

Pleiotropy and facilitation of local adaptation in the silverleaf sunflower *Helianthus argophyllus*

Uzezi E. Okinedo¹, Brook T. Moyers¹

1. Department of Biology, University of Massachusetts Boston, Boston, Massachusetts,
United States of America

Current Address: 100 Morrissey Blvd., Central Receiving–Biology, Boston, Massachusetts,
02125, United States of America

* Corresponding author

E-mail: lizzyokinedo@gmail.com (UEO)

Abstract

Local adaptation drives changes in population phenotypes that confer survival or increased reproduction in certain environments. Local adaptation may be hindered or facilitated by pleiotropy, which is the control of multiple traits by a single genetic locus. In this study, we explored the connections between pleiotropy and local adaptation in the Texas endemic silverleaf sunflower, *Helianthus argophyllus*. Populations of *H. argophyllus* exhibit a bimodal life history strategy, consisting of tall, late-flowering forms and short, early-flowering forms that occur in close geographic proximity. The expression of life history traits within *H. argophyllus* populations is linked to local adaptation and controlled by a highly pleiotropic locus. Still, we do not know how local adaptation and pleiotropy interact at the transcriptomic level. We identified putatively locally adapted genes using whole RNA sequencing data and two selection outlier approaches. We assessed transcriptomic pleiotropy by evaluating whether allelic variants within genes control the expression of other genes (an eQTL approach) and gene co-expression network connectivity. Our results show that candidate locally adaptive genes identified by both methods are enriched for eQTL loci. Candidate genes identified through associations with environmental variables exhibit modular expression and have significantly lower network connectivity than non-adapted genes. In contrast, candidate genes identified only by controlling for population structure have significantly higher network connectivity. Our results suggest that the role of pleiotropy in local adaptation, at least at the transcriptomic level, depends on the function of the specific loci under selection.

Author Summary

This study investigates how genes contribute to local adaptation in *Helianthus argophyllus*, a sunflower species with distinct ecotypes. Using transcriptomic data and two methods involving environmental association and population structure analysis, we found that adaptive genes can be categorized into two types: pleiotropic "hub" genes, which affect multiple traits, and modular genes, which affect fewer, more specific traits. Environment-associated genes tended to be modular, while population-structure-associated genes were more pleiotropic. These patterns support a dynamic model of adaptation, where pleiotropy is beneficial early in adaptation and modularity becomes increasingly crucial as populations refine their traits. The findings provide new insights into how gene network architecture influences evolutionary responses to environmental change, particularly in diverging populations, such as those of *H. argophyllus*.

Introduction

Pleiotropy is an important concept in adaptive evolution. It describes how a single genetic locus influences multiple phenotypic traits [1,2]. It can arise from the effects of shared cis-regulatory elements on gene expression or from the joint influence of gene networks, forming variational modules [3]. These modules often represent gene co-expression networks, where connections between trait nodes and gene nodes indicate genetic influence on phenotype expression [3,4]. Understanding how pleiotropy shapes adaptation is critical for assessing whether adaptive loci function as regulatory hubs or if selection favors more specialized genetic architectures.

Evolutionary models, such as Fisher's geometric model and the cost of pleiotropy theory, suggest that highly pleiotropic or large-effect loci may constrain adaptation by generating trade-offs that shift an organism from its fitness optimum [5,6]. Consequently, adaptation often proceeds

through small-effect loci or modular subunits, which minimize detrimental consequences on an organism's fitness. However, extensions of Fisher's theory propose that pleiotropic loci can accelerate adaptation, mainly when populations are far from their fitness optimum, such as during habitat expansion or environmental shifts [1,6]. Experimental evidence supports this by showing that adaptive loci with larger effect sizes often correlate with increased distance from optimal fitness [4,7]. This raises the question of whether locally adaptive loci tend to be pleiotropic or more functionally specialized, a key focus of this study.

Helianthus argophyllus (Silverleaf sunflower) is an annual species in the Asteraceae family, endemic to Texas and parts of Florida, and grows in sandy soils across diverse habitats [8][9]. It shares morphological similarities with *Helianthus annuus* but is distinguishable by dense silvery-white pubescence on its leaves, stems, and phyllaries [10,11]. A key feature of *H. argophyllus* is its bimodal life history strategy, which has led to the emergence of two distinct ecotypes: a tall, late-flowering ecotype with low branching found in drier inland Texas and a short, early-flowering, highly branched ecotype, distributed on barrier islands along the Gulf of Mexico [12,13]. Despite their narrow geographical range, these ecotypes coexist within the same populations but differ in flowering phenology [13]. The tall, late-flowering ecotype blooms from late August to early September and is the primary ecotype in inland populations. In contrast, the short, early-flowering ecotype blooms in June and is more common in coastal populations [9]. This bimodal flowering strategy is controlled by HaFT1, a FLOWERING LOCUS T homolog introgressed from *H. annuus*. HaFT1 is located in a 10Mb low-recombination region on chromosome 6 of *H. annuus*, and likely in the same position in *H. argophyllus*, although this chromosome has a reciprocal translocation with *H. annuus*

chromosome 15. The short, early-flowering ecotype possesses one or both copies of HaFT1, either homozygous or heterozygous for its presence, representing the ancestral allele. In contrast, the tall, late-flowering ecotype lacks HaFT1 entirely (homozygous for its absence, the derived allele) [8,14].

Some studies suggest that adaptive loci are modular and minimize pleiotropic constraints, while others argue that pleiotropy is a key feature of adaptive evolution, allowing correlated trait changes [4,15–17]. However, few studies have directly examined whether adaptive loci in naturally diverging populations tend to be pleiotropic or functionally specialized. Understanding this relationship is critical for determining whether selection favors genetic architectures that minimize trade-offs or integrate multiple adaptive changes through pleiotropic loci [18].

Helianthus argophyllus provides an ideal system for studying pleiotropy in adaptation because it exhibits clear ecotypic divergence with a known genetic basis. The HaFT1 locus is strongly associated with flowering time, which correlates with fitness-related traits such as seed size, seed mass, and plant height [13,14,19–21]. If pleiotropic loci drive adaptation in this system, we would expect adaptive genes to function as regulatory hubs influencing multiple traits. Alternatively, if selection favors modularity, adaptive loci may exhibit low pleiotropy, with distinct genes controlling different adaptive traits.

In this study, we apply a transcriptomic and network-based approach to investigate the role of pleiotropy in the local adaptation of *H. argophyllus*. To estimate genetic diversity indices, we sequence the transcriptomes of individuals representing both tall, late-flowering, and short, early-flowering ecotypes. We identify adaptive loci using an environment-dependent model and a population structure-based model. We assess pleiotropy by examining associations between gene expression and genetic variants. We construct gene coexpression networks to determine whether adaptive variants cluster in specific gene modules.

Our results provide new insights into the role of pleiotropy in local adaptation by examining the network properties and functional enrichment of adaptive genes in *Helianthus argophyllus*. We found that candidate locally adaptive genes identified using PCAdapt [22,23], which detects selection based solely on population structure, exhibit higher network connectivity but are not enriched in gene co-expression network modules. Adaptive loci identified through PCAdapt may function as highly connected regulatory genes whose involvement in multiple biological processes restrains their evolvability, as suggested by their low between-population genetic differentiation. In contrast, adaptive genes identified using LFMM [24], incorporating environmental associations, had lower network connectivity but were significantly enriched in specific gene network modules, suggesting a potential trade-off between pleiotropy and modular adaptation [25]. This set of candidate adaptive genes is enriched in primary metabolic pathways such as histidine biosynthesis, glycolysis, and pyruvate metabolism that appear to function within modular genetic architectures, facilitating local adaptation through specialized functional units rather than broad regulatory control. The two methods appear to identify candidate genes that likely have experienced different routes of local adaptation, possibly underlying traits that

were closer to or distant from their respective fitness optima. These findings support that local adaptation involves both functional specialization (modularity) and highly pleiotropic genes acting as central hubs in gene networks [15]. This aligns with recent revisions to Fisher’s model, which propose that while pleiotropic loci can drive adaptation when phenotypes are far from their fitness optimum, modular architectures may be favored in fine-scale local adaptation, where selection acts on genes with specific metabolic or physiological roles, rather than broadly connected regulatory genes [3,4,26].

Results

RNA sequencing

We analyzed total RNA sequence data for 19 *Helianthus argophyllus* samples collected from North inland (late flowering, n = 9) and coastal sub-populations (early flowering, n = 10) of its range in Texas to identify genomic footprints of local adaptation. Hereafter, we refer to these populations as “north” and “coast”. A total of 6,283,383 SNPs and Indels were called from Freebayes [27]. After filtering, we retained 106,133 biallelic SNPs in 19,394 genes.

