Genomics applied to the study of adaptation in pine species

M.A. Guevara, A. Soto, Carmen Collada, Christophe Plomion, O. Savolainen, D.B. Neale, Santiago C. González-Martínez, M.T. Cervera

Invest Agrar: Sist Recur For (2005) 14(3), 292-306

Genomics applied to the study of adaptation in pine species

M. A. Guevara[1], A. Soto[2], C. Collada[3], C. Plomion[4], O. Savolainen[5], D. B. Neale[6], S. C. González-Martínez[1] and M. T. Cervera[1]*

[1] *Genética y Ecofisiología Forestal, Departamento de Sistemas y Recursos Forestales. CIFOR - INIA. Carretera de La Coruña km 7. 28040 Madrid. Spain*

[2] *Unidad de Anatomía, Fisiología y Genética Vegetal. Escuela Técnica Superior de Ingenieros de Montes (UPM). Ciudad Universitaria s/n. 28040 Madrid. Spain*

[3] *Departamento de Biotecnología. ETSIM. Ciudad Universitaria s/n. 28040 Madrid. Spain*

[4] *UMR BIOGECO - INRA. Equipe de génétique. 69 route d'Arcachon. F-33612 Cestas Cédex. France*

[5] *Department of Biology. University of Oulu. FIN-90014 Oulu. Finland*

[6] *Department of Plant Sciences. University of California. One Shields Avenue. 95616 Davis CA. USA*

Abstract

This manuscript provides an overview on the use of genomics to study pine adaptation to environmental conditions, such as drought or photoperiod. We briefly discuss molecular tools that permit to acquire a comprehensive view of pine genome, such as cytogenetic and genetic linkage maps. We will also discuss the application of two complementary strategies to dissect complex adaptive traits based on linkage disequilibrium between neutral molecular markers and quantitative trait loci: QTL mapping and association analysis. An important step in these studies is the identification of candidate genes involved in the response of pine trees to environmental variation using functional genomic approaches.

Key words: *Pinus*, genetic maps, QTLs, association mapping, ESTs.

Resumen

Aplicaciones de la genómica al estudio de la adaptación de los pinos

Este artículo constituye una revisión de las aplicaciones de la genómica en el estudio de la adaptación de los pinos a diferentes condiciones medioambientales, tales como la sequía o el fotoperíodo. Discutiremos brevemente las herramientas moleculares desarrolladas para adquirir una visión integrada del genoma de pino como por ejemplo la proporcionada por los mapas citológicos y genéticos. También discutiremos la aplicación de dos estrategias complementarias para diseccionar caracteres cuantitativos complejos, basadas en el desequilibrio de ligamiento entre marcadores moleculares neutrales y los *loci* que causan variación cuantitativa: mapeo de QTLs y estudios de asociación. Otro aspecto importante en estos estudios es la identificación de genes candidatos involucrados en la respuesta de los pinos a variaciones ambientales empleando estrategias de genómica funcional.

Palabras clave: *Pinus*, mapas genéticos, QTLs, mapeo de asociación, ESTs.

Introduction

Plant genomics aims to understand how genes function in plants and how plant genomes evolved. Genomics has become an integral part of forest tree research and been applied to unravel mechanisms controlling tree development and interaction with its environment. During their life-span trees have to adapt to environmental stress situations, which influence the expression of their full genetic potential for growth and reproduction. Tree survival depends on the capacity to develop plastic responses to the changing environmental conditions. Trees, such as pines, are good models to study the genetics of adaptation because of their large and weakly structured natural populations, random mating, high level of nucleotide diversity, rapid

* Corresponding author: cervera@inia.es
Received: 30-06-05; Accepted: 25-07-05.

decay of linkage disequilibrium and long history of ecological genetic research (Morgenstern, 1996; Neale and Savolainen, 2004). A complete view of the molecular basis of adaptation requires cross-disciplinary collaboration among experts in molecular biology, population and quantitative genetics, ecophysiology, forest management, statistics and bioinformatics. An understanding of the principles involved in tree adaptation to environmental stress will enable optimisation of practices for sustainable forest management, thereby minimising environmental impact.

The aim of this chapter is to present an overview of the most common genomic tools and strategies to approach studies on pine adaptation. We will briefly describe pine genome structure (i.e., composition, molecular markers as well as physical, genetic and cytological maps, and comparative mapping to integrate information). Secondly, methods for the analysis of genetic control of complex adaptive traits, including both QTL and association mapping, are presented. Finally, large-scale EST sequencing and gene expression profiling are discussed as key tools for gene discovery and analysis of gene function in conifer species, for which the whole genome DNA sequence is not expected to be available in the foreseeable future.

Pine genome structure

Some notes on pine genome

Pines are diploid organisms characterized by large and complex genomes, which make difficult genomic analysis. Nuclear DNA content, as estimated by 2C DNA value, is highly variable among pine species (Grotkopp *et al.,* 2004), ranging from 43.96 pg (*P. banksiana*) to 75.36 pg (*P. gerardiana*) (1 pg = 960 kb, Arumuganathan and Earle, 1991). However, pines are known to have very slow chromosomal evolution: all of them share 12 pairs of chromosomes, which are morphologically similar (Saylor 1961). Although 75% of the pine genome corresponds to highly repeated sequences, there is 25% of low- to single-copy DNA. Pine sequencing has revealed a high number of repeats including retrotransposons (Friesen *et al.,* 2001 and listed references), simple sequence repeats (Schmidt *et al.,* 2000) and complex gene families resulting from duplication events (Kinlaw and Neale, 1997).

Molecular markers used to analysed pine genome

Due to development in molecular genetics, a variety of different molecular techniques have emerged during the last few decades to genotype any plant species. These genetic markers differ in important features such as genomic abundance, level of polymorphism, locus specificity, reproducibility, technical requirements and cost. Therefore, the key-question is to choose the appropriate marker for each study. Molecular markers can be classified according to their dominant / codominant nature (see reviews Karp and Edwards, 1998; Cervera *et al.,* 2000 and Table 1 in Krutovsky and Neale, 2001). Dominant markers, such as RAPDs (Random Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), SAMPLs (Selective Amplificafion of Microsatellite Polymorphic Loci), S-SAPs (Sequence Specific Amplification Polymorphisms) and ISSRs (Inter-Simple Sequence Repeat amplification) (acronyms and references are provided in http://www.dpw.wau.nl/PV/aflp/acronyms.html), require no prior sequence information and therefore are used to generate a high number of anonymous markers in any pine species. Codominant markers, such as RFLPs (Restriction Fragment Length Polymorphism), SSRs (Simple Sequence Repeats), SCARs (Sequence Characterized Amplified Regions) and ESTPs (Expressed Sequence Tag Polymorphisms) or gene-based markers, are sequence-dependent and therefore can be transferred among species, easily exchanged across labs and used to identify homologous linkage groups. Although codominant markers are laborious to identify, up to now, a remarkable number of RFLPs, SSRs and ESTPs markers have been developed in pines species (for further information, see Table 1 from Chagné *et al.,* 1994a and http://dendrome.ucdavis.edu/Gen_res.htm). It is still always important to assure that the different loci really are orthologous in all species to be compared, rather than closely related paralogous regions.

