

# Inversions contribute disproportionately to parallel genomic divergence in dune sunflowers

Received: 19 April 2024

Accepted: 30 October 2024

Published online: 4 December 2024

 Check for updates

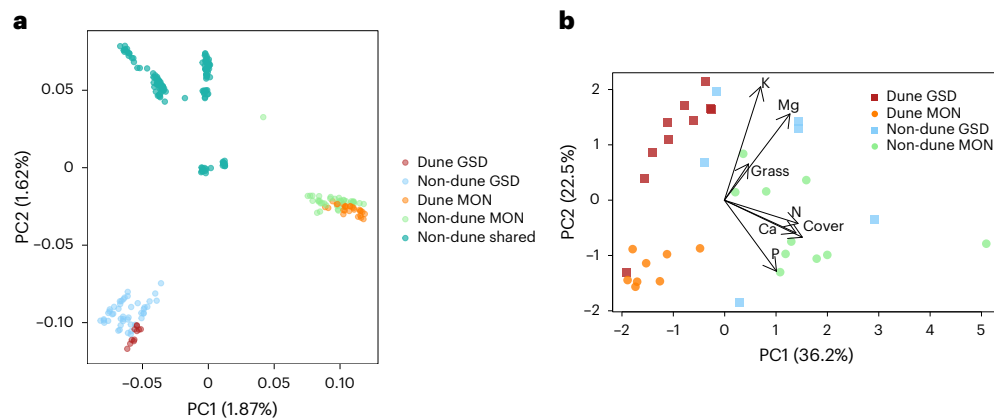
Kaichi Huang <sup>1,2,8</sup>✉, Kate L. Ostevik <sup>2,3,8</sup>✉, Mojtaba Jahani <sup>2</sup>,  
Marco Todesco <sup>2,4,5</sup>, Natalia Bercovich<sup>2</sup>, Rose L. Andrew <sup>6</sup>,  
Gregory L. Owens <sup>7</sup> & Loren H. Rieseberg <sup>2</sup>

The probability of parallel genetic evolution is a function of the strength of selection and constraints imposed by genetic architecture. Inversions capture locally adapted alleles and suppress recombination between them, which limits the range of adaptive responses. In addition, the combined phenotypic effect of alleles within inversions is likely to be greater than that of individual alleles; this should further increase the contributions of inversions to parallel evolution. We tested the hypothesis that inversions contribute disproportionately to parallel genetic evolution in independent dune ecotypes of *Helianthus petiolaris*. We analysed habitat data and identified variables underlying parallel habitat shifts. Genotype–environment association analyses of these variables indicated parallel responses of inversions to shared selective pressures. We also confirmed larger seed size across the dunes and performed quantitative trait locus mapping with multiple crosses. Quantitative trait loci shared between locations fell into inversions more than expected by chance. We used whole-genome sequencing data to identify selective sweeps in the dune ecotypes and found that the majority of shared swept regions were found within inversions. Phylogenetic analyses of shared regions indicated that within inversions, the same allele typically was found in the dune habitat at both sites. These results confirm predictions that inversions drive parallel divergence in the dune ecotypes.

Parallel evolution is the repeated evolution of the same phenotype or genotype in different populations<sup>1–3</sup>. Although it is clear that organisms experiencing similar selective pressures often display parallel phenotypic changes<sup>4–6</sup>, less is known about the extent of parallel evolution at the genomic level. Theoretical studies have shown that the probability of parallel genetic evolution increases with the strength of selection, decreases with the number of possible alleles and is particularly likely to occur for genes with large phenotypic effects<sup>7–9</sup>. Therefore, architectures that limit genetic variability, or the independence of

this variability, will reduce the number of possible adaptive solutions, thereby increasing the repeatability of evolution.

Chromosomal inversions have received much recent attention as important genomic architectures facilitating adaptation and speciation<sup>10–12</sup>. By capturing favourable alleles and suppressing recombination between them, inversions are thought to permit local adaptation in the presence of gene flow. However, because locally adapted alleles are tightly linked to the inversions, the range of adaptive responses is limited when populations encounter a new environment<sup>13</sup>.



**Fig. 1 | PCA of genomic and habitat data for dune and non-dune ecotypes of *H. petiolaris* ssp. *fallax*.** **a**, PCA of 259 samples of *H. petiolaris* ssp. *fallax* using WGS data. The shared non-dune samples are samples from non-dune populations located more than 100 km away from either location. **b**, PCA of the habitat

characteristics for each site. The arrows represent the relative contributions of percent ground cover (Cover), percent ground cover made up of grass (Grass), nitrogen (N), calcium (Ca), phosphorus (P), magnesium (Mg) and potassium (K) for each site measured.

Moreover, the combined phenotypic effect of alleles within inversions is likely to be greater than that of individual alleles<sup>10,14</sup>. Therefore, inversions should increase the likelihood of parallel genomic evolution in different populations. Although inversions have been associated with traits and loci involved in parallel evolution in various taxa<sup>15–17</sup>, whether inversions have contributed disproportionately to parallel genetic evolution has not been formally tested.

We explored the role of inversions in parallel genetic evolution by studying two dune ecotypes of prairie sunflower, *Helianthus petiolaris* ssp. *fallax*, that inhabit active sand dunes in the Great Sand Dunes National Park and Preserve (GSD), Colorado, and around Monahans Sandhills State Park (MON), Texas<sup>18–20</sup>. The dunes at GSD are characterized by lower nitrogen levels and lower vegetation cover than the surrounding sand sheet<sup>18</sup>. The dunes at MON probably share these features but have not been thoroughly surveyed. The dune ecotype in GSD has been shown to have seeds that are three times heavier than the non-dune ecotype at the same location, a trait that contributes importantly to dune adaptation<sup>21</sup>. The plants from the dunes in MON also tend to have large seeds<sup>22</sup>, but the adaptive relevance of this trait has not been systematically investigated at this location. Demographic analyses have suggested recent origins of the dune ecotypes and substantial gene flow between dune and non-dune ecotypes in GSD and MON<sup>18,20</sup>. However, there is no evidence of direct gene flow between the two dune ecotypes. Previous studies have identified large inversions in *H. petiolaris* ssp. *fallax* segregating in one or both of the dune ecotypes (Supplementary Table 1) and found that several of these inversions are associated with seed size, flowering time and soil fertility<sup>22,23</sup>, suggestive of a role of inversions in controlling locally adapted traits. Here, we used diverse comprehensive datasets across the habitats to examine the extent of parallelism and how inversions respond to similar environments and contribute to phenotypic shifts in the two dune ecotypes (Extended Data Fig. 1). We paired newly collected habitat information with newly generated genotyping-by-sequencing (GBS) data across populations in MON and performed genotype–environment association (GEA) analyses to test whether inversions were associated with shared selective pressures in MON and GSD. We then systematically investigated seed size of dune and non-dune populations in MON and conducted quantitative trait locus (QTL) mapping to identify the genetic regions underlying seed size differences in both GSD and MON. Last, we made use of the available whole-genome sequencing (WGS) data for wild sunflowers and applied population genomics approaches to identify selective sweeps (loci under positive selection) shared by the dune ecotypes to test the hypothesis that inversions contribute disproportionately to parallel genetic evolution. The phylogenetic

relationships of the shared regions were also examined to understand the source(s) of genetic variation that contribute to parallel genetic evolution.

## Results and discussion

### Independent origins of the dune ecotypes

To assess patterns of genomic divergence of the dune ecotypes, we used a WGS dataset of 259 *H. petiolaris* from previous studies<sup>22,24</sup>. In a principal component analysis (PCA) based on 387,619 unlinked single-nucleotide polymorphisms (SNPs) in this dataset, dune and non-dune samples from the same location clustered together, implying independent origins of the dune plants (Fig. 1a). This pattern was also supported by a neighbour-joining tree of the samples (Supplementary Fig. 1).

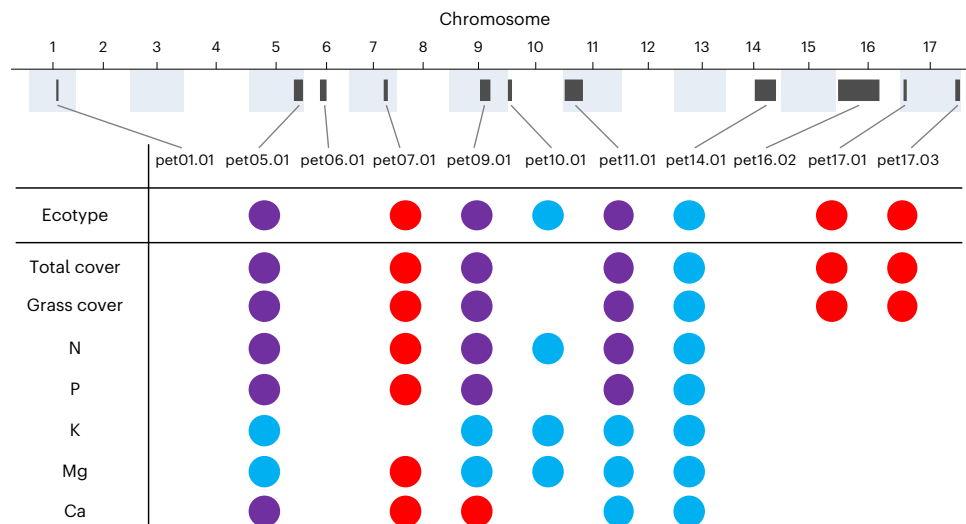
### Parallel habitat shifts in dunes

We quantified environmental parallelism between the dune and non-dune habitats at each location. Contrary to the results based on genetic data, PCA on all seven soil and vegetation-related variables across the two locations revealed that the first axis of environmental variance (PC1) separated dune and non-dune habitats (Fig. 1b). PC2 mostly separated the dune sites from the two locations, but the non-dune sites clustered together along this axis. At both locations, non-dune sites had significantly higher levels of plant cover, total nitrogen, phosphorus and calcium. In addition, the non-dune sites at MON, but not GSD, had significantly higher levels of grass cover, potassium and magnesium (Fig. 1b and Supplementary Table 2).

### Parallelism in GEA

To compare GEA patterns in the dunes at MON with previous results for the samples at GSD<sup>23</sup>, we generated new GBS data for a wide collection of samples (Supplementary Table 3) and performed SNP calling and GEA analyses for eight variables: percentage cover, percentage cover that is grass, percentage cover of sunflowers, nitrogen, phosphorus, potassium, magnesium and calcium<sup>23</sup> (Methods). In the GEA of MON samples, we found multiple large genomic regions with consistently high association, most of which overlapped with the inversions (Supplementary Figs. 2–9).

We compared the GEA results from MON with those in GSD for the same variables. To assess genomic parallelism, we summarized the results of SNPs in non-overlapping windows of 0.025 cM and compared the observed parallel windows against randomly permuted null distributions<sup>25</sup> (Methods). A total of 123 ecotype-associated windows ( $N_{\text{exp}} = 11.8297$ ;  $P < 1 \times 10^{-5}$ ) were shared between the two locations



**Fig. 2 | GEA of the inversions in *H. petiolaris* ssp. *fallax*.** For each variable, coloured circles indicate significant associations of the inversions in GSD (red), MON (blue) or both (purple). Names of inversions are consistent with those used in ref. 22.

