Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Core genes driving climate adaptation in plants

Samuel Yeaman (samuel.yeaman@ucalgary.ca)
University of Calgary
James Whiting
University of Calgary
Tom Booker
University of British Columbia
Clement Rougeux
University of Calgary
Brandon Lind
University of Calgary
Pooja Singh
University of Pretoria
Mengmeng Lu
University of Calgary
Kaichi Huang
University of British Columbia https://orcid.org/0000-0002-0378-5988
Michael Whitlock
University of British Columbia https://orcid.org/0000-0002-0782-1843
Sally Aitken
University of British Columbia
Rose Andrew
University of New England
Justin Borevitz
Australian National University
Jeremy J. Bruhl
Botany and N.C.W. Beadle Herbarium, University of New England, New South Wales, Australia
Timothy Collins
Department of Planning and Environment
Martin Fischer
ETH Zurich https://orcid.org/0000-0002-1888-1809
Kathryn Hodgins
Monash University
Jason Holliday
Virginia Tech

Pär K Ingvarsson
Swedish University of Agricultural Sciences https://orcid.org/0000-0001-9225-7521
Jasmine Janes
Vancouver Island University
Momena Khandaker
University of New England
Daniel Koenig
University of California at Riverside https://orcid.org/0000-0002-1037-5346
Julia Kreiner
University of British Columbia
Antoine Kremer
INRAE https://orcid.org/0000-0002-3372-3235
Martin Lascoux
Uppsala University https://orcid.org/0000-0003-1699-9042
Thibault Leroy
University of Vienna
Pascal Milesi
Uppsala Universitet
Kevin Murray
Australian National University
Christian Rellstab
Swiss Federal Research Institute WSL
Loren Rieseberg
University of British Columbia https://orcid.org/0000-0002-2712-2417
Fabrice Roux
INRA
John Stinchcombe
University of Toronto https://orcid.org/0000-0003-3349-2964
Ian R. H. Telford
Botany and N.C.W. Beadle Herbarium, University of New England, New South Wales, Australia
Marco Todesco
University of British Columbia https://orcid.org/0000-0002-6227-4096
Baosheng Wang
South China Botanical Garden, Key Laboratory of Plant Resources Conservation and Sustainable Utilization, Chinese Academy of Sciences https://orcid.org/0000-0002-0934-1659
Detlef Weigel
Max Planck Institute for Biology Tubingen
Yvonne Willi
University of Basel

Stephen Wright
University of Toronto https://orcid.org/0000-0001-9973-9697
Lecong Zhou
Virginia Tech

Biological Sciences - Article

Keywords:

Posted Date: November 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3434061/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Closely-related species often use the same genes to adapt to similar environments[1,2]. However, we know little about why such genes possess increased adaptive potential, and whether this is conserved across deeper evolutionary time. Classic theory suggests a “cost of complexity”: adaptation should occur via genes affecting fewer traits to reduce deleterious side-effects (i.e. lower pleiotropy)[3]. Adaptation to climate presents a natural laboratory to test this theory, as even distantly-related species must contend with similar stresses[4]. Here, we re-analyse genomic data from thousands of individuals from 25 plant species to identify a suite of 108 genes enriched for signatures of repeated local adaptation to climate. This set includes many genes with well-known functions in abiotic stress response, identifying key genes that repeatedly drive adaptation in species as distantly-related as lodgepole pine and Arabidopsis (~ 300 My). Using gene co-expression networks to quantify each gene’s pleiotropy, we find enrichment for greater network centrality/interaction strength and broader expression across tissues (i.e. higher pleiotropy), contrary to the ”cost of complexity” theory. These genes may be particularly important in helping both wild and crop species cope with future climate change, representing a set of important candidates for future study.

INTRODUCTION

Is evolution repeatable? This question, captured by Stephen Jay Gould’s ‘replaying the tape of life’ analogy[5], has been the subject of decades of empirical research (reviewed in [6]). The answer appears to be context-dependent, with variation in repeatability among taxa, populations within species, and genes in the genome, which leads to new questions about which evolutionary forces govern the continuum between identical parallel change and divergent responses[7,8]. The relative importance of deterministic vs. stochastic/contingent explanations has been the subject of great interest, and have been explored extensively at short evolutionary timescales among populations (e.g. in stickleback[9]) or closely-related species (e.g. *Arabidopsis* spp.[2] and *Heliconius* spp.[10]), highlighting predominant roles of the availability of common adaptive variation either through shared inheritance[11–13], gene flow[10] or recurrent mutation[14]. Experimental evolution has further emphasised the inverse associations between genetic repeatability and the complexity of adaptive trait architecture[15]. We know much less, however, about the processes and contingencies shaping genetic repeatability at much longer evolutionary distances, or whether such distances impose hard limits on observing any repeatability at all. The *MC1R* gene provides a textbook example of repeatability across deep time, driving adaptive colour polymorphism in distantly-related vertebrates, from fish to mammoths[16]. Some have speculated the widespread re-use of this gene is key due to minimal interactions with other genes, facilitating its modification whilst incurring minimal pleiotropic disruption[1]. We have little understanding of whether the kinds of factors affecting repeatability for candidate genes such as *MC1R* also generalise to drive repeatability at genome-wide scales, or generalise to more complex adaptive phenotypes beyond simpler traits such as colour.

Climatic variation across the species ranges of plants is a ubiquitous selection pressure among distantly-related species, presenting a natural laboratory to study repeated adaptation. Such variation exerts strong selection pressure for local adaptation and genotypic responses[17], demonstrable through common garden experiments, provenance trials and reciprocal transplants[18]. Numerous candidate genes for drought and thermal tolerance have been identified in *Arabidopsis thaliana*[19,20], *Panicum hallii*[21], and conifers[22]. Plants may use a wide array of strategies to adapt to climate, including phenological shifts, such as flowering earlier or later to avoid frost, or modification of structures, such as smaller leaves to mitigate the effect of air temperature in hotter climates[4]. While such studies have advanced our understanding of climate adaptation within individual species, and have provided examples of a few individual candidate genes with evidence of adaptiveness in multiple species, there has been little combined analysis of whole-genome patterns across many species and large phylogenetic distances. Repeatability in adaptive genetic responses to climate across independently evolving species is expected if biological adaptation is limited to conserved functions, whereas a lack of repeatability might indicate highly polygenic, alternative adaptive strategies to the problem of climate adaptation[23]. The survival of plant populations challenged by anthropogenic climate change may depend upon adaptation, which itself may depend upon the maintenance of adaptive genetic variation within a species' metapopulation through local adaptation[24,25]. Thus, studying repeatability of climate adaptation will give insights into the flexibility of such genetic responses, which will both deepen our understanding of evolution and help predict how species may respond to further changes in the future.