Pine genome mapping

Physical maps of low-copy regions

Physical mapping consists in placing nucleotide sequences with respect to a DNA matrix. The large

Table 1. List of genetic linkage maps constructed for different pine species (modified and completed from «Inheritance and mapping studies in gymnosperms», available at http://www.pierroton.inra.fr/genetics/labo/mapreview.html)

Subgenus (Section)	Species	Marker type	Population*	References
Pinus (Pinus)	*Pinus sylvestris*	RAPDs	OB	Yazdani *et al.,* 1995
		RAPDs	F1	Hurme and Savolainen, 1999
		AFLPs	F1	Lerceteau *et al.,* 2000
		AFLPs	F1	Yin *et al.,* 2003
		ESTPs	F1	Komulainen *et al.,* 2003
	Pinus thunbergii	AFLPs	OB	Hayashi et al., 2001
Pinus (Pinea)	*Pinus pinaster*	Proteins	OB	Bahrman and Damerval, 1989
		Proteins	OB	Gerber *et al.,* 1993
		Proteins, RAPDs	F2 S	Plomion *et al.,* 1995a
		RAPDs	F2 S	Plomion *et al.,* 1995b
		RAPDs	F2	Plomion *et al.,* 1996
		Proteins	F2 S	Plomion *et al.,* 1997
		Proteins, RAPDs, AFLPs	F2 S	Costa *et al.,* 2000
		AFLPs	TG OB	Chagné *et al.,* 2002
		AFLPs	F1	Ritter *et al.,* 2002
		SSRs	F2	Mariette *et al.,* 2001
		ESTPs	F2 S	Plomion *et al.,* 1999
		ESTPs	TG OB	Chagné *et al.,* 2003
	Pinus brutia	RAPDs	OB	Kaya and Neale, 1995
	Pinus halepensis	RAPDs	OB	Gómez *et al.,* 2002
Pinus (Trifoliis)	*Pinus contorta*	RAPDs	OB	Li and Yeh, 2001
		AFLPs, ESTPs	F1	Hayashi *et al*., 2002
	Pinus caribaea	RAPDs	F1	Dale and Teasdale, 1995
	Pinus elliottii × *Pinus caribaea*	AFLPs	F1	Shepherd *et al.,* 2003
	Pinus elliottii	RAPDs	OB	Nelson *et al.,* 1993
		RAPDs	F1	Kubisiak *et al.,* 1995
		RAPDs	F1	Dale and Teasdale, 1995
	Pinus palustris × *Pinus elliottii*	RAPDs	F1	Weng *et al.,* 2002
	Pinus palustris	RAPDs	OB	Nelson *et al.,* 1994
		RAPDs	F1	Kubisiak *et al.,* 1995
		RAPDs	OB	Kubisiak *et al.,* 1996
	Pinus taeda	Isozymes, RFLPs	TG OB	Devey *et al.,* 1994
		Isozymes, RFLPs	TG OB	Groover *et al.,* 1994
		RFLPs	TG OB	Sewell *et al.,* 1999
		RFLPs	TG OB	Devey *et al.,* 1999
		AFLPs	OB	Remington *et al.,* 1999
		SSRs	TG OB	Zhou *et al.,* 1994
		ESTPs	TG OB	Harry *et al.,* 1998
		ESTPs	TG OB	Temesguen *et al.,* 2001
		ESTPs	TG OB	Brown *et al.,* 2001
		ESTPs	TG OB	Krutovsky *et al.,* 2004

Subgenus (Section)	Species	Marker type	Population*	References
	Pinus radiata	Proteins, RAPDs, SSRs	TG OB	Devey *et al.,* 1996
		Proteins	TG OB	Devey *et al.,* 1999
		RAPDs	OB	Kuang *et al.,* 1999
		SSRs	F1	Fisher *et al.,*1996
		SSRs	F1	Fisher *et al.,* 1998
		SSRs	F1	Wilcox *et al.,* 2001
		ESTPs	F1	Cato et al., 2001
Ducampopinus (*Parryana*)	*Pinus edulis*	AFLPs	OB	Travis *et al.,* 1998
Strobus (*Strobus*)	*Pinus strobus*	RAPDs, SSRs, ESTPs	OB	Echt and Nelson, 1997

*OB = outbred pedigree (megagametophytes)
F2 S = F2 selfed family (megagametophyte)
TG OB = three-generation outbred pedigree
F1 = full-sib progeny
F2 = self-pollinated progeny

genome size of pines makes unaffordable to develop a high throughput sequencing effort to obtain their complete DNA sequences in the near-future. However, this strategy could be focussed on sequencing the low-copy fraction of the genome through the construction and characterization of BAC (Bacterial Artificial Chromosome) libraries enriched in this genome fraction. Up to now, partial BAC libraries are under construction for *Pinus taeda* (http://www.srs.fs.usda.gov/about/newsrelease/nr_2004-07-27-pinegene.htm) and *Pinus pinaster* (Francisco Cánovas, personal communication).

Genetic linkage maps

A genetic map results from the linkage analysis of markers segregating in a mapping population according to Mendelian ratios. Marker loci are ordered along linkage groups based on genetic distances (related to recombination distances). Therefore, a good genetic linkage map consists of a set of Mendelian marker loci that are evenly spaced and span the full genome. Different software packages have been developed to carry out statistical analysis estimating recombination frequency (distance) and locus ordering (further information on mapping softwares at URL http://linkage.rockefeller.edu/soft/list.html).

Different strategies have been followed to construct genetic linkage maps of *Pinus* species, all of them based on the high level of genetic variability detected in most of these outbreed species:

I) *Haploid mapping strategy.* This strategy is based on the construction of a linkage map of a single conifer tree analysing marker segregation from megagametic haploid DNAs extracted from progeny seeds. This haploid tissue represents a single meiotic event, being genetically equivalent to a maternal gamete. All heterozygous markers segregate 1:1. In 1981, Conkle proposed and used megagametophytes to study co-segregation of allozymes in conifer species. Since then, a high number of genetic maps have been constructed for single trees of representative pine species using the haploid mapping strategy combined with different molecular tools, including protein and DNA markers (listed in Table 1).

II) *Diploid mapping strategies*, which include the *two-way pseudo-testcross strategy*, the *F2 inbred model* and the *three-generation outbred model*. The *two-way pseudo-testcross strategy* is based on selection of polymorphic markers heterozygote in one parent and homozygote null in the other parent and therefore segregating 1:1 in their F1 progeny as in a testcross configuration (Grattapaglia and Sederoff, 1994). Two linkage maps can be generated, one for each parent.

The *F2 inbred model* is based on a three-generation pedigree for which the grandparents are treated as inbred lines (represented by AA and BB), and therefore three genotypes occur at any locus: AA, AB and BB, segregating 1:2:1. A combined parental map from the F2 progeny data using the intercross mating type can be assembled. The *three-generation outbred model* (Sewell *et al.,* 1999) is an extension of the pseudo-testcross strategy. Within a single outbred pedigree, all co-dominant markers will segregate in one of three different ways:

— 1:1 (i.e., testcross mating type): one progenitor is heterozygous and the other is homozygous.
— 1:2:1 (i.e., intercross mating type): both parents are heterozygous and have the same genotype.
— 1:1:1:1 (i.e., fully informative mating type): both parents are heterozygous and have different genotypes.

Segregation data is subdivided into two independent data sets, one per progenitor, and independent maps are constructed for each parent. A sex-average map is then constructed using fully informative and intercross markers as common anchor-points between each parental data set.

A high number of linkage maps have been constructed for a representative number of pine species using these strategies combined with different marker techniques (listed in Table 1).

