Complex genomic landscape of inversion polymorphism in Europe's most destructive forest pest

Anastasiia Mykhailenko[1,2], Piotr Zieliński[1], Aleksandra Bednarz[1], Fredrik Schlyter[3,4], Martin N. Andersson[5], Bernardo Antunes[1], Zbigniew Borowski[6], Paal Krokene[7], Markus Melin[8], Julia Morales-García[1,2], Jörg Müller[9,10], Zuzanna Nowak[1], Martin Schebeck[11], Christian Stauffer[11], Heli Viiri[12], Wiesław Babik[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

[3]Chemical Ecology, Department of Plant Protection Biology, Swedish University of Agricultural Sciences Alnarp, PO Box 190, 234 22 Lomma, Sweden

[4] ETM, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 00, Praha, Czechia

[5]Department of Biology, Lund University, Sölvegatan 37, 223 62, Lund, Sweden

[6] Forest Research Institute, Sękocin Stary, Braci Leśnej 3, 05–090 Raszyn, Poland

[7] Division of Biotechnology and Plant Health, Norwegian Institute of Bioeconomy Research, Høgskoleveien 8, Ås, 1433, Norway

[8]Natural Resources Institute Finland, Yliopistokatu 6b, 80100 Joensuu, Finland

[9]Field Station Fabrikschleichach, Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, Glashüttenstrasse 5, 96181 Rauhenebrach, Germany

[10]Bavarian Forest National Park, Freyungerstrasse 2, 94481 Grafenau, Germany

[11]Institute of Forest Entomology, Forest Pathology and Forest Protection, Department of Forest and Soil Sciences, University of Natural Resources and Life Sciences Vienna (BOKU), Peter-Jordan-Straße 82/I, 1190, Vienna, Austria

[12]UPM-Kymmene, UPM Forest, BOX 85, 33100, Tampere, Finland

Corresponding author: Krystyna Nadachowska-Brzyska

Key words: polymorphic inversions, odorant receptors, diapause, spruce bark beetle, *Ips typographus*, forest pest

Abstract

In many species, polymorphic inversions underlie complex phenotypic polymorphisms and appear to facilitate local adaptation in the face of gene flow. Multiple polymorphic inversions can co-occur in a genome, but the prevalence, evolutionary significance, and limits to complexity of genomic inversion landscapes remain poorly understood. Here, we examine genome-wide variation in one of Europe's most destructive forest pests, the spruce bark beetle *Ips typographus*, scan for polymorphic inversions, and test whether inversions are involved in key adaptations in this species. We sampled 244 individuals from 18 populations across the species' European range and, using a whole-genome resequencing approach, identified 27 polymorphic inversions covering at least 28% of the genome. The inversions vary in size and in levels of intra-inversion recombination, are highly polymorphic across the species range, and often overlap, forming an extremely complex genomic architecture. We show that the heterogeneous inversion landscape is likely maintained by the combined action of several evolutionary forces and that inversions are enriched in odorant receptor genes encoding key elements of recognition pathways for host plants, mates, and symbiotic fungi. Our results indicate that the genome of this major forest pest of growing social, political, and economic importance harbors one of the most complex inversion landscapes described to date and is pushing the limits of genomic architecture complexity.

Introduction

The eight spined spruce bark beetle (*Ips typographus*; hereafter spruce bark beetle) is one of the most destructive insect pests in Europe, causing mass mortality in spruce-dominated forests[1,2]. The extent of recent outbreaks is unprecedented and impacts will likely increase in the coming decades in response to climate change[3–5]. Increasing bark beetle attacks and other forest disturbances have already triggered social and political conflicts in parts of Europe and have highlighted the urgent need for improved management strategies[2,6,7]. Indeed, a rapidly growing body of research focuses on the species' ecology, the causes and consequences of outbreaks, and their social aspects[2,4,8,9]. Despite this enormous interest, one aspect of the species' biology remains largely unexplored: we know almost nothing about the species' genome-wide variation and the evolutionary mechanisms that shape this variation. The lack of population genomics studies restricts our understanding of the genomic basis of adaptation and adaptive potential in

the spruce bark beetle. Such information could provide a critical missing link between applied and basic research and serve as a foundation for effective management. This is particularly important because, as discovered and described in this study, the spruce bark beetle genome harbors an extremely complex inversion polymorphism landscape that may play a critical role in many evolutionary processes, including key species adaptations[10–12].

Polymorphic chromosomal inversions are chromosomal segments that occur in two orientations within populations: collinear and inverted haplotypes/arrangements. Inversions have been shown to be involved in speciation, local adaptation, and/or maintenance of complex phenotypes[12–15]. This is due to a key property of inversions: they suppress recombination within heterozygotes and thereby prevent separation of coadapted variants. Because of their role as recombination modifiers, inversions can act as supergenes, i.e., large elements of genomic architecture with multiple linked functional elements[16]. Supergenes keep coadapted alleles together in the face of gene flow and prevent the formation of maladaptive recombinant genotypes.

Classic examples of supergenes include inversions associated with different mating strategies in ruffs (*Calidris pugnax*)[15], mimicry phenotypes in Heliconius butterflies[17], and social organization in fire ants (*Solenopsis* spp.)[18,19]. In many other species, polymorphic inversions define locally adapted ecotypes[20,21] or exhibit spatial frequency differences, e.g. by forming geographic and climatic gradients[22,23]. While the vast majority of described cases are organisms with one or a few inversions, several recent studies have reported species with many polymorphic inversions[24–26]. These recent findings raise questions about the prevalence and evolutionary significance of polymorphic inversions in natural populations. Are inversion-rich genomes the exception or the rule? How much of the genome can be situated within polymorphic inversions and, consequently, how large can the fraction of the genome with reduced recombination be? The latter question is particularly important because, in addition to suppressing recombination and keeping coadapted alleles together, inversion heterozygotes will also prevent the formation of new allelic combinations and thus reduce the efficacy of natural selection[27].

Given the importance of recombination in purging deleterious mutations[28], the presence of multiple inversions that cover a substantial fraction of the genome raises other questions. First, as the degree of recombination suppression depends on the number of heterozygous genotypes, how

frequent are inversion haplotypes in inversion-rich species? Second, are balanced polymorphisms (where both inversion arrangements are equally common across the species' range) more common than inversions with one common and one rare haplotype? Third, how common are mechanisms that mitigate negative consequences of reduced recombination? Such mechanisms include double crossover-events and gene conversion in heterozygotes, which can reshuffle allelic content between large parts of two inversion arrangements and thereby reduce mutational load and create new haplotypes[10,29].

Long-term persistence of two inversion haplotypes can be facilitated by two main types of selection: divergent and balancing[30]. Divergent selection can favor different inversion genotypes in different environments and, when coupled with reduced migration between divergent populations, can lead to speciation. Even when intraspecific gene flow is high, divergent selection can lead to divergent ecotypes associated with locally advantageous inversion genotypes. Alternatively, balancing selection may maintain balanced inversion polymorphisms over time via several, not mutually exclusive, mechanisms, such as overdominance, negative frequency dependence, antagonistic pleiotropy, and spatially or temporally varying selection[30,31]. Importantly, regardless of the selection mechanism, inversions will accumulate mutations independently since recombination is suppressed in inversion heterozygotes. This will lead to differentiated allelic content and increased differentiation between inversion haplotypes over time[30]. Each inversion haplotype can be treated as a separate "population" with a size that corresponds to the frequency of that arrangement within studied population or species. Rare inversion arrangements will suffer from a high mutational load due to reduced recombination and limited purging since homozygotes are rare. However, deleterious mutations can also accumulate on more frequent inversion haplotype[10] and lead to associative overdominance. This contributes to the maintenance of inversion polymorphisms since independent accumulation of mutations continues over the time but recessive deleterious alleles private to an inversion arrangement are not visible to selection in inversion heterozygotes.

Here, we investigated genome-wide variation across spruce bark beetle populations with a special focus on chromosomal inversion polymorphisms. We found one of the most complex polymorphic inversion architectures described to date and tested several evolutionary mechanisms, including directional and balancing selection, that can maintain inversion

polymorphisms in the genome. We also tested associations between inversion polymorphisms and the two key fitness-related traits diapause and olfaction. Our results suggest that inversions are involved in key species adaptations and raise questions about the prevalence and role of inversion polymorphisms in species characterized by large effective population sizes, little geographic subdivision, and no clear phenotypic variation across the species range.

Results and Discussion

After careful filtering of the whole-genome re-sequencing data, we used 240 individuals in downstream analyses. The mean per individual sequencing depth was 23.2× (range: 5 - 53×). Sequencing coverage on IpsContig9 was consistently lower in individuals sexed as males (on average 0.57 individual coverage). Thus, we considered IpsContig 9 to be a sex (X) chromosome. After quality filtering, we retained 5.245 million SNPs covering the entire genome assembly but analyzed a subset of 5.067 million SNPs located on the 35 longest contigs and representing 75% of the genome assembly.

Complex genomic inversion landscape

Twenty-nine candidate inversions ("inversions" henceforth) were identified following the criteria described in the Methods section (Table 1, Figure 1, Figure S1, and Figure S2). Two possible inversions (Inv16.1 and Inv23.1) were most likely part of the same inversion: the same beetle individuals were genotyped as both homo- and heterozygotes, which is unlikely to occur for two unlinked inversions. The same situation was found for Inv16.2 and Inv23.2. No other inversions were in strong linkage disequilibrium (LD) with each other, which would suggest co-segregation (Figure S3). Thus, overall, we found 27 inversions in 17 contigs, including one located on the X chromosome (Inv9). Inversion size varied from 0.1 to 10.8 Mb (Table 1) and inversions constituted 28% of the analyzed part of the genome. Estimated inversion age ranged from 0.5 to 2.6 My (Table 1, assuming a mutation rate of 2.9×10^{-9}; see Table S1 for results for different mutation rates). Inversion regions exhibited a reduced population recombination rate (rho, Figure S4) and moderate to high genetic differentiation between inversion arrangements (F_{ST} between homozygous individuals was 0.15-0.64; Figure 1; Figure S1).

While 12 contigs contained single inversions, five contigs showed patterns consistent with multiple adjacent or overlapping inversions (Figure 1, Figure S1, Figure S2). IpsContig14 and IpsContig22 contained complexes of multiple adjacent and sometimes overlapping inversions (Figure 1 j-o), and three other contigs contained two overlapping inversions each (IpsContigs: 7, 16 and 23, Figure S1). Additionally, four putative double crossover-events were identified within four inversions (Inv5, Inv18, Inv22.3, Inv22.4; Figure 1, Figure S1, Figure S2). In these cases, LD clusters were separated by regions of lower LD and intermediate groups of individuals were visible between the main clusters along the first principal component.