Here, we analyse sequencing data from thousands of individuals from 25 plant species (Fig. 1) across a distribution spanning diverse biomes across four continents (Fig. 1A) to investigate repeatability in the genetic basis of adaptation to climate variation. By processing all raw data through a common pipeline, we compiled by far the largest population genomics dataset to date in terms of phylogenetic breadth and sequencing effort to examine the phenomenon of repeated local adaptation across the genome. We seek to identify genes that exhibit evolutionary associations with climate across multiple species, employing novel statistical analyses tailored specifically to detect repeatable genotype-environment associations. Finally, we explore what properties of these genes may facilitate their repeated evolution and significance for climate adaptation across the plant kingdom.

ASSEMBLING DATASETS AND DEFINING ORTHOLOGY ACROSS SPECIES

Raw sequencing data were downloaded from the SRA/ENA for 29 datasets, covering 25 unique species (Table S1). These data were either individual whole-genome sequencing (WGS), capture-based sequencing (CAPTURE), or pool-sequencing (POOL) data, with a minimum of five sampling locations (see methods for full selection criteria). Data were processed using a common SNP-calling pipeline (Fig S2; see methods), to minimise the influence of bioinformatic technical artefacts. The number of SNPs and individuals sampled per dataset ranged between 173,119 to 23,406,976 and 46 to ~ 1300 respectively (Table S1). To enable comparisons of multiple species, we reconstructed orthology relationships using

OrthoFinder2[27] and classified genes into orthologous sets across species (orthogroups), which could include multiple copies of a gene in a given species due to gene duplication. In total, 44,861 orthogroups were classified, 14,328 were species-specific, and 92.8% of all genes from all genomes were assigned to an orthogroup. Importantly, orthogroups were predominantly characterised by low rates of paralogy (Fig. 1E) and high occupancy (Fig. 1F), with rates of single-copy genes ranging from 39.8%-62.4% (Table S2).

To identify genes with signatures of local adaptation within species, we performed genotype-environment association (GEA) scans by testing the association among per-site allele frequencies and per-site climatic variation (BioClim variables 1–19) taken from worldclim (v2.1, 2.5 minutes)[28] using non-parametric Kendall's Tau correlations (see methods for full details). We also defined and quantified two variables that capture change in local climate (max temperature and precipitation) over the last 50 years. Briefly, these represent the effect size per sampling site of the difference between climate quantified in either the 1960s or 2010s (example in Fig. 1C, see methods). We then combined per-SNP p-values to calculate per-gene p-values using the weighted-Z analysis (WZA)[29] correcting for associations between SNP count and WZA variance. This WZA GEA method exhibits increased power and reduced error for identifying adaptive genes across realistic and extreme spatially-correlated climatic variation compared with other commonly-used methods[29]. To identify orthogroups with repeated signatures of association across multiple species, we applied PicMin[30] (see methods supp) across a focal 8,470 orthogroups with at least 20 species represented for each of the 21 climate variables. These focal orthogroups were slightly more likely to include genes with stronger associations with climate relative to untested orthogroups (see supplementary text). For orthogroups with multiple paralogs within a given species, we include the paralog with the strongest evidence of association to test for repeatability after correction for multiple testing. This approach, therefore, tests for repeated adaptation driven by any member of a gene family.

CLIMATE ADAPTATION INVOLVES REPEATED GENE RE-USE

To cast a broad net and capture a large number of genes driving adaptation to climate, we adopt a False Discovery Rate of 50%, and identify a set of 108 unique orthogroups with evidence of repeated association to climate. While this lenient FDR threshold will include a substantial number of false positives, it is chosen specifically to capture as many true-positives as possible to enable downstream analyses of gene properties (as false positives will simply add noise to any analysis of enrichment). This set is the union from tests across all climate variables, many of which are correlated, representing a collection of 141 significant results at FDR < 0.5. Tests of our method on simulated data yield a median of 36 significant 'repeatable' orthogroups, representing 36 unique orthogroups at FDR < 0.5 (see Supplementary Materials). Indeed, 108 unique orthogroups was greater than any simulated outcome across 1,000 permutations (max = 102). Our results therefore constitute a three-fold enrichment of orthogroups than would be expected by chance (Fig. 2A; permuted $p < 0.001$) for causal loci driving

adaptation in species spanning both gymnosperms and angiosperms (Fig. 2E), despite the ~ 300 million years separating these clades.

These Repeatedly Associated Orthogroups (hereafter, RAOs) were observed across most climate variables, although variation in the number and strength of statistical support varied. Generally, temperature variables exhibited a greater number of RAOs with stronger evidence of repeatability, in particular mean temperature in the warmest month (Fig. 2B). This suggests that the adaptive molecular response to temperature variation across plants may be more repeatable at the level of individual genes, compared with precipitation, which might reflect adaptive constraint or the added complexity of how precipitation interacts with soil to modulate drought effects. Variability in the number of RAOs identified across climate variables for a given species was not linked to inferred GEA power or estimated niche breadth (see supplementary results, Figs S3-S5). For downstream analyses, we focus on two sets of orthogroups: those with FDR < 0.5 to explore general trends in the distribution of RAOs among species and climate variables and to discuss specific candidate genes; and those with a PicMin p-value < 0.005 (based on the top decile cut-off of strongest evidence of repeatability across all climate variables per-orthogroup) within each climate variable to explore general properties of RAOs.

For RAOs (FDR < 0.5), we identified species contributing towards the signature of repeatability on the basis of their per-orthogroup p-value (see methods). Visualising the abundance of these contributions by species and climate variables (Fig. 2C) demonstrates importantly that no specific species, or cluster of species, contributes excessively to the repeatability signatures that we observe. In agreement, we failed to observe any evidence of phylogenetic signal of GEA results in our RAOs with respect to random orthogroups (see supplementary results; Fig S6). This shows that our identification of significant orthogroups is not driven by groups of closely-related species, but rather a signature observed broadly across the phylogenetic tree. This is reinforced by visualising the contribution to repeatability signatures for pairwise contrasts among species (Fig. 2D), where if groups of closely-related species were driving effects, we would expect heat signatures to cluster near the diagonal, which is not observed.