Comparative mapping

Comparative mapping allows to compare genome structure not only across pedigrees within pine species and among pine species but also with other conifers, and sheds light on genome organization, species differentiation and evolutionary history. Additionally, comparative mapping allows to transfer information, such as QTLs and candidate gene location, across species. It requires the use of a common set of co-dominant markers (orthologous loci) to analyse marker content (synteny) and order (colinearity). This approach is crucial to acquire a compressive view of the structure of large and complex genomes as those of pine species. These genomes contain more «junk» DNA (i.e., portions of the DNA sequence for which no function has been identified) and more duplications than other plant species and, in the absence of a complete genome sequence, much more efforts are required to locate and characterize specific genes.

The Conifer Comparative Genomics Project was created to compare genetic maps between several pine species and other conifers using low-copy cDNA markers developed primarily in loblolly pine (http://dendrome.ucdavis.edu/ccgp/). Thus, different pairwise alignments between the loblolly pine reference map and other pine species such as slash pine (*Pinus elliottii*; Brown *et al.,* 2001), maritime pine (*P. pinaster*; Chagné *et al.,* 2003), scots pine (*P. sylvestris*; Komulainen *et al.,* 2003), Monterey pine (*P. radiata*; Devey *et al.,* 1999), and even more distant species of Pinaceae such as Douglas fir (*Pseudotsuga menziesii*; Krutovsky *et al*., 2004), sugi (*Cryptomeria japonica*; Naoki Tani, in progress) and Norway spruce (*Picea abies*; Michela Troggio, in progress) revealed a high degree of synteny and colinearity for the set of markers used. Moreover, Chagne *et al.* (2003) observed conservation of QTLs for wood properties and their colocation with candidate genes between maritime pine and loblolly pine, which revealed a common genetic control of wood traits between pine species that diverged approximately 120 My ago.

Cytogenetic maps

The division of genomic DNA into independent chromosomes is a fundamental feature of genome architecture. Karyotype studies of different pine species have been performed using FISH and fluorescence banding pattern analysis (Doudrick *et al.,* 1995; Jacobs *et al.,* 2000; Hizume *et al.,* 2002; Friesen *et al.,* 2001; Shibata *et al.,* 2005). Different probes, such as telomere and centromeric repeats as well as rDNA and retrotransposon sequences were used to analyse their distribution. Both rDNA and DAPI (a blue fluorescent DNA stain applied in cytochemical research) positive bands showed chromosome-specific distribution patterns and thus allowed discrimination of all the chromosomes. Comparative analysis of three American and four Eurasian pine species (subgenus *Pinus*) revealed the existence of homeologous chromosomes, although significantly different FISH and fluorescent banding patterns among subsections (Hizume *et al.,* 2002). Comparative karyotypic analysis may be an important tool to understand species

differentiation. Also, karyological studies have shown aberrations in karyotypes of Siberian stone pines growing in bogs, which could be associated to environmental stress (Sedelnikova and Muratova, 2002). Finally, construction of karyotypes and unambiguous identification of chromosomes is a requirement for the integration of genetic and partial physical maps of pine species, which might be accomplished by the use of BAC clones as FISH probes.

Analysis of the genetic control of complex adaptive traits

Quantitative trait loci (QTL) mapping

The genetic dissection of complex adaptive traits has traditionally relied on identification of *quantitative trait loci* (QTL), with application both in molecular breeding and evolutionary biology (see reviews in Mauricio, 2001; Doerge, 2002). Quantitative genetics usually defines a quantitative trait in terms of variances, partitioning the total phenotypic variance into a genetic and an environmental component. Then, QTLs are defined as genetic loci or chromosomal regions that contribute to the genetic variance in complex quantitative traits, as identified by statistical analysis. The identification of QTLs requires (1) relatively dense genetic maps with evenly distributed markers, (2) appropriate statistical tools for the specific cross, and (3) an experimental population segregating for both genetic markers and phenotypic traits. A variety of experimental designs have been proposed, generally based on inbred line crosses or outbred populations (see review in Lynch and Walsh, 1998 and links to software for QTL analyses at http://www.stat.wisc.edu/~yandell/qtl/software/). In pines, typical outcrossing mating systems prevents the construction of inbred lines, and outbred strategies for QTL mapping have been favoured. Because QTL effects are expressed as means with inbred lines, but as genetic variances in outbred approaches, estimates from outbred populations are expected to be inherently less precise.

In pines, QTL mapping is typically performed using general linear models and/or likelihood methods such as interval mapping (e.g., Brown *et al.,* 2003; Yazdani *et al.,* 2003; Pot, 2004). Simple *Interval Mapping* (IM) uses one or two pairs of flanking markers to assess the probability that an interval between two markers is associated with a QTL (Lander and Botstein, 1989). *Composite Interval Mapping* (CIM) assesses the probability that an interval between two markers is associated with a QTL, while controlling for the effects of other markers on the trait of interest, and *Multiple Interval Mapping* (MIM) uses multiple marker intervals simultaneously to construct multiple putative QTLs in the mapping model. MIM can also analyse epistatic QTLs and estimate the individual genotypic value and the heritabilities of quantitative traits (Kao *et al.,* 1999). Power and accuracy in QTL detection is closely related to the size of the mapping population. In most cases, more than 500 plants should be studied (Beavis, 1994, 1998).

Several QTL mapping studies have been developed for growth, drought-tolerance and wood quality traits in pine (Table 2). Most reliable QTLs are those that have been validated at different times, space, and/or genetic backgrounds (see, for example, Brown *et al.,* 2003). Research in QTL mapping has produced informative results with respect to (1) the number and genome location of major genes affecting a quantitative trait, (2) the magnitude and mode of action of major genes, (3) analysis of epistatic interactions and (4) development of molecular breeding programs. However, the precision of mapped QTLs is usually low in pines, and the QTLs intervals are very wide. Indeed, given the large genome size, a typical QTL interval in pines can cover thousands of genes. For this reason, positional or map-based cloning to find genes underlying quantitative traits is not feasible in pines and other approaches such as association mapping are rapidly becoming popular.

Association mapping

While QTL identification is based on linkage disequilibrium (LD) generated in a few generations of crossing, *association mapping* takes advantage of events that created association in the distant past to find statistical association between a neutral marker allele and a phenotype (Hirschhorn and Daly, 2005; Jannink and Walsh, 2002; Neale and Savolainen, 2004 for conifers). Indeed, after many generations of recombination and random mating only tightly linked loci will show statistical association, allowing a finer

Table 2. Selected QTL studies for four major pine crops: *Pinus sylvestris*, *P. pinaster*, *P. taeda* and *P. radiata* (modified and completed from *Pine Genomics Workshop*, May 21-22, 2003, University of California, Davis, CA; available at http://dendrome.ucdavis.edu/lpgp/); na: information not available.

