A FOUNDER PROJECT: MARKING THE DOMESTICATION BASELINE FOR FOREST TREES [1]

by
C.G. Williams[2], F.E. Bridgwater[3] and C.D. Nelson[4]

INTRODUCTION

One of the most apparent benefits of forest genomics programmes is to provide genotypic information on the original selections of tree improvement programmes worldwide. In many breeding programmes, branches from these selections were grafted onto seedlings and the grafted seedlings composed the first seed orchards for planting programmes. With advanced generation orchards or new vegetative propagation technology, these original orchards have become genetic archives. These archives conserve the original selections or founders of domesticated forests thus providing an opportunity for genotyping the entire population of founders. The window of opportunity, known as a Founder Project, is narrowing for obtaining DNA samples from founders because archives are being lost to pests, pathogens, extreme weather events, climate change, constricting budgets and increasing demands for arable land.

MEASURING THE GENETIC CONSEQUENCES OF FOREST DOMESTICATION

The foremost reason for genotyping founders as part of a Founder Project is to provide a baseline for tracking genetic consequences of forest domestication. Genotypic information on the founders provides a baseline against which later generations or various breeding strategies can be tested for loss or increase in genetic diversity. Early generations of recurrent genetic improvement are expected to maintain high levels of genetic diversity, since population-level improvement conserves allelic diversity as a means of ensuring long-term adaptation (Williams *et al.*, 1994; McKeand and Bridgwater, 1998; White *et al.*, 1993). In more intensive breeding programmes, genotyping founders allows for an estimate of the loss of genetic diversity that can occur the tradeoff between enhanced genetic gains and decreased genetic diversity. In those programmes that emphasize gene stewardship rather than maximum genetic gain, genotyping founders provides a baseline for monitoring effects of large shifts in breeding or production population size as a result of climate change, reduction in population sizes, use of vegetative propagules or selective seed collection.

IDENTIFYING DIAGNOSTIC ALLELES FOR TRACING GERMPLASM ORIGIN FOR EXOTIC SPECIES

Another application for a Founder Project is to determine the original provenance(s) of an exotic species or landrace. Again, a founder genotype database provides a baseline or a foundation for comparison. There are numerous exotic forest tree species for which provenance introductions are not known. In this situation, founders for the species within its indigenous range are first genotyped for hypervariable molecular markers. A search for diagnostic alleles at common frequencies is conducted and declared successful if a common allele appears in one provenance but not in another. These diagnostic allele patterns are determined in the indigenous range and then compared with the exotic populations in question. An example for *Pinus taeda* L. introductions in Zimbabwe is shown in Box 1.

---

1 Received May 2003.
2 Duke University. School of the Environment & Earth Sciences. Box 90658. Durham NC 27708. claire.williams@duke.edu
3 USDA Forest Service, Southern Research Station, 3041 Cornwallis Road, Research Triangle Park, NC 27709 USA
4 USDA Forest Service, Southern Institute of Forest Genetics, 23332 MS Highway 67, Saucier MS 39564 USA

Box 1. The provenance origin of *Pinus taeda* in Zimbabwe

Pinus taeda comprises historically large, interconnecting populations, which extend along the Unit States Atlantic seaboard from Maryland to Florida and westward to central Texas. The Mississippi River Val roughly delineates the eastern and the western parts of the species' range. A small founder genotyping data was developed with 36 nuclear microsatellite markers and 171 founders collected from grafted archives natural stand selections made from 1950 to 1970 (Al-Rababah and Williams, 2002). These *P. taeda* found represent the species in its natural range prior to intensive plantation establishment and domestication in the la 20th century. Similarly, a small set of Zimbabwe selections were assayed for the same microsatellites th compared with the founder dataset.

The first step is to find alleles that distinguish one provenance from another within the natural range the species. Diagnostic or unique alleles within stands or subpopulations are often frequent with molecu markers than have large numbers (>10) of alleles per locus. Diagnostic alleles must occur in frequencies above percent. Less frequent marker alleles may also be absent in one population and present in another populatio but this is more likely due to sampling so these alleles are not treated as diagnostic.

The second step is to check for these diagnostic alleles in the exotic population in question. T provenance origin of *P. taeda* in Zimbabwe has been an open question for decades. By comparing eastern a western *P. taeda* sources, four marker alleles appear diagnostic (Table 1). The larger founder sample size Table 1 shows that Zimbabwe population is an admixture of eastern and western sources although the diagnos alleles reported in an earlier study (Williams *et al.*, 2000) changed with the fourfold increase in founder populatio size.

Table 1. *Pinus taeda* in Zimbabwe is still an admixture based on the larger survey (N =171) across the indigenous speci range.

