "response to stimulus," which are potentially linked to odorant receptor function. These findings are of particular interest considering recent studies showing that inversions are

enriched in genes encoding odorant receptors (Mykhailenko et al. 2024), and that these inversions may contain receptor variants that respond differently to different components of pheromones (Johny et al. 2025).

Conclusions and future directions

Our study provides new insights into the selection landscape across the spruce bark beetle genome, identifying candidate genes that may guide future research into the genomic basis of adaptation in this ecologically and economically important species. We show that most polymorphic inversions in the genome are enriched for selection signals, supporting their role as key contributors to adaptation, while also highlighting the importance of accounting for structural variation — particularly inversions — when performing selection inference. As illustrated by the changing selection patterns in the homozygote datasets, inversions can affect the detection of selection within their regions. Nevertheless, their presence does not obscure the identification of selection signals outside inversions, emphasizing the feasibility of conducting genome-wide selection scans even in inversion-rich genomes. More broadly, our results underscore the need to systematically evaluate the performance and robustness of evolutionary inference methods in the presence of structural variation — a factor still largely overlooked in standard methodological assessments, with notable exceptions such as the recent evaluation of GONE, which used simulations to assess how balanced inversion polymorphisms affect effective population size estimates (Novo et al. 2023). However, such efforts remain rare, and further research is needed to understand how inversions influence evolutionary inference, particularly given that their effects likely vary across species with differing numbers and evolutionary histories of large polymorphic inversions.

Acknowledgments

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

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 analyse, 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.

References

Adrion, J. R., et al (2020). Predicting the landscape of recombination using deep learning. *Molecular Biology and Evolution*, vol. 37(6):1790–808. https://doi.org/10.1093/molbev/msaa038.

Alexa, A., et al. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure'. *Bioinformatics*, vol. 22(13):1600–16007. https://doi.org/10.1093/bioinformatics/btl140.

Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv*, 27(1-26): 776.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Braham Bioinformatics.

Barrett, R., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, vol. 23(1):38–44. https://doi.org/10.1016/j.tree.2007.09.008.

Berdan, E. L., et al. (2023) How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology*, vol. 36(12):1761–1782. https://doi.org/10.1111/jeb.14242.

Bolger, A. M., et al. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, vol. 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bomblies, K., and Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences*, vol. 119(30):1-10. https://doi.org/10.1073/pnas.2122152119.

Browning, B. L., et al. (2021) Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, vol. 108(10):1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Cheng, X., and DeGiorgio, M. (2022). BalLeRMix +: mixture model approaches for robust

joint identification of both positive selection and long-term balancing selection. *Bioinformatics*, vol. 38(3):861–63. https://doi.org/10.1093/bioinformatics/btab720.

Christmas, M. J., et al. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*, vol. 28(6):1358–1374. https://doi.org/10.1111/mec.14944.

Connallon, T., and Olito, C. (2022). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, vol. 31(13):3627–3641. https://doi.org/10.1111/mec.16091.

Crown, K. N., et al. (2018). Local inversion heterozygosity alters recombination throughout the genome. *Current Biology*, vol. 28(18):2984-2990. https://doi.org/10.1016/j.cub.2018.07.004.

Danecek, P., et al. (2011). The Variant Call Format and VCFtools'. *Bioinformatics*, vol. 27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

DeGiorgio, M., and Szpiech. Z. A. (2022). A spatially aware likelihood test to detect sweeps from haplotype distributions. *PLOS Genetics*, vol. 18(4):1-37 https://doi.org/10.1371/journal.pgen.1010134.

Dobor, L., et al. (2024). The interacting effect of climate change and herbivory can trigger large-scale transformations of European temperate forests. *Global Change Biology*, vol. 30(2):1-20. https://doi.org/10.1111/gcb.17194.

Enbody, E. D., et al. (2023). Community-wide genome sequencing reveals 30 years of Darwin's finch evolution. *Science*, vol. 381(6665):1-9. https://doi.org/10.1126/science.adf6218.

Faria, R., et al. (2019). Evolving inversions. *Trends in Ecology & Evolution*, vol. 34(3):239–48. https://doi.org/10.1016/j.tree.2018.12.005.

Ferrer-Admetlla, A., et al. (2014). On detecting incomplete soft or hard selective sweeps

using haplotype structure. *Molecular Biology and Evolution*, vol. 31(5):1275–1291. https://doi.org/10.1093/molbev/msu077.

Fuller, Z. L., et al. (2018). Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLOS Genetics*, vol. 14(7):1-26. https://doi.org/10.1371/journal.pgen.1007526.

Gohli, J., et al. (2024). Climatic and management-related drivers of endemic European spruce bark beetle populations in Boreal forests. *Journal of Applied Ecology*, vol. 61(4):809–820. https://doi.org/10.1111/1365-2664.14606.

Guerrero, R. F., et al. (2012). Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367(1587):430–438. https://doi.org/10.1098/rstb.2011.0246.

Harringmeyer, O. S., and Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, vol. 6(12):1965–1979. https://doi.org/10.1038/s41559-022-01890-0.

Hlásny, T., et al. (2021). Bark beetle outbreaks in Europe: state of knowledge and ways forward for management. *Current Forestry Reports*, vol. 7(3):138–165. https://doi.org/10.1007/s40725-021-00142-x.

Hoffmann, A. A., and 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*, vol. 39(1):21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532.

Höllinger, I., et al. (2019). Polygenic adaptation: from sweeps to subtle frequency shifts. *PLOS Genetics*, vol. 15(3):1-26. https://doi.org/10.1371/journal.pgen.1008035.

Johny, J., et al. (2025). Not all bark beetles smell the same: population-level functional olfactory polymorphisms in *Ips Typographus* pheromone receptor ItypOR33.

Molecular Ecology, vol. 34(6):1-19. https://doi.org/10.1111/mec.17693.

Kapun, M., et al. (2023). An ancestral balanced inversion polymorphism confers global adaptation. *Molecular Biology and Evolution*, vol. 40(6):1-28. https://doi.org/10.1093/molbev/msad118.

Kautz, M., et al. (2024). Early detection of bark beetle (*Ips typographus*) infestations by remote sensing – a critical review of recent research. *Forest Ecology and Management*, vol. 556:1-17. https://doi.org/10.1016/j.foreco.2023.121595.

Kern, A. D., and Schrider, D. R. (2018). diploS/HIC: an updated approach to classifying selective sweeps. *G3 Genes|Genomes|Genetics*, vol. 8(6):1959–1970. https://doi.org/10.1534/g3.118.200262.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, vol. 8(9):1-5. https://doi.org/10.1371/journal.pbio.1000501.

Kirkpatrick, M., and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, vol. 173(1):419–434. https://doi.org/10.1534/genetics.105.047985.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, vol. 9(4):357–359. https://doi.org/10.1038/nmeth.1923.

Lotterhos, K. E. (2019). The effect of neutral recombination variation on genome scans for selection. *G3 Genes|Genomes|Genetics*, vol. 9(6):1851–1867. https://doi.org/10.1534/g3.119.400088.

Mérot, C., et al. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Molecular Biology and Evolution*, vol. 38(9):3953–3971. https://doi.org/10.1093/molbev/msab143.

Mykhailenko, A., et al. (2024). Complex genomic landscape of inversion polymorphism in Europe's most destructive forest pest. *Genome Biology and Evolution*, vol. 16(12):1-

23. https://doi.org/10.1093/gbe/evae263.