Species	Population*	Traits	Number of QTLs per trait / variance explained per QTL	References
Pinus sylvestris				
	F1 full-sib (backcross)	Bud set Frost hardiness	1-8 QTLs / 3-13%	Hurme *et al.,* 2000
	F1 OB	Frost hardiness Wood density Branch diameter Growth	0-3 QTLs / 9-23% (phenotypic)	Lerceteau *et al.,* 2000
	F1 OB	Frost hardiness Growth	0-4 QTLs / 7-79% (per trait)	Yazdani *et al.,* 2003
Pinus pinaster				
	F2 S	Growth	1-3 QTLs / 5-14%	Plomion *et al.,* 1996
	F2 OB	Ring width Water-use efficiency	1-3 QTLs / 6-12% (phenotypic)	Brendel *et al.,* 2002
	F1 OB	Height and diameter Wood physical and chemical properties	10-40 QTLs / 4-18% (phenotypic)	Markussen *et al.,* 2003
	na	Wood physical and chemical properties	na / 4-12% (phenotypic)	Pot, 2004
Pinus taeda				
	F2 OB	Specific gravity	5 QTLs / 23% (per trait, phenotypic)	Groover *et al.,* 1994
	F1 S	Inbreeding depression Growth Survival	3-4 QTLs / 2-14% (phenotypic)	Remington and O'Malley, 2000
	F2 OB	Wood density Microfibril angle % late wood	5-9 QTLs / 5-15% (phenotypic)	Sewell *et al.,* 2000
	F2 OB	Wood chemistry traits	8 QTLs / 5-13% (phenotypic)	Sewell *et al.,* 2002
	F2 OB	Wood physical and chemical properties	5-12 QTLs / 2-16% (phenotypic)	Brown *et al.,* 2003
Pinus radiata				
	F1 OB	Growth	2 QTLs / 9-10%	Emebiri *et al.,* 1998
	F1 S	Inbreeding depression Survival	9 regions / sub-lethal to lethal	Kuang *et al.,* 1999
	F1 OB	Wood density	1-2 QTLs	Kumar *et al.,* 2000
	F1 OB	Stem diameter Juvenile wood density	2-8 QTLs / 1-4% (phenotypic)	Devey *et al.,* 2004

* OB = outbred, S = selfing.

mapping than standard QTL approaches. Association mapping is also useful to select nucleotide polymorphisms for population genetic studies aiming to reveal adaptive divergence in the wild (see Box 1). Conifers, and in particular pines, are well-suited for association studies due to (1) high levels of genetic variation, (2) low level of domestication, (3) large natural populations and outcrossing mating systems, resulting in a low level of population structure, (4) high population recombination rates and rapid decay of intragenic LD and (5) accurate evaluation of phenotypes due to the possibility of establishing clonal replicates or large progenies. Pines are also a widespread group with a major ecological role in terrestrial ecosystems (Richardson, 1998), which highlights the interest of studying adaptive variation patterns in this group of species.

Current statistical methods used in association mapping include comparison of phenotype means across groups showing different marker genotypes and case-control studies, where allele frequencies are compared across different phenotypes. Application of mixed models to identify SNPs associated with differences in phenotypic means for disease traits in pine has been relatively successful (Dudley Huber, personal communication). Case-control studies were initially developed in the field of human genetics, based on the comparison of disease-affected (*case*) and healthy (*control*) individuals. Population structure is the most common systematic bias producing false-positive associations in association studies (Marchini *et al.,* 2004; Hirschhorn and Daly, 2005). As a consequence, methods that account for population genetic structure have been developed. One of such methods, the *Transmision/Disequilibrium Test* (TDT) of Spielman *et al.* (1993) is probably the most popular. The TDT is based on the transmission of alleles from heterozygous parents to affected offspring and can be readily extended to quantitative traits and multiple markers (see Jannink and Walsh, 2002 for details). The power of association mapping is related to (1) the magnitude of the LD between the marker allele and the loci underlying the quantitative trait, (2) the marker allele frequency, power decreasing substantially with lower minor allele frequencies, (3) the effect size of the causative allele itself and (4) the mode of gene action.

Box 1. Population genetics methods to detect adaptive divergence in nature.

The detection of outlier loci as a means of distinguishing between single-locus and genome-wide effects constitutes a powerful tool to detect adaptive divergence in nature. The tests most widely used are based on the detection of outlier loci for multiple-population genetic differentiation estimates. A first simple method consists on the comparison of differentiation estimates, such as F_{st}, for putatively neutral molecular markers (usually nuclear SSRs) and candidate adaptive ones provided, for instance, by association studies. Then, those markers showing higher (or lower) differentiation than the putatively neutral ones might be consider as being under diversifying (or stabilizing) selection. A more sophisticated approach, which removes any hypothesis about neutrality of a certain kind of marker loci, consists on the use of the coalescent theory to build, by means of simulation, a neutral expectation. Two competing approaches of the later approximation are becoming rapidly widespread. Firstly, Beaumont and Nichols (1996) developed a method based on the analytical framework of Lewotin and Krakauer (1973). This method, given the global genetic differentiation found in a sample, constructs a theoretical neutral expectation of F_{st} for each value of expected heterozygosity (H_e). Studies based on simulated populations have shown an acceptable rate of identification of loci under positive selection but also showed this method generalized failure to detect loci under balancing selection (Beaumont and Balding 2004). A competing method was developed by Vitalis *et al*. (2001) based on estimates of shared ancestry among populations (F). This method computes estimates for pairs of populations, which is advantageous to detect selection acting at a local scale. The method developed by Vitalis *et al*. (2001) might be more adequate than Beaumont and Nichols' for its application to forest trees. A major advantage is that the former allows for historical changes in effective population size, for instance those resulting from range expansion/reduction. This approach is also robust when moderate gene flow among populations is present, a common feature in forest tree populations.

Sample sizes of about 500 individuals are required, in most cases, to have sufficient power to detect causative polymorphisms (Long and Langley, 1999). Up to date, major association mapping projects in pine are only being developed for *Pinus taeda* (see http://dendrome.ucdavis.edu/adept/), and *P. pinaster* and *P. sylvestris* (http://cc.oulu.fi/~genetwww/treesnips/), but these studies will probably be extended to other major pine crops, such as *Pinus radiata*, in the near-future (see http://www.scionresearch.com/).

Functional Genomics

Expressed Sequence Tags (ESTs)

Partial sequences of cDNA selected from libraries are known as *Expressed Sequence Tags* (ESTs). According to the specificity of such libraries, high throughput random sequencing of ESTs can become a useful strategy for the identification of genes with potential roles in the tissue, organ and/or process of interest. This approach is feasible in pines as far as properly annotated sequence databases in model species and adequate search tools are available, thus allowing the annotation of pine ESTs. EST libraries have been developed for forest trees since the late 90's.

When libraries are constructed using randomly selected cDNAs, ESTs from highly expressed genes are more frequently obtained than from genes that are poorly transcribed in the source material. ESTs included just once in a database are known as *singletons*. The term *redundancy* refers to the level of repeated genes in an EST database, and can be expressed as the ratio of total to singleton ESTs.

Most efforts devoted to EST analysis in forest trees have been performed in economically important genus, such as *Eucalyptus, Populus* and *Pinus*. Efforts have been mainly focused in xylem development (e.g., Allona *et al.,* 1998; Kinlaw *et al.,* 1996; Sterky *et al.,* 1998) and, more recently, biotic and abiotic stress-response traits (see, for instance, http://fungen.org/Projects/Pine/NSF%20abstract.htm). For obvious commercial reasons, information about EST-sequencing projects on certain species is out of public access. Up to date, the two biggest forest EST collections are for *Eucalyptus grandis* and *Pinus radiata*. EST sequencing projects for these two species began in 1995, and currently account for more than 500,000 ESTs. A detailed description of these projects can be found in Strabala (2004). Relevant EST collections for traits related to host-pathogen interactions and environmental stress responses have also been produced (e.g., drought stress in *Pinus pinaster* Ait. and *Pinus taeda* L., Dubos and Plomion, 2003; Frigerio *et al.,* 2004; http://fungen.org/Projects/Pine/NSF%20abstract.htm; or *Pinus sylvestris/ Heterobasidion annosum* infection, Karlsson, 2005).