It was not surprising that we detected polymorphic inversions in the spruce bark beetle genome, as there are many well-known examples of polymorphic inversions in natural populations[12]. What was striking, however, was the extremely complex genomic landscape of polymorphic inversions we found in this species. The spruce bark beetle has at least 27 large inversions covering a substantial part of the genome (28% of the analyzed part of the genome, but this is probably an underestimate since we focused on ca. 75% of the genome and on large inversions > 0.1 Mb only). Numerous (a dozen or more) polymorphic inversions have so far only been described for a few species (e.g.[26,32,33]). It is still an open question whether many polymorphic inversions within species is the exception or the rule.

The exceptionally complex inversion architecture we found in the spruce bark beetle, with multiple adjacent and often overlapping inversions, resembles well known examples from *Heliconius* butterflies[34] or fire ants[35]. In these insects multiple adjacent inversions are the basis for mimicry phenotypes and complex social organization, respectively. The presence of clusters of adjacent inversions and inversion overlaps are consistent with theoretical expectations of stepwise extension of recombination suppression on supergenes[36] and with a highly polygenic architecture of adaptation[37].

Genome-wide variation and its geographic structuring – collinear vs. inversion regions

Analyses based on both the whole genome and collinear parts only revealed a clear latitudinal structuring of genetic variation in the spruce bark beetle (Figure 2). PCA and NGSadmix supported the presence of two distinct genetic groups corresponding to southern and northern

populations, with Polish populations showing varying degrees of admixture between the two clusters. Based on these results, we divided the 18 studied populations into a northern, southern, and Polish group. Despite unambiguous NGSadmix-division into two genetic clusters, the genome-wide genetic differentiation between the northern and southern group was extremely low (F_{ST} = 0.021). Similarly, F_{ST} between all population pairs showed low levels of differentiation, ranging from 0.000 to 0.035 (Table S2). Mean genome-wide nucleotide diversity was moderate (π = 0.0062) and per population π ranged from 0.0055 to 0.0066 (Figure S5). There was a weak negative correlation between nucleotide diversity and latitude (r^2 = 0.32, p = 0.033, Figure S6), as northern populations had slightly lower genetic variation than southern populations ($\pi_{southern}$ = 0.0065; $\pi_{northern}$ = 0.0061; π_{Polish} = 0.0066). Southern populations had an excess of rare alleles and, consequently, had more negative Tajima's D values along the genome than northern populations (mean Tajima's D was -0.458 and -0.062 in the southern and northern group, respectively; Figure S7). All these results are consistent with previous phylogeographic studies of the spruce bark beetle that analyzed a much smaller number of genetic markers. The data suggests high levels of connectivity among spruce bark beetle populations and a very recent differentiation into two genetic clusters[38–41]. More recent RADseq data confirms a very weak genetic structure in the spruce bark beetle across much of Sweden[42], as is expected in a species with high dispersal[43] and recent divergence.

Inversion regions in the spruce bark beetle did not structure geographically into a southern and northern group. Almost all identified inversions were polymorphic across the European species range, except for one inversion (Inv9) that was polymorphic only within northern populations. For three inversions (Inv2, Inv5, Inv16.2+Inv23.2) unambiguous genotyping was only possible across part of the species range. The differentiation between inversion haplotypes was high, suggesting long-term persistence of inversion polymorphisms within this species (Figure 1; Figure S1). According to our age estimates, the origin of the inversions in the spruce bark beetle predates the Last Glacial Period and many inversions may be several million years old, likely also predating the within-species differentiation into a southern and northern group. This was not unexpected, as many known inversion polymorphisms have been segregating within species for hundreds of thousands or millions of years (see table in[12]), sometimes even persisting through multiple speciation events[44].

Inversions and key fitness-related traits

Inversion polymorphism is often associated with the maintenance of complex polymorphic phenotypes[15,45]. Although the spruce bark beetle does not exhibit easily identifiable phenotypes, such as distinct color patterns or mating types, we were able to test for associations between inversions and two complex traits of key importance for many insects: diapause and olfaction. Diapause allows species to suspend development during unfavorable conditions. While multiple environmental factors can influence this complex process, many aspects of diapause, such as its induction and termination, are heritable[46] and may be controlled by a small number of loci or be highly polygenic[47–50]. The spruce bark beetle exhibits two diapause strategies that could be associated with polymorphic inversions: a facultative photoperiod-regulated diapause and an obligate photoperiod-independent diapause[51]. However, we found no association between diapause phenotypes and inversion genotypes (exact G test, Table S3, Figure S8), nor did we find any highly differentiated genomic regions between facultatively and obligately diapausing individuals (Figure S9). This suggests a polygenic nature of diapause phenotypes in the spruce bark beetle.

Olfaction is another key fitness-related trait in many insects, including bark beetles. Insect odorant receptors (ORs) are encoded by a large and dynamically evolving gene family. Some of the receptors are evolutionarily conserved across species within insect orders, however, many are species- or genus-specific. In the spruce bark beetle, detection of odorants is essential for host and mate finding, as well as recognition and maintenance of symbiosis with specific fungi[52,53]. We examined 73 antennally expressed ORs[54] and found that 46% of these were located within inverted regions (Table 1), even though inverted regions constituted only 28% of the analyzed part of the genome. A permutation test confirmed this over-representation of OR genes within inversions ($p = 0.03$). In addition, several *Ips*-specific ORs (5 out of 7 ORs from an *Ips*-specific OR clade[55]) were located in inverted regions, specifically on IpsContig13 (4 out of 7 ORs). These 5 ORs (ItypOR23, ItypOR27, ItypOR28, ItypOR29, and ItypOR49) have been functionally characterized, responding to selectively compounds primarily produced by beetles (pheromones), the host tree, or fungal symbionts, respectively (Powell et al., 2021; Hou et al.,

2021). We found no difference in the OR composition (no OR deletions) between inversion haplotypes.

Two inversions harboring multiple OR genes showed significant latitudinal variation (Figure 3; Inv13 and Inv16.1+23.1, Table 1). Interestingly, one of the inversion, Inv13, include a gene encoding ItypOR23, receptor that has been previously shown to primarily respond to an odor from fungi[55]. We hypothesize that different OR alleles associated with these inversions are involved in spruce bark beetle interactions with fungal associates present in different parts of Europe. Not only are spruce bark beetle populations exposed to different fungal species throughout the beetle's range, but individual beetles may also have preferences for different fungal species[53,56]. German spruce bark beetles have, for example, been shown to be more attracted to fungal species that are common in Germany (*Grosmannia penicillata*, *Endoconidiophora polonica*) than to rarer species (*Leptographium europhioides*). Unpublished data from Swedish beetles suggests that they are more attracted to *L. europhioides*, which is common in Sweden (personal communication D. Kandasamy). Although preliminary, these observations suggest that beetle preferences may be tuned to the local fungal flora and we speculate that inversions may be involved in recognition of region-specific fungal species.

Several other interesting behavioral strategies are polymorphic among spruce bark beetle individuals, including the existence of pioneer individuals that are the first to infest host trees, and re-emergence of females after egg laying to establish so-called sister broods in new trees. Other less obvious/visible phenotypes could also be associated with inversion polymorphisms. Comprehensive research, both bottom-up and top-down, is needed to understand the relationship between spruce bark beetle phenotypes and the inversion polymorphism landscape.

Evolutionary mechanisms maintaining inversion polymorphism in the spruce bark beetle

Several non-mutually exclusive evolutionary processes can maintain polymorphic inversions within species, particularly divergent and balancing selection[12,30]. The importance of divergent selection has been postulated based on allele frequency patterns and associations of polymorphic inversions with local adaptations that persist despite extensive intraspecific gene flow[57]. For example, a recent study of deer mice (*Peromyscus maniculatus*)[26] identified multiple polymorphic inversions with clinal variation across environmental gradients in two distinct

habitats. Such frequency changes have also been reported across hybrid zones[33] or latitudinal gradients[23]. Although the spruce bark beetle does not occupy distinct environmental niches, it inhabits forests across a very wide latitudinal gradient (spanning at least 16 degrees). It is also a species with high dispersal capacity and extensive gene flow, as indicated by low F_{ST} across its range. We found a significant correlation between the frequency of inversion haplotypes in populations and geographic location (latitude) for five inversions (Figure 3; r^2 ranged from 0.34 to 0.68). There were no significant correlations between haplotype frequencies and longitude (Figure S10). In most cases, differences in inversion haplotype frequencies were small (Figure 3), except for the two inversions with the strongest correlations ($r^2 > 0.6$; Inv12 and Inv13). We found no association between inversion genotypes and climate or land cover variables (Figure 4), but several SNPs showed significant correlation with the first principal component in a PCA of many environmental variables. These results indicated that there probably is selection across environmental gradients in the spruce bark beetle but that this is not the only, or even a major, force maintaining inversion polymorphism within the species.

While the 'local adaptation' hypothesis is a major hypothesis proposed to explain inversion polymorphism[26,58,59], balancing selection and related mechanisms may also be important in maintaining polymorphic arrangements[10,30,60]. Such mechanisms include overdominance, associative overdominance, frequency-dependent selection, and spatially and temporally varying selection. We found no support for overdominance playing a role in the spruce bark beetle, as no excess of inversion heterozygotes was detected in any of the populations, geographic regions or across the whole species range. We therefore looked more closely at mutation load, which can say something about the role of associative overdominance in the maintenance of inversion polymorphism. Theory predicts that recessive deleterious mutations will accumulate on both inversion arrangements but that most of these mutations will be private to only one arrangement[10,30,61]. This would lead to associative overdominance, as in heterozygotes the effects of deleterious recessive alleles on one arrangement would be masked by the wild-type alleles on the other arrangement. The result would be long-term maintenance of the inversion polymorphism, resulting in strong divergence between inversion haplotypes[10,61,62].