Nadachowska-Brzyska, K., et al. (2019). Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. *Molecular Ecology*, vol. 28(9):2290–2304. https://doi.org/10.1111/mec.15021.

Novo, I., et al. (2023). Impact of population structure in the estimation of recent historical effective population size by the software GONE. *Genetics Selection Evolution*, vol. 55(1):1-15. https://doi.org/10.1186/s12711-023-00859-2.

O'Reilly, G. D., et al. (2024). A new method for ecologists to estimate heterozygote excess and deficit for multi-locus gene families. *Ecology and Evolution*, vol. 14(7):1-13. https://doi.org/10.1002/ece3.11561.

Pegueroles, C., et al. (2010). Recombination and selection in the maintenance of the adaptive value of inversions. *Journal of Evolutionary Biology*, vol. 23(12):2709-2717. https://doi.org/10.1111/j.1420-9101.2010.02136.x.

Pennings, P. S., and Hermisson, J. (2006). Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, vol. 23(5):1076-1084. https://doi.org/10.1093/molbev/msj117.

Powell, D., et al. (2021). A highly contiguous genome assembly of the Eurasian spruce bark beetle, *Ips typographus*, provides insight into a major forest pest. *Communications Biology*, vol. 4(1):1-9. https://doi.org/10.1038/s42003-021-02602-3.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics*, vol. 160(3):1179–1189. https://doi.org/10.1093/genetics/160.3.1179.

Rane, R. V., et al. (2015). Genomic evidence for role of inversion 3RP of *Drosophila Melanogaster* in facilitating climate change adaptation. *Molecular Ecology*, vol. 24(10):2423–2432. https://doi.org/10.1111/mec.13161.

Reeve, J., et al. (2024). Chromosomal inversion polymorphisms are widespread across the

species ranges of rough periwinkles (*Littorina Saxatilis* and *L. arcana*). *Molecular Ecology*, vol. 33(24):1-18. https://doi.org/10.1111/mec.17160.

Roberts Kingman, G. A., et al. (2021). Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, vol. 7(25):1-12. https://doi.org/10.1126/sciadv.abg5285.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, vol. 66(1):1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x.

Schaal, S. M., et al. (2022). Inversion invasions: when the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 377(1856):1-14. https://doi.org/10.1098/rstb.2021.0200.

Seich Al Basatena, N. K., et al. (2013). The effect of genomic inversions on estimation of population genetic parameters from SNP data. *Genetics*, vol. 193(1):243-253. https://doi.org/10.1534/genetics.112.145599.

Seidl, R., et al. (2016). Small beetle, large-scale drivers: how regional and landscape factors affect outbreaks of the European spruce bark beetle. *Journal of Applied Ecology*, vol. 53(2):530-540. https://doi.org/10.1111/1365-2664.12540.

Sella, G., and Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics*, vol. 20(1):461-493. https://doi.org/10.1146/annurev-genom-083115-022316.

Stephan, W. (2019). Selective sweeps. *Genetics*, vol. 211(1):5–13. https://doi.org/10.1534/genetics.118.301319.

Stevison, L. S., et al. (2011). Effects of inversions on within- and between-species

recombination and divergence. *Genome Biology and Evolution*, vol. 3:830-841. https://doi.org/10.1093/gbe/evr081.

Szpiech, Z. A., et al. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution Letters*, vol. 5(4):408-421. https://doi.org/10.1002/evl3.232.

Szpiech, Z. A. (2024). Selscan 2.0: scanning for sweeps in unphased data. *Bioinformatics*, vol. 40(1):1-6. https://doi.org/10.1093/bioinformatics/btae006.

Thompson, M. J., and Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, vol. 113(1):1–8. https://doi.org/10.1038/hdy.2014.20.

Vitti, J. J., et al. (2013). Detecting natural selection in genomic data. Annual Review of Genetics, vol. 47(1):97-120. https://doi.org/10.1146/annurev-genet-111212-133526.

Yang, X., et al. (2022). Lineage-specific targets of positive selection in three leaf beetles with different defence capacity against a parasitic wasp. *Authorea.* https://doi.org/10.22541/au.166754711.16782535/v1.

Yeaman, S., et al. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, vol. 353(6306):1431–1433. https://doi.org/10.1126/science.aaf7812.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, vol. 110(19):1743-1751. https://doi.org/10.1073/pnas.1219381110.

Zieliński, P., et al. (2025). Demographic history inferred from an inversion-rich spruce bark beetle genome. Submitted.