Genetic diversity and population structure

Nucleotide diversity (θ_π) and Watterson theta (θ_w) had an approximate value of 0.2, indicating high levels of genetic diversity within the *H. argophyllus* population. Tajima’s D was slightly negative, $D = -0.09$, possibly due to the presence of rare alleles or population expansion. PCA (Fig. 1A) and F_{ST} (mean $F_{ST} = 0.21$) demonstrated genetic differentiation between the sampled sub-populations. PC1 (18.4 PVE) and PC2 (13 PVE) showed a clear differentiation between north and coast samples. Admixture analysis using NGSAdmix suggests that the population is

stratified into two distinct genetic groups at $K = 2$, representing the two sampled populations (Fig. 1B), with differences between the populations remaining at higher K values.

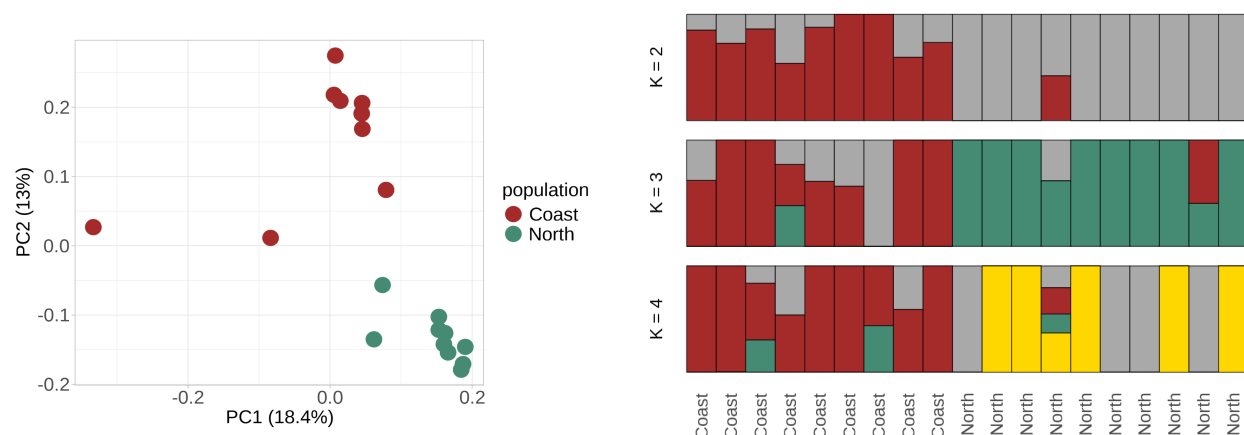


Fig 1. Genetic differentiation and population structure in *H. argophyllus*.

A: Genetic differentiation along the first two PCs. Percentages in brackets show the proportion of variance explained by each PC axis. **B:** Approximate admixture proportions for each sampled individual at three different levels of K .

Selection scans

We used two methods to identify genes underlying the local adaptation of *H. argophyllus*:

LFMM 2 and PCAdapt [22,24]. LFMM 2 takes an environment-genotype association approach that combines population structure and environmental variables, while PCAdapt uses information from population structure only. For our LFMM model, we controlled for population structure using $K = 4$ latent factors estimated from the scree plot of the genotype-based principal component analysis (PCA) (Fig. S5). We defined the main axes of environmental variation using a principal component analysis (PCA) of 19 BIOCLIM variables. We used the first five PCs, which explain 98.95% of the total variation. Out of all 106,133 biallelic SNPs, 1,438 distributed

across 560 genes bore signatures of selection in the LFMM model ($q < 0.05$). BIOCLIM PC1, PC2, PC3, PC4, and PC5 were associated with 44, 125, 1,271, 15, and 138 genes, respectively. In the PCAdapt analysis, using $K = 4$ latent factors, 4,968 SNPs distributed across 1,122 genes were identified as selection outliers ($q < 0.05$). Twelve SNPs in 10 genes were identified by both methods, fewer than expected by chance (p -value $< 2.2 \times 10^{-16}$, Fisher's exact test). For further analyses, we considered genes with outlier SNPs from either approach as candidates for local adaptation ("selection outliers"). Genes containing bi-allelic SNPs not identified as selection outliers make up our control comparison group.

Proportion of eQTLs and eGenes in selection outliers

To investigate whether selection outlier loci were more pleiotropic than control genes, we tested associations between SNPs and gene expression to identify *trans*-regulatory variants that control the expression of other genes. 42,638 genes were used for the eQTL analysis. The model included admixture proportions ($K = 4$) as cofactors to control for population structure. A total of 1,824 (4.3%) eQTL genes (genes containing eQTL SNPs) were identified and associated with the expression of 3,129 (7.3%) genes (eGenes) ($FDR < 0.01$). Twenty-six (0.06%) genes were identified as both eQTL and eGenes. Of the 560 LFMM outlier genes, 114 (20.4%) were eQTLs, while 208 (18.5%) PCAdapt outliers were eQTLs. Eleven (2%) LFMM outlier genes and 24 (2.1%) PCAdapt outlier genes were eGenes. Compared to the rest of the transcriptome, both outlier gene sets harbored more eQTLs ($p < 2.2e-16$, Fisher's exact test) and fewer eGenes, respectively ($p < 2.2e-16$, Fisher's exact test; Fig. 2).

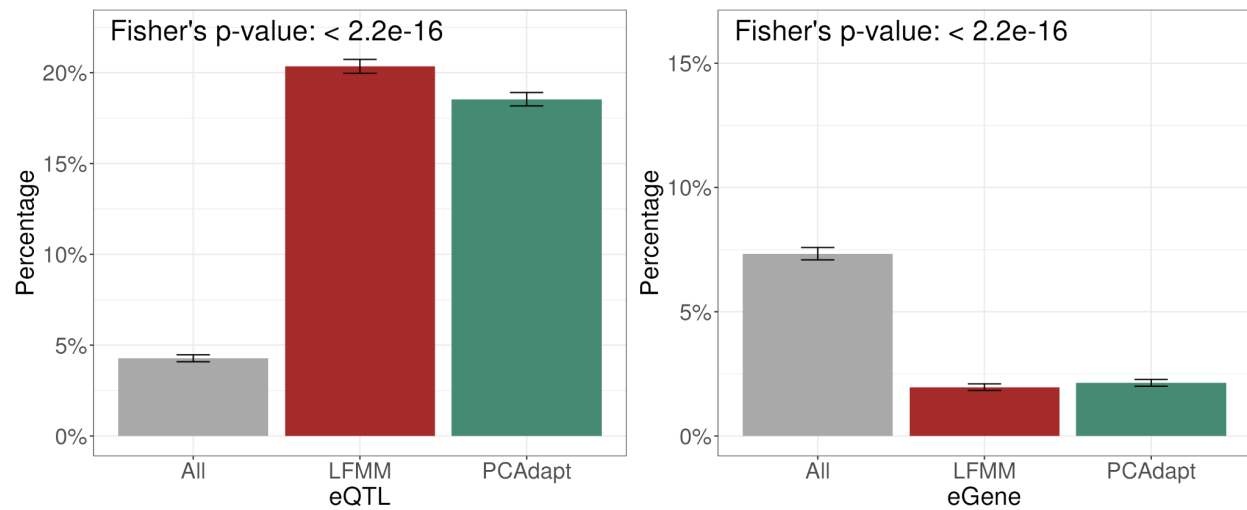


Figure 2. Compared to control genes, the proportion of eQTLs and eGenes in candidate adaptive genes identified by the LFMM and PCAdapt methods is significantly higher for eQTLs and lower for eGenes, respectively. Error bars were calculated using a 95% bootstrapped confidence interval.

