An initial ecotypic divergence admixes species dispersal and within island diversification in a parallel Galapagos beetle radiation.

Carl Vangestel [1,2], Steven M. Van Belleghem [3], Zoë De Corte[1], Frederik Hendrickx [1,2]

[1] Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[2] Terrestrial Ecology Unit, Biology Department, Ghent University, Gent, Belgium

[3] Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, Puerto Rico

Abstract

Within insular systems, selection along similar ecological gradients may result in the repeated evolution of strikingly similar ecotypes on different islands. Whether these repeated *in-situ* divergences reflect independent evolutionary events remains, however, poorly understood because the origin of the alleles underlying ecotypic differentiation and their mode of dispersal between islands is generally unknown. Here, we investigate the genomic underpinning of a parallel caterpillar-hunter beetle (*Calosoma*) radiation from the Galapagos wherein short-winged highland species and populations seemingly evolved repeatedly out of a widespread winged lowland species. The radiation represents a unique divergence gradient that progresses from populations with only partially reduced wings on the youngest islands to distinct short-winged highland species on the oldest islands. Although patterns of genetic divergence largely follow this pattern of repeated evolution, adaptation towards highland habitats is driven by repeated selection of the same alleles that are shared across all highland species and populations. These alleles comprise extensive chromosomal inversions whose origin is traced back to the oldest high-lowland divergence within this radiation. Highland alleles subsequently spread along this island progression by both dispersal of highland individuals as well as lowland individuals that are polymorphic at adaptive loci. These results highlight the importance of an ancient singular divergence in driving the direction of species radiations. This mechanism intermixes. We argue that this mechanism admixes the role of ecotype dispersal and within-island diversification, and more generally the contribution of ecological and evolutionary processes, in the genesis of similar species assemblages in island archipelagos.

Introduction

Insular radiations, like those found on island archipelagos and isolated lakes, provide key insight into the ecological and evolutionary drivers of adaptive radiation and species diversification [1–3]. Important progress in this field is gained from settings wherein species repeatedly and independently proliferate along similar environmental gradients. These cases of parallel within island evolution of ecological and phenotypic similar species demonstrated that the direction of evolution can be surprisingly predictable [2], that natural selection can drive rapid speciation (i.e. ecological speciation), even in absence of a geographic barrier [4,5] and that the interaction between evolutionary and ecological dynamics drives the assembly of species communities within islands [6,7].

The extent to which these repeated radiations represent independent replicates of an evolutionary process remains, however, less understood [8–10]. Independent evolution within each island presumes that the alleles subject to divergent selection evolved independently and, hence, by unique mutations within each island (Fig. 1a). Yet, there is accumulating evidence that parallel ecological differentiation often involves repeated selection on the same alleles [11–13] and, hence, a shared evolutionary origin of the alleles underlying parallel evolution. Whether these alleles evolved during a singular divergence and how they are subsequently transported across islands remains currently poorly understood. First, adaptive alleles could be introduced from nearby islands by founding individuals that are polymorphic at adaptive loci (Fig. 1b). Second, alleles involved in ecotypic differentiation could be introduced more directly into the gene pool of a resident ecotype by admixture from occasional immigrants from an alternative ecotype (Fig. 1c). However, if a larger number of immigrants founds a new ecotypic population and reproductive isolation with the resident ecotype is incomplete, admixture may erase the initial genetic differences between these

lineages, while maintaining differentiation at loci involved in ecotypic differentiation (Fig 1d). Under this latter scenario, the repeated occurrence of ecologically similar species on the different islands primarily involves dispersal of ecotypes between islands (Fig. 1e) and no within island diversification is actually involved [8].

Despite these scenarios involve highly different contributions of ecological (ecotype dispersal) versus evolutionary (within island diversification) processes in generating ecotypic species pairs, patterns of genetic differentiation at neutral and adaptive loci are expected to be remarkably similar under these scenarios (Fig. 1) [14,15]. This issue even questions the degree of evidence for parallel evolution for some of the most iconic examples of adaptive radiation [8,10].

Studying parallel divergence at different stages of the divergence process has been proposed to offer the best solution to infer the role of historical processes and the source of the alleles involved in ecotypic divergence [16]. Here, we reconstruct the evolutionary history of a unique caterpillar hunter beetle (*Calosoma* sp.) radiation showing repeated and gradual highland divergence along an island progression at the Galapagos archipelago [17]. At low elevations, a single long-winged species, *C. grantense*, is found on all major islands, while high elevations of the old and intermediate-aged islands San Cristobal, Santa Cruz and Santiago are each occupied by a distinct highland species (*C. linelli, C. leleuporum and C. galapageum* respectively)(Fig. 2a). These highland species share morphological traits such as a marked reduction in wing size, but the degree of morphological divergence with the lowland species decreases towards more recent islands (Fig. 2B). Highlands of the youngest islands Isabela and Fernandina are occupied by populations of the lowland species *C. granatense* but evolved a wing size reduction in line with the divergence of the highland species on the older islands (Fig. 2B).

By assessing genetic variation at a genome-wide scale we show that this apparent repeated divergence is driven by repeated selection of the same alleles, encompassing massive chromosomal inversions, that likely evolved during a singular highland divergence on the oldest island. Both inter-island dispersal of highland populations and polymorphic lowland individuals contributed to the progressive spread of these inversions across the islands. These findings demonstrate how an ancient ecotypic divergence and inter-island dispersal of ecotypes contribute to the genesis of parallel species assemblages in insular systems.

Results

Genome assembly, sampling, and sequencing data

We first assembled the genome of the lowland species *C. granatense* using both paired-end short reads (170bp, 500bp and 2kb insert sizes) from a single individual sampled at Santa Cruz and mate-paired libraries (10kb and 20kb insert sizes) from a pool of nine individuals originating from the same population. The resulting genome assembly comprised 6,045 scaffolds with an N50 of 5.6 Mb and a cumulative length of 167Mb, being 97% of the *k*-mer based estimated genome size (Table S3).

We included all islands from the Galapagos that harbor highland species or populations and sampled between 11 and 22 individuals from each species and population, except for the most recent island Fernandina where only five individuals could be obtained from both the high- and lowland habitat. Wing-sizes of the sampled individuals clearly matched the earlier reported wing sizes of these species and populations [18] (Fig. 2b) and confirmed the gradual reduction in wing-size of the highland species and populations in response to island age. For the most recent island Fernandina, however, wing sizes of highland individuals overlapped with those from lowland individuals. Genomic variation of all individuals was assessed by restriction-site associated DNA sequencing (RADseq) and by resequencing the genomes of four individuals of each highland species and population as well as two individuals from each *C. granatense* lowland population, except for the oldest island San Cristobal for which four individuals of the lowland population were sequenced (Table S1, Table S2). We further sequenced the genome of a single specimen of the related mainland species *Calosoma sayi* [19,20], which was used as an outgroup species.

Species phylogeny, population structure and species admixture

We reconstructed the phylogenetic relationship among the species and populations based on a set of 500 random genomic windows of 20kb that were extracted from the resequenced individuals. Maximum likelihood trees depicted multiple incongruent relationships illustrating that divergence within this radiation does not follow a clear bifurcating pattern (Fig 3a). Yet, a species tree integrating these individual trees supported the following deeper relationships: (i) an initial divergence between the highland species of the oldest island San Cristobal (*C. linelli*) and all other species, (ii) a subsequent divergence that separates the lowland population of San Cristobal from the remaining species, (iii) a subsequent divergence that separates the lowland species of the second oldest island (Santa Cruz) from all other species and populations and (iv) a shared common ancestry of the highland species *C. leleuporum* (Santa Cruz) and *C. galapageium* (Santiago). Clustering of these latter two highland species within the populations of the lowland species *C. granatense* suggests that at least two separate highland-lowland divergences took place within this radiation being a first one that split the highland species *C. linelli* from the lowland species *C. granatense* and a second that split the clade with the highland species *C. leleuporum* and C. galapageium from the lowland species *C. granatense*. Although most nodes in our multilocus species tree were generally strongly supported, gene concordance factors (gCF), which express the percentage of individual trees containing a particular node, were low and ranged between 12% to 36% only, except for the oldest divergence of *C. linelli* (Fig. 3a). Low gCF's were further noted for nodes that clustered the individuals of the highland species *C. leleuporum* (42%) and *C. galapageium* (37%), indicating that these species share a substantial proportion of their haplotypes with other species for many the investigated loci..

Genetic relationship between the populations as inferred with TreeMix [21] based on the RADtag sequencing data confirmed the strong divergence of the highland species of San Cristobal (*C. linelli*) and revealed a very close relationship among the remaining species. Congruent with our multispecies phylogeny, the highland species *C. leleuporum* (SCZ) and *C. galapageium* (SAN) grouped in a clade that was situated within the populations of the lowland species *C. granatense*. Despite the strong genetic divergence of *C. linelli* from all other species, this species appeared subject to substantial admixture from the *C. granatense* population from on the same island. Genetic admixture was further supported from *C. linelli* into the two other highland species (*C. leleuporum* and *C. galapageium*) and from these latter highland species into the highland populations of *C. granatense* on the recent islands Isabela and Fernandina.

In contrast to the multispecies tree, a principal coordiante analysis (PCoA) based on SNP frequencies obtained from RADseq depicted a gradual differentiation of highland species along the island progression with *C. linelli* again being the most divergent species, but now followed by differentiation of the highland species *C. leleuporum* of Santa Cruz and subsequently *C. galapageium* of Santiago (Fig. 2c). Thus, SNP allele frequencies of *C. galapageium* tend to be closer to those of the lowland species *C. granatense* compared with those of its sister species *C. leleuporum*. This discrepancy with the clustering of those two highland species in the species tree suggests that the ancestral population of *C. galapageium* originates from the ancestral highland population of Santa Cruz, but experienced considerable admixture with the lowland species *C. granatense* after colonization.

Genome-wide divergence between high- and lowland species and populations

Using our RADtag sequencing data, we calculated for each SNP the differentiation (F_{st}) between the high- and lowland species or population within each island (Fig. 4). Average F_{st}-values between the high- and lowland species within each island were generally low and ranged from 0.035 on Fernandina to 0.10 between the morphologically clearly distinct species on Santa Cruz. Only for the oldest island San Cristobal, more substantial differentiation (F_{st} = 0.24) was observed between the highland species *C. linelli* and the lowland species *C. granatense*. The increase in

genetic differentiation towards older islands was primarily caused by an increased frequency of SNPs with high F_{st} values, rather than due to a shift in the mode of the F_{st} distribution. For example, the proportion of SNPs with an F_{st} > 0.4 in the highland-lowland comparison increased consistently from 2% on the most recent island Fernandina to 20% on the oldest island San Cristobal (Fig. SX).

To identify potential SNPs associated with high-lowland divergence within each island, we used BayeScan [22] to search for outlier SNPs that are significantly more differentiated between the high- and lowland species or populations than expected from the genome wide background (Q-value < 0.1). Except for the most recent island Fernandina for which a lower number of individuals could be sampled, between 196 (IVA) and 458 (SAN) outlier SNPs could be identified in the within-island comparisons. These outlier SNPs were not randomly distributed across the genome, but generally clustered into large genomic blocks extending up to several megabases (Fig. 4). Moreover, these genomic regions characterized by an elevated divergence and high density of outlier SNPs were remarkably similar across the different islands, which demonstrates that largely similar genomic regions are associated with high-lowland divergence in the different species and islands.

Extensive chromosomal inversions at outlier loci

Because most outlier SNPs were concentrated into large genomic blocks, we asked if the alleles under divergent selection potentially differ by extensive structural variations (SV). We investigated the presence of such SV for the twelve longest contiguous regions with elevated divergence (Fst > 0.1) in an overall high-lowland comparison. Each region contained at least one RADtag with a SNP identified as outlier in minimum two within-island ecotype comparisons, with a total of 109 outlier RADtags (57%) across all twelve regions combined, which supports their association with high-lowland divergence (Fig. 4, Table S2). Using a structural variation analysis based on anomalies in the orientation and insert size of read pairs, we identified chromosomal inversions that perfectly overlapped with the five largest (3.7Mb to 5.9Mb) and one smaller (213kb) region of elevated divergence (Fig. 5a, Fig. S2, Table S2). Two additional regions of elevated divergence included the scaffold start and read pairs situated at the potential SV breakpoint mapped to another scaffold and thus likely represent partially assembled inversions. One last region was flanked by deletions and potentially comprises a more complex SV. Three remaining regions did not show evidence for SV based on anomalous read mappings at the flanking regions (Table X).