Sixteen orthogroups were associated with repeatability across multiple climate variables (Fig S3). Most strikingly, the orthogroup including the *A. thaliana* genes *PRR3* and *PRR7*, with instrumental functional roles in circadian rhythm[31] and flowering time[32], was repeatedly adaptive across 10 of 21 variables (most strongly with mean temperature in the dry quarter, FDR = 0.102) and across multiple variables within the same species (difference between red and blue bars in Fig. 2E; Fig S7). The role of this gene family in circadian rhythm may contribute towards its repeated association with multiple climate variables if these also vary with latitude. Another notable RAO includes a family of four ubiquitin-related proteins (*RUB1/ RUB3/ UBQ7/ AT1G11970*), which was associated with six climate variables and most strongly with minimum temperature in the coldest month (FDR = 0.051). The RUB-conjugation pathway has been implicated in the auxin response, embryo development and growth[33]. A summary of genes in RAOs that are associated with phenotypes known to be adaptive under climatic stress[34] is shown in Table 1 (full details of RAO gene contents in Table S3-S5)

Table 1
Summary of known climatic adaptive phenotypes associated with RAOs. Each line represents the A. thaliana paralogs within a given RAO.

Adaptive Phenotype	***Genes (A. thaliana paralogs)***
Flowering time, development, and photoperiodism	*ATC/ TSF/ TFL1/ BFT/ FT*
	AT4G04260/ EBS/ SHL1
	PDP5/ PDP2
	AHBP-1B/ BF5/ TGA6/ PAN
	ELF9
	GIGANTEA
	BRD1/ BRD2
	RVE1/ RVE2/ AT3G10113/ EPR1
	GPRI1/ GLK2
	CUL4
Circadian rhythm	*PRR7/ PRR3*
	ATH13
	RVE1/ RVE2/ AT3G10113/ EPR1
	GIGANTEA
Auxin signaling	*GH3.1/ GH3.3/ GH3.9/ WES1/ BRU6/ DFL1/ GH3.17/ GH3.4*
	UBQ7/ AT1G11970/ RUB3/ RUB1
	IAA33
	PEX5
	ABCG33/ ABCG41/ ABCG30/ ABCG42/ ABCG43/ ABCG37
	RVE1/ RVE2/ AT3G10113/ EPR1
Salicylic Acid (SA) signaling	*AHBP-1B/ OBF5/ TGA6/ PAN*
Abscisic Acid (ABA) signaling	*CIPK3/ CIPK8/ CIPK26/ SOS2/ CIPK9/ CIPK23, SPP1*
Seed dormancy and vegetative timing	*DRG*
	RPN8A/ MEE34
	GPS1
	AT4G04260/ EBS/ SHL1

Adaptive Phenotype	*Genes (A. thaliana paralogs)*
	POD1
	CPN60A/ Cpn60alpha2
Root growth and development	*ZP1*
	ATSAC1B/ RHD4/ SAC8
	ABCB24/ ABCB23/ ABCB25
	ABCG33/ ABCG41/ ABCG30/ ABCG42/ ABCG43/ ABCG37
	BCHA2/ SPI
	WAV2
	GH3.1/ GH3.3/ GH3.9/ WES1/ BRU6/ DFL1/ GH3.17/ GH3.4
	HRD/ AT5G52020/ AT1G12630, ANL2/ HDG1
Cold- and thermo-tolerance	*CIPK3/ CIPK8/ CIPK26/ SOS2/ CIPK9/ CIPK23*
	HCF106
	GIGANTEA
	HSF4
Salt stress	*SOS1/ NHX8, ANL2/ HDG1*
	CIPK3/ CIPK8/ CIPK26/ SOS2/ CIPK9/ CIPK23
	ATSAC1B/ RHD4/ SAC8

The functions of these genes reflect the expected responses to climatic variation, such as phenological avoidance of drought or frost through changes to flowering time or seed dormancy[4], or modification to root hair number and structure in response to temperature changes[35]. Changes to growth or stomatal function through hormone signaling[36] meanwhile may facilitate tolerance to drought by reducing water loss, and genes associated with salt stress may be involved in surviving salt accumulation in soils due to aridity.

We found only a single RAO associated with our two climate change variables at FDR < 0.5; harbouring the *A. thaliana* genes *ATKPNB1*, *AT3G08943* and *AT3G08947*. *ATKPNB1* is sensitive to abscisic acid and is involved in drought tolerance through stomatal closure[37]. The limited number of RAOs here likely reflects the relatively short amount of time that our climate change variables are calculated over (~ 50 years), and the limited time to respond to selection subsequently, particularly in longer-lived species.

REPEATED ADAPTATION ACROSS ORTHOGROUPS WITH SIMILAR FUNCTIONS

Repeated adaptation may also happen beyond the gene level, where multiple genes from within the same molecular pathway are used for adaptation across multiple species. Examples of this 'functional repeatability' have been documented in adaptation to whole-genome duplication in *Arabidopsis*[38] and highland adaptation in maize[39]. To explore this phenomenon, we used the STRING database[40] (v11.5) to provide a network-based representation of protein-protein interactions (direct and indirect associations compiled from genomic context, experimental evidence such as co-expression, and text-mining of literature), and tested whether RAOs formed networks with more interactions than expected (see methods). We grouped RAOs all together and into temperature- and precipitation-related groups, and tested each group to see whether RAOs as a group contained genes that were more likely to interact with one another than random orthogroups.

Each group of RAOs tended to include more protein-protein interactions than random orthogroups, and this was particularly clear for RAOs identified through precipitation variables (permuted p-value = 0.015) (Fig. 3A). Gene-ontology (GO) enrichment over the same groups of RAOs reflected this, as precipitation-related RAOs were highly enriched for biological processes with clear adaptive roles (Fig. 3B-C; Table S6). orthogroups associated with 'root hair tip' were enriched across all RAOs, and temperature-RAOs were enriched for several processes, notably 'regulation of photoperiodism, flowering' and 'brassinosteroid biosynthetic process' which are known to be involved in thermotolerance[4,41]. These results suggest that independently identified RAOs contain genes involved in similar functional processes, which implies that repeated adaptation is occurring beyond the level of the gene. This may occur because a given adaptive response requires the coordinated modification of multiple functionally-related genes in all species involved. Alternatively, this signal could also be driven by different subsets of functionally-related genes contributing to adaptation in different subsets of species (e.g. if genes A, B, C, and D are functionally related, species 1, 2, and 3 adapt via genes A and D while species 4, 5, and 6 adapt via genes B and D). In either case, our results show that certain pathways or functional groups of genes are particularly important for adaptation to these climatic stressors, particularly with regards to adaptation to precipitation variation.