Evidence of selection at outlier loci

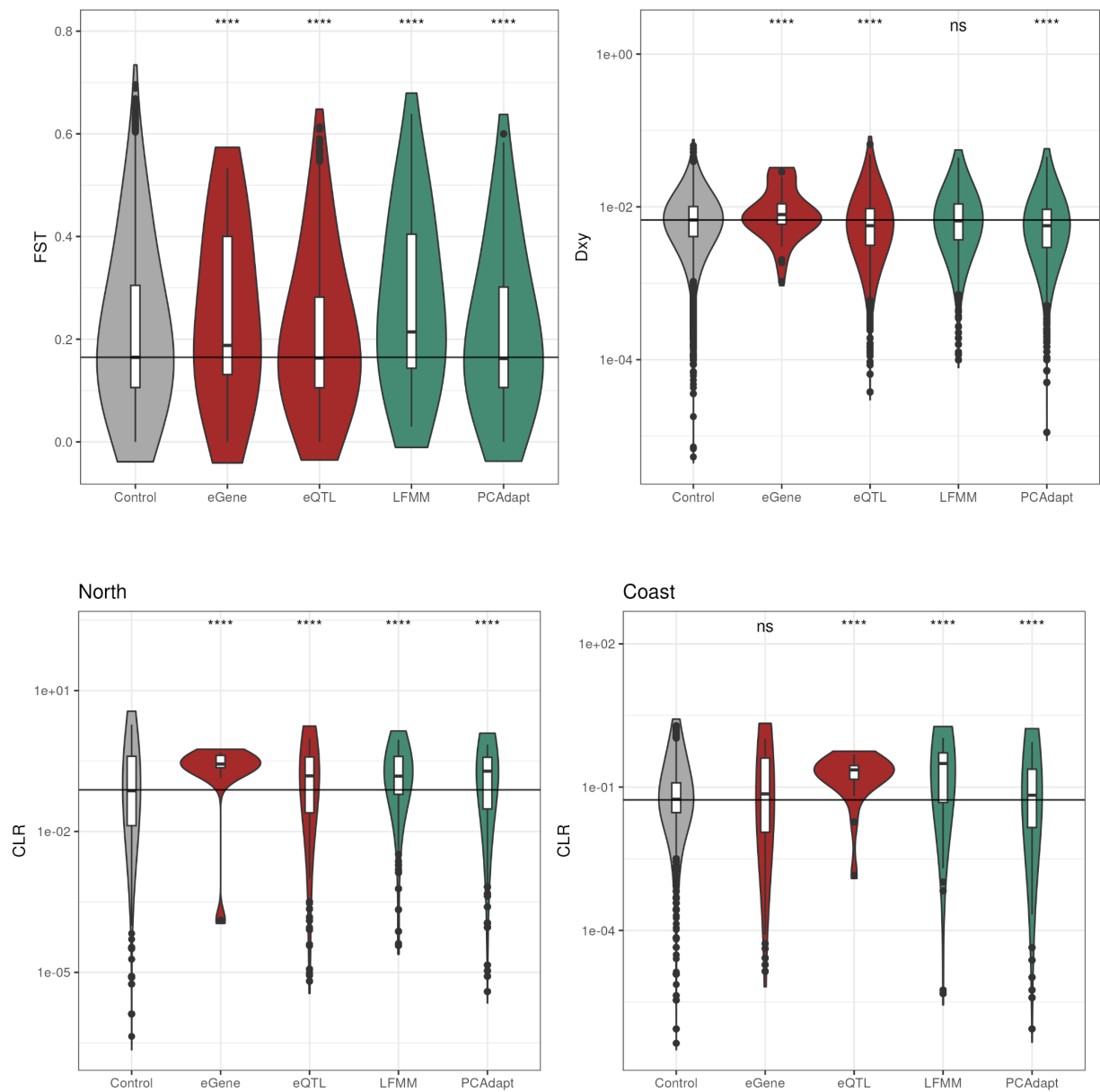
Genetic differentiation, as estimated by F_{ST} , was considerably higher for LFMM outlier genes than for control genes (p -value < 2.2e-16, Wilcoxon rank sum test). However, they were not significantly different in D_{xy} values (Fig. 3). In contrast, PCAdapt outlier genes had lower F_{ST} and lower D_{xy} compared to control genes (p -value = 3.941e-06 and < 2.2e-16, respectively, Wilcoxon rank sum test) (Fig. 3). These results emphasize that our two selection outlier methods appear to be identifying genes with substantially different signatures of selection. We also find that F_{ST} and D_{xy} are significantly elevated in eQTL genes and reduced in eGenes (Fig. 3).

We estimated two measures of genetic diversity, θ_π (nucleotide diversity) and θ_W (Watterson's estimator), to explore additional evolutionary patterns in outlier genes. Nucleotide diversity and

Watterson's θ should be approximately equal in a population under equilibrium conditions (evolving neutrally with constant demographic size) [28]. We would expect genes experiencing positive or purifying selection to exhibit reduced genetic diversity, while balancing selection (or a higher local mutation rate) might result in increased genetic diversity, and transcriptome-wide differences between θ_π (nucleotide diversity) and θ_w likely reflect population size changes. In the north population, PCAdapt outlier genes had significantly lower θ_w compared to the control genes (p -value = 0.003819, Wilcoxon rank sum test) (Fig. S3), but θ_π was not significantly different (p -value = 0.3752, Wilcoxon rank sum test). A similar pattern was observed for PCAdapt outliers in the coast population, although θ_π was significantly larger than in the control (θ_π ; p -value = 0.004505, θ_w ; p -value = 0.001016, Wilcoxon rank sum test). In the north population for LFMM outlier genes, θ_π was significantly higher than the control, while θ_w was not significantly different (θ_π ; p -value < 2.2×10^{-16} , θ_w ; p -value = 0.06571, Wilcoxon rank sum test). An opposite pattern was observed in the coast population: LFMM outlier genes exhibited significantly lower θ_w compared to the control (p -value = 1.57×10^{-14} , Wilcoxon rank sum test) and no significant difference in θ_π (p -value = 0.2015, Wilcoxon rank sum test).

Finally, we further validated our outlier genes using Fay and Wu's H neutrality measure and the Composite Likelihood Ratio (CLR) test for selective sweeps (Fig. 2B). CLR estimates at the candidate adaptive genes identified by the LFMM and PCAdapt methods, as well as eQTL and eGenes, were significantly more positive compared to the control genes (p -value < 2.2×10^{-16} , Wilcoxon rank sum test for both LFMM and PCAdapt) (Fig. 2B). PCAdapt and LFMM outliers in both populations had more negative Fay & Wu's H estimates compared to control genes (LFMM-north; p -value = 0.001; PCAdapt-north; p -value = 6.623×10^{-6} ; LFMM-coast; p -value =

4.998e-05, PCAdapt-coast; p -value = 5.977e-05) (Fig. S4). These results suggest that both candidate locally adapted gene sets are enriched for selective sweeps in both populations.



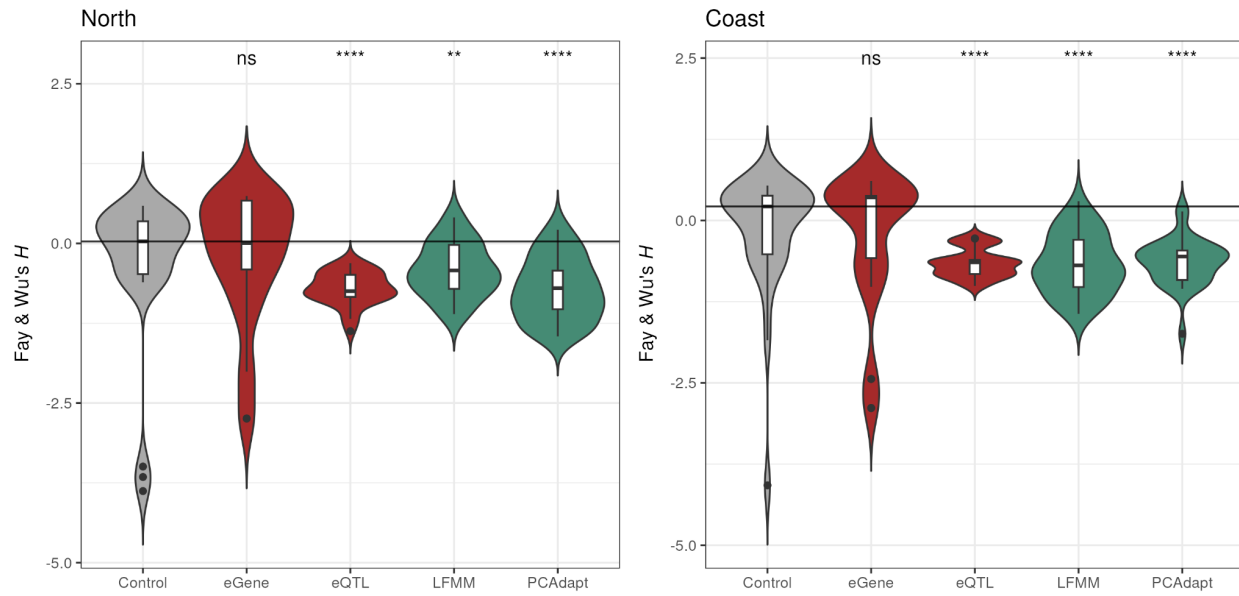


Fig 3. Genetic differentiation and selective sweep at adaptive genes. **A:** Weighted F_{ST} estimates across adaptive genes, eQTLs, eGenes, and the rest of the transcriptome (control). **B:** Distribution of CLR and **C:** Fay and Wu's H values. The horizontal line represents the median value of the transcriptome-wide background for all estimated values. Differences significant at $p < 0.05$ are indicated, even where they appear visually similar.