SNPs located on each SV generally showed an identical segregation pattern with no obvious decay in linkage disequilibrium across the entire length of the SV (Fig 5b, Fig S2), providing additional support that extensive non-recombining and highly divergent haplotype clusters underlie these regions of elevated divergence. A local PCoA analysis based on the SNP genotypes within each SV consistently clustered individuals into four distinct groups (Fig. 5c, Fig S2): (I) a cluster of individuals from the highly divergent species *C. linelli*, (II) a cluster of individuals from the highland species *C. leleuporum*, *C. galapageium* and the highland populations of *C. granatense*, (III) a cluster that mainly comprises individuals from the lowland species *C. granatense* and (IV) a cluster containing individuals of both high-and lowland species and populations that was situated intermediate clusters II and IV. This clustering is consistent with the presence of a distinct high (H) - and lowland (L) allele, with clusters II and III comprising individuals homozygous for the high- and lowland allele respectively, and cluster IV corresponding to individuals being heterozygous for both alleles. Heterzogosity for a distinct high- and lowland allele in the individuals of this latter cluster (IV) was supported by their markedly higher nucleotide diversity at the SV compared with those considered homozygote for one of the two alleles (clusters II and III, Fig. 5d). These elevated levels of nucleotide diversity in heterozygotes were moreover maintained across the entire length of the SV, which provides additional support for the lack of recombination among those divergent haplotypes (Fig. SX).

Phylogenetic relationship at chromosomal inversions associated with high-lowland divergence

To infer the evolutionary history of these haplotype clusters, we constructed maximum likelihood phylogenies of the haplotypes present at each SV. Haplotypes associated with all highland species and populations, including the highly distinct haplotypes of the most divergent highland species *C. linelli* from San Cristobal, consistently clustered with high confidence into a monophyletic clade (Fig. 5e, Fig SX). Thus, haplotypes selected in all highland species and populations appear to have a single evolutionary origin and subsequently spread across all islands. Spread of highland alleles generally followed the island progression as shown by a consecutive split of highland individuals according to island age for six SV (Fig5e, Fig SX). This progressive spread was further corroborated by a highly significant and consistent decrease in nucleotide diversity of the highland alleles towards younger islands for all but one SV (Fig. 5f, Fig SX, Table S2).

The highly consistent phylogenetic and nucleotide diversity patterns across the different SV could, at least partially, be caused by a tight physical linkage among the investigated SV. However, none of the SV showed an identical segregation pattern across the 32 investigated individuals, which demonstrates that they represent independently evolved loci with a shared evolutionary history (Fig. SX).

We investigated if the divergences between the high- and lowland associated alleles across all SVs evolved during a singular high-lowland divergence event by comparing their splitting times (Fig. 6). Estimated divergence times, expressed relative to the divergence from the mainland species *C. sayi*, ranged between 0.5 and 0.75 and 95% posterior density intervals of the splitting times overlapped for six of the nine SV. These estimated divergence times were centered around the estimated divergence time of the most ancient high-lowland divergence in our species phylogeny, being the split that gave rise to the highland species *C. linelli* and the remaining species.

Discussion

Archipelagos where island emerged following a known chronosequence have proven key to reconstruct the processes and mechanisms of species diversification [2,3,23]. Here, we infer the genomic underpinning of a gradual divergence of multiple highland species and populations from a single lowland species along the progression of the Galapagos islands. Patterns of genome-wide divergence corroborated the decreasing genomic divergence of highland species and populations from the lowland species towards younger islands, which hints towards a gradual, repeated, and independent within-island evolution of distinct highland species and populations on the different islands.

However, adaptation towards highland habitats generally involved selection within the same genomic regions across all highland species. These genomic regions were characterized by extensive chromosomal inversions, often extending multiple megabases, and resulted in distinct and non-recombining haplotype groups associated with high-lowland divergence. The well-supported monophyletic clustering of haplotypes within each group provides strong evidence that a single evolutionary origin underlies the evolution of highland alleles. Across loci, high- and lowland selected alleles moreover diverged within a similar time frame that overlapped with the estimated split time of the most anciently diverged highland species *C. linelli* from San Cristobal. Thus, evolution of highland species and more recent highland populations appears driven by selection on alleles that evolved during a singular high-lowland divergence that can be traced back to the most ancient high-lowland species divergence. Whether this initial high-lowland divergence occurred *in-situ* on San Cristobal or at separate islands remains an open question. This initial divergence might even have resulted from a secondary colonization at Galapagos, wherein a long-winged species came into secondary contact with a short-winged highland species that evolved in isolation on the Galapagos.

The phylogenetic pattern of haplotypes associated with highland habitats depicted consecutive splits consistent with the chronosequence of the islands. Combined with a consistent decrease in nucleotide variation towards younger islands, this strongly supports a progressive spread of highland alleles towards younger islands. This propagation of highland alleles towards more recent islands might either have occurred through (i) dispersal of highland species between islands (Fig. 1d) (ii) immigration of few highland individuals that introduced highland alleles in the present

lowland species ('adaptive introgression' [24], [24], Fig. 1c) or (iii) dispersal of lowland individuals that are polymorphic at loci involved in high-lowland divergence (Fig. 1b). Our data suggest that several of these mechanisms took place within this radiation. Dispersal of the highland species towards younger islands was strongly supported by the sister relationship of the highland species *C. leleuporum* (Santa Cruz) and *C. galapageium* (Santiago), which indicates that the highland species of Santiago evolved from the ancestral highland species of Santa Cruz rather than through an *in-situ* high-lowland divergence on Santiago. This appears plausible given the connection between both islands about 1Mya [25]. However, subsequent gene-flow with the lowland species *C. granatense* appeared to be substantial and almost completely erased the phylogenetic signature of the sister relationship between those two highland species. Indeed, our genome-wide phylogeny revealed that only about 12% of the genome (gCF) still supports the sister relationship between these highland species. In contrast, spread of highland alleles by dispersal of polymorphic individuals of the lowland species *C. granatense* likely resulted in the introduction of highland alleles took at least place between the more recent islands Isabela and, particularly, Fernandina. For these islands, we found that winged lowland individuals of *C. granatense* exhibit a high degree of polymorphism at loci under divergent selection, which may easily result in the transport of highland alleles between islands. Moreover, highland habitats of these islands are not populated by distinct highland species, but by individuals of the lowland species with reduced wings and a higher frequency of highland alleles, which precludes dispersal of 'pure' highland individuals between these islands. While decreasing levels of within-island ecotypic divergence towards more recent island are often considered to represent different stages of the speciation process, our results rather point towards a reversal of the high-lowland divergence along this island progression. More precisely, while high- and lowland species are morphologically and genetically well differentiated on the oldest island San Cristobal, their divergence decreases towards the younger islands Santa Cruz and Santiago. Provided the sister relationship of the highland species on these latter two islands, this lower genomic divergence on Santiago implies increased introgression from the lowland species into the highland on this younger island (cfr Fig. 1d). Increased introgression between high- and lowland species on more recent islands may ensue from the stepwise colonization of highland species on islands with a resident lowland population. This leads to decreasing founder population sizes of highland species on more recent islands, which on its turn increases asymmetric introgression from the large population of the resident lowland species *C. granatense* into the smaller colonizing population of highland individuals [26,27]. If the number immigrant highland individuals becomes very small, these individuals may even be more likely to exchange genes with the resident lowland population rather than with other highland immigrants, resulting in adaptive introgression [24] of highland alleles in the resident lowland population (Fig. 1c). Selection on these introgressed alleles in highland populations may then result in a repeated evolution of the highland ecotype. Because introgression of adaptive alleles leads to polymorphisms at adaptive loci, highland alleles may ultimately be introduced by the immigration of such polymorphic individuals (Fig. 1b), a process that likely took place at the youngest islands Isabela and Fernandina of the archipelago.

Patterns of divergence along this island progression therefore provide unique support and an extant illustration of the presumed stages of the emerging 'two-time frames' model of ecotypic divergence, which proposes that contemporary ecotypic evolution is driven by selection on alleles that originate from an old ecotypic divergence event [12,13].

Conclusion

The repeated occurrence of parallel species pairs on insular systems is generally presumed to either result from the dispersal of ecotypes between islands or through repeated *in-situ* radiations [2]. Results from this study demonstrate that distinguishing both mechanisms may be inherently difficult due to the complex introgression patterns between as well as within islands. Habitat patches on islands are highly dynamic with respect to their spatial and temporal distribution, driven by climatic and geological dynamics, resulting in multiple episodes of fission and fusion between

ecotypes that might either erase the original phylogenetic signals of species divergence and drive the inter-island exchange of alleles involved in ecotypic differentiation. Given the frequently reported evidence of interspecific gene-flow in island radiations like Darwin finches [28,29], giant tortoises [30,31], *Hogna* wolf spiders [32] and Hawaiian *Metrisoderos* trees [33], these complex introgression patterns can be expected to be prevalent. [28,29], giant tortoises [30,31], *Hogna* wolf spiders [32] and Hawaiian *Metrisoderos* trees [33], these complex introgression patterns can be expected to be prevalent.

While species dispersal and within-island diversification are generally interpreted as two clearly distinct mechanisms that generate parallel species assemblages on islands, our findings suggest that the distinction between these mechanisms could be more gradual and admixed as currently assumed. Key factors that determine the relative contribution of these two mechanisms in generating parallel species pairs are the number of founding individuals or haplotypes of the alternative ecotype that colonize an island and the amount of interspecific gene flow between the two ecotypes after colonization. Comprehending the admixture between these two mechanisms could help to better predict biodiversity dynamics on islands [34].

Methods

Sampling

We sampled all highland and lowland species or populations from the major islands of the Galapagos archipelago being San Cristobal (SCB), Santa Cruz (SCZ), Santiago (SAN), Isabela (IVA) and Fernandina (FER) (Table S1, Fig 2). These are the only islands within the archipelago from which distinct highland ecotypes or species have been reported. The island Isabela is the only island within the Galapagos that comprises multiple volcanoes and we restricted our sampling to the most centrally located Volcan Alcedo. Individuals were sampled during different sampling campaigns between 1996 and 2014 (Table S1) and stored live in liquid nitrogen or pure ethanol shortly after sampling. While high- and lowland species are easily identified in the field on the oldest islands San Cristobal, Santa Cruz and to a lesser extent Santiago [35], divergence of highland ecotypes varies more gradually on the younger islands Isabela and Fernandina. To ensure that individuals of the highland ecotype were sampled at these islands, only those individuals sampled at the outermost volcano summit (1110m and 1290m at Isabela and Fernandina respectively) were considered as highland ecotypes. We included between 11 and 22 individuals per population for genetic analysis, except for Fernandina where only five individuals of the high- and lowland population could be sampled. A single individual of the mainland species *Calosoma sayi*, sampled by M. Husemann in Texas, USA, was included as outgroup species. This *Calosoma* species is one of the most closely related species with those found at the Galapagos and taxonomically classified within the same subgenus (*Castrida*) [36].