REPEATABILITY IS ASSOCIATED WITH INCREASED PLEIOTROPY

We next ask whether the genes we identified that contribute to repeated adaptation tend to share particular characteristics with respect to their degree of pleiotropy. Pleiotropy is a fundamental attribute of a gene describing the number of traits it affects. Based on Fisher's model of universal pleiotropy[42], the 'Cost of Complexity' hypothesis[3] posits a reduced adaptive potential for genes with greater pleiotropy, as constraint increases with organismal 'complexity'. In keeping with this, greater fitness consequences are predicted by the degree of pleiotropy in yeast[43]. However, empirical evidence from mice, nematodes and yeast suggests this cost may be counteracted by a greater mutational effect size per-trait observed for genes with greater pleiotropy[44,45]. In line with this, Rennison and Peichel[46] found genes repeatedly involved in stickleback adaptation exhibited elevated levels of pleiotropy. To test the importance of

pleiotropy by multiple definitions in our dataset, we used public databases of gene expression for *A. thaliana* and *Medicago truncatula* genes extracted from Expression Atlas[47] and ATTED-II[48]. We explored pleiotropy by two definitions, tissue specificity[49] and condition-independent co-expression with other genes[50]. Tissue specificity of gene expression is inversely associated with pleiotropy, has previously been linked to increased rates of evolution[51], and was estimated here according to the τ metric[52] (Fig. 4A). τ describes tissue specificity, the inverse of pleiotropy, so to avoid confusion we will describe changes to the breadth of expression, which we define as lower τ. Contrary to the 'Cost of Complexity' prediction, we found that RAOs with the strongest evidence of repeatability were strongly associated with increased expression breadth ($p = 5.44e^{-4}$), and expression breadth tended to decrease in subsets of orthogroups with increasingly weaker evidence of repeatability, such that orthogroups with the weakest evidence of repeatability were enriched for genes with high specificity ($p = 4.74e^{-6}$; Fig. 4C).

Alternatively, pleiotropic constraint can be considered as various node centrality statistics within a gene coexpression network (Fig. 4B), where the 'distance' between two gene nodes is lower when co-expression of those genes increases (e.g. across experimental treatments or tissue types). Node degree and strength describe the number and summed weights of interactions for a given node, betweenness describes a gene's role in bridging subnetworks, while closeness represents nodes with the shortest distance (edge length) to all other nodes. All of these centrality measures have been inversely linked to rate of evolution in several eukaryotic protein-protein networks, and changes to genes with high centrality are more likely lethal[53,54], indicative of evolutionary constraint. However, in contrast to these negative associations with evolvability, we observed clear positive associations between evidence of repeatability and co-expression centrality for node closeness, degree, and strength across both co-expression networks (Fig. 4C; node betweenness was not significantly associated with evidence of repeatability in the *A. thaliana* network). Similar to results for specificity of expression, centrality was significantly greater ($p < 0.05$) in orthogroups with the strongest evidence of repeatability and significantly lower ($p < 0.05$) in orthogroups with the weakest evidence. These clear trends highlight a robust association between increased pleiotropy and evidence of adaptive repeatability across all tested orthogroups.

The association we find between repeatability of local adaptation and increased pleiotropy stands in apparent contrast to previous findings of increased contributions to adaptation by genes with reduced pleiotropy[51,53,54]. This difference may arise because of the scale at which adaptation is occurring; here we have focused on local adaptation, which involves a tension between migration and spatially divergent selection that tends to favour the contribution of alleles of large effect, as they can overcome migration swamping[23,55]. As the phenotypic effect size of mutations tends to increase with pleiotropy [44,45], increased pleiotropy may therefore be favoured in local adaptation. By contrast, when a species adapts to a temporal change in environment across its whole range, there is no tension between migration and selection and no additional advantage for alleles of larger effect[23]. This may explain the reduced pleiotropy previously observed in rapidly evolving genes[51,53,54]. Our results suggest a robust impact of pleiotropy on local adaptation across multiple plant species, consistent with similar observations in stickleback[46], ragweed[56], and *A. thaliana*[57]. It is unknown whether the association between pleiotropy

and repeatability is monotonic, or if intermediate pleiotropy promotes repeatability and extreme pleiotropy remains constraining, as suggested by[46,57]. In the *A. thaliana* tissue specificity and node degree data, RAO sets were diminished for the least pleiotropic genes, but also enriched for the most pleiotropic (Fig S8). It is important to make clear that we cannot rule out an association between repeatability and pleiotropy due to an increased likelihood of detecting large-effect alleles with GEA methods. A comparable analysis centred on global adaptation through selective sweeps could distinguish these, as selective sweep methods are similarly biassed towards large-effect alleles but there is no assumption of a biological role for increased pleiotropy facilitating global adaptation at the species-level[3,42].

We were also interested in whether pleiotropy enrichment was isolated to specific climate variables. We therefore repeated enrichment analyses for orthogroups exhibiting the strongest evidence of repeatability within climate variables (PicMin p-value < 0.005) for tissue expression specificity and *A. thaliana* node degree (Fig. 4D). The majority of these sets of orthogroups exhibited elevated pleiotropy. Strikingly, orthogroups with the strongest evidence of repeatability associated with our climate change variables were highly enriched ($p < 0.05$) for pleiotropic genes by both measures. Given these variables only capture environmental change over ~ 50 years, the genes with the strongest evidence of repeatability associated with these variables may be highly pleiotropic due to genes with greater effect sizes facilitating rapid adaptation to shifting fitness optima[58].

CONCLUDING REMARKS

We demonstrated repeatability in the genetic basis of local adaptation to climate variation across plant species separated by > 300 million years of evolution. The gene families identified here will be of significant adaptive importance as climates change and plants find themselves under increased climatic selection. Our results demonstrate that the adaptive responses to such changing selection involves conserved, core functional responses. We have also shown that, contrary to expectations that pleiotropic constraint impedes adaptation, the genes identified as repeatedly adaptive bear signatures of increased pleiotropy. These results support the model of pleiotropy driving increasing phenotypic effect sizes, and such large effects being a major driver of local adaptation in the genome. Whether these results extend beyond the dominant biomes of boreal and temperate forest found in the Global North to species in the tropics and Global South will be of pertinent interest as global sequencing effort continues to increase.

Declarations

ACKNOWLEDGEMENTS

We would like to thank Outi Savolainen, Tuomas Hämälä, Tom Mitchell-Olds, David Lowry, Matias Kirst, Peter Tiffin, Shosei Kubota and Alex Widmer who were involved in the collection and preservation of data used in this study. Funding was provided by NSERC Discovery and Alberta Innovates, with computational resources and support provided by the Digital Research Alliance of Canada.

DATA AVAILABILITY

All scripts for SNP-calling and analyses are available and documented at https://github.com/JimWhiting91/RepAdapt. The SRA codes for all raw sequencing data are in Table S1. All VCFs will be made available on dryad.

