

CONVERGENT EVOLUTION

Convergent local adaptation to climate in distantly related conifers

Sam Yeaman,^{1,2*} Kathryn A. Hodgins,^{3*} Katie E. Lotterhos,⁴ Haktan Suren,⁵ Simon Nadeau,² Jon C. Degner,² Kristin A. Nurkowski,³ Pia Smets,² Tongli Wang,² Laura K. Gray,⁶ Katharina J. Liepe,⁶ Andreas Hamann,⁶ Jason A. Holliday,⁵ Michael C. Whitlock,⁷ Loren H. Rieseberg,⁸ Sally N. Aitken^{2†}

When confronted with an adaptive challenge, such as extreme temperature, closely related species frequently evolve similar phenotypes using the same genes. Although such repeated evolution is thought to be less likely in highly polygenic traits and distantly related species, this has not been tested at the genome scale. We performed a population genomic study of convergent local adaptation among two distantly related species, lodgepole pine and interior spruce. We identified a suite of 47 genes, enriched for duplicated genes, with variants associated with spatial variation in temperature or cold hardiness in both species, providing evidence of convergent local adaptation despite 140 million years of separate evolution. These results show that adaptation to climate can be genetically constrained, with certain key genes playing nonredundant roles.

Evolutionary convergence has provided a window into the constraints that shape adaptation (1, 2). Studies of convergent local adaptation among closely related lineages commonly find evidence of many shared genetic changes (3), but such evidence may be a result of shared standing variation, rather than shared constraints in how genotypes give rise to phenotypes (4–6). Across time scales where shared standing variation is precluded, adaptation sometimes arises by mutations in the same genes, such as melanism via *Mclr* and *agouti* (7). However, such examples of adaptation via large-effect loci may not be representative of the true spectrum of phenotypes (8). Highly polygenic traits may have greater genetic redundancy than traits governed by a single molecular pathway, and might therefore exhibit less repeatable signatures of adaptation (9). Relatively little is known about the genome-wide repeatability of local adaptation in more highly polygenic traits in distantly related species, where shared standing variation is precluded.

We compared signatures of local adaptation in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca*, *Picea engelmannii*, and their hybrids), which inhabit similar environmental gradients across montane and boreal regions of western

North America, and last shared a common ancestor more than 140 million years ago (10). Like many conifers, these species show patterns of local adaptation to climate that reflect a tradeoff between competition for light resources and acquisition of freezing tolerance (11, 12). Although some candidate genes have been identified that may drive these phenotypic responses (13, 14), we still know little about the genomic basis of adaptation. Comparative gene expression studies indicate that plastic responses to temperature and moisture are highly conserved in spruce and pine, with ~70% of differentially expressed orthologs showing parallel responses in both species (15) and lower rates of protein evolution (16). Plastic responses to climate appear to be relatively conserved and highly polygenic, but the extent to which local adaptation involves similar responses at the genomic level is unknown.

To characterize the basis of adaptation in these large genomes (~20 Gb), we sampled individuals from >250 populations across their geographic ranges and identified more than 1 million single-nucleotide polymorphisms (SNPs) in ~23,000 genes (17). We searched for correlations between individual SNPs and (i) 17 phenotypes measured

in growth chambers [genotype-phenotype association (GPA)] and (ii) 22 environmental variables [genotype-environment association (GEA)]. We identified top-candidate genes as those with an exceptional proportion of their total SNPs being GPA or GEA outliers (99th percentile) (17) (Fig. 1A).

The strongest phenotypic signatures of local adaptation to climate were for correlations between fall and winter cold injury traits and low-temperature stress-related environmental factors, including latitude (12) (Fig. 1B). The strength of these correlations was similar in pine and spruce, providing evidence of convergent phenotypic local adaptation. These two phenotypic traits and five environmental factors (hereafter the “main variables”) also showed the strongest signatures of selection, with the largest number of top-candidate genes in both species (Fig. 1, C and D) and greatest mean strength of association (ρ^2) across all SNPs (fig. S1). Although these results suggest that adaptation to climate is highly polygenic, not all variables had similar genomic signatures in both species. Many top GEA candidates were found for longitude in pine but not spruce, whereas the converse was true for precipitation falling as snow (Fig. 1D), indicating that these species are divergent in some aspects of adaptation (17).

