Modularity of genes involved in local adaptation to climate despite physical linkage

**Authors**

Katie E. Lotterhos1,\*, Sam Yeaman2, Jon Degner3, Sally Aitken3, Kathryn A. Hodgins4

**Affiliations**

1Department of Marine and Environmental Sciences, Northeastern Marine Science Center, 430 Nahant Rd, Nahant, MA 01908

2Department of Biological Sciences, University of Calgary, AB, T2N1N4

3Department of Forest and Conservation Sciences, Faculty of Forestry, Vancouver, BC V6T 1Z4 Canada

4School of Biological Sciences, Monash University, Wellington Rd, Clayton VIC 3800

\* Corresponding Author, [k.lotterhos@neu.edu](mailto:k.lotterhos@neu.edu)

**Keywords**: landscape genomics, genetic-environment associations, genome-wide associations (GWAS), conifers, linkage disequilibrium, ion antiporters, auxin biosynthesis, flowering time

**Running title**: Modularity of adaptation despite physical linkage

Abstract

**Background:** Physical linkage among genes shaped by different sources of selection is a fundamental aspect of genetic architecture. Theory predicts that evolution in complex environments selects for modular genetic architectures and high recombination rates among loci shaped by different sources of selection. However, limited data exist to test these predictions because the field lacks consensus for how to control for intercorrelated environmental variables. Here, we introduce a co-association network analysis, which clusters loci based on differing environmental associations, and use it to study the genetic architecture of local adaptation to climate in lodgepole pine (*Pinus contorta*).

**Results:** We identified many modules of genes associated with distinct environments (aridity, freezing, geography), and discovered low recombination rates among some candidate genes in different modules. We observed limited evidence for environmental pleiotropic effects on distinct aspects of climate. We also found limited correspondence between the modularity of co-association networks and gene regulatory networks. We compared co-association networks to associations with principal components, and found the latter can lead to misinterpretation. Finally, we used simulations to illustrate the benefits and caveats of co-association networks.

**Conclusions**: Co-association networks provided a useful framework for studying modularity. Our results supported the prediction that evolution to complex environments selects for modular genetic architectures, but some of our results went against the prediction that selection would increase recombination among loci experiencing different sources of selection.  These results give new insight into evolutionary debates about the extent of modularity and pleiotropy in the evolution of genetic architectures.

Background

Pleiotropy and linkage are fundamental aspects of genetic architecture [[1]](https://paperpile.com/c/vNqLuE/2ocb). Genes that have effects on seemingly unrelated traits may influence the rate of adaptation [[2, 3]](https://paperpile.com/c/vNqLuE/FdfN+siUQ), and linkage among genes experiencing different kinds of selection can facilitate or hinder adaptation [[4–6]](https://paperpile.com/c/vNqLuE/9sr1+KcDn+d75E). Despite progress in understanding the underlying pleiotropic nature of phenotypes and the influence of pleiotropy on the rate of adaptation to specific conditions [[7]](https://paperpile.com/c/vNqLuE/nmjU), we have an incomplete understanding of the extent and magnitude of linkage and pleiotropy in the local adaptation of natural populations to complex environments.

Here, we aim to characterize the number of separate components of the environment in which a gene affects fitness (a form of “selectional pleiotropy”, Table 1)[[8]](https://paperpile.com/c/vNqLuE/M4Tc). A major hurdle in achieving this characterization is defining the environment by the action of selection and not by the intrinsic attributes of the organism or by the environmental variables we happen to measure. In local adaptation to climate, an allele that has different effects on fitness at different extremes of an environmental variable (e.g., positive effects on fitness in cold environments and negative effects in warm environments, often called “antagonistic pleiotropy”, Table 1[[9]](https://paperpile.com/c/vNqLuE/nGPC)) will evolve to have a clinal relationship between the allele frequency and environment [[10–15]](https://paperpile.com/c/vNqLuE/NNGV+sRrP+xgno+3Pai+qpAW+yp41). While associations between allele frequencies and environments have been well characterized across many taxa [[16]](https://paperpile.com/c/vNqLuE/Ite8), whether genes affect fitness in multiple distinct aspects of the environment, which we call “environmental pleiotropy” (e.g., has effects on fitness in both cold and dry environments, , Table 1), has not been well characterized  [[17]](https://paperpile.com/c/vNqLuE/Tmmz). This is because of conceptual issues that arise from defining environments along the univariate axes that we measure, because “cold” and “dry” might be a single selective optimum (“cold-dry”) to which a gene adapts [[7]](https://paperpile.com/c/vNqLuE/nmjU). Moreover, climate variables such as temperature and precipitation are highly correlated across landscapes, and this correlation structure makes inferring pleiotropy from signals of selection to climate difficult. Indeed, in their study of adaptation of *Arabidopsis* to climate, Hancock et al. [[17]](https://paperpile.com/c/vNqLuE/Tmmz) noticed that candidate loci showed signals of selection in multiple environmental variables, potentially indicating pleiotropic effects. However, they also found that a substantial proportion of this overlap was due to correlations among climate variables on the landscape, and as a result they were unable to fully describe pleiotropic effects.

Because of the conceptual issues described above, certain aspects of genetic architecture have not been well characterized in adaptation to multivariate environments: particularly modularity (i.e., when mutations affect traits that are part of the same functional complex, Table 1) and recombination rates among genes in different modules (i.e., physical proximity in the genome). These aspects of genetic architecture are important to characterize in order to test the theoretical predictions described below, and to inform the considerable debate about whether organisms have a modular organization of gene effects on phenotypes (or aspects of fitness) versus universal effects of genes on all phenotypes (or aspects of fitness) [[18–24]](https://paperpile.com/c/vNqLuE/htcT+k0Rn+BRvk+FI10+SGJc+ayb9+Tkxk).

Modular genetic architectures are characterized by extensive pleiotropic effects among elements within a module, and a suppression of pleiotropic effects between different modules [[25]](https://paperpile.com/c/vNqLuE/6Zft).  Modular genetic architectures are predicted to be favored by theory when genomes face complex spatial and temporal environments [[26]](https://paperpile.com/c/vNqLuE/hmua) or when multiple traits are under a combination of directional and stabilizing selection (because modularity allows adaptation to take place in one trait without undoing the adaptation achieved by another trait [[25, 27]](https://paperpile.com/c/vNqLuE/6Zft+nGKW). Adaptation to climate on a landscape fits these criteria because environmental variation among populations is complex - with multiple abiotic and biotic challenges occurring at different spatial scales - and traits are thought to be under stabilizing selection within populations but directional selection among populations [[28]](https://paperpile.com/c/vNqLuE/N75Q).

Clusters of physically linked loci subject to the same selective environment and a lack of physical linkage among loci subject to different selection pressures are expected based on theory. When mutations are subject to the same selection pressure, recombination can bring variants with similar effects together and allow evolution to proceed faster [[29]](https://paperpile.com/c/vNqLuE/MsZ2). Clusters of adaptive loci can also arise through genomic rearrangements that bring existing mutations together [[30]](https://paperpile.com/c/vNqLuE/ONV3), or because new causal mutations linked to adaptive alleles have an increased establishment probability [[31]](https://paperpile.com/c/vNqLuE/IiAA). Similarly, clusters of locally adaptive loci are expected to evolve in regions of low recombination, such as inversions, because of the reduced gene flow these regions experience [[32, 33]](https://paperpile.com/c/vNqLuE/N8fZ+HvzF). In general, these linked clusters of adaptive loci are favored over evolutionary time because low recombination rates increase the rate at which they are inherited together. Conversely, selection will also act to disfavour linkage and increase recombination rates between genes adapting to different selection pressures [[34–36]](https://paperpile.com/c/vNqLuE/y6Th+cSvc+1CYn). Thus, genes adapting to different selection pressures would be unlikely to be physically linked or to have low recombination rates between them. In practice, issues can arise in inference because physical linkage will cause correlated responses to selection in neutral loci flanking a causal locus. Large regions of the genome can share similar patterns of association to a given environmental factor, such that many loci within a given candidate region are probably not causally responding to selection. Conversely, if linked genes are associated with completely different aspects of the selective environment, this is unlikely to arise by chance.

In summary, current analytical techniques have given limited insight into the genetic architectures of adaptation to the multivariate environmental drivers of selection. Characterizing the different aspects of the multivariate environment that act on genomes is difficult because measured variables are univariate and may not be representative of selection from the perspective of the organism, and because of spatial correlations among environmental variables. Even when many variables are summarized with ordination such as principal components that are orthogonal, the axes that explain the most variation in physical environment don’t necessarily correspond to the axes that cause selection [[37]](https://paperpile.com/c/vNqLuE/KF5g). Secondly, the statistical methods widely used for inferring adaptation to climate are also univariate in the sense that they test for significant correlations between frequency of a single allele and a single environmental variable [[e.g., 38, 39, 40]](https://paperpile.com/c/vNqLuE/vwhx+1IaM+Eiy0/?prefix=e.g.%2C%20,,). While some multivariate regression methods like redundancy analysis have been used to understand how multiple environmental factors shape genetic structure [[41, 42]](https://paperpile.com/c/vNqLuE/WHAY+p6hC), they still rely on ordination and have not been used to identify distinct evolutionary modules of loci.

Here, we aim fill this gap by presenting a framework for characterizing the genetic architecture of adaptation to the multivariate environment, through the joint inference of modules of loci that associate with distinct aspects of the selectional environment that we call “environmental response modules” (Table 1, Figure 1), as well as the distinct aspects of the environment to which they associate. Using this framework, we can characterize some aspects of genetic architecture, including modularity and linkage, that have not been well studied in the adaptation of genomes to complex environmental landscapes. This framework is illustrated in Figure 1 for four hypothetical genes adapted to two distinct aspects of climate (freezing and aridity). In this figure we compare the patterns expected for a modular architecture (left column, where pleiotropic effects of a gene are limited to adaptation to one particular aspect of climate) to a highly environmentally pleiotropic architecture (right column, where genes have pleiotropic effects on adaptation to multiple distinct aspects of climate). Candidate SNPs are first identified by the significance of the associations between allele frequency and the measured environmental variables, evaluated against what would be expected by neutrality. Then, hierarchical clustering of candidate-SNP allele associations with environments is used to identify environmental response modules (Figure 1B) [[43–45]](https://paperpile.com/c/vNqLuE/oPtk+KFAA+AVgs). These modules can be visualized with a co-association network analysis, which identifies groups of loci that may covary with one environmental variable but covary in different ways with another, revealing patterns that are not evident through univariate analysis (Figure 1C). By defining the distinct aspects of the selectional environment (Table 1) for each module through their environmental associations, we can infer pleiotropic effects of genes through the associations their SNPs have with distinct selective environments (Figure 1D). In this approach, the genetic effects of loci on different traits (e.g., developmental/functional pleiotropy or modules) under selection are unknown, and we assume that each aspect of the multivariate environment selects for a trait or suite of traits that can be inferred by connecting candidate loci directly to the aspects of the environment that select for particular allele combinations.

We apply this new approach to characterize the genetic architecture of local adaptation to climate in lodgepole pine (*Pinus contorta*) using a previously published exome capture dataset [[46–48]](https://paperpile.com/c/vNqLuE/UhqG+HGG2+val2) from trees that inhabit a wide range of environments across their range, including freezing temperatures, precipitation, and aridity [[49–52]](https://paperpile.com/c/vNqLuE/bVD2+JkoP+vUkI+yaFi). Lodgepole pine is a coniferous species inhabiting a wide range of environments in northwestern North America and exhibits isolation by distance population structure across the range [[46]](https://paperpile.com/c/vNqLuE/UhqG). Previous work based on reciprocal transplants and common garden experiments has shown extensive local adaptation [[46, 53, 54]](https://paperpile.com/c/vNqLuE/cZUs+UhqG+FBGF). We recently used this dataset to study convergent adaptation to freezing between lodgepole pine and the interior spruce complex (*Picea glauca* x *Picea engelmannii*) [[46–48]](https://paperpile.com/c/vNqLuE/UhqG+HGG2+val2). However, the comparative approach was limited to discovering parallel patterns between species, and did not examine selective factors unique to one species. As in most other systems, the genomic architecture in pine underlying local adaptation to the multivariate environment has not been well characterized, and our reanalysis yields several new biological insights overlooked by the comparative approach.