The Institute for Genomic Research, TIGR (Rockville, Maryland, USA; http://www.tigr.org), publishes *Gene Indexes* that integrate research data from international EST-sequencing and gene-research projects for several taxa, including animals and plants. The goal of these public-access indexes is to provide a non-redundant view of available genes and data on their expression patterns, functions and evolutionary relationships. Up to date, this task is being performed for three forest genera: *Pinus, Populus* and *Picea*. In the last release of the Pine Gene Index (5.0, October 2004) 191,229 ESTs and 1,381 cDNA sequences obtained by «classical» methods (ETs, Expressed Transcripts) are included. Bioinformatic analyses of these sequences have yielded 137 singleton ETs and 18,250 singleton ESTs, as well as 16,666 tentative consensus (TCs) defined from repeated sequences (a total of 35,053 unique sequences, versus 54,756 in the Poplar Gene index and 27,194 in the Spruce Gene Index). The Pine Gene Index web interface provides user-friendly specific search tools (based on sequence similarity, original EST library, annotation, expression, and many more). The available number of *Pinus* ESTs is steadily increasing, reaching the total of 276,366 in GenBank public database (National Center for Biotechnology Information, NCBI, USA; http://www.ncbi.nlm.nih.gov) in June, 2005 (Table 3).

Table 3. Pine ESTs available at GenBank DataBase

Species	No. of ESTs
Pinus taeda	257,822
Pinus pinaster	18,254
Pinus sylvestris/ /Heterobasidion annosum	1,663
Pinus sylvestris	2
Pinus elliottii	150
Pinus radiata	69
Pinus banksiana	46
Pinus patula	23

The development of ESTs for maritime pine (*Pinus pinaster* Ait.) was initiated at INRA (Pierroton, France), in collaboration with other institutions, within the «LIGNOME» program, in 2001. This program aims to study the genome and proteome of ligneous species, with a focus on environmental adaptation and wood quality. Concerning maritime pine, specific cDNA libraries have been constructed for the identification of genes involved in wood formation and response to drought stress (Cantón *et al.,* 2004; Frigerio *et al.,* 2004). To date, nearly 20,000 ESTs obtained from these libraries have been included in public databases. Recently, in collaboration with Universidad Politécnica de Madrid and INIA (Spain), a cDNA library from shoots has yielded 9,000 additional ESTs (ongoing work). A bioinformatics suite is being implemented in collaboration with the University of Bordeaux (http://cbi.labri.fr) for the analysis and annotation of maritime pine ESTs.

EST sequencing is considered the most cost-efficient strategy of gene discovery for organisms whose genome sequence is not yet available, as it is the case for conifer species. ESTs also constitute a valuable resource for the development of molecular markers. For instance, in maritime pine SNPs and SSRs have been developed from ESTs databases (Chagné *et al.,* 2004b; Le Dantec *et al.,* 2004).

Transcriptional profiling

This strategy combines the principles of differential display and expression arrays. Differential display allows detection of genes that are up- or down-regulated, for instance during a physiological process or after a treatment, by comparing their expression levels in different samples. On its side DNAs from the genes under study are fixed to a solid support in *expression arrays*; the array is then hybridized with RNA-derived labelled probes isolated from the samples of interest. Analysis of the signal emitted from the array allows the identification of the genes present in the array that are expressed in the sample and even to estimate their expression levels (Figure 1.A).

Hybridization of an array with samples obtained during the course of a physiological process, for example, wood formation, provides information about the temporal pattern of gene transcription for genes included in the array (i.e., their *transcriptional profile*; Figure 1.B). The key point for this technique is the appropriate choice of candidate and control genes to be included in the array. Glass-slide microarrays and photolitographic oligo arrays enable testing a very high number of genes, which is useful for the analysis of several biological processes.

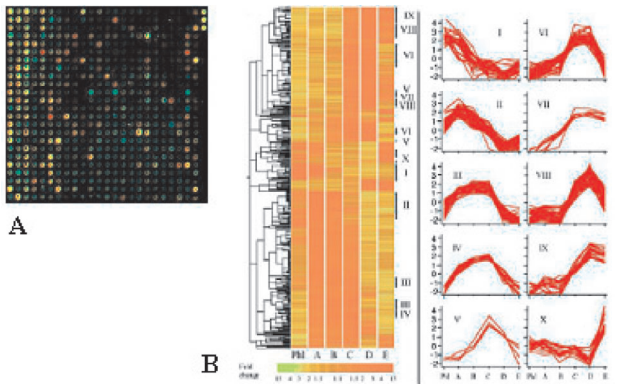

Figure 1.A. Detail of a hybridized microarray. **B:** Transcriptional profiling of different genes included in a microarray (Hertzberg *et al.,* 2001). Gene clustering (left) based on their expression profiles (right).

As stated before regarding ESTs from cDNA libraries, xylem formation is one of the processes to which most attention is paid in forest trees, and transcriptional profiling is being widely used in its analysis. In maritime pine, array analyses are being applied to identify genes involved in the formation of compression wood, or related to the transition between early and late wood, or juvenile and mature wood (Paiva *et al.,* 2005). This strategy has also been applied to investigate other physiological processes such as frost tolerance in Scots pine (Balk *et al.,* 2005) and drought-stress in *Pinus taeda* (Watkinson *et al.,* 2003).

High levels of sequence conservation among taxa allow the use of arrays from one taxon for studies on gene expression in another. Transferability of arrays among *Picea* and *Pinus* (and among species within this genus) has been documented (van Zyl *et al.,* 2002; Stasolla *et al.,* 2003).

Conclusions and future perspectives

In general, forest tree capacity to develop adaptive responses to changing environmental conditions is based on how forests have behaved under past climatic conditions. An integrated approach combining

expertise in different disciplines such as genomics, population and quantitative genetics, evolution, ecophysiology, pathology and forest management is required to acquire a comprehensive knowledge of forest tree adaptation mechanisms.

Recent development of genomic tool kits, including EST sequencing, microarrays, BAC libraries, high-density genetic maps, QTL and association mapping as well as genetic transformation techniques will allow to decode complex pine genomes for a better understanding of the genetic basis of adaptation. Conifer genomes, such as those of pine species, are so large that they are refractive to complete genome sequencing. However, comparative genomics coupled with appropriate sampling strategies gives us the ability to understand the structure and function of these genomes. Combination of methylation and high-C0t filtration techniques is an excellent sampling strategy for sequencing other large and complex plant genomes. Exploration of gene-rich regions, without having to dilute sequencing efforts in highly-repetitive DNA around them, provides a preview of genome structure, and may turn out to be a rapid and cost-effective alternative to whole genome approach. The sequences resulting from this approach will be fragmented. However, comparing them to sequences of model species will allow ordering a large percentage of such sequences. It is important to point out that from a functional point of view, analysis of gene families in conifer species has revealed that evolution and resulting family structure differ from that of angiosperms. Therefore, assignment of gene function cannot only be based on sequence similarities and has to be validated by candidate gene and reverse genetic approaches.

Acknowledgements

This review was carried out in the frame of the Spanish research project RTA03-213, from the Programa Nacional de Recursos y Tecnologías Agroalimentarias (Ministerio de Educación y Ciencia, MEC), and the grants from the European Union (TREESNIPS project: QLK3-CT-2002-01973) and Trilateral Co-Operation in Plant Genomics between Spain (MEC), France (GÉNOPLANTE) and Germany (GABI) (DIGENFOR project: Gen2003-20216-C02).

References

ALLONA I., QUINN M., SHOOP E., SWOPE K., ST. CYR S., CARLIS J., RIEDL J., RETZEL E., CAMPBELL M.M., SEDEROFF R., WHETTEN R.W., 1998. Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci USA 95, 993-9698.

ARUMUGANATHAN K., EARLE E.D., 1991. Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9, 208-218.

BAHRMAN N., DAMERVAL C., 1989. Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait.). Heredity 63, 267-274.