Interestingly, several stable evolutionary scenarios that maintain polymorphic inversions are possible (for details see Figure 4 in[10]). These scenarios differ in the expected mutation load, fitness, and frequency of the corresponding genotypes. Given the haplotype frequencies we observed in spruce bark beetle inversions, two scenarios are likely. First, that minority arrangements suffer from higher mutation load (due to reduced recombination and lower population size) but are maintained in the population at low frequency due to, e.g., associative overdominance. Such a mechanism would favor balanced inversion polymorphisms of intermediate to large sizes[63,64] and has been shown to play a role in maintaining polymorphic inversions in several insect species[34,65]. Second, mutation load may accumulate on one or both inversion arrangements but be mitigated by the haplotype structuring process, i.e., the existence of multiple diverged sub-haplotypes among inversion homozygotes that reduces the mutation load within homozygotes. If this process operates within one or both inversion haplotypes it may result in more equal frequencies of alternative inversion haplotypes. However, such a mechanism is only possible when genetic variation and mutation load is high[10].

In contrast to these theoretical expectations, we observed no sign of increased mutation load (measured by the π_N/π_S ratio) in inversion regions compared to the collinear part of the spruce bark beetle genome (Table S4; Figure 5). We also found no sign of haplotype structuring capable of reducing mutation load within inversion homozygotes (Figure S11). The only significant within-homozygote clustering we observed was in a few inversion haplotypes, divided individuals into southern and northern clades and suggests that divergent selection has been acting on one of the inversion arrangements, rather than haplotype structuring being associated with mutation load (Figure S11). These results are consistent with observations of no significant mutation load in other species where inversion haplotypes are subject to divergent selection that results in geographic structuring[26,66]. In such cases, inversions facilitate adaptive divergence but do not tend to accumulate a mutation load. However, geographic clustering of inversion polymorphisms is weak in the spruce bark beetle, and many inversions appear to have a slightly lower mutation load associated with the more common inversion haplotype (Figure 5, Table S4). It is possible that accumulation of a mutation load in the spruce bark beetle is mitigated by high effective population size in this species and/or gene conversion and double crossover events, which despite their apparent rarity have been detected in several spruce bark beetle inversions.

Overall, the absence of heterozygote excess observed in the spruce bark beetle does not support a role of overdominance in the maintenance of inversion polymorphism in this species. Likewise, the absence of an elevated mutation load in inverted regions does not support a role of associative overdominance either. However, since we only have genomic data available, we cannot conclusively rule out that other potential mechanisms, such as negative frequency-dependent selection or antagonistic pleiotropy, could maintain balanced inversion polymorphisms. Additional temporal data are needed to test whether temporally varying selection has affected the frequencies of inversion haplotypes in the spruce bark beetle. Our results indicate that inversions in this species are maintained as polymorphic by a complex interaction of different, not mutually exclusive mechanisms. Further research is essential to determine the role of different mechanisms.

Far-reaching consequences of having an inversion-rich genome

The presence of multiple polymorphic inversions can have significant consequences for the evolution of a species, as well as for evolutionary inferences based on genome-wide polymorphism data. Importantly, polymorphic inversions are a reservoir of genetic variation that can facilitate adaptation to rapidly changing environments. Indeed, several studies have shown that polymorphic inversions support rapid adaptation to changing climatic conditions[67–69] or adaptive tracking of fluctuating environments[70]. Spruce bark beetle populations are subject to seasonal weather changes and a rapidly changing environment due to strong anthropogenic pressures. Warmer weather and drought periods have been associated with a predicted intensification of bark beetle outbreaks[3,4,6], which may act as a strong selection factor within bark beetle populations. Whether inversions are involved in rapid adaptations in the spruce bark beetle is an open question that requires further investigation.

Abundant polymorphic inversions within the genome can have far-reaching consequences for inferences about demographic history and selection. It is well known that non-equilibrium demography and selection can leave similar genomic signatures. Traditionally, demographic analyses have used non-coding parts of the genome, based on the assumption that directional selection mostly affects protein-coding regions. However, growing evidence for the importance

of background selection in shaping genome-wide diversity is moving the field towards incorporating linked selection into inferences of demographic history[71,72]. We believe that new approaches should also consider the potential influence of polymorphic inversion landscapes, because variation patterns of inverted regions can be shaped by different types of balancing selection. In addition, genomics scans for selection in inversion-rich genomes may be biased due to reduced recombination within inversion regions. Importantly, the effect of reduced recombination may extend outside inversions[73,74].

Materials and methods

Study system

The spruce bark beetle, *Ips typographus* (Coleoptera: Curculionidae: Scolytinae), plays a key role in Eurasian forest ecosystems. Under normal (endemic) conditions, this forest pest attacks mainly weakened Norway spruce (*Picea abies*) trees. However, if tree resistance is compromised by certain abiotic disturbances (e.g. snowbreaks, windfalls, high temperatures, drought), an increased availability of stressed trees can trigger mass-propagation, leading to rapid population increase and devastating outbreaks. During outbreaks, the beetles attack healthy trees and cause massive tree mortality with hundreds or thousands of hectares of dead spruce stands.

During the first decade of the 21st century the spruce bark beetle killed an estimated 14.5 million m^3 of timber per year on average, and this number is expected to increase due to climate change[75]. The Czech Republic provides a particularly striking example of the beetles' destructive potential. During the peak outbreak years 2017-2019 the beetles killed annually 3.1-5.4% of the country's growing stock of Norway spruce, which in 2019 translated to 23 million m^3. In some regions there was an almost total depletion of Norway spruce[76]. From a socio-economic perspective, bark beetle outbreaks and subsequent intensive salvage cuttings negatively affects the quality of life for people living in outbreak areas and cause serious disturbances to the wood market. In the Czech Republic, decreasing timber prices caused severe revenue losses for forest owners and required state interventions in the amount of 260 million EUR in 2018-2019[76].

Historically, spruce forests in Central Europe have been most heavily affected by bark beetle outbreaks, while in northern Europe outbreaks have been less frequent and destructive[1]. However, this may change with climate warming that probably will make the boreal forests of northern Europe more vulnerable to bark beetle outbreaks[3,5,7,77,78]. As an example, heatwaves and severe summer drought in Sweden in 2018 initiated a bark beetle outbreak killing over 30 million m^3 Norway spruce the next years[79,80].

Sampling

Adult spruce bark beetles were collected with pheromone-baited traps in the spring and summer of 2020. In total, we sampled 18 populations throughout Europe with 13-14 individuals per locality (244 individuals in total) (Figure 2; Table S5). Throughout the text, we use the term 'population' to refer to a particular site or a collection of closely situated sites (within about 50 km). In Austria, we pooled individuals from three localities that were up to 120 km apart because of small sample sizes (five beetles or less per site). Populations from Scandinavia will be referred to as northern populations (or the northern group) and populations from central Europe will be referred to as southern populations (or the southern group). Polish populations are considered separately from other central European populations (due to high admixture proportions from northern group identified in downstream analysis). Beetles were brought alive to the laboratory, kept on a paper diet for several days, dissected, sexed based on genitalia morphology, and subjected to DNA extraction (described below).

DNA extraction and genome re-sequencing

DNA was extracted from the whole body of dissected beetles using the Wizard Genomic DNA Purification Kit (Promega). The concentration of extracted DNA was estimated using a Qubit fluorometer (Thermo Fisher Scientific). Genomic libraries were prepared with NEBNext Ultra II FS DNA Library Prep with Beads (New England Biolabs), with single indexes. Individual libraries were combined into three pools and 2×150 bp paired-end sequenced in three lanes of a S4 flowcell using the NovaSeq 6000 instrument and v1 sequencing chemistry (Illumina Inc.). Sequencing was done by the National Genomics Infrastructure, SNP&SEQ Technology Platform (Uppsala, Sweden). To assess the overall genotyping error, we prepared and sequenced duplicate libraries for nine individuals.

Data preparation and filtering

Details of raw data processing and filtering are described in the Supplementary Files. Shortly, raw reads were mapped to the reference genome[81] using Bowtie 2[82]. Duplicated reads were removed using Picard MarkDuplicates (Broad Institute 2019). To detect and correct systematic errors in base quality scores, recalibration was done using the Genome Analysis Toolkit (GATK), BaseRecalibrator, and ApplyBQSR[83,84]. Variant calling and genotyping was done using GATK HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs. GATK VariantRecalibrator and ApplyVQSR were used to calculate and filter (by variant) quality score log-odds (VQSLOD). Bcftools[85] was used to remove insertions and deletions (indels) as well polymorphisms five bases up- and downstream. GATK VariantFiltration was applied to mask all genotypes with low sequencing depth or low genotype quality[83,84]. Variants which were not biallelic single nucleotide polymorphisms or did not meet the recommended hard filtering thresholds (GATK Team, see Supplementary materials) were filtered out. To filter out polymorphisms that could come from duplicated regions we removed variants located within reapeat-masked regions of the genome[81], variants with excessive overall coverage, and variants with heterozygote excess. Variants for which genotypes could be detected in less than half of the individuals were also removed. We used PLINK[86] to detect sample contaminations, swaps and duplications, and unknown familial relationships (e.g. sibling pairs present in the data) which might bias downstream analyses. Individuals with excessive coverage were removed, as these could be a result of human errors during library preparation or pooling. We focused on contigs longer than 1 Mb that together constituted 78% of the genome assembly, i.e. a total of 186 Mb. Since part of IpsContig33 had high similarity to mtDNA this contig was not included in the downstream analysis. Genotyping error was assessed using GATK Genotype Concordance.

Genome-wide genetic variation and its geographic structuring

Genome-wide genetic structuring was explored by PCA using PLINK. The most likely number of genetic clusters and admixture proportions was estimated using NGSadmix[87]. The analysis was run for five different K-values (1-5; 10 replicates per K-value), using a minor allele frequency (MAF) filter of 0.05 and 10,000 iterations, and a SNP dataset that was pruned for linkage disequilibrium using PLINK (–indep-pairwise 50 10 0.1 option). To choose the most likely number of genetic clusters, the results were examined using CLUMPAK

(http://clumpak.tau.ac.il/index.html). To examine the influence of inversions on genetic clustering and to facilitate inversion genotyping, NGSadmix was run separately for (1) all autosomal contigs without potential inversions and (2) each potential inversion.

To assess genetic differentiation among different spruce bark beetle populations Weir and Cockerham's[88] F_{ST} was estimated using VCFtools[89]. F_{ST} was calculated between population groups identified by NGSadmix, as well as among all population pairs. Additionally, we summarized F_{ST} values in 100 kb overlapping windows (using 20 kb steps) along the contigs. Window-based analyses were done for each contig and each population pair. In addition, absolute sequence divergence (d_{xy}), nucleotide diversity, and Tajima's D statistic were estimated and summarized for 100 kb non-overlapping windows using ANGSD[90]. These statistics were calculated for each population separately and were based on a maximum likelihood estimate of the folded site frequency spectrum (SFS). We excluded sites with mapping quality below 30 phred, quality score below 20, and coverage less than three times the population sample size and more than three times the average coverage, following the approach used in Delmore et al.[91]. The ANGSD calculations were based on allele frequencies estimated from genotype likelihoods[92] and ngsPopGen scripts (https://github.com/mfumagalli/ngsPopGen).