Figures

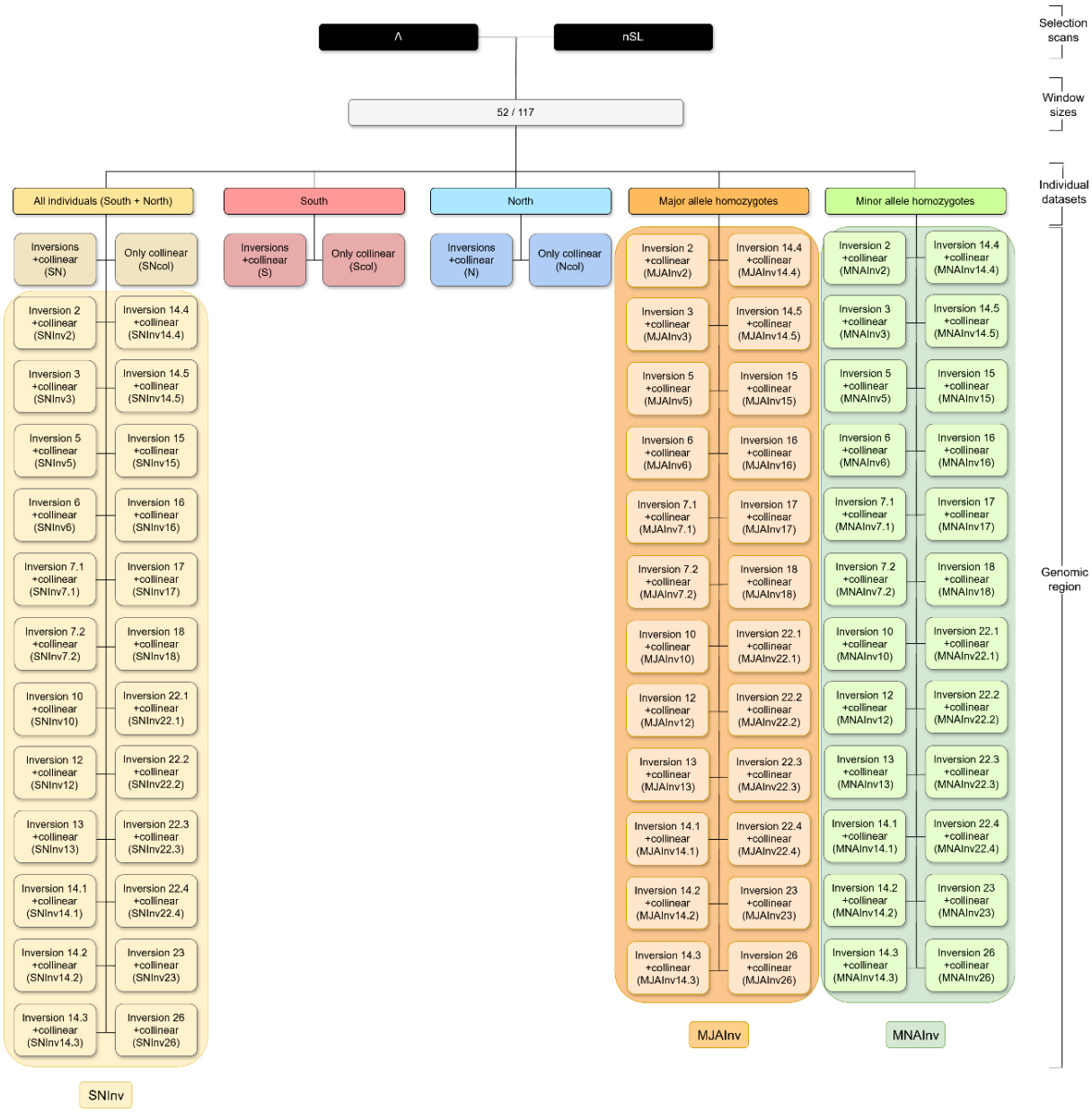


Figure 1. A graphical summary of the datasets used for selection scans in the spruce bark beetle. nSL and Λ selection scans were performed with two window sizes: 52 and 117 SNPs. Individual beetles were assigned to different groups based on the following two factors: 1) geographic origin: southern populations (S and Scol ("col" abbreviation of "collinear" to refers to datasets that included only collinear parts of the genome, excluding inversions), red), northern populations (N and Ncol, blue), or all individuals (SN and SNcol, southern and northern region combined, yellow), and 2) inversion genotype: major allele homozygotes (MJAInv, orange), minor allele homozygotes (MNAInv, green), and groups including all

possible genotypes (SNInv, yellow). Each dataset included versions that included (e.g., SN) or excluded (e.g., SNcol) inversions from the analysis.

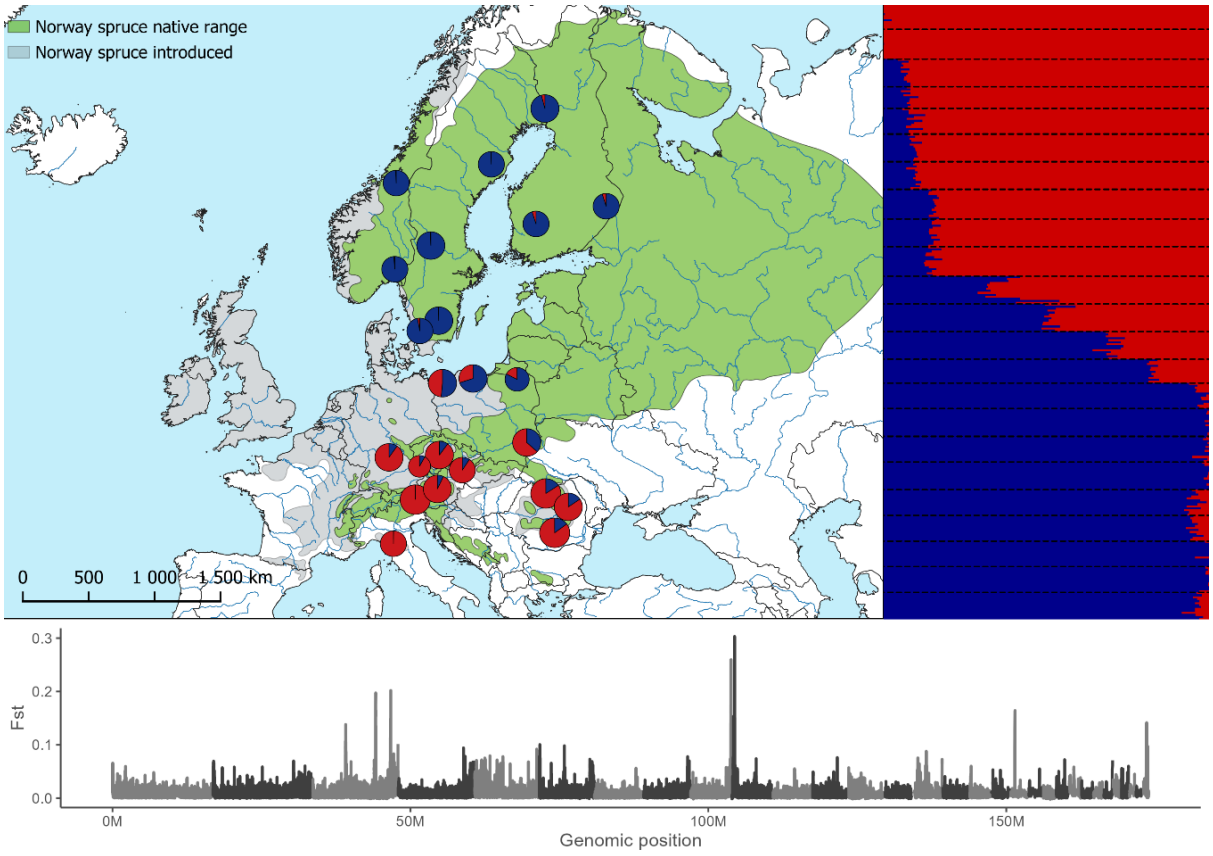


Figure 2. The geographic distribution, admixture, and genetic differentiation of *Ips typographus* populations as inferred in Zielinski et al. 2025. Top left: map of sampled populations, with admixture proportions illustrated by coloured circles. The colours represent the admixture proportions that were 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.