BALK P.A., VAN DER GEEST L., BRØNNUM P., PERKS M., STATTIN E., VAN WORDRAGEN M., 2005. Analogous shifts in gene expression accompany frost tolerance development in Scots pine and common beech during autumn. Plant & Animal Genomes XIII Conference San Diego, CA, January 15-19.

BEAUMONT M.A., NICHOLS R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London B 263, 1619-1626.

BEAUMONT M.A., BALDING D.J., 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13, 969-980.

BEAVIS W.D., 1994. The power and deceit of QTL experiments, lessons from comparative QTL studies. Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference, ASTA, Washington DC, pp. 250-266.

BEAVIS W.D., 1998. QTL analyses, Power., precision and accuracy. In: Molecular analysis of complex traits (Paterson A.H, ed). CRC Press, Boca Raton, FL, pp. 145-161.

BRENDEL O., POT D., PLOMION C., ROZENBERG P., GUEHL J.M., 2002. Genetic parameters and QTL analysis of delta C-13 and ring width in maritime pine. Plant Cell Environ 25, 945-953.

BROWN G.R., BASSONI D.L., GILL G.P., FONTANA J.R., WHEELER N.C., MEGRAW R.A., DAVIS M.F., SEWELL M.M., TUSKAN G.A., NEALE D.B., 2003. Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.) III. QTL verification and candidate gene mapping. Genetics 164, 1537-1546.

BROWN G.R., KADEL E.E III., BASSONI D.L., KIEHNE K.L., TEMESGEN B., VAN BUIJTENEN J.P., SEWELL M.M., MARSHALL K.A., NEALE D.B., 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159, 799-809.

CANTÓN F.R., LE PROVOST G., GARCIA V., BARRÉ A., FRIGÉRIO J.M., PAIVA J., FEVEREIRO P., ÁVILA C., MOURET J. F., DE DARUVAR A., CÁNOVAS F.M., PLOMION C., 2004. Transcriptome analysis of wood formation in maritime pine. In: Sustainable Forestry, Wood products & Biotechnology, BIOFOR Proceeding, Vitoria, Spain, pp. 334-347.

CATO S.A., GARDNER R.C., KENT J., WARR A.A., ECHT C.S., RICHARDSON T.E., 2001. A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102, 296-306.

CHAGNÉ D., LALANNE C., MADUR D., KUMAR S., FRIGERIO J.M., KRIER C., DECROOCQ S., SAVOURÉ A., BOU-DAGHER-KHARRAT M., BERTOCCHI E., BRACH J., PLOMION C., 2002. A high density genetic map of Maritime pine based on AFLPs. Ann For Sci 59, 627-636.

CHAGNÉ D., BROWN G., LALANNE C., MADUR D., POT. D., NEALE D., PLOMION C., 2003. Comparative genome and QTL mapping between maritime and loblolly pines. Mol Breeding 12, 185-195.

CHAGNÉ D., CHAUMEIL P., RAMBOER A., COLLADA C., GUEVARA A., CERVERA M.T., VENDRAMIN G.G., GARCIA V., FRIGERIO J.M., ECHT C., RICHARDSON T., PLOMION C., 2004a. Cross species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109, 1204-1214.

CHAGNÉ D., LE DANTEC L., RAMBOER A.., FRIGERIO J.M., POT D., COLLADA C., CERVERA M.T., VENDRAMIN G.G., PLOMION C., 2004b. Development of molecular marker in *Pinus pinaster* based on pine ESTs. Plant & Animal Genomes XII Conference San Diego, CA, January 10-14.

CERVERA M.T., PLOMION C., MALPICA C., 2000. Molecular markers and genome mapping in woody plants. In: Molecular Biology of Woody Plants (Jain S.M., and Minocha S.C., eds) Kluwer Academic Publishers, Dordrecht, Holland, pp. 375-394.

COSTA P., POT D., DUBOS C., FRIGERIO J.M., PIONNEAU C., BODÉNÈS C., BERTOCCHI E., CERVERA M., REMINGTON DL., PLOMION C., 2000. A genetic map of Maritime pine based on AFLP., RAPD and protein markers. Theor Appl Genet 100, 39-48.

DALE G., TEASDALE B., 1995. Analysis of growth., form and branching traits in an F2 population of the *Pinus elliottii x Pinus caribaea* interspecific hybrid using RAPD markers. Proceedings of the 25th Southern Forest Tree Improvement Conference, pp. 242-253.

DEVEY M.E., BELL J.C., SMITH D.N., NEALE D.B., MORAN G.F., 1996. A genetic map for *Pinus radiata* based on RFLP, RAPD and microsatellite markers. Theor Appl Genet 92, 673-679.

DEVEY M.E., CARSON S.D., NOLAN M.F., MATHESON A.C., RIINI C.T., HOHEPA J., 2004. QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection. Theor Appl Genet 108, 516-524.

DEVEY M.E., FIDDLER T.A., LIU B.H., KNAPP S.J., NEALE B.D., 1994. An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. Theor Appl Genet 88, 273-278.

DEVEY M.E., SEWELL M.M., UREN T.L., NEALE D.B., 1999. Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. Theor Appl Genet 99, 656-662.

DOERGE R.W., 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3, 43-52.

DOUDRICK R.L., HESLOP-HARRISON J.S., NELSON C.D., SCHMIDT T., NANCE W.L., SCHWARZACHER T., 1995. Karyotype of slash pine (*Pinus elliottii* var. *elliottii*) using patterns of fluorescence in situ hybridization and fluorochrome banding. J Hered 86, 289-296.

ECHT C.S., NELSON C.D., 1997. Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). Theor Appl Genet 94, 1031-1037.

EMEBIRI L.C., DEVEY M.C., MATHESON A.C., SLEE M.U., 1998. Age-related changes in the expression of QTLs for growth in radiata pine seedlings. Theor Appl Genet 97, 1053-1061.

FISHER P.J., GARDNER R.C., RICHARDSON T.E., 1996. Single locus microsatellites isolated using 5' anchored PCR. Nuc Acid Res 24, 4369-4371.

FISHER P.J., RICHARDSON T.E., GARDNER R.C., 1998. Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. Theor Appl Genet 96, 969-979.

FRIESEN N., BRANDES A., HESLOP-HARRISON J.S., 2001. Diversity, origin and distribution of retrotransposons (gypsy and copia) in conifers. Mol Biol Evol 18, 1176-1188.

FRIGERIO J.M., DUBOS C., CHAUMEIL P., SALIN F., GARCIA V., BARRÉ A., PLOMION C., 2004. Using transcriptome analysis to identify osmotic stress candidate genes in maritime pine (*Pinus pinaster* Ait.). In: Sustainable Forestry., Wood products & Biotechnology, BIOFOR Proceeding, Vitoria, Spain, pp. 348-362.

GERBER S., RODOLPHE F., BAHRMAN N., BARADAT P., 1993. Seed-protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis, genetic determinism and construction of a linkage map. Theor Appl Genet 85, 521-528.

GÓMEZ A., ARAVANOPOULOS F.A., BUENO M.A., ALIA R., 2002. Linkage of random amplified polymorphic DNA markers in *Pinus halepensis* Mill. Silvae Genet 51, 196-201.

GRATTAPAGLIA D., AND SEDEROFF R., 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross, mapping strategy and RAPD markers. Genetics 137, 1121-1137.

GROTKOPP E., REJMANEK M., SANDERSON M.J., ROST T.L., 2004. Evolution of genome size in pines *(Pinus),* and its life-history correlates, supertree analysis. Evolution 58, 1705-1729.