Locus	Allele	Zimbabwe	Western		Eastern	
			BA	WG	SC-FL	NE
PtTX2037	9	0.042	0	0	0.071	0.017
PtTX2128	2	0.167	0	0	0.065	0.018
PtTX2146	9	0.136	0	0	0.068	0.034
PtTX2164*	10	0.136	0	0	0.091	0.204
PtTX3011	3	0.042	0	0	0.068	0
PtTX3030	1	0.111	0.125	0.024	0	0
PtTX3030	14	0.278	0	0	0.143	0.089
PtTX3037	4	0.042	0	0	0	0.086
PtTX3037	14	0.125	0	0	0.024	0.069
PtTX3037	27	0.083	0	0.014	0	0

The asterisk (*) represents the one diagnostic allele common to the smaller and larger sample sizes. Western regions a Bastrop County (BA) and Western Gulf (WG). Eastern regions are South Carolina to Florida (SC-FL) and Northeast (NE)

A shared founder genotyping database for *P. taeda* provides a resource where other investigators ca compare founder genotypic data against their own exotic populations. Others can determine provenances fo other countries planting *P. taeda* as an exotic plantation species. Sharing a founder genotype database at th international level reduces the amount of time and lab resources committed to redundant genotyping.

OTHER APPLICATIONS FOR A FOUNDER PROJECT

In those breeding programmes which have extensive advanced generation breeding, a Founder Project useful for finding the right pedigrees for linkage map construction and for tracing haplotypes betwee populations and generations. If two molecular markers prove to be closely linked, then they delineate a segmen of chromosome, defined as a haplotype. Each diploid individual has two haplotypes, one contributed by it maternal parent and one contributed by its paternal parent (Figure 1).

A Founder Project would provide a useful database for locating one or more copies of a particula haplotype to trace among a founder's descendants. The principal value to tracing haplotyping occurs if there ar pedigrees for each founder in question. Tracing particular haplotypes in related pedigrees is quite relevant to tre breeders or ecologists if a given haplotype has been shown to be correlated with major phenotypic changes adverse or beneficial (Gwaze *et al.*, 2003).

If the haplotype spans a chromosomal interval that includes one or more genes influencing the phenotypic trait in question then this type of haplotype is known as a quantitative trait loci or QTL (see review in Williams, 1998). Starting with the haplotype information already available in the founder genotype database saves precious time and lab resources otherwise spent searching for a pedigree suited to an important experimental question. Applications for a Founder Projects also extend to teaching, professional and practical training in breeding, conservation and ecological genetics of forest trees.

Figure 1. An example of tracing haplotypes from four unrelated founders through a three-generation pedigree.

Founder 1	Founder 2	Founder 3	Founder 4
$\frac{A_7 - B_2}{A_6 - B_2}$	$\frac{A_4 - B_2}{A_{14} - B_2}$	$\frac{A_{12} - B_2}{A_4 - B_5}$	$\frac{A_4 - B_5}{A_6 - B_1}$

Parent 12	Parent 34
$\frac{A_7 - B_2}{A_{14} - B_2}$	$\frac{A_{12} - B_5}{A_6 - B_1}$

Offspring

1	2	3	4	5	6	7	8
$\frac{A_{14} - B_2}{A_{12} - B_1}$	$\frac{A_{14} - B_2}{A_{12} - B_5}$	$\frac{A_7 - B_2}{A_6 - B_1}$	$\frac{A_{14} - B_2}{A_6 - B_5}$	$\frac{A_7 - B_2}{A_{12} - B_1}$	$\frac{A_7 - B_2}{A_{12} - B_5}$	$\frac{A_7 - B_2}{A_6 - B_1}$	$\frac{A_{14} - B_2}{A_6 - B_1}$

Note that each founder has two unique haplotypes or chromosomal segments. With recombination or crossing-over, each founder can contribute one of four haplotypes to the next generation. In this example, Founder 3 contributes a recombinant haplotype (A_{12}-B_5) and Founder 4 contributes a parental haplotype (A_6-B_1) to Parent 34.

GENOME COVERAGE AND CHOICE OF MOLECULAR MARKERS FOR A FOUNDER PROJECT

There is no ideal molecular marker for a Founder Project so the best strategy is to include several types of markers as funding permits: this might include microsatellites, isozymes, and polymorphic gene sequences. Including mitochondrial or chloroplast markers is quite vital for parental identification, for phylogeography studies and for measuring gene flow and population differentiation. The best marker systems for a Founder Project have the following properties:

- the marker's polymorphism corresponds to a known change in its DNA sequence,
- marker assays are low-cost,
- for nuclear markers, it is important that a marker score discern between homozygotes and heterozygotes.

This property is important because heterozygosity is the basis for measuring genetic diversity changes during domestication. This becomes especially important for highly heterozygous species that typically have multiple allele systems (see Box 2, Figure 2). Dominant marker systems such as RAPDs or AFLPs have no prominent role in founder genotype databases because neither heterozygosity nor multiple alleles can be discerned.

Box 2. An example of Founder Project showing a nuclear microsatellite.

Polymorphism levels of nuclear microsatellites tend to be very high for focal species (Echt *et al.,* 1996). The high number of alleles per locus offers a method of measuring allelic diversity with increased resolution. Marker PtTX4114 shows a nuclear microsatellite for *Pinus taeda* which has a multiple-allele intercross configuration (Figure 2).