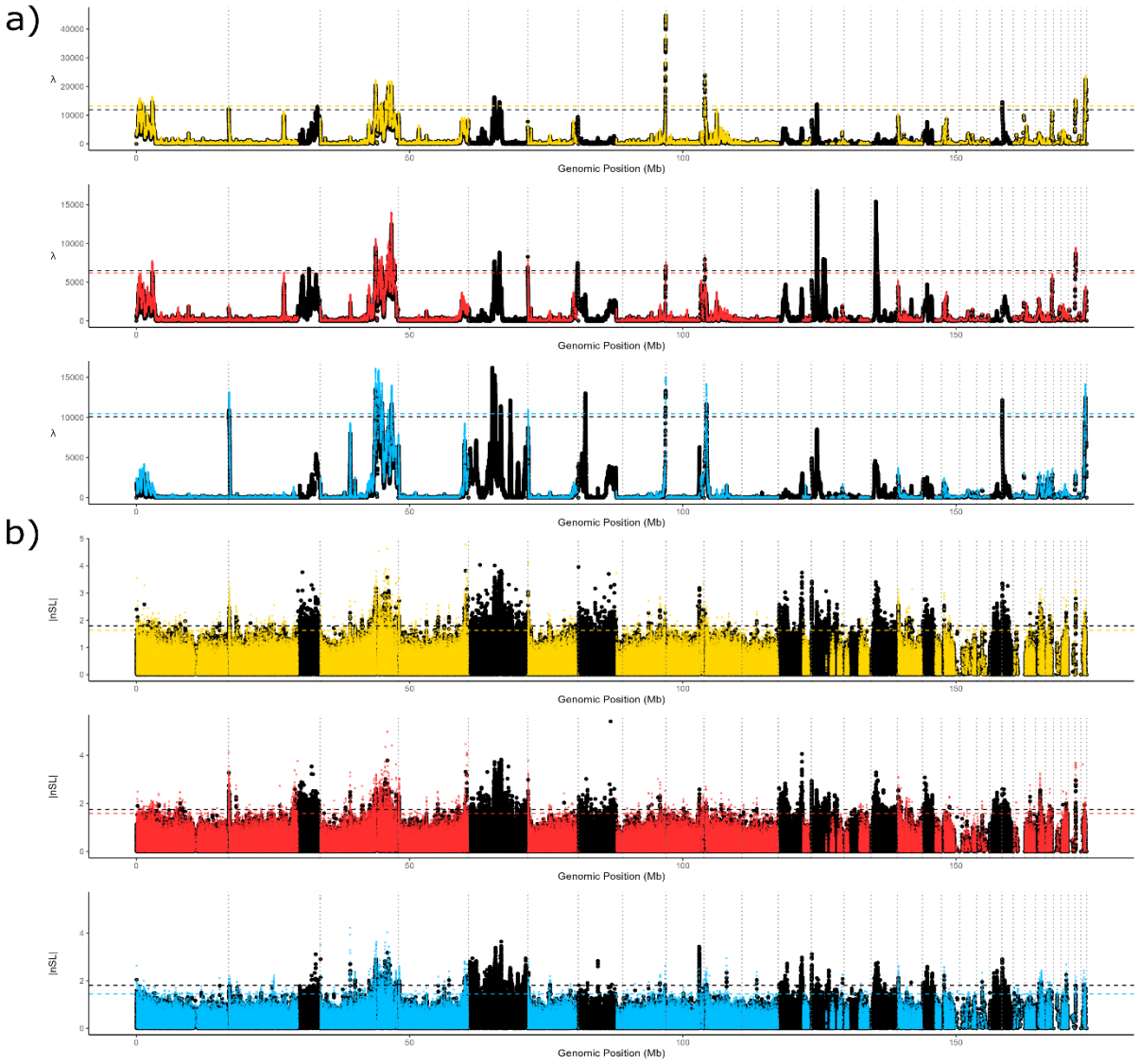


Figure 3. Results of the selection scans for the 52 SNP window sizes and SN, S, and N datasets for both selection scans (a) Λ and (b) nSL (absolute values). Results using both collinear and inversion regions are shown in black and the results of the same analysis performed excluding inversion regions 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.

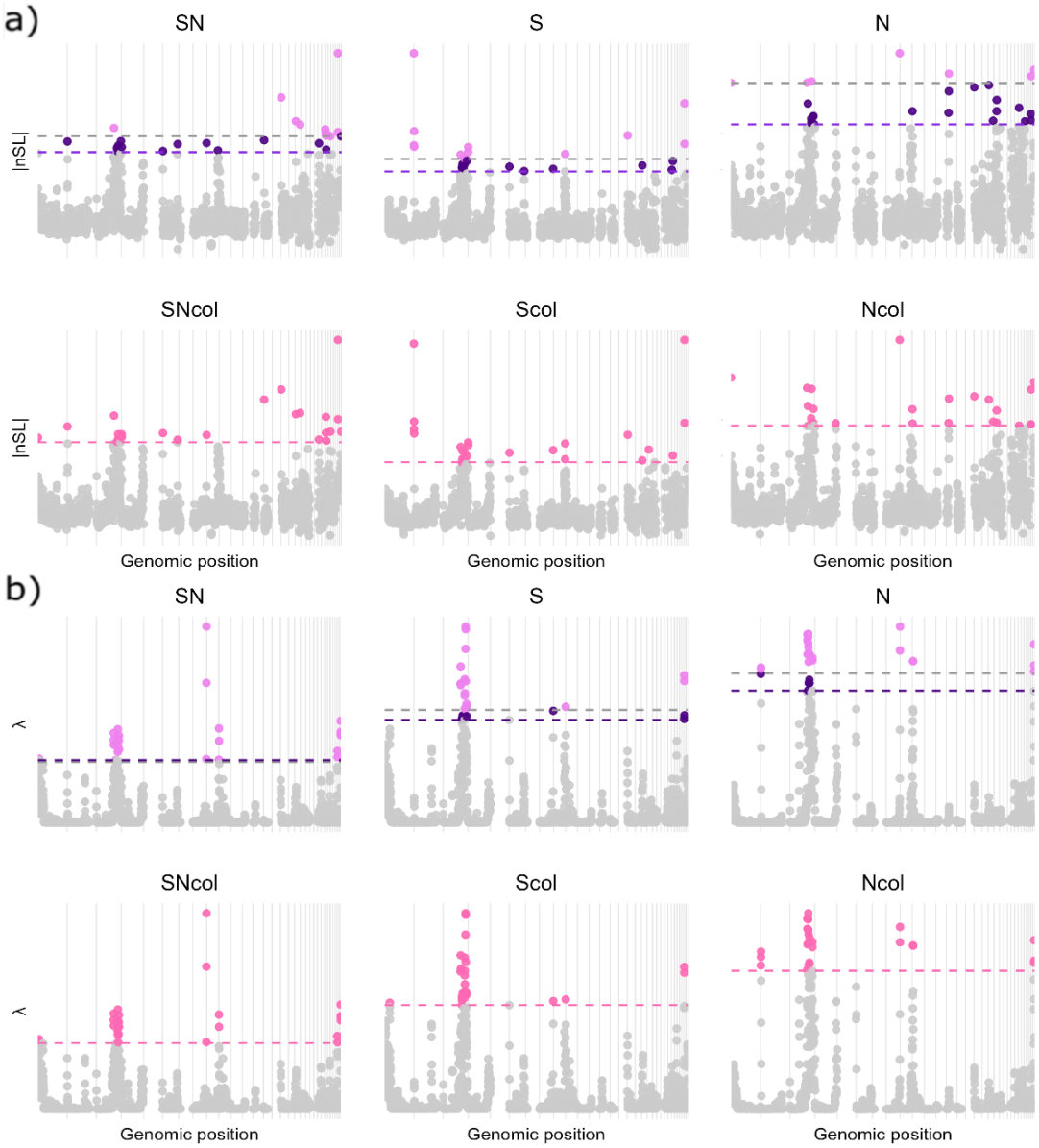


Figure 4. Comparison of the selection landscape in a collinear genome across different datasets, including or excluding inversion regions. (a) nSL selection scans results (absolute values); (b) Λ selection scans results. Only collinear regions are shown. The dashed lines represent the outlier thresholds calculated as 1% of the values in each respective dataset. SN, S, and N correspond to datasets in which selection scans were performed using the entire genome (including collinear and inverted regions). These datasets have two outlier thresholds: one including (grey) and one excluding (orange) inversions in the threshold calculation. SNcol, Scol, and Ncol represent datasets in which only collinear regions were included in the selection scans and the outlier threshold is shown in pink. Outliers values are colour-coded as follows: violet: outliers in collinear regions identified based on datasets that included inversions, and inversions were not removed before identifying the outliers; purple: outliers in collinear regions

identified based on datasets that included inversions and inversions were removed before identifying the outliers; and pink: outliers in collinear regions identified based on datasets that included only collinear regions. The visual comparison of coloured dots indicates that inversions have negligible effect on the identification of outliers in the collinear genome.