We evaluated the benefits and caveats of this new framework by comparing it with other multivariate approaches (based on principal components) and by evaluating it with simulated data. The evaluation with simulations yielded several important insights, including the importance of using strict criteria to exclude loci with false positive associations with the environment.  Thus, a key starting point for inferring environmental response modules is a good set of candidate SNPs for adaptation. We developed this candidate set by first identifying top candidate genes for local adaptation (from a previously published set of genes that contained more outliers for genotype-environment associations and genotype-phenotype associations than expected by chance, [[46]](https://paperpile.com/c/vNqLuE/UhqG)). We then identified top candidate SNPs within these top candidate genes as those whose allele frequencies were associated with at least one environmental variable above that expected by neutrality (using a criterion that excluded false positives in the simulated data described below). To this set of top candidate SNPs, we applied the framework outlined in Figure 1 to characterize environmental modularity and linkage of the genetic architecture. The power of our dataset comes from including a large number of populations inhabiting diverse environments (>250), the accurate characterization of climate for each individual with 22 environmental variables, a high-quality exome capture dataset representing more than 500,000 single-nucleotide polymorphisms (SNPs) in ~29,000 genes [[46–48]](https://paperpile.com/c/vNqLuE/UhqG+HGG2+val2), a mapping population that allows us to study recombination rates among genes, and an outgroup species that allowed us to determine the derived allele for most candidate SNPs. When such data is available, we find that this framework is useful for characterizing the environmental modularity and linkage relationships among candidate genes for local adaptation to multivariate environments.

Results

*Top candidate genes and top candidates SNPs*

The study of environmental pleiotropy and modularity is relevant only to loci under selection. In this study we identified a SNP as a top candidate based on whether (i) it was located within a top-candidate gene, and (ii) its allele frequency was associated with at least one environmental variable above and beyond what may be expected for neutrality. Our “top candidate” approach identified a total of 117 candidate genes out of a total of 29,920 genes. These contigs contained 801 top-candidate SNPs (out of 585,270 exome SNPs) that were strongly associated with at least one environmental variable and were likely either causal or tightly linked to a causal locus. This set of top candidate SNPs was enriched for *XTX* outliers (Supplemental Figure 1:  *XTX* is an analog of *FST* that measures differentiation in allele frequencies across populations). To elucidate patterns of multivariate association, we applied the framework described in Figure 1 to these 801 top candidate SNPs.

*Environmental response modules*

Hierarchical clustering and co-association network analysis of top candidate SNPs revealed a large number of environmental response modules, each of which may contain SNPs from one or more genes. For the purposes of presentation, we grouped SNPs into 4 main groups, each with several environmental response modules, classified according to the kinds of environmental variables they were most strongly associated with: Aridity, Freezing, Geography, and an assorted group we bin as “Multi” (Figure 2A, B).  While we use these four main environmental factor groupings for the purpose of illustration, it is the underlying clustering of the SNPs revealed by co-association networks that is relevant to the study of modularity (Figure 2B-F). In the co-association networks, top candidate SNPs are represented by nodes, and SNPs are connected by edges based on their similarity in associations with the environment. We chose a threshold for drawing edges based on simulations that showed selected loci adapting to the same environmental variable would be more likely to be connected and false positive neutral loci would be more likely to be unconnected in the network (see *Results: Simulated datasets*). This division of data into groups was necessary to produce coherent visual network plots and to make data analysis more computationally efficient (we found when there were more than ~20,000 edges in the data, computation and plotting of the network were not feasible with the package). Because SNPs in the 4 groups are more dissimilar to SNPs in other groups than the criteria we use to calculate modules, they would not be connected by edges in a co-association network. Interestingly, this clustering by association signatures does not closely parallel the correlation structure among environmental variables themselves. For example, temperature difference (TD), degree-days below 0 (DD\_0), and latitude (LAT) are all relatively strongly correlated (> 0.5), but the “Freezing” SNPs are strongly correlated with TD and DD\_0 but not LAT (Figure 2A, 2B).

The environmental response modules are shown in Figures 2C-F. Each connected network of SNPs can be considered a group of loci that shows associations with a distinct aspect of the multivariate environment.  The “Multi” group stands for multiple environments because these SNPs showed associations with 19 to 21 of the 22 environmental variables. This group consisted of 60 top candidate SNPs across just 3 genes and undirected graph networks revealed 2 environmental response modules within this group (Figure 2C, G, Supplementary Figure 2). The “Aridity” group consisted of 282 SNPs across 28 genes and showed associations with climate moisture deficit, annual heat:moisture index, mean summer precipitation, and temperature variables excluding those that were frost-related (Figure 2B). All these SNPs were very similar in their patterns of association and grouped into a single network module (Figure 2D, G, Supplementary Figure 3). The “Freezing” group consisted of 176 SNPs across 21 genes and showed associations with freezing variables including number of degree-days below 0oC, mean coldest month temperature, and variables related to frost occurrence (Figure 2B, G). SNPs from eight of the genes in this group formed a single module (genes #35-42), with the remaining SNPs mainly clustering by gene (Figure 2E, G Supplementary Figure 4). The final group, “Geography,” consisted of 282 SNPs across 28 genes that showed consistent associations with the geographical variables elevation and longitude, but variable associations with other climate variables (Figure 1B). This was a loosely connected network consisting of several submodules containing 1 to 9 genes (Figure 1F, G, Supplementary Figure 6). Network analysis using structure-corrected associations between allele frequency and the environmental variables resulted in broadly similar patterns, although the magnitude of the correlations was reduced (Supplemental Figure 6).

The pleiotropy barplot is visualized in Figure 1G, where each gene is listed along the x-axis, the bar color indicates the environmental response module, and the bar height indicates the number of SNPs clustering with that module. If each environmental response module is considered to associate with a distinct aspect of the multivariate environment, then genes whose SNPs associate with different environmental response modules (e.g., genes with different colors in their bars in Figure 1G) might be considered to be environmentally pleiotropic. However, conceptual issues remain in inferring the extent of pleiotropy, because environmental response modules within the Geography group, for instance, will be more similar to each other in their associations with environments than between a module in the Geography group and a module in the Multi group. For this reason, we are only inferring that our results are evidence of environmental pleiotropy when genes have SNPs in at least 2 of the 4 major groups in the data. For instance, gene #1, whose majority of SNPs cluster with the Multi group, also has 8 SNPs that cluster with the Freezing group (but are not located in environmental response modules with any genes defined by Freezing).  In the Aridity group, gene #11 has three SNPs that also cluster with the Geography group (but are not located in environmental response modules with any genes defined by Geography). In the Freezing group, some genes located within the same environmental response module (#35-40) also have SNPs that cluster with another environment response module in the Geography group (with genes #75-76; these are not physically linked to genes 35-37, see below). Whether or not these are “true” instances of environmental pleiotropy remains to be determined by experiments. For the most part, however, the large majority of SNPs located within genes map to a specific environmental response module or to modules located within one of the four main groups, so environment-associated pleiotropy at the gene-level appears to be generally quite limited.

*Statistical and physical linkage disequilibrium*

To determine if grouping of SNPs into environmental response modules correspond to associations driven by statistical linkage disequilibrium (LD), we calculated mean LD among all SNPs in the top candidate genes (as the correlation in allele frequencies) and found that the co-association network visualization captured patterns of statistical LD among the genes through their common associations with environmental variables (Supplementary Figure S7). There was higher than average statistical LD within the environmental response modules of the Multi, Aridity, and Freezing groups, and very low statistical LD between the Aridity group and the other groups (Supplementary Figure S7). The statistical LD among the other three groups (Multi, Freezing, and Geography) was small, but higher with each other than with Aridity. Thus, co-association networks capture the same information as simple LD-based clustering with the important additional benefit of linking LD clusters to likely environmental drivers of selection.

The high statistical LD observed within the four main climate modules could arise via selection by the same aspect of the multivariate environment, via physical linkage on the chromosome, or both. We used a mapping population to disentangle these two hypotheses, by calculating recombination rates among the top candidate genes (see *Methods: Recombination rates*). Of the 117 top candidate genes, 66 had SNPs that were represented in our mapping population. The recombination data revealed that all the genes in the Aridity group have strong LD and are physically linked (Figure 3). Within the other three groups, we found physical proximity for only a few genes within the same environmental response module (but note that our mapping analysis does not have high power to infer recombination rate when loci are physically unlinked; see *Methods*). Other notable exceptions were a few environmental response modules in the Geography group (comprised of genes #53-54, #60-63, or #75-76), which also had very low recombination rates among them. Of the three genes the largest environmental response module in the Freezing Group represented in our mapping panel (#35-37), two of them were physically linked to each other.

Strikingly, low recombination rates were estimated between some genes belonging to different environmental response modules across the four main groups, even though there was little statistical LD among SNPs in these genes (Figure 2). This included a large LD block comprised of genes from all 4 groups: 8 genes from the Aridity environmental response module, 1 gene from the large module in the Multi group, 2 genes from different environmental response modules in the Freezing group, and 7 genes from different environmental response modules in the Geography group (see Supplementary Figure S8 for a reorganization of the recombination data and more intuitive visualization).

*Comparison to conclusions based on principal components of environments*

We compared the results from the co-association network analysis to associations with principal components (PC) of the environmental variables. Briefly, all environmental variables were input into a PC analysis (Supplementary Figure S9), and associations between allele frequencies and PC axes were analyzed. We used the same criteria (log10 BF > 2 in bayenv2) to determine if a locus was a significant outlier and compared (i) overlap with top candidate SNPs based on outliers from univariate associations with environments, and (ii) interpretation of the selective environment based on loadings of environments to PC axes. The first three PC axes explained 44% (PC1), 22% (PC2), and 15% (PC3) of the variance in environments (80% total). Overall, 80% of the geography SNPs, 75% of the Freezing SNPs, 20% of the Aridity SNPs, and 10% of the Multi SNPs were *not* outliers along the first 10 PC axes and would have been missed by a study based on PC axes.

Next, we evaluated whether interpretation of selective environment based on PCs was consistent with that based on associations with environments. Some of the temperature and frost variables (MAT: mean annual temperature, EMT: extreme minimum temperature, DD0: degree days below 0C, DD5: degree days above 5C, bFFP: begin frost-free period, FFP: frost free period, eFFP: end frost free period, labels in Figure 1A) had the highest loadings for PC1 (Supplementary Figure S9). Almost all of the SNPs in the Multi group (90%) and 19% of SNPs in the Freezing group were outliers along this axis (Supplementary Figure 10, less than 2% of candidate SNPs in the other groups were outliers). For PC1, interpretation of the selective environment (e.g., MAT, DD0, FFP, eFFP, DD5) is somewhat consistent with the co-association network analysis (both Multi SNPs and Freezing SNPs show associations with all these variables, Figure 1B). However, the Multi SNPs and Freezing SNPs had strong associations with other variables (e.g., Multi SNPs showed strong associations with Latitude and Freezing SNPs showed strong associations with longitude, Figure 1B) that did not load strongly onto this axis, and would have been missed in an interpretation based on associations with principal components.

