FGC extension note 06

Applications of DNA Markers in BC Tree Improvement Programs

April 2004

Author:

Craig Newton, PhD
Director, Molecular Diagnostics
Vizonscitec
3650 Westbrook Mall
Vancouver, BC V6S 2L2

Prepared for the Forest Genetics Council of British Columbia

Published April 2004

Library and Archives Canada Cataloguing in Publication Data

ISBN 0-7726-5338-0

About the Forest Genetics Council of British Columbia

The Forest Genetics Council of BC (FGC) is a multi-stakeholder group representing the forest industry, Ministry of Forests, Canadian Forest Service, and universities. Council's mandate is to champion forest gene resource management in British Columbia, to oversee strategic and business planning for a cooperative provincial forest gene resource management program, and to advise the Chief Forester on forest gene resource management policies.

Acknowledgements

Many thanks to our collaborators in the Ministry of Forests Research Branch, especially Joe Webber for initiating these studies, Roger Painter at the Operational Tree Improvement Program (OTIP), Dr. Carol Ritland at the Genetic Data Centre (UBC), and Dr. Salley Aitken at the Forest Conservation Centre (UBC). Special thanks to Mr. Tam Vo and to the memories of Steve Ross and Robert Gawley.

Abstract

DNA markers are tools for detecting variations in DNA that differentiate all living organisms, and thus play an increasing role in defining biological *identity*. The Forest Genetic Council (FGC) has long recognized a need for DNA marker-based identification in Tree Improvement programs and has supported a range of projects focused on BC species. This article will summarize some recent developments in this area and give examples that illustrate their use.

Introduction

The earliest applications of DNA markers in BC forestry were to address seed planning questions in introgression zones (Sutton et al. 1994). While useful in distinguishing populations at the species level (e.g., sitka spruce, interior spruce and hybrids thereof), these first generation DNA markers could not reveal *intra*specific variation (i.e., the ability to differentiate individuals *within* a species). As a result, they were not applicable to questions such as paternity analysis, clonal verification, pedigree analysis, and other issues related to gene flow between individuals within a population. To address this limitation, projects with BC Research were initiated by the Ministry of Forests to develop *intra*specific type markers based both on organelle DNA (chloroplasts and mitochondria) and nuclear or chromosomal DNA. Each type of DNA marker has specific applications suited to different needs.

Organelle DNA markers and gamete gene flow

Unlike most plants, conifer chloroplasts (cp) are inherited from the paternal pollen parent and therefore are ideal targets to develop DNA markers for monitoring paternal gene contribution. The wind-borne male gamete is the single most variable component in orchard gene dynamics and cannot be easily monitored without some kind of marker associated with the pollen. Therefore, if parental clones in an orchard are distinguishable by their cpDNA, the paternal contribution (or a supplemental pollen mix) can be measured simply by counting the corresponding parental cpDNAs in seed embryos. This provides estimates of pollen contamination, selfing rates, and gametic imbalance (e.g., non-random mating) in the fertilized seedlots. These molecular approaches can replace potentially inaccurate alternatives such as male strobili surveys, pollen sticky traps, female cone counts, and provide data needed for estimating *actual* seedcrop composition. As tree improvement programs evolve so will the need for verification methods to ensure that the genetic integrity of seedcrops are being achieved. It is this niche that DNA markers uniquely fill.

However, practical application of these approaches has been slow, first because developing DNA markers in highly conserved chloroplast genomes is not simple and second, because the of the number of different species used in BC Tree improvement programs. For example, DNA markers that work well in one species, say Douglas-fir, do not work in *Pinus* spp., nor do *Pinus* markers always work in *Picea* spp. Therefore, each different tree species often requires its own specific sets of DNA markers.

With funding support from FGC, BC Research has used genome screening techniques such as Restriction Fragment Length Polymorphism RFLP-PCR (Parducci and Szmidt 1999) to develop species-specific cpDNA markers for many of the major tree improvement programs. The focus of these studies was identification of specific regions in the cpDNA that showed high rates of evolution or mutation, so that unrelated individuals have a high probability of being distinguishable by a simple gel electrophoresis assay. Conifer chloroplast DNA markers of this type now exist for Douglas-fir (Stoehr et al. 1998), lodgepole pine (Stoehr and Newton 2002), spruce, larch (Newton and Vo 2001), and non-native species such as loblolly pine (Newton and Vo 2002). Markers for hemlock and the cedars (red and yellow) are in