(Extended Data Fig. 2). All of the parallel windows were found within inversions pet09.01 and pet11.01. For individual environmental variables, we observed significantly more parallel windows than expected for five environmental variables (plant cover, grass cover, total nitrogen, phosphorus and potassium; Extended Data Fig. 2).

To determine the possible role of inversions themselves in the evolution of the dune ecotypes, we also genotyped all 11 inversions found in *H. petiolaris* ssp. *fallax* using a supervised machine learning algorithm and performed GEA analyses in which putative inversions were treated as single biallelic loci. We used inversion genotypes from 427 WGS samples of *H. petiolaris*<sup>22</sup> as training sets and extracted all shared SNPs between the GBS samples of GSD and MON and the WGS samples within each inversion as predictor variables. This circumvented the issue of low SNP density in the window-based analysis. In total, we obtained 48–377 overlapping SNPs between the GBS and the WGS dataset for each inversion in GSD and 51–631 SNPs in MON. After training, the out-of-bag error rates of the models were generally lower than 5%, except for pet07.01 and pet17.03 in GSD, for which the numbers of overlapping SNPs were small (Supplementary Table 4).

Results from the GEA of inversions were generally correlated well with those for the SNPs inside them (Supplementary Figs. 2–9). When comparing the results for MON and GSD, we found that three inversions (pet05.01, pet09.01 and pet11.01) were associated with dune ecotypes in both GSD and MON (Fig. 2 and Supplementary Fig. 2). However, five inversions were associated with dune ecotypes in only one of the locations: pet07.01, pet17.01 and pet17.03 with dunes at GSD; and pet10.01 and pet14.01 with dunes at MON (Fig. 2 and Supplementary Fig. 2). For individual environmental variables, the three inversions that were associated with dune ecotypes in both locations were also associated with cover, nitrogen and phosphorus in both locations, whereas these inversions were associated with potassium and magnesium only in MON (Fig. 2 and Supplementary Figs. 2–9). These results were consistent with the PCA of environmental variables, in which cover, nitrogen and phosphorus contributed to the parallel habitat shift of the dune and non-dune habitats, whereas potassium and magnesium only varied across habitat types at MON, suggesting parallel responses of chromosomal inversions to shared selective pressures. Calcium, another variable that showed parallel shifts between the dune and non-dune habitats, was associated with pet05.01 at both locations but with pet09.01 only at GSD and pet11.01 only at MON, probably owing to generally weaker associations with this variable, possibly coupled with genotyping error of the inversions. Besides these parallel associated inversions, five

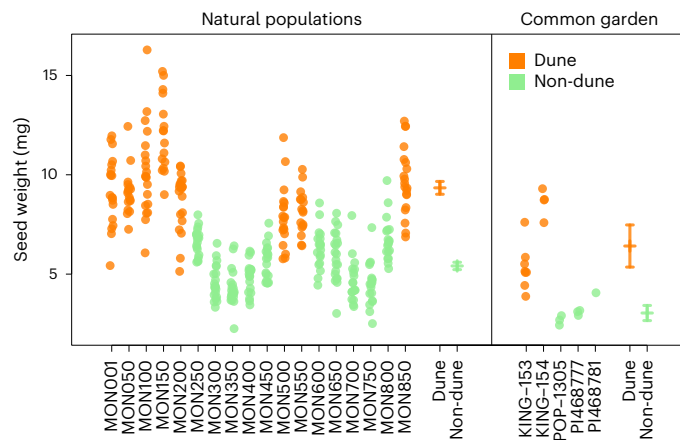
other inversions that were associated with dunes in only one location were also associated with several or all environmental variables at the corresponding location. For instance, in GSD, pet07.01 was associated with cover, nitrogen, phosphorus and potassium and magnesium; and pet17.01 and pet17.03 were consistently correlated with cover variables in GSD. For MON, pet10.01 probably contributes to adaptation to low nitrogen, potassium and magnesium content in the dunes, whereas pet14.01 was strongly associated with all variables tested.

### Parallel genetic changes underlying seed size differences

We examined whether similar shifts in seed weight had taken place in dune ecotypes in MON by investigating plants in natural populations (Supplementary Table 3) and in a common garden. Dune seeds were on average 3.9 mg heavier than non-dune seeds (likelihood ratio (LR) = 30, d.f. = 1,  $P < 0.001$ ) across both experiments (ecotype  $\times$  experiment interaction was not significant; LR = 2.4, d.f. = 1,  $P = 0.12$ ), and seeds collected from natural populations were 2.4 mg heavier than those collected under common greenhouse conditions (Fig. 3; LR = 12, d.f. = 1,  $P < 0.001$ ). The substantial difference in seed size in dune versus non-dune populations of MON was consistent with the seed size difference in GSD<sup>21</sup>, suggesting that natural selection is responsible for larger dune seeds<sup>21,25</sup>.

We made two mapping populations for each location by reciprocally crossing a large-seeded dune individual with a small-seeded non-dune individual (Supplementary Tables 3 and 5). In all cases,  $F_1$  plants produced intermediate seed sizes, and each  $F_2$  mapping population produced a roughly normal distribution spanning most of the variation between the grandparent populations (Extended Data Fig. 3). Cytoplasm type did not have a detectable effect on seed size except for the GSD1  $F_2$ s, for which there was an extremely minor effect (Extended Data Fig. 3 and Supplementary Table 6).

We analysed  $F_2$  samples using a selective genotyping approach<sup>26</sup> and found many genomic regions that were overrepresented in large-seeded plants (Extended Data Fig. 4), consistent with seed size being highly polygenic<sup>27–32</sup>. These regions were more likely to be from the dune parent than the non-dune parent in each mapping population ( $P < 0.001$ ; Extended Data Fig. 4). The strongest signals were generally found within inversions, such as pet11.01 in cross GSD2 and pet14.01 in cross MON2 (Extended Data Fig. 4); these inversions are likely to be large-effect QTL that maintain trait divergence in the face of substantial gene flow<sup>33</sup>. We also inferred the inversion genotypes of the samples using the same algorithm as in GEA analyses. Despite fewer overlapping



**Fig. 3 | The weight of seeds produced by MON plants in natural populations and grown in a common garden.** Each point represents the average seed weight of an individual plant based on at least five seeds. Lines with error bars represent the mean and 95% confidence intervals of the seed weight produced by plants from each habitat type grown under different conditions.

SNPs between the GBS and the WGS data and higher out-of-bag error rates than those in GEA analyses (Supplementary Table 7), we found that pet09.01 and pet11.01 showed significant frequency differences between large-seeded samples and small-seeded samples in MON1, whereas pet07.01, pet10.01 and pet14.01 were detected in MON2. In GSD mapping populations, pet11.01 was the only inversion that showed a significant association (Extended Data Fig. 4).

By summarizing genotypes in windows of 1,000 variants and conducting randomization tests, we found that QTL with significant seed size associations fell into the inversions more than expected by chance in all four mapping populations (Extended Data Figs. 4 and 5), and that more significant QTL windows were shared than expected by chance in pairs within and between locations (Table 1 and Extended Data Figs. 4 and 5). The extent of parallelism within versus outside the inversions varied for each pair of mapping populations from different locations. However, in three of the four pairs, there was significantly more parallelism within versus outside the inversions (Table 1 and Extended Data Fig. 6), suggesting that the inversions (in particular, pet05.01, pet09.01, pet11.01 and pet14.01) disproportionately contribute to parallel shifts in seed size.

Although inversions displayed the strongest signals in our analysis of seed size, the levels of associations for most inversions varied among the crosses. There are multiple explanations for this. First, although several inversions probably contained genes for seed size, selective genotyping may have biased QTL detection towards the largest effect allele that was segregating in each population. A second issue is that the  $F_2$  mapping populations were monomorphic for certain inversions (Extended Data Fig. 4). This was because the crosses were made before we were aware of the existence of segregating inversions. However, even for monomorphic inversions, we sometimes found an association with seed size. For example, in MON1, SNPs within the region of pet05.01 showed significant frequency differences between large-seeded and small-seeded samples, even though the inversion was not segregating in the mapping population. Thus, some of the loci for seed size could still make substantial contributions without the presence of inversions, as has been demonstrated previously in *Helianthus*<sup>34</sup>. Regardless of the cause, we still found more shared QTL windows within the inversions.

### Disproportionate contribution of inversions to parallel adaptive divergence

We identified selective sweeps in each of the dune ecotypes using three approaches that take advantage of different characteristics of genetic polymorphisms. For the composite likelihood ratio (CLR)

statistic computed using SweepFinder2 (ref. 35), we obtained a total of 3,909 outlier regions in GSD and 8,960 in MON. Using the other two approaches that compared dune and non-dune populations, we found 7,159 and 7,742 outlier regions of reduction of diversity (ROD) and 5,907 and 4,569 outlier regions of  $F_{ST}$ , in GSD and MON, respectively. We found 652 and 851 selective sweep regions supported by all three statistics in GSD and MON, averaging 145,339 base pairs (bp) and 46,515 bp, respectively.

After searching for regions that were shared between the two locations, we identified 56 parallel regions across these independently derived ecotypes. Among the parallel regions, 44 regions covering 38.7 Mbp were found to reside within known inversions (Fig. 4), with several large regions covering substantial proportions of inversions pet11.01, pet17.01 and pet17.03, and a number of scattered regions found in inversion pet09.01. This is consistent with previous findings of high sequence divergence associated with inversions in the dune ecotypes<sup>22,23,36</sup>. By comparing the number and length of parallel regions within inversions with the null distribution from permutations, we also found significant enrichment of parallel regions in three inversions that showed parallel divergence between GSD and MON (pet05.01, pet09.01 and pet11.01;  $P < 1 \times 10^{-4}$  for number of regions, and  $P = 0.024$  for total length). The enrichment remained significant for number of regions when testing on all inversions ( $P < 1 \times 10^{-4}$ ), but not for total length ( $P = 0.17$ ). After linkage disequilibrium (LD) pruning with a 95% threshold of the genomic null distribution, we kept 37 regions and still found significant enrichment of parallel regions in pet05.01, pet09.01 and pet11.01 ( $P < 1 \times 10^{-4}$  for number of regions, and  $P = 0.003$  for total length) and across all inversions ( $P < 1 \times 10^{-4}$  for number of regions, and  $P = 0.018$  for total length). This indicates a disproportionate contribution of these three inversions to parallel genomic divergence of GSD and MON.