Many precipitation and aridity variables loaded strongly onto PC2, including mean annual precipitation, annual heat:moisture index, climate moisture deficit, and precipitation as snow (Supplementary Figure 9). However, few top candidate SNPs were outliers along this PC axis: only 13% of Freezing SNPs, 10% of Aridity SNPs, and less than 3% of Multi or Geography SNPs were outliers (Supplementary Figure 10).

For PC3, latitude, elevation, and two frost variables (beginning frost-free period and frost-free period) had the highest loadings (Supplementary Figure 9). The majority (78%) of the Aridity SNPs were outliers in PC3 (Supplementary Figure 10). Based on the PC association, this would lead one to conclude that the Aridity SNPs show associations with latitude, elevation, and frost-free period. While the Aridity SNPs do have strong associations with latitude (5th row in Figure 1B), they show very weak associations with the beginning of frost-free period, elevation, and frost-free period (3rd, 4th, and last row in Figure 1B, respectively). Thus, interpretation of the environmental drivers of selection based on associations with PC3 would have been very different from the univariate associations.

*Interpretation of multivariate allele associations*

While the network visualization gave insight into patterns of LD among loci, it does not give insight into the patterns of allele frequency change on the landscape, relative to the ancestral state. As illustrated above, principal components would not be useful for this latter visualization. Instead, we accomplished this by plotting the association of a derived allele with one environmental variable against the association of that allele with a second environmental variable. Note that when the two environmental variables themselves are correlated on the landscape, an allele with a larger association in one environment will also have a larger association with a second environment, regardless of whether or not selection is shaping those associations.  We can visualize (i) the expected genome-wide covariance using shading of quadrants and (ii) the observed genome-wide covariance using a 95% prediction ellipse (Figure 4). Since alleles were coded according to their putative ancestral state in loblolly pine (*Pinus taeda*), the location of any particular SNP in the plot represents the bivariate environment in which the derived allele is found in higher frequency than the ancestral allele (Figure 4). Visualizing the data in this way allows us to understand the underlying correlation structure of the data, as well as to develop testable hypotheses about the true selective environment and the fitness of the derived allele relative to the ancestral allele.

We overlaid the top candidate SNPs, colored according to their grouping in the co-association network analysis, on top of this genome-wide pattern (for the 668 of 801 top candidates for which the derived allele could be determined). We call these plots “galaxy biplots” because of the characteristic patterns we observed when visualizing data this way (Figure 5). Galaxy biplots revealed that SNPs in the Aridity group showed associations with hot/dry versus cold/wet environments, while SNPs in the Multi and Freezing groups showed patterns of associations with hot/wet versus cold/dry environments (Figure 5Aa). These outlier patterns became visually more extreme for some SNPs and environments after correcting the associations for population structure (Figure 5B). Most SNPs in the Freezing group showed associations with elevation but not latitude (Figure 5 C,E). Conversely, the large environmental response module in the Multi group (gene #1) showed associations with latitude but not elevation, while the second environmental response module in the Multi group (genes #2-3) showed associations with both latitude and elevation (Figure 5C, E). Note how the structure correction polarized these patterns somewhat without changing interpretation, suggesting that the structure-corrected allelic associations become more extreme when their pattern of allele frequency went against the background population structure (Figure 5D, F).

Some modules were particularly defined by the fact that almost all the derived alleles changed frequency in the same direction (e.g., sweep-like signatures). For instance, for the environmental response module in the Multi group defined by genes #2-3, 14 of the 16 derived SNPs were found in higher frequencies at colder temperatures, higher elevations, and higher latitudes. Contrast this with a group of SNPs from an environmental response module in the Freezing group defined by gene #32, in which 14 of 15 derived SNPs were found in higher frequencies in warmer temperatures and lower elevations, but showed no associations with latitude. These may be candidates for genotypes that have arisen in frequency to adapt to particular environmental conditions on the landscape.

Conversely, other modules showed different combinations of derived alleles that arose in frequency at opposite values of environmental variables. For instance, derived alleles in the Aridity environmental response module were found in higher frequency in either warm, dry environments (88 of 155 SNPs) or in cold, moist environments (67 of 155 SNPs). Similarly, for the Multi environmental response module defined by gene #1, derived alleles were found in higher frequency in either cold, dry environments (15 of 37 SNPs) or in warm, moist environments (22 of 37 SNPs). These may be candidates for genes acted on by antagonistic pleiotropy within a locus (Table 1), in which one genotype is selected for at one extreme of the environment and another genotype is selected for at the other extreme of the environment. Unfortunately, we were unable to fully characterize the relative abundance of sweep-like vs. antagonistically pleiotropic patterns across all top candidate genes due to (i) the low number of candidate SNPs for most genes, and (ii) for many SNPs the derived allele could not be determined (because there was a SNP or missing data in the ancestral species).

We also visualized the patterns of allele frequency on the landscape for two representative SNPs, chosen because they had the highest number of connections in their environmental response module (and were more likely to be true positives, see *Results: Simulated datasets*). Geographic and climatic patterns are illustrated with examples for two such SNPs: (i) a SNP in the Multi environmental response module defined by gene #1 is shown in Figure 6A (with significant associations with latitude and mean annual temperature), and (ii) a SNP in the Aridity environmental response module (Figure 6B, gene #8 from Figure 1, with significant associations with annual heat:moisture index and latitude). These landscapes show the complex environments that may be selecting for particular combinations of genotypes despite potentially high gene flow in this widespread species.

*Candidate gene annotations*

Although many of the candidate genes were not annotated, as is typical for conifers, the genes underlying adaptation to these environmental gradients had diverse putative functions. The top candidate SNPs were found in 3’ and 5’ untranslated regions and open reading frames in higher proportions than all exome SNPs (Supplemental Figure S11). A gene ontology (GO) analysis using previously assigned gene annotations [[46, 55]](https://paperpile.com/c/vNqLuE/UhqG+3Dcq) found that a single molecular function, solute:cation antiporter activity, was over-represented across all top candidate genes (Supplemental Table S1). In the Aridity and Geography groups, annotated genes included sodium or potassium ion antiporters (one in Aridity, a KEA4 homolog, and two in Geography, NHX8 and SOS1 homologs), suggestive of a role in drought, salt or freezing tolerance [[56]](https://paperpile.com/c/vNqLuE/RNnp). Genes putatively involved in auxin biosynthesis were also identified in the Aridity (YUCCA 3) and Geography (Anthranilate synthase component) groups (Supplemental Table S2), suggestive of a role in plant growth. In the Freezing and Geography groups, several flowering time genes were identified [[57]](https://paperpile.com/c/vNqLuE/GCnW) including a homolog of CONSTANS [[58]](https://paperpile.com/c/vNqLuE/tUPf) in the Freezing group and a homolog of FY, which affects FCA mRNA processing, in the Geography group [[58]](https://paperpile.com/c/vNqLuE/tUPf) (Supp Table 2). In addition, several putative drought/stress response genes were identified, such as DREB transcription factor [[59]](https://paperpile.com/c/vNqLuE/CWVc) and an RCD1-like gene (Supplemental Table 2). RCD-1 is implicated in hormonal signaling and in the regulation of several stress-responsive genes in *Arabidopsis* *thaliana* [[57]](https://paperpile.com/c/vNqLuE/GCnW). In the Multi group, the only gene that was annotated functions in acclimation of photosynthesis to the environment in *A. thaliana* [[60]](https://paperpile.com/c/vNqLuE/TVGb).

2016 as undergoing convergent evolution in lodgepole pine with the interior spruce hybrid complex for adaptation to low temperatures, 10 were retained with our stringent criteria for top candidates. All of these genes grouped into the Freezing and Geography groups (shown by “\*” in Figure 1G), which were the two groups that had many SNPs with significant associations with elevation. This is consistent with the pattern of local adaptation in the interior spruce hybrid zone, whereby Engelmann spruce is adapted to higher elevations and white spruce is adapted to lower elevations [[61]](https://paperpile.com/c/vNqLuE/h5Ty).

*Comparison of co-expression modules to co-association modules*

To further explore if adaptation clusters have similar gene functions, we examined their gene expression patterns in response to climate treatments using previously published RNAseq data of 10,714 differentially expressed genes that formed 8 distinct co-expression modules [[55]](https://paperpile.com/c/vNqLuE/3Dcq). Of the 108 top candidate genes, 48 (44%) were also differentially expressed among treatments in response to factorial combinations of temperature (cold, mild, or hot), moisture (wet vs. dry), and/or day length (short vs. long day length). We found limited correspondence between co-association modules and co-expression clusters. Most of the top-candidate genes that were differentially expressed mapped to 2 of the 10 co-expression clusters previously characterized by [[55]](https://paperpile.com/c/vNqLuE/3Dcq) (Figure 7, blue circles are the P2 co-expression network and green triangles are the P7 co-expression network previously described by [[55]](https://paperpile.com/c/vNqLuE/3Dcq)). Genes in the P2 co-expression cluster had functions associated with the regulation of transcription and their expression was strongly influenced by all treatments, while genes in the P7 co-expression cluster had functions relating to metabolism, photosynthesis, and response to stimulus [[55]](https://paperpile.com/c/vNqLuE/3Dcq). Genes from the closely linked Aridity group mapped to 4 distinct co-expression modules, contigs from the Freezing group mapped to 3 distinct co-expression modules, and genes from the Geography group mapped to 3 distinct co-expression modules.

We used a Fisher exact test to determine if any co-expression cluster was over-represented in any of the the four major co-association groups shown in Figure 2. We found that the Freezing group had an over-representation of the P2 co-regulated gene expression cluster (*P* < 0.05) with seven (58%) of the Freezing genes found within the P2 expression network, revealing coordinated expression in response to climate conditions. Homologs of four of the seven genes  were present in *A. thaliana*, and three of these genes consisted of transcription factors involved in abiotic stress response (*DREB* transcription factor), flowering time (*CONSTANS*, pseudoresponse regulator) or the circadian clock (pseudo-response regulator 9). No other significant over-representation of gene expression class was identified for the four association groups or for all adaptation candidate genes.

*Simulated datasets*

Specifically, we used simulations with random sampling designs from three replicates across three demographic histories: (i) isolation by distance at equilibrium, and non-equilibrium range expansion from a (ii) single refuge or from (iii) two refugia. These landscape simulations were similar to lodgepole pine in the sense that they simulated large effective population sizes and resulted in similar *FST* across the landscape as that observed in pine ([[62, 63]](https://paperpile.com/c/vNqLuE/rgwy+LE3w), *FST*in simulations ~ 0.05, vs. FST in pine ~ 0.016 [[46]](https://paperpile.com/c/vNqLuE/UhqG)). To explore how the allele frequencies that evolved in these simulations might yield spurious patterns under the co-association network analysis, we overlaid the 22 environmental variables used in the lodgepole pine dataset onto the landscape genomic simulations [[62, 63]](https://paperpile.com/c/vNqLuE/rgwy+LE3w). To simulate the unmeasured environment, a small proportion of SNPs (1%) were subjected to computer-generated spatially varying selection along a weak latitudinal cline [[62, 63]](https://paperpile.com/c/vNqLuE/rgwy+LE3w).environmental response modules in the data given the correlation structure of observed environments), and (ii) loci were unlinked.

The *P*-value and Bayes factor criteria for choosing top candidate SNPs in the empirical data produced no false positives with the simulated datasets (Supplemental Figure 12), although using these criteria also reduced the proportion of true positives. Therefore, we used less stringent criteria to analyze the simulations so that we could also better understand patterns created by unlinked, false positive neutral loci.

We found that loci under selection by the same environmental factor generally formed a single tightly connected environmental response module even though they were unlinked, and that the degree of connectedness of selected loci was greater than among neutral loci (Figure 8). Thus, a single co-association module typically resulted from adaptation to the single selective environment in the simulations. This occurred because the distance threshold used to define connections in the environmental response modules was enriched for connections among selected loci that showed non-random associations in allele frequencies due to selection by a common environmental factor (Supplementary Figure 13A).