To study the repeatability of local adaptation on a gene-by-gene basis, for each gene identified as a top candidate for at least one of the seven main variables in one species, we examined the strength of associations in orthologous gene(s) in the other species (Fig. 2). To quantify similarity in signatures of association underlying convergent adaptation (hereafter “signatures of convergence”), we compared the strength of association (ρ^2) for all SNPs within each of these top-candidate orthologs to a null distribution constructed from all non-top-candidate orthologs [which we term the “null-*W* method” (17)]. For the one-to-one orthologs, 22.3 to 27.5% of tested orthologs (spruce) and 5.7 to 11.6% of tested orthologs (pine) were in the 5% tail of the null distribution, and for most variables tested, the observed proportion was significantly higher than expected by chance (Fig. 3; see also fig. S2).

If the observed overlap of gene involvement in local adaptation occurs because of fundamental constraints in how genotype gives rise to phenotype, then duplication and neofunctionalization may increase flexibility in the genetic program (18). Consistent with this prediction, genes

¹Department of Biological Sciences, University of Calgary, Calgary, Canada. ²Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada. ³School of Biological Sciences, Monash University, Melbourne, Australia. ⁴Department of Marine and Environmental Science, Northeastern University, Nahant, MA, USA. ⁵Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. ⁶Department of Renewable Resources, University of Alberta, Edmonton, Canada. ⁷Department of Zoology, University of British Columbia, Vancouver, Canada. ⁸Department of Botany, University of British Columbia, Vancouver, Canada.

*These authors contributed equally to this work. †Corresponding author. Email: sally.aitken@ubc.ca

Table 1. Number of genes with signatures of convergence. Columns report the number of cases where a gene from one species that was orthologous to a top candidate in the other species was significantly associated to at least one of the seven main variables by the null-*W* test after adjusting for false discovery rate.

False discovery rate	One-to-one orthologs	One-to-many: Both duplicates convergent	One-to-many: One duplicate convergent	Total number of genes with strong signatures of convergence
0.01	6	0	0	6
0.05	40	0	7	47
0.10	71	2	8	83

duplicated in either species were also more likely to have strong signatures of convergence. Across all comparisons, signatures of convergence were 65% more common in cases where one ortholog was duplicated than in one-to-one orthologs (Fig. 3 shows results on a copy-for-copy basis, mean ratio 1.65, range 0.67 to 4.0; fig. S5 shows results on a per-orthogroup basis, mean ratio 1.98, range 0.67 to 4.3). Independent of convergence signatures, duplicated genes also had a higher probability of being top candidates, although the effect was nonsignificant in most cases (fig. S3). Linkage disequilibrium (LD) between tandem duplicates may be responsible for these patterns, as LD is high among paralogs with at least one member that is a top candidate (fig. S4). However, LD and tandem duplication cannot explain the enrichment of association signatures in the single-copy orthologs to duplicated top candidates (Fig. 3 and fig. S2, orange bars), nor the duplicated orthologs, because binning duplicates before repeating the analysis yielded similar results (fig. S5). Thus, convergent local adaptation and gene duplication are associated in these conifers, possibly as a means to increase genetic flexibility. Alternatively, duplications of genes involved in local adaptation may have been favored under migration-selection balance, due to changes in linkage relationships (19) or dominance-associated masking of migration load (20).