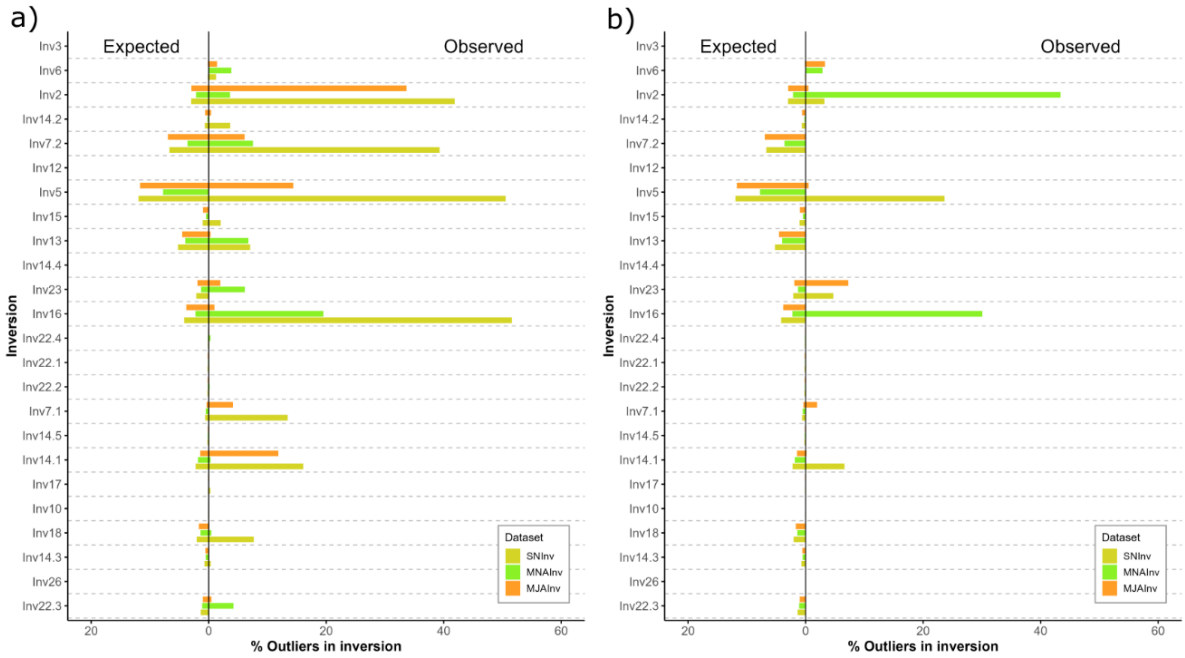


Figure 5. Comparison of the expected and the observed percentages of outliers detected in inversions. The expected values were calculated based on the length of the inversion and a 1% outlier threshold. The results are shown for nSL (a) and Λ (b). All selection scans were performed using 52 SNP windows 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).

Tables

Table 1. The comparison between Λ and nSL selection scans using different datasets and a 52-SNP window size. 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. Inversion-associated outliers: the percentage of outlier windows located within inversion regions, calculated from the total number of outliers. Collinear-region outliers: percentage of outliers located outside inversion regions calculated from the total number of outliers. Shared outliers: percentage of outliers that are shared between Λ and nSL results. Shared outliers in inversions: the percentage of outliers shared between Λ and nSL results that are in inversion regions. Shared outliers in collinear: the percentage of outliers shared between Λ and nSL results that are in the collinear genome.

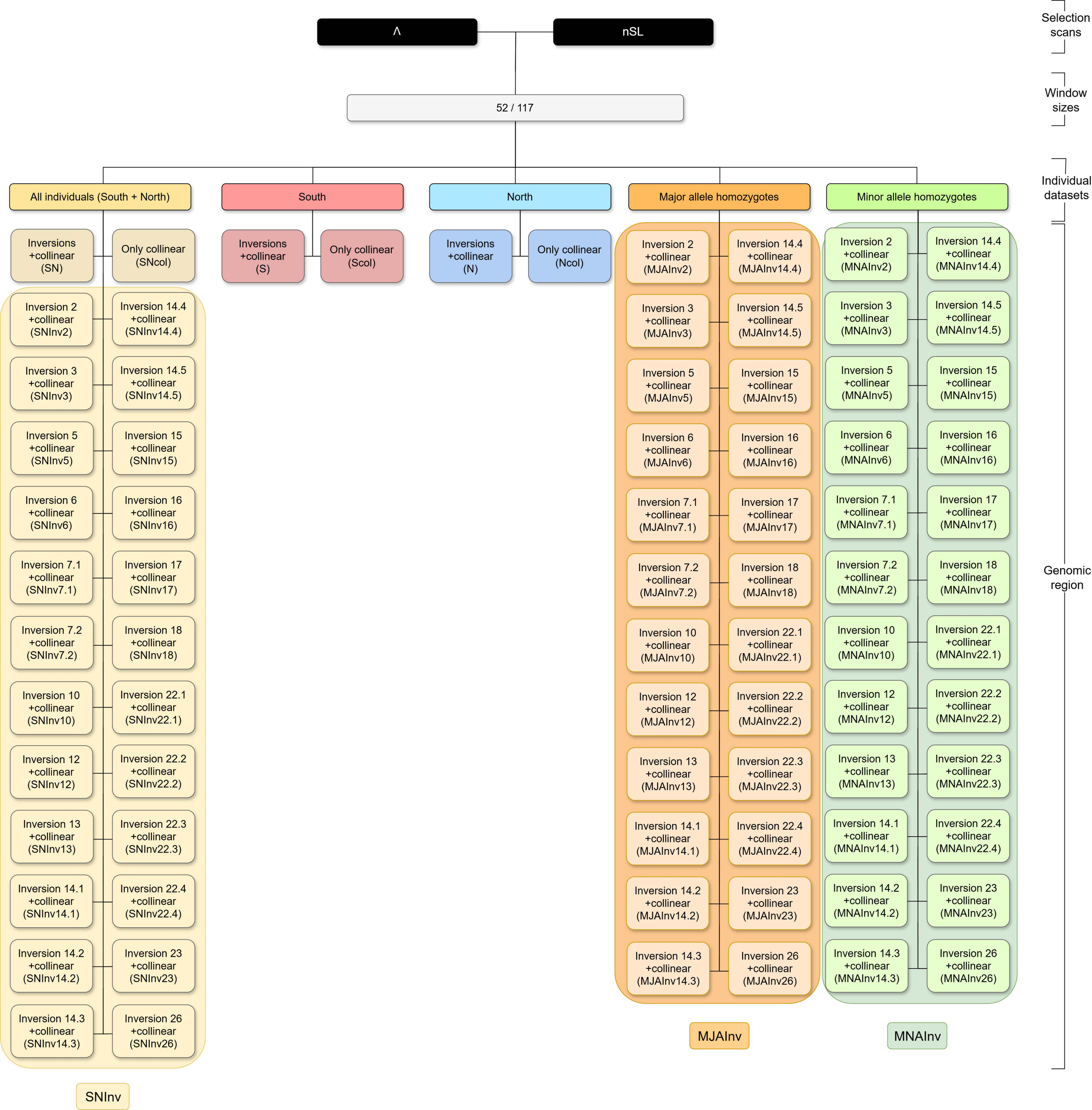
Dataset	Genomic region	Statistic	Inversion associated outliers	Collinear region outliers	Shared outliers	Shared outliers in inversions	Shared outliers in collinear
SN	Inversions + collinear	Λ nSL	29% 86%	71% 14%	15%	12%	3%
SNcol	Only collinear	Λ nSL		100% 100%	11%		11%
S	Inversions + collinear	Λ nSL	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%