GROOVER A., DEVEY M., FIDDLER T., LEE J., MEGRAW R., MITCHEL-OLDS T., SHERMAN B., VUJCIC S., WILLIAMS C., NEALE D., 1994. Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138, 1293-1300.

HARRY D.E., TEMESGEN B., NEALE D., 1998. Codominant PCR-based markers of *Pinus taeda*

developed from mapped cDNA clones. Theor Appl Genet 97, 327-336.

HAYASHI E., KONDO T., TERADA K., KURAMOTO N., GOTO Y., OKAMURA M., KAWASAKI H., 2001. Linkage map of Japanese black pine based on AFLP and RAPD markers including markers linked to resistance against the pine needle gall midge. Theor Appl Genet 102, 871-875.

HAYASHI E., KONDO T., TERADA K., KURAMOTO N., GOTO Y., OKAMURA M., KAWASAKI H., 2002. Linkage map based on AFLP and RAPD markers in Japanese black pine. Bulletin of the National Forest Tree Breeding Center, 18, 49-57.

HERTZBERG M., ASPEBORG H., SCHRADER J., ANDERSSON A., ERLANDSSON R., BLOMQVIST K., BHALERAO R., UHLEN M., TEERI T.T., LUNDEBERG J., SUNDBERG B., NILSSON P., SANDBERG G., 2001. A transcriptional roadmap to wood formation. Proc Natl Acad Sci USA 98, 14732-14737.

HIRSCHHORN J.N., DALY M.J., 2005. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6, 95-108.

HIZUME M., SHIBATA F., MATSUSAKI Y., GARAJOVA Z., 2002. Chromosome identification and comparative karyotypic analyses of four *Pinus* species. Theor Appl Genet 105, 491-497.

HURME P., SAVOLAINEN O., 1999. Comparison of homology and linkage of random amplified polymorphic DNA (RAPD) markers between individual trees of Scots pine (*Pinus sylvetris* L.). Mol Ecol 8, 15-22.

HURME P., SILLANPAA M.J., ARJAS E., REPO T., SAVOLAINEN O., 2000. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. Genetics 156, 1309-1322.

JACOBS M.D., GARDNER R.C., MURRAY B.G., 2000. Cytological characterization of heterochromatin and rDNA in *Pinus radiata* and *P. taeda*. Plant Syst Evol 223, 71-79.

JANNINK J.L., WALSH J.B., 2002. Association mapping in plant populations. In: Quantitative Genetics, Genomics and Plant Breeding (Kang M.S., ed). CABI, Wallingford, UK, pp. 59-68.

KAO C.H., ZENG Z.B., TEASDALE R.D., 1999. Multiple interval mapping for quantitative trait loci. Genetics 152, 1203-1216

KARP A., EDWARDS K.J., 1998. DNA markers, A global overview. In: DNA Markers: Protocols, Applications and Overviews (Caetano-Anollés G., Gresshoff P.M., eds). John Wiley & Sons, New York, NY, pp.1-13.

KAYA Z., NEALE D.B., 1995. Utility of random amplified polymorphic DNA (RAPD) markers for linkage mapping in Turkish red pine (*Pinus brutia* Ten.). Silvae Genet 44, 110-116.

KINLAW C.S., HO T., GERTTULA S.M., GLADSTONE E., HARRY D.E., QUINTANA L., BAYSDORFER C., 1996. Gene discovery in loblolly pine through cDNA sequencing. In: Somatic cell genetics and molecular genetics of trees (Ahuja M.R., Boerjan W., Neale D.B., eds.). Kluwer Academic Publishers. Dordrecht., The Netherlands.

KINLAW C.S., NEALE D.B., 1997. Complex gene families in pine genomes. Trends in Plant Sci 2, 356-359.

KOMULAINEN P., BROWN G.R., MIKKONEN M., KARHU A., GARCIA-GIL M.R., O'MALLEY D., LEE B., NEALE D.B., SAVOLAINEN O., 2003. Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. Theor Appl Genet 107, 667-678.

KRUTOVSKY K., NEALE D.B., 2001. Forest genomics for conserving adaptive genetic diversity. Forest Genetic Resources Working Papers 3, FAO, Rome, available on-line at http//www.fao.org/DOCREP/003/X6884E/x6884e00.htm.

KUANG H. RICHARDSON T.E., CARSON S., WILCOX P.L., BONGARTEN B., 1999. Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. I. Genetic map with distorted markers. Theor Appl Genet 98, 697-703.

KUBISIAK T.L., NELSON CD., NANCE W.L., STINE M., 1995. RAPD linkage mapping in a longleaf pine x slash pine F1 family Theor Appl Genet 90, 1110-1127.

KUBISIAK T.L., NELSON C.D., NANCE W.L., STINE M., 1996. Comparison of RAPD linkage maps constructed for a single longleaf pine from both haploid and diploid mapping populations. For Genet 3, 203-211.

KUMAR S., SPELMAN R.J., GARRICK D.J., RICHARDSON T.E., LAUSBERG M., WILCOX P.L., 2000. Multiple marker mapping of wood density loci in an outbred pedigree of radiata pine. Theor Appl Genet 100, 926-933.

KRUTOVSKY K.V., TROGGIO M., BROWN G.R., JERMSTAD K.D., NEALE D.B., 2004. Comparative mapping in the Pinaceae. Genetics 168, 447-461.

LANDER E.S., BOTSTEIN D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185-199.

LE DANTEC L., CHAGNE D., POT D., CANTIN O., GARNIER-GERE P., BEDON F., FRIGERIO J.M., CHAUMEIL P., LEGER P., GARCIA V., LAIGRET F., DE DARUVAR A., PLOMION C., 2004. Automated SNP detection in expressed sequence tags, statistical considerations and application to maritime pine sequences Plant Mol Biol 54, 461-470.

LERCETEAU E., PLOMION C., ANDERSSON B., 2000. AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*, a preliminary study. Mol Breeding 6, 451-458.

LEWOTIN R.C., KRAKAUER J.K., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175-195.

LI C., YEH F.C., 2001. Construction of a framework map in *Pinus contorta* subsp. *latifolia* using random amplified polymorphic DNA markers. Genome 44, 147-153.

LONG A.D., LANGLEY C.H., 1999. The power of association studies to detect the contribution of candidate

genetic loci to variation in complex traits. Genome Res 9, 720-731.

LYNCH M., WALSH B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

KARLSSON M., 2005. Transcriptional responses during the pathogenic interaction between *Heterobasidion annosum* s.l. and Conifers. PhD Thesis, Swedish University of Agricultural Sciences, Uppsala, Swedish.

MARCHINI J., CARDON L.R., PHILLIPS M.S., DONNELLY P., 2004. The effects of human population structure on large genetic association studies. Nat Genet 36, 512-517.

MARIETTE S., CHAGNE D., DECROOCQ S., VENDRAMIN G.G., LALANNE C., MADUR C., PLOMION C., 2001. Microsatellite markers for *Pinus pinaster* Ait. Ann For Sci 58, 203-206.

MARKUSSEN T., FLADUNG M., ACHERE V., FAVRE J.M., FAIVRE-RAMPANT P., ARAGONES A., DA SILVA PÉREZ D., HAVENGT L., RITTER E., 2003. Identification of QTLs controlling growth, chemical and physical wood property traits in *Pinus pinaster*, Ait. Silvae Genet 52, 8-15.

MAURICIO R., 2001. Mapping quantitative traits loci in plants, uses and caveats for evolutionary biology. Nat Rev Genet 2, 370-381.