References

1. Mundy, N. I. A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proc. Biol. Sci.***272**, 1633–1640 (2005).
2. Bohutínská, M. *et al.* Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proc. Natl. Acad. Sci. U. S. A.***118**, (2021).
3. Orr, H. A. Adaptation and the cost of complexity. *Evolution***54**, 13–20 (2000).
4. James, M. E., Brodribb, T., Wright, I. J., Rieseberg, L. H. & Ortiz-Barrientos, D. Replicated Evolution in Plants. *Annu. Rev. Plant Biol.* (2022) doi:10.1146/annurev-arplant-071221-090809.
5. Gould, S. J. *Wonderful life: the Burgess Shale and the nature of history*. (WW Norton & Company, 1990).
6. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science***362**, (2018).
7. Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J. & Stuart, Y. E. (Non)Parallel Evolution. *Annu. Rev. Ecol. Evol. Syst.***49**, 303–330 (2018).
8. Arendt, J. & Reznick, D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.***23**, 26–32 (2008).
9. Magalhaes, I. S. *et al.* Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nat Ecol Evol* (2020) doi:10.1038/s41559-020-01341-8.
10. Montejo-Kovacevich, G. *et al.* Repeated genetic adaptation to altitude in two tropical butterflies. *Nat. Commun.***13**, 4676 (2022).
11. Waters, J. M. & McCulloch, G. A. Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. *Mol. Ecol.***30**, 4162–4172 (2021).
12. Konečná, V. *et al.* Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. *Nat. Commun.***12**, 4979 (2021).
13. Louis, M. *et al.* Selection on ancestral genetic variation fuels repeated ecotype formation in bottlenose dolphins. *Sci Adv***7**, eabg1245 (2021).
14. Zou, D. *et al.* Vulture genomes reveal molecular adaptations underlying obligate scavenging and low levels of genetic diversity. *Mol. Biol. Evol.* (2021) doi:10.1093/molbev/msab130.
15. Long, A., Liti, G., Luptak, A. & Tenaillon, O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat. Rev. Genet.***16**, 567–582 (2015).

16. Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B. & Hoekstra, H. E. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.***365**, 2439–2450 (2010).
17. Exposito-Alonso, M. *et al.* Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature***573**, 126–129 (2019).
18. Lortie, C. J. & Hierro, J. L. A synthesis of local adaptation to climate through reciprocal common gardens. *J. Ecol.***110**, 1015–1021 (2022).
19. Monroe, J. G. *et al.* Drought adaptation in Arabidopsis thaliana by extensive genetic loss-of-function. *Elife***7**, (2018).
20. Exposito-Alonso, M. *et al.* Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology & Evolution***2**, 352–358 (2017).
21. Lovell, J. T. *et al.* The genomic landscape of molecular responses to natural drought stress in Panicum hallii. *Nat. Commun.***9**, 5213 (2018).
22. MacLachlan, I. R. *et al.* Genome-wide shifts in climate-related variation underpin responses to selective breeding in a widespread conifer. *Proc. Natl. Acad. Sci. U. S. A.***118**, (2021).
23. Yeaman, S. Evolution of polygenic traits under global vs local adaptation. *Genetics***220**, (2022).
24. Anderson, J. & Song, B.-H. Plant adaptation to climate change - Where are we? *J. Syst. Evol.***58**, 533–545 (2020).
25. Christmas, M. J., Breed, M. F. & Lowe, A. J. Constraints to and conservation implications for climate change adaptation in plants. *Conserv. Genet.***17**, 305–320 (2016).
26. Kubota, S. *et al.* A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild Arabidopsis Species. *PLoS Genet.***11**, e1005361 (2015).
27. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.***20**, 238 (2019).
28. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.***37**, 4302–4315 (2017).
29. Booker, T. R., Yeaman, S., Whiting, J. R. & Whitlock, M. C. The WZA: A window-based method for characterizing genotype-environment association. *Mol. Ecol. Resour.* (2023) doi:10.1111/1755-0998.13768.
30. Booker, T. R., Yeaman, S. & Whitlock, M. C. Using genome scans to identify genes used repeatedly for adaptation. *bioRxiv* 2022.03.24.485690 (2022) doi:10.1101/2022.03.24.485690.
31. Liu, T., Carlsson, J., Takeuchi, T., Newton, L. & Farré, E. M. Direct regulation of abiotic responses by the Arabidopsis circadian clock component PRR7. *Plant J.***76**, 101–114 (2013).
32. Guan, J. *et al.* Genomic analyses of rice bean landraces reveal adaptation and yield related loci to accelerate breeding. *Nat. Commun.***13**, 5707 (2022).
33. Bostick, M., Lochhead, S. R., Honda, A., Palmer, S. & Callis, J. Related to ubiquitin 1 and 2 are redundant and essential and regulate vegetative growth, auxin signaling, and ethylene production in

Arabidopsis. *Plant Cell***16**, 2418–2432 (2004).

34. Gray, S. B. & Brady, S. M. Plant developmental responses to climate change. *Dev. Biol.***419**, 64–77 (2016).
35. Calleja-Cabrera, J., Boter, M., Oñate-Sánchez, L. & Pernas, M. Root Growth Adaptation to Climate Change in Crops. *Front. Plant Sci.***11**, 544 (2020).
36. Acharya, B. R. & Assmann, S. M. Hormone interactions in stomatal function. *Plant Mol. Biol.***69**, 451–462 (2009).
37. Luo, Y. *et al.* An Arabidopsis homolog of importin β1 is required for ABA response and drought tolerance. *Plant J.***75**, 377–389 (2013).
38. Bohutínská, M. *et al.* Novelty and convergence in adaptation to whole genome duplication. *Mol. Biol. Evol.* (2021) doi:10.1093/molbev/msab096.
39. Wang, L. *et al.* Molecular Parallelism Underlies Convergent Highland Adaptation of Maize Landraces. *Mol. Biol. Evol.* (2021) doi:10.1093/molbev/msab119.
40. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.***49**, D605–D612 (2021).
41. Li, N. *et al.* Plant Hormone-Mediated Regulation of Heat Tolerance in Response to Global Climate Change. *Front. Plant Sci.***11**, 627969 (2020).
42. Fisher, R. A. *The genetical theory of natural selection*. vol. 272 (Clarendon Press, 1930).
43. Cooper, T. F., Ostrowski, E. A. & Travisano, M. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evolution***61**, 1495–1499 (2007).
44. Wang, Z., Liao, B.-Y. & Zhang, J. Genomic patterns of pleiotropy and the evolution of complexity. *Proc. Natl. Acad. Sci. U. S. A.***107**, 18034–18039 (2010).
45. Wagner, G. P. *et al.* Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature***452**, 470–472 (2008).
46. Rennison, D. J. & Peichel, C. L. Pleiotropy facilitates parallel adaptation in sticklebacks. *Mol. Ecol.***31**, 1476–1486 (2022).
47. Papatheodorou, I. *et al.* Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.***46**, D246–D251 (2018).
48. Obayashi, T., Hibara, H., Kagaya, Y., Aoki, Y. & Kinoshita, K. ATTED-II v11: A Plant Gene Coexpression Database Using a Sample Balancing Technique by Subagging of Principal Components. *Plant Cell Physiol.***63**, 869–881 (2022).
49. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.***51**, 1339–1348 (2019).
50. Proulx, S. R., Promislow, D. E. L. & Phillips, P. C. Network thinking in ecology and evolution. *Trends Ecol. Evol.***20**, 345–353 (2005).