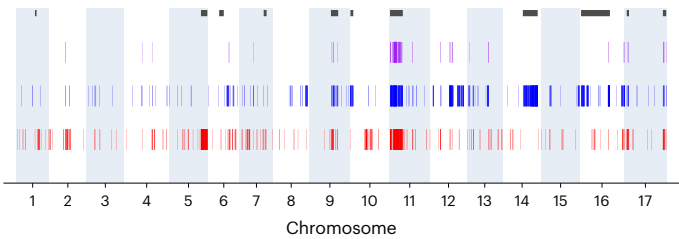
A general caveat for interpreting our results is that because of the block inheritance of inversions, parallel adaptation on any gene within an inversion could leave an inversion-wide signal. This could increase the likelihood of detecting the signature of parallelism, as well as the number and length of parallel regions<sup>37</sup>. Although this may have contributed to the large inversion signal reported here, there were also several large clusters of swept regions outside the inversions seen in each location, such as those found near the middle of chromosomes 2 and 10 in GSD, and in the middle of chromosome 12 and at the distal end of chromosome 16 at MON (Extended Data Fig. 7). This clustering could be viewed as evidence of ‘divergence hitchhiking’, in which new locally adapted mutations establish preferentially in physical linkage to already diverged genes, thereby generating genomic islands of differentiation<sup>38–40</sup>. However, there may be more prosaic reasons for these clusters. The mid-chromosome swept regions in GSD were in areas of low recombination, which probably explains their large size. The clusters of swept regions on chromosomes 12 and 16 were coincident with a translocation between the two chromosomes, which appears to be polymorphic at MON<sup>41</sup>. Significantly, unlike inversions, the low recombination regions on chromosomes 2 and 12 (and elsewhere in the genome) did not show a strong signature of parallelism. Likewise, despite the large number of chromosome translocations differentiating sunflower taxa<sup>41</sup>, translocations do not appear to have contributed to parallel genetic evolution in the dune systems analysed here. Thus, inversions stand out as one of the most influential genetic architectures facilitating parallel genetic evolution.

Despite the disproportionate contribution of inversions to parallel genetic evolution overall, not all showed signals of parallel adaptive divergence. Although inversion pet05.01 was found in both dunes and displayed associations with all variables in MON in GEA analysis, its frequency was relatively lower in the MON dunes (Extended Data Fig. 8); thus, it did not show a detectable signal in the WGS dataset, which had fewer samples from this region. Likewise, pet14.01 increased significantly in frequency in the dunes in one location but only marginally

**Table 1 | The extent of shared windows between seed size QTL in pairs of mapping populations**

	Significance of tests for more shared windows than expected by chance				Presence of shared windows inside specific inversions			
	Overall	Within inversions	Outside inversions	Within versus outside	pet05.01	pet09.01	pet11.01	pet14.01
MON1 versus MON2	<0.001	0.539	0.400	0.376	Yes	Yes	Yes	Yes
GSD1 versus GSD2	<0.001	0.020	0.019	0.064		Yes	Yes	Yes
MON1 versus GSD1	<0.001	0.007	0.290	0.010	Yes	Yes	Yes	Yes
MON1 versus GSD2	<0.001	<0.001	<0.001	<0.001		Yes	Yes	Yes
MON2 versus GSD1	<0.001	0.014	0.178	0.020	Yes	Yes	Yes	Yes
MON2 versus GSD2	0.175	0.521	0.958	0.487			Yes	Yes

Names of inversions are consistent with those used in ref. 22.



**Fig. 4 | Parallel regions in GSD and MON.** Selective sweep regions and overlapping regions (parallel regions) are indicated by bars of different colours: red, GSD; blue, MON; purple, parallel regions. Parallel regions of less than 1 Mbp are represented by thicker bars for better visualization. Black bars show the locations of large non-recombining haplotype blocks (mainly due to inversions) identified in ref. 22 that are polymorphic within *H. petiolaris* ssp. *fallax*.

in the other (Extended Data Fig. 8). The inversion clearly facilitates ecological divergence in MON according to GEA, QTL and selective sweep analyses but was at very low frequency in GSD<sup>22,23</sup>. The lack of parallelism for these inversions might be due to environmental differences between GSD and MON such as relative magnesium and potassium content (Fig. 1b), dune size, dune height, temperature and biotic communities. These differences may generate differences in selective pressures on these inversions between the two sites. Another possibility is that the inversion haplotypes found in GSD and MON no longer share the same adaptive variants, leading to different responses to selection.

**Sources of parallel regions**

Knowing the source(s) of genetic variation giving rise to parallel sweeps is important to interpret patterns of genomic parallelism. We conducted phylogenetic analyses for 12 parallel regions that contain sufficient numbers of SNPs for reliable phylogenetic reconstruction (6 within and 6 outside of inversions). For most regions outside of inversions, the dune samples from each site were closest to those from local non-dune populations, and the alleles found in the dunes likely originated from intraspecific genetic variation (Fig. 5). This suggests that the alleles found in the dunes were independently drawn from the non-dune populations in each of the dune ecotypes. In some regions outside of inversions, such as a region found on chromosome 7, the dune samples at MON showed deep divergence with other samples from the same species, suggesting the dune allele may come from introgression from another species, while the one at GSD arose from within *H. petiolaris* (Fig. 5).

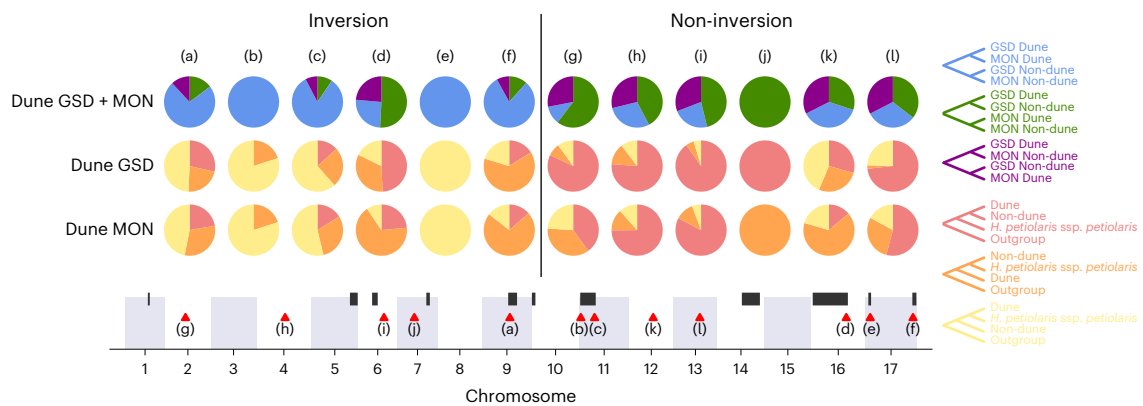
For all regions within inversions (except pet16.02), we consistently found strong support for parallel evolution of the same dune haplotypes in GSD and MON (Fig. 5). In the parallel regions, the dune haplotypes of the inversions were often found to group with *H. petiolaris* ssp. *petiolaris*, suggesting the existence of this adapted dune haplotype in the other subspecies. The results of the region in inversion pet17.03 indicated a more ancient origin of the dune haplotypes (Fig. 5).

This was consistent with previous discoveries that the dune haplotypes of these inversions probably originated via introgression with an unknown and possibly extinct donor species<sup>22</sup>. Regardless of the donor, the same haplotypes of these inversions were targeted by selection at the two sites, indicating parallel evolution of the inversions from standing genetic variation. The dune-adapted inversions were probably introduced to the species before the establishment of the GSD and MON populations, segregating at low frequency in the non-dune habitat<sup>22</sup> and raised to high frequency in parallel by natural selection in the dunes. Selected alleles at the same loci in different populations could have multiple independent mutational origins<sup>42–44</sup> or could arise by selection on the standing variation present in the ancestral populations<sup>45,46</sup>. In contrast to independent de novo mutation, adaptation from standing genetic variation is likely to be rapid<sup>47</sup>. The retention of inversion polymorphisms within *H. petiolaris* might have provided a directly available pool of adapted alleles for the dune environment, which promoted the reuse of these inversions in each site. As these inversions may represent ancient introgressions<sup>22</sup>, these alleles might have been pretested by selection in past environments, further increasing the probability of parallel evolution.

**Conclusions and future directions**

Frequent reuse of inversions during the repeated colonization of dune habitats suggests strong biases and constraints in adaptive evolution. Through comprehensive sampling across habitats and genetic mapping, our GEA and QTL analyses demonstrated that these inversions contain locally adapted alleles of multiple ecologically relevant traits. By maintaining genotypic combinations at loci affecting adaptation, inversions limit the genetic options for local adaptation, thereby increasing the potential for parallel genetic evolution<sup>13</sup>. Furthermore, the combined phenotypic effect of alleles in an inversion is likely to be larger than that of individual alleles, increasing the likelihood of inversions being established by natural selection and maintained under gene flow selection balance<sup>9,14,48</sup>. In our GEA and QTL analyses, the strongest associations were usually found at SNPs within inversions or with the inversions themselves, indicating that these inversions are major contributors to adaptation to low vegetation cover and lack of nutrients in the dunes, consistent with theoretical predictions of parallel genetic evolution. Overall, we show that inversions are more strongly involved than the rest of the genome in GEAs, the genetic basis of a locally adapted phenotype and selective sweeps in the dune ecotypes. Together, these results show that inversions are disproportionately responsible for the parallel adaptation of sunflowers to two dune habitats.

Notably, our phylogenetic analyses showed that although shared sweeps within inversions were mostly parallel in the strictest sense (that is, swept haplotypes within the inversions derived from the same ancestral source), this was generally not the case for shared sweeps outside of the inversions, which mostly derived from different variants found in the local non-dune population (Fig. 5). This suggests that we



**Fig. 5 | Topological weighting of parallel regions.** a–l, Ha412HOC hr09:107701884–116410572 (a), Ha412HOCChr11:5890343–5916951 (b), Ha412HOCChr11:59695130–62373169 (c), Ha412HOCChr16:140217167–142747183 (d), Ha412HOCChr17:19185748–19215393 (e), Ha412HOCChr17:190415052–191878407 (f), Ha412HOCChr02:82397894–82927898 (g), Ha412HOCChr04:91990564–139990182 (h), Ha412HOCChr06:101167874–106607943 (i), Ha412HOCChr07:68115435–68247437 (j), Ha412HOCChr12:96918451–97184453 (k), Ha412HOCChr13:105437209–105575210 (l). Regions are grouped by their

positions relative to inversions. The tested scenarios regarding the source(s) of dune-adapted variation are displayed on the right: habitat, indicative of parallel genotypic evolution (blue); geography, indicative of non-parallel genotypic evolution (green); other (purple); within-species diversity (red); introgression/ancient variation (orange); subspecies introgression (yellow). The genomic positions of the parallel regions (red triangles) and the inversions (black bars) are shown at the bottom.

might be underestimating the extent of parallelism within inversions compared with other regions of the genome.

Work in other systems has also found evidence that inversions play a role in parallel adaptation (for example, refs. 49–51). However, this is not a universal pattern<sup>52,53</sup>. For systems where parallel reuse of inversions in repeated adaptation has been observed, analyses similar to those that we conducted here would offer stronger support for the importance of inversions in adaptive evolution. When compared across systems, inversions appear to be particularly abundant and relevant to local adaptation in systems such as sunflowers, *Littorina* snails<sup>54</sup>, mice<sup>55</sup> and monkeyflowers<sup>56</sup>, in which there are sharp ecological transitions and ongoing gene flow, consistent with theoretical prediction<sup>10</sup>.

Our study represents a first step towards understanding the evolutionary and ecological factors that contribute to parallel genetic adaptation. We recommend that future studies estimate the strength of selection on parallel swept regions in the two dune ecotypes using a combination of modelling<sup>9,57</sup> and empirical approaches (for example, ref. 25). Studies that dissect the genetic architectures of other ecologically relevant traits<sup>58,59</sup>, including identification of the causal genes and mutations, would illuminate both the cause(s) of parallel evolution of adaptive traits and the particular evolutionary and demographic histories of the genetic variation underlying parallel adaptation<sup>45</sup>. Lastly, dune adaptation has occurred many times in *Helianthus*, involving half a dozen different species or subspecies<sup>60</sup>. Extending the approach taken in the present study to these other systems would be likely to lead to further insights into the factors that facilitate or constrain parallel genetic evolution.