The propensity of neutral loci to form tightly-clustered co-association networks increased with the complexity of the demographic history: the false positive neutral loci from the two refugia model formed tightly connected networks (Figure 8 right column), despite the fact that all simulated loci were unlinked. This occurred because of non-random associations in allele frequency due to a shared demographic history. In some cases, selected loci formed separate or semi-separate modules according to their strengths of selection, but the underlying patterns of association were the same (e.g. Figure 8A, Supplementary Figure 14).

Discussion

Co-association networks provided a valuable framework for interpreting the genetic architecture of adaptation to a multivariate environment in lodgepole pine. Our most interesting result was the discovery of low recombination rates among genes putatively adapting to different and distinct aspects of climate, which was unexpected because selection is predicted to increase recombination between loci acted on by different sources of selection as discussed below. While the top candidate SNPs from most genes had associations with only a single aspect of the multivariate environment, for some genes we discovered evidence of environmental pleiotropy, i.e., candidate SNPs associated with multiple distinct aspects of climate. Within environmental response modules, we observed a combination of local sweep-like signatures (in which derived alleles at a locus were all found in a particular climate) and antagonistically pleiotropic patterns (in which some derived alleles at a locus were found at one environmental extreme and others found at the opposite extreme) underlying adaptation to climate, although we could not evaluate the relative importance of these patterns. Finally, we observed that the modularity of plastic responses to climate factors did not correspond to the modularity of the genetic architecture to climate, as evidenced through comparing co-association networks with co-expression networks. These results give insight into evolutionary debates about the extent of modularity and pleiotropy in the evolution of genetic architecture [[18–24]](https://paperpile.com/c/vNqLuE/htcT+k0Rn+BRvk+FI10+SGJc+ayb9+Tkxk).

*Genetic architecture of adaptation: pleiotropy and modularity*

Most of the top candidate genes in our analysis do not exhibit universal pleiotropy to distinct aspects of climate as defined by the expected pattern outlined in Figure 1B. Our results are more consistent with the the Hypothesis of Modular Pleiotropy [[19]](https://paperpile.com/c/vNqLuE/k0Rn), in which loci may have extensive effects on aspects of fitness *within* a distinct aspect of the environment (as defined by the variables that associate with each environmental response module), but few pleiotropic effects *among* distinct aspects of the environment. These results are in line with theoretical predictions that modular architectures should be favored in complex environments [[26]](https://paperpile.com/c/vNqLuE/hmua). But note also that if many pleiotropic effects are weak, the stringent statistical thresholds used in our study to reduce false positives may also reduce the extent to which pleiotropy is inferred [[20, 21]](https://paperpile.com/c/vNqLuE/BRvk+FI10). Therefore in our study, any pleiotropic effects of genes on fitness detected in multiple aspects of climate are likely to be large effects, and we refrain to making any claims as to the extent of environmental pleiotropy in the genome.

The extent of pleiotropy *within* individual environmental response modules is hard to quantify with our analysis, as for any given module we observed associations between genes and several environmental variables. Associations between a SNP and multiple environmental variables may or may not be interpreted as extensive environmental pleiotropic effects, depending on whether univariate environmental variables are considered distinct climates or collectively represent a single multivariate optimum. In many cases, these patterns are certainly affected by correlations among the environmental variables themselves.

Our results also highlight conceptual issues with the definition of and interpretation of pleiotropic effects on distinct aspects of fitness from real data: namely, what constitutes a “distinct aspect” (be it among traits, aspects of fitness, or aspects of the environment)? In this study we defined the selective environment through the perspective of the environmental variables that SNPs are associated with, using a threshold that produced reasonable results in simulation. But even with this definition, some environmental response modules are more similar in their multivariate environmental “niche” than others. For instance, genes within the Geography group could be interpreted to have extensive pleiotropic effects if the patterns of associations of each individual module were taken to be “distinct,” or they may be considered to have less extensive pleiotropic effects if their patterns of associations were too similar to be considered “distinct”. While the framework we present here is a step toward understanding and visualizing this hierarchical nature of “distinct aspects” of environmental factors, more quantitative frameworks are needed to frame pleiotropic effects in terms of their distinctness.

*Genetic architecture of adaptation: linkage*

We also observed physical linkage between genes that were associated with very distinct aspects of the climate. This was somewhat unexpected from a theoretical perspective, as selection would be expected to disfavour linkage and increase recombination between genes adapting to selection pressures with different spatial patterns of variation [[34–36]](https://paperpile.com/c/vNqLuE/y6Th+cSvc+1CYn). Interestingly, while the linkage map suggests that these loci are sometimes located relatively close together on a single chromosome, this does not seem to be sufficient physical linkage to also cause a noticeable increase in LD. In other words, it is possible that the amount of physical linkage sometimes observed between genes in different environmental response modules is not strong enough to constrain adaptation to these differing gradients. Genetic maps and reference genomes are not yet well developed for the large genomes of conifers; improved genetic maps or assembled genomes will be required to explore these questions in greater depth. If this finding is robust and not compromised by false positives, physical linkage among genes adapting to different climates could either facilitate or hinder a rapid evolutionary response as the multivariate environment changes [[4, 5]](https://paperpile.com/c/vNqLuE/9sr1+KcDn).

Within environmental response modules, we observed varying patterns of physical linkage among genes. The Aridity group, in particular, consisted of several tightly linked genes that may have arisen for a number of different reasons. Clusters of physically linked genes such as this may act as a single large-effect QTL [[64]](https://paperpile.com/c/vNqLuE/dqlN) and may have evolved due to competition among alleles or genomic rearrangements [[30, although these are rare in conifers]](https://paperpile.com/c/vNqLuE/ONV3/?suffix=%2C%20although%20these%20are%20rare%20in%20conifers), increased establishment probability due to linked adaptive alleles [[4]](https://paperpile.com/c/vNqLuE/9sr1), or divergence within inversions [[32]](https://paperpile.com/c/vNqLuE/N8fZ). Alternatively, if the Aridity region was one of low recombination, a single causal variant could create the appearance of linked selection [[65]](https://paperpile.com/c/vNqLuE/EtUT), a widespread false positive signal may have arisen due to genomic variation such as background selection and increased drift [[66–68]](https://paperpile.com/c/vNqLuE/l0wo+aqQx+fCND), or a widespread false signal may have arisen due to a demographic process such as allele surfing [[69, 70]](https://paperpile.com/c/vNqLuE/hG3M+zObF).

*Genetic architecture of adaptation: modularity of plasticity vs. fitness*

We also compared the modularity of plastic responses to climate (as measured by co-expression networks) to the modularity of evolved genetic responses to climate (as measured by co-association networks). Genes that show similar responses in expression to experimental treatments form a co-expression network. Co-expression networks have been successful at identifying genes that respond the same way to environmental stimuli [[71]](https://paperpile.com/c/vNqLuE/n4bC), so it might be reasonable to expect that if these genes were adapting to climate that they would also show similar patterns of associations with climate variables. However, differential expression analyses only identify genes with transcription (i.e., plastic) responses to climate - plasticity is not a prerequisite for adaptation and may be an alternative strategy to adaptation. This is illustrated by our result that only half of our top candidate contigs for adaptation to climate were differentially expressed in response to climate conditions.

Interestingly, we found limited correspondence between co-expression modules and our co-association modules that are putatively favored by natural selection. Specifically, genes that appeared to be adapting to different aspects of the multivariate environment (because they were located in different co-association modules) could none-the-less be co-expressed in response to specific conditions. However, we observed that loci from the tightly linked Aridity module had many distinct expression patterns in response to climate treatments. These observations support of the idea that the developmental/functional modularity of plasticity may not correspond to the modularity of the genotype to fitness map. However, the power of the analysis could be low due to stringent statistical cutoffs and these patterns warrant further investigation.

*Physiological adaptation of lodgepole pine to climate*

It is challenging to disentangle the physiological effects and importance of freezing versus drought in the local adaptation of conifers to climate. We found distinct groups of candidate genes along an axis of warm/wet to cold/dry (environmental response modules in the Freezing and Multi groups), and another distinct group along an axis of cold/wet to warm/dry (the Aridity environmental response module). Selection by drought conditions in winter may occur through extensive physiological remodeling that allows cells to survive intercellular freezing by desiccating protoplasts - but also results in drought stress at the cellular level [[55]](https://paperpile.com/c/vNqLuE/3Dcq). Another type of winter drought injury in lodgepole pine - red belt syndrome - is caused by warm, often windy events in winter, when foliage desiccates but the ground is too cold for roots to be able to supply water above ground [[72]](https://paperpile.com/c/vNqLuE/59Ru). This may contrast with drought selection in summer, when available soil water is lowest and aridity highest. The physiological and cellular mechanisms of drought and freezing response have similarities but also potentially important differences that could be responsible for the patterns we have observed.

Our results provide a framework for developing hypotheses that will disentangle the specific drivers of selection and provide genotypes for assisted gene flow in reforestation [[73]](https://paperpile.com/c/vNqLuE/njVD). While climate change is expected to increase average temperatures across this region, some areas are experiencing more precipitation than historic levels and others experiencing less [[74]](https://paperpile.com/c/vNqLuE/1kUZ). Tree mortality rates are increasing across North America due to increased drought and vapour pressure deficit for tree species including lodgepole pine, and associated increased vulnerability to damaging insects, but growth rates are also increasing with warming temperatures and increased carbon dioxide [[75, 76]](https://paperpile.com/c/vNqLuE/LTw8+tbtz). Hot, dry valleys in southern BC are projected to have novel climates emerge that have no existing analogues in North America [[77]](https://paperpile.com/c/vNqLuE/vOfl). The considerable standing adaptive variation we observe here involving many genes could facilitate adaptation to new temperature and moisture regimes, or could hinder adaptation if novel climates are at odds with the physical linkage among alleles adapted to different climate stressors.

*Limitations of associations with principal components*

For these data, testing associations of genes with PC-based climate variables would have led to a very limited interpretation of the environmental drivers of selection because the ordination is not biologically informed as to what factors are driving divergent selection [[37]](https://paperpile.com/c/vNqLuE/KF5g). First, many putative candidates in the Freezing and Geography groups would have been missed. Second, strong associations between the Multi SNPs and environmental variables that did not load strongly onto PC1, such as latitude, would have also been missed. Finally, many Aridity SNPs were outliers in PC3, which was strongly correlated with variables that the Aridity SNPs did not have any significant associations with. This occurred because no single variable loaded strongly onto PC3 (the maximum loading of any single variable was 0.38) and many variables had moderate loadings on to PC3, such that no single environmental variable explained the majority of the variance (the maximum variance explained by any one variable was 15%). Thus, associations with higher PC axes become increasingly difficult to interpret when the axis itself explains less variance of the multivariate environment and the environmental factors loading onto that axis explain similar amounts of variance in that axis. While principal components will capture the environmental factors that covary the most, this may have nothing to do with the combinations that drive divergent selection and local adaptation. This needlessly adds a layer of complexity to an analysis that may not reveal anything biologically important. In contrast, co-association networks highlight those combinations of environments that are biologically important for those genes likely involved in local adaptation.

*Benefits and caveats of co-association networks*

Co-association networks provide an intuitive framework to understand patterns of associations across many potentially correlated variables. By parsing loci into different groups based on their associations with multiple variables, this framework offers a more informative approach than grouping loci according to their outlier status based on associations with single environmental variables. While in this study we have used them to infer groups of loci that adapt to distinct aspects of the multivariate environment, co-association networks could be widely applied to a variety of situations, including genotype-phenotype associations. They offer the benefit of jointly identifying modules of loci and the groups of environmental variables that the modules are associated with. While the field may still have some disagreement about how modularity and pleiotropy should be defined, measured, and interpreted [[19–21, 23, 24]](https://paperpile.com/c/vNqLuE/k0Rn+BRvk+FI10+ayb9+Tkxk), co-association networks at least provide a quantitative framework to define modularity and infer pleiotropic effects among modules and a visual means of illustrating these.