Genome assembly

We assembled the genome of *Calosoma granatense* using both paired-end libraries with short insert sizes of 170bp, 500bp and 800bp and mate-paired libraries with insert sizes of 2kb, 5kb, 10kb and 20kb. Short-insert size libraries were all constructed from a single individual sampled at Santa Cruz at 350m altitude, while long-insert mate paired libraries were constructed from DNA extracts from nine different individuals that all originated from this same locality (Table S1). Total DNA was extracted from these individuals with the NucleoSpin ® Tissue Kit, Macherey-Nagel GmBH and library construction and sequencing performed at the Bejing Genomic Institute, Hongkong. Sequencing errors were corrected based on the *k*-mer frequency spectrum with SOAPec [37], specifying a *k*-mer value of 17. Corrected reads were then used as input for genome assembly with Platanus [38] using default settings. Contigs were constructed based on the short-insert libraries only with the 'platanus assemble' tool, and subsequently combined into scaffolds with 'platanus scaffold' using both short- and long-insert libraries. Gaps between the scaffolds were

finally filled with the 'platanus gap_close' tool using both short- and long-insert libraries. The final assembly consisted of 6045 scaffolds summing to a size of 167,880,245bp (Table S3). We estimated the size of the genome by obtaining the *k*-mer frequency spectrum of all resequenced individuals separately (see Methods section – Whole genome resequencing) with Jellyfish v2.3.0 [39] and analyzed the frequency distribution with GenomeScope [40] . These analyses yielded an estimated genome size of 173.3 Mb (± 6.0 SD) for a *k*-mer = 21 and highly similar estimates for other tested *k*-mer sizes (*k*-mer= 17: 172.6 Mb ± 8.2SD; *k*-mer = 31 : 172.6 Mb ± 8.2SD). Based on these estimates, the assembled genome represents 97% of the estimated genome size. Completeness of the assembly was further assessed based on a set of 1,658 benchmarked single copy orthologs (BUSCO's) from Insecta [41]. Screening the draft genome for these BUSCO's revealed that 85.3% were present in our assembly, of which 0.3% duplicated and 8.4% fragmented (Table S3).
We screened the genome for repetitive elements with RepeatMasker v1.295 [42] specifying 'Coleoptera' as species and constructed a library of *de novo* repetitive elements with RepeatScout v1.0.5 [43]. Both methods resulted in a total repeat content of 11.04% (Table S3).

Phylogeny

We inferred the phylogenetic relationship between the species and populations based on 500 windows of 20kb each. Windows were selected with a custom python script (SelectRandomWindows.py) and written to gff format specifying the start and end-positions of each window. Fasta files containing the individual sequences for each window were then extracted with the vcf2fasta.pl tool (https://github.com/santiagosnchez/vcf2fasta), using the Cgran1.0 reference genome, genome wide VCF and gff-file specifying the locations of the 20kb windows as input files. We estimated maximum likelihood trees for each window-specific fasta with IQ-TREE [44], specifying 1000 ultra-fast bootstrap samplings [45]. Before tree estimation, best substitution models for each window were selected using ModelFinder [46] as implemented in IQ-TREE. ML trees were visualized with the densiTree function implemented in the *phangorn* v2.5.5 [47] package in R v4.0.3. To ease visualization, trees were first made ultrametric by transforming the branches proportionally in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree). We then reconstructed a multispecies phylogeny that accounts for the potential discordance in the window-specific phylogenies (e.g. due to incomplete lineage sorting) using ASTRAL [48,49]. Besides the calculation of branch support values as implemented in ASTRAL [49], we also calculated gene concordance factors (gCF) [50], which express the percentage of gene (window) trees containing this branch. gCF's were obtained from IQ-TREE by specifying the multi-species consensus tree from ASTRAL as reference tree.

Structural variation analysis

Patterns of genomic divergence (F_{ST}) and outlier analysis revealed that sites with elevated divergence between high- and lowland species were generally clustered into contiguous genomic regions, potentially indicating the presence of structural variations (SV) like inversions or translocations. We searched for the presence of SV in all scaffolds containing at least ten consecutive 20kb windows (200kb in total) with an $F_{ST} > 0.1$. Average weighted F_{ST} for each 20kb window, calculated with VCFtools v0.1.16 [51], was based on an overall comparison between all high- versus lowland species and populations of all resequenced individuals. This selection procedure resulted in twelve scaffolds with continuous regions of elevated divergence that all contained at least two RADtags with an outlier SNP in one of the within-island high-lowland comparisons (see "Outlier loci detection"), which additionally supports that these regions are associated with high-lowland differentiation in at least one of the islands.
We used BreakDancer v1.3.6 [52] to screen for anomalies in the insert size or orientation of read pairs that flanked genomic region of elevated divergence. More precisely, BreakDancer v1.3.6 was run on all individual bam files and we searched for SV whose breakpoints (i) are located within ±20kb of the boundaries of the region with elevated

divergence; (ii) have a maximal quality score of Q = 99 and (iii) are supported by a significantly different sequencing coverage between high- and lowland individuals (Welch *t*-test; $P < 0.05$). Regions matching these criteria were considered as continuous SV for further analysis.

We tested if the sequence composition at SV correspond to the presence of distinct high- and lowland alleles by means of a PCoA analysis on SNPs located within each SV. If haplotypes at the SV represent distinct alleles, we expect individuals in the PCoA to be clustered into three distinct groups corresponding to individuals homozygote for the highland allele, individuals homozygote for the lowland allele and a potential group of heterozygote individuals that are situated intermediate between both homozygote groups. To further confirm that the three clusters correspond to individuals with different genotypes for distinct highland or lowland alleles, we calculated average nucleotide diversity (π) at the SV and tested if π is significantly higher in the individuals in the heterozygote cluster compared to those in the two homozygote clusters. PCoA analysis was perfomed with the adegenet 2.1.3 [53] package in R v.4.0.3 using SNPs located within each SV and filtered with VCFtools v0.1.16 [51] for a genotype quality > 30 and presence in all 32 resequenced individuals and kept one out of 1000 SNPs to reduce computational time.

We tested if the SV reduced recombination between high- and lowland associated alleles by calculating pairwise r^2 values between SNP genotypes across the entire length of the selected scaffolds and plotted the distribution of SNPs that are in perfect linkage disequilibrium ($r^2=1$). If the SV suppresses recombination between both alleles, $r^2 = 1$ values are expected across the entire length of the SV. We further compared patterns of r^2 between this set that includes all individuals and a set that only includes individuals that are homozygous for the allele associated with the lowland ecotype and, thus, expected to show patterns of free recombination. r^2 calculations were performed with VCFtools v0.1.16 [51] based on the same vcf file as used for the PCoA analysis (min GQ > 30, genotypes present in all individuals), but additionally filtered for a minimum allele frequency of 0.05 and an additional SNP thinning of either 1/1000 or 1/10,000 to reduce the number of SNPs to less than 500. Suppression of recombination along the SV of interest was further tested by comparing profiles of nucleotide diversity (π) between individuals that are homo- en heterozygous for the SV, wherein heterozygous individuals are expected to consistently show higher nucleotide diversity compared to homozygotes across the entire length of the SV.

Phylogenetic relationships between the haplotypes located at the SV were estimated by maximum likelihood with IQ-TREE [44], specifying 1000 ultra-fast bootstrap samplings [45]. Before tree estimation, best substitution models for each window were selected using ModelFinder [46] as implemented in IQ-TREE. We only included individuals that are homozygous for the SV to estimate phylogenetic relationships because phasing errors in individuals that are heterozygote for the highly divergent alleles may lead to erroneous recombinant haplotypes and highly inaccurate phylogenies [54]. The sequence of *C. sayi* was specified as outgroup.

We compared the timing of the divergence between high- and lowland alleles across the different SV using the divergence between the *Calosoma* species from the Galapagos and the outgroup species *C. sayi* as relative calibration point. Analyses were performed with BEAST v2.6.0 [55], specifying the best fiiting substitution model for each SV as selected by ModelFinder in our IQ-TREE analysis, a random local clock model, empirical base frequencies, and Yule tree prior. We further specified a clade with individuals with the highland allele and a clade with individuals with the lowland allele, which were both constrained to be monophyletic if supported (>95%) by our maximum likelihood analysis. Analyses were run for 50M generations and only used samples from the stationary phase of the Markov chain, comprising at least 25M generations, for consensus tree construction and divergence time estimation. Mean and 95% HPD of the height of the split between the clades comprising individuals with the high- or lowland alleles was then divided by the height of the split between *C. sayi* and the Galapagos radiation to compare the relative divergence time between both alleles among the different scaffolds.

Relative divergence of this split between high-and lowland alleles was compared with the relative divergence time of the different highland species. Timing of the split of these highland species, relative to the divergence from the outgroup *C. sayi*, was estimated by running a multispecies multilocus coalescence analysis with *BEAST [55] based on a random selection of 50 windows of 20kb (see Phylogeny in Methods section for details on the extraction of fasta

files for these 20kb windows) that were located outside the SV. We only selected those windows for which ModelFinder reported a HKY substitution model as this allowed us to specify the same HKY substitution model for all 50 windows simultaneously in the *BEAST analysis. Like our previous analysis on individual SV, we specified a random local clock model, empirical base frequencies, and a Yule tree prior.

Literature

1. Schluter, D. *The ecology of adaptive radiation*. (Oxford University Press, 2000).
2. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
3. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
4. Schluter, D. Evidence for Ecological Speciation. *Science* **323**, 737–741 (2009).
5. Nosil, P. *Ecological speciation*. (Oxford University Press, 2012).
6. Losos, J. B. *et al.* Niche lability in the evolution of a Caribbean lizard community. *Nature* **424**, 542–545 (2003).
7. Gillespie, R. Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356–359 (2004).
8. Bierne, N., Gagnaire, P. A. & David, P. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr. Zool.* **59**, 72–86 (2013).
9. Faria, R. *et al.* Advances in Ecological Speciation: An integrative approach. *Mol. Ecol.* **23**, 513–521 (2014).
10. Welch, J. J. & Jiggins, C. D. Standing and flowing: The complex origins of adaptive variation. *Mol. Ecol.* **23**, 3935–3937 (2014).
11. Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol Evol* **23**, 38–44 (2008).
12. Van Belleghem, S. M. *et al.* Evolution at two time frames : Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genet.* 1–26 (2018). doi:10.1371/journal.pgen.1007796
13. Marques, D. A., Meier, J. I. & Seehausen, O. A Combinatorial View on Speciation and Adaptive Radiation. *Trends Ecol. Evol.* **xx**, 1–14 (2019).
14. Foote, A. D. Sympatric Speciation in the Genomic Era. *Trends Ecol. Evol.* **xx**, 1–11 (2017).
15. Yang, M., He, Z., Shi, S. & Wu, C. I. Can genomic data alone tell us whether speciation happened with gene flow ? *Mol. Ecol.* 2845–2849 (2017). doi:10.1111/mec.14117
16. Seehausen, O. *et al.* Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–92 (2014).
17. Hendrickx, F. *et al.* Persistent inter- and intraspecific gene exchange within a parallel radiation of caterpillar hunter beetles (Calosoma sp .) from the Gal apagos. *Mol. Ecol.* **24**, 3107–3121 (2015).
18. Desender, K., Baert, L. & Maelfait, J.-P. Evolutionary systematics of Calosoma WEBER carabid beetles of the Galápagos Archipelago, Ecuador (Coleoptera: Carabidae). in *Advances in Coleopterology* (eds. Zunino, M., Bellés, X. & Blas, M.) 193–200 (AEC, 1991).
19. Gidaspow, T. The genus Calosoma in Central America, the Antilles and South Amercica (Coleoptera, carabidae). *Bull. Am. Museum Nat. Hist.* **124**, 1–208 (1963).
20. Sota, T., Takami, Y., Ikeda, H., Liang, H. & Karagyan, G. Molecular Phylogenetics and Evolution Global dispersal and diversification in ground beetles of the subfamily Carabinae. *Mol. Phylogenet. Evol.* **167**, 107355 (2022).
21. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* **8**, e1002967 (2012).
22. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–93 (2008).
23. Gillespie, R. G. Island time and the interplay between ecology and evolution in species diversification. *Evol. Appl.* **9**, 53–73 (2015).
24. Edelman, N. B. & Mallet, J. Prevalence and Adaptive Impact of Introgression. *Annu. Rev. Genet.* 1–19 (2021).