Table 1: *DNA markers for BC Forest species*

		Marker type				
		Nuclear			Organelle	
Species	Common name	nuc SSR[a]	EST-SSR[b]	Ribosomal IGS[c]	cp SSR[d]	mtSSR[e]
Picea glauca	white spruce	X		X	X	dev
Picea sitchensis	Sitka spruce	X		X	X	dev
P. glauca x engelmannii	interior spruce	X	X	X	X	dev
Thuja plicata	western redcedar	X		X	dev	
Chamaecyparis nootkatensis	yellow-cedar	X			dev	
Pseudotsuga menziesii	Douglas-fir	?		X	X	
Tsuga heterophylla	western hemlock	X			X	
Larix occidentalis	western larch	X			X	
Pinus monticola	white pine	X				
Pinus taeda	loblolly pine	X			X	
Pinus contorta	lodgepole pine	X			X	

a X indicates minimum of five microsatellite markers available either from BC Research or the Genetic Data Centre at UBC. Contact C. Ritland. ? indicates not verified.

b Microsatellite markers made from expressed genes (est) in somatic embryo cDNA. Contact K. Ritland, Dept. Forest Sciences (UBC).

c High copy polymorphic probes for quantitative analysis.

d Multiplex (locus) chloroplast marker systems for paternity analysis. dev indicates genome screening is in development.

e Mitochondrial multiplex markers for maternity analysis, in development.

development and the same techniques can easily be applied to other coniferous species (Table 1).

Figure 1 shows an example of a cpDNA marker set in spruce. The figure shows a gel electrophoresis assay of cpDNA from embryos of fertilized seed. Each lane contains an embryo 'DNA fingerprint' composed of five variably sized pieces of spruce chloroplast DNA, each derived from a unique region or *locus* of the spruce cpDNA. The size differences visible between different lanes (embryos) arise from additions or deletions of DNA caused by mistakes (e.g., mutations) during DNA replication. Each of the five cpDNA loci use a letter code to define the respective size variants, which together comprise a five-letter cpDNA haplotype for each seed (e.g., sample #17 haplotype is ADCAA). Using this multilocus (e.g., five cpDNA loci) approach typical gene diversities for markers of this type are >0.9 which in practice means that 50–75% of the orchard parents will have cpDNAs that are distinguishable. At this level of resolution a good deal can be learned about orchard pollen dynamics and its impact on seedlot makeup.

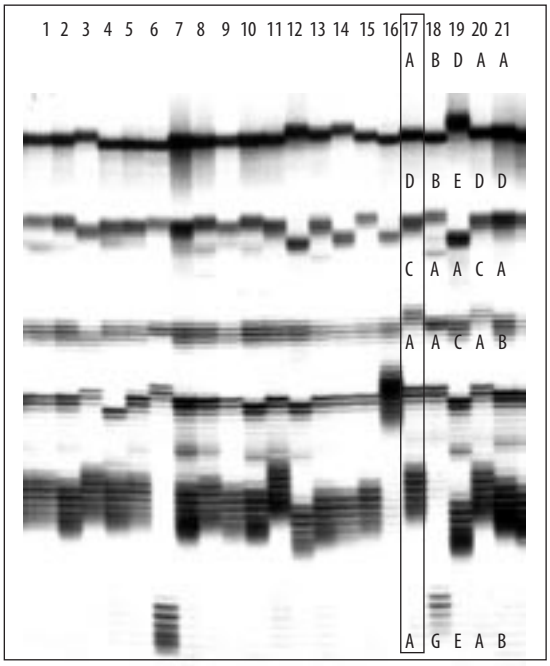

Figure 1: ***Spruce cpDNA markers.*** *Each lane (1–21) shows the haplotype profile for a seed embryo using five polymorphic loci. Letter codes for each locus size variant are shown for lanes 17–21 and a corresponding haplotype for lane 17 is boxed vertically.*

Figure 2 shows the pollen contribution in four open pollinated lodgepole pine seedlots from the Grandview Orchard #308 between 1996 and 2000 (Newton et al. in prep). The figure shows the contribution for each of the 31 distinguishable parental haplotypes (out of a total of 40 clones) compared to theoretical panmixis expected if all clones contributed randomly. In any one year, wide variation is evident between different parental clones, ranging from zero to almost three times the expected panmictic mating. The cpDNA-based pollen contribution shows good correlation with pollen strobili survey data ($R^2 = 0.65$) and confirms that fertilization success is pollen-density dependant. Note that high and low pollen contributors are consistent in each of the four seedlot years and suggests that clonal differences in pollen load show little annual variation, at least over the time span of this study. The fact that differences in parental contribution are not averaged over different years points to systemic reductions in effective population size (N_e = 18-22) compared to N_e based on census number (N = 40) or strobili surveys (N_e = 32).