Connectivity of selection outliers

We estimated gene co-expression networks to further evaluate the pleiotropy of genes harboring selection outliers. Our gene co-expression analysis identified 44 modules, with a median of 41 genes per module (range, 9–2,209). Three modules contained more than expected LFMM genes (p -value < 0.05 , Fisher's exact test), while no WGCNA module had significant enrichment for PCAdapt genes. We used the connectivity of our candidate adaptive genes, as well as our eQTLs and eGenes, to estimate transcriptomic pleiotropy. More highly connected genes are putatively more pleiotropic in that changes to them will have broader effects on the expression of other

genes. PCAdapt outlier genes and eQTLs had higher connectivity than control genes (p -value $< 2.2 \times 10^{-16}$, Wilcoxon rank sum test). In contrast, LFMM outlier genes and eGenes were less connected than the control (p -value $< 2.2 \times 10^{-16}$, Wilcoxon rank sum test) (Fig. 4).

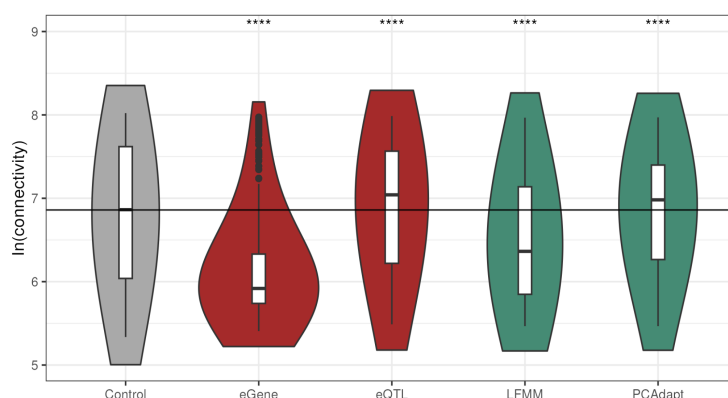


Fig 4. Gene network connectivity at candidate adaptive genes. Estimated connectivity of candidate adaptive genes, eGenes, and eQTLs compared to the transcriptome-wide control.

Annotation of adaptation loci and gene co-expression networks

Our gene overrepresentation analysis revealed significant enrichment for pollen recognition and protein serine/threonine kinase activity in both LFMM and PCAdapt candidate genes (Fig. 5). LFMM genes were additionally enriched in functions related to sulfotransferase, catalase, and oxidoreductase activities. In contrast, PCAdapt genes were significantly enriched for signal transduction, defense response, and dehydrogenase activity. The few genes identified by both methods included members of the bZIP family of transcription factors, whose functional orthologues in *Arabidopsis thaliana* show high expression in reproductive tissues such as pollen, inflorescence meristem, carpels, and petals. Several of these genes also exhibited high expression in guard cells, the shoot apex, and the shoot system [29]. WGCNA modules with more than expected LFMM candidate genes were involved in pathways annotated as predominantly

primary metabolism, including histidine biosynthesis, de novo purine biosynthesis, pyruvate oxidation, triacylglycerol biosynthesis, and leucine biosynthesis (Table 1).

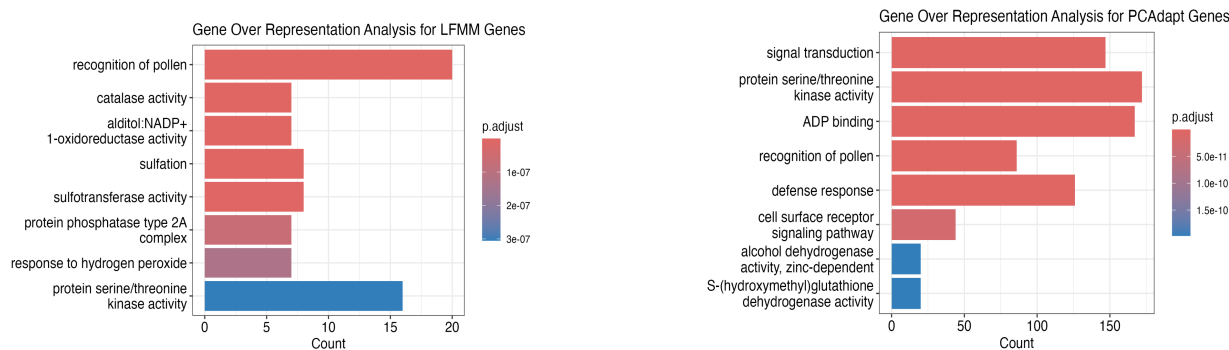


Fig 5. Gene ontology enrichment for biological processes of adaptive genes identified by LFMM (top), PCAdapt (middle), and both LFMM and PCAdapt (bottom).

Table 1. KEGG Module Annotation of the three network modules containing more than expected LFMM adaptive genes.

WGCNA Module	KEGG ID	Description	Adjusted p-value
Dark grey	M00026	Histidine biosynthesis, PRPP => histidine	0.00063984
	M00129	Ascorbate biosynthesis, glucose-1P => ascorbate	0.0014631
Dark turquoise	M00002	Glycolysis, a core module involving three-carbon compounds	0.0006257
	M00048	De novo purine biosynthesis, PRPP + glutamine => IMP	0.00117279

	M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	0.00228412
	M00130	Inositol phosphate metabolism, PI=> PIP2 => Ins(1,4,5)P3 => Ins(1,3,4,5)P4	0.00288918
Green	M00307	Pyruvate oxidation, pyruvate => acetyl-CoA	0.00107016
	M00089	Triacylglycerol biosynthesis	0.00954748
	M00432	Leucine biosynthesis, 2-oxo isovalerate => 2-oxo isocaproate	0.03465363

Discussion

Pleiotropy is a key factor in adaptive evolution, influencing how genetic loci contribute to phenotypic variation and shaping evolutionary trajectories [1,2]. Fisher's geometric model posits that adaptation is more efficient through small-effect loci, as large-effect mutations may overshoot the fitness optimum and cause maladaptation [3]. However, Orr's extension of Fisher's model suggests that pleiotropic loci can facilitate rapid adaptation in novel environments, particularly following range expansion [7,30]. The role of pleiotropy in adaptation has been debated, with some models suggesting that highly pleiotropic loci constrain evolution by generating trade-offs [25,31]. In contrast, others propose that such loci can accelerate adaptation [32], particularly when populations are far from optimal fitness [4,33].

Our findings provide new insights into the interplay between pleiotropy and local adaptation in the silverleaf sunflower *Helianthus argophyllus*. We identified distinct sets of candidate adaptive loci using two methods: PCAdapt and LFMM selection scans. Both sets are enriched for loci associated with the expression of other genes (eQTL) and signals of positive selection. PCAdapt outliers, identified through controlling for population structure, exhibited significantly higher gene co-expression network connectivity than the transcriptome-wide control without specific enrichment in any co-expression module. This suggests that the candidate adaptive loci identified by PCAdapt fit Orr's extension of Fisher's model and represent regulatory hub genes with broad pleiotropic effects, influencing multiple biological processes but potentially facing evolutionary constraints due to their central role in gene networks.

In contrast, LFMM outliers, identified by incorporating environmental variables, had significantly lower gene network connectivity and were enriched in gene co-expression modules involved in primary metabolic pathways such as histidine biosynthesis, glycolysis, and pyruvate metabolism. This suggests these candidate genes may contribute to adaptation through modular genetic architectures rather than extensive pleiotropy [34]. These two distinct sets demonstrate that adaptation may take multiple avenues, even within the same populations of the same species.

Our results are consistent with those of Lotterhos *et al.* (2018), who examined how genetic architecture influences local adaptation to climate in lodgepole pine (*Pinus contorta*) [35]. Their

results support a modular pleiotropy hypothesis, suggesting that candidate locally selected genes play a significant role in specific environmental conditions but have little effect across environments [3,15]. For example, a gene might regulate multiple traits related to temperature adaptation (e.g., cold tolerance and seasonal growth timing) but have little or no effect on drought response or nutrient uptake. This modular genetic structure suggests that adaptation can be specialized, with different genes fine-tuned for particular environmental challenges rather than broadly shaping responses. Such modular architectures may be favored in complex environments where multiple selective pressures, such as temperature, precipitation, and soil composition, act independently. This makes it more advantageous for genes to affect specific traits peculiar to a particular environmental context than to influence many unrelated factors universally [15, 16].

The higher proportion of eQTLs in candidate adaptive genes compared to non-adaptive genes suggests that regulatory variation plays a significant role in the local adaptation of *Helianthus argophyllus*. This could be because adaptive genes are subject to more plastic gene expression control, potentially enhancing their ability to respond to environmental selection [36].

Interestingly, while PCAdapt outliers exhibited higher network connectivity, they were not significantly enriched in any co-expression modules. This suggests that although they function as regulatory hubs, they may not form distinct adaptive modules, a pattern potentially reflecting evolutionary constraints imposed by widespread pleiotropy [25]. In contrast, LFMM outliers were more likely to be involved in specific co-expression modules, reinforcing that environmental selection may act through functionally specialized genetic networks.