## Methods

### Habitat analysis

Habitat characteristics for GSD populations were from ref. 18. In 2015, we visited ten non-dune and six dune populations within Monahans Sandhills State Park, Texas, and two dune populations outside the park (Supplementary Table 3). We established a transect through each population and picked five sites for habitat analysis at even intervals along those transects. At each site, we took a photograph of a 0.65 m<sup>2</sup> quadrat and used ImageJ<sup>61</sup> to determine the proportions of vegetative cover and grass cover. Soil samples were taken from 25 cm deep, dried at 60 °C for 24 h and pooled from the five sites at each population. Available phosphorus and exchangeable potassium, magnesium and calcium were measured at A&L Eastern Laboratories and total nitrogen

content was determined by Micro-Dumas Combustion (NA1500, Carlo Erba Strumentazione) at the University of Georgia Analytical Chemistry Laboratory. For each habitat trait at each location, we fit linear models in R v.4.0.3 (ref. 62) to determine whether ecotype was a significant explanatory variable. To see the effects of all variables at once, we imputed missing data (a maximum of three individuals had missing data for any variable) using R package missMDA<sup>63</sup> and conducted PCA on scaled variables using the prcomp function.

### GEA analyses

We collected mature seeds from 20 individuals at each MON population mentioned above and germinated seeds from 10–12 individuals per population, making a total of 195 samples for GEA analysis (Supplementary Table 3). We extracted DNA from these samples using DNeasy Plant kits (Qiagen) and prepared a GBS library following a modified version of the ref. 64 protocol<sup>19</sup>. The library was paired-end sequenced on an Illumina HiSeq. We called variants against the HA412-HOV2 *Helianthus annuus* from ref. 65 and filtered the SNPs for downstream analyses using the parameters reported in ref. 23, except that individual genotypes with depth less than 5 were set as missing owing to the lower depth of this library. Samples with excessive missing data (>50%) were removed from the dataset, leaving 174 samples for downstream analyses.

The GEA analysis was conducted using BayPass v.2.1 (ref. 66). Population structure was estimated by choosing 1,000 random SNPs in low LD ( $R^2 < 0.1$ ) and running BayPass in core model mode. The covariance matrix from this analysis was used as a control for population structure. We then ran BayPass under the standard covariate model using the default importance sampling estimator approach. Each environmental variable was scaled using the -scalecov option. Owing to missing data for total nitrogen in population MON001, the analysis was run separately for total nitrogen and other variables. We also simulated pseudo-observed data (POD) with 1,000 SNPs using the 'simulate.baypass' function, and analysed the newly created POD for each environmental variable as described above. SNPs that had a BF above the top 1% quantile of those of the POD were treated as associated SNPs. To identify genomic parallel divergence between dune and non-dune ecotypes in GSD and MON, we also coded the populations according to their ecotypes and ran BayPass by treating the ecotypes as covariates. No scaling was performed for the binary variable.

We then compared the GEA results of MON and GSD for shared variables. To test for parallelism between locations, we divided the

chromosomes into non-overlapping windows of 0.025 cM and calculated the average BF value for each window by summarizing the SNPs within a range of 0.5 cM. Windows with a BF above the top 1% quantile of the corresponding POD were treated as associated windows. As a null distribution, we randomly drew  $N$  0.025 cM windows from each location, where  $N$  is the count of associated windows for each location, and then assessed the overlap of randomly selected associated windows between the locations over 10,000 iterations<sup>26</sup>.

### Seed size analysis

To investigate whether similar shifts in seed size had taken place in the dune ecotype in MON, we weighed groups of five seeds collected from 16–21 individuals from the same populations used in the GEA analyses above (Supplementary Table 3). To explore phenotypic plasticity, we grew 1–4 plants from each of two dune and three non-dune populations that had been collected previously (Supplementary Table 3) under standard greenhouse conditions and weighed 6–20 individual seeds produced by these plants. We fit a linear mixed effect model using R package nlme<sup>67</sup> to describe mean individual seed weight across both datasets. The model included ecotype, experiment (wild collected versus common garden), and their interaction as fixed effects and population as a random effect. We identified significant model terms by comparing nested models using LR tests.

### QTL mapping

We made two mapping populations for each location by reciprocally crossing a large-seeded dune individual with a small-seeded non-dune individual (Supplementary Table 3). As sunflowers are self-incompatible, we grew up several  $F_1$  seeds produced by each of these crosses and crossed pairs of reciprocal  $F_1$ s. This resulted in  $F_2$  seed lots with dune and non-dune cytoplasms for each of the four cross types (Supplementary Table 5). We grew up 186–325  $F_2$  plants of each cross-by-cytoplasm type under our standard greenhouse conditions (Supplementary Table 5). In addition, we grew a few plants from wild collected seeds (half siblings of the grandparent plants) and  $F_1$  seeds (full siblings of the parent plants) used to make the  $F_2$  seed lots.  $F_2$ s were mainly open pollinated, and groups of five seeds per plant were weighed.

Using a selective genotyping approach<sup>26</sup>, we genotyped the 12 plants that produced the heaviest seeds and the 12 plants that produced the lightest seeds from each of the eight cross-by-cytoplasm types, for a total of 192 plants. Genotyping was done using the methods described above, and biallelic variants were filtered for heterozygosity < 95%, minor allele frequency > 15% and coverage > 75%. We also used the parents of each mapping population to determine which allele came from the large-seeded dune parent. For each mapping population, we performed Fisher's exact tests at each locus to determine whether there was a higher proportion of dune alleles in large-seeded samples than in small-seeded samples. We used Bonferroni corrections based on the total number of variants and the number of independent regions in the dataset as estimated by PCA<sup>68</sup> to define significance thresholds.

We tested for parallelism in seed size QTL by determining the most frequent genotype (homozygous dune, homozygous non-dune or heterozygous) in windows of 1,000 variants for each mapping population and performing Fisher's exact tests on these windows as above. This yielded 1,548 genotyped windows in all four mapping populations. We randomized the results of these tests 10,000 times and counted the number of windows significantly associated with larger seeds ( $P < 0.05$ ) in pairs of mapping populations and in inversions. We did a second set of randomizations across the windows inside and outside inversions separately and counted the numbers of significant windows shared across pairs of mapping populations. This allowed us to determine the extent of parallelism inside and outside the inversions and whether there was a significant difference between those rates.

### Genotyping and parallelism analyses of inversions

We also performed GEA and QTL analyses in which inversions were treated as single biallelic loci. For each of the GBS samples in the aforementioned GEA and QTL analyses, we determined their genotypes at all 11 inversions found in *H. petiolaris* ssp. *fallax* (Supplementary Table 1) using a supervised machine learning algorithm, with SNPs and inversion genotypes from 427 WGS samples of *H. petiolaris*<sup>22</sup> as training sets. We extracted shared SNPs between the GBS and WGS samples within each inversion as predictor variables. The inversion genotypes of the WGS samples were marked as categorical variables and used as the response variable. A random forest model was constructed for each inversion using the function 'randomForest' in R package randomForest<sup>69</sup>. The number of decision trees was set to 151, and the number of randomly sampled variables for each tree was determined by traversing different numbers of SNPs and choosing the value with the lowest prediction error rate in the training set. The trained model was subsequently used to infer inversion genotypes in the GBS samples.

We performed GEA analyses for the inversions using the same procedure for SNPs as described above. We also reran inversion GEA analyses for the GSD samples using the new selection of inversions and inversion genotypes determined by the random forest method. Inversions undergoing parallel divergence between dune and non-dune ecotypes in GSD and MON were identified using the procedures mentioned above. Inversions that had a Bayes factor (BF) above the top 1% quantile of those of the POD were treated as significantly associated. To determine associations between the inversions with seed size, we treated them as single biallelic loci and conducted Fisher's exact test analysis using the procedure described above for SNPs.

### WGS and SNP calling

To assess genome-wide patterns of adaptive divergence, we used WGS samples and an SNP dataset for *H. petiolaris* from previous studies<sup>22,24</sup>. The previous intraspecific SNP set contained 7,107,746 biallelic SNPs across 259 WGS samples of *H. petiolaris* ssp. *fallax*, including 11 dune ecotype samples from GSD, 27 dune ecotype samples from MON, and 221 non-dune samples from the sand sheet below the dunes and from other populations of *H. petiolaris* ssp. *fallax* (Supplementary Table 8).

For analyses that required polarization of the SNPs or outgroups, we generated another variant dataset for the species using all available samples for *H. petiolaris*, as well as samples from four outgroup perennial sunflowers (664647\_GIG from *Helianthus giganteus*, DEC\_1895 from *Helianthus decapetalus*, DIV\_1956 from *Helianthus divaricatus*, and GRO\_2043 from *Helianthus grosseserratus*; ref. 22). Sequences of these samples were aligned to a reference genome of *H. annuus* (Ha412-HOV2; ref. 65) using NextGenMap v.0.5.3 (ref. 70). For each sample, a GVCF file was produced with 'HaplotypeCaller' in GATK v.4.1.4.1 (ref. 71) with parameter --heterozygosity 0.01. After individual variant calling, all samples of *H. petiolaris* and the four perennial samples were jointly genotyped using GATK's 'GenomicsDBImport' and 'GenotypeGVCFs' on 1 Mbp genomic regions without transposable elements, and the raw VCF chunks were then gathered by chromosome using 'GatherVcfs'. Variant quality score recalibration was performed using the top 20 samples with the highest sequencing depth in *H. petiolaris* as a 'gold set'. The raw variants were filtered to remove sites with extremely high heterozygosity (ExcessHet z-score < -4.5), and the gold set was then applied against this filtered set of variants to produce recalibration models for SNPs and indels using 'VariantRecalibrator'. The 90% tranche for each cohort was selected based on these recalibration models using 'ApplyVQSQR'. After variant quality filtration, the variants were further filtered to retain only biallelic SNPs with minor allele frequency > 0.01 and genotyping rate > 50% across all samples. This generated a total of 21,836,165 biallelic SNPs. By comparison with the outgroup species, we obtained information on ancestral states for about 58% of all SNPs. Unpolarized sites that were monomorphic within the target

group were discarded, leaving a total of 5,996,827 sites in GSD and 9,405,793 in MON for input into SweepFinder2.

### Population structure and phylogenetic tree

To investigate sample relationships, we conducted PCA and phylogenetic analysis using the SNPs from WGS of all dune and non-dune plants. The SNPs were pruned with a LD threshold of 0.2, and PCA then conducted using the R package SNPrelate<sup>72</sup>. A neighbour-joining tree was constructed using MEGA 10.1.6 (ref. 73) with the maximum composite likelihood method using the same SNP dataset. The neighbour-joining tree was visualized with the R package ggtree<sup>74</sup>.

### Identification of selective sweeps

Selective sweeps in each of the dune ecotypes were identified using three approaches that take advantage of different characteristics of genetic polymorphisms following selective sweeps.