Other applications include measuring pollen contamination and determining supplemental mass pollination (SMP) efficiency. Any pollen haplotypes found in seed that have no counterparts in orchard parents result from outside contamination and can be used to measure pollen inflow from adjacent orchards or pollen sources outside an orchard complex. In a coastal Douglas-fir orchard contamination rates were found to be 40% (Stoehr et al. 1998) while in lodgepole pine seedlots the levels were much lower (average 5%), consistent with the absence of nearby wild lodgepole stands (Stoehr and Newton 2002). For SMP studies, applied pollen can be collected from parental sources with unique cpDNA haplotypes. Therefore, seed found possessing these haplotypes could only have arisen from the applied pollen. In Douglas-fir, SMP success ranged from 39–73% depending on the maternal clone flowering phenology (Stoehr et al. 1998). Recent studies in lodgepole pine indicate that SMP treatments approximately double the reproductive success of applied pollen, although overall SMP efficiency is still relatively low (15%) (Stoehr et al. in prep.).

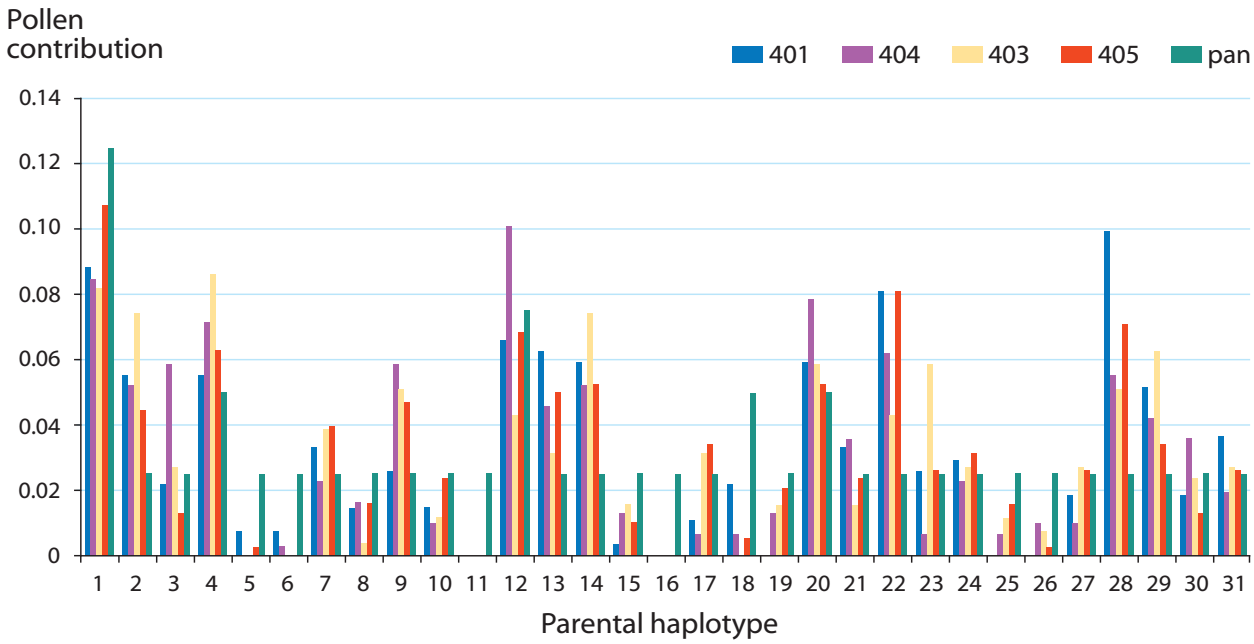

Figure 2: ***Lodgepole pine seedlot analysis using cpDNA haplotypes.*** *Embryos (N = 256–382, av. 304) from open pollinated seedlots (401, 404, 403, 405, Grandview Orchard 308) were haplotyped using a six-locus multiplex. Pollen contribution is expressed as the proportion of total seed embryos carrying an identical parental haplotype (#1–31). Panmictic contribution (pan) is the expected proportion if each orchard parent (N = 40) contributes equally to the active pollen cloud (0.025).*