Population recombination rates were estimated following the approach used in Jones et al.[93], using 20 individuals from the southern and northern group (10 individuals per group). Watterson theta estimates were used to create a custom likelihood lookup table using the Ldhat program *complete*[94]. The *Interval* program was used to estimate the population recombination rate across investigated contigs in 1 Mb segments. The *interval* algorithm was run for 2 million iterations and the chain was sampled every 10,000 iterations with a burn-in of 100,000 generations (*stat* program in Ldhat package, block penalty = 5). Population recombination rates were estimated for each contig, and the results were summarized in non-overlapping 100 kb windows using a custom perl script.

Identification of inversions, their geographic distribution, and variation patterns

Potential chromosomal inversion regions were identified based on the results of per contig PCAs, local PCAs, patterns of heterozygosity, and LD clustering. Per contig PCAs were performed

using PLINK[86]. Local PCAs were performed using the *lostruct* R package[95], following the approach described in Huang et al.[96]. Linkage disequilibrium among SNPs (thinned by selecting one SNP every 10 kb; MAF > 5%) was calculated for each contig using PLINK. We considered a genomic region to be an inversion region if (1) local PCA analysis identified the region as an outlier, and/or (2) the region exhibited high LD (most SNPs having $r^2 > 0.4$), and (3) PCA performed on SNPs from this region separated individuals into two (or more) distinct groups with heterozygosity patterns matching the expectation of at least one group having lower levels of heterozygosity (i.e., the group consisting of individuals with homozygous inversion). Genotyping of individual beetles with respect to the inversion haplotypes they carried was done based on inversion region PCA1 loading scores and/or inversion-specific NGSadmix clustering (with K = 2 inversion heterozygotes having mixed ancestry in approximate 50/50 proportions; Figure S12). In a few special cases (such as overlapping inversions or complex double crossover patterns) genotyping was done based on inversion region PCA2 loading scores or was limited to either the southern or northern group. Contigs with less than 10,000 variants provided ambiguous results and were excluded from genotyping. Putative inversion boundaries were defined based on local PCAs and sharp borders detected in LD clusters.

Inversion genotype and haplotype frequencies were calculated using an in-house R-script. Frequencies were calculated (1) within each population, (2) within the southern and northern group, and (3) for all sampled populations combined. Deviations from Hardy-Weinberg equilibrium were estimated for all three datasets. Inversions with only two haplotypes were tested using exact Fisher tests. Inversions with more than two haplotypes (including recombinant haplotypes between two inversion arrangements) were tested using permutation test, and sex chromosome inversions were tested as described in Graffelman & Weir[97]. All tests were done using the R package *HardyWeinberg*[98]. To investigate if inversion haplotypes differed in frequency along environmental gradients, Pearson correlation coefficients between inversion haplotype frequencies and latitude/longitude was calculated.

To assess levels of genetic differentiation between inversion haplotypes, F_{ST} and d_{xy} between alternative inversion haplotypes (AA, BB) were estimated following the approach described

above. Deletions present in alternative inversion haplotypes were not included in d_{xy} calculations but were summarized separately using custom scripts.

Inversion age estimates

Absolute sequence divergence between alternative inversion haplotypes was used to calculate the approximate time of divergence of inverted and non-inverted haplotypes. We used the equation $T = d_{xy}/2\mu$, where T is the divergence time in generations, μ is the mutation rate per site per generation, and d_{xy} is a mean d_{xy} calculated based on per SNP values estimated in ANGSD. Since mutation rates of the spruce bark beetle are unknown, we used a selection of mutation rate estimates available for some diploid, sexually reproducing insects including *Drosophila melanogaster*[99], *Heliconius melpomene*[100], and *Chironomus riparius*[101]. Per generation mutation rate estimates varied from 2.1 to 11.7×10^{-9}. This approach could only give us rough inversion age estimates due to the uncertainty of the mutation rate estimates, probable intraspecific variation in mutation rate[99], and a (likely) substantial influence of gene flux[29].

Mutation load estimation

To estimate mutation load we calculated the ratio of nucleotide diversity at non-synonymous sites (π_N) vs. synonymous sites (π_S). Mutation load (π_N/π_S) was calculated separately for each inversion homozygote and for the collinear part of the genome. We computed nucleotide diversity for each site using SNPGenie[102]. The π_N/π_S ratio was estimated in windows of 200 kb using an in-house R script. To account for the fact that inversions can greatly suppress recombination in surrounding parts of the genome[74] the collinear part of the genome was divided into two groups: (1) a group including all collinear 200 kb windows outside inversions and (2) a group including all collinear windows outside inversions but excluding windows that came from contigs with inversions (so called strict filtering). Both collinear datasets were used to test for overall differences in mutation load between inversions and the collinear part of the genome (using two-sided t-test). One-sided t-tests were used to test whether minor (less frequent) homokaryotypes had higher mutation loads than major (more frequent) homokaryotypes. Homokaryotypes that contained fewer than four 200 kb windows and were present in few individuals (two thresholds were tested: < 4 and < 10 individuals) were excluded from the

analysis. Additionally, windows with a small number of genes were excluded (two thresholds were tested: < 5 and < 10 genes).

Haplotype structuring, i.e., the existence of two or more distinct sub-haplotypes among inversion homozygote haplotypes, can halt fitness degeneration on one or both inversion haplotypes by carrying partially complementary sets of deleterious recessive alleles[10,103]. To check if any inversion homozygotes exhibited haplotype structuring we first phased the data using Beagle 5.2 (default settings[104]). Next, in each inverted region and homokaryotype we filtered out all variants with MAF < 0.1 and used PGDSpider 2.1.1.0[105] to convert variant call formats (VCF) to full length sequences. Finally, we constructed neighbor-joining trees for alleles within haplotypes using MEGA7[106].

Phenotype-genotype associations

To test whether inversion polymorphisms were associated with diapause phenotypes we genome sequenced 18 individuals from a spruce bark beetle diapause study by Schebeck et al.[51] (10 beetles expressing facultative diapause and 8 beetles expressing obligatory diapause). DNA sequencing was performed using the DNBseq platform (BGI Tech Solutions, Poland) to a mean coverage of 20×. The data was processed in the same way as described above and combined with other sequenced individuals before performing PCA in PLINK. To test for differentiation in inversion haplotype frequencies between diapause phenotypes an exact G test was run using Genepop 4.1.2[107]. F_{ST} between individuals expressing facultative and obligate diapause was estimated using VCFtools[89] and summarized in 100 kb overlapping windows (20 kb steps).

To check if spruce bark beetle odorant receptors (ORs) genes were associated with inversions we examined 73 OR genes recently annotated by Yuvaraj et al.[54]. OR sequences were mapped to the bark beetle reference genome using minimap2[108], and 71 out of 73 ORs were located in the 186 Mb covered by the 36 contigs we analyzed. Three ORs mapped to more than one contig. ItypOR9 and ItypOR58 mapped to the end of IpsContig16 and 23, suggesting possible assembly error and duplication of end-of-contig sequences, which are difficult to assemble. ItypOR59NTE mapped to three nearby locations on IpsContig6, suggesting either assembly error or recent duplication. For these three ORs we used one randomly chosen location in downstream analyses.

To test if inversions were enriched in OR genes, we ran permutation tests (10,000 iterations; permutating inversions locations). To check if alternative inversion arrangements harbored different numbers of OR genes (e.g. that one arrangement carried a deletion) we compared sequence coverage within ORs in individual beetles identified as inversion homozygotes.

Genotype-environment association

Genotype-environment association (GEA) analyses were done to test if allele frequency changes in SNPs were associated with the beetle populations' local environmental. Many different environmental variables were summarized along two principal components (see below). To control for confounders due to the overall genetic differentiation, we used Latent Factors Mixed Models (LFMM[109]) as implemented in the lfmm2() function from the R package LEA[110]. We used only SNPs with < 20% missing data and MAF >= 0.1, and that occurred in individuals with < 30% missing data. Because LFMM cannot handle missing data, we imputed missing genotypes with impute() from LEA. We ran LFMM for (1) all SNPs that passed the filters described above and (2) for a dataset where each inversion was represented as a single "SNP" inversion genotype. We used five latent factors (K = 5) in lfmm2(). P-values were calculated using lfmm2.test() from LEA and false discovery rate (FDR)-corrected using the p.adjust() R function with method = "fdr".

Each beetle population's local environment was characterized according to climate and land-cover data. We used all 19 bioclimatic variables from WorldClim version 2.1.[111] with a resolution of ~1 km^2. These variables are averaged over the years 1970-2000. Proportions of forest, cropland, and built-up areas were downloaded from https://lcviewer.vito.be/ for 2015 with a spatial resolution of ~100 m^2[112]. These global land-cover maps are part of the Copernicus Land Service, derived from PROBA-V satellite observations, and have an accuracy of 80% as measured by the CEOS land product validation subgroup. We also included the proportion of land area covered by spruce trees (genus *Picea*) at a resolution of ~1 km^2, as obtained by Brus et al.[113] using a statistical mapping approach. All environmental variables were reprojected to a final resolution of ~1 km^2 using the Lambert azimuthal equal area method. We then extracted mean values for all environmental variables (Figure S13) within a 50 km radius from each population location using the R package terra[114]. Finally, PCA was used to summarize the multi-

scale environmental variation among populations. The first two PCA components (PC1 and PC2) explained 25% and 23% of the environmental variation, respectively, and were used as the final input for the GEA analyses. PC1 represented environmental variation mainly related to latitude, with northern populations showing higher values indicative of higher temperature seasonality and lower temperatures during the coldest months. PC2 represented environmental variation mostly related to temperature and amount of cropland, with higher values representing localities with higher temperatures during the warmest months and a higher proportion of cropland (and conversely less forest cover and spruce) (Figure S14).

Data Availability

All DNA sequences have been deposited to the European Nucleotide Archive under the BioProject ID PRJNA1013983.

