

# The use of DNA technology to advance the Sitka spruce breeding programme

**Steve Lee, Stuart A'Hara and Joan Cottrell**

Improved planting stocks are now available for a number of conifer species including Sitka spruce (*Picea sitchensis*), Scots pine (*Pinus sylvestris*), Corsican pine (*Pinus nigra*) and hybrid larch (*Larix x eurolepis*) (Lee, 2004). The greatest effort and progress has been made with Sitka spruce for which volume gains of improved planting stocks at rotation age have been predicted to be between 21% and 29% (Lee and Matthews, 2004). The process of selection, testing and breeding is long-term and costly. Final selections for breeding populations are currently made around 9 years after planting, based on data collected in replicated field trials. The cycle length between selection, testing and making improved material available to the industry is around 20 to 25 years. Therefore, despite the fact that the conifer breeding effort commenced in the early 1960s, it was not until the late 1980s that improved planting stocks became generally available.

Biotechnological approaches are continually developing and are recognised as potentially powerful tools for tree breeding and conservation (Lee *et al.*, 2004). The development of DNA based molecular markers may offer a specific tool to enable tree breeders to shorten this breeding cycle and improve the accuracy of selection. If DNA markers can be identified that are consistently associated with characteristics that can only be measured when a tree is much older, then there is the opportunity for early selection of trees in the laboratory by screening for those markers.

The Conifer Tree Breeding Unit in Forest Research (FR) has recently started a project aimed at utilising DNA technologies within the breeding of Sitka spruce. The project is referred to as Marker Aided Selection (MAS) and its objective is to shorten the breeding cycle and improve the rate of genetic gain that reaches the forestry industry. If success in the development of clonal forestry techniques continues, it could be possible to mass produce trees selected in the laboratory and make them available to forest managers in as little as 2 to 5 years following initial screening.

### Linking phenotypic performance with DNA markers

Suitable field trials are required to identify and verify trait and DNA marker associations. Early attempts to apply this technology elsewhere in the world relied on existing field trials with just a restricted number of progeny (100 to 200) per family. These studies often found marker and trait associations in a limited number of trees but there was no facility for verifying the findings in larger numbers of progeny or in other families (cf. Devey *et al.*, 2004). Elsewhere, simulation studies suggested that larger progenies of the order of 1000 trees were actually required to verify marker/trait associations with improved accuracy (Beavis, 1994, 1998). One successful study of identification and verification involved measurements on nearly 4500 trees from a single full-sibling radiata pine (*Pinus radiata*) family planted in a commercial forestry plantation (Devey *et al.*, 2004).

It was important that the design of any new trial should contain sufficient genotypes to allow accurate trait and marker associations. Soller and Beckmann (1990) emphasised the additional accuracy gained from the use of clonal replication of progenies, but even then, as many as 1260 genotypes may be required per family to provide Type I error 0.01 and power 0.80. Bradshaw and Foster (1992) agreed that the successful application of MAS depends upon precise estimation of phenotypic variation attributable to the markers and that clonal testing is the best experimental design to achieve this goal.

In spring 2005, FR planted three large clonal tests designed to investigate the association between markers and phenotypic variation in Sitka spruce. The experiments involve the same 1500 progeny from each of three full-sibling families. Each progeny (or genotype) is replicated four times at a single site and the whole experiment is repeated at three climatically contrasting sites (Figures 1, 2 and 3). In this way,

breeders can investigate whether the marker/trait associations are stable across a range of environmental conditions. If markers are found to be specific only to a certain climate type or family, then the technology becomes of limited value. Each family consists of contrasting parents for which we have good knowledge of their genetic quality, e.g. a parent of good stem straightness crossed with a crooked parent; or a parent of high wood density crossed with one of low wood density.

The replicated clones representing each family in the field will be measured in the normal way for a range of characteristics, such as wood density, stem straightness or growth rate. The top and bottom echelon for a given trait (perhaps 5%, equivalent to 75 clones) will have tissue collected from them for further investigation in the laboratory. Breeders will then look for markers which are consistently associated with a given trait. If a high proportion of the total variation in, for example, wood density is associated with a limited number of markers, then in the future selection could take place in the laboratory by choosing trees which possess this complement of markers. Devey *et al.* (2004) found that eight markers explained more than 14% of the variation of juvenile wood density in radiata pine.