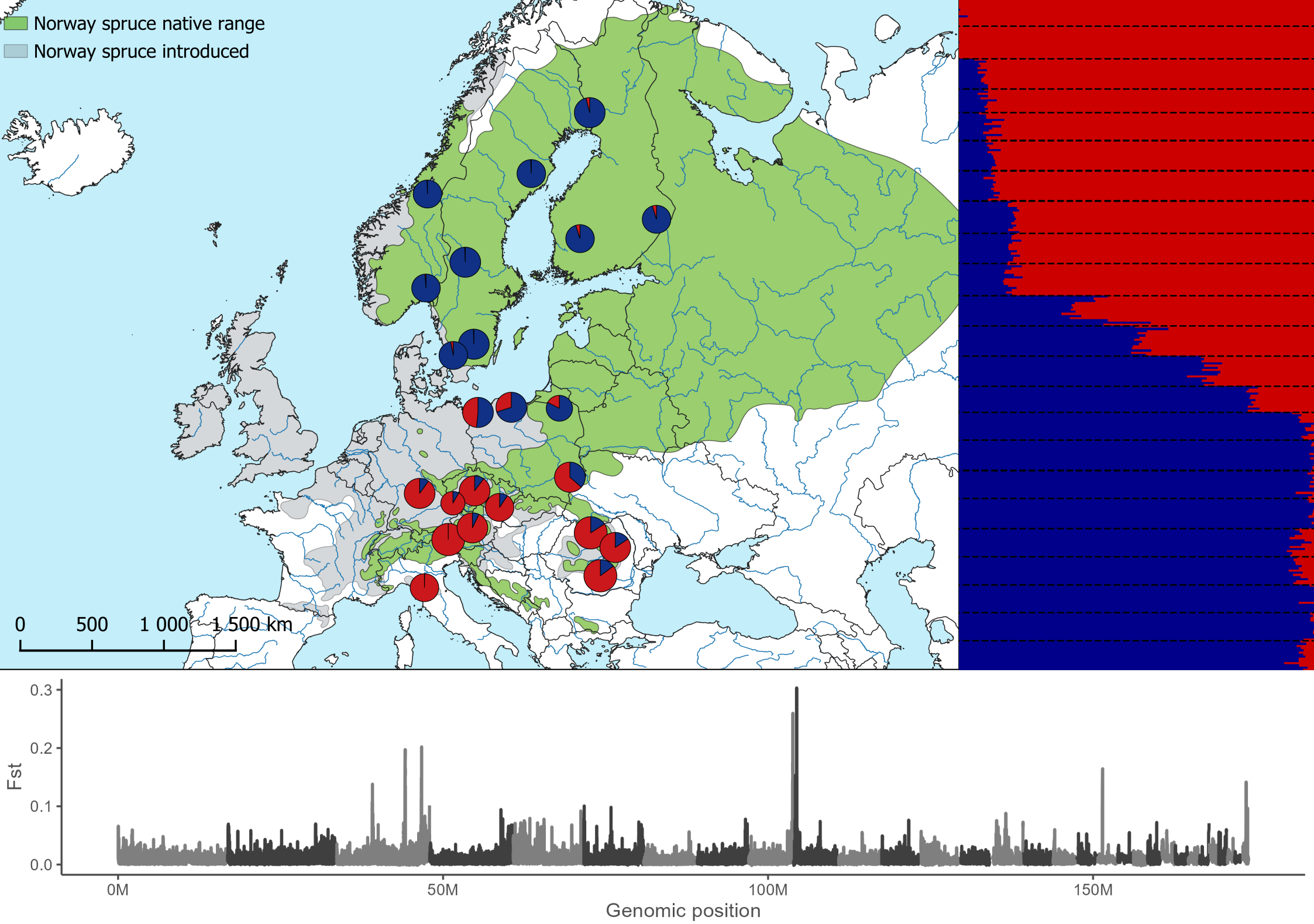
Table 2. Comparison of the results of Λ and nSL selection scans obtained using different datasets. Calculations from nSL analysis are based on the absolute values of the statistic. Genomic region: part of the genome that is included in each dataset. Statistic: selection scan which comparisons were done. 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

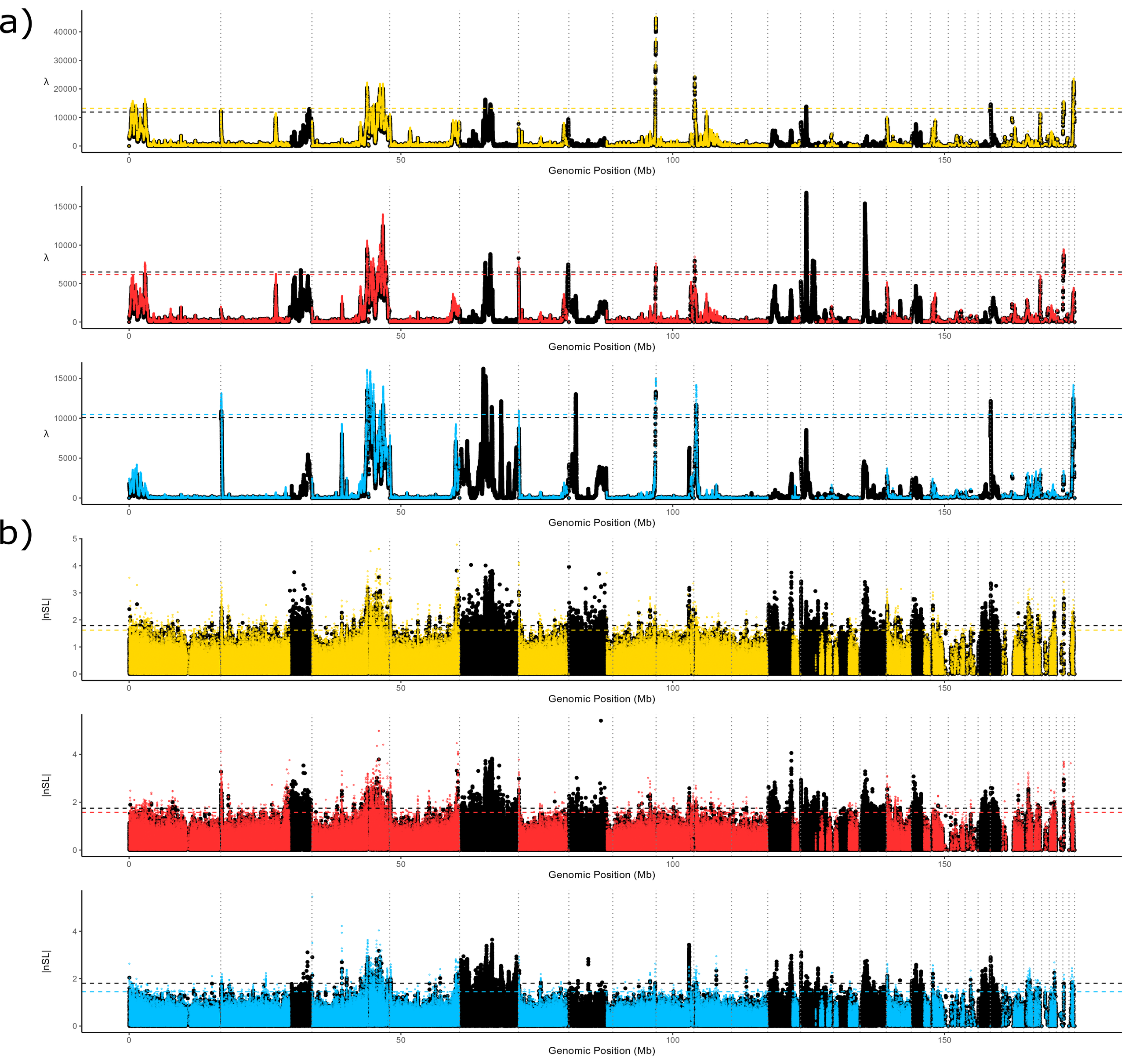
Compared datasets	Genomic region	Statistic	Shared outliers	Shared outliers in inversions	Shared outliers in collinear
SN, N	Inversions + collinear	Λ	23%	5%	18%
		nSL	33%	31%	2%
SNcol, Ncol	Only collinear	Λ	19%		19%
		nSL	32%		32%
SN, S	Inversions + collinear	Λ	30%	6%	24%
		nSL	48%	39%	9%
SNcol, Scol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