First, we employed the CLR statistic, which contrasts the likelihood of the null hypothesis based on the genome-wide site frequency spectrum with the likelihood of a model where the site frequency has been altered by a recent selective sweep. The CLR statistic was computed using SweepFinder2 (ref. 35), which includes invariant (fixed) sites and accounts for the effects of negative selection on diversity<sup>75</sup>.

In the analyses of each dune ecotype, we excluded samples from within 100 km of the other dune ecotype to remove their impact on sweep detection. SNPs of the dune samples (target group) were extracted and polarized by comparison with the perennial samples. For each SNP, the outgroup allele was inferred as ancestral state if all four perennials were called and homozygous and the allele matched one of the alleles of the target group; otherwise the site was defined as unpolarized. Unpolarized sites that were monomorphic within the target group were discarded. All unpolarized sites that were polymorphic within the target group were kept and assigned folded = 1 in the input files for SweepFinder2. We generated the allele frequency input files for SweepFinder2 using a custom script, and the recombination rate of each position was calculated based on an integrated genetic map for cultivated sunflower<sup>22</sup>. Empirically derived allele frequency spectra across all chromosomes were calculated using the -f option of SweepFinder2. CLR values were then calculated using the recombination map and precomputed empirical spectrum with a grid size of 2 kb. No correction for background selection was conducted because genome-wide estimation of background selection has not yet been conducted in sunflowers.

As SweepFinder2 only examines sequence polymorphism within populations, some outlier regions that it identifies might come from broader selective sweeps (that is, regional or species-wide sweeps). Therefore, the other two approaches we employed to detect sweeps permitted comparisons between dune and non-dune populations to filter for those restricted to the dune populations.

A well-known characteristic of hard sweeps is a reduction in sequence diversity in the selected region and at linked neutral sites<sup>76,77</sup>. Thus, for each site, sequence diversity ( $\pi$ ) was calculated for the dune and non-dune groups separately using VCFtools v.0.1.14 (ref. 78) with 20 kb sliding windows and a step size of 2 kb. The ROD<sup>79,80</sup> in the dunes was then estimated by summarizing the ratio of  $\pi$  for the dune group and non-dune group ( $\pi_{\text{dune}}/\pi_{\text{non-dune}}$ ). Another common feature of selective sweeps is an increase in genetic differentiation in the selected region and closely linked SNPs<sup>81,82</sup>. Therefore, we used VCFtools to calculate Weir and Cockerham's  $F_{ST}$  (ref. 83) using 20 kb sliding windows and a step size of 2 kb to estimate genetic divergence between the dune and non-dune groups.

For each of the three statistics mentioned above, the top 5% of genomic windows were classified as outliers, and groups of adjacent outlier windows were merged into regions. Finally, outlier regions that were identified by SweepFinder2 and overlapped with regions identified by both ROD and  $F_{ST}$  methods were considered selected regions.

### Regions of parallel evolution and enrichment analyses

Genomic regions with outlier values in population genomics calculations may result from the same selective sweep owing to genetic linkage. To account for clustering of regions with signatures of selective sweep, we extracted the SNPs in each selected region and calculated the pairwise LD between each pair of SNPs between windows using PLINK v.1.9 (refs. 84,85). The LD between regions was summarized with the average value of  $R^2$  between SNPs from each window. We also generated genome-wide null distributions of LD for the same physical distance. Specifically, we randomly selected 10,000 pairs of 5 kb windows with the same physical distance, and the LD between these windows was summarized using the same approach. Adjacent outlier regions were merged if the LD between windows exceeded 95% of the null distribution ( $P < 0.05$ ) of the same physical distance.

After merging based on LD, overlaps of the selective sweep regions in the two dune ecotypes were identified using BEDTools<sup>86</sup>, and these overlapping regions were treated as regions involved in parallel evolution (hereafter 'parallel regions'). After identification of the parallel regions, we merged the regions again based on the LD between them using the method mentioned above and the genetic map distance between them. Clusters of regions were merged into a single larger region if the LD between closest regions was greater than 95% of the null distribution and the map distance was less than 5 cM. We also pruned for LD based on the same criteria and kept only one parallel region with the greatest size for each LD cluster.

To test whether inversions showed a stronger pattern of parallel evolution than non-inverted regions, we used a randomization approach to find the numbers and sizes of regions overlapping between parallel regions and chromosomal inversions and to estimate the significance of this overlap. Specifically, we held the position of the parallel regions constant and randomized the position of the inversions across the genome 10,000 times. For each permutation, we summarized the numbers and lengths of the regions that overlapped with inversions and built null distributions. Significance of enrichment was estimated using a one-tailed test by comparing the observed values with these null distributions. Using this statistical test, we asked whether the observed overlap between parallel sweeps and inversions was greater than would be expected by chance. The test was conducted on the parallel regions before and after merging based on LD and genetic map distance.

### Phylogeny of parallel regions

To estimate the source(s) of genetic variation giving rise to parallel sweeps, we conducted phylogenetic analyses on 12 merged parallel regions containing more than 100 SNPs. For each region, we chose five samples from each dune population, five from non-dune populations in proximity to each dune site, respectively, five samples from *H. petiolaris* ssp. *petiolaris* and two perennial samples as the outgroup. We imputed and phased the SNPs using Beagle<sup>87</sup> and divided the genome into 20 kb windows. For each genomic window, we randomly chose one haplotype for each individual and calculated a maximum likelihood phylogeny using IQ-TREE<sup>88</sup>. We then used Twisst<sup>89</sup>, which takes into account phylogenetic position variation among individuals within a species and calculates topological weighting for each possible gene tree, to quantitatively determine the source of dune-adapted variation. We tested different scenarios in which dune haplotypes are shared or from adjacent non-dune populations (Fig. 5) and also explored where dune haplotypes in GSD and MON came from within-species diversity, *H. petiolaris* sp. *petiolaris* or a more distant source (Fig. 5). We extracted all 20 kb windows that had >50% overlap with each parallel region and normalized weights such that each parallel region had a total weight of 1 and counted the total weight for each topology.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Environmental data of GSD have been published previously (<https://doi.org/10.5061/dryad.158pb518>, ref. 90); those for MON generated herein have been deposited in Dryad (<https://doi.org/10.5061/dryad.bcc2fqznn>, ref. 91). Sequencing data of GSD samples have been published previously (<https://doi.org/10.5061/dryad.j2448>, ref. 92); those for MON samples generated herein have been deposited in the NCBI Sequence Read Archive (SRA) under accession [PRJNA1145483](https://doi.org/10.5061/dryad.bcc2fqznn). Seed weight data and sequencing data of F<sub>2</sub> plants in QTL mapping experiments have been deposited in Dryad (<https://doi.org/10.5061/dryad.bcc2fqznn>, ref. 91) and in the NCBI SRA under accession [PRJNA1145296](https://doi.org/10.5061/dryad.bcc2fqznn), respectively. WGS data for *H. petiolaris* have been published previously and are stored in the NCBI SRA under accessions [PRJNA322345](https://doi.org/10.5061/dryad.bcc2fqznn) and [PRJNA532579](https://doi.org/10.5061/dryad.bcc2fqznn).

## Code availability

The codes used for all analyses have been archived through GitHub ([https://github.com/hkchi/Dune\\_parallelism](https://github.com/hkchi/Dune_parallelism)) and are available via Zenodo at <https://doi.org/10.5281/zenodo.13983266> (ref. 93).

## References

- Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822 (2004).
- Arendt, J. & Reznick, D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008).
- James, M. E., Brodribb, T., Wright, I. J., Rieseberg, L. H. & Ortiz-Barrientos, D. Replicated evolution in plants. *Annu. Rev. Plant Biol.* **74**, 697–725 (2023).
- Jones, R., Culver, D. C. & Kane, T. C. Are parallel morphologies of cave organisms the result of similar selection pressures? *Evolution* **46**, 353–365 (1992).
- Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308 (2000).
- Rajakaruna, N., Siddiqi, M. Y., Whitton, J., Bohm, B. A. & Glass, A. D. Differential responses to Na<sup>+</sup>/K<sup>+</sup> and Ca<sup>2+</sup>/Mg<sup>2+</sup> in two edaphic races of the *Lasthenia californica* (Asteraceae) complex: a case for parallel evolution of physiological traits. *New Phytol.* **157**, 93–103 (2003).
- Orr, H. A. The probability of parallel evolution. *Evolution* **59**, 216–220 (2005).
- Bailey, S. F., Blanquart, F., Bataillon, T. & Kassen, R. What drives parallel evolution?: how population size and mutational variation contribute to repeated evolution. *BioEssays* **39**, 1–9 (2017).
- MacPherson, A. & Nuismer, S. L. The probability of parallel genetic evolution from standing genetic variation. *J. Evol. Biol.* **30**, 326–337 (2017).
- Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
- Hoffmann, A. A. & Rieseberg, L. H. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**, 21–42 (2008).
- Feder, J. L. & Nosil, P. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63**, 3061–3075 (2009).
- Roesti, M., Gilbert, K. J. & Samuk, K. Chromosomal inversions can limit adaptation to new environments. *Mol. Ecol.* **31**, 4435–4439 (2022).
- Westram, A. M., Faria, R., Johannesson, K., Butlin, R. & Barton, N. Inversions and parallel evolution. *Phil. Trans. R. Soc. B* **377**, 20210203 (2022).
- Arostegui, M. C., Quinn, T. P., Seeb, L. W., Seeb, J. E. & McKinney, G. J. Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. *Mol. Ecol.* **28**, 1412–1427 (2019).
- Fang, B., Kemppainen, P., Momigliano, P., Feng, X. & Merilä, J. On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).
- Balanyà, J. et al. Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. *Evolution* **57**, 1837–1845 (2003).
- Andrew, R. L., Ostevik, K. L., Ebert, D. P. & Rieseberg, L. H. Adaptation with gene flow across the landscape in a dune sunflower. *Mol. Ecol.* **21**, 2078–2091 (2012).
- Ostevik, K. L. *The Ecology and Genetics of Adaptation and Speciation in Dune Sunflowers*. PhD thesis, Univ. British Columbia (2016).
- Raduski, A. R., Rieseberg, L. H. & Strasburg, J. L. Effective population size, gene flow, and species status in a narrow endemic sunflower, *Helianthus neglectus*, compared to its widespread sister species, *H. petiolaris*. *Int. J. Mol. Sci.* **11**, 492–506 (2010).
- Ostevik, K. L., Andrew, R. L., Otto, S. P. & Rieseberg, L. H. Multiple reproductive barriers separate recently diverged sunflower ecotypes. *Evolution* **70**, 2322–2335 (2016).
- Todesco, M. et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
- 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).
- Huang, K. et al. Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. *Mol. Biol. Evol.* **39**, msac101 (2022).
- Goebel, A. M., Kane, N. C., Doak, D. F., Rieseberg, L. H. & Ostevik, K. L. Adaptation to distinct habitats is maintained by contrasting selection at different life stages in sunflower ecotypes. *Mol. Ecol.* **33**, e16785 (2024).
- Lander, E. S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
- Tang, S., Leon, A., Bridges, W. C. & Knapp, S. J. Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. *Crop Sci.* **46**, 721–734 (2006).
- Baack, E. J., Sapir, Y., Chapman, M. A., Burke, J. M. & Rieseberg, L. H. Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Mol. Ecol.* **17**, 666–677 (2007).
- Chapman, M. A. et al. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* **20**, 2931–2945 (2008).
- Bert, P.-F. et al. Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.). 2. Characterisation of QTL involved in developmental and agronomic traits. *Theor. Appl. Genet.* **107**, 181–189 (2003).
- Wills, D. M. & Burke, J. M. Quantitative trait locus analysis of the early domestication of sunflower. *Genetics* **176**, 2589–2599 (2007).
- Mokrani, L. et al. Mapping and analysis of quantitative trait loci for grain oil content and agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **106**, 149–156 (2002).
- Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration-selection balance: the genetic architecture of local adaptation. *Evolution* **65**, 1897–1911 (2011).