Acknowledgments

We thank A. Hietala, E. Stengel, K. Zub, Å. Lindelöw, O. Langvall, M. Holmlund, E. Kristensen, U. Johansson, R. Modlinger, J. Reisenberger, K. Szreder, W. Skowroński L. Stanecki, and M. Ahlström for help in sampling spruce bark beetles across Europe. T. Mokrzycki helped with beetle sexing and identification. We thank members of the Genomics and Experimental Evolution Group at Jagiellonian University for their help in improving this manuscript. We thank Tomasz Gaczorek for the help in writing scripts and optimizing data analysis. This work was funded by a Polish National Science Center 2018/30/E/NZ8/00105 grant to K.N.B, and the Foundation in Memory of Oscar and Lili Lamm to M.N.A.

Author contributions

A.M., P.Z. and K.N.B. conceived the study, performed the main analyses, and wrote the manuscript. A.B. managed sample shipment, sexed beetles, isolated DNA, and prepared samples for sequencing. F.S., P.K. and M.N.A. provided an unpublished version of the spruce bark beetle genome and valuable insights on the species' ecology and sensory biology. M.M., J.M., Z.B., M.S., P.K., and C.S. organized sampling, provided beetles for analysis and insights on sampled populations. B.A. and W.B. performed GEA analysis. Z.N. analyzed diapause data. J.M. phased

the data. M.S. provided diapause samples. W.B. helped in data interpretation and provided feedback on all manuscript versions. All authors read and approved the final manuscript.

References

1. Hlásny, T. *et al.* Living with bark beetles: impacts , outlook and management options. *From Science to Policy* vol. 8 (2019).
2. Vega, F. & Hofstetter, R. Bark Beetles: Biology and Ecology of Native and Invasive Species. Elsevier (2015).
3. Bentz, B. J., Hansen, E. M., Davenport, M. & Soderberg, D. Complexities in predicting mountain pine beetle and spruce beetle response to climate change. In *Bark Beetle Management, Ecology, and Climate Change*. Elsevier (2021).
4. Biedermann, P. H. W. *et al.* Bark Beetle Population Dynamics in the Anthropocene: Challenges and Solutions. *Trends Ecol. Evol.* **34**, 914–924 (2019).
5. Müller, M., Olsson, P. O., Eklundh, L., Jamali, S. & Ardö, J. Features predisposing forest to bark beetle outbreaks and their dynamics during drought. *For. Ecol. Manage.* **523**, 120480 (2022).
6. Hlásny, T. *et al.* Bark Beetle Outbreaks in Europe: State of Knowledge and Ways Forward for Management. *Curr. For. Reports* **7**, 138–165 (2021).
7. Müller, M. How natural disturbance triggers political conflict: Bark beetles and the meaning of landscape in the Bavarian Forest. *Glob. Environ. Chang.* **21**, 935–946 (2011).
8. Wermelinger, B. & Jakoby, O. Bark beetles. in *Disturbance Ecology* (ed. Thomas Wohlgemuth, Anke Jentsch, R. S.) 271–293 (2022).
9. Wermelinger, B. Ecology and management of the spruce bark beetle Ips typographus - A review of recent research. *For. Ecol. Manage.* **202**, 67–82 (2004).
10. Berdan, E. L., Blanckaert, A., Butlin, R. K. & Bank, C. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genet.* **17**, (2021).
11. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
12. Wellenreuther, M. & Bernatchez, L. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol. Evol.* **33**, 427–440 (2018).
13. Lohse, K., Clarke, M., Ritchie, M. G. & Etges, W. J. Genome-wide tests for introgression

between cactophilic Drosophila implicate a role of inversions during speciation. *Evolution* **69**, 1178–1190 (2015).

14. Fuller, Z. L., Koury, S. A., Phadnis, N. & Schaeffer, S. W. How chromosomal rearrangements shape adaptation and speciation: Case studies in Drosophila pseudoobscura and its sibling species Drosophila persimilis. *Mol. Ecol.* **28**, 1283–1301 (2019).
15. Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2015).
16. Thompson, M. J. & Jiggins, C. D. Supergenes and their role in evolution. *Heredity* **113**, 1–8 (2014).
17. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
18. Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**, 2728–2732 (2014).
19. Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L. & Slotte, T. The Genomic Architecture and Evolutionary Fates of Supergenes. *Genome Biol. Evol.* **13**, 1–19 (2021).
20. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
21. Kirubakaran, T. G. *et al.* Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol. Ecol.* **25**, 2130–2143 (2016).
22. Kapun, M., Fabian, D. K., Goudet, J. & Flatt, T. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. *Mol. Biol. Evol.* **33**, 1317–1336 (2016).
23. Mérot, C. *et al.* Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. *Mol. Biol. Evol.* **38**, 3953–3971 (2021).
24. Porubsky, D. *et al.* Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* **185**, 1986-2005.e26 (2022).
25. Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
26. Harringmeyer, O. S. & Hoekstra, H. E. Chromosomal inversion polymorphisms shape the

genomic landscape of deer mice. *Nat. Ecol. Evol.* **6**, 1965–1979 (2022).

27. Roesti, M., Gilbert, K. J. & Samuk, K. Chromosomal inversions can limit adaptation to new environments. 1–16 (2022).

28. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).

29. Charlesworth, B. The effects of inversion polymorphisms on patterns of neutral genetic diversity. *Genetics* **224**, 1–21 (2023).

30. Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. Evolving Inversions. *Trends Ecol. Evol.* **34**, 239–248 (2019).

31. Connallon, T. & Clark, A. G. Balancing selection in species with separate sexes: Insights from fisher's geometric model. *Genetics* **197**, 991–1006 (2014).

32. Tigano, A. *et al.* Chromosome-Level Assembly of the Atlantic Silverside Genome Reveals Extreme Levels of Sequence Diversity and Structural Genetic Variation. *Genome Biol. Evol.* **13**, 1–17 (2021).

33. Faria, R. *et al.* Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Mol. Ecol.* **28**, 1375–1393 (2019).

34. Jay, P. *et al.* Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat. Genet.* **53**, 288–293 (2021).

35. Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).

36. Jay, P., Tezenas, E., Véber, A. & Giraud, T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLoS Biology* **20**, e3001698 (2022).

37. Schaal, S. M., Haller, B. C. & Lotterhos, K. E. Inversion invasions: When the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philos. Trans. R. Soc. B Biol. Sci.* **377**, (2022).

38. Stauffer, C., Lakatos, F. & Hewitt, G. M. Phylogeography and postglacial colonization routes of Ips typographus L. (Coleoptera, Scolytidae). *Mol. Ecol.* **8**, 763–773 (1999).

39. Sallé, A., Arthofer, W., Lieutier, F., Stauffer, C. & Kerdelhué, C. Phylogeography of a host-specific insect: Genetic structure of Ips typographus in Europe does not reflect past fragmentation of its host. *Biol. J. Linn. Soc.* **90**, 239–246 (2007).

40. Bertheau, C. *et al.* Divergent evolutionary histories of two sympatric spruce bark beetle species. *Mol. Ecol.* **22**, 3318–3332 (2013).
41. Mayer, F. *et al.* Comparative multilocus phylogeography of two Palaearctic spruce bark beetles: Influence of contrasting ecological strategies on genetic variation. *Mol. Ecol.* **24**, 1292–1310 (2015).
42. Ellerstrand, S. J. *et al.* Weak population genetic structure in Eurasian spruce bark beetle over large regional scales in Sweden. *Ecol. Evol.* **12**, 1–12 (2022).
43. Nilssen, A. C. Long-range aerial dispersal of bark beetles and bark weevils (Coleoptera, Scolytidae and Curculionidae) in northern Finland. *Ann. Entomol. Fenn.* **50**, 37–42 (1984).
44. Brelsford, A. *et al.* An Ancient and Eroded Social Supergene Is Widespread across Formica Ants. *Curr. Biol.* **30**, 304-311.e4 (2020).
45. Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes. *Curr. Biol.* **24**, 288–294 (2014).
46. Roff, D. a. The Evolution of Threshold Traits in Animals. *Q. Rev. Biol.* **71**, 3–35 (1996).
47. Pruisscher, P., Nylin, S., Gotthard, K. & Wheat, C. W. Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. *Mol. Ecol.* **27**, 3613–3626 (2018).
48. Kozak, G. M. *et al.* Genomic Basis of Circannual Rhythm in the European Corn Borer Moth. *Curr. Biol.* **29**, 3501-3509.e5 (2019).
49. Paolucci, S., Salis, L., Vermeulen, C. J., Beukeboom, L. W. & van de Zande, L. QTL analysis of the photoperiodic response and clinal distribution of period alleles in Nasonia vitripennis. *Mol. Ecol.* **25**, 4805–4817 (2016).
50. Denlinger, D. *Insect diapause*. Cambridge University Press (2022).
51. Schebeck, M., Dobart, N., Ragland, G. J., Schopf, A. & Stauffer, C. Facultative and obligate diapause phenotypes in populations of the European spruce bark beetle Ips typographus. *J. Pest Sci. (2004).* **95**, 889–899 (2022).
52. Hansson, B. S. & Stensmyr, M. C. Evolution of insect olfaction. *Neuron* **72**, 698–711 (2011).
53. Kandasamy, D., Gershenzon, J., Andersson, M. N. & Hammerbacher, A. Volatile organic compounds influence the interaction of the Eurasian spruce bark beetle (*Ips typographus*)

with its fungal symbionts. *ISME J.* **13**, 1788–1800 (2019).

54. Yuvaraj, J. K. *et al.* Putative ligand binding sites of two functionally characterized bark beetle odorant receptors. *BMC Biol.* **19**, 1–21 (2021).

55. Hou, X. Q. *et al.* Functional Evolution of a Bark Beetle Odorant Receptor Clade Detecting Monoterpenoids of Different Ecological Origins. *Mol. Biol. Evol.* **38**, 4934–4947 (2021).

56. Kandasamy, D. *et al.* Conifer-killing bark beetles locate fungal symbionts by detecting volatile fungal metabolites of host tree resin monoterpenes. *PLoS Biology* **21**, e3001887 (2023).

57. Tigano, A. & Friesen, V. L. Genomics of local adaptation with gene flow. *Mol. Ecol.* **25**, 2144–2164 (2016).

58. Akopyan, M. *et al.* Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. *Mol. Ecol.* **31**, 3323–3341 (2022).

59. Christmas, M. J. *et al.* Chromosomal inversions associated with environmental adaptation in honeybees. *Mol. Ecol.* **28**, 1358–1374 (2019).

60. Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L. & Wellenreuther, M. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nat. Commun.* **11**, (2020).