Co-association networks differ from the application of bipartite network theory for estimating the degree of classical pleiotropic effects of genes on traits [[3]](https://paperpile.com/c/vNqLuE/siUQ). Bipartite networks are two-level networks where the genes form one type of nodes and the traits form the second type of nodes, and a connection is drawn from a gene to a trait if there is a significant association [[3]](https://paperpile.com/c/vNqLuE/siUQ). The degree of pleiotropy of a locus is then inferred by the number of traits that gene is connected to. With the bipartite network approach, trait nodes are defined by those traits measured, and not necessarily the multivariate effects from the perspective of the gene (e.g., a gene that affects organism size will have effects on height, weight and several other variables - if all these traits are analyzed, this gene would be inferred to have large pleiotropic effects). Even if highly correlated traits are removed, simulations have shown that even mild correlations in mutational effects can bias estimates of pleiotropy from bipartite networks [[20, 21]](https://paperpile.com/c/vNqLuE/BRvk+FI10). The advantage of co-association networks is their ability to identify combinations of variables (be they traits, environments, or other variables) that associate with modules. Correlated variables that measure essentially the same environment or phenotype will simply cluster together in a module, which can facilitate interpretation. On the other hand, correlated variables that measure different aspects of the environment or phenotype may cluster into different modules (as we observed in this study). The observed combinations of associations can then be used to develop and test hypotheses as to whether the combination represents a single multivariate environment that the gene is adapting to (in the case of allele associations with environment or fitness) or a single multivariate trait that the gene has effects on (in the case of allele associations with phenotypes).

While co-association networks hold promise for elucidating the modularity and pleiotropy of the genotype-phenotype-fitness map, there are some caveats to their application that should be noted. First, correlation among variables will make it difficult to infer the exact conditions that select for or the exact traits that associate with particular allelic combinations. Results from this framework can make it easier, however, to generate hypotheses that can be tested with future experiments. Second, the analysis of simulated data shows that investigators should consider demographic history and choose candidates with caution for data analysis to exclude false positives, as we have attempted here. Co-association networks can arise among unlinked neutral loci by chance, and it is almost certain that some proportion of the “top candidates SNPs” in this study are false positives due to linkage with causal SNPs or due to demographic history. The simulated data also showed, however, that causal SNPs tend to have a higher degree of connection in their co-association network than neutral loci, and this might help to prioritize SNPs for follow up experiments, SNP arrays, and genome editing. Third, it may be difficult to draw conclusions about the level of modularity of the genetic architecture. The number of modules may be sensitive to the statistical thresholds used to identify top candidate SNPs [[20, 21]](https://paperpile.com/c/vNqLuE/BRvk+FI10) as well as the distance threshold used to identify modules. While inferred modules composed of a single SNP could be interpreted as unique, our simulations also show that neutral loci are more likely to be unconnected in co-association networks. Many alleles of small effect may be just below statistical detection, and whether or not these alleles are included could profoundly change inference as to the extent of pleiotropy [[20, 21]](https://paperpile.com/c/vNqLuE/BRvk+FI10). This presents a conundrum common to most population genomic approaches for detecting selection, because lowering statistical thresholds will almost certainly increase the number of false positives, while only using very stringent statistical thresholds may decrease the probability of observing pleiotropy if many pleiotropic effects are weak [[20]](https://paperpile.com/c/vNqLuE/BRvk).  Thus, while co-association networks are useful for identifying SNP modules associated with correlated variables, further work is necessary to expand this framework to quantitatively measure pleiotropic effects and the strength of modularity.

Conclusions

In this study we discovered physical linkage among loci putatively adapting to different aspects of the climate. These results give rare insight into both the ecological pressures that favor the evolution of modules by natural selection [19] and into the organization of the genetic architecture itself. As climate changes, the evolutionary response will be determined by the extent of physical linkage among these loci, in combination with the strength of selection and phenotypic optima across the environmental gradient, the scale and pattern of environmental variation, and the details of migration and demographic fluctuations across the landscape. While theory has made strides to provide a framework for predicting the genetic architecture of local adaptation under divergence with gene flow to a single environment [[4, 30, 31, 78–82]](https://paperpile.com/c/vNqLuE/ONV3+IiAA+CxPj+Dxw1+zPtc+lItH+9sr1+Ea1g), as well as the evolution of correlated traits under different directions and/or strengths of selection when those traits have a common genetic basis [[35, 36]](https://paperpile.com/c/vNqLuE/cSvc+1CYn), how genetic architectures evolve on complex heterogeneous landscapes has not been clearly elucidated. Furthermore, it has been difficult to test theory because the field still lacks a framework for evaluating empirical observations of adaptation in many dimensions. Here, we have attempted to develop an initial framework for understanding adaptation to several complex environments with different spatial patterns, which may also be useful for understanding the genetic basis of multivariate phenotypes from genome-wide association studies. This framework lays the foundation for future studies to study modularity across the genotype-phenotype-fitness continuum.

This study uses the same dataset analyzed by Yeaman et al. [[46]](https://paperpile.com/c/vNqLuE/UhqG), but with a different focus as explained in the introduction. Briefly, we obtained seeds from 281 sampling locations of lodgepole pine from reforestation collections for natural populations, and these locations were selected to represent the full range of climatic and ecological conditions within the species range in British Columbia and Alberta based on ecosystem delineations. Seeds were grown in a common garden and  2-4 individuals were sampled from each sampling location. The environment for each sampling location was was characterized by estimating climate normals for 1961-1990 from geographic coordinates using the software package ClimateWNA [[83]](https://paperpile.com/c/vNqLuE/605b). The program extracts and downscales the moderate spatial resolution generated by PRISM [[84]](https://paperpile.com/c/vNqLuE/DKgI) to scale-free and calculates many climate variables for specific locations based on latitude, longitude and elevation.

The methods for this section are identical to those reported in [[46]](https://paperpile.com/c/vNqLuE/UhqG). Briefly, [[46 for more details, see 47]](https://paperpile.com/c/vNqLuE/HGG2+UhqG/?prefix=see,&suffix=,for%20more%20details) and the resulting captured fragments were amplified using the protocol and reagents from the NEXTflex kit.

Sequenced reads were filtered and aligned to the loblolly pine genome [[85]](https://paperpile.com/c/vNqLuE/vVBR) using bwa mem [[86]](https://paperpile.com/c/vNqLuE/iZUG) and variants were called using GATK Unified Genotyper [[87]](https://paperpile.com/c/vNqLuE/m1ss), with steps included for removal of PCR duplicates, realignment around indels, and base quality score recalibration [[46, 87]](https://paperpile.com/c/vNqLuE/m1ss+UhqG).

[[46]](https://paperpile.com/c/vNqLuE/UhqG).We also performed a BLASTX against the nr database screened for green plants and used Blast2GO [[88]](https://paperpile.com/c/vNqLuE/Vpkc) to assign GO terms and enzyme codes [[46 for details, see 55]](https://paperpile.com/c/vNqLuE/3Dcq+UhqG/?prefix=see,&suffix=,for%20details).To identify if genes with particular molecular function and biological processes were over-represented in top candidate genes, we performed a GO enrichment analysis using topGO [[89]](https://paperpile.com/c/vNqLuE/1ih5). All GO terms associated with at least two candidate genes were analyzed for significant over-representation within each group and in all candidate genes (FDR 5%).

First, top candidate genes were obtained from [[46]](https://paperpile.com/c/vNqLuE/UhqG). For this study, genes with unusually strong signatures of association from multiple association tests (uncorrected genotype-phenotype and genotype-environment correlations, for details see [[46]](https://paperpile.com/c/vNqLuE/UhqG))

For this study, we identified top candidate SNPs within the set of top candidate genes.Note that because candidate SNPs are limited to the top candidate genes

[[39]](https://paperpile.com/c/vNqLuE/1IaM).As detailed in [[46]](https://paperpile.com/c/vNqLuE/UhqG)set of non-coding SNPs to[[90]](https://paperpile.com/c/vNqLuE/c4Bo).

groups in the data. For each of these main groups[[91]](https://paperpile.com/c/vNqLuE/yXK5).

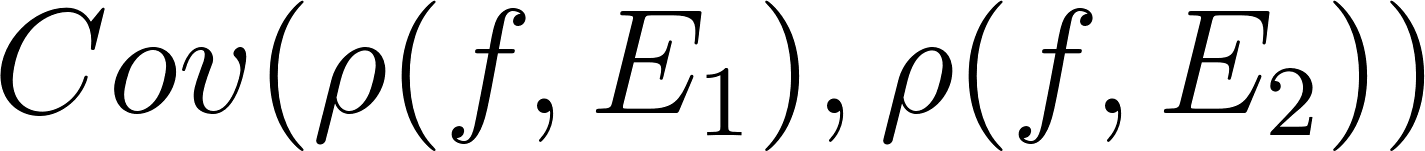
*Linkage disequilibrium*

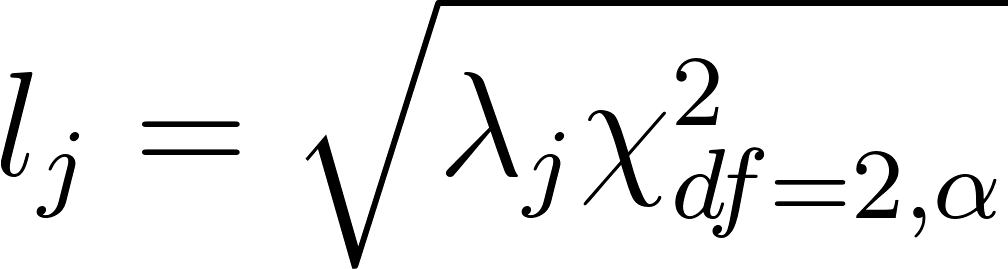
Linkage disequilibrium was calculated among pairwise combinations of SNPs within genes (genes). Mean values of Pearson’s correlation coefficient squared (*r2*) were estimated

[[92]](https://paperpile.com/c/vNqLuE/5EFO).

Note that this criterion is less conservative than that used to identify candidate SNPs for the network analysis (because it did not require the additional criteria of a significant Bonferroni-corrected *P*-value), so it should result in greater overlap between PC candidate SNPs and top candidate SNPs based on univariate associations.

The co-expression data used in this study was previously published by [[55]](https://paperpile.com/c/vNqLuE/3Dcq). To determine if adaptation cluster members had similar gene functions, we examined their gene expression patterns in response to seven growth chamber climate treatments using previously published RNAseq data [[55]](https://paperpile.com/c/vNqLuE/3Dcq). Expression data was collected on 44 seedlings from a single sampling location, raised under common conditions, and then exposed to growth chamber environments that varied in their temperature, moisture and photoperiod regimes. We used a Fisher’s exact test to determine if genes with a significant climate treatment effect were over-represented in each of the 4 major groups and across all adaptation candidates relative to the other sequenced and expressed genes. In addition, Yeaman et al 2014 used weighted gene co-expression network analysis (WGCNA) to identify eight clusters of co-regulated genes among the seven climate treatments. We used a Fisher’s exact test to determine if these previously identified expression clusters were over-represented in the any of the 4 major groups relative to the other sequenced and expressed genes.