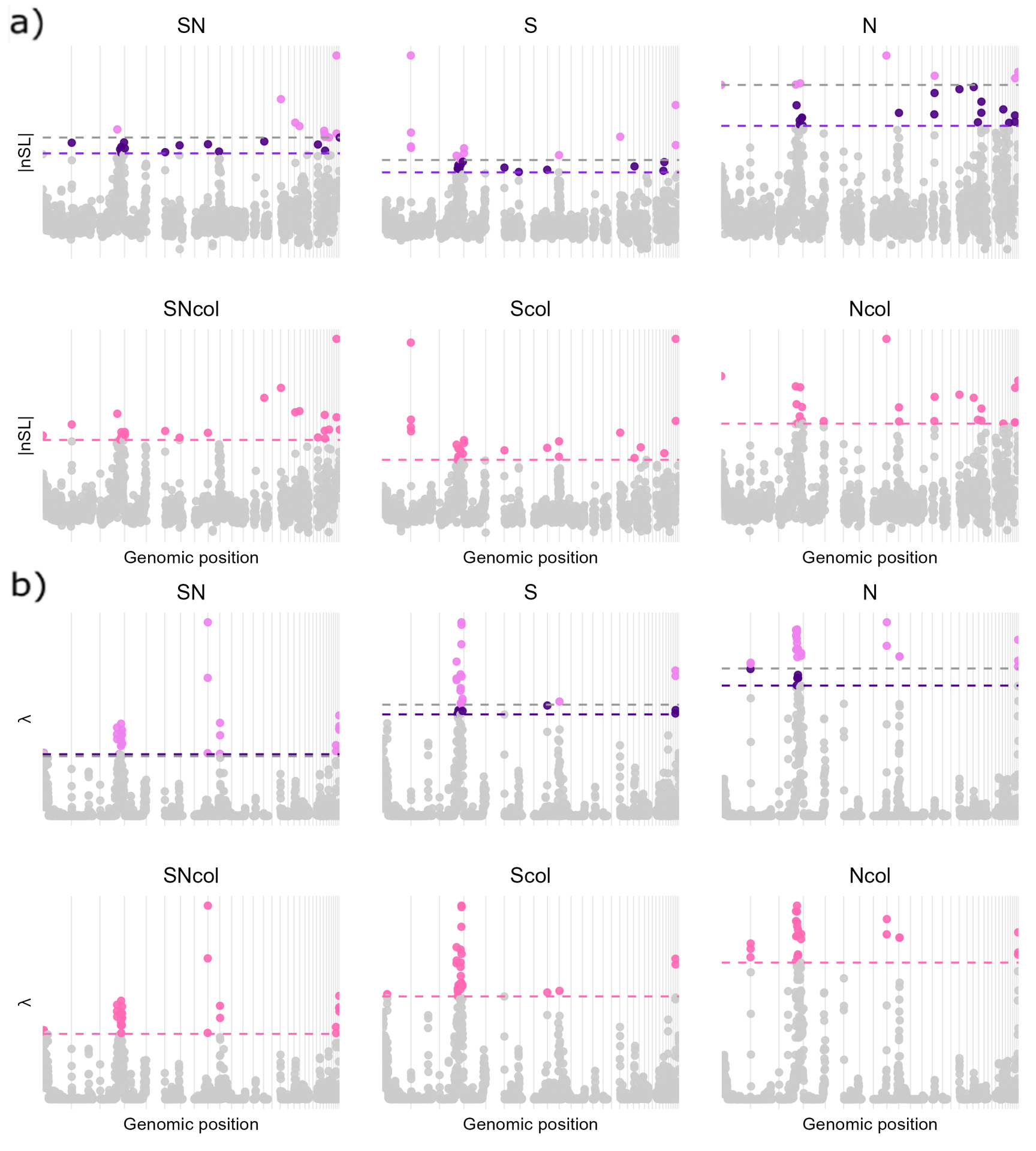
Table 3. Changes in percentage of outliers found in collinear parts of the genome when including or excluding inversions. Results are shown for three datasets (SN, N, and S) analysed using both selection scan methods (Λ and nSL) with results summarised in 50kb windows. Calculations from nSL analysis 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.

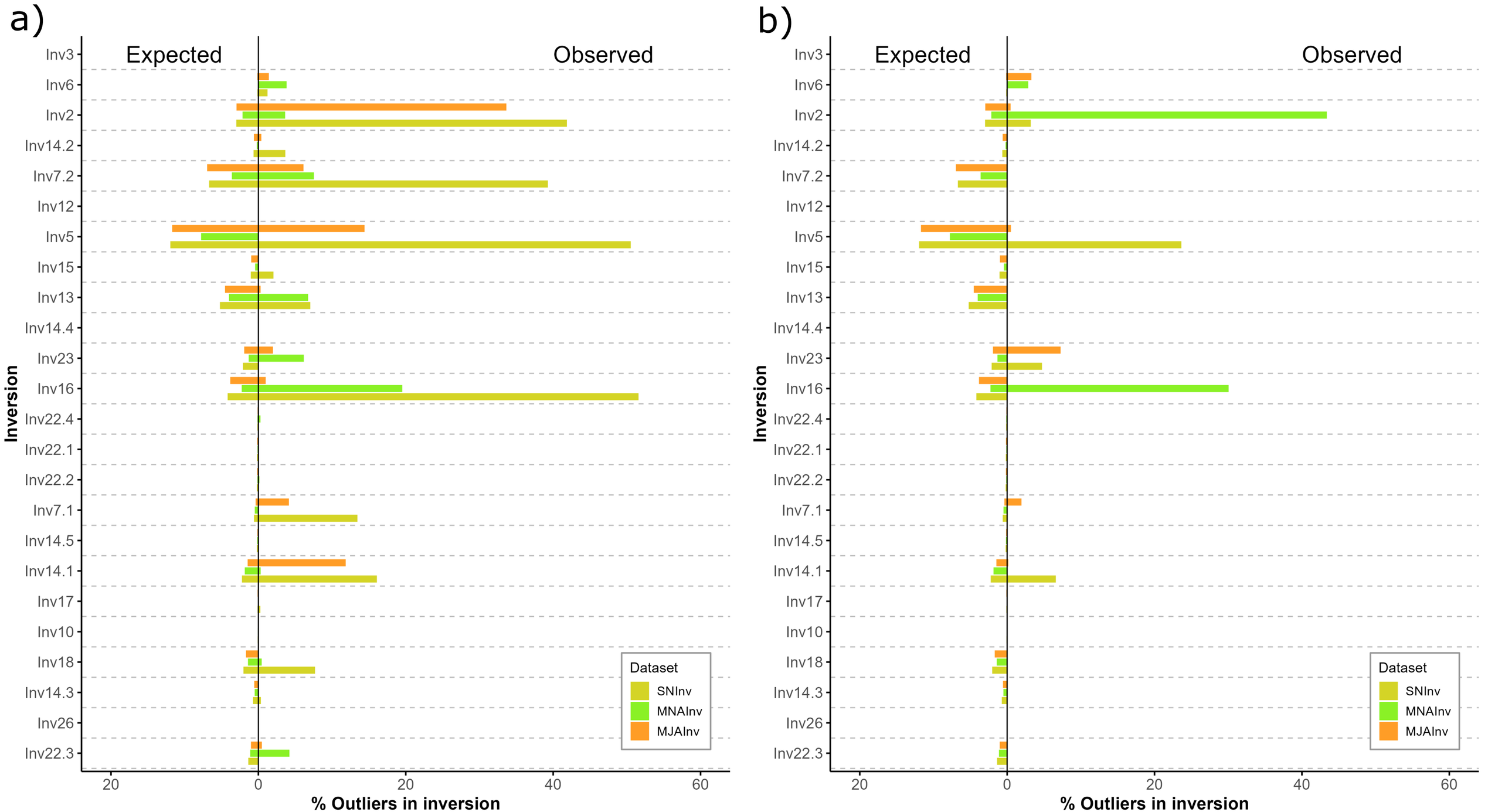
Dataset	Statistic	Outliers in collinear regions before excluding inversions	Lost outliers in collinear parts after excluding inversions	New outliers in collinear parts after excluding inversions
SN	Λ	83%	9%	0%
	nSL	30%	0%	42%
S	Λ	48%	0%	26%
	nSL	36%	0%	36%
N	Λ	60%	0%	14%
	nSL	21%	0%	53%