In a similar way, it is hoped that other markers will be identified which are associated with good stem straightness, fast growth rate, good insect resistance or perhaps high frost tolerance. It is possible that for some correlated traits, such as wood density and growth rate, the same markers for one trait may show an association with another. However, it is more likely that individual traits will have associations to a unique set of markers. This means that breeders using this technology will require a comprehensive set of markers, evenly spread across the genome, to enable these different marker/trait associations to be identified.

# The use of DNA technology to advance the Sitka spruce breeding programme

**Figure 1** Location of the three trials, planted in spring 2005 at Huntly (north Scotland), Llandoverly (south Wales) and Holsworthy (north Devon).



**Figure 2** Planting on the upland site near Huntly.



**Figure 3** Successful establishment on the mid-ranged Llandoverly site.

## Composing a genetic linkage map for Sitka spruce

Although the trial has only recently been planted and the plants are too young for the traits of interest to be measured, work has begun on marker development and the construction of a rudimentary genetic linkage map. The objective of map development is to determine the relative position of DNA markers in the genome, with the goal of building up a suite of markers that provide an even coverage across all chromosomes. The positions of the markers on the map are based on the frequencies of recombination

between markers during crossover of homologous chromosomes at meiosis. At a simplistic level, the greater the frequency of recombination (segregation) between two genetic markers, the farther apart they are assumed to be. Conversely, the higher the frequency of association between any given markers, the smaller the physical distance between them. A rudimentary map has already been created (Figure 4).

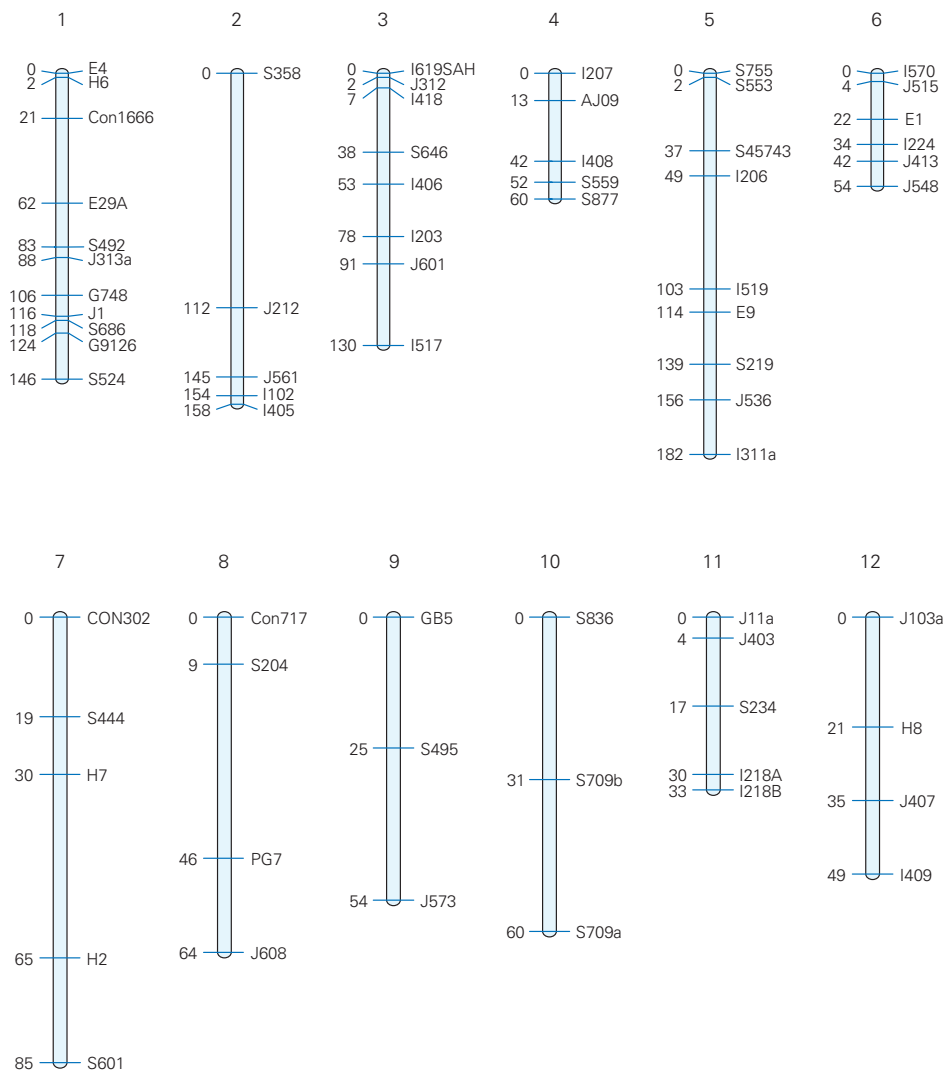
As field data for different traits become available for analysis, the map will be used to identify regions of the genome which are consistently associated with the traits under investigation. This part of the work