For two variables, the 2 x 2 variance-covariance matrix of [](https://www.codecogs.com/eqnedit.php?latex=Cov(%5Crho(f,%20E_1),%20%5Crho(f,%20E_2)))

[](https://www.codecogs.com/eqnedit.php?latex=l_j%20=%20%5Csqrt%7B%20%5Clambda_j%20%5Cchi%5E2_%7Bdf=2,%5Calpha%7D),

where *lj*= {1, 2} represents the lengths of the major and minor axes on the ellipse, respectively, *λj*  represents the eigenvalues of the covariance matrix, and χ2df=2,α represents the value of the χ2 distribution for the desired *α* value [[93–95]](https://paperpile.com/c/vNqLuE/u7x6+OzFa+H3oc).

environmental response modules.[[83]](https://paperpile.com/c/vNqLuE/605b) and shaded with colour gradients scaled to the range of climates across the sampling locations.

The simulations used in this study are identical a subset of those previously published by [[63]](https://paperpile.com/c/vNqLuE/LE3w). Briefly, the simulator uses forward-in-time recurrence equations to model the evolution of independent haploid SNPs on a quasi-continuous square landscape. We modelled three demographic histories that resulted in the same overall neutral *FST* for each demography, but demographic history determined the distribution of *FST*’s around that mean (IBD had the lowest variance, followed by 1R, and 2R had the highest variance). The landscape size was 360 x 360 demes and migration was determined by a discretized version of a Gaussian dispersal kernel. Carrying capacity per deme differed slightly for each scenario to give the same overall neutral *FST* = 0.05. IBD was run until equilibrium at 10,000 generations, but 1R and 2R were only run for 1,000 generations in order to mimic the the expansion of lodgepole pine since the last glacial maximum [[96]](https://paperpile.com/c/vNqLuE/bYgs). All selected loci adapted to computer generated landscape with a weak north-south cline and spatial heterogeneity at smaller spatial scales. See or more details.

The simulations were then expanded in the following way: for

List of abbreviations

* LD: Linkage disequilibrium
* PC: Principal components
* SNP: single nucleotide polymorphism

Declarations

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and material**

The dataset(s) supporting the conclusions of this article are available in the Dryad repository [unique persistent identifier and hyperlink to dataset(s) in http:// format will be archived upon acceptance of manuscript].

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

KEL was supported by a grant from the National Science Foundation (502483). This research was part of the AdapTree project, led by SNA and by Genome Canada (LSARP2010\_161REF), Genome BC, Genome Alberta, Alberta Innovates BioSolutions, the Forest Genetics Council of British Columbia, the British Columbia Ministry of Forests, Lands and Natural Resource Operations (BCMFLNRO), Virginia Polytechnic University, and the University of British Columbia.

**Authors' contributions**

KEL conceived of the analysis, conducted analyses, and lead writing of the manuscript. KH and SY did the bioinformatics and various specific analyses. JD created the allele frequency landscape plots. SA led the AdapTree project. All authors contributed to writing of the manuscript.

**Acknowledgements**

We thank Sebastian E. Ramos-Onsins, Tanja Pyhäjärvi, an anonymous reviewer and PCI Evol Biol for comments that greatly improved this manuscript. Mike Whitlock provided valuable advice and feedback on various aspects of the research. We thank Jeremy Yoder for organizing the SNP chip data used for calculating the recombination rates. Pia Smets, Connor Fitzpatrick and Sarah Markert assembled and grew genetic materials, and Kristin Nurkowski prepared sequence capture libraries. Tongli Wang and Andreas Hamann selected populations based on climatic distribution of species. Seeds were kindly donated by 63 forest companies and agencies in Alberta and British Columbia (listed at http://adaptree.forestry. ubc.ca/seed-contributors/).

References

[1. Hansen TF. The evolution of genetic architecture. Annu Rev Ecol Evol Syst. 2006;37:123–57.](http://paperpile.com/b/vNqLuE/2ocb)

[2. Orr HA. Adaptation and the cost of complexity. Evolution. 2000;54:13–20.](http://paperpile.com/b/vNqLuE/FdfN)

[3. Wang Z, Liao B-Y, Zhang J. Genomic patterns of pleiotropy and the evolution of complexity. Proc Natl Acad Sci U S A. 2010;107:18034–9.](http://paperpile.com/b/vNqLuE/siUQ)

[4. Aeschbacher S, Bürger R. The effect of linkage on establishment and survival of locally beneficial mutations. Genetics. 2014;197:317–36.](http://paperpile.com/b/vNqLuE/9sr1)

[5. Reeve J, Ortiz-Barrientos D, Engelstädter J. The evolution of recombination rates in finite populations during ecological speciation. Proc Biol Sci. 2016;283. doi:](http://paperpile.com/b/vNqLuE/KcDn)[10.1098/rspb.2016.1243](http://dx.doi.org/10.1098/rspb.2016.1243)[.](http://paperpile.com/b/vNqLuE/KcDn)

[6. Barton NH. Genetic linkage and natural selection. Philos Trans R Soc Lond B Biol Sci. 2010;365:2559–69.](http://paperpile.com/b/vNqLuE/d75E)

[7. Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet. 2011;12:204–13.](http://paperpile.com/b/vNqLuE/nmjU)

[8. Paaby AB, Rockman MV. The many faces of pleiotropy. Trends Genet. 2013;29:66–73.](http://paperpile.com/b/vNqLuE/M4Tc)

[9. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20.](http://paperpile.com/b/vNqLuE/nGPC)

[10. Slatkin M. Gene flow and selection in a cline. Genetics. 1973;75:733–56.](http://paperpile.com/b/vNqLuE/NNGV)

[11. Slatkin M. Spatial patterns in the distributions of polygenic characters. J Theor Biol. 1978;70:213–28.](http://paperpile.com/b/vNqLuE/sRrP)

[12. Barton NH. Clines in polygenic traits. Genet Res. 1999;74:223–36.](http://paperpile.com/b/vNqLuE/xgno)

[13. Felsenstein J. The theoretical population genetics of variable selection and migration. Annu Rev Genet. 1976;10:253–80.](http://paperpile.com/b/vNqLuE/3Pai)

[14. Haldane JBS. The theory of a cline. J Genet. 1948;48:277–84.](http://paperpile.com/b/vNqLuE/qpAW)

[15. Haldane JBS. A mathematical theory of natural and artificial selection (Part VI, Isolation). Math Proc Cambridge Philos Soc. 1930;26:220.](http://paperpile.com/b/vNqLuE/yp41)

[16. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24:4348–70.](http://paperpile.com/b/vNqLuE/Ite8)

[17. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, et al. Adaptation to climate across the](http://paperpile.com/b/vNqLuE/Tmmz) *[Arabidopsis thaliana](http://paperpile.com/b/vNqLuE/Tmmz)* [genome. Science. 2011;334:83–6.](http://paperpile.com/b/vNqLuE/Tmmz)

[18. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. Cell. 2017;169:1177–86.](http://paperpile.com/b/vNqLuE/htcT)

[19. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet. 2007;8:921–31.](http://paperpile.com/b/vNqLuE/k0Rn)

[20. Hill WG, Zhang X-S. Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. Nat Rev Genet. 2012. doi:](http://paperpile.com/b/vNqLuE/BRvk)[10.1038/nrg2949-c1](http://dx.doi.org/10.1038/nrg2949-c1)[.](http://paperpile.com/b/vNqLuE/BRvk)

[21. Hill WG, Zhang X-S. On the pleiotropic structure of the genotype–phenotype map and the evolvability of complex organisms. Genetics. 2012;190:1131–7.](http://paperpile.com/b/vNqLuE/FI10)

[22. Rockman MV. The QTN program and the alleles that matter for evolution: all that’s gold does not glitter. Evolution. 2012;66:1–17.](http://paperpile.com/b/vNqLuE/SGJc)

[23. Paaby AB, Rockman MV. Pleiotropy: what do you mean? Reply to Zhang and Wagner. Trends Genet. 2013;29:384.](http://paperpile.com/b/vNqLuE/ayb9)

[24. Wagner GP, Zhang J. Universal pleiotropy is not a valid null hypothesis: reply to Hill and Zhang. Nat Rev Genet. 2012;13:296.](http://paperpile.com/b/vNqLuE/Tkxk)

[25. Wagner GP. Homologues, natural kinds and the evolution of modularity. Am Zool. 1996;36:36–43.](http://paperpile.com/b/vNqLuE/6Zft)

[26. Le Nagard H, Chao L, Tenaillon O. The emergence of complexity and restricted pleiotropy in adapting networks. BMC Evol Biol. 2011;11:326.](http://paperpile.com/b/vNqLuE/hmua)

[27. Griswold CK. Pleiotropic mutation, modularity and evolvability. Evol Dev. 2006;8:81–93.](http://paperpile.com/b/vNqLuE/nGKW)

[28. Le Corre V, Kremer A. Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. Genetics. 2003;164:1205–19.](http://paperpile.com/b/vNqLuE/N75Q)

[29. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8:269–94.](http://paperpile.com/b/vNqLuE/MsZ2)

[30. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci U S A. 2013;110:E1743–51.](http://paperpile.com/b/vNqLuE/ONV3)

[31. Yeaman S, Aeschbacher S, Bürger R. The evolution of genomic islands by increased establishment probability of linked alleles. Mol Ecol. 2016;25:2542–58.](http://paperpile.com/b/vNqLuE/IiAA)

[32. Kirkpatrick M. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173:419–34.](http://paperpile.com/b/vNqLuE/N8fZ)

[33. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Curr Biol. 2014;24:R288–94.](http://paperpile.com/b/vNqLuE/HvzF)

[34. Lenormand T, Otto SP. The evolution of recombination in a heterogeneous environment. Genetics. 2000;156:423–38.](http://paperpile.com/b/vNqLuE/y6Th)

[35. Guillaume F. Migration-induced phenotypic divergence: the migration-selection balance of correlated traits. Evolution. 2011;65:1723–38.](http://paperpile.com/b/vNqLuE/cSvc)

[36. Chebib J, Guillaume F. What affects the predictability of evolutionary constraints using a G-matrix? The relative effects of modular pleiotropy and mutational correlation. Evolution. 2017. doi:](http://paperpile.com/b/vNqLuE/1CYn)[10.1111/evo.13320](http://dx.doi.org/10.1111/evo.13320)[.](http://paperpile.com/b/vNqLuE/1CYn)

[37. Houle D, Mezey J, Galpern P. Interpretation of the results of common principal components analyses. Evolution. 2002;56:433–40.](http://paperpile.com/b/vNqLuE/KF5g)

[38. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30:1687–99.](http://paperpile.com/b/vNqLuE/vwhx)

[39. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. Genetics. 2013;195:205–20.](http://paperpile.com/b/vNqLuE/1IaM)

[40. Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics. 2015;201:1555–79.](http://paperpile.com/b/vNqLuE/Eiy0)

[41. Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. Characterizing genomic variation of](http://paperpile.com/b/vNqLuE/WHAY) *[Arabidopsis thaliana](http://paperpile.com/b/vNqLuE/WHAY)*[: the roles of geography and climate. Mol Ecol. 2012;21:5512–29.](http://paperpile.com/b/vNqLuE/WHAY)

[42. Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, et al. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (](http://paperpile.com/b/vNqLuE/p6hC)*[Homarus americanus](http://paperpile.com/b/vNqLuE/p6hC)*[). Mol Ecol. 2016;25:5073–92.](http://paperpile.com/b/vNqLuE/p6hC)

[43. Hedrick PW. Genetic polymorphism in heterogeneous environments: a decade later. Annu Rev Ecol Syst. 1986;17:535–66.](http://paperpile.com/b/vNqLuE/oPtk)

[44. Hedrick PW, Ginevan ME, Ewing EP. Genetic polymorphism in heterogeneous environments. Annu Rev Ecol Syst. 1976;7:1–32.](http://paperpile.com/b/vNqLuE/KFAA)

[45. Barton NH. Multilocus clines. Evolution. 1983;37:454–71.](http://paperpile.com/b/vNqLuE/AVgs)