If seed are collected from individual maternal clones, as opposed to bulked seedlots, questions such as deviations from random mating or selfing can be addressed. For example, if a maternal clone has a unique haplotype, seed collected from that clone which carry this haplotype could only have arisen from self-fertilization. For some clones this may comprise as much as 50% of its seed, although average rates are between 2 and 5% (Stoehr and Newton 2002). The fact that operational seedlots are generally *not* collected on a clonal basis means these questions can only be addressed if the maternal origin of a seed is known. This can be done either by analyzing the chloroplast DNA of the seed megagametophyte (which are maternally derived) or by analyzing the mitochondrial DNA of the embryos which, like other plants, are also maternally inherited in conifers. Similar RFLP-PCR genome screening techniques have recently been applied to mitochondrial genomes of conifers (Newton and Gigou 2003). The availability of these complementary organelle DNA marker systems will allow determining the origin of both maternal and paternal gametes regardless of how the seed is collected or stored.

The goal of these programs is to develop species-wide organelle DNA markers that can be used first, to validate current assumptions on seed orchard efficiency (e.g., N_e, % contamination, % selfing, and orchard design) and second, as a management tool for optimizing these parameters using research trials with now quantifiable outcomes. The benefits will be higher quality seed crops with improved genetic worth that will have higher, more quantifiable market value. Lastly these tools also can be applied to more academic questions concerning forest species genetic diversity, pollen migration, and phylogeographic studies (Marshall et al. 2002).

Nuclear DNA markers and clonal identity

While organelle markers are well suited to many applications in seed orchards, there are instances where even higher levels of genetic resolution are needed (e.g., to confirm clonal identities, analysis of specific crosses, progeny selection, gene/trait mapping and others such as wood forensics and theft cases). To achieve true *fingerprinting* levels of genetic resolution, such as in human forensic applications where 100% of a population can be distinguished, requires DNA markers derived from the nuclear or genomic DNA. However, the shear size and complexity of conifer genomes, approximately 10 times that of humans, has until recently hindered the development of useful nuclear genetic markers for BC forest species.

However, with FGC support, nuclear DNA markers are now available for most species used in BC Tree improvement programs (Table 1). The DNA markers of choice are called *microsatellites* or simple sequence repeats (SSRs) and are the nuclear equivalent of the mutation-prone or 'fast evolving' DNA sequences used in organelle marker development. Their mutation-prone nature means that individual microsatellite loci can possess as many as 30 or more alleles, have heterozygosities >0.7 and, are often co-dominant (e.g., able to detect both alleles at a diploid locus). When used in sets of 5–10 independent microsatellite loci, the combined allelic variation provides sufficient diagnostic power to discriminate virtually all individuals in a population and are currently the standard for use in human and wildlife forensic cases.

Figure 3 shows an example of forensic application of nuclear markers in Sitka spruce involving an alleged case of wood theft. The question was whether DNA extracted from the stumps of allegedly stolen trees could be matched to the DNA from tree stems found in possession of the suspect. Figure 3 shows the results obtained with one (Ss56) of the five microsatellites used to analyze 17 stem and stump DNA samples.

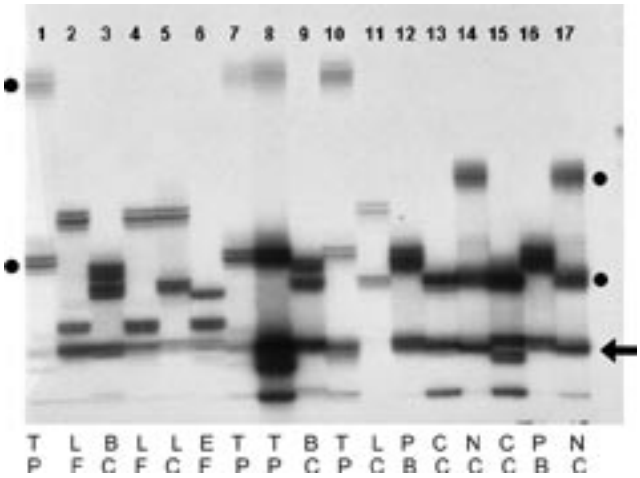

Figure 3: ***Microsatellite analysis of wood in spruce.*** *Wood DNA samples (lanes 1–17) were amplified with microsatellite marker Ss56. Alleles for lanes 1 and 17 are shown with dots and diploid genotypes are indicated beneath each lane.*