34. Soudi, S. et al. Repeatability of adaptation in sunflowers reveals that genomic regions harbouring inversions also drive adaptation in species lacking an inversion. *eLife* **12**, RP88604 (2023).
35. DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I. & Nielsen, R. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2016).
36. Andrew, R. L. & Rieseberg, L. H. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution* **67**, 2468–2482 (2013).
37. Charlesworth, B. The effects of inversion polymorphisms on patterns of neutral genetic diversity. *Genetics* **224**, iyad116 (2023).
38. Feder, J. L., Gejji, R., Yeaman, S. & Nosil, P. Establishment of new mutations under divergence and genome hitchhiking. *Phil. Trans. R. Soc. B* **367**, 461–474 (2012).
39. Via, S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil. Trans. R. Soc. B* **367**, 451–460 (2012).
40. Yeaman, S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl Acad. Sci. USA* **110**, E1743–E1751 (2013).
41. Ostevik, K. L., Samuk, K. & Rieseberg, L. H. Ancestral reconstruction of karyotypes reveals an exceptional rate of nonrandom chromosomal evolution in sunflower. *Genetics* **214**, 1031–1045 (2020).
42. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
43. Pearce, R. J. et al. Multiple origins and regional dispersal of resistant *dhps* in african *Plasmodium falciparum* malaria. *PLoS Med.* **6**, e1000055 (2009).
44. Chan, Y. F. et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *pitx1* enhancer. *Science* **327**, 302–305 (2010).
45. Colosimo, P. F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
46. Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W. & Berner, D. The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**, 3944–3956 (2014).
47. Barrett, R. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
48. Yeaman, S. & Otto, S. P. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**, 2123–2129 (2011).
49. Kapun, M. & Flatt, T. The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol. Ecol.* **28**, 1263–1282 (2019).
50. Zong, S.-B., Li, Y.-L. & Liu, J.-X. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol. Biol. Evol.* **38**, 1317–1329 (2021).
51. Ma, L.-J. et al. Rapid and repeated climate adaptation involving chromosome inversions following invasion of an insect. *Mol. Biol. Evol.* **41**, msae044 (2024).
52. Hager, E. R. et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**, 399–405 (2022).
53. Weinstein, S. Y., Thrower, F. P., Nichols, K. M. & Hale, M. C. A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. *PLoS ONE* **14**, e0223018 (2019).
54. Reeve, J., Butlin, R. K., Koch, E. L., Stankowski, S. & Faria, R. Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (*Littorina saxatilis* and *L. arcana*). *Mol. Ecol.* <https://doi.org/10.1111/mec.17160> (2023).
55. Harringmeyer, O. S. & Hoekstra, H. E. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat. Ecol. Evol.* **6**, 1965–1979 (2022).
56. 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).
57. Messer, P. W. & Neher, R. A. Estimating the strength of selective sweeps from deep population diversity data. *Genetics* **191**, 593–605 (2012).
58. Colosimo, P. F. et al. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* **2**, e109 (2004).
59. Conte, G. L. et al. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. *Genetics* **201**, 1189–1200 (2015).
60. Heiser, C. B., Smith, D. M., Clevenger, S. B. & Martin, W. C. *The North American Sunflowers (Helianthus)* (Seeman Printery, 1969).
61. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
62. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
63. Josse, J. & Husson, F. missMDA: a package for handling missing values in multivariate data analysis. *J. Stat. Softw.* **70**, 1–31 (2016).
64. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253 (2012).
65. Huang, K. et al. The genomics of linkage drag in inbred lines of sunflower. *Proc. Natl Acad. Sci. USA* **120**, e2205783119 (2023).
66. Gautier, M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**, 1555–1579 (2015).
67. Pinheiro, J. & Bates, D. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000).
68. Johnson, R. C. et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genom.* **11**, 724 (2010).
69. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
70. Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).
71. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
72. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
73. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
74. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
75. Huber, C. D., DeGiorgio, M., Hellmann, I. & Nielsen, R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.* **25**, 142–156 (2016).
76. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
77. Hahn, M. W. *Molecular Population Genetics* (Oxford Univ. Press; Sinauer Associates, 2019).
78. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

79. Cheng, F. et al. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat. Genet.* **48**, 1218–1224 (2016).
80. Lu, K. et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* **10**, 1154 (2019).
81. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu. Rev. Genet.* **47**, 97–120 (2013).
82. Weigand, H. & Leese, F. Detecting signatures of positive selection in non-model species using genomic data. *Zool. J. Linn. Soc.* **184**, 528–583 (2018).
83. Weir, B. S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sinauer Associates, 1996).
84. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
85. 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).
86. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014).
87. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
88. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
89. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
90. Andrew, R. L. et al. Data from: Adaptation with gene flow across the landscape in a dune sunflower. *Dryad* <https://doi.org/10.5061/dryad.158pb518> (2011).
91. Huang, K. et al. Data from: Inversions contribute disproportionately to parallel genomic divergence in dune sunflowers. *Dryad* <https://doi.org/10.5061/dryad.bcc2fqznn> (2024).
92. Rieseberg, L. H. et al. Data from: Recent non-hybrid origin of sunflower ecotypes in a novel habitat. *Dryad* <https://doi.org/10.5061/dryad.j2448> (2012).
93. Huang, K. et al. hkchi/Dune\_parallelism: v1.0.0. *Zenodo* <https://doi.org/10.5281/zenodo.13983266> (2024).

## Acknowledgements

We thank S. Yeaman for helpful discussion of the analyses, as well as D. Irwin, A. Angert, K. Adams, S. Aitken and M. Johnson for their careful and critical reading of an earlier version of the manuscript. We also thank L. Donovan and C. Mason for help with soil analyses. This work was supported by start-up funding from Sun Yat-sen University to K.H. (no. 77010-12240014), an NSERC Postgraduate Scholarship awarded to K.L.O. (no. 393112) and an NSERC Discovery grant to L.H.R. (no. 327475).

## Author contributions

K.H., K.L.O. and L.H.R. conceived the study. K.L.O. collected and produced habitat data, seed size data and sequence data for samples at Monahans Sandhills State Park. M.T., N.B. and R.L.A. contributed other sequence data. K.H., K.L.O. and M.J. analysed the data. G.L.O. helped with analyses. K.H. K.L.O. and L.H.R. wrote the paper. All authors approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-024-02593-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02593-4>.

**Correspondence and requests for materials** should be addressed to Kaichi Huang or Kate L. Ostevik.

**Peer review information** *Nature Ecology & Evolution* thanks Emma Berdan, Angela Hancock and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

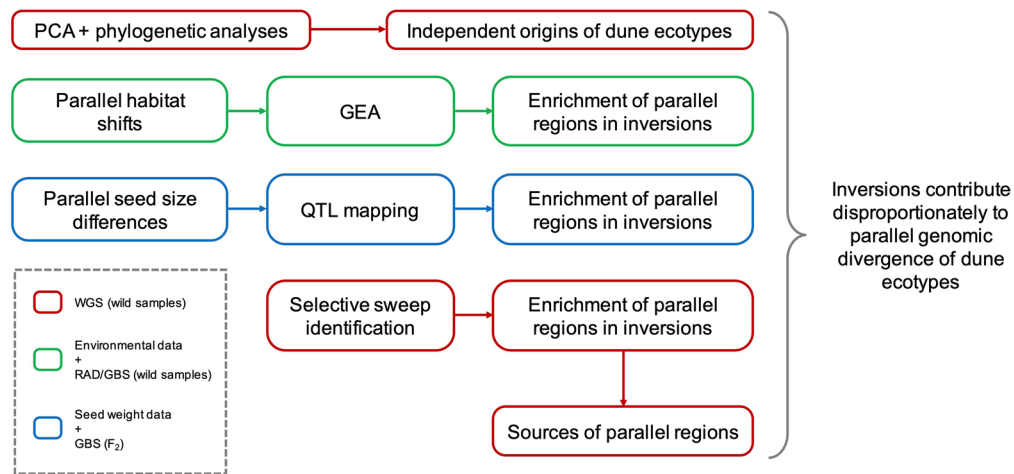
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

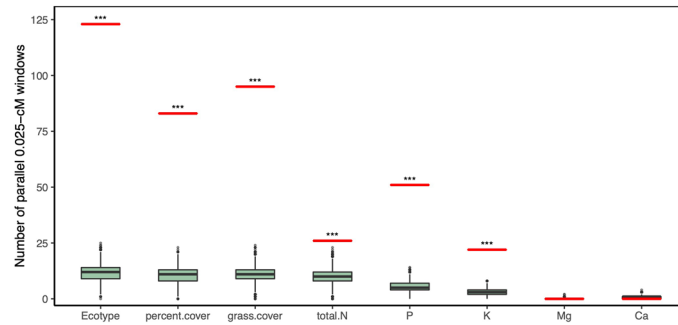
**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

<sup>1</sup>School of Ecology, Sun Yat-sen University, Shenzhen, China. <sup>2</sup>Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada. <sup>3</sup>Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA, USA. <sup>4</sup>Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada. <sup>5</sup>Department of Biology, University of British Columbia, Kelowna, British Columbia, Canada. <sup>6</sup>School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia. <sup>7</sup>Department of Biology, University of Victoria, Victoria, British Columbia, Canada. <sup>8</sup>These authors contributed equally: Kaichi Huang, Kate L. Ostevik. ✉ e-mail: [kaichi.huang@botany.ubc.ca](mailto:kaichi.huang@botany.ubc.ca); [kostevik@ucr.edu](mailto:kostevik@ucr.edu)

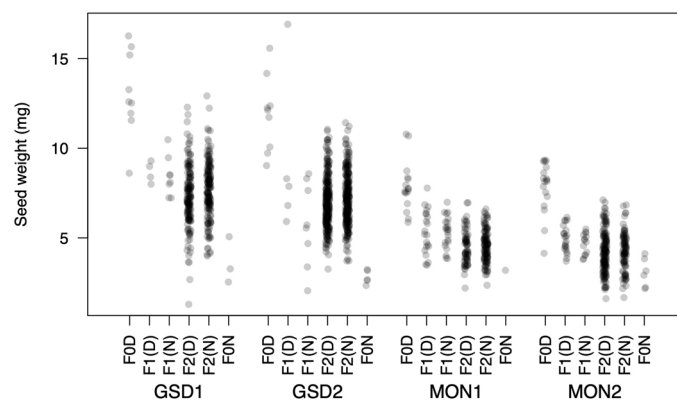


**Extended Data Fig. 1 | Illustration of the analyses in this study.** Colors of boxes indicate the data sets used in the analyses.