[46. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, et al. Convergent local adaptation to climate in distantly related conifers. Science. 2016;353:1431–3.](http://paperpile.com/b/vNqLuE/UhqG)

[47. Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, et al. Exome capture from the spruce and pine giga-genomes. Mol Ecol Resour. 2016;16:1136–46.](http://paperpile.com/b/vNqLuE/HGG2)

[48. Hodgins KA, Yeaman S, Nurkowski KA, Rieseberg LH, Aitken SN. Expression divergence Is correlated with sequence evolution but not positive selection in conifers. Mol Biol Evol. 2016;33:1502–16.](http://paperpile.com/b/vNqLuE/val2)

[49. Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. Back to nature: ecological genomics of loblolly pine (](http://paperpile.com/b/vNqLuE/bVD2)*[Pinus taeda](http://paperpile.com/b/vNqLuE/bVD2)*[, Pinaceae). Mol Ecol. 2010;19:3789–805.](http://paperpile.com/b/vNqLuE/bVD2)

[50. Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, et al. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (](http://paperpile.com/b/vNqLuE/JkoP)*[Pinus taeda](http://paperpile.com/b/vNqLuE/JkoP)* [L., Pinaceae). Genetics. 2010;185:969–82.](http://paperpile.com/b/vNqLuE/JkoP)

[51. Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, et al. Potential for evolutionary responses to climate change - evidence from tree populations. Glob Chang Biol. 2013;19:1645–61.](http://paperpile.com/b/vNqLuE/vUkI)

[52. Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Can J Bot. 2003;81:1247–66.](http://paperpile.com/b/vNqLuE/yaFi)

[53. Liepe KJ, Hamann A, Smets P, Fitzpatrick CR, Aitken SN. Adaptation of lodgepole pine and interior spruce to climate: implications for reforestation in a warming world. Evol Appl. 2016;9:409–19.](http://paperpile.com/b/vNqLuE/cZUs)

[54. Illingworth K. Study of lodgepole pine genotype-environment interaction in B.C. In: Proceedings International Union of Forestry Research Organizations (IUFRO) Joint Meeting of Working parties: Douglas-fir provenances, Lodgepole Pine Provenances, Sitka Spruce Provenances and Abies Provenances. Vancouver, British Columbia, Canada; 1978. p. 151–8.](http://paperpile.com/b/vNqLuE/FBGF)

[55. Yeaman S, Hodgins KA, Suren H, Nurkowski KA, Rieseberg LH, Holliday JA, et al. Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (](http://paperpile.com/b/vNqLuE/3Dcq)*[Pinus contorta](http://paperpile.com/b/vNqLuE/3Dcq)*[) and interior spruce (](http://paperpile.com/b/vNqLuE/3Dcq)*[Picea glauca×Picea engelmannii](http://paperpile.com/b/vNqLuE/3Dcq)*[). New Phytol. 2014;203:578–91.](http://paperpile.com/b/vNqLuE/3Dcq)

[56. Blumwald E, Aharon GS, Apse MP. Sodium transport in plant cells. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2000;1465:140–51.](http://paperpile.com/b/vNqLuE/RNnp)

[57. Ahlfors R, Lång S, Overmyer K, Jaspers P, Brosché M, Tauriainen A, et al.](http://paperpile.com/b/vNqLuE/GCnW) *[Arabidopsis](http://paperpile.com/b/vNqLuE/GCnW)* [RADICAL-INDUCED CELL DEATH1 belongs to the WWE protein-protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses. Plant Cell. 2004;16:1925–37.](http://paperpile.com/b/vNqLuE/GCnW)

[58. Amasino RM, Michaels SD. The timing of flowering. Plant Physiol. 2010;154:516–20.](http://paperpile.com/b/vNqLuE/tUPf)

[59. Singh D, Laxmi A. Transcriptional regulation of drought response: a tortuous network of transcriptional factors. Front Plant Sci. 2015;6:895.](http://paperpile.com/b/vNqLuE/CWVc)

[60. Walters RG, Shephard F, Rogers JJM, Rolfe SA, Horton P. Identification of mutants of](http://paperpile.com/b/vNqLuE/TVGb) *[Arabidopsis](http://paperpile.com/b/vNqLuE/TVGb)* [defective in acclimation of photosynthesis to the light environment. Plant Physiol. 2003;131:472–81.](http://paperpile.com/b/vNqLuE/TVGb)

[61. De La Torre A, Ingvarsson PK, Aitken SN. Genetic architecture and genomic patterns of gene flow between hybridizing species of Picea. Heredity . 2015;115:153–64.](http://paperpile.com/b/vNqLuE/h5Ty)

[62. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of](http://paperpile.com/b/vNqLuE/rgwy) *[F](http://paperpile.com/b/vNqLuE/rgwy)[ST](http://paperpile.com/b/vNqLuE/rgwy)* [outlier tests. Mol Ecol. 2014;23:2178–92.](http://paperpile.com/b/vNqLuE/rgwy)

[63. Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015;24:1031–46.](http://paperpile.com/b/vNqLuE/LE3w)

[64. Christians JK, Senger LK. Fine mapping dissects pleiotropic growth quantitative trait locus into linked loci. Mamm Genome. 2007;18:240–5.](http://paperpile.com/b/vNqLuE/dqlN)

[65. Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 1997;70:155–74.](http://paperpile.com/b/vNqLuE/EtUT)

[66. Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012;190:5–22.](http://paperpile.com/b/vNqLuE/l0wo)

[67. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134:1289–303.](http://paperpile.com/b/vNqLuE/aqQx)

[68. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188:379–97.](http://paperpile.com/b/vNqLuE/fCND)

[69. Klopfstein S, Currat M, Excoffier L. The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol. 2006;23:482–90.](http://paperpile.com/b/vNqLuE/hG3M)

[70. Hofer T, Ray N, Wegmann D, Excoffier L. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet. 2009;73:95–108.](http://paperpile.com/b/vNqLuE/zObF)

[71. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4:Article17.](http://paperpile.com/b/vNqLuE/n4bC)

[72. Bella IE, Navratil S. Growth losses from winter drying (red belt damage) in lodgepole pine stands on the east slopes of the Rockies in Alberta. Can J For Res. 1987;17:1289–92.](http://paperpile.com/b/vNqLuE/59Ru)

[73. Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. Annu Rev Ecol Evol Syst. 2013;44:367–88.](http://paperpile.com/b/vNqLuE/njVD)

[74. Mbogga MS, Hamann A, Wang T. Historical and projected climate data for natural resource management in western Canada. Agric For Meteorol. 2009;149:881–90.](http://paperpile.com/b/vNqLuE/1kUZ)

[75. Hember RA, Kurz WA, Coops NC. Relationships between individual-tree mortality and water-balance variables indicate positive trends in water stress-induced tree mortality across North America. Glob Chang Biol. 2017;23:1691–710.](http://paperpile.com/b/vNqLuE/LTw8)

[76. Hember RA, Kurz WA, Coops NC. Increasing net ecosystem biomass production of Canada’s boreal and temperate forests despite decline in dry climates. Global Biogeochem Cycles. 2017;31:2016GB005459.](http://paperpile.com/b/vNqLuE/tbtz)

[77. Mahony CR, Cannon AJ, Wang T, Aitken SN. A closer look at novel climates: new methods and insights at continental to landscape scales. Glob Chang Biol. 2017. doi:](http://paperpile.com/b/vNqLuE/vOfl)[10.1111/gcb.13645](http://dx.doi.org/10.1111/gcb.13645)[.](http://paperpile.com/b/vNqLuE/vOfl)

[78. Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. Evolution. 2011;65:1897–911.](http://paperpile.com/b/vNqLuE/CxPj)

[79. Kremer A, Le Corre V. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. Heredity . 2012;108:375–85.](http://paperpile.com/b/vNqLuE/Dxw1)

[80. Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 2012;21:1548–66.](http://paperpile.com/b/vNqLuE/zPtc)

[81. Flaxman SM, Feder JL, Nosil P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. Evolution. 2013;67:2577–91.](http://paperpile.com/b/vNqLuE/lItH)

[82. Bürger R, Akerman A. The effects of linkage and gene flow on local adaptation: A two-locus continent–island model. Theor Popul Biol. 2011;80:272–88.](http://paperpile.com/b/vNqLuE/Ea1g)

[83. Wang T, Hamann A, Spittlehouse DL, Murdock TQ. ClimateWNA—high-resolution spatial climate data for western North America. J Appl Meteorol Climatol. 2012;51:16–29.](http://paperpile.com/b/vNqLuE/605b)

[84. Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int J Climatol. 2008;28:2031–64.](http://paperpile.com/b/vNqLuE/DKgI)

[85. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. 2014;15:R59.](http://paperpile.com/b/vNqLuE/vVBR)

[86. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.](http://paperpile.com/b/vNqLuE/iZUG)

[87. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.](http://paperpile.com/b/vNqLuE/m1ss)

[88. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008;2008:619832.](http://paperpile.com/b/vNqLuE/Vpkc)

[89. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. 2009.](http://paperpile.com/b/vNqLuE/1ih5) <https://bioconductor.riken.jp/packages/3.2/bioc/vignettes/topGO/inst/doc/topGO.pdf.> [Accessed 1 Jan 2017.](http://paperpile.com/b/vNqLuE/1ih5)

[90. Blair LM, Granka JM, Feldman MW. On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics. 2014;8:1.](http://paperpile.com/b/vNqLuE/c4Bo)

[91. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695:1–9.](http://paperpile.com/b/vNqLuE/yXK5)

[92. Margarido GRA, Souza AP, Garcia AAF. OneMap: software for genetic mapping in outcrossing species. Hereditas. 2007;144:78–9.](http://paperpile.com/b/vNqLuE/5EFO)

[93. Pison G, Struyf A, Rousseeuw PJ. Displaying a clustering with CLUSPLOT. Comput Stat Data Anal. 1999;30:381–92.](http://paperpile.com/b/vNqLuE/u7x6)

[94. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons; 2009.](http://paperpile.com/b/vNqLuE/OzFa)

[95. Titterington DM. Algorithms for computing D-optimal design on finite design spaces. Proceedings of the 1976 Conference on Information Science and Systems. 1976;:213–6.](http://paperpile.com/b/vNqLuE/H3oc)

[96. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405:907–13.](http://paperpile.com/b/vNqLuE/bYgs)

Tables

Table 1. Overview of terminology used in the literature regarding pleiotropy and modularity.

|  |  |  |
| --- | --- | --- |
| Term | References | Meaning |
| Selectional pleiotropy | [[8]](https://paperpile.com/c/vNqLuE/M4Tc) | Traits are defined by the action of selection and not by the intrinsic attributes of the organism |
| Antagonistic pleiotropy | [10] | An allele has different effects on fitness at different extremes of an environmental variable (e.g., positive effects on fitness in cold environments and negative effects in warm environments), which results in an association between the allele frequency and the environmental variable |
| Environmental pleiotropy | This study | Genes affect fitness in multiple distinct aspects of the multivariate environment, where each aspect is defined by the action of selection |
| Modularity or modular genetic architecture | [[25]](https://paperpile.com/c/vNqLuE/6Zft) | Pleiotropic effects are limited to elements within a module, with a suppression of pleiotropic effects between different modules (Figure 1A, left column) |
| Co-association network | This study | An application of network theory used to identify modules of loci that are similar in their associations across many variables. |
| Environmental response module | This study | A group of SNPs that adapt to a distinct aspect of the selectional environment. These modules can be thought of as “variational” modules [[sensu 19]](https://paperpile.com/c/vNqLuE/k0Rn/?prefix=sensu), which are composed of features that vary together and are relatively independent of other such sets of features. In practice, environmental response modules are inferred by their similarity in associations with multiple environmental variables. |
| Distinct aspect of the selectional environment | This study | A multivariate environmental fitness landscape to which a SNP adapts on a geographic landscape. In practice, these are inferred by the environmental variables that associate with candidate SNPs within environmental response modules. |