Overall, 47 genes exhibit signatures of convergence at a false discovery rate (FDR) of 5% (or 83 at FDR = 10%), out of 260 and 450 top candidates with identified orthology relationships in pine and spruce, respectively (Table 1; see fig. S6 for phylogenies). This suggests that ~10 to 18% of locally adapted genes are evolving convergently, a lower rate than typically found for candidate genes or quantitative trait loci (3); however, the true proportion may be much higher. Many of the top candidates identified within either species (Fig. 1, C and D) are likely false positives due to the lack of control for population structure (21) or because they are physically linked to a causal locally adapted gene but are not themselves locally adapted. The former artifact is not expected to affect the convergence candidates significantly above the rate represented by our null hypothesis (horizontal gray line, Fig. 3), as drift is unlikely to give rise to the same false positive in both species (17). Although we found evidence of considerable LD among some top candidates (fig. S4), the convergence candidates were not usually in strong LD with each other; hence, this latter artifact is also not causing many false positives (figs. S7 and S8). Because these artifacts are likely to inflate the number of top candidates identified within species but not to significantly affect signatures of convergence, the true proportion of genes adapting convergently may be higher than 10 to 18%.

Data on gene expression in response to climate stress [from (18)] revealed that 61 convergence candidates with expression data had conserved patterns of differential expression in both species, while 17 had divergent patterns (a factor of ~3.5 difference). This is approximately twice the ratio of conserved:divergent expression observed in non-

convergently adapted genes ($P = 0.014$, Fisher's exact test; table S6). Genes with signatures of convergence were also enriched for transcription factors and genes involved in biological regulation and RNA metabolism (enrichment significant in spruce convergence candidates but not pine; tables S8 and S9). Thus, although genes involved in convergent local adaptation are disproportionately conserved in their expression, they are also more likely to affect the expression of other genes. Evidence from *Arabidopsis* suggests that the protein products of several of these convergent genes could be relevant to seasonal transitions and abiotic stress (table S9). For example, PSEUDO-RESPONSE REGULATOR 5 (PRR5) directly regulates the circadian clock and associated developmental transitions (22); FY regulates processing of *FCA* mRNA, which in turn regulates accumulation of *FLOWERING LOCUS C (FLC)* mRNA (23); and REGULATORY COMPONENT OF ABA RECEPTOR 1 (RCAR1) functions as a sensor of abscisic acid, a key abiotic stress-related phytohormone (24).

Taken together, our results indicate that local adaptation is more repeatable at the genomic level than might be expected, given the highly polygenic basis of these traits (8, 9) and the potential for considerable genetic redundancy.

Furthermore, gene duplication appears to contribute importantly to convergence, although the reason for this is unknown. Whether gene duplication is a common facilitator of convergent genotypic evolution across the domains of life remains to be seen.

Our results suggest that long-diverged conifers share a suite of genes that play an important role in adaptation to temperature, and should enable functional annotation and tools for candidate-augmented genomic selection. However, they also show that adaptation is highly polygenic and involves heterogeneous, nonconvergent responses