# The use of DNA technology to advance the Sitka spruce breeding programme

involves the use of powerful computer software, which analyses the molecular data in conjunction with the field data and identifies robust marker/trait associations.

We are initially using a type of molecular marker known as a microsatellite. A microsatellite is a small length of DNA comprising a short motif which is repeated a variable number of times, e.g. (CA)<sub>5</sub> would represent the dinucleotide repeat CACACACACA and (ATT)<sub>5</sub> would represent the trinucleotide repeat ATTATTATTATTATT. These are highly informative markers as they are codominant (the alleles from both parents can be tracked in the offspring) and highly

polymorphic (variable between individual genotypes). Microsatellites were historically expensive to develop because cloning and sequencing from custom-made genomic libraries was required for their discovery. In addition, the large, repetitive nature of conifer genomes made single-locus microsatellite discovery a low return to effort process. However, the recent public release of expressed sequence tag (EST) information – the sequences of transcribed genes – from a number of *Picea* species in DNA databases has opened up this avenue of microsatellite development. ‘Datamining’ is the term used to describe discovery of microsatellite sequences from DNA databases.



**Figure 4** Preliminary genetic linkage map for Sitka spruce comprising 68 microsatellite markers. There are markers on all 12 chromosomes and as more are found the number on each chromosome will increase to ensure even coverage across the genome.

## The use of DNA technology to advance the Sitka spruce breeding programme

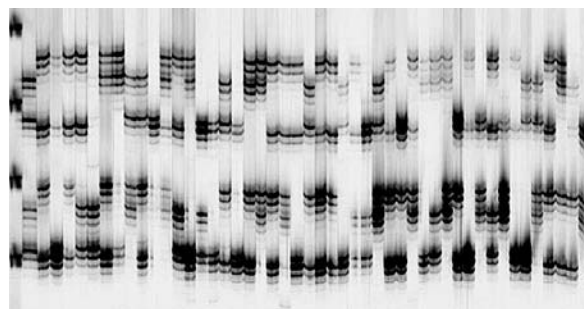
FR is searching for 200 microsatellites to create the framework map (Figure 5). Once this major step is completed, other classes of marker will be used to increase the marker density and ensure there are no gaps on the map. Microsatellite markers are often cross-species transferable, thereby allowing maps with other species to be combined. The other classes of markers, known as dominant markers (e.g. AFLPs and ISSRs) are not as informative as microsatellites and cannot be transferred across species, but can be generated in greater numbers and are less costly to develop. There are also plans to explore the use of a further type of marker known as single nucleotide polymorphisms (SNPs).



**Figure 5** Preparing Sitka spruce needles for DNA extraction prior to PCR. FR scientists hope to amplify 200 microsatellite markers and position them on a framework map.

Currently, we have 180 of our 200 microsatellite target: 50 genomic and 150 EST based following datamining of *Picea* databases. Given that there were only six available in the literature when the project started in 2004, substantial progress has been made.

All of these molecular techniques are made possible by a process called the polymerase chain reaction (PCR), which can amplify many millions of copies of



**Figure 6** Example of a microsatellite gel. Ninety-six individual trees (two banks of 48) are scored for one microsatellite marker. A DNA size standard is run at either side of the gel.

specific fragments of DNA, allowing scientists to visualise the fragments on a gel (Figure 6). PCR uses short pieces of DNA, known as primers, which bind to unique, non-variable, regions of DNA on either side of a variable microsatellite motif. The process then cycles through a number of heating and cooling steps which accurately multiply the targeted region of DNA at an exponential rate. Primers are designed from the DNA sequence information to be specific for a particular microsatellite. Each microsatellite therefore has its own unique primer pair which will amplify only that microsatellite. The difference in the number of repeats between the primers produces a length difference that can be seen on the gel.