Our results also provide insights into the selective pressures acting on these candidate adaptive loci. Genetic differentiation (F_{ST}) was significantly higher for LFMM outliers than for PCAdapt and control genes. Still, these loci did not show significant differences in D_{xy} values from the control, suggesting that they may be undergoing selection within locally adapted populations without widespread divergence [38]. This pattern is consistent with a modular adaptation model, where environmental selection acts on functionally specialized loci without extensive genetic differentiation [15,35]. PCAdapt outliers exhibited lower F_{ST} and D_{xy} than the control, suggesting that these loci may be subject to purifying or balancing selection [15,35].

The patterns of diversity observed between PCAdapt and LFMM outliers in both populations suggest the different ways selection may act on these loci. The elevated θ_π and lower θ_W and the strong sweep signals observed in CLR and Fay & Wu's H in LFMM outliers in the north population suggest that selection may be acting on standing genetic variation. In such cases, a hard sweep can still arise if one beneficial allele confers a stronger advantage or if demographic constraints reduce haplotypic diversity, causing selection to favor a single variant [39, 40]. The coast population, on the other hand, shows reduced diversity at LFMM loci, indicative of a hard sweep where a new advantageous mutation has rapidly fixed. These contrasting patterns suggest that while both populations might adapt to similar selective pressures, the source and dynamics of the beneficial alleles differ, possibly due to variations in demographic history, gene flow, or the timing of selection events.

Gene ontology analysis revealed functional differences between adaptive loci identified by LFMM and PCAdapt. LFMM outliers were significantly enriched in primary metabolic pathways, suggesting that modular adaptation may be driven by genes involved in biochemical processes [38]. In contrast, PCAdapt outliers were enriched in pathways related to broader regulatory functions, such as signal transduction, defense response, and protein serine/threonine kinase activity. This is consistent with the role of pleiotropic regulatory genes in mediating the evolution of complex traits.

The distinction between the adaptive loci identified by PCAdapt and LFMM underscores a potential trade-off between pleiotropy and modularity in local adaptation. Due to their broad regulatory functions, highly pleiotropic loci, such as those identified by PCAdapt, may play a role in long-term evolutionary processes. In contrast, adaptive loci functioning within modular pathways may allow for more flexible and rapid responses to environmental selection. This is consistent with recent extensions of Fisher's geometric model, which propose that while pleiotropic loci can facilitate adaptation in populations far from their fitness optimum, modular architectures may be more advantageous for fine-scale adaptation, where selection acts on genes with specific metabolic or physiological roles, rather than broad regulatory influence [7,18].

While our study provides new insight into *H. argophyllus* adaptive variation, some limitations are to be considered. First, our transcriptomic data were collected at a single developmental time point (3 weeks), and our analyses might only show how *H. argophyllus* adapts during this growth period and would not reflect the entire trajectory of *H. argophyllus* adaptation. Secondly, the

differences we observe might be due to neutral processes, since our analysis approach only captured relevant genomic variation within genes.

Our study shows how important it is to think about both pleiotropy and modularity when studying local adaptation. Highly pleiotropic genes can act as central hubs, which may limit a species' ability to evolve. In contrast, modular genetic systems allow for adaptation by focusing on specific functions. Backed by genetic differentiation data, eQTL analysis, selective sweep signatures, and gene ontology enrichment, our findings demonstrate how gene networks and specialized functions shape local adaptation in *Helianthus argophyllus*. This research contributes to the ongoing discussion of how selection, pleiotropy, and modularity work together to drive evolution in complex organisms.

Conclusions

This study provides new insights into the relationship between pleiotropy and adaptation in a naturally diverging plant species. Our findings suggest that candidate locally adaptive loci in *Helianthus argophyllus* are under strong selection and more likely to regulate gene expression than non-adaptive loci. Our results also suggest the need to use multiple approaches when identifying potential loci under selection, as each method may identify unique variations. In our case, one method identified genes that are highly connected within gene co-expression networks, as well as other genes at the periphery of these networks. Local adaptation may be facilitated and hindered by pleiotropy, which likely depends on the distance of the populations from fitness optima and the consistency of those optima. Understanding this dynamic helps us understand

how genetic structures drive adaptation, with important implications for evolutionary biology, conservation, and plant breeding.

Acknowledgments

We thank Dr. Andrew Whitehead (UC Davis) for his valuable insights on trait evolvability. We are also grateful to the members of the Moyers Lab at UMass Boston for their support and contributions. The Unity Cluster hosted by UMass Amherst was used for all computational analyses.

Material and methods

Sample collection and RNA sequencing

Sample collection and sequencing are described in Renaut *et al.* 2013. Here, we include 19 populations representing two distinct subpopulations of *H. argophyllus*, identified by previous research [41]. Briefly, individual seeds representing unique populations of *H. argophyllus* collected by BTM in 2011 were grown for 3 weeks in a growth chamber with 12 hours of daylight at 22 °C [41]. All above-ground tissue was then frozen and stored at −80° until RNA was extracted using a modified TRIzol Reagent protocol. Libraries were prepared using mRNASeq (Illumina, San Diego, CA) approach and sequenced on a GAI (paired-end sequencing, 2 × 100 bp reads). The average number of reads per individual was 1864375, with an average read depth of 17.8.

Sequence processing

We ran FastQC/0.11.5 [42] on the RNA sequencing reads to assess quality and Trimmomatic/0.39 [43] to perform sliding window trimming with a 4-base window size. We removed reads with an average quality below 20 and reads shorter than 25 bases after trimming. We aligned reads to the Ha412HOv2 *H. annuus* reference genome (<https://www.sunflowergenome.org/assembly-data/>) [44] using STAR [45]. After alignment, we used Picard/2.23.3 [46] to add read group information and remove duplicated reads. We called variants on aligned reads using Freebayes/1.3.5 [27] and filtered the VCF file with vcftools 0.1.14 [47]. We applied the following filtering criteria: site quality ≥ 30 , genotype quality ≥ 20 , read coverage ≥ 6 , and missing data $< 20\%$.

Genetic diversity and population structure

We filtered genotype variants for minor allele frequency (MAF) ≥ 0.05 for these analyses. We estimated θ_w , θ_π , Tajima's D , and F_{ST} with ANGSD/0.935 [48] and used Pixy to estimate D_{xy} for all sampled individuals [49], and used NGSAdmix to identify admixture proportions [50]. We estimated θ_w , θ_π , and Tajima's D within each population. Watterson's θ (θ_w) is an estimator of population mutation rate and can be compared to nucleotide diversity (π or θ_π), a measure of within-population polymorphism, to identify genetic loci that have experienced non-neutral evolution or demographic changes (Watterson 1975, Tajima 1989). We estimated mean F_{ST} within 5000 kb sliding windows. We used the *H. annuus* reference as an outgroup when calculating Fay and Wu's H to estimate the proportion of derived alleles.

Selection outliers

We used the LFMM 2 and PCAdapt R packages to identify selection outliers at a q -value threshold of <0.05 . LFMM 2 is a genotype-environment association (GEA) method that incorporates more than one independent variable in a multivariate model. These variables can be climatic, methylation, or gene expression data [24]. We included 19 BIOCLIM variables in our LFMM 2 model to identify genotype-environment associations [51]. PCAdapt identifies locally adapted candidates after controlling for population structure. PCAdapt is optimized to handle admixed individuals, account for population divergence, handle large sequencing datasets, and has a relatively low FDR [22].

eQTL and eGenes

We counted the number of reads expressed for each gene using FeatureCounts in the Subread/1.6.2 package [52]. We normalized the gene count file generated by Subread for our eQTL analysis with the Deseq2 R package [53], which uses the median-ratio normalization method [54]. We used the MatrixEql R package to identify loci (eQTL) associated with changes in the expression of specific genes (eGenes) [55]. We increased our filter MAF threshold from 0.05 to 0.1 to reduce the occurrence of spurious associations [56]. To control for expression heterogeneity, we performed a principal component analysis on our expression data following [57] and selected the first two principal components that explained 92% of the variation in our expression data. We included the residuals from the regression analysis using the first two principal components in our eQTL model to control for the effect of population structure [57].

Selective sweeps

We subset the VCF file by the genomic locations of selection outlier SNPs or genes, eQTL, eGenes, and control genes for each population. Control genes were the expressed genes that were not identified by any of the other analyses. We identified putative selective sweeps using two approaches: Fay and Wu's H [58] and the Composite Likelihood Ratio (CLR) method using the SweepFinder2 package [59] (Vy and Kim, 2015). In estimating CLR, we subset each VCF file by chromosome and used all positions per chromosome to create grid files for the analysis.

Statistical comparisons

We compared values for population genetic estimators across selection outlier genes, eQTL, eGenes, and control genes using the Wilcoxon signed-rank test in R [60].