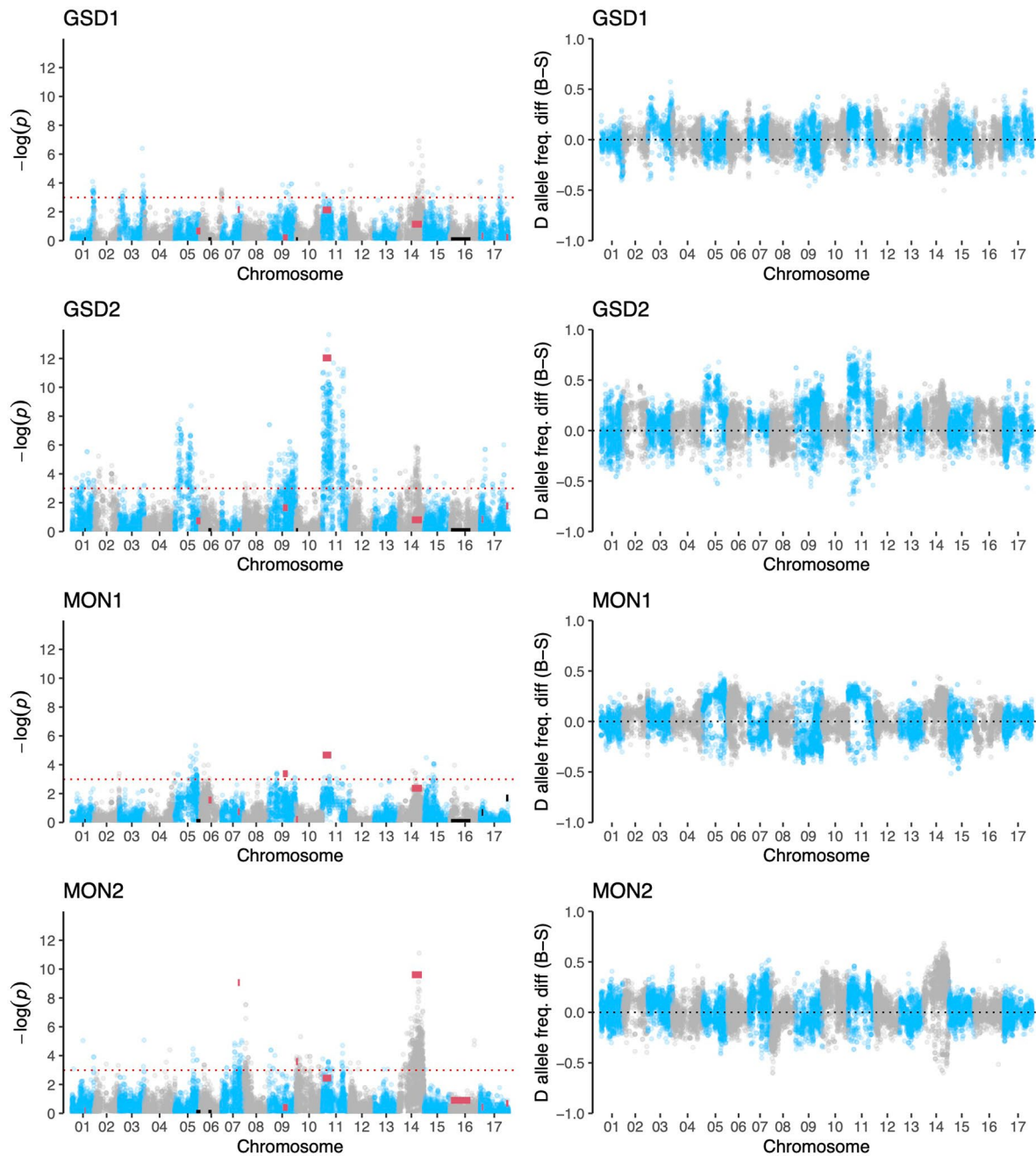


**Extended Data Fig. 2 | Expected and observed counts of parallel 0.025-cM windows associated with each environmental variable between GSD and MON.** Each boxplot shows the simulated counts of parallel windows across  $N = 10,000$  iterations of random draws of  $m$  0.025-cM windows from each location, where  $m$  is the count of associated windows for each location.

Each boxplot shows the median, quartiles, standard range ( $1.5 \times \text{IQR}$ ) and points show outliers beyond the standard range. The red bars show the observed counts of parallel associated windows. The asterisks denote the significance of one-tailed tests between the observed counts and the simulated counts (\*\*\*:  $P < 0.01$ ).

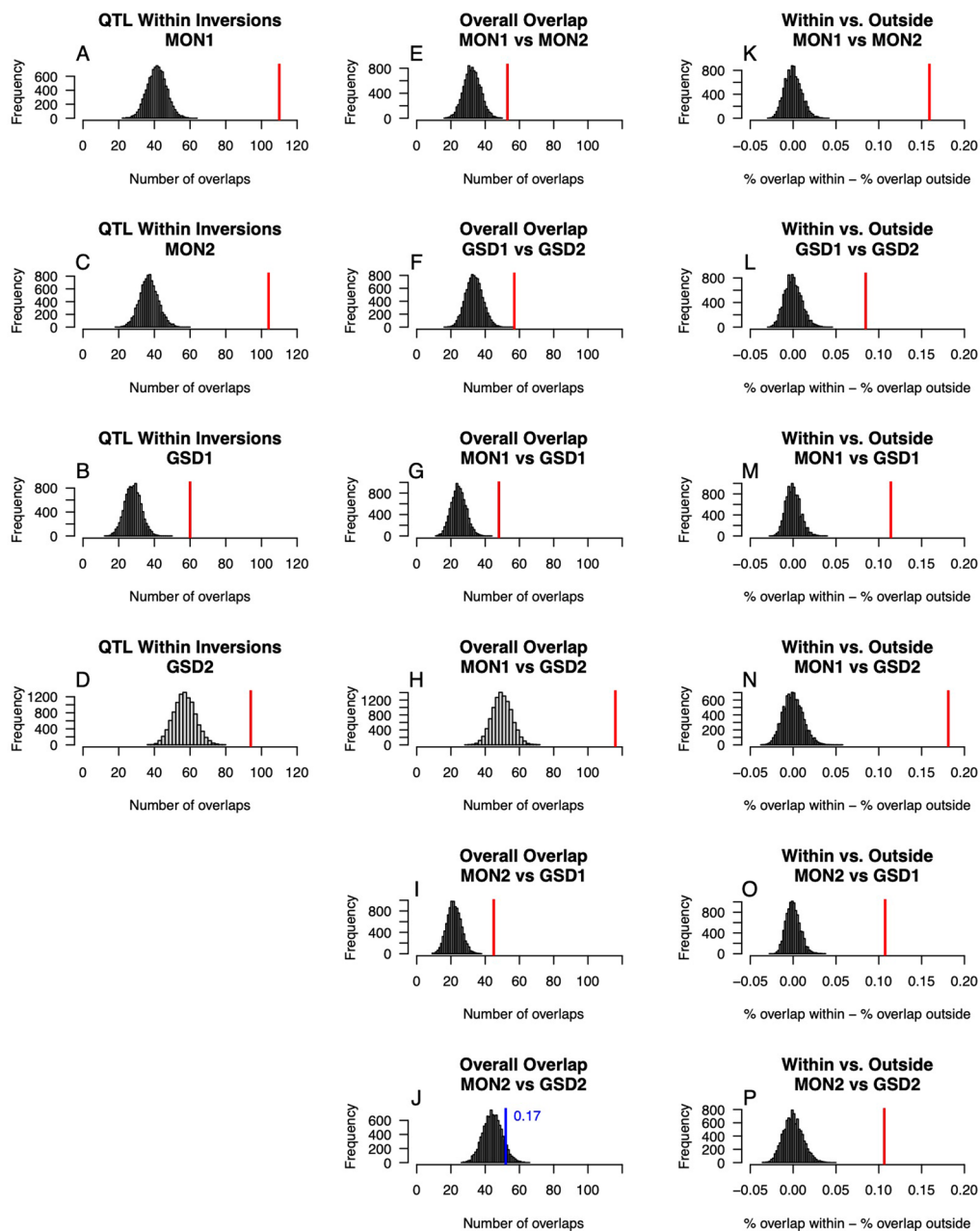


**Extended Data Fig. 3 | Seed weight of parents,  $F_1$ s, and  $F_2$ s in QTL mapping populations.** The letter in brackets shows the cytoplasm type of the  $F_1$  and  $F_2$  individuals.

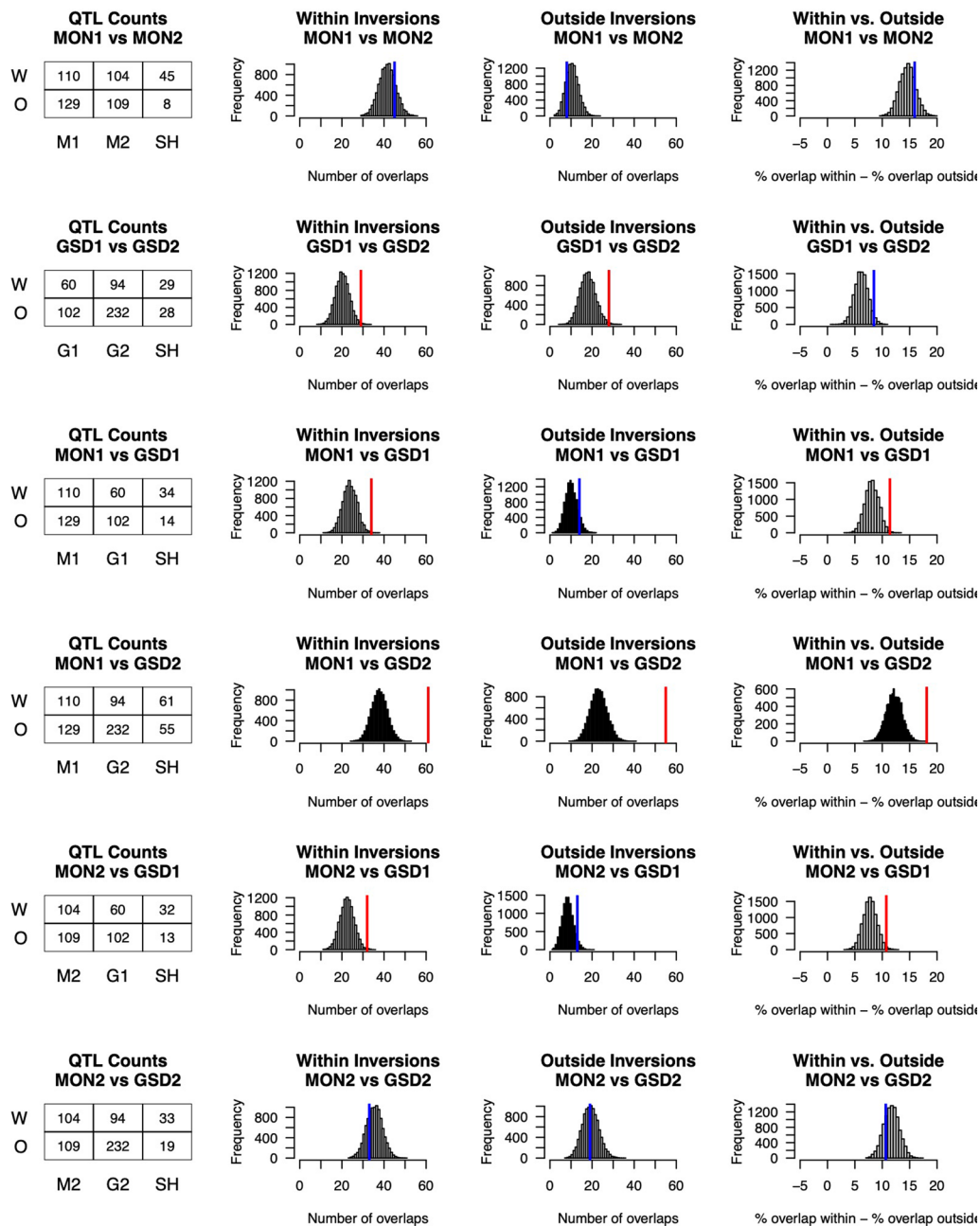


**Extended Data Fig. 4 |**Manhattan plots showing the significance (left) and effect size (right) of one-sided Fisher's exact tests performed to determine whether dune alleles are more often found in large-seeded  $F_2$ s and the allele frequency difference between the two types. Black and red solid bars indicate the locations of monomorphic (minor allele frequency  $< 0.05$ ) and polymorphic

inversions in each mapping population, respectively. The height of each bar represents the significance of Fisher's exact tests when they were treated as single bi-allelic loci. The red horizontal dashed line is a significance threshold based on a Bonferroni correction of the number of independent regions of the genome.