Λ
nSL
52 / 117
Selection scans
Window sizes
Individual datasets
Genomic region
All individuals (South + North)
South
North
Major allele homozygotes
Minor allele homozygotes
Inversions +collinear (SN)
Only collinear (SNcol)
Inversions +collinear (S)
Only collinear (Scol)
Inversions +collinear (N)
Only collinear (Ncol)
Inversion 2 +collinear (SNInv2)
Inversion 14.4 +collinear (SNInv14.4)
Inversion 3 +collinear (SNInv3)
Inversion 14.5 +collinear (SNInv14.5)
Inversion 5 +collinear (SNInv5)
Inversion 15 +collinear (SNInv15)
Inversion 6 +collinear (SNInv6)
Inversion 16 +collinear (SNInv16)
Inversion 7.1 +collinear (SNInv7.1)
Inversion 17 +collinear (SNInv17)
Inversion 7.2 +collinear (SNInv7.2)
Inversion 18 +collinear (SNInv18)
Inversion 10 +collinear (SNInv10)
Inversion 22.1 +collinear (SNInv22.1)
Inversion 12 +collinear (SNInv12)
Inversion 22.2 +collinear (SNInv22.2)
Inversion 13 +collinear (SNInv13)
Inversion 22.3 +collinear (SNInv22.3)
Inversion 14.1 +collinear (SNInv14.1)
Inversion 22.4 +collinear (SNInv22.4)
Inversion 14.2 +collinear (SNInv14.2)
Inversion 23 +collinear (SNInv23)
Inversion 14.3 +collinear (SNInv14.3)
Inversion 26 +collinear (SNInv26)
SNInv
Inversion 2 +collinear (MJAInv2)
Inversion 14.4 +collinear (MJAInv14.4)
Inversion 3 +collinear (MJAInv3)
Inversion 14.5 +collinear (MJAInv14.5)
Inversion 5 +collinear (MJAInv5)
Inversion 15 +collinear (MJAInv15)
Inversion 6 +collinear (MJAInv6)
Inversion 16 +collinear (MJAInv16)
Inversion 7.1 +collinear (MJAInv7.1)
Inversion 17 +collinear (MJAInv17)
Inversion 7.2 +collinear (MJAInv7.2)
Inversion 18 +collinear (MJAInv18)
Inversion 10 +collinear (MJAInv10)
Inversion 22.1 +collinear (MJAInv22.1)
Inversion 12 +collinear (MJAInv12)
Inversion 22.2 +collinear (MJAInv22.2)
Inversion 13 +collinear (MJAInv13)
Inversion 22.3 +collinear (MJAInv22.3)
Inversion 14.1 +collinear (MJAInv14.1)
Inversion 22.4 +collinear (MJAInv22.4)
Inversion 14.2 +collinear (MJAInv14.2)
Inversion 23 +collinear (MJAInv23)
Inversion 14.3 +collinear (MJAInv14.3)
Inversion 26 +collinear (MJAInv26)
MJAInv
Inversion 2 +collinear (MNAInv2)
Inversion 14.4 +collinear (MNAInv14.4)
Inversion 3 +collinear (MNAInv3)
Inversion 14.5 +collinear (MNAInv14.5)
Inversion 5 +collinear (MNAInv5)
Inversion 15 +collinear (MNAInv15)
Inversion 6 +collinear (MNAInv6)
Inversion 16 +collinear (MNAInv16)
Inversion 7.1 +collinear (MNAInv7.1)
Inversion 17 +collinear (MNAInv17)
Inversion 7.2 +collinear (MNAInv7.2)
Inversion 18 +collinear (MNAInv18)
Inversion 10 +collinear (MNAInv10)
Inversion 22.1 +collinear (MNAInv22.1)
Inversion 12 +collinear (MNAInv12)
Inversion 22.2 +collinear (MNAInv22.2)
Inversion 13 +collinear (MNAInv13)
Inversion 22.3 +collinear (MNAInv22.3)
Inversion 14.1 +collinear (MNAInv14.1)
Inversion 22.4 +collinear (MNAInv22.4)
Inversion 14.2 +collinear (MNAInv14.2)
Inversion 23 +collinear (MNAInv23)
Inversion 14.3 +collinear (MNAInv14.3)
Inversion 26 +collinear (MNAInv26)
MNAInv


Norway spruce native range
Norway spruce introduced
0
500
1 000
1 500 km
Fst
0.3
0.2
0.1
0.0
0M
50M
100M
150M
Genomic position


a)
λ
Genomic Position (Mb)
b)
|nSL|
Genomic Position (Mb)


a)
SN
S
N
SNcol
Scol
Ncol
|nSL|
Genomic position
b)
λ


a)
Expected
Observed
Inv3
Inv6
Inv2
Inv14.2
Inv7.2
Inv12
Inv5
Inv15
Inv13
Inv14.4
Inv23
Inv16
Inv22.4
Inv22.1
Inv22.2
Inv7.1
Inv14.5
Inv14.1
Inv17
Inv10
Inv18
Inv14.3
Inv26
Inv22.3
Inversion
20
0
20
40
60
% Outliers in inversion
Dataset
SNInv
MNAInv
MJAInv
b)
Expected
Observed
Inv3
Inv6
Inv2
Inv14.2
Inv7.2
Inv12
Inv5
Inv15
Inv13
Inv14.4
Inv23
Inv16
Inv22.4
Inv22.1
Inv22.2
Inv7.1
Inv14.5
Inv14.1
Inv17
Inv10
Inv18
Inv14.3
Inv26
Inv22.3
Inversion
20
0
20
40
60
% Outliers in inversion
Dataset
SNInv
MNAInv
MJAInv