

INVITED SPEAKER PRESENTATION

Open Access

Population and conservation genomics of forest trees: seeing the forest for the trees

Andrew Eckert

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d'Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

Forest trees exhibit striking adaptations to the environments in which they grow. A long history of quantitative genetic experimentation has established the genetic basis for many traits, which are likely adaptive since many of them are also correlated with environmental heterogeneity. The genes underlying these traits, however, have largely remained elusive. Recent applications of high-throughput sequencing and genotyping technologies to natural populations of forest trees have identified several promising candidates for genes underlying complex and adaptive traits (reviewed by [1]). The diversity of analytical approaches employed in those studies, however, begs the question of the generality of reported results. Here, I exploited the diversity of analytical approaches used previously to identify genes underlying adaptive traits for two conifer species to assess the logical consistency among results generated from different conceptual frameworks.

Materials and methods

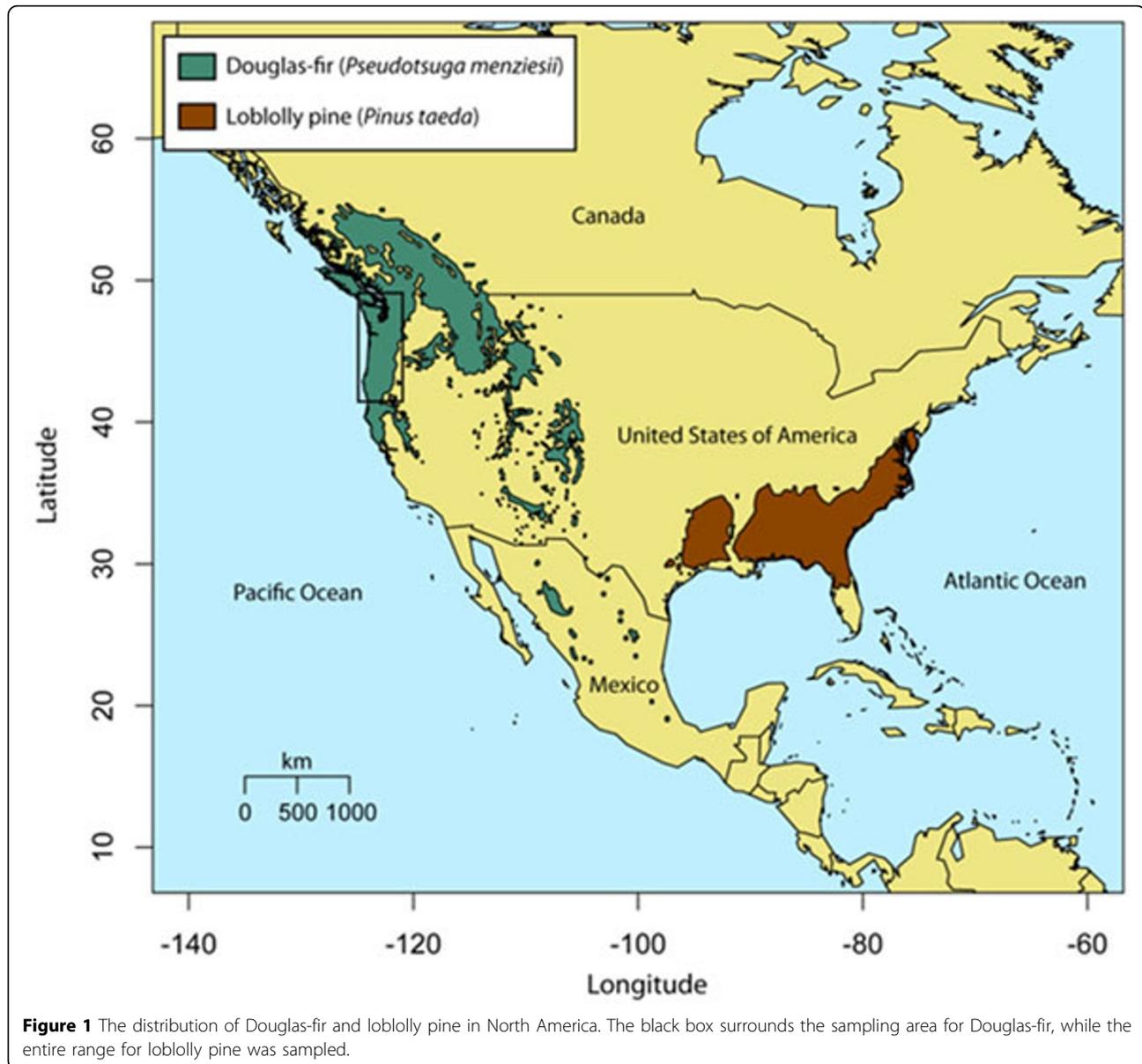
Data sets comprised of genotyped single nucleotide polymorphisms (SNPs) were gathered for two North American conifers (Fig. 1): loblolly pine (*Pinus taeda* L.) and coastal Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*). These data sets were generated through resequencing of diversity panels ($n = 18$ -24 megagametophytes) for a sample of expressed sequence tag (EST) unigenes ($n = 7,535$) and cold-hardiness related candidate genes ($n = 121$), respectively. Genotyping was performed using Illumina's Infinium (loblolly pine) or GoldenGate (coastal Douglas-fir) array technologies. For each data set, associations to phenotypes and environmental variables were gathered from the