51. Mack, K. L., Phifer-Rixey, M., Harr, B. & Nachman, M. W. Gene Expression Networks Across Multiple Tissues Are Associated with Rates of Molecular Evolution in Wild House Mice. *Genes* **10**, (2019).
52. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics***21**, 650–659 (2005).
53. Hahn, M. W. & Kern, A. D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.***22**, 803–806 (2005).
54. Mähler, N. *et al.* Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.***13**, e1006402 (2017).
55. Lenormand, T. Gene flow and the limits to natural selection. *Trends Ecol. Evol.***17**, 183–189 (2002).
56. Hämälä, T., Gorton, A. J., Moeller, D. A. & Tiffin, P. Pleiotropy facilitates local adaptation to distant optima in common ragweed (Ambrosia artemisiifolia). *PLoS Genet.***16**, e1008707 (2020).
57. Frachon, L. *et al.* Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time. *Nat Ecol Evol***1**, 1551–1561 (2017).
58. Jain, K. & Stephan, W. Rapid Adaptation of a Polygenic Trait After a Sudden Environmental Shift. *Genetics***206**, 389–406 (2017).

Figures

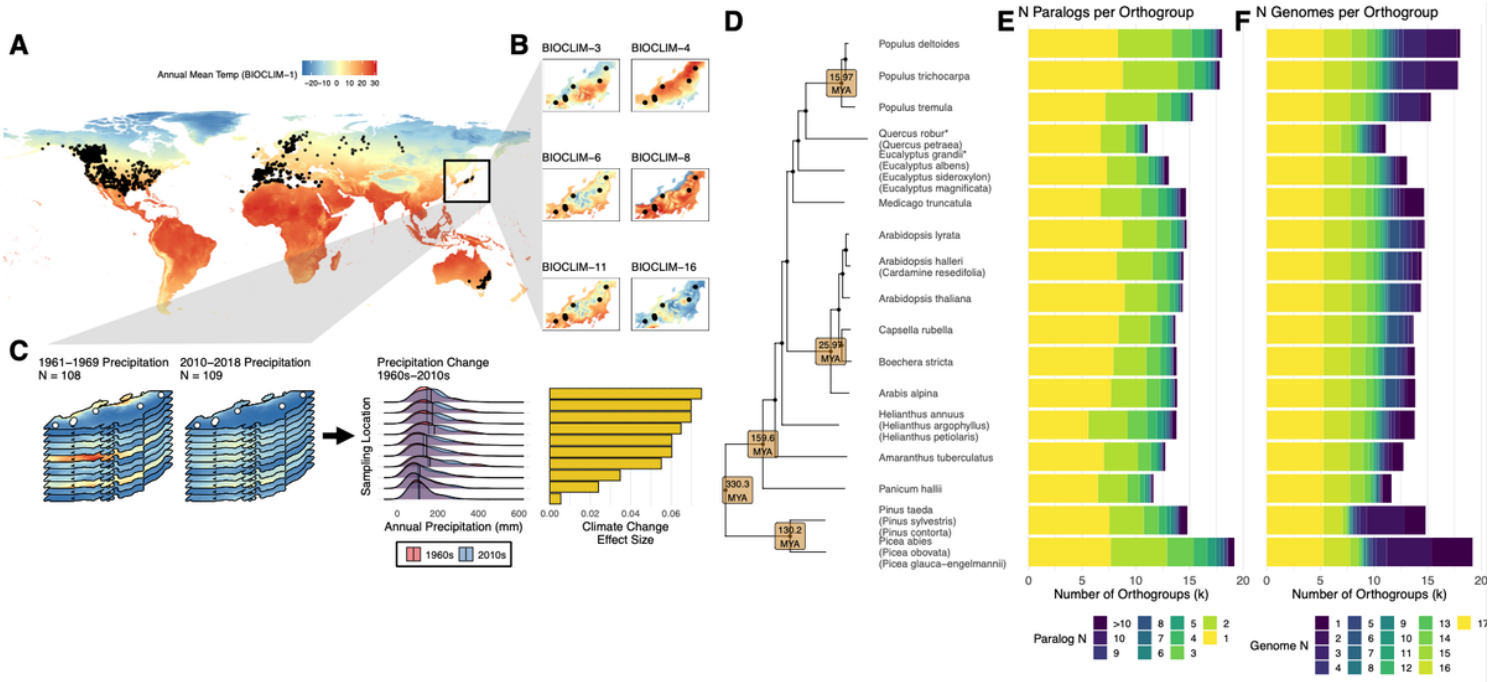

Figure 1

Summary of study design and orthology assignment among species. Sampling locations from all 29 datasets (25 species) relative to global annual mean temperature (A). An example of six bioclim variables across a single dataset with ten sampling locations from Kubota et al.[26] (B). Panel C shows an example for the same dataset of how climate change variables were calculated, taking monthly climate data from

the 1960s and 2010s and taking the effect size of the difference between decades. The species tree derived from the OrthoFinder2 analysis for the 17 reference genomes used in this study is shown in D; species that provided a reference genome but were not included in analyses are marked with asterisks and the age of high-confidence nodes is shown with values pulled from TimeTree (Fig S1). Species analysed here are placed according to the reference genome used and the reference's position in the tree. The per-genome distribution of the number of paralogs per orthogroup (E) and number of representative genomes per orthogroup (F) are depicted as stacked bars, with bar fill scaling from light-to-dark according to most-to-least desirable value.