The mating type configuration is multiple-allele intercross with four alleles. This means that there are four different heterozygote classes for the full-sib progeny. Ladder (M) shows two molecular bands sizes 120 and 105 bp, respectively. Lanes 1 to 4 show genotypes for four unrelated grandparents and Lanes 5 to 6 show genotypes for two unrelated parents descended from grandparents. Lanes 7, 8, 9 and 14 show the four different types of heterozygotes that occur with the mating of the original highly heterozygous founders (or grandparents in this pedigree).

Figure 2. Microsatellite marker PtTX4114 is shown for a founder genotype and its descendants for *Pinus taeda.*

M: molecular weight marker.

Marker PtTX4114 shows different alleles segregating among four different founders. In this advanced-generation programme, the founders were mated pairwise and their respective offspring, shown as parents (Figure 1) produced offspring. The four alleles in the founders and the parents were transmitted to the offspring. The four offspring heterozygotes can be seen by observing the offspring genotypes in lanes 7, 8, 9 and 14.

PUBLIC CURATORSHIP OF A FOUNDER PROJECT

Curatorship of a Founder Project should extend from database management to storage of voucher specimens. Many founders are archived in remote, unprotected areas so a voucher specimen from each founder should be archived in a safe place as an insurance policy. Long-term storage of freeze-dried leaf tissue and seeds tends to be more stable than DNA in the event that a founder is lost in the field. Voucher specimens ensure future availability of founder DNA for assaying new types of marker systems not yet available for use today.

Database management should be standardized so that genotyping records are easily retrieved. With time and added resources, phenotypic data can be combined with a Founder Project's genotypic data. This might include original descriptions of the founder, its offspring's measurements and breeding values as well as records on pest and pathogen attack. The goal is to provide a streamlined record of genotypic and phenotypic descriptions that will outlast a human career. Continuity of a database with founder genotyping and record keeping can be priceless in the event of political upheaval, extreme weather events, or financial shortfall.

Genotypic data from the founder projects should be made available for public use through Internet access and government database storage. This is a political obstacle for so many programmes in private and public sectors alike. One must ask whether it is better to have no genotype database resources for an important species rather than share resources. It is important to consider that forestry research and education worldwide has had a tradition of modest funding to the extent that cooperation and sharing have been a necessity. Sharing a founder genotype database is consistent with a tradition of shared genetic resources.

A founder genotype project in the public domain serves as a unifying framework for tree breeding, genetics, genomics and molecular technology. It can be viewed a foundation for other applications of markers

into breeding programmes (Williams and Byram, 2001). A founder genotype database will no doubt raise some important questions about intellectual property that will require careful consensus-building among different organizations contributing DNA samples from founders.

Agricultural conservation programmes should be studied closely as examples of framework and infrastructure for long-term database maintenance. For example, the United State Department of Agriculture has a Germplasm Resources Information Network (GRIN) database that can be accessed via the Web and it now includes forest trees (http://www.ars-grin.gov/). A Founder Project for *Pinus taeda* is now underway at the USDA Forest Service's Southern Forest Genetics Institute. The plan is to post its founder genotype database on the Web but the level of detailed information permitted by founder contributors will ultimately determine its value to the larger forest genetics community.

REFERENCES

Al-Rababah, M. & Williams, C.G. 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. *Forest Ecology and Management* 163: 263-271.

Auckland, L.D., Bui, T., Zhou, Y., Shepherd, M. & Williams, C.G. 2002. *Conifer Microsatellite Handbook.* Corporate Press, Raleigh NC. 57 p.

Echt, C. S., May-Marquardt P., Hseih, M. & Zahorchak, R. 1996. Characterization of microsatellite markers in eastern white pine. Genome 39: 1102-1108.

McKeand, S.E. & Bridgwater, F.E. 1998. A strategy for the third breeding cycle of loblolly pine in the southeastern U.S. Silvae Genetica 47: 223-234.

Gwaze, D.P., Zhou, Y., Reyes-Valdes, M.H., Al-Rababah, M.A. & Williams, C.G. 2003. Haplotypic QTL mapping in an outbred pedigree. *Genetical Research* 81: 43-50.

White, T.L., Hodge, G.R. & Powell, G.L. 1993. An advanced generation tree improvement plan for slash pine in the southeaster United States. Silvae Genetica 42: 359-371.

Williams, C.G. 1998. QTL mapping in outbred pedigrees. *In* Molecular Dissection of Complex Traits. *Edited* by A.H. Paterson. CRC Press. Boca Raton, Fla. pp. 81-94.

Williams, C.G., Hamrick, J.L. & Lewis, P.O. 1994. Genetic diversity levels in a multiple population breeding strategy: a case study using *Pinus taeda* L. *Theoretical and Applied Genetics* 90: 384-394.

Williams, C.G., Elsik, C.G. & Barnes, R.D. 2000. Microsatellite analysis of *Pinus taeda* in Zimbabwe. Heredity 84: 261-268.

Williams, C.G. & Byram, T.D. 2001. Forestry's third revolution: integrating biotechnology into *Pinus taeda* L. breeding programmes. *Southern Journal of Applied Forestry* 25: 116-121.