literature or performed as described elsewhere (see references in [1]). For each associated gene, I asked two questions: (1) Are genes associated to phenotypes more often also associated to environmental variables than randomly chosen genes? (2) Are associated genes also outliers for nucleotide diversity and site-frequency spectrum based statistics, nucleotide divergence or F_{ST} more often than randomly chosen genes? Permutation tests were used to assess whether or not observed patterns were different than those produced by chance.

Results

A total of ~30 genes were associated with cold-hardiness phenotypes or with climate data for coastal Douglas-fir, while ~850 associated genes were identified for loblolly pine. Genes associated with phenotypes for coastal Douglas-fir were more often associated to environmental variables than randomly chosen genes ($P = 0.005$), which was only moderately apparent for loblolly pine ($P = 0.067$). Genes associated to phenotypes or environmental variables were not more likely to be outliers for nucleotide diversity, site-frequency spectrum based statistics, nucleotide divergence or F_{ST} for Douglas-fir ($P > 0.05$). There was some evidence, however, for non-neutral evolution for associated genes using a statistic based on the McDonald-Kreitman table [2]. In this case, associated genes had too many extreme values for the direction of selection (DoS) statistic than expected from randomly resampling the available genes ($P < 0.05$). Associated genes on average had skewed site-frequency spectra for loblolly pine, especially for synonymous sites, as well as too many extreme values for the DoS statistic. Further classification of genes for loblolly pine into functional categories revealed striking trends indicative of non-neutral processes underlying some of the associations (Fig. 2). A total of 11 gene categories were consistent with positive ($n = 7$) or negative ($n = 4$)

Correspondence: ajeckert@ucdavis.edu
Department of Ecology and Evolution, University of California at Davis, Davis, CA, 95616, USA



selection. In both cases, the frequency of associated loci increased for these categories and the synonymous site frequency spectra became more skewed. Many associations for loblolly pine may thus reflect linked selection, with molecular phenotypes (e.g. gene expression) accounting for all the associations in gene categories indicative of negative selection and environmental associations (e.g. aridity) being largely located in gene categories consistent with positive selection.

Conclusions

Many of the genes associated to phenotypes for both species were also correlated to environmental variables and exhibited patterns of non-neutral evolution. Thus,

associated genes are prime targets for conservation efforts. The questions posed here, however, make the strong assumption that genes associated to phenotypes or environmental variables should also show non-neutral patterns of evolution. This is not always expected to be the case [3,4], yet the lack of consistency is often interpreted as such and is one explanation for those genes or sets of genes reported here as lacking non-neutral signals. The search for logical consistency among analytical approaches, however, often focuses on uninformative patterns. To illustrate this point, I employed a novel environmental association approach that correlates genetic divergence to environmental change and show that most of the site-frequency spectrum based outliers