25. Geist, D. J., Snell, H., Snell, H., Goddard, C. & Kurz, M. D. A Paleogeographic Model of the Galápagos Islands and Biogeographical and Evolutionary Implications. *Galápagos A Nat. Lab. Earth Sci. Geophys. Monogr. 204* 145–166 (2014).
26. Currat, M., Ruedi, M., Petit, R. J. & Excoffier, L. The hidden side of invasions: Massive introgression by local genes. *Evolution* **62**, 1908–1920 (2008).
27. Petit, J. & Excoffier, L. Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393 (2009).
28. Farrington, H. L., Lawson, L. P., Clark, C. M. & Petren, K. The evolutionary history of darwin's finches: Speciation, gene flow, and introgression in a fragmented landscape. *Evolution* 2932–2944 (2014). doi:10.1111/evo.12484
29. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* (2015). doi:10.1038/nature14181
30. Garrick, R. C. *et al.* Lineage fusion in Galápagos giant tortoises. *Mol. Ecol.* **23**, 5276–5290 (2014).
31. Emerson, B. C. & Faria, C. M. A. Fission and fusion in island taxa – serendipity , or something to be expected ? *Mol. Ecol.* **23**, 5132–5134 (2014).
32. De Busschere, C., Van Belleghem, S. M. & Hendrickx, F. Inter and intra island introgression in a wolf spider radiation from the Galápagos, and its implications for parallel evolution. *Mol. Phylogenet. Evol.* **84**, (2015).
33. Choi, Y. J. *et al.* Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands. *Proc. Natl. Acad. Sci.* **118**, 1–10 (2021).
34. Valente, L. *et al.* A simple dynamic model explains the diversity of island birds worldwide. *Nature* **579**, 92–96 (2022).
35. Desender, K. & Dijn, B. De. The Calosoma species of the Galápagos archipelago. I. Redescription and distribution of the species. *Bull. R. Belgian Inst. Nat. Sci. - Entomol.* **59**, 131–144 (1989).
36. Gidaspow, T. The genus Calosoma in Central America, the Antilles, and South America (Coleoptera, Carabidae). *Bull. Am. Museum Nat. Hist.* **124**, 275–314 (1963).
37. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
38. Kajitani, R. *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 1384–1395 (2014). doi:10.1101/gr.170720.113.Freely
39. Marçais, G. & Kingsford, C. A fast , lock-free approach for efficient parallel counting of occurrences of k - mers. *Bioinformatics* **27**, 764–770 (2011).
40. Vurture, G. W. *et al.* Genome analysis GenomeScope : fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
41. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E. M. Genome analysis BUSCO : assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
42. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker. http://www.repeatmasker.org/. (2014).
43. Price, A. L., Jones, N. C. & Pevzner, P. a. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351-358 (2005).
44. Nguyen, L., Schmidt, H. A., Haeseler, A. Von & Minh, B. Q. IQ-TREE : A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
45. Hoang, D. T., Chernomor, O., Haeseler, A. Von, Minh, B. Q. & Vinh, L. S. UFBoot2 : Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
46. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. Von & Jermiin, L. S. ModelFinder : fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
47. Schliep, K. P. phangorn : phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
48. Mirarab, S. *et al.* ASTRAL : genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, 541–548 (2014).
49. Rabiee, M., Sayyari, E. & Mirarab, S. Multi-allele species reconstruction using ASTRAL. *Mol. Phylogenet. Evol.* **130**, 286–296 (2019).
50. Minh, B. Q., Hahn, M. W. & Lanfear, R. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Mol. Biol. Evol.* **37**, 2727–2733 (2020).
51. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
52. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat.*

Methods **6**, 677–81 (2009).
53. Jombart, T. & Ahmed, I. adegenet 1 . 3-1 : new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
54. Posada, D. & Crandall, K. A. The Effect of Recombination on the Accuracy of Phylogeny Estimation. *J. Mol. Evol.* **54**, 396–402 (2002).
55. Bouckaert, R. *et al.* BEAST 2 . 5 : An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, 1–28 (2019).

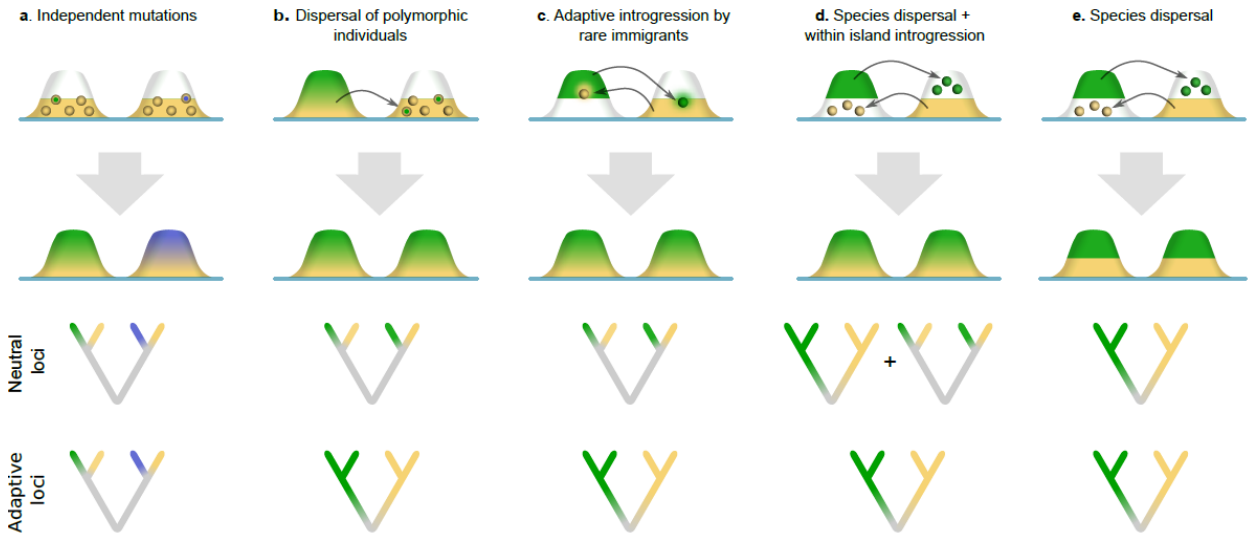


Figure 1 | Schematic overview of different scenarios that may result in similar ecotypic species pairs on islands and the expected haplotype relationships at neutral and adaptive loci. a. Repeated evolution by independent mutations. A unique mutation within each island results in the independent evolution of ecotypes within each island. Ecotypes residing on the same island will be most closely related at both neutral and adaptive loci. **b. Dispersal of polymorphic individuals.** Introgression between both ecotypes within an island result in polymorphic individuals that may disperse between islands. Divergent selection on these polymorphisms result in the repeated evolution of both ecotypes. **c. Adaptive introgression by rare immigrants.** Dispersal of a single or few individuals of the alternative ecotype that hybridize with the resident ecotype results in adaptive introgression of the alternative alleles. At neutral loci, introgressed haplotypes will be lost by drift due their low frequency. At adaptive loci, selection may maintain introgressed haplotypes and drive the repeated evolution of both ecotypes. **d. Dispersal of ecotypic species between island, followed by subsequent within island introgression.** If reproductive isolation between the ecotypic species is incomplete, gene-flow within islands may result in a closer relationship between species from the same island at a part of the neutral loci. At adaptive loci, introgression is reduced due to divergent selection and maintains the close relationship of ecotypic similar species from different islands. **e. Dispersal of ecotypic species between islands.** If both ecotypes are reproductively isolated, populations from the same ecotypic species will be most closely related at both neutral as adaptive loci.

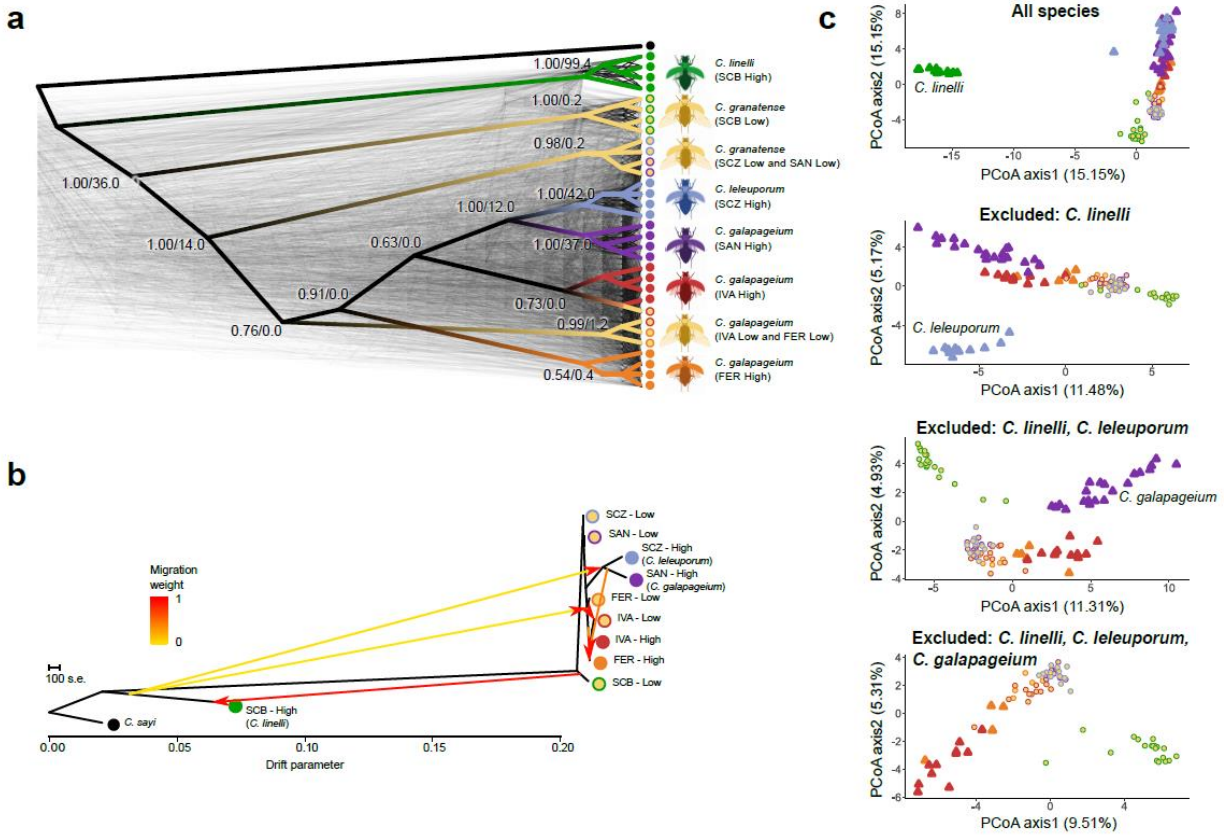


Figure 3 | Genetic relationship among the *Calosoma* species and populations at the Galapagos. a. Phylogenetic relationship between the species and populations as inferred from maximum likelihood (ML) trees from 500 random genomic windows of 20kb (light grey trees), with the multispecies phylogeny (ASTRAL [49]) superimposed. Branch labels show branch support values (before slash) and gene concordance factors (gCF [50]), which express the percentage of ML trees containing this branch (after slash). **b**. Population tree and population admixture (arrows) of the different *Calosoma* populations and species, based on SNP data obtained from RADtag sequencing, using TreeMix with six migration edges. **c**. Principal coordinate analysis of the species and populations based on SNP data obtained from RADtag sequencing. The most divergent species was sequentially excluded towards lower panels. Color codes of the different species and populations are as in Fig. 2.