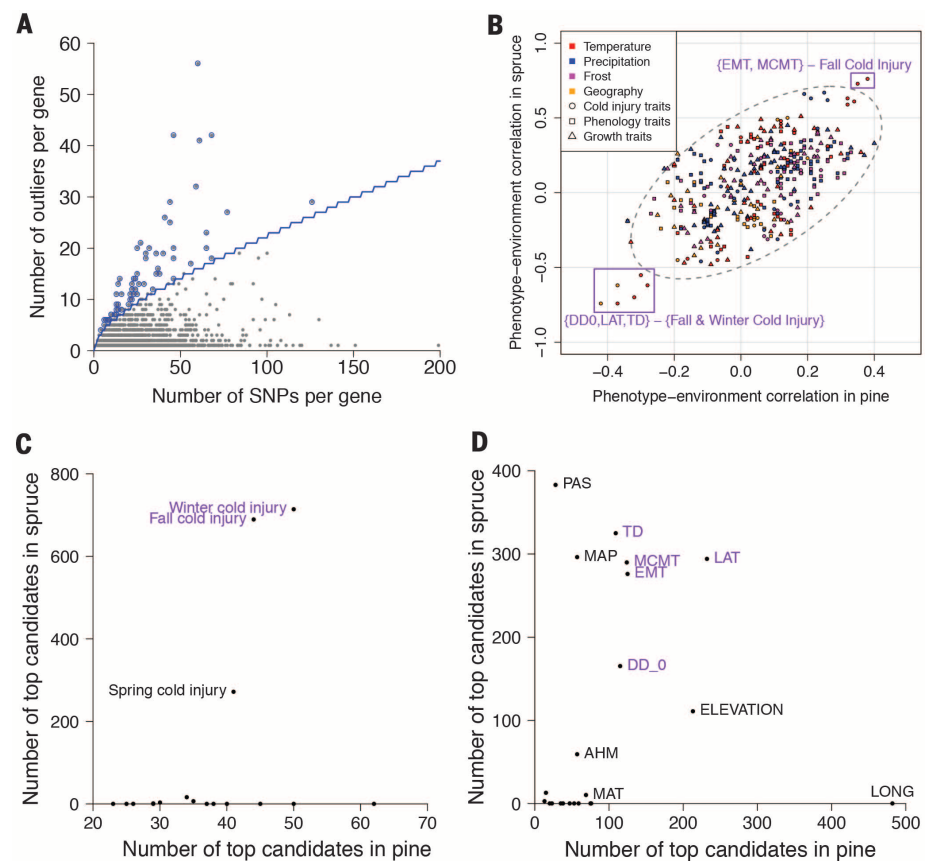


Fig. 1. Signatures of convergent adaptation at the phenotypic and genomic level. Spearman correlations were calculated between each SNP and the 22 environmental and 17 phenotypic variables. **(A)** Top-candidate genes for each of the 39 tests were identified as those with an extreme number of outlier SNPs relative to a binomial expectation, shown in blue (mean annual temperature in pine). **(B)** Cold injury response phenotypes were strongly correlated to temperature variables in both lodgepole pine and interior spruce, with the most strongly correlated cases shown in purple ("main variables"). **(C and D)** The seven main variables with strong phenotype-environment correlations also had the largest number of top-candidate genes for phenotypes (C) and environments (D); labels are omitted for data points near the axes for clarity. EMT, extreme minimum temperature; MCMT, mean coldest-month temperature; DD_0, degree-days below 0°C; LAT, latitude; TD, temperature difference; MAT, mean annual temperature; PAS, precipitation as snow; MAP, mean annual precipitation; AHM, annual heat-moisture index; LONG, longitude (see tables S1 and S2).