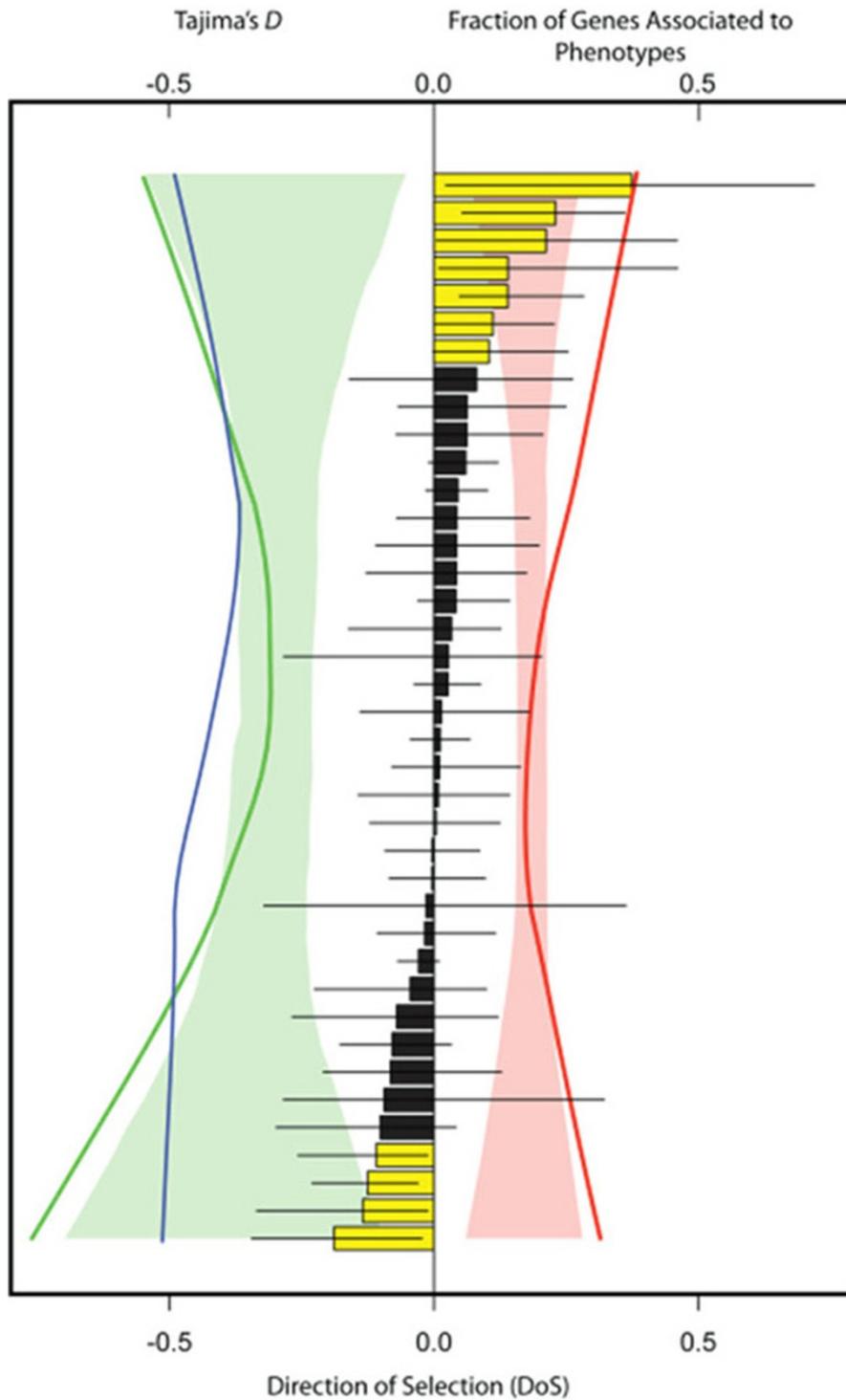


Figure 2 An example of logical consistency for loblolly pine between genes associated to phenotypes or environmental variables and gene categories that have average deviations from neutrality consistent with positive or negative selection. Error bars give 99% bootstrap confidence intervals for the direction of selection statistic (DoS), with yellow bars having confidence intervals excluding zero. Lines give average Tajima's *D* for nonsynonymous (blue) and synonymous sites (green), as well as the proportion of genes associated to at least one phenotype (red). Shaded areas give null distributions (99% quantiles) generated via permutations of genes among categories ($n = 10,000$ permutations). All lines, including those forming the null distributions, were smoothed using lowess smoothing.

for coastal Douglas-fir [5] are correlated to change in climate variables but not to extant climate patterns. Taken together these results illustrate that non-neutral genes are often identified during association analyses, that departures from neutrality for genes driving associations are not only those due to recent directional selection, and that further work is needed to understand the population genetic processes underlying associations between genotypes, phenotypes and the environment.

Acknowledgements

I would like to thank J. L. Liechty, B. N. Figueroa G. Rosa, and J. L. Wegrzyn for bioinformatics and computational support. I would like to thank all of the collaborators on the ADEPT2 project (<http://dendrome.ucdavis.edu/NealeLab/adept2/>). This work was funded by the National Science Foundation (IOS-PGRP-0501763, IOS-PGRP 0638502) and the United States Department of Agriculture (NRI Plant Genome 04-712-0084).

Published: 13 September 2011

References

1. Neale DB, Kremer A: **Forest tree genomics: Growing resources and applications.** *Nat Rev Genet* 2011, **12**:111-122.
2. Stoletzki N, Eyre-Walker A: **Estimation of the neutrality index.** *Mol Biol Evol* 2011, **28**:63-70.
3. Le Corre V, Kremer A: **Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection.** *Genetics* 2003, **164**:1205-1219.
4. Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.** *Curr Biol* 2010, **20**:R208-R215.
5. Eckert AJ, Bower AD, Pande B, Jermstad KD, Krutovsky KV, St. Clair JB, Neale DB: **Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits.** *Genetics* 2009, **182**:1289-1302.

doi:10.1186/1753-6561-5-S7-12

Cite this article as: Eckert: Population and conservation genomics of forest trees: seeing the forest for the trees. *BMC Proceedings* 2011 **5** (Suppl 7):12.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