Gene coexpression networks

We used the WGCNA R package to construct gene coexpression networks [61]. We set a power threshold of six to get the adjacency matrix of the signed network using the 'softConnectivity' function. This helps to estimate the number of genes connected to each node in a network module [61,62]. We compared the log-transformed connectivity of node genes that were selection outliers, eQTLs, or eGenes to the connectivity of node genes in the transcriptome-wide background using the Wilcoxon signed-rank test [60]. We visualized gene coexpression modules using Cytoscape [63].

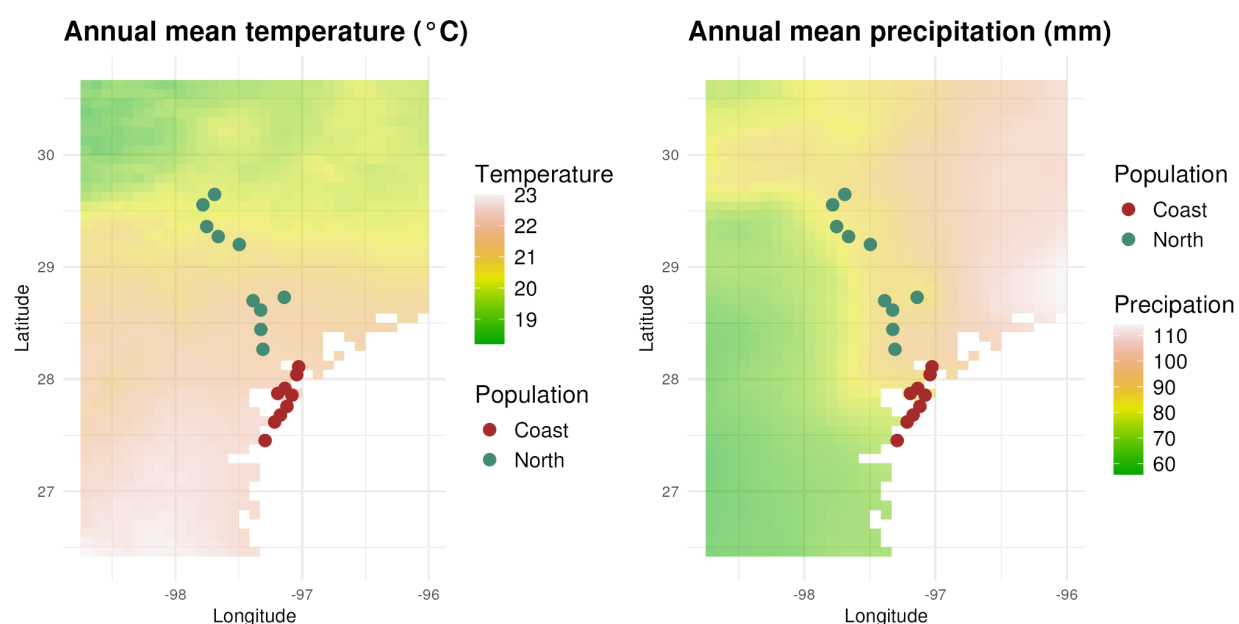
Annotation of genes and network modules

We used the clusterProfiler R package for gene and KEGG module overrepresentation analyses of selection outlier genes and coexpression network modules [64]. Since VCF was aligned to the

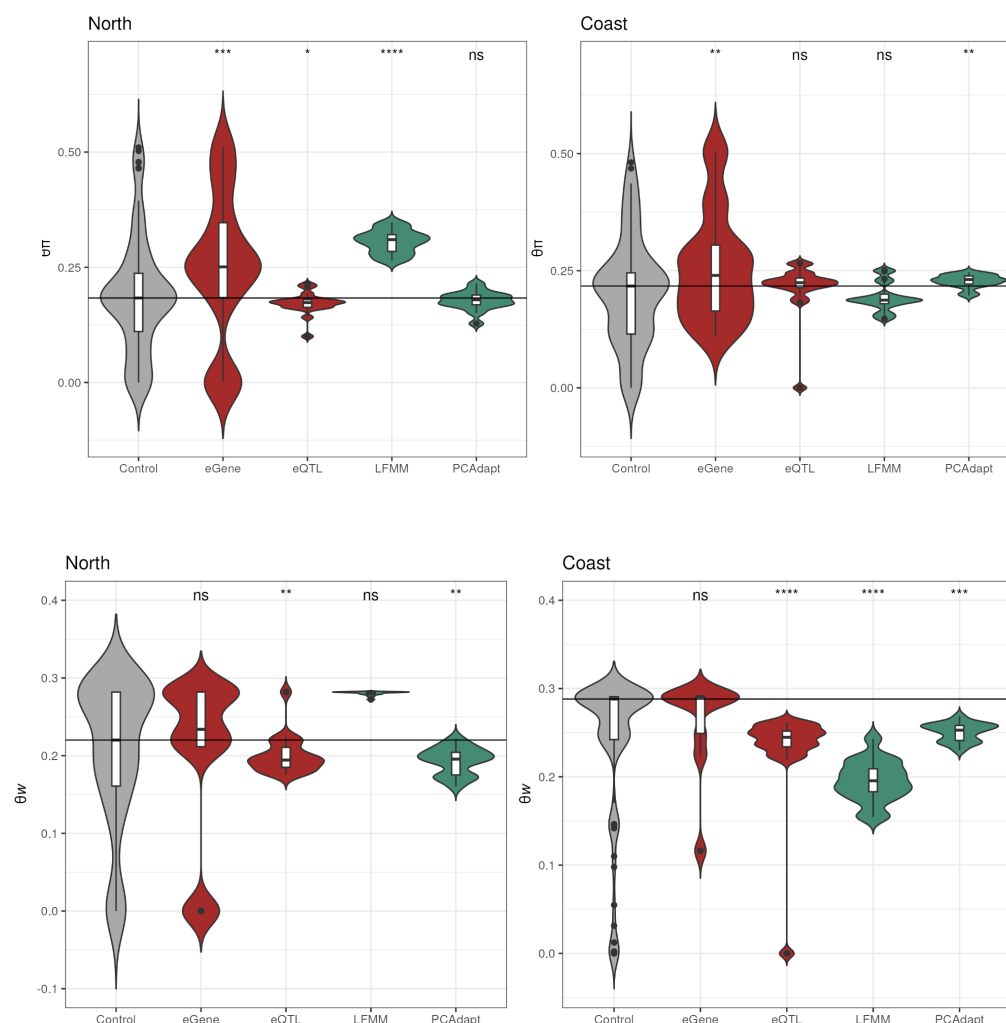
HA412-HO reference genome, but the HanXRQr2.0-SUNRISE genome is used in the KEGG database, we also aligned our transcriptome to the HanXRQr2.0-SUNRISE genome to extract NCBI gene IDs for each gene sequence using the NCBI BLAST command line tool [65]. We then used these gene IDs as queries in the gene overrepresentation analysis using the ‘enricher’ function in the clusterProfiler package. We used gene IDs as queries for the KEGG module analysis using the ‘enrichMKEGG’ function in clusterProfiler. All scripts can be accessed here: <https://github.com/Uzezi93/Pleiotropy-and-Local-Adaptation-Silverleaf-Sunflower>.

Supporting information

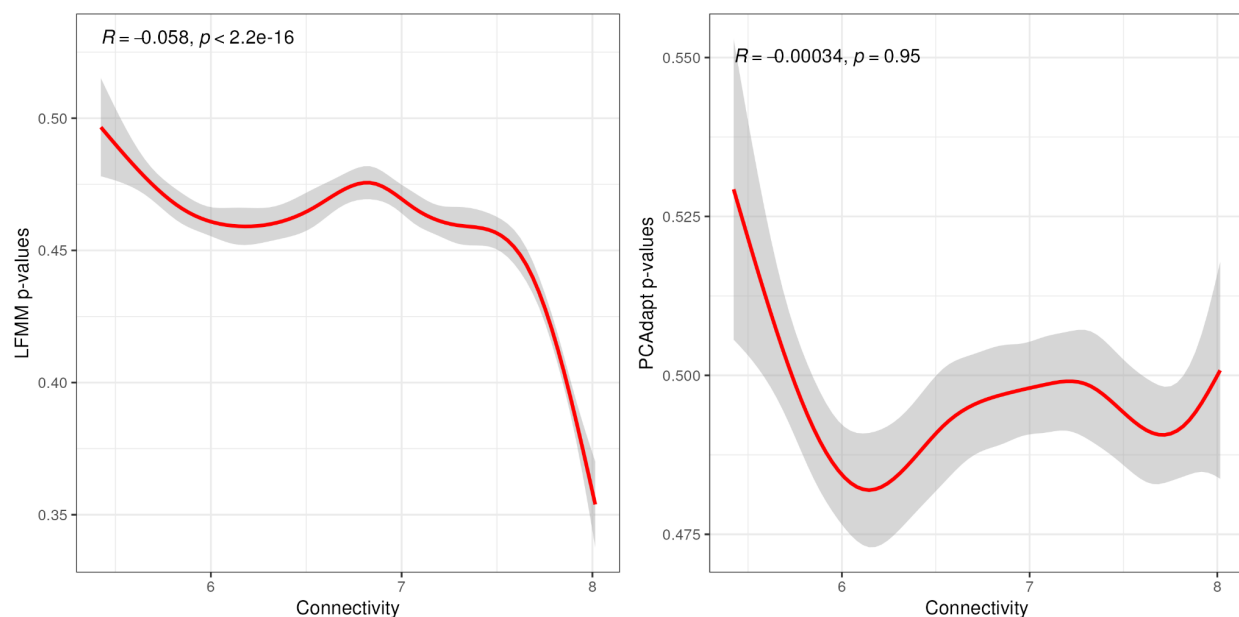
S1 Fig. Variation in temperature and precipitation in *H. argophyllus* sampling distribution