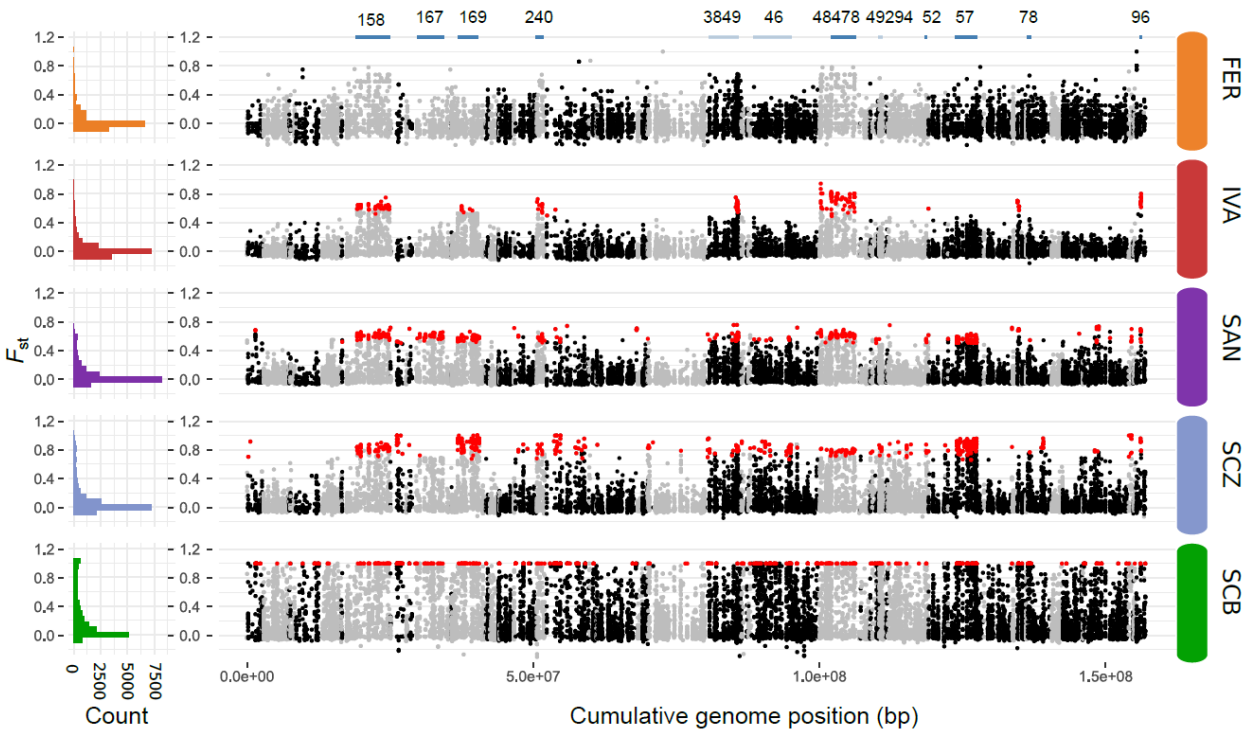


Figure 4 | Genomic divergence between the high- and lowland species or populations within each island. Differentiation (F_{st}) between high- and lowland individuals within each island at individual SNPs obtained from RADtag sequencing. Left histograms show the F_{st} frequency distribution across the entire genome and right panels shows their location at the different genomic scaffolds. SNPs indicated in red are outlier SNPs with a significantly higher differentiation than expected by chance within each island comparison (BayeScan [22] ; Q-value < 0.1). Upper blue line segments show the location of genomic regions tested for the presence of structural variations (SV), with those in dark blue having SV that are supported based on anomalies in the orientation and insert size of read mappings as detected by *breakdancer* [52]. Scaffold ID's are given above each SV.

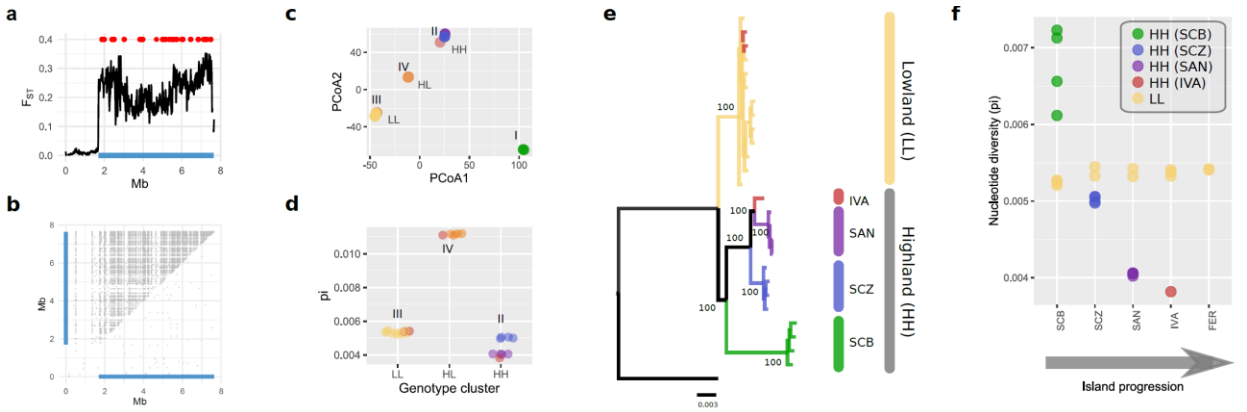


Figure 5 | Structural variations (SV) underlie distinct alleles associated with high-lowland divergence. Results of a single scaffold (scaffold158) are shown. Results for the other scaffolds with SV associated with high-lowland divergence are shown in Fig. SX. **a**. F_{ST} distribution (20kb windows) based on a comparison between all resequenced high- versus lowland individuals. Blue bottom line shows the location of the chromosomal rearrangements detected by *breakdancer* [52]. Red dots show the location of RADtags identified as outlier loci in at least one within-island highland-lowland comparison. **b**. Location of SNPs in perfect linkage disequilibrium (r^2 = 1). Grey dots above diagonal show r^2 = 1 values for all 32 resequenced individuals. Grey dots below diagonal show r^2 = 1 values for homozygous LL individuals only. **c**. PCoA based on SNPs located at the inversion (blue line in panel A). HH, LL and HL refer to the cluster of individuals genotyped at the inversion as homozygous for the highland allele (HH), lowland allele (LL) and heterozygous (HL). **d**. Differences in nucleotide diversity at the inversion between individuals genotyped as HH, HL and LL in panel C. **e**. Maximum likelihood tree of the nucleotide sequence at the inversion excluding individuals genotyped as heterozygotes (HL). Node values represent bootstrap values based on 1000 replicates. The tree was rooted with the mainland species *C. sayi*. **f**. Relationship between individual nucleotide diversity at the inversion and the progression of the islands. Only individuals genotyped as homozygous for the lowland allele (LL, yellow) and highland allele (HH, remaining colors) are included. Color codes are as in Fig. 2.

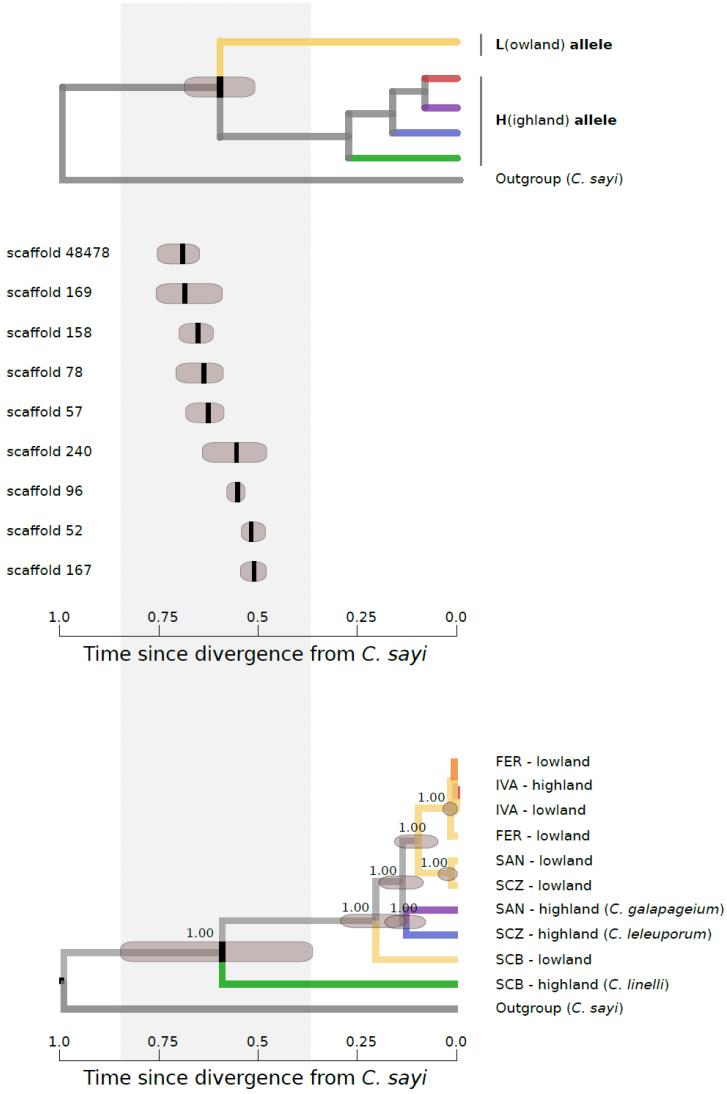


Figure 6 | Relative divergence time between high- and lowland alleles for nine different SV compared to the relative divergence times of the species and populations. Upper panel shows the estimated divergence time (black vertical bars: mean; dark grey bar: 95% highest posterior density) between high- and lowland alleles for nine different structural variations (SV) associated with high-lowland divergence. Lower panel shows the relative divergence times of the different species and populations based on random selection of 50 genomic windows of 20kb that are located outside the SV. Node values represent posterior probabilities of the clades. Both panels are scaled to the same relative time scale expressed as time since the divergence from the mainland species *C. sayi*.

Table S4. – Summary of the SV detected by *breakdancer* v1.3.6 that overlap with the largest genomic regions with elevated divergence. Scaffold: scaffold with the inversion. SV type, start, end and length: type, start and end of the SV reported by *breakdancer*. Q: quality-score of the SV reported by *breakdancer*, with Q = 99 being the maximal score. $P_{high\text{-}low}$: *P*-value (Welch-test) of the difference in the number reads that support the inversion between high- and lowland individuals. *n* RADtags: number of RADtags with an outlier SNP that is shared in at least two within island ecotype comparisons. P_{geno}, $P_{HH\text{-}HL}$, $P_{HH\text{-}LL}$ and $P_{HL\text{-}LL}$: *P*-value (one-way ANOVA) of the difference in nucleotide diversity at the SV between individuals identified as homozygous highland (HH), homozygous lowland (LL) and heterozygous (HL) and the respective pairwise comparisons between these genotypes by mean of a Tukey HSD test. r_S(HH) and r_S(LL): Spearman rank order correlation between island age and nucleotide diversity of individuals homozygous for the highland allele (HH) or lowland (LL) allele. Next columns show corresponding *P* -values of the correlation.

Scaffold	SV type	SV start	SV end	SV length	Q	$P_{high\text{-}low}$	n RADtag	P_{geno}	$P_{HL\text{-}LL}$	$P_{HH\text{-}LL}$	$P_{HL\text{-}HH}$	r_S(HH)	$P_{rS(HH)}$	r_S(LL)	$P_{rS(LL)}$
scaffold158	Inversion	1,701,112	7,624,939	5,923,827	99	7.20E-05	23	<2E-16	<2E-07	2.2E-06	<2E-07	-0.96	2.24E-07	0.65	0.017
scaffold167	Inversion	760,528	5,294,525	4,533,997	99	0.0038	9	<2E-16	<2E-07	<2E-07	<2E-07	-0.96	3.40E-06	0.07	0.78
scaffold169	Inversion	1,057,791	4,614,186	3,556,395	99	0.01	21	<2E-16	<2E-07	2.7E-05	<2E-07	-0.97	3.28E-07	-0.28	0.33
scaffold48478	Inversion	1,985,964	6,286,057	4,300,093	99	2.09E-07	22	<2E-16	<2E-07	2.1E-06	<2E-07	-0.95	6.80E-06	-0.16	0.61
scaffold57	Inversion	1,291,095	5,011,189	3,720,094	99	4.01E-07	20	<2E-16	<2E-07	<2E-07	<2E-07	-0.93	7.50E-05	-0.5	0.06
scaffold96	Inversion	896,427	1,110,117	213,690	99	4.20E-02	2	<2E-16	<2E-07	<2E-07	<2E-07	-0.97	2.30E-08	-0.13	0.64
scaffold240	Interchromosomal translocation	NA	1,176,655	1,176,655	99	8.00E-04	8	<2E-16	<2E-07	4.6E-03	<2E-07	0.05	0.8	-0.71	0.01
scaffold78	Interchromosomal translocation	NA	788,332	788,332	99	2.11E-06	2	<2E-16	<2E-07	1.1E-07	<2E-07	-0.46	0.04	0.88	0.002
scaffold52[1]	Undetermined	6,571,448	6,783,082	211,634	99	6.47E-04	1	<2E-16	<2E-07	<2E-07	<2E-07	-0.97	1.90E-08	0.21	0.49
Notes															
1	The region of elevated divergence on this scaffold was flanked by a deletion at position 6,571,448 and a deletion at position 6,783,082 that were both statistically supported by a significantly different number of reads between high- and lowland individuals ($P_{high\text{-}low}$ = 0.000647 and $P_{high\text{-}low}$ = 0.0000125 respectively).														