The MAS clonally replicated field trials are now entering their second growing season and the linkage map is about 12 months away from completion. It is intended that around six years from now we shall be trying to find an association between particular markers and the first trait measured in the field – stem straightness or wood density are the prime candidates. Markers for other characteristics are expected to follow.

### Collaboration with other partners

This area of research is the focus of considerable international interest. Forest Research has joined forces with Genome Canada based at the University of British Columbia, Vancouver to develop a joint linkage map between Sitka spruce and white spruce (*Picea glauca*): ([www.genomecanada.ca/xresearchers/researchPrograms/projects](http://www.genomecanada.ca/xresearchers/researchPrograms/projects)). In addition, the new clonal tests are also seen as a robust resource for

verifying marker and trait associations which may have been identified by any third party.

Elsewhere, a collaborative project with Southampton University aims to determine the genes associated with traits such as wood density. This work will hopefully provide the basis for selecting those genes in which SNP markers will be developed and has the potential of finding a shortcut marker to a candidate gene directly influencing a given trait.

## Future value of these new tools

If it proves possible to identify genetic markers in the laboratory which explain a sufficient proportion of the total phenotypic variation for a given trait in the field, then expensive and time-consuming field trials may become a thing of the past. In practice, breeders will need to be cautious and thorough verification will be necessary before deployment of selected genotypes.

DNA marker technology is developing rapidly. The computer software to help find the markers is also becoming cheaper and more sophisticated. Forestry can benefit from the molecular advances made in more valuable crops such as rice, tomato and poplar. By planting the field trials and developing a linkage map of the Sitka spruce genome, we are at the forefront of biotechnology on a global scale and are already collaborating with partners who share our interests.

Our vision is that ten years from now screening for clones of high wood density, good stem straightness, perhaps fine branches and good growth rate, will be aided by work in the laboratory. Once selected, many copies will be made of the chosen clones for rapid deployment to the field. The forest industry will benefit from planting a more uniform stock which grows faster and is better targeted to satisfy the construction timber market. Stands of trees will not only produce more volume of timber at a given age but will also produce a greater proportion which satisfies higher log grades in the forest and construction timber grades at the sawmill.

## Acknowledgement

This article is based on an article published in *Forestry and British Timber*, January 2006, pages 14–16.

## References

- Beavis, W. D. (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC, 250–266.
- Beavis, W. D. (1998). QTL analysis: power, precision and accuracy. In: *Molecular dissection of complex traits*, ed. A. H. Paterson. CRC Press, Boca Raton, USA, 145–162.
- Bradshaw, H. D. and Foster, G. S. (1992). Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. *Canadian Journal of Forest Research* **22**, 1022–1049.
- Devey, M. E., Carson, S. D., Nolan, M. F., Matheson, A. C., Te Riini, C. and Hohepa, J. (2004). QTL associations for density and diameter in *Pinus radiata* and the potential for Marker Aided Selection. *Theoretical and Applied Genetics* **108**, 516–524.
- Lee, S. J. (2001). Selection of parents for the Sitka spruce breeding population in Britain and the strategy for the next breeding cycle. *Forestry* **74** (2), 129–143.
- Lee, S. J. (2004). *The products of conifer tree breeding in Britain*. Forestry Commission Information Note 58. Forestry Commission, Edinburgh.
- Lee, S. J. and Matthews, R. (2004). *An indication of the likely volume gains from improved Sitka spruce planting stock*. Forestry Commission Information Note 55. Forestry Commission, Edinburgh.
- Lee, S. J., Cottrell, J. and John, A. (2004). *Advances in biotechnology: powerful tools for tree breeding and genetic conservation*. Forestry Commission Information Note 50. Forestry Commission, Edinburgh.
- Soller, M. and Beckmann, J. S. (1990). Marker-aided mapping of quantitative trait loci using replicated progenies. *Theoretical and Applied Genetics* **80**, 205–208.