61. Navarro, A., Barbadilla, A. & Ruiz, A. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in drosophila. *Genetics* **155**, 685–698 (2000).

62. Guerrero, R. F., Rousset, F. & Kirkpatrick, M. Coalescent patterns for chromosomal inversions in divergent populations. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 430–438 (2012).

63. Ohta, T. Associative overdominance caused by linked detrimental mutations. *Genet. Res.* **18**, 277–286 (1971).

64. Connallon, T. & Olito, C. Natural selection and the distribution of chromosomal inversion lengths. *Mol. Ecol.* **31**, 3627–3641 (2022).

65. Yang, Y. Y., Lin, F. J. & Chang, H. Y. Comparison of recessive lethal accumulation in inversion-bearing and inversion-free chromosomes in Drosophila. *Zool. Stud.* **41**, 271–282 (2002).

66. Huang, K. *et al.* Mutation Load in Sunflower Inversions Is Negatively Correlated with Inversion Heterozygosity. *Mol. Biol. Evol.* **39**, 1–17 (2022).
67. McCulloch, G. A. & Waters, J. M. Rapid adaptation in a fast-changing world: Emerging insights from insect genomics. *Glob. Chang. Biol.* **29**, 943–954 (2023).
68. Rane, R. V., Rako, L., Kapun, M., Lee, S. F. & Hoffmann, A. A. Genomic evidence for role of inversion 3RP of *Drosophila melanogaster* in facilitating climate change adaptation. *Mol. Ecol.* **24**, 2423–2432 (2015).
69. Kapun, M. & Flatt, T. The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol. Ecol.* **28**, 1263–1282 (2019).
70. Nunez, J. C. B. *et al.* A cosmopolitan inversion drives seasonal adaptation in overwintering Drosophila. *bioRxiv* 2022.12.09.519676 (2022).
71. Li, J. *et al.* Joint analysis of demography and selection in population genetics: Where do we stand and where could we go? *Mol. Ecol.* **21**, 28–44 (2012).
72. Johri, P. *et al.* The Impact of Purifying and Background Selection on the Inference of Population History: Problems and Prospects. *Mol. Biol. Evol.* **38**, 2986–3003 (2021).
73. Li, H. *et al.* Heterozygous inversion breakpoints suppress meiotic crossovers by altering recombination repair outcomes. *PLoS Genet.* **19**, 1–20 (2023).
74. Koury, S. A. Predicting recombination suppression outside chromosomal inversions in Drosophila melanogaster using crossover interference theory. *Heridity* **130**, 196–208 (2023).
75. Seidl, R., Schelhaas, M., Rammer, W. & Verkerk, P. J. Increasing forest disturbances in Europe and their impact on carbon storage. *Nat. Clim. Chang.* **4**, 806–810 (2014).
76. Hlásny, T. *et al.* Devastating outbreak of bark beetles in the Czech Republic: Drivers, impacts, and management implications. *For. Ecol. Manage.* **490**, 119075 (2021).
77. Lange, H., Økland, B. & Krokene, P. Thresholds in the life cycle of the spruce bark beetle under climate change. *Interjournal Complex Syst.* **1648**, 1–10 (2006).
78. Jönsson, A. M., Harding, S., Bärring, L. & Ravn, H. P. Impact of climate change on the population dynamics of Ips typographus in southern Sweden. *Agric. For. Meteorol.* **146**, 70–81 (2007).
79. Wulff, S. & Roberge, C. Inventering av granbarkborreangrepp i Götaland och Svealand. (2022).

80. Öhrn, P., Berlin, M., Elfstrand, M., Krokene, P. & Jönsson, A. M. Seasonal variation in Norway spruce response to inoculation with bark beetle-associated bluestain fungi one year after a severe drought. *For. Ecol. Manage.* **496**, 119443 (2021).
81. Powell, D. *et al.* A highly-contiguous genome assembly of the Eurasian spruce bark beetle, Ips typographus, provides insight into a major forest pest. *Commun. Biol.* **4**, 1–9 (2021).
82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
83. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
84. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
85. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
86. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
87. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
88. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
89. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
90. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 1–13 (2014).
91. Delmore, K. E. *et al.* Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* **2**, 76–87 (2018).
92. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
93. Jones, J. C. *et al.* Extreme Differences in Recombination Rate between the Genomes of a Solitary and a Social Bee. *Mol. Biol. Evol.* **36**, 2277–2291 (2019).
94. McVean, G., Awadalla, P. & Fearnhead, P. A coalescent-based method for detecting and

estimating recombination from gene sequences. *Genetics* **160**, 1231–1241 (2002).

95. Li, H. & Ralph, P. Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019).

96. Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L. & Rieseberg, L. H. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol. Ecol.* **29**, 2535–2549 (2020).

97. Graffelman, J. & Weir, B. S. Multi-allelic exact tests for Hardy–Weinberg equilibrium that account for gender. *Mol. Ecol. Resour.* **18**, 461–473 (2018).

98. Graffelman, J. Exploring diallelic genetic markers: The HardyWeinberg package. *J. Stat. Softw.* **64**, 1–23 (2015).

99. Krasovec, M. The spontaneous mutation rate of Drosophila pseudoobscura. *G3 Genes, Genomes, Genet.* **11**, (2021).

100. Keightley, P. D. *et al.* Estimation of the Spontaneous Mutation Rate in Heliconius melpomene. *Mol. Biol. Evol.* **32**, 239–243 (2015).

101. Oppold, A. M. & Pfenninger, M. Direct estimation of the spontaneous mutation rate by short-term mutation accumulation lines in Chironomus riparius. *Evol. Lett.* **1**, 86–92 (2017).

102. Nelson, C. W., Moncla, L. H. & Hughes, A. L. SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**, 3709–3711 (2015).

103. Charlesworth, B. & Charlesworth, D. Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet. Res.* **70**, 63–73 (1997).

104. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

105. Lischer, H. E. L. & Excoffier, L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).

106. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

107. Raymond, M. & Rousset, F. An Exact Test for Population Differentiation. *Evolution* **49**,

1280 (1995).

108. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

109. Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* **36**, 852–860 (2019).

110. Gain, C. & François, O. LEA 3: Factor models in population genetics and ecological genomics with R. *Mol. Ecol. Resour.* **21**, 2738–2748 (2021).

111. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

112. Buchhorn, M. *et al.* Copernicus global land service: Land cover 100m: Epoch 2015: Globe. Version V2. 0.2. (2019).

113. Brus, D. J. *et al.* Statistical mapping of tree species over Europe. *Eur. J. For. Res.* **131**, 145–157 (2012).

114. Hijmans, R. terra: Spatial Data Analysis. R package version 1.7-18. https://CRAN.R-project.org/package=terra. (2023).

Tables

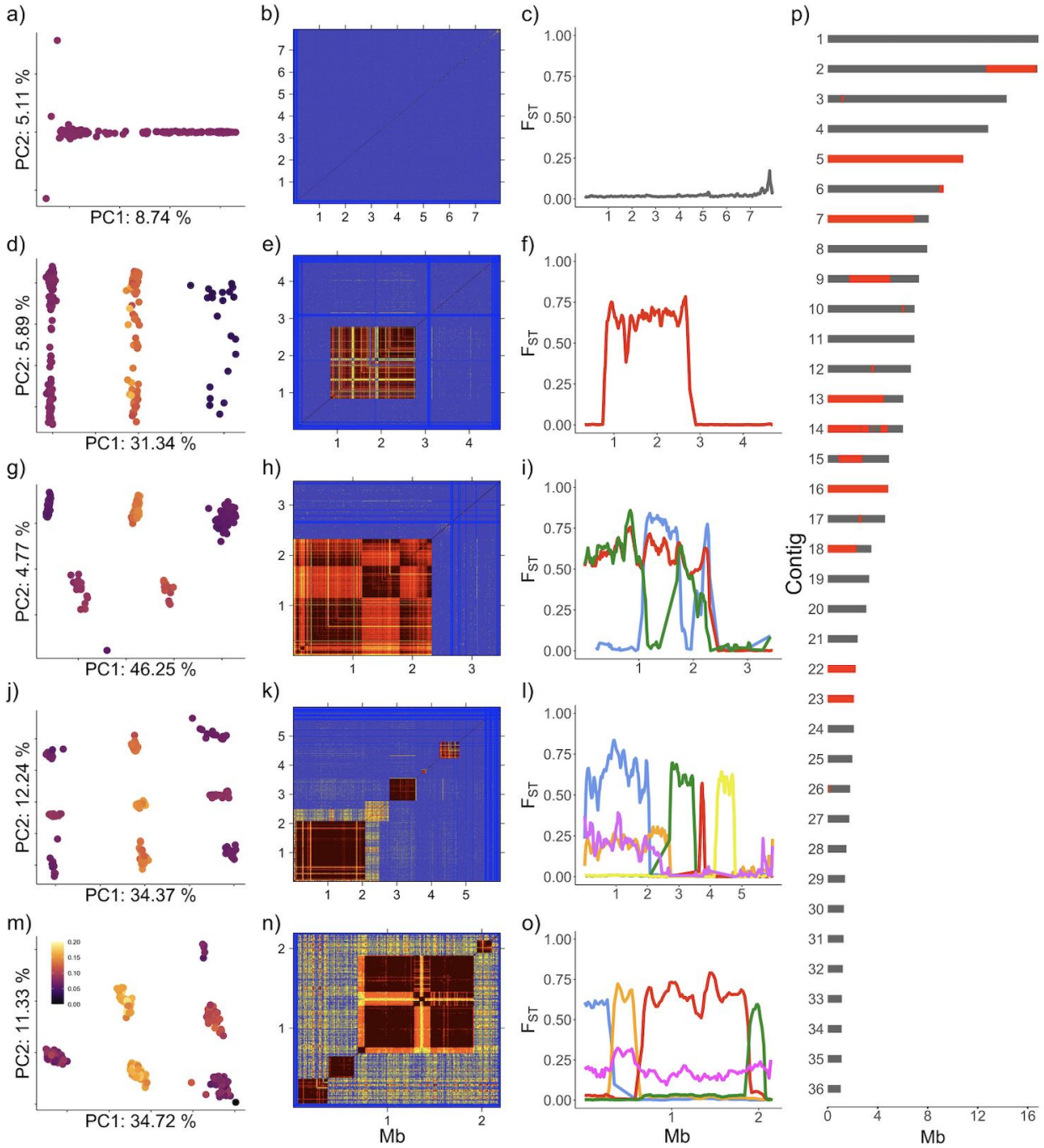
Table 1 List of identified chromosomal inversions in the *Ips typographus* genome. ID: inversion name; Contig: contig name; Size: size of the inversions (Mb); Start and End: coordinates of the inversion (Mb); Age: approximate age of the inversion in Myr; Odorant receptors: odorant receptors present within inversion. Note that Inv16.1 and Inv23.1 are parts of the same inversion and Inv16.2 and Inv23.2 are a part of another single inversion.