Fig S1. – **Genetic divergence between high- and lowland species or populations versus island age.** Relationship between the percentage of SNPs with an F_{st} > 0.4 in a within island comparison between high- and lowland species or populations in relation to the estimated age of the island. Fst values are based on SNP data obtained from RADtag sequencing. Estimated ages of each island are based on [25] and given between square brackets.

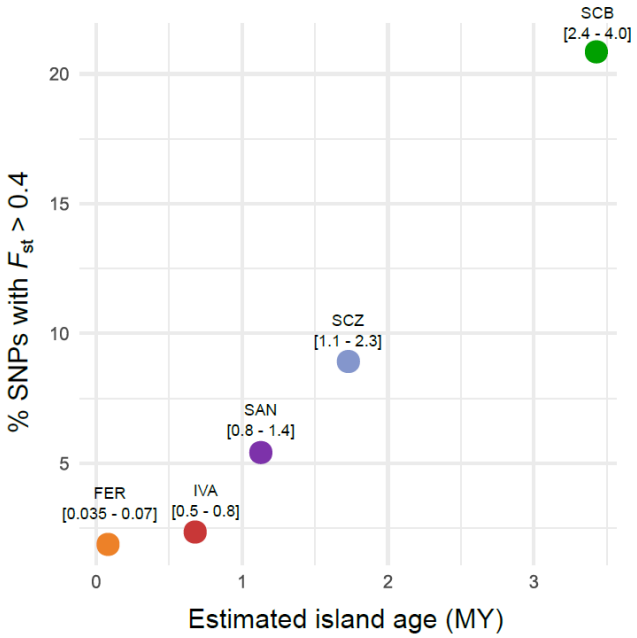
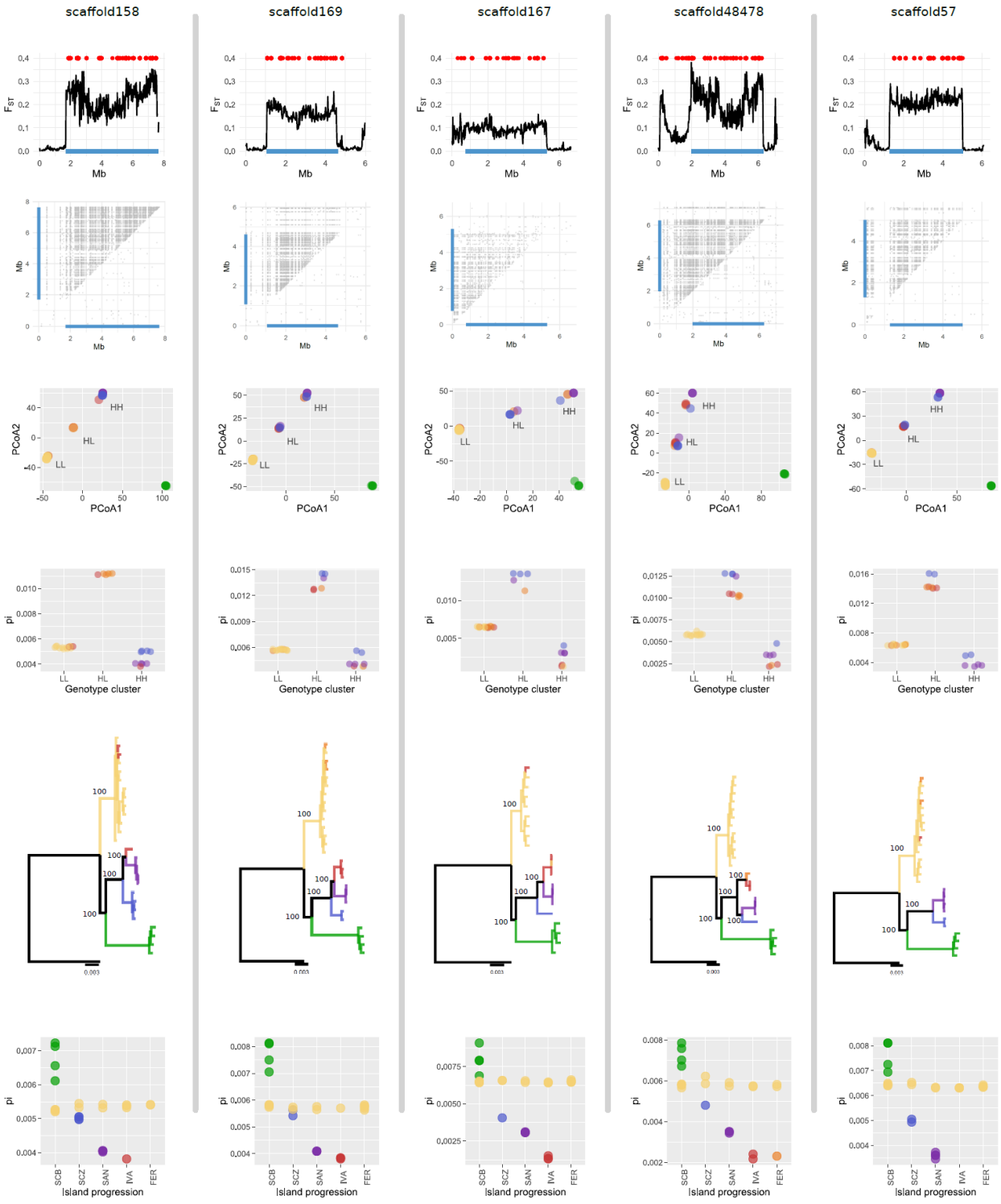


Fig S2. – **Structural variations (SV) underlie distinct alleles associated with high-lowland divergence.** Each column shows the result of a single scaffold on which the SV was located. **First row**: F_{ST} distribution (20kb windows) based on a comparison between all high- versus lowland individuals. Blue bottom line shows the location of the chromosomal inversion detected by *breakdancer* [52]. Red dots show the location of RADtags identified as outlier loci in at least one within-island highland-lowland comparison. **Second row:** Location of SNPs in perfect linkage disequilibrium (r^2 = 1). Grey dots above diagonal show r^2 = 1 values for all 32 resequenced individuals. Grey dots below diagonal show r^2 = 1 values for homozygous LL individuals only. **Third row:** PCoA based on SNPs located at the inversion (blue line in panel A). HH, LL and HL refer to the cluster of individuals genotyped at the inversion as homozygous for the highland allele (HH), lowland allele (LL) and heterozygous (HL). **Forth row**: Differences in nucleotide diversity at the inversion between individuals genotyped as HH, HL and LL in panel C. **Fifth row**: Maximum likelihood tree of the nucleotide sequence at the inversion excluding individuals genotyped as heterozygotes (HL). Node values represent bootstrap values based on 1000 replicates. The tree was rooted with the mainland species *C. sayi*. **Sixth row**: Relationship between individual nucleotide diversity at the inversion and the progression of the islands. Only individuals genotyped as homozygous for the lowland allele (LL, yellow) and highland allele (HH, remaining colors) are included. Color codes are as in Fig. 2.


scaffold158
scaffold169
scaffold167
scaffold48478
scaffold57
FST
Mb
PCoA1
PCoA2
HH
HL
LL
pi
Genotype cluster
Island progression
SCB
SCZ
SAN
IVA
FER

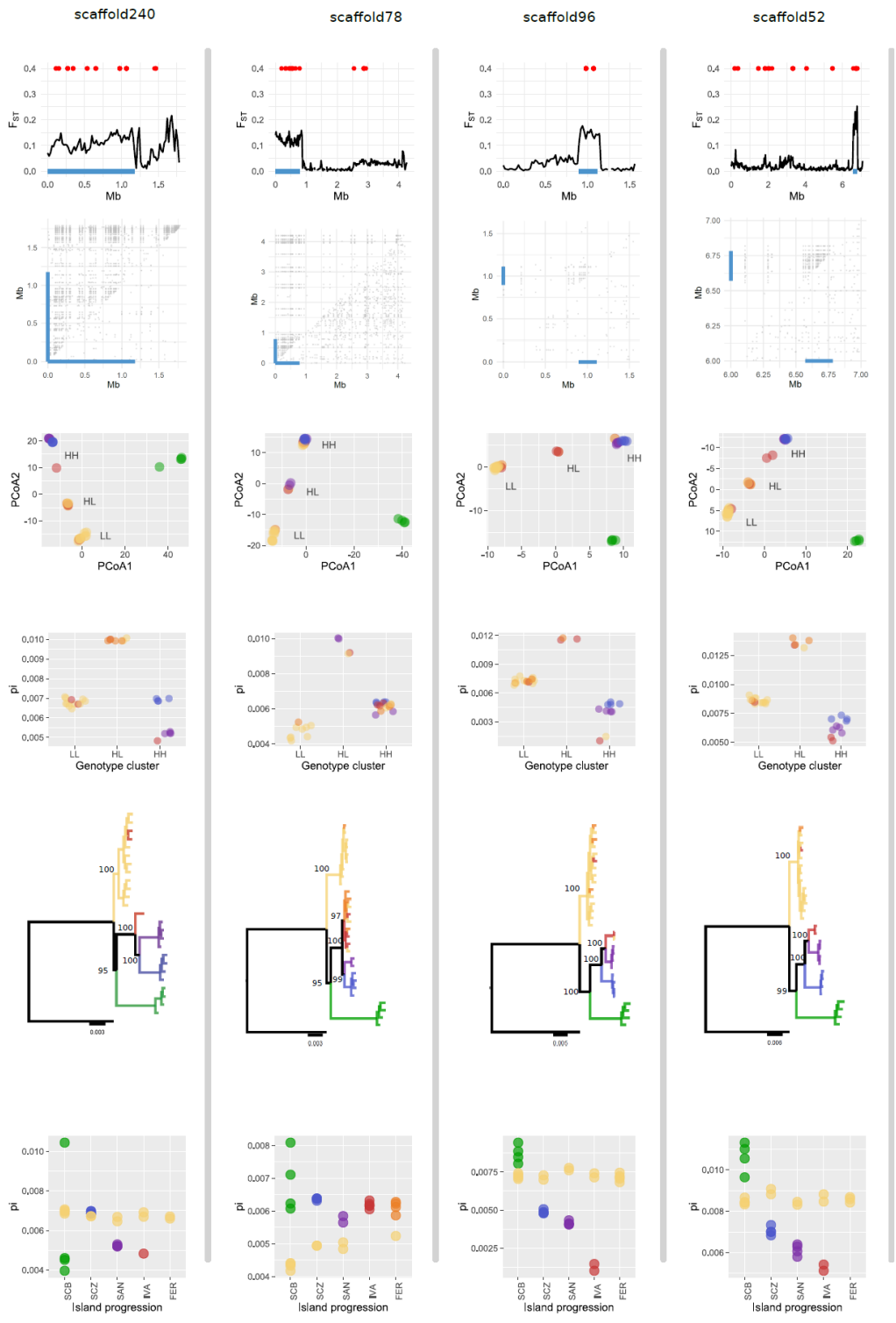

scaffold240
scaffold78
scaffold96
scaffold52
FST
Mb
PCoA1
PCoA2
HH
HL
LL
pi
Genotype cluster
Island progression
SCB
SCZ
SAN
IVA
FER

Fig S3. – Segregation pattern of high- and lowland alleles at the nine investigated SV associated with high-lowland divergence. Yellow squares are individuals homozygote for the lowland allele; grey squares are individuals heterozgote for the high and lowland allele; remaining colored squares are individuals homozygote for the highland allele, with the colors corresponding to the highland population and species colors shown in Fig. 2.