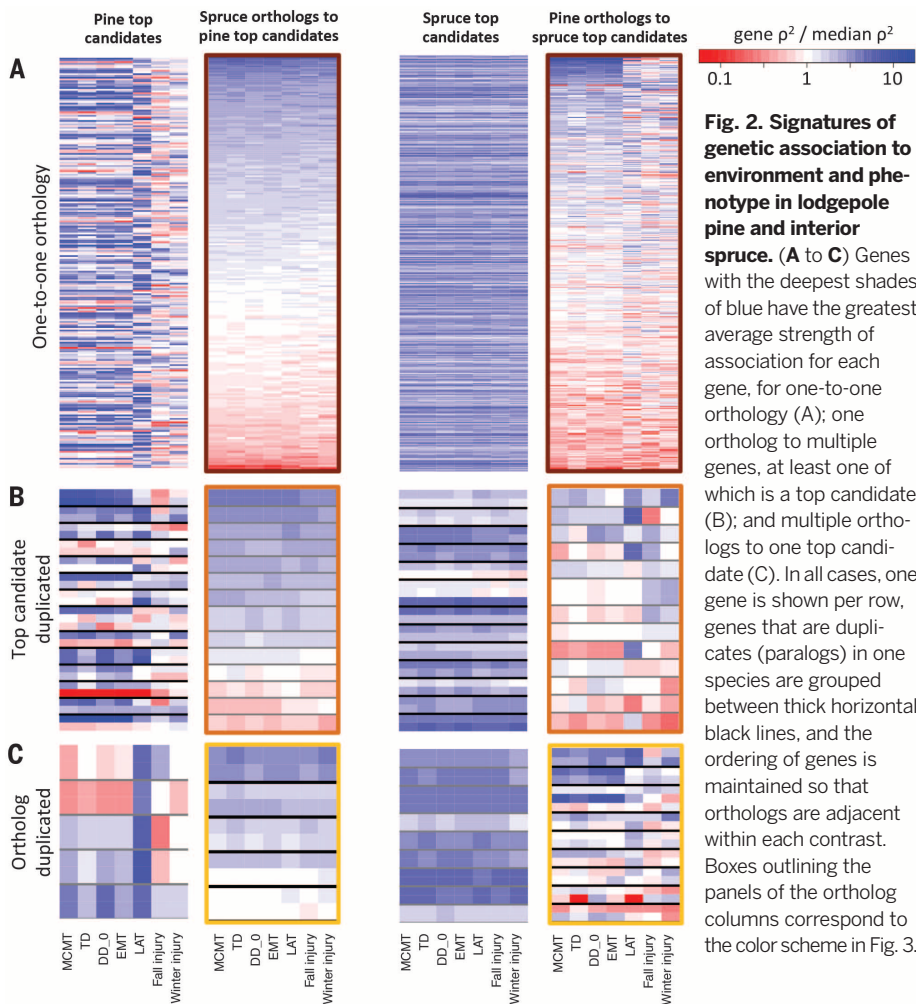
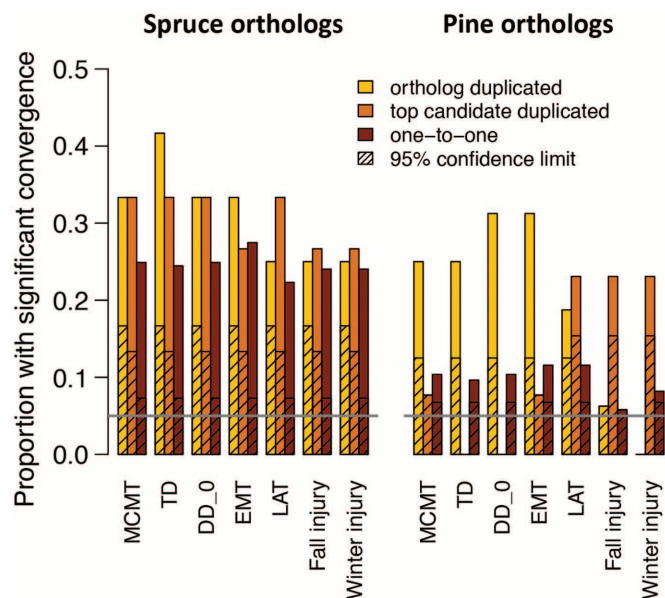


Fig. 2. Signatures of genetic association to environment and phenotype in lodgepole pine and interior spruce. (A to C) Genes with the deepest shades of blue have the greatest average strength of association for each gene, for one-to-one orthology (A); one ortholog to multiple genes, at least one of which is a top candidate (B); and multiple orthologs to one top candidate (C). In all cases, one gene is shown per row, genes that are duplicates (paralogs) in one species are grouped between thick horizontal black lines, and the ordering of genes is maintained so that orthologs are adjacent within each contrast. Boxes outlining the panels of the ortholog columns correspond to the color scheme in Fig. 3.

Fig. 3. Proportion of top-candidate orthologs with significant signatures of convergent local adaptation.

All orthologs from Fig. 2 were tested with the null- W test with $\alpha = 0.05$; colors correspond to the outlines in the respective panels. The horizontal gray line at 0.05 indicates the expected number of significant results under the null hypothesis of pure drift. Hatching indicates the upper 95% confidence limit for this null hypothesis (based on a binomial test with $P = 0.05$).



at many other genes. The success of climate change mitigation strategies such as assisted migration and breeding for new climates will depend on a thorough understanding of adaptation to climate (25), and exploration of the genomic basis of adaptation will inform these activities.