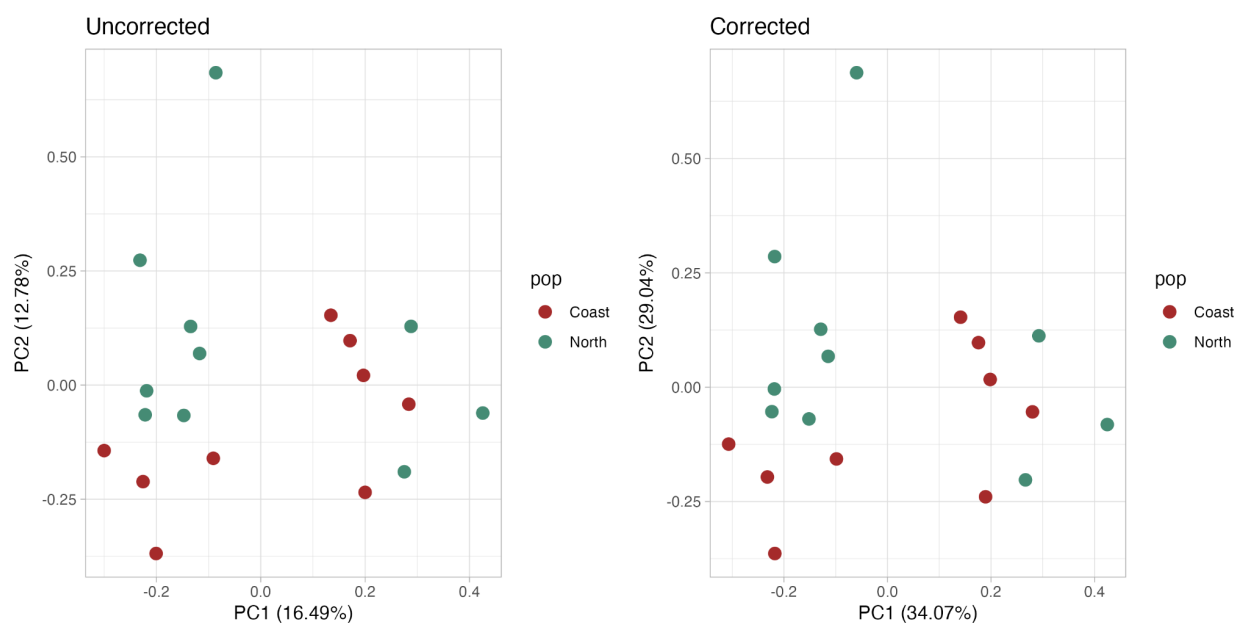
S2 Fig Genetic diversity in selection outliers in North and Coast Populations



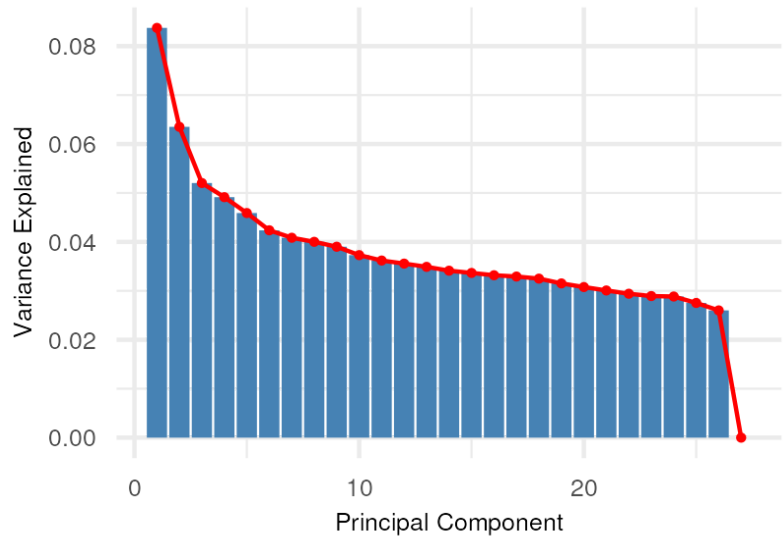
S3 Fig. Association between LFMM and PCAdapt connectivity and P -values at 95% CI



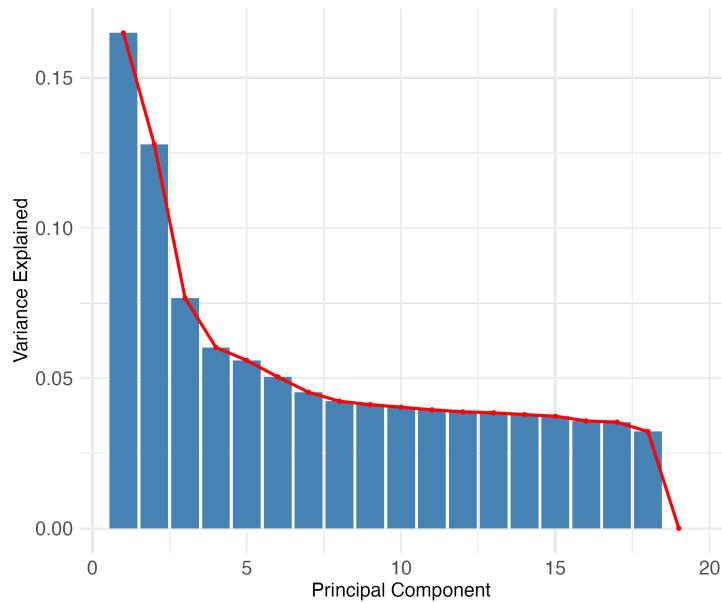
S4 Fig. Corrected and Uncorrected Expression-level PCA with first two PCs.



S5 Fig. Screeplot from genotype-based PCA



S6 Fig. Screeplot from expression-based PCA



S1 Table. Mean estimates of genetic diversity across all expressed genes in sampled populations.

	$\theta\pi$	θw	Tajima's D
--	-------------	------------	--------------

North	0.1937444	0.2076168	-0.1745136
Coast	0.2060006	0.2353111	-0.2985429

S2 Table. Proportion of Variance Explained (PVE) by the first six Principal Components (PC)

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	15.23	2.22	0.73	0.41	0.2	0.11
Proportion Explained	0.8	0.12	0.04	0.02	0.01	0.01
Cumulative Proportion	0.8	0.92	0.96	0.98	0.99	1

References

1. Tenaillon O. The Utility of Fisher’s Geometric Model in Evolutionary Genetics. *Annu Rev Ecol Evol Syst.* 2014;45: 179. doi:10.1146/ANNUREV-ECOLSYS-120213-091846
2. Stearns FW. One hundred years of pleiotropy: A retrospective. *Genetics.* 2010;186: 767–773. doi:10.1534/genetics.110.122549
3. Wagner GP, Zhang J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* 2011 12:3. 2011;12: 204–213. doi:10.1038/nrg2949
4. Hämälä T, Gorton AJ, Moeller DA, Tiffin P. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLoS Genet.* 2020;16: e1008707. doi:10.1371/JOURNAL.PGEN.1008707
5. Edwards AWF. The Genetical Theory of Natural Selection. *Genetics.* 2000;154: 1419–1426. doi:10.1093/GENETICS/154.4.1419
6. Fisher RA, Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930. doi:10.5962/bhl.title.27468
7. Orr HA. THE POPULATION GENETICS OF ADAPTATION: THE DISTRIBUTION OF FACTORS FIXED DURING ADAPTIVE EVOLUTION. *Evolution.* 1998;52: 935–949. doi:10.1111/J.1558-5646.1998.TB01823.X
8. Pascual Robles MA. The effects of a FLOWERING LOCUS T polymorphism on phenology, reproductive isolation and hurricane survival in the silverleaf sunflower. University of British Columbia. 2023.
9. Heiser CB. Hybridization in the Annual Sunflowers: *Helianthus annuus* x *H. argophyllus*. *Am Nat.* 1951;85: 65–72. doi:10.1086/281651
10. Charles B. Heiser JDMSSBCWCM Jr. THE NORTH AMERICAN SUNFLOWERS (*HELIANTHUS*). *Memoirs of the Torrey Botanical Club.* 1969; 1–218.
11. Gausman HW, Menges RM, Escobar DE, Everitt JH, Bowen RL. Pubescence Affects Spectra and Imagery of Silverleaf Sunflower (*Helianthus argophyllus*). *Weed Sci.* 1977;25: 437–440. doi:10.1017/S0043174500033816
12. Moyers BT. The landscape of divergence in silverleaf sunflowers. University of British Columbia. 2015.
13. Moyers BT, Rieseberg LH. Remarkable life history polymorphism may be evolving under divergent selection in the silverleaf sunflower. *Mol Ecol.* 2016;25: 3817–3830. doi:10.1111/MEC.13723
14. Todesco M, Owens GL, Bercovich N, Légaré JS, Soudi S, Burge DO, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature.* 2020;584: 602–607. doi:10.1038/s41586-020-2467-6
15. Kinsler G, Geiler-Samerotte K, Petrov DA. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *Elife.* 2020;9: e61271. doi:10.7554/eLife.61271