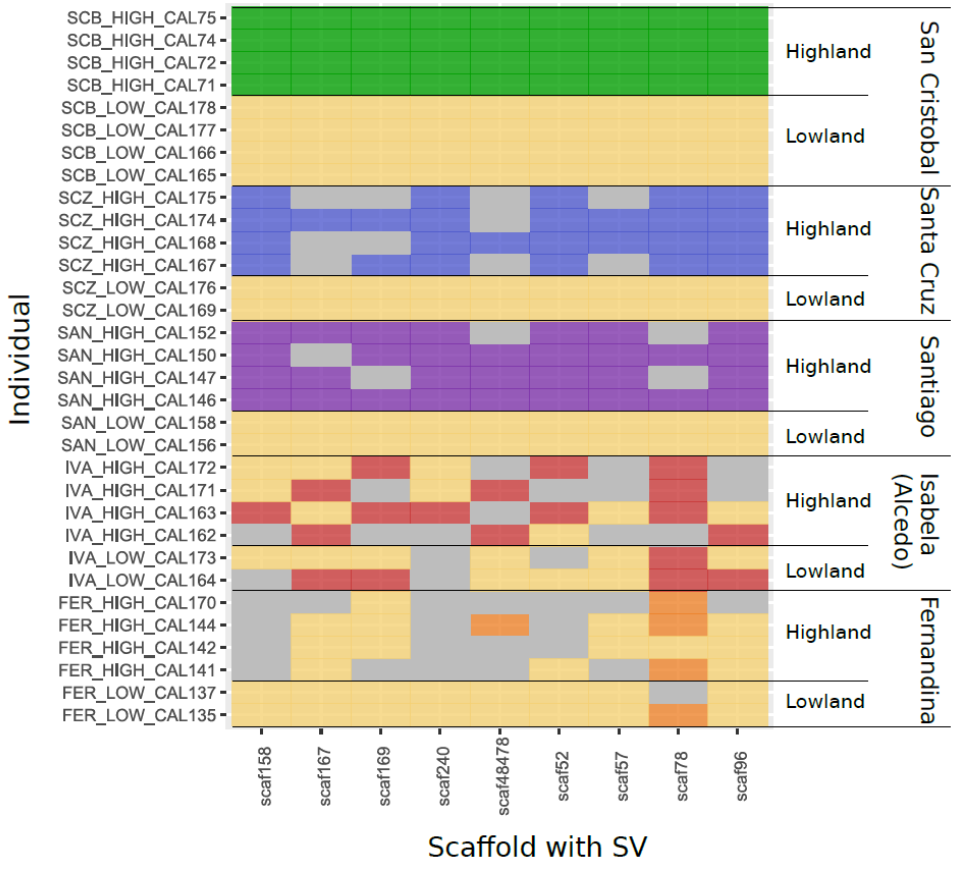
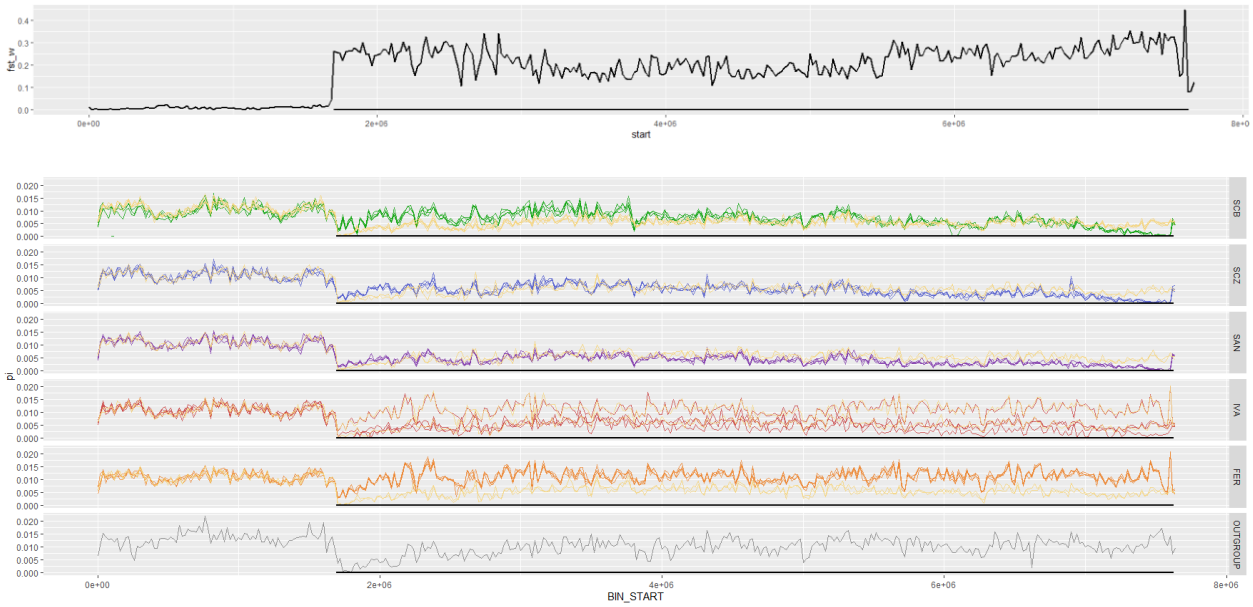
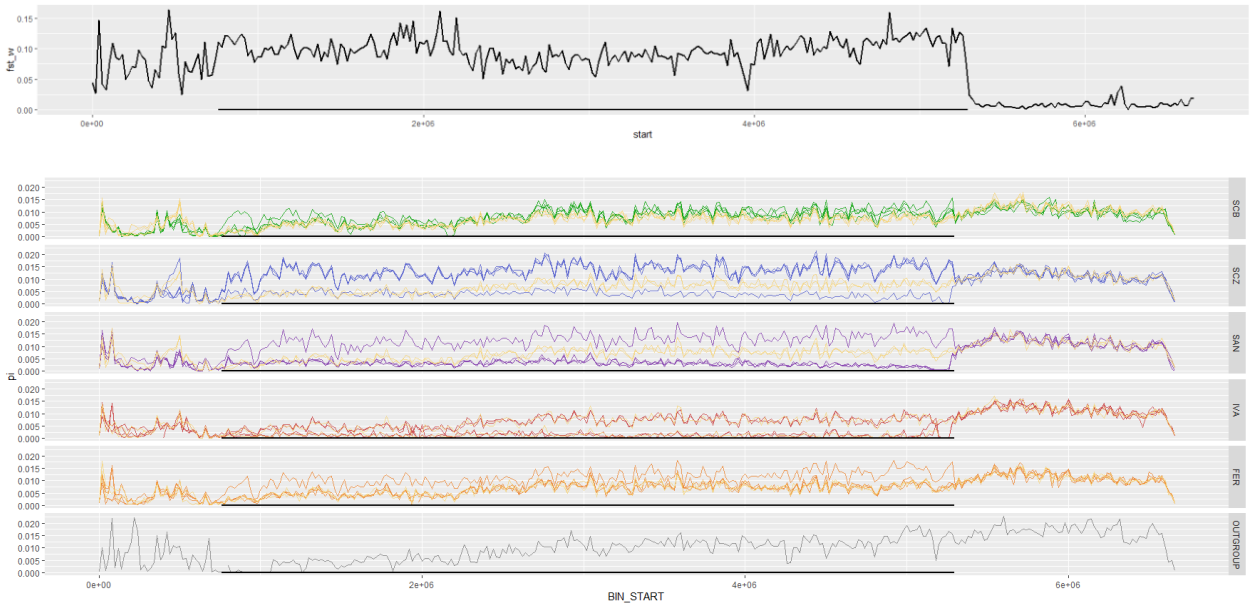


Fig S4. – Detailed patterns of genetic differentiation and nucleotide diversity at scaffolds with structural variations associated with highland-lowland divergence.