ID	Contig	Size	Start	End	Age	Odorant receptors
Inv2	IpsContig2	4.04	12.67	16.71	0.5	
Inv3	IpsContig3	0.14	1.11	1.25	1.1	
Inv5	IpsContig5	10.84	0.00	10.84	1.3	ItypOR33, ItypOR41, ItypOR40, ItypOR10, ItypOR47, ItypOR50, ItypOR29, ItypOR43JOI, ItypOR34, ItypOR52NTE, ItypOR4, ItypOR3, ItypOR53, ItypOR2, ItypOR19
Inv6	IpsContig6	0.34	8.93	9.27	1.0	
Inv7.1	IpsContig7	0.67	0.00	0.67	1.7	
Inv7.2	IpsContig7	6.92	0.00	6.92	1.1	ItypOR1, ItypOR17
Inv9	IpsContig9	3.30	1.71	5.01	0.6	
Inv10	IpsContig10	0.08	6.05	6.13	1.8	
Inv12	IpsContig12	0.07	3.63	3.70	1.5	
Inv13	IpsContig13	4.50	0.00	4.50	1.7	ItypOR28, ItypOR23, ItypOR49, ItypOR27
Inv14.1	IpsContig14	2.08	0.00	2.08	1.9	ItypOR36, ItypOR44, ItypOR18JF, ItypOR20NTE
Inv14.2	IpsContig14	0.67	2.08	2.75	1.0	
Inv14.3	IpsContig14	0.76	2.78	3.54	2.1	
Inv14.4	IpsContig14	0.11	3.73	3.84	2.1	
Inv14.5	IpsContig14	0.57	4.23	4.80	2.3	
Inv14.6	IpsContig14	2.48	0.00	2.48	0.6	ItypOR36, ItypOR44, ItypOR18JF, ItypOR20NTE
Inv15	IpsContig15	1.92	0.86	2.78	1.9	
Inv16.1	IpsContig16	4.83	0.00	4.83	1.6	ItypOR58, ItypOR9, ItypOR11, ItypOR31, ItypOR30, ItypOR16, ItypOR35JF
Inv16.2	IpsContig16	4.83	0.00	4.83	0.7	ItypOR58, ItypOR9, ItypOR11, ItypOR31, ItypOR30, ItypOR16,

						ItypOR35JF
Inv17	IpsContig17	0.21	2.48	2.69	2.2	
Inv18	IpsContig18	2.32	0.00	2.32	2.1	
Inv22.1	IpsContig22	0.32	0.00	0.32	2.1	
Inv22.2	IpsContig22	0.23	0.42	0.65	2.1	
Inv22.3	IpsContig22	1.23	0.68	1.91	2.0	ItypOR22CTE
Inv22.4	IpsContig22	0.20	1.92	2.12	2.1	
Inv22.5	IpsContig22	2.24	0.00	2.24	0.6	ItypOR22CTE
Inv23.1	IpsContig23	2.10	0.00	2.10	1.4	ItypOR58, ItypOR9
Inv23.2	IpsContig23	1.84	0.26	2.10	0.6	ItypOR58, ItypOR9
Inv26	IpsContig26	0.10	0.12	0.22	2.6	

Figures

Figure 1 Identification of chromosomal inversions in *Ips typographus*. Each row of figure panels shows the results of per contig PCA, per contig linkage disequilibrium analysis, and genetic differentiation (F_{ST}) analysis for a selected contig. Panels a-c show results for a contig with no inversions (IpsContig8) and little differentiation between southern and northern populations. Panels d-o show different contigs with increasingly complex inversion patterns: a single inversion (d-f, Inv15); a single inversion with a double-crossover signal (g-i, Inv18); multiple adjacent inversions with one inversion overlapping with the first two inversions on the contig (j-l, Inv14.1, Inv14.2, Inv14.3, Inv14.4, Inv14.5, and Inv.14.6); multiple adjacent inversions with one large inversion overlapping with several smaller ones (m-o, Inv22.1, Inv22.2, Inv22.3, Inv22.4, and Inv22.5). The large overlapping inversion is visible as a yellow background in the LD plot (n). The PCA grouping shown in (m) corresponds to genotypes of the largest inversion on IpsContig22 (Inv22.3), including genotypes that include haplotypes produced with the double crossover event. The last panel (p) shows the 36 largest contigs of the spruce bark beetle genome with inversions indicated in red. Dots in the PCA plots represent individual beetles and are colored according to the heterozygosity of the individual (darker color represents low heterozygosity). Both axis on LD plots (b,e,h,k,n) represent conting's positions in megabases (Mb); low levels of linkage are shown in blue and higher levels in yellow to dark red. F_{ST} values in panels f,i,l,o show genetic differentiation between inversion haplotypes along the contig in question.


a)
PC2: 5.11 %
PC1: 8.74 %
b)
c)
F_{ST}
p)
d)
PC2: 5.89 %
PC1: 31.34 %
e)
f)
g)
PC2: 4.77 %
PC1: 46.25 %
h)
i)
Contig
j)
PC2: 12.24 %
PC1: 34.37 %
k)
l)
m)
PC2: 11.33 %
PC1: 34.72 %
n)
o)
Mb
Mb
Mb

938

Figure 2 Genomic structure and differentiation in *Ips typographus*. (a) Whole-genome PCA, where colors correspond to genetic clustering of beetle individuals revealed by NGSadmix analysis. (b,c) Geographical distribution and genetic differentiation of the 18 beetle populations analyzed. Blue dots and bars: northern populations (N); red dots and bars: southern populations (S); violet dots: Polish populations (P). (d) Genome-wide genetic differentiation (F_{ST}) calculated between northern and southern populations. Vertical lines separate different contigs shown in grey and black.

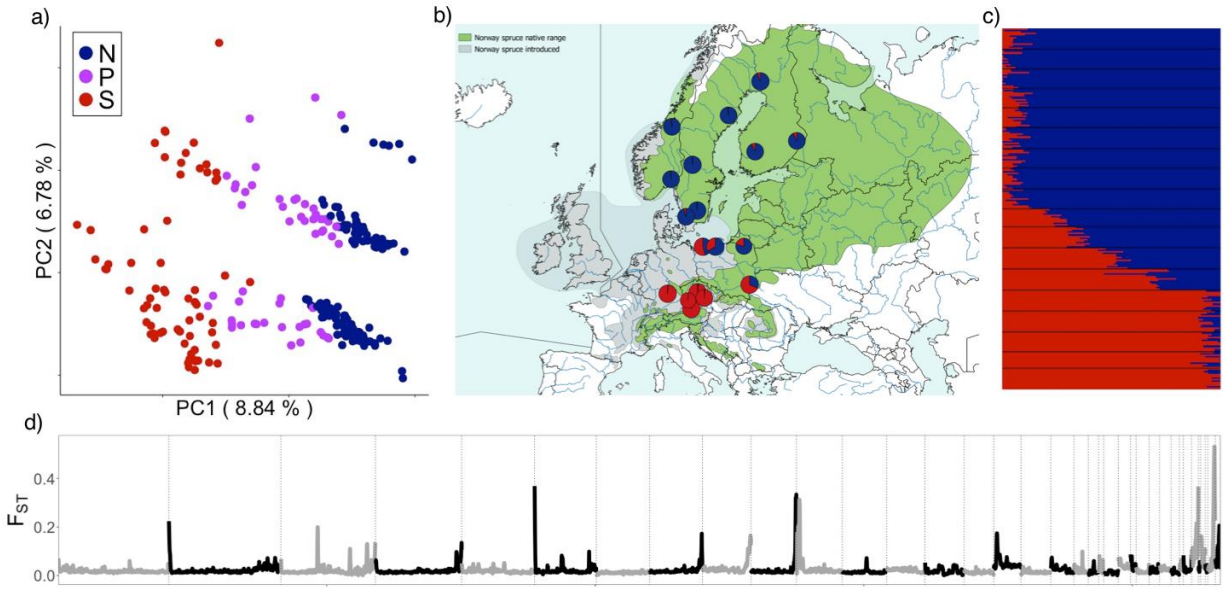


Figure 3 Correlation between inversion haplotype frequency and population origin (latitude) for chromosomal inversions detected in European *Ips typographus* populations. Significant correlations are indicated with asterisks (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). Inversions harboring odorant receptors are indicated in green.