REFERENCES AND NOTES

1. D. L. Stern, *Nat. Rev. Genet.* **14**, 751–764 (2013).
2. A. Martin, V. Orgogozo, *Evolution* **67**, 1235–1250 (2013).
3. G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schluter, *Proc. R. Soc. B* **279**, 5039–5047 (2012).
4. P. F. Colosimo *et al.*, *Science* **307**, 1928–1933 (2005).
5. V. Soria-Carrasco *et al.*, *Science* **344**, 738–742 (2014).
6. J. A. Holliday, L. Zhou, R. Bawa, M. Zhang, R. W. Oubida, *New Phytol.* **209**, 1240–1251 (2016).
7. M. Manceau, V. S. Domingues, C. R. Linnen, E. B. Rosenblum, H. E. Hoekstra, *Philos. Trans. R. Soc. B* **365**, 2439–2450 (2010).
8. M. V. Rockman, *Evolution* **66**, 1–17 (2012).
9. S. Yeaman, *Am. Nat.* **186** (suppl. 1), S74–S89 (2015).
10. X. Q. Wang, D. C. Tank, T. Sang, *Mol. Biol. Evol.* **17**, 773–781 (2000).
11. O. Savolainen, T. Pyhäjärvi, T. Knürr, *Annu. Rev. Ecol. Evol. Syst.* **38**, 595–619 (2007).
12. K. J. Liepe, A. Hamann, P. Smets, C. R. Fitzpatrick, S. N. Aitken, *Evol. Appl.* **9**, 409–419 (2016).
13. T. L. Parchman *et al.*, *Mol. Ecol.* **21**, 2991–3005 (2012).
14. B. Hornoy, N. Pavy, S. Gérard, J. Beaulieu, J. Bousquet, *Genome Biol. Evol.* **7**, 3269–3285 (2015).
15. S. Yeaman *et al.*, *New Phytol.* **203**, 578–591 (2014).
16. K. A. Hodgins, S. Yeaman, K. A. Nurkowski, L. H. Rieseberg, S. N. Aitken, *Mol. Biol. Evol.* **33**, 1502–1516 (2016).
17. See supplementary materials on Science Online.
18. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, 1970).
19. S. Yeaman, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1743–E1751 (2013).
20. A. Yanchukov, S. Proulx, *Evolution* **66**, 1543–1555 (2012).
21. K. E. Lotterhos, M. C. Whitlock, *Mol. Ecol.* **24**, 1031–1046 (2015).
22. N. Nakamichi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17123–17128 (2012).
23. G. G. Simpson, P. P. Dijkwel, V. Quesada, I. Henderson, C. Dean, *Cell* **113**, 777–787 (2003).
24. Y. Ma *et al.*, *Science* **324**, 1064–1068 (2009).
25. S. N. Aitken, M. C. Whitlock, *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).

ACKNOWLEDGMENTS

We thank D. Bachelet, E. Buckler, G. Howe, O. Savolainen, P. Ingvarsson, J. Mee, T. Parchman, R. Barrett, and D. Schluter for comments, and R. Baranowski for support on Westgrid. Seeds were kindly provided by 63 forest companies and agencies in British Columbia and Alberta (listed at adapttree.sites.olt.ubc.ca/seed-contributors), facilitated by the BC Tree Seed Centre and the Alberta Tree Improvement and Seed Centre. D. Neale and J. Bohlmann generously provided access to loblolly pine and white spruce draft genomes prior to their release. This research was part of the AdapTree Project (S.N.A. and A.H., co-project leaders) funded by Genome Canada, Genome BC, Genome Alberta, Alberta Innovates BioSolutions, the Forest Genetics Council of British Columbia, Virginia Tech, the University of British Columbia, NSF Plant Genome Research Program grant IOS:1054444 (J.A.H.), USDA National Institute of Food and Agriculture, McIntire Stennis Project grant 10005394 (J.A.H.), and the British Columbia Ministry of Forests, Lands and Natural Resource Operations. Sequence data are deposited in the Short Read Archive (SRP071805; PRJNA251573) and data and analysis scripts are deposited in Dryad (doi:10.5061/dryad.0t407). The authors declare no conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6306/1431/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 to S10
References (26–61)

31 March 2016; accepted 11 August 2016
10.1126/science.aaf7812