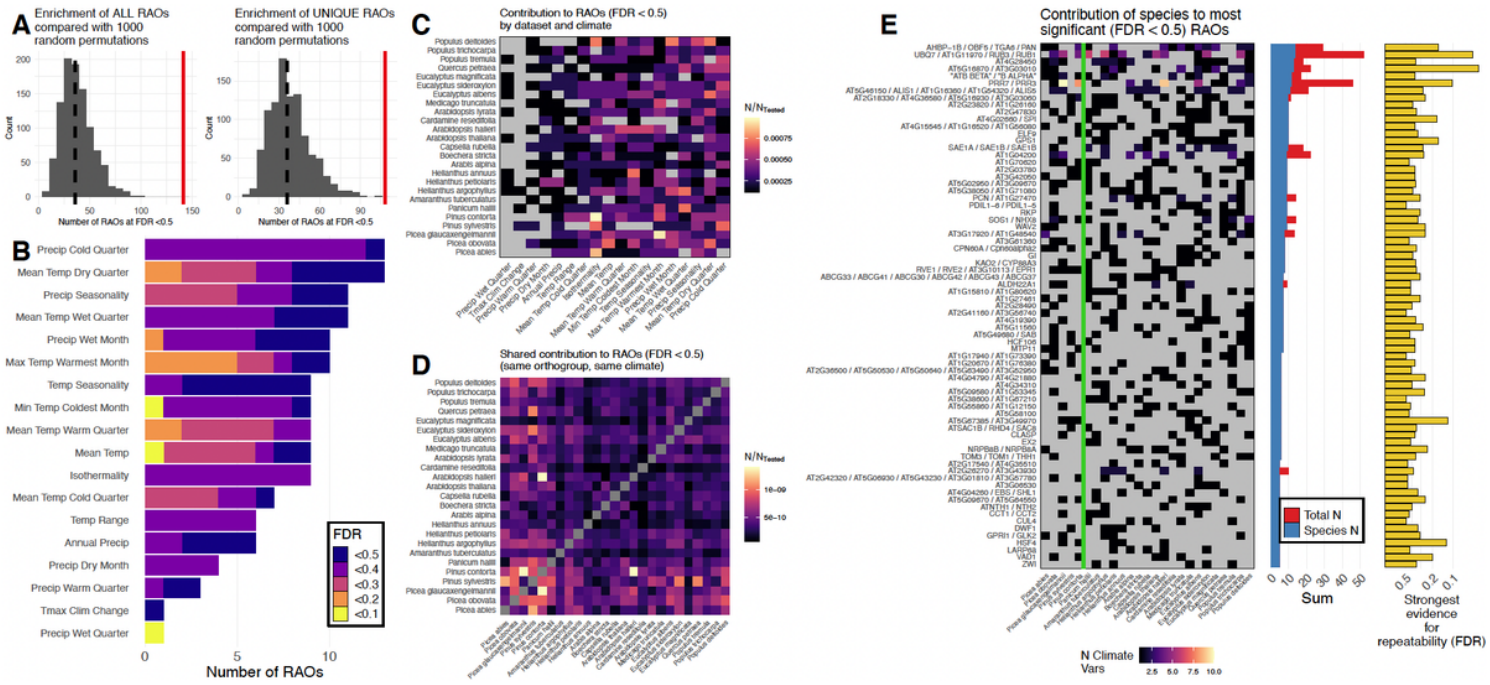

Figure 2

Signatures of repeatability for 8,470 orthogroups across 21 climate variables and 25 species. Enrichment of the total and unique number of RAOs observed at FDR <0.5 relative to random permutations is shown in A. The number of orthogroups exhibiting significant repeatability at different FDR thresholds per climate variable is shown in panel B. Heatmaps show the contribution of individual species to orthogroup repeatability at FDR < 0.5 for different climate variables (C) and among pairs of species (D), with species ordered phylogenetically. In each case, the fill of each cell represents the proportion of orthogroups where a given species contributes towards the signature of repeatability based on its minimum GEA p-value. The number of times a species contributes towards repeatability for a given RAO (FDR < 0.5) is also shown (E). In this panel, cell fill denotes the number of climate variables where a species contributes a low p-value to a given RAO, with grey = 0. The vertical green line separates gymnosperms and angiosperms. Row-wise summations are shown as bars, where the blue bar shows the number of species associated with a given orthogroup (Species N), and the red bar shows the total number of species and climate variables (Total N). The strongest statistical support for repeatability is also shown per-orthogroup as -log10-transformed FDR. Only 73 RAOs with at least 5 contributing species are displayed.