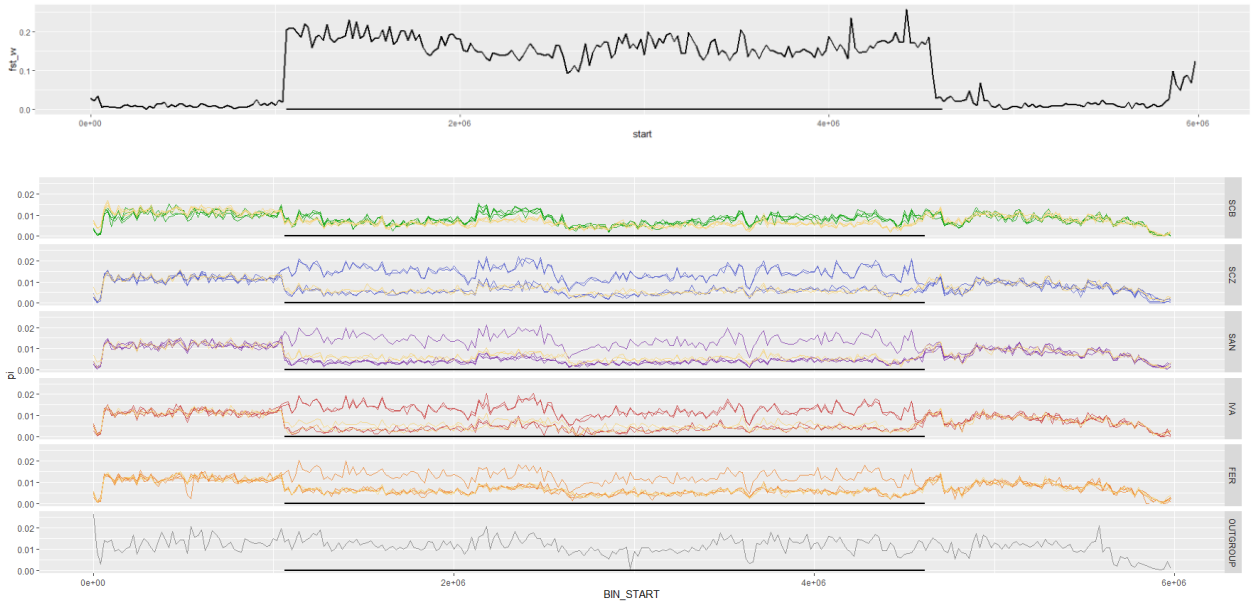
Scaffold158



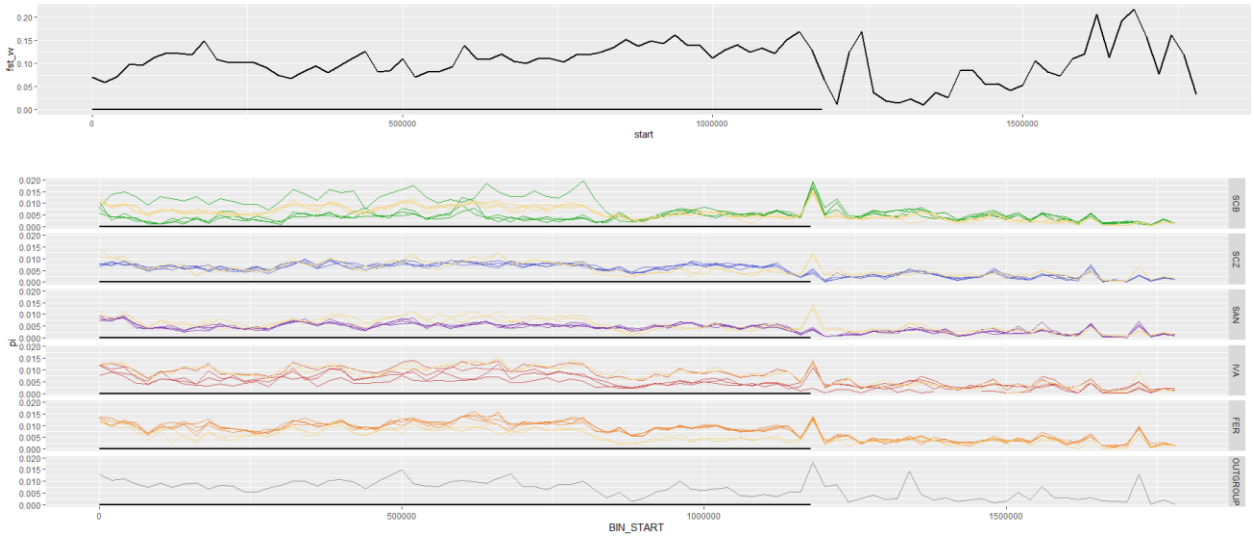
Scaffold167



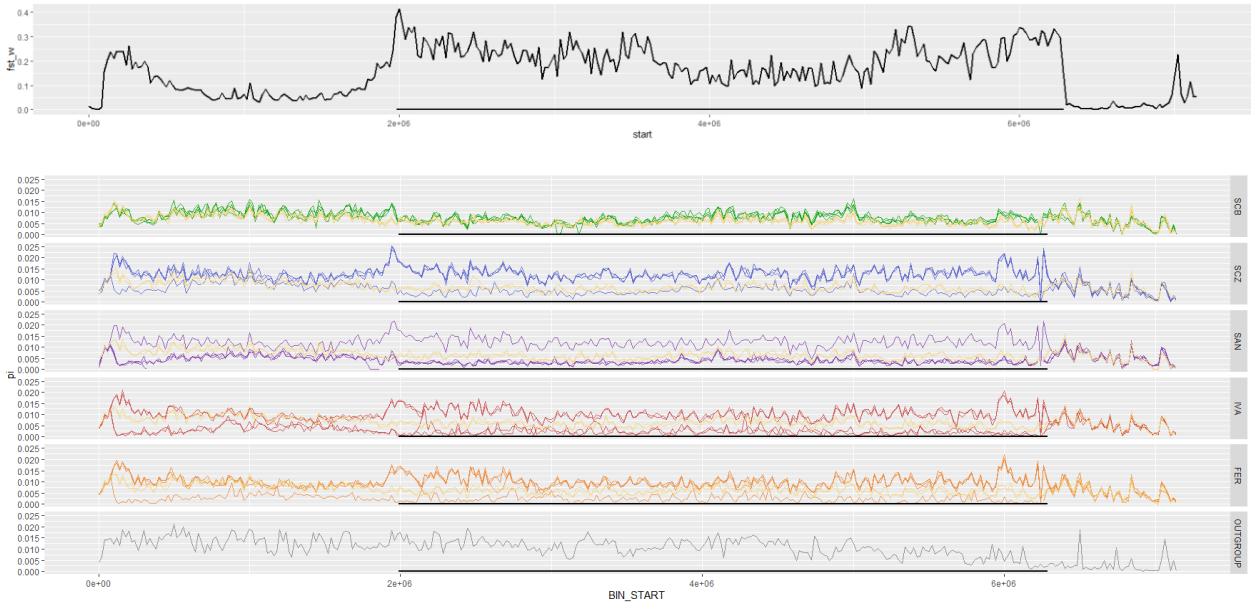
Scaffold169



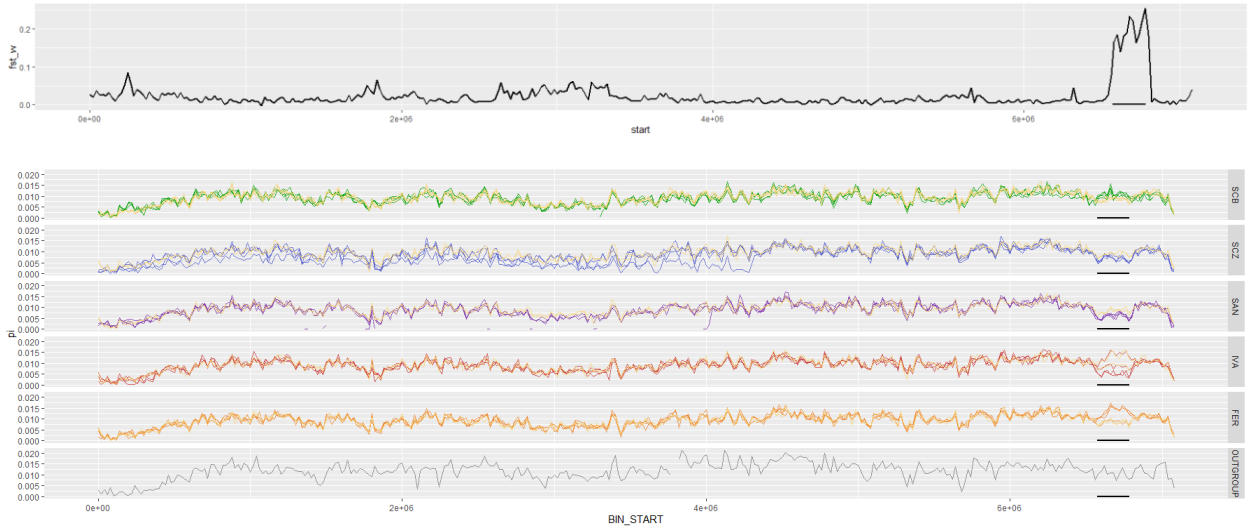
Scaffold240



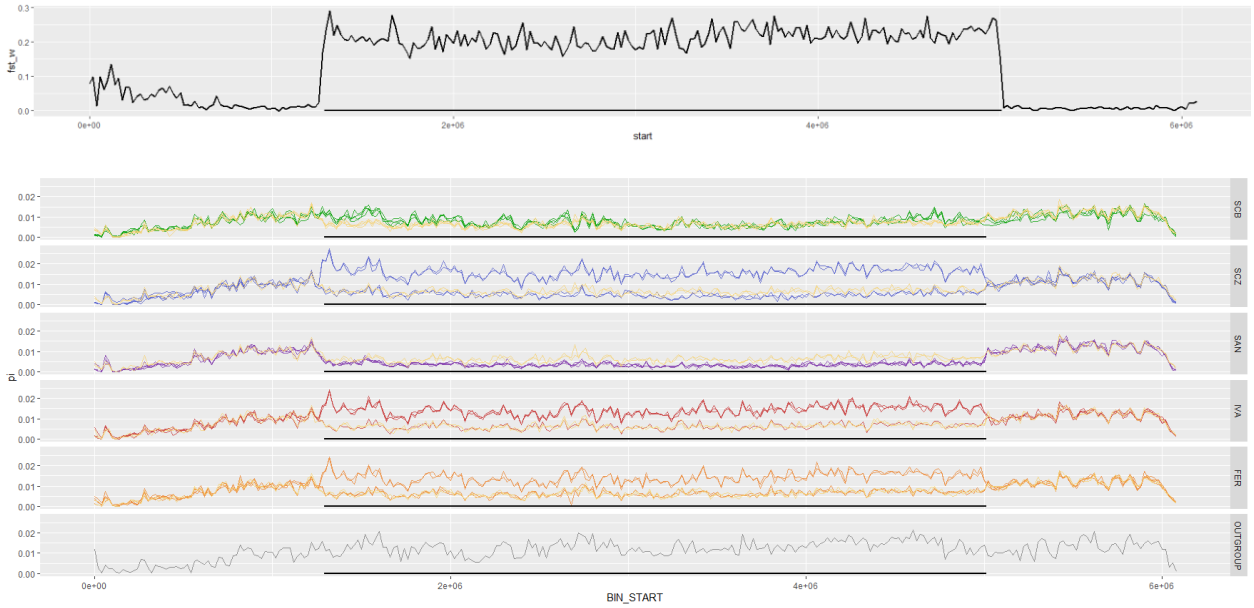
Scaffold48478



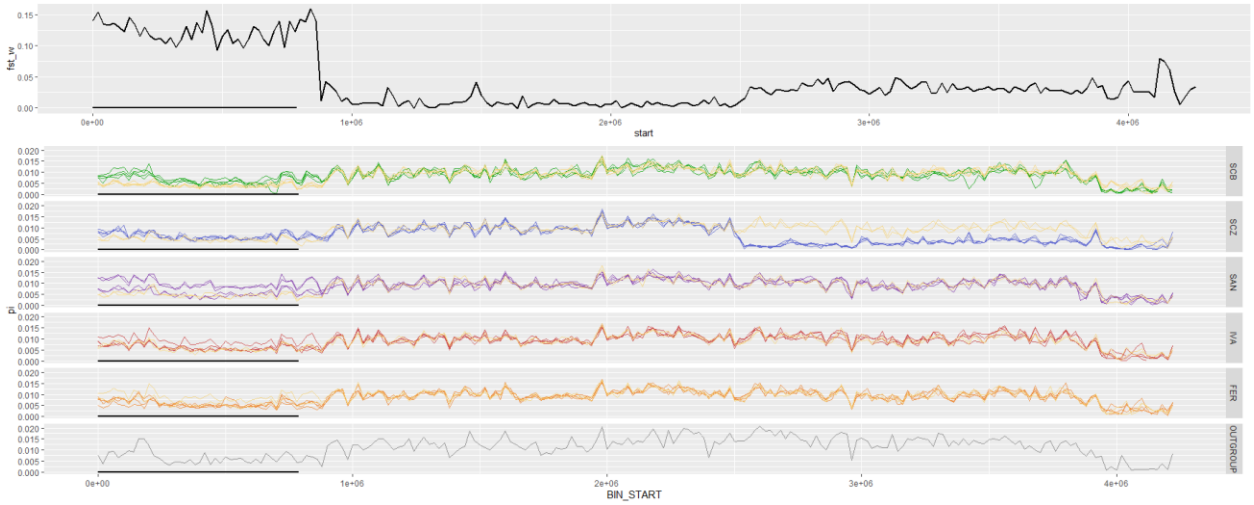
Scaffold52



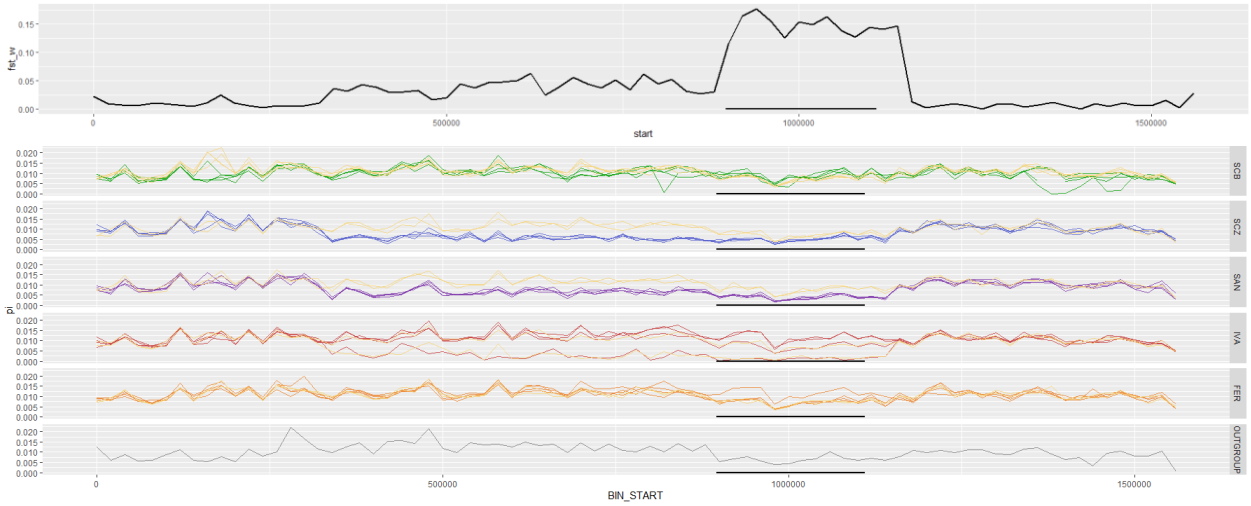
Scaffold57



Scaffold78



Scaffold96


fst_wr
0.15
0.10
0.05
0.00
0
500000
1000000
1500000
start
pi
0.020
0.015
0.010
0.005
0.000
SCB
SCZ
SAN
IVA
FER
OUTGROUP
BIN_START