As expected for wild trees, most samples are heterozygous and show two different sized bands corresponding to both *alleles* of the diploid tree genome. The genotypes for each sample are shown with letter codes beneath each lane. The identical genotypes seen between different lanes in the figure (e.g., lanes 2 and 4, 5 and 11, 1, 7, 8 and 10, etc.) were reproduced with all five microsatellite markers and corresponded exactly to stump/stem DNA sample pairs. These results give sample 'match probabilities' on the order of 10^{-10} to10^{-12} and are strong evidence that the stolen tree stems were indeed derived from the suspected tree stumps (Newton and Vo 2002b).

Related applications include clonal verification to identify mislabelling or grafting anomalies (J. King, MOF, pers. comm.), family analysis to confirm pedigree identification (B. Jaquish, MOF, pers. comm.), and verification of clonal propagation technologies (e.g., somatic embryogenesis, M. Gilbert, CellFor Inc).

The long-term application of these, and even newer classes of genetic markers (e.g., single nucleotide polymorphism [SNP] and expressed sequence tag [EST] markers), is to provide DNA markers that are associated with specific tree genes. Current gene identification studies in conifers have progressed rapidly (Genome BC, K. Ritland and J. Bolhman, UBC) and the ability to link specific genes with associated phenotypic traits (e.g., disease resistance, growth rates, form, and wood density) may not be far off. This class of gene-specific DNA markers will provide tree breeders with tools to specifically tailor tree genotypes to maximize forest potential and to meet the needs of a changing environment.

In summary, DNA markers as tools for identification in forestry are now available to address a wide range of questions that previously could only be assumed or estimated indirectly. These new management and research tools will lead to increased confidence in tree improvement activities and the accuracy of long-term forecasts on forest health and harvest yields.

References

Marshall, H.D., C. Newton, and K. Ritland. 2002. Chloroplast phylogeography and microsatellite evolution in lodgepole pine (*Pinus contorta* var. Dougl.) Theor. App. Gen. 104:367–378.

Newton, C.H. and G. Gigou. 2003. Tools to measure sustainable forestry: Organelle markers in conifers. Forestry Innovation Initiative, R02–103 (year 1).

Newton, C.H. and T. Vo. 2001. Development of chloroplast markers in spruce and larch. B.C. Min. For., Res. Br., Victoria, BC.

Newton, C.H. and T. Vo. 2002a. Development of paternity systems for loblolly pine. Weyerhaeuser Corp., Tacoma, Wa.

Newton, C.H. and T. Vo. 2002b. Application of mircosatellite markers to identification of Sitka spruce wood samples. B.C. Min. For. and Dept. Fisheries and Oceans, Victoria, BC.

Newton, C.H., M. Carlson, and G. O'Neil. [2004] Analysis of open pollinated seedlots from a lodgepole pine seed orchard using chloroplast DNA markers. (In prep.).

Parducci, L. and A.E. Szmidt. 1999. PCR-RFLP analysis of cpDNA in the genus *Abies*. Theor. Appl. Gen. 98:802–808.

Stoehr, M.U. and C. H Newton. 2002. Evaluation of mating dynamics and pollen contamination in a lodgepole pine seed orchard using chloroplast DNA markers. Can. J. For. Res. 32:469–476.

Stoehr M., H. Mehl, G. Nicholson, G. Pieper, and C. Newton. [2004] Determination of SMP efficiency in a lodgepole pine orchard in British Columbia using chloroplast DNA markers. B.C. Min. For., Res. Br., Victoria, BC. (In prep.).

Stoehr, M.U., B. Orvar, T. Vo, B. Gawley, J. Webber, and C. H. Newton. 1998. Application of a chloroplast DNA marker in Seed Orchard Management: evaluations of Douglas-fir. Can. J. For. Res. 28:187–195.

Sutton, B.C.S., S.C. Pritchard, J.R. Gawley, C.H. Newton, and G.K. Kiss. 1994. Analysis of Sitka/interior spruce introgression in British Columbia using cytoplasmic and nuclear DNA probes. Can. J. For. Res. 24:278–285.