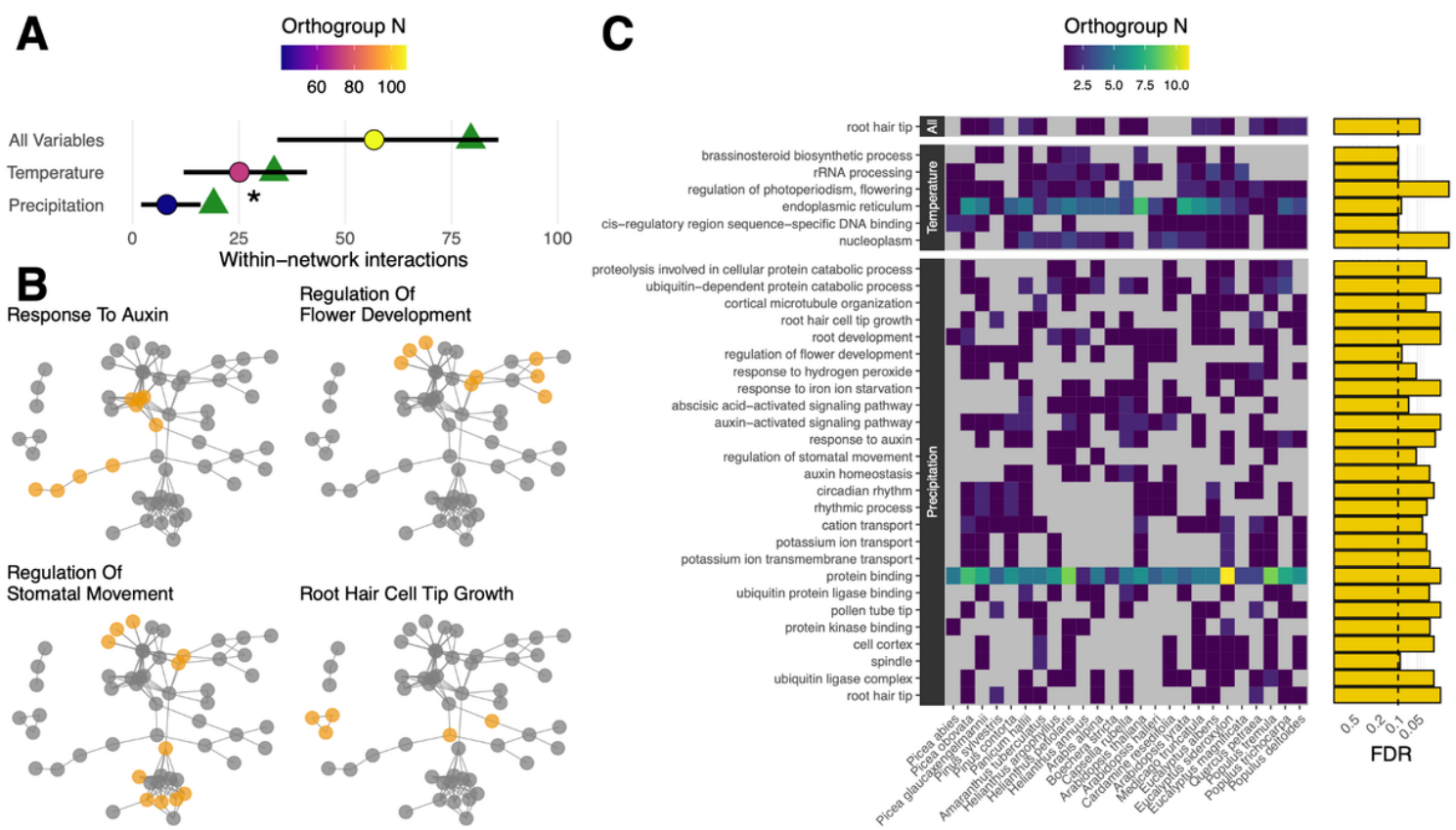

Figure 3

Enrichment of functional interactions among RAOs. (A) Mean number of STRING interactions within single-gene-per-orthogroup networks based on grouped RAOs (triangles) compared to results from permutation tests. Circles and lines represent the mean and 5/95% quantiles STRING interactions among 10,000 random gene sets sampled for single genes in the same way. The colour of circles shows the number of orthogroups in each group of RAOs. (B) The network derived from all genes in precipitation-related RAOs, with orange highlighting indicating which genes are members of 4 enriched GO terms. (C) Enriched GO terms (FDR < 0.1) showing the number of contributing orthogroups from each species, with FDR-adjusted significance shown as adjacent bars. GO terms are ordered on the y-axis based on semantic similarity clustering, to group GO terms associated with similar genes together. Cell fill shows the number of orthogroups per species and GO term, with grey = 0.

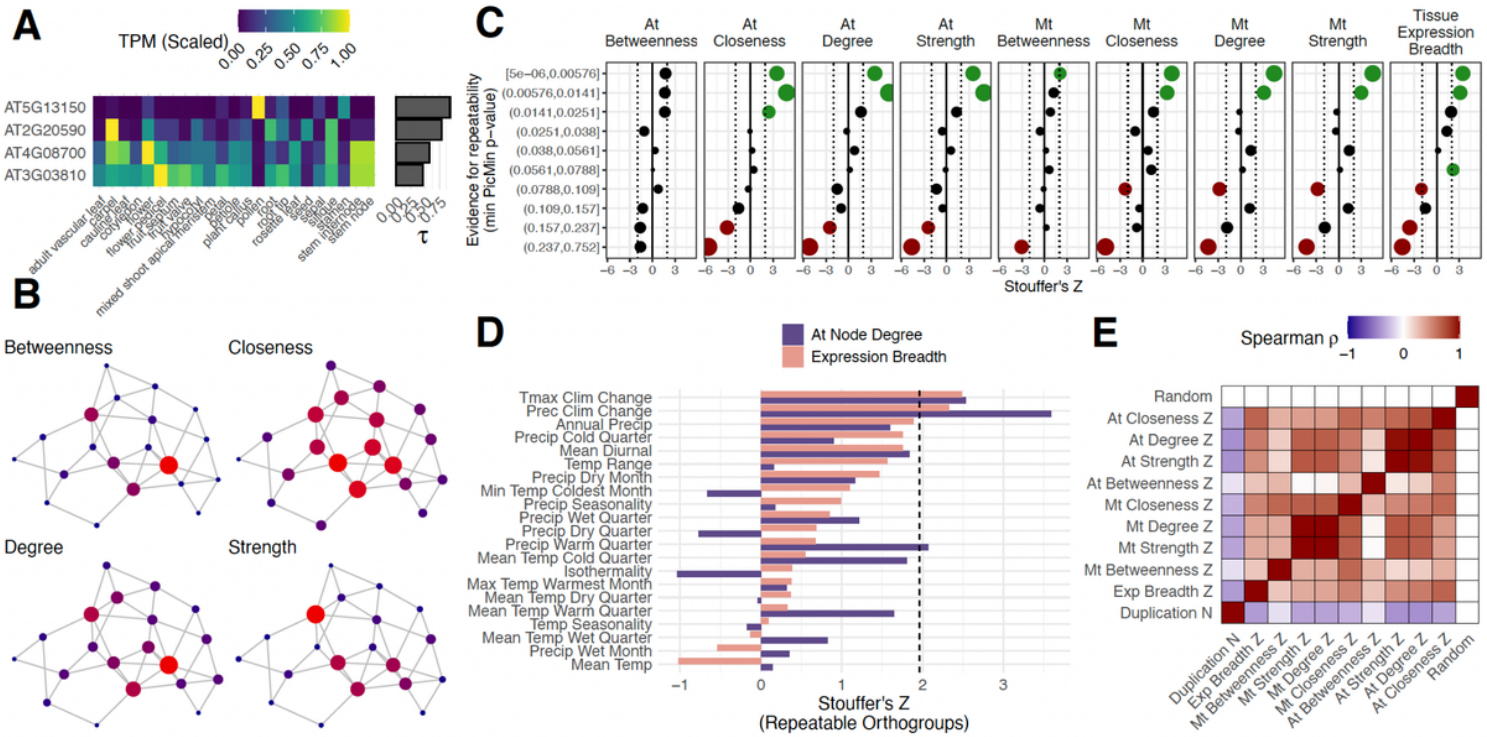

Figure 4

Associations between pleiotropy and orthogroup repeatability. Panels A and B show examples of the variables considered, breadth of expression across tissues (A) and gene co-expression (B). In panel A, five A .thaliana genes are shown with their tau statistic of tissue specificity (bars). Each cell in the heatmap shows tissue-specific gene expression scaled by the maximum across all tissues. Panel B shows an example network of 20 genes with four centrality measures calculated across nodes. Node size and colour (small-blue = low, large-red = high) denotes centrality by each metric. Panels C and D show pleiotropy enrichment in RAO. Panel D shows Stouffer's Z calculated for orthogroups grouped into deciles on the basis of the strongest evidence of repeatability across all climate variables (green = p < 0.05 more pleiotropic than expected, red = p < 0.05 less pleiotropic than expected). Panel D shows Stouffer's Z for node degree and tissue specificity metrics within climate variables, focussing on orthogroups with at least one PicMin p-value < 0.005. Panel E shows non-parametric correlations of per-orthogroup pleiotropy and duplication metrics, along with a randomised control variable.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- repadaptmethodsandsuppsvNature.docx