16. Kinsler G, Li Y, Sherlock G, Petrov DA. A shift from pleiotropic to modular adaptation revealed by a high-resolution two-step adaptive walk. *bioRxiv*. 2024. p. 2024.04.17.589938. doi:10.1101/2024.04.17.589938
17. Smith SD. Pleiotropy and the evolution of floral integration. *New Phytol*. 2016;209: 80–85. doi:10.1111/nph.13583
18. Hansen TF. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems*. 2003;69: 83–94. doi:10.1016/s0303-2647(02)00132-6
19. Gnan S, Priest A, Kover PX. The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* magic lines. *Genetics*. 2014;198: 1751–1758. doi:10.1534/GENETICS.114.170746/-/DC1
20. Osnato M, Cota I, Nebhnani P, Cereijo U, Pelaz S. Photoperiod Control of Plant Growth: Flowering Time Genes Beyond Flowering. *Front Plant Sci*. 2022;12: 805635. doi:10.3389/FPLS.2021.805635/BIBTEX
21. Ziebell AL, Barb JG, Sandhu S, Moyers BT, Sykes RW, Doeppke C, et al. Sunflower as a biofuels crop: An analysis of lignocellulosic chemical properties. *Biomass Bioenergy*. 2013;59: 208–217. doi:10.1016/J.BIOMBIOE.2013.06.009
22. Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour*. 2017;17: 67–77. doi:10.1111/1755-0998.12592
23. Wang J, Hu Z, Upadhyaya HD, Morris GP. Genomic signatures of seed mass adaptation to global precipitation gradients in sorghum. *Heredity*. 2019;124: 108–121. doi:10.1038/s41437-019-0249-4
24. Caye K, Jumentier B, Lepeule J, François O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Mol Biol Evol*. 2019;36: 852–860. doi:10.1093/MOLBEV/MSZ008
25. Mauro AA, Ghalambor CK. Trade-offs, pleiotropy, and shared molecular pathways: A unified view of constraints on adaptation. *Integr Comp Biol*. 2020;60: 332–347. doi:10.1093/icb/icaa056
26. Wagner G, Pavličev M, Cheverud J. The road to modularity. *Nat Rev Genet*. 2007;8: 921–931. doi:10.1038/nrg2267
27. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. Available: <https://github.com/freebayes/freebayes>
28. Cutter AD. A primer of molecular population genetics. London, England: Oxford University Press; 2019.
29. Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, et al. TAIR: a resource for integrated *Arabidopsis* data. *Funct Integr Genomics*. 2002;2: 239–253. doi:10.1007/s10142-002-0077-z
30. Orr HA. The genetic theory of adaptation: a brief history. *Nat Rev Genet*. 2005;6: 119–127. doi:10.1038/nrg1523
31. Guillaume F, Otto SP. Gene functional trade-offs and the evolution of pleiotropy. *Genetics*. 2012;192: 1389–1409. doi:10.1534/genetics.112.143214

32. Whiting JR, Booker TR, Rougeux C, Lind BM, Singh P, Lu M, et al. The genetic architecture of repeated local adaptation to climate in distantly related plants. *Nat Ecol Evol.* 2024;8: 1933–1947. doi:10.1038/s41559-024-02514-5
33. Battlay P, Yeaman S, Hodgins KA. Impacts of pleiotropy and migration on repeated genetic adaptation. *Genetics.* 2024. doi:10.1093/genetics/iyae111/7712980
34. Wang Z, Liao B-Y, Zhang J. Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci U S A.* 2010;107: 18034–18039. doi:10.1073/pnas.1004666107
35. Lotterhos KE, Yeaman S, Degner J, Aitken S, Hodgins KA. Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biol.* 2018;19: 157. doi:10.1186/s13059-018-1545-7
36. Hedrick P, McDonald J. Regulatory gene adaptation: An evolutionary model. *Heredity (Edinb).* 1980;45: 83–97. doi:10.1038/hdy.1980.52
37. Durkin SM, Ballinger MA, Nachman MW. Tissue-specific and cis-regulatory changes underlie parallel, adaptive gene expression evolution in house mice. *PLoS Genet.* 2024;20: e1010892. doi:10.1371/journal.pgen.1010892
38. Arnqvist G, Sayadi A. A possible genomic footprint of polygenic adaptation on population divergence in seed beetles? *Ecol Evol.* 2022;12: e9440. doi:10.1002/ece3.9440
39. Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. *Evolution.* 2005;59: 2312–2323. doi:10.1111/j.0014-3820.2005.tb00941.x
40. Booker TR, Jackson BC, Keightley PD. Detecting positive selection in the genome. *BMC Biol.* 2017;15. doi:10.1186/S12915-017-0434-Y
41. Moyers BT, Rieseberg LH. DIVERGENCE IN GENE EXPRESSION IS UNCOUPLED FROM DIVERGENCE IN CODING SEQUENCE IN A SECONDARILY WOODY SUNFLOWER. *Int J Plant Sci.* 174. doi:10.1086/671197
42. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30: 2114. doi:10.1093/BIOINFORMATICS/BTU170
44. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature.* 2017;546: 148–152. doi:10.1038/nature22380
45. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29: 15–21. doi:10.1093/BIOINFORMATICS/BTS635
46. GitHub - broadinstitute/picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. Available: <https://github.com/broadinstitute/picard>
47. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27: 2156–2158. doi:10.1093/BIOINFORMATICS/BTR330

48. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 2014;15: 1–13. doi:10.1186/S12859-014-0356-4/TABLES/4
49. Korunes KL, Samuk K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 2021;21: 1359–1368. doi:10.1111/1755-0998.13326
50. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 2013;195: 693–702. doi:10.1534/GENETICS.113.154138/-/DC1
51. Noce S, Caporaso L, Santini M. A new global dataset of bioclimatic indicators. *Scientific Data* 2020 7:1. 2020;7: 1–12. doi:10.1038/s41597-020-00726-5
52. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30: 923–930. doi:10.1093/BIOINFORMATICS/BTT656
53. Bioconductor - DESeq2. Available: <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>
54. Evans C, Hardin J, Stoebe DM. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief Bioinform*. 2018;19: 776–792. doi:10.1093/bib/bbx008
55. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28: 1353–1358. doi:10.1093/BIOINFORMATICS/BTS163
56. Wang T, Liu Y, Ruan J, Dong X, Wang Y, Peng J. A pipeline for RNA-seq based eQTL analysis with automated quality control procedures. *BMC Bioinformatics*. 2021;22: 403. doi:10.1186/s12859-021-04307-0
57. Tian L, Quitadamo A, Lin F, Shi X. Methods for population-based eQTL analysis in human genetics. *Tsinghua Sci Technol*. 2014;19: 624–634. doi:10.1109/TST.2014.6961031
58. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*. 2000;155: 1405–1413. doi:10.1093/genetics/155.3.1405
59. Degiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*. 2016;32: 1895–1897. doi:10.1093/BIOINFORMATICS/BTW051
60. Rey D, Neuhaus M. Wilcoxon-Signed-Rank Test. *International Encyclopedia of Statistical Science*. 2011; 1658–1659. doi:10.1007/978-3-642-04898-2_616
61. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9: 1–13. doi:10.1186/1471-2105-9-559/FIGURES/4
62. Zeng F, Shi M, Xiao H, Chi X. WGCNA-based identification of hub genes and key pathways involved in nonalcoholic fatty liver disease. *Biomed Res Int*. 2021;2021: 5633211. doi:10.1155/2021/5633211
63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003;13: 2498. doi:10.1101/GR.1239303

64. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16: 284–287. doi:10.1089/omi.2011.0118
65. > N. BLAST® Command Line Applications User Manual. National Center for Biotechnology Information (US); 2008. Available: <https://www.ncbi.nlm.nih.gov/books/NBK279690/>