npg

NEWS AND COMMENTARY

Molecular genetics

Trees' genes and traits link up

B Heinze

A formidable question has driven forest geneticists ever since the first genetic markers became available: there is strong clinal and ecotypic variation in the growth traits of forest trees, but this is hardly ever reflected in similar patterns at genetic markers (or patterns of similar magnitude)—why? In this issue of *Heredity*, González-Martínez *et al.*, 2008 present an elegant way of getting closer to an answer. Using statistical methods originally intended for studying human and livestock genetics, they show how two particular nucleotide changes influence drought tolerance in pines.

Tree genetics, similar to plant genetics in general, has traditionally been advanced by studies of parent–offspring relationships, especially those in artificial crosses. Yet there are a number of practical reasons why these have had little impact in tree research: flowers can often only be reached by climbing the canopy, only few seeds are produced per flower in many species, and performing such crossings on more manageable grafted trees takes years of preparations. Even then, performing backcrosses or multigeneration crosses may take many researcher generations, as trees often have extended juvenile phases—just similar to humans.

But here is the trick of González-Martínez *et al.*, 2008: if trees are so similar to humans in these ways—and others, such as widespread populations and high heterozygosities—why not deploy the approaches used to study complex traits as in humans?

Association genetics—relating variation in traits to variation in genetic markers in large populations—is one such approach. It seems a straightforward thing to do, but there are many reasons why it is not so easy in practice, neither for humans nor for trees. Ever since genetic markers became available in greater numbers—first isoenzymes, then anonymous markers like random amplified polymorphic DNA (RAPD, and similar systems), researchers have tried to make this link from phenotype to genotype, and have often claimed success (for example, Bergmann, 1978;

Jiang *et al.*, 2003; Xu *et al.*, 2004; including my own—in hindsight, rather vain—attempts, Heinze and Geburek, 1995). But statistics all too often fooled us: with tens of thousands of genes present, and large, unstructured populations that are essentially in linkage equilibrium, there is no way for a precision landing with a marker in or very near the gene underlying a trait of interest.

This is why the last decade has seen many studies using artificial crosses in trees, even in rather intractable organisms like oak trees. Families show enough linkage disequilibrium so that one in a few hundred markers may provide evidence of linked genes in their vicinity. But this wider chromosomal vicinity may still encompass hundreds of genes. Moreover, it has been clear for quite a long time that complex traits such as growth form (straightness and branching) or quantitative traits (such as sheer size) are influenced by many genes with smaller effects. For some of them, it may be possible to postulate, in family studies, quantitative trait loci (QTL), that is, chromosome regions statistically associated with trait differences—but these usually prove very dependent on the genetic background.

With DNA sequencing becoming cheaper, researchers have now returned to the wild forests, this time with a number of good candidate genes in their pockets—the 'usual suspects' including genes of known function, genes with interesting expression patterns or chromosomal locations (for example, those in linkage with QTL). However, there is often ample polymorphism at the sequence level, linkage disequilibrium seldom extends over more than 100 bases (in genomes of several gigabasepairs), and many of the good candidates fail to find support for their supposed involvement in certain traits (for example, Heuertz *et al.*, 2006). Even if a candidate gene is supported, there can still be doubts, for example whether patterns could have been produced by clinal demography (for example, Ingvarsson *et al.*, 2006).

The quantitative transmission disequilibrium test employed by González-Martínez *et al.*, 2008 exploits 'a little bit of linkage disequilibrium' caused by population structure, which is otherwise unwanted in association studies, as it creates false associations. In contrast, it can turn into an asset if this population (or family) structure is astutely chosen to allow the effects of individual alleles to be assessed in many genetic backgrounds (or in a constant background). With this approach, they have found two particularly interesting variable genes among their candidates. Variation in water use efficiency in pine trees at two locations in the south-eastern US involves single nucleotide polymorphisms in a dehydrin and a cell wall re-enforcement gene.

This is an exciting beginning of a probably long and tedious new path towards fully understanding the control of complex traits in natural tree populations. There is reason to believe that the polygene model still holds, as the two candidates can only explain a few percent of the total phenotypic variation present. This small value is striking, as one can assume that best candidates have already been chosen, and any other ones will be more difficult to select. But more importantly, gene interactions may play an important role in local adaptation. For the polygenic case, it may turn out as David Neale once remarked, many years ago, after I presented another (doomed) RAPD association study at a workshop—'Shouldn't you have at least a little bit of linkage disequilibrium?'

Dr B Heinze is at the Department of Genetics, Federal Research Centre for Forests, Vienna, Austria.

e-mail: Berthold.Heinze@bfw.gv.at

Bergmann F (1978). The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*) along similar climatic gradients. *Theor Appl Genet* **52**: 57–64.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008). Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101**: 19–26.

Heinze B, Geburek T (1995). Searching for DNA markers linked to leaf colour in copper beech, *Fagus sylvatica* L. var *atropunicea*. *Silvae Genet* **44**: 339–343.

Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I, Morgante M *et al.* (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce (*Picea abies* (L.) Karst). *Genetics* **174**: 2095–2105.

Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006). Clinal variation in phyB2, a candidate gene for day-length-induced growth

cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* **172**: 1845–1853.

Jiang L, You R-L, Li M-X, Shi C (2003). Identification of a sex-associated RAPD marker in *Ginkgo biloba*. *Acta Bot Sin* **45**: 742–747.

Xu W-J, Wang B-W, Cui K-M (2004). RAPD and SCAR markers linked to sex determination in *Eucommia ulmoides* Oliv. *Euphytica* **136**: 233–238.

Editor's suggested reading

Garcia MV, Ingvarsson PK (2007). An excess of nonsynonymous polymorphism and extensive haplotype structure at the *PtABI1B* locus in European aspen (*Populus tremula*): a case of balancing selection in an obligately outcrossing plant? *Heredity* **99**: 381–388.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007). Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* **98**: 74–84.

Woolbright SA, DiFazio SP, Yin T, Martinsen GD, Zhang X, Allan GJ *et al.* (2008). A dense linkage map of hybrid cottonwood (*Populus fremontii* × *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity* **100**: 59–70.