Table 2. Environmental variables measured for each sampling location, ordered by their abbreviations shown in Figure 2 A and B.

|  |  |  |
| --- | --- | --- |
| Abbreviation | Definition | Category |
| MSP | May to September precipitation (mm) | Aridity |
| LONG | Longitude | Geography |
| bFPP | Day of the year frost-free period begins | Freezing |
| ELEVATION | Elevation | Geography |
| LAT | Latitude | Geography |
| TD | Temperature difference (MWMT-MCMT) (°C) | Freezing or Aridity |
| DD\_0 | Degree-days below 0°C | Freezing |
| PAS | Precipitation as snow (mm) | Aridity or Freezing |
| MAP | Mean annual precipitation (mm) | Aridity |
| CMD | Hargreaves climate-moisture deficit | Aridity |
| SHM | Summer heat-moisture index ((MWMT)/(MSP/1000)) | Aridity |
| AHM | Annual heat-moisture index (MAT+10)/(MAP/1000)) | Aridity |
| MWMT | Mean warmest month temperature (°C) | Aridity |
| DD5 | Degree-days above 5°C | Aridity |
| Eref | Hargreaves reference evaporation | Aridity |
| EXT | Extreme maximum temperature over 30 years (°C) | Aridity |
| MCMT | Mean coldest month temperature (°C) | Freezing |
| EMT | Extreme minimum temperature over 30 years (°C) | Freezing |
| MAT | Mean annual temperature (°C) | Aridity or Freezing |
| eFFP | Day of the year frost-free period ends | Freezing |
| NFFD | Number of days without frost | Freezing |
| FFP | Frost-free period (bFFP-eFFP) | Freezing |

Figure Legends

**Figure 1. Conceptual framework for evaluating modularity of genetic architectures adapting to environment.**In this example, each gene (identified by numbers) contains two causal SNPs (identified by letters) where mutations affect fitness in potentially different aspects of the environment. The two aspects of the environment that affect fitness are aridity and freezing. A) The true underlying genetic architecture adapting to multiple aspects of climate. The left column represents a modular genetic architecture in which any pleiotropic effects of genes are limited to a particular aspect of the environment. The right column represents a non-modular architecture, in which genes have pleiotropic effects on multiple aspects of the environment. Universal pleiotropy occurs when a gene has effects on all the multiple distinct aspects of the environment. Genes in this example are unlinked in the genome, but linkage among genes is an important aspect of the environmental response architecture. B) Hierarchical clustering is used to identify the “environmental response modules,” which jointly describe the groups of loci that adapt to a distinct aspects of climate as well as the distinct aspects of climate to which they adapt. For example, the “aridity module” is a group of SNPs within two genes adapting to aridity, and shows associations with both temperature and climate-moisture deficit. C) Co-association networks are used to visualize the results of the hierarchical clustering with regards to the environment, and connections are based on similarity in SNPs in their associations with environments. In this example all SNPs within a module are have the same associations with multiple environmental variables. D) Pleiotropy barplots are used to visualize the results of the hierarchical clustering with regards to the genetic architecture, represented by the proportion of SNPs in each candidate gene that affects different aspects of the environment (as defined by the environmental response module).

**Figure 2. Environmental response modules for *Pinus contorta*.**

A) Correlations among environments measured by Spearman's ⍴. Abbreviations of the environmental variables can be found in Table 2. B) Hierarchical clustering of associations between allele frequencies (of SNPs in columns) and environments (in rows) measured by Spearman's ⍴. C-F) Each co-association network represents a distinct environmental response module, with color schemes according to the four major groups in the data. Each node is a SNP and is labeled with a number according to its exome contig, and a color according to its module - with the exceptions that modules containing a single SNP are all give the same color within a major group.  G) The pleiotropy barplot, where each bar corresponds to a contig, and the colors represent the proportion of SNPs in each environmental response module. Note that contig IDs are ordered by their environmental response module, and the color of contig-IDs along the x-axis is determined by the environmental response module that the majority of SNPs in that contig cluster with.  Contigs previously identified as undergoing convergent evolution with spruce by Yeaman et al. 2016 are indicated with “\*''. Abbreviations: “Temp": temperature, “Precip": precipitation, “freq": frequency.

**Figure 3. Comparison of linkage disequilibrium (lower diagonal) and recombination rates (upper diagonal) for exome contigs.**

Only contigs with SNPs in the mapping panel are shown. Rows and column labels correspond to Figure 1G. Darker areas represent either high physical linkage (low recombination) or high statistical linkage disequilibrium.

**Figure 4. Overview of galaxy biplots.**

The association between allele frequency and one variable is plotted against the association between allele frequency and a second variable. The Spearman’s *ρ* correlation between the two variables (mean annual temperature or MAT and mean annual precipitation or MAP in this example) is shown in the lower right corner. When the two variables are correlated, genome-wide covariance is expected to occur in the direction of their association (shown with quadrant shading in light grey). The observed genome-wide distribution of allelic effects is plotted in dark grey and the 95% prediction ellipse is plotted as a black line. Because derived alleles were coded as 1 and ancestral alleles were coded as 0, the location of any particular SNP in bivariate space represents the type of environment that the derived allele is found in higher frequency, whereas the location of the ancestral allele would be a reflection through the origin (note only derived alleles are plotted).

**Figure 5. Galaxy biplots for different environmental variables for regular (left column) and structure-corrected (right column) associations.**

Top candidate SNPs are highlighted against the genome-wide background. The internal color of each point corresponds to its environmental response module (as shown in Figure 2 C-F). Top row: mean annual temperature (MAT) vs. mean annual precipitation (MAP), middle row: MAT and Elevation, bottom row: MAT and latitude (LAT).

**Figure 6. Pie charts representing the frequency of derived candidate alleles across the landscape.**

Allele frequency pie charts are overlain on top of an environment that the SNP shows significant associations with. The mean environment for each population is shown by the color of the outline around the pie chart. A) Allele frequency pattern for a SNP from contig 1 in the Multi cluster from Figure 1. The derived allele had negative associations with temperature but positive associations with latitude. B) Allele frequency pattern for a SNP from contig 8 in the Aridity cluster. The derived allele had negative associations with annual:heat moisture index (and other measures of aridity) and positive associations with latitude. SNPs were chosen as those with the highest degree in their submodule.

**Figure 7. Co-association modules mapped to co-expression clusters determined by climate treatments.**

Contig ID, color, and order shown on the bottom correspond to co-association modules plotted in Figure 2. Co-expression clusters from [[55]](https://paperpile.com/c/vNqLuE/3Dcq) are shown at the top.

**Figure 8. Comparison of co-association networks resulting from simulated data for 3 de- mographies.**

A) Isolation by distance (IBD), B) range expansion from a single refuge, and C) range expansion from two refugia. All SNPs were simulated unlinked and 1% of SNPs were simulated under selection to an unmeasured weak latitudinal cline. Boxplots of degree of connectedness of a SNP as a function of its strength of selection, across all replicate simulations (top row). Examples of networks formed by datasets that were neutral-only (middle row) or neutral+selected (bottom row) outlier loci.

Supplementary Tables

**Table S1. Results from GO analysis for all top candidate genes and for each group.**

The top 5 processes are shown for each category. “P” represents the *P*-value from parent-child Fisher test, while "fdr" represents significance after correction for false discovery rate.

**Table S2. Top candidate genes and their annotations.**

For each gene the following information is indicated: the environmental response module ID (“group\_subMod”), the number of outlier SNPs in each of the four major groups (“Multi”, “Aridity”, “Freezing”, or “Geography”), the Gene ID used in the main paper (“NewContigIDMod”), the color used for plotting (“module\_col”), whether or not its homolog shows convergent signals of adaptation with spruce (“is.covergent”), TAIR ID (“tair”), putative gene function (“Annotations”), whether or not the gene was differentially expressed (“diffExp”), and the co-expression cluster (“coexCluster”).

|  |
| --- |
|  |

Supplementary Figures

Figure S1. Histogram of *XTX* estimated from Bayenv2 for all SNPs (top) and for top candidate SNPs (bottom).

Figure S2. Undirected graph network for the Multi group (enlarged version of Figure 2C).

Figure S3. Undirected graph network for the Aridity group (enlarged version of Figure 2D).

Figure S4. Undirected graph network for the Freezing group (enlarged version of Figure 2E).

Figure S5. Undirected graph network for the Geography group (enlarged version of Figure 2F).

Figure S6. Heatmap of structure-corrected allele associations with the environment, analogous to Figure 2B in the main paper. Note that although the pattern is very similar, the magnitude of allele correlations is smaller in the structure-corrected data.

Figure S7. Gametic disequilibrium heatmap. Mean correlation among allele frequencies between top candidate genes. Genes are ordered the same as Figure 2G in the main paper.

Figure S8. Recombination heatmap, clustered by recombination rates. The same data as is shown in Figure 3, except re-clustered by recombination rates to more easily see the patterns of physical linkage.

Figure S9. Loadings of environments onto PC axes. The length and direction of each vector represents the scaled loading of that environmental variable onto the PC axis. The color of each vector represents the mean proportion of variance explained by that environment in the two axes plotted.

Figure S10. Outliers on PC axes. The distribution of Bayes Factors for the association between SNPs and environments along the first three PC axes. Each point is a SNP colored according to its environmental response module in Figure 2C-F. Vertical and horizontal lines represent criteria for significance, and the black ovals represent the 95% prediction ellipse. Note that candidate SNPs all had BF > 2 with at least one univariate environmental variable.

Figure S11. Proportion of exome SNPs falling into various categories for genomic features compared to in the top candidate list. 3primeFLANK: 3’ flanking region; 3primeUTR: 3’ untranslated region; 5primeFLANK: 5’ flanking region; 5primeUTR: 5’ untranslated region; non-tcontig: not located in a transcriptomic contig (intergenic); nonsyn: non-synonymous substitution; unk-adj: unknown adjacent region; unk-flank: unknown flanking region; UNKNOWN-ORF: unknown open reading frame.

Figure S12. Error rates from the simulations given a less stringent criteria (Bonferroni, left) and a more stringent criteria (Bonferroni and Bayes Factors from bayenv2, right). The less stringent criteria was used for the simulations because it had some false positives (A), while the more stringent criteria was used for the empirical data because it didn’t have any false positives (B). While using the more stringent criteria resulted in no false positives, it also reduced the number of true positives (compare C and D), with the most severe reduction under isolation by distance.

Figure S13. Pairwise distances among loci as a function of selection for simulated data. Evaluation of 0.1 as a distance threshold for creating an environmental response module. The three demographies are isolation by distance (IBD), range expansion from one refuge (1R), and range expansion from two refugia (2R). For the simulated data, top candidates were chosen as described in the methods. Multivariate euclidean distance was calculated among the loci based on their associations with environments, and the proportion of pairwise distances above the distance threshold of 0.1 (used for the empirical data) was calculated for each type of comparison. We evaluated four types of pairwise comparisions: neutral loci with each other ("Neut-Neut"), neutral loci with selected loci ("Neut-Sel"), all selected loci with each other ("Sel-Sel"), and only loci under strong selection with each other (*s* > 0.1, "strongSel-strongSel"). A higher proportion of pairwise distances above the threshold indicates that these loci would be more connected to each other in the co-association network.

Figure S14. Examples of networks from simulations. The simulated datasets were nested within randomly generated selective environments, such that different demographic histories were simulated on the same environmental landscape. For this randomly generated environment, loci simulated under stronger selection had a propensity to cluster differently than loci simulated under weaker selection.