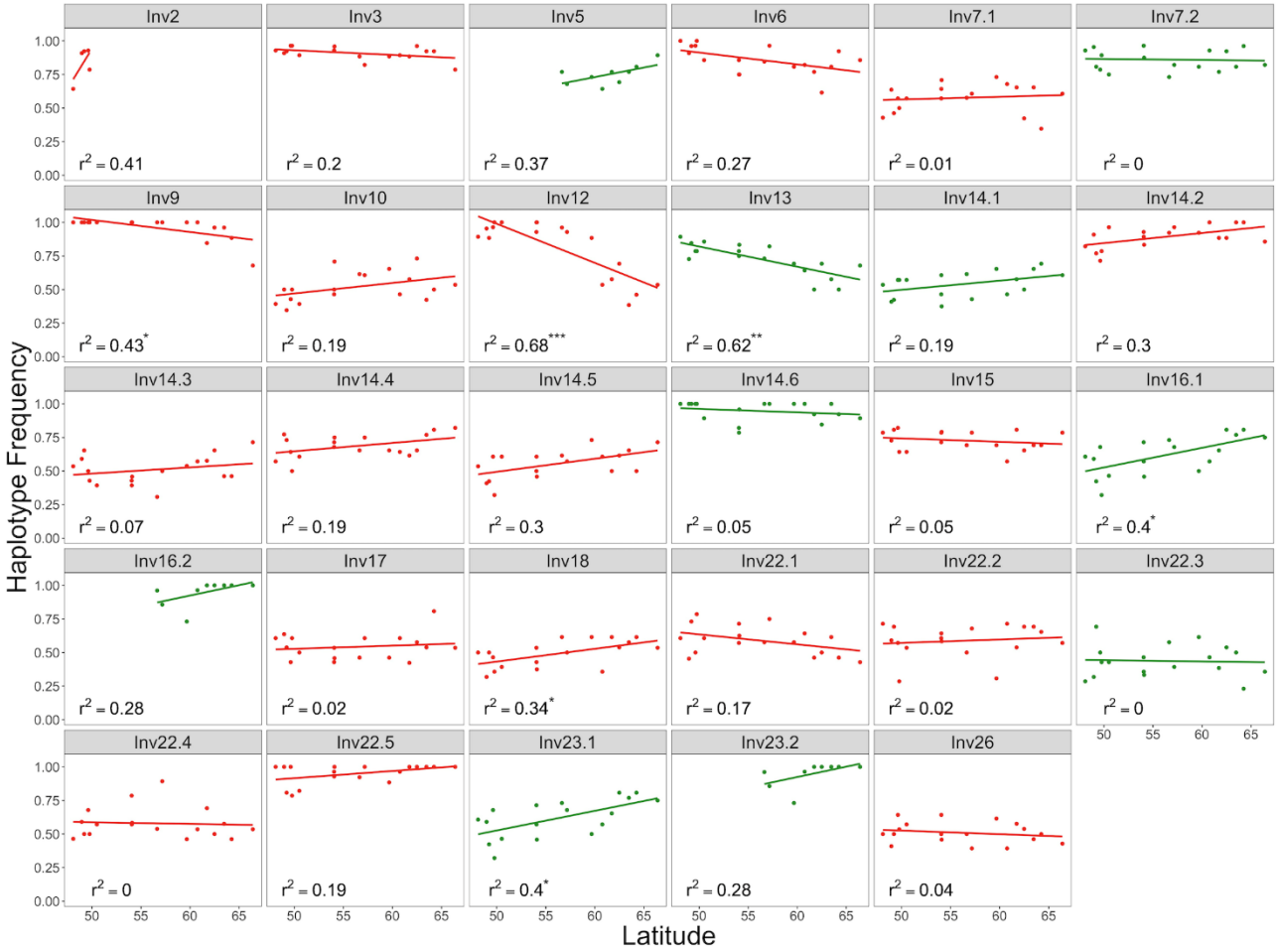


Figure 4 Genotype-environment associations across 36 contigs in the *Ips typographus* genome. Different colors represent different contigs. The upper three panels show results for all SNPs and the lower three panels show results using data where each inversion was coded as a single locus (shown in a blue circle). MAF: Minor Allele Frequency. PC1, PC2, and PC3 represent the first three of PCA used to summarize the environmental variation among populations.

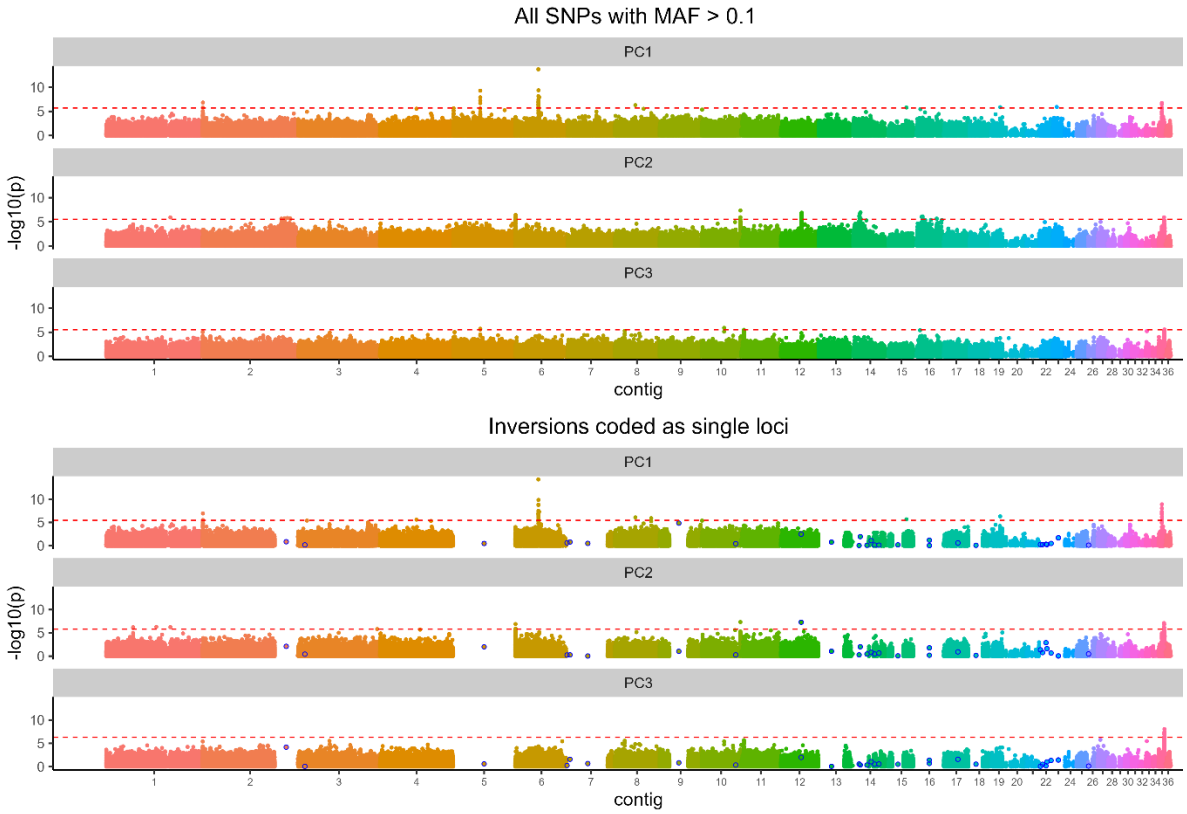
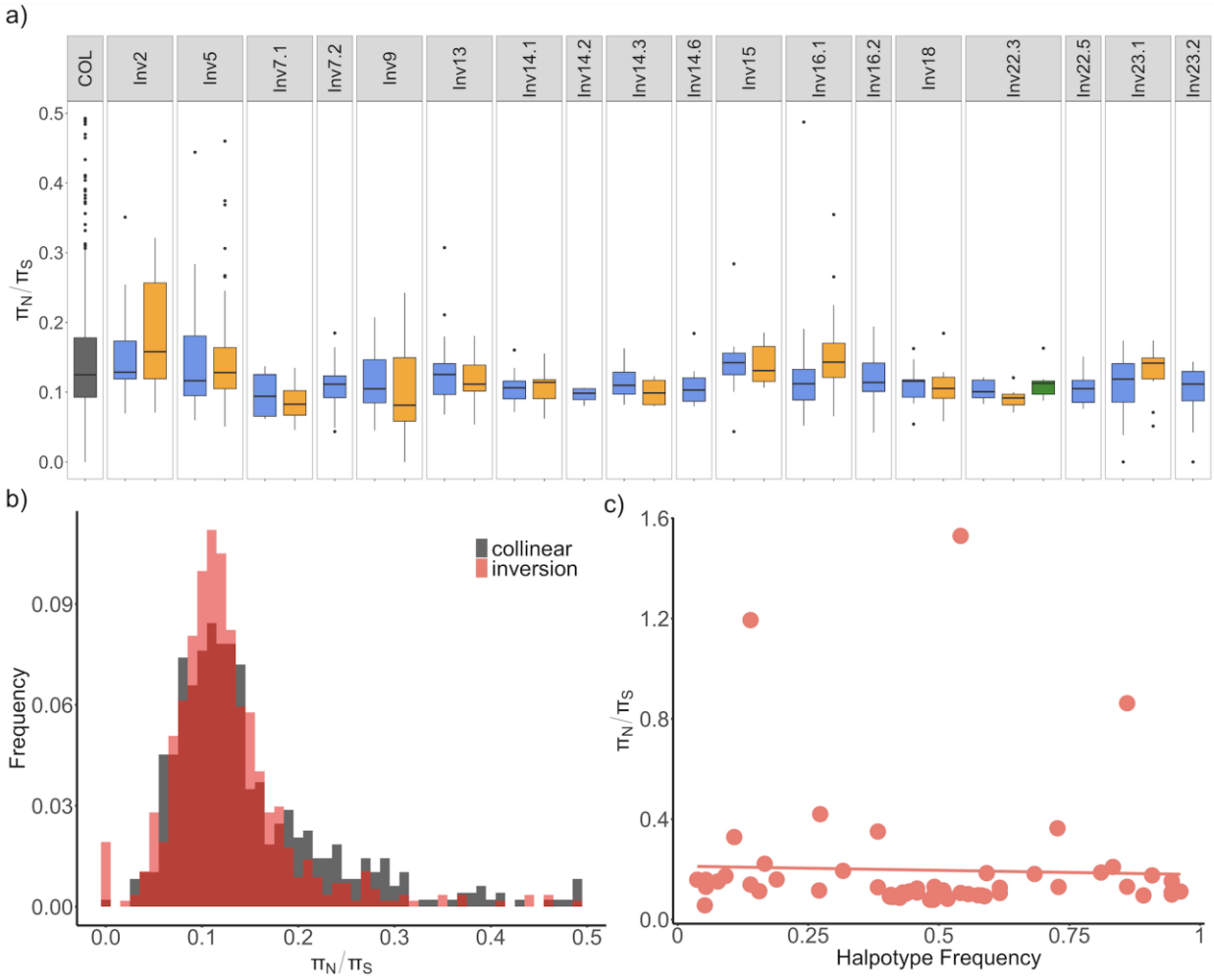



Figure 5 Mutation load analysis in *Ips typographus*. (a) The ratio of nonsynonymous to synonymous nucleotide diversity π_N/π_S computed for 200 kb windows along collinear parts of the genome (COL) and different inversion haplotypes (blue: major haplotype; yellow: minor haplotype; green: haplotype produced by double crossover events between minor and major haplotypes). n = 4 to 55 windows per inversion haplotype and n = 531 for the collinear genome. Inversion haplotypes that were present in less than four individuals, did not include four or more 200 kb windows or include less than five genes were not included in the analysis. For better visibility, π_N/π_S outliers above 0.5 are not shown (a total of 58 π_N/π_S values, 45 of them belonging to COL). The lower and upper box hinges correspond to the first and third quartiles, whiskers show 1.5 * the inter-quartile range. (b) Distribution of π_N/π_S values per 200 kb window (c) Correlation between mean π_N/π_S for all inversion haplotypes and their frequency across spruce bark beetle populations.