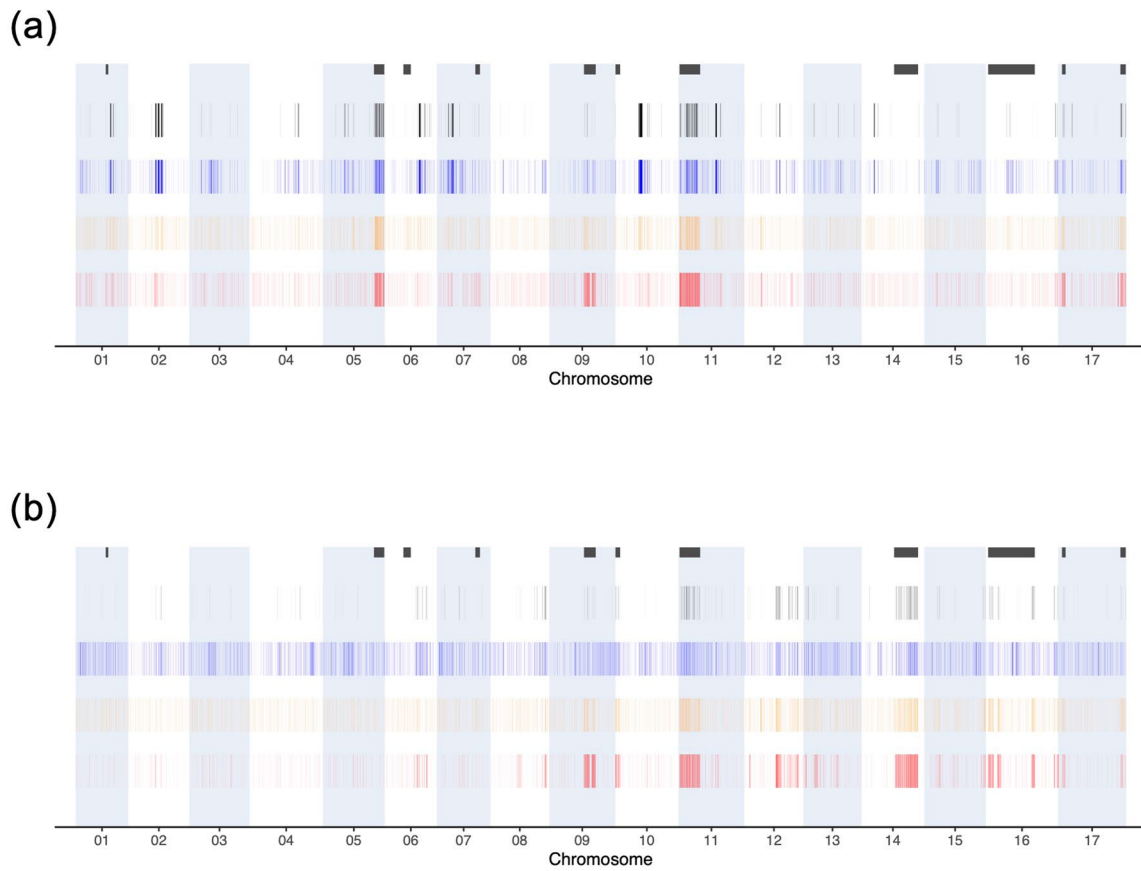


**Extended Data Fig. 5 | Seed size QTL are found within inversions (A-D), shared across mapping populations (E-J), and shared within inversions (K-P) more often than expected by chance.** Histograms represent the number of QTL within inversions (A-D) or shared QTL windows (E-P) across 10,000 full QTL randomizations. The vertical lines show the observed number of those values.



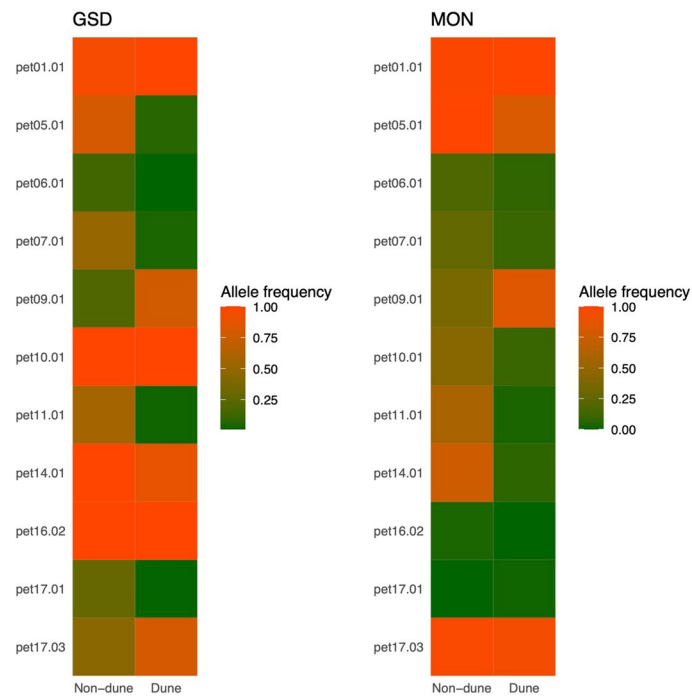
**Extended Data Fig. 6 | Seed size QTL are shared more often than expected by chance, especially within inversions.** Tables show the count of QTL that are significantly associated with seed size within (W) and outside (O) inversions for each mapping population and shared (SH) between mapping populations. Histograms represent the number of shared windows with significant seed size

associations or the difference in percent overlap within versus outside inversions across 10,000 randomizations that preserve the number of QTL within vs between inversions. The vertical lines show the observed value of each statistic. Red lines denote values outside expectations based on the randomization, while blue lines denote values that fall within expectations.



**Extended Data Fig. 7 | Selective sweep regions in (a) GSD system and (b) MONsystem.** Outlier regions for each statistic and overlapping selective sweep regions are indicated by bars of different colors: black - selective sweep regions, blue - outlier regions of CLR, orange - outlier regions of ROD, red - outlier regions

of  $F_{ST}$ . Black bars show the locations of large non-recombining haplotype blocks (mainly due to inversions) identified in Todesco et al. (2020), several of which were shown to contribute to dune adaptation.



**Extended Data Fig. 8** | Allele frequency of inversions in dune and non-dune ecotypes in each of the dune populations.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used to collect data.
Data analysis	<div>R (v4.0.3) BayPass (v2.1) missMDA R package (v1.19) nlme R package (v3.1) randomForest R package (v4.6.14) NextGenMap (v0.5.3) GATK (v4.1.4.166) MEGA (v10.1.6) SNPRelate R package (v1.18.1) ggtree R package (v3.1.3) SweepFinder2 (v1.0) VCFtools (v0.1.14) PLINK (v1.9) BEDTools (v2.29.2) Beagle (v5.1) IQ-TREE (v1.6.10) The software Twisst is not versioned but is available on GitHub: <a href="https://github.com/simonhmartin/twisst/">https://github.com/simonhmartin/twisst/</a>.</div>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Environmental data of GSD were published previously: <https://doi.org/10.5061/dryad.158pb518>; those of MON generated herein have been deposited in Dryad: <https://doi.org/10.5061/dryad.bcc2fqznn>. Sequencing data of GSD samples were published previously: <https://doi.org/10.5061/dryad.j2448>; those of MON samples generated herein have been deposited in NCBI Sequence Read Archive (SRA) under accession PRJNA1145483. Seed weight data and sequencing data of F2 plants in QTL mapping experiments have been deposited in Dryad: <https://doi.org/10.5061/dryad.bcc2fqznn> and in NCBI SRA under accession PRJNA1145296, respectively. Whole genome sequencing data of *H. petiolaris* were published previously and are stored in NCBI SRA under accessions PRJNA322345 and PRJNA532579.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Sequencing wild-collected plants and performing cross experiments to explore the patterns of parallel genetic evolution in two ecotypes of <i>Helianthus petiolaris</i> .
Research sample	Wild <i>Helianthus petiolaris</i> (Asteraceae) plants from across the range. Population details provided in supplementary material.
Sampling strategy	We used 195 samples for data genotype-environment association analyses, which is in line with standard practice. In quantitative trait locus mapping, we adopted a selective genotyping approach and selected 12 offspring individuals with the largest seeds and 12 with the smallest in each cross type. For population genomic analyses, we included 259 samples across 52 populations, which is in line with standard practice.
Data collection	Sequencing data were collected by sequencing service providers, as described in the methods.
Timing and spatial scale	Samples were collected in 2015. Details of locations provided in supplementary material.
Data exclusions	In the analyses of selective sweep in each dune ecotype, samples from within 100 km of the other dune ecotype were excluded to remove their impact.
Reproducibility	Samples were randomly chosen from wild collections.
Randomization	Samples were grouped into populations based on the design of the original papers, which used geography to cluster them.

Blinding

Did the study involve field work? ☒ Yes ☐ No

## Field work, collection and transport

Field conditions

Location

Access & import/export

Disturbance

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input checked="" type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Public health
<input checked="" type="checkbox"/>	<input type="checkbox"/> National security
<input checked="" type="checkbox"/>	<input type="checkbox"/> Crops and/or livestock
<input checked="" type="checkbox"/>	<input type="checkbox"/> Ecosystems
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other significant area

## Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Demonstrate how to render a vaccine ineffective
<input checked="" type="checkbox"/>	<input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent
<input checked="" type="checkbox"/>	<input type="checkbox"/> Increase transmissibility of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Alter the host range of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable evasion of diagnostic/detection modalities
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable the weaponization of a biological agent or toxin
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other potentially harmful combination of experiments and agents

## Plants

Seed stocks	The seeds used in this study were collected from Monahans Sandhills State Park, Texas in 2015 and stored at the University of British Columbia.
Novel plant genotypes	N/A
Authentication	N/A