MORGENSTERN E.K., 1996. Geographic Variation in Forest Trees, Genetic basis and application of knowledge in silviculture. UBC Press, University of British Columbia, Vancouver, BC, Canada.

NEALE D.B., SAVOLAINEN O., 2004. Association genetics of complex traits in conifers. Trends Plant Sci 9, 325-330.

NELSON C.D., KUBISIAK T.L., STINE M., NANCE WL., 1994. A genetic linkage map of longleaf pine (*Pinus palustris* Mill) based on random amplified polymorphic DNAs. J Hered 85, 433-439.

NELSON CD., NANCE W.L., DOUDRICK R.L., 1993. A partial genetic linkage map of Slash pine (*Pinus elliotti* Englem var elliottii) based on random amplified polymorphic DNAs. Theor Appl Genet 87, 145-151.

PAIVA J.P., GION J.M., EVENO E., GARNIER-GÉRE P., POT D., CANTON F., LE PROVOST G., BRACH J., CANOVAS F., FEVEREIRO P., PLOMIONC., 2005. Expression profiling of wood forming tissues in maritime pine. Plant & Animal Genomes XIII Conference San Diego, CA, January 15-19.

PLOMION C., BAHRMAN N., DUREL C.E., O'MALLEY D.M., 1995a. Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. Heredity 74, 661-668.

PLOMION C., O'MALLEY D.M., DUREL C.E., 1995b. Genomic analysis in maritime pine (*Pinus pinaster*) Comparison of two RAPD maps using selfed and open-pollinated seeds of the same individual. Theor Appl Genet 90, 1028-1034.

PLOMION C., O'MALLEY D.M., 1996. Recombination rate differences for pollen parents and seed parents in pine. Heredity 77, 341-350.

PLOMION C., DUREL C.E., O'MALLEY D.M., 1996. Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. Theor Appl Genet 93, 849-858.

PLOMION C., COSTA P., BAHRMAN N., 1997. Genetic analysis of needle protein in Maritime pine 1 Mapping dominant and codominant protein markers assayed on diploid tissue, in a haploid-based genetic map. Silvae Genet 46, 161-165.

PLOMION C., HURME P., FRIGERIO J.M., RIDOLPHI M., POT D., PIONNEAU C., AVILA C., GALLARDO F., DAVID H., NEUTLINGS G., CAMPBELL M., CANOVAS F.M., SAVOLAINEN O., BODÉNÈS C., KREMER A., 1999. Developing SSCP markers in two *Pinus* species. Mol Bred 5, 21-31.

POT D., 2004. Déterminisme génétique de la qualité du bois chez le pin maritime, du phénotype aux gènes. PhD Thesis, ENSA de Rennes, France.

REMINGTON D.L., O'MALLEY D.M., 2000. Whole genome characterization of embryonic state inbreeding depression in a selfed loblolly pine family. Genetics 155, 337-348.

REMINGTON D.L., WHETTEN R.W., LIU B.H., O'MALLEY D.M., 1999. Construction of an AFLP genetic map with nearly complete coverage in *Pinus taeda*. Theor Appl Genet 98, 1279-1292.

RITTER E., ARAGONES A., MARKUSSEN T., ACHERE V., ESPINEL S., FLADUNG M., WROBEL S., FAIVRE-RAMPANT P., JEANDROZ S., FAVRE J. M., 2002. Towards construction of an ultra high density linkage map for *Pinus pinaster*. Ann For Sci 59, 637-643.

SAYLOR L.C., 1961. A karyotypic analysis of selected species of *Pinus.* Silvae Genet 10, 77-84.

SCHMIDT A., DOUDRICK R.L., HESLOP-HARRISON J.S., SCHMIDT T., 2000. The contribution of short repeats of low sequence complexity to large conifer genomes. Theor Appl Genet 101, 7-14.

SEDELNIKOVA T.S., MURATOVA E.N., 2002. Specific Karyological Features of Siberian Stone Pine (*Pinus sibirica* Du Tour) in Western Siberian Bogs. Russian J Ecol 33, 303-308.

SEWELL M.M., SHERMAN B.K., NEALE D.B., 1999. A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. Genetics 151, 321-330.

SEWELL M.M., BASSONI D.L., MEGRAW R.A., WHEELER N.C., NEALE D.B., 2000. Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) I. Physical wood properties. Theor Appl Genet 101, 1273-1281.

SEWELL M.M., DAVIS M.F., TUSKAN G.A., WHEELER N.C., ELAM C.C., BASSONI DL., NEALE D.B., 2002. Identificaiton of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. Theor Appl Genet 104, 214-222.

SHEPHERD M., CROSS M., DIETERS M.J., HENRY R., 2003. Genetic maps for *Pinus elliottii* var. elliottii and *P.*

caribaea var. hondurensis using AFLP and microsatellite markers. Theor Appl Genet 106, 1409-1419.

SHIBATA F., MATSUSAKI Y., HIZUME M., 2005. AT-rich sequences containing *Arabidopsis*-type telomere sequence and their chromosomal distribution in *Pinus densifolia*. Theor Appl Genet 110, 1253-1258.

SPIELMAN R.S., MCGINNIS R.E., EWENS W.J., 1993. Transmission test for linkage disequilibrium, the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52, 506-516.

STASOLLA C., VAN ZYL L., EGERTSDOTTER U., CRAIG D., LIU W., SEDEROFF R.R., 2003. The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. Plant Physiol 131, 49-60.

STERKY F., REGAN S., KARLSSON J., HERTZBERG M., ROHDE A., HOLMBERG A., AMINI B., BHALERAO R., LARSSON M., VILLARROEL R., VAN MONTAGU M., SANDBERG G., OLSSON O., TEERI T.T., BOERJAN W., GUSTAFSSON P., UHLÉN M., SUNDBERG B., LUNDEBERG J., 1998. Gene discovery in the wood-forming tissues of poplar, Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci USA 95, 13330-13335.

STRABALA T.J., 2004. Expressed sequence tag databases from forestry tree species. In: Molecular genetics and breeding of forest trees Food Products (Kumar S., Fladung M., eds). Press, NY, USA.

TEMESGUEN B., BROWN G., HARRY DE., KINLAW CS., SEWELL M.M., NEALE D.B., 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 102, 664-675.

TRAVIS S.E., RITLAND K., WHITMAN T.G., KEIM P., 1998. A genetic linkage map of Pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. Theor Appl Genet 97, 871-880.

VAN ZYL L., VON ARNOLD S., BOZHKOV P., CHEN Y., EGERTSDOTTER U., MACKAY J., SEDEROFF R., SHEN J., ZELENA L., CLAPHAM D., 2002. Heterologous array analysis in Pinaceae, hybridization of *Pinus taeda* cDNA arrays with cDNA from needles and embryogenic cultures of *P. taeda*, *P. sylvestris* or *Picea abies*. Comp Funct Genom 3, 306-318.

VITALIS R., DAWSON K., BOURSOT P., 2001. Interpretation of variation across marker loci as evidence of selection. Genetics 158, 1811-1823.

WATKINSON J.I., SIOSON A.A., VASQUEZ-ROBINET C., SHUKLA M., ELLIS M., HEATH L.S., RAMAKRISHNAN N., CHEVONE B., WATSON L.T., VAN ZYL L., EGERTSDOTTER U., SEDEROFF R.R., GRENE R., 2003. Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. Plant Physiol. 133, 1702-1716.

WENG C., KUBISIAK T.L., NELSON C.D., STINE M., 2002. Mapping quantitative trait loci controlling early growth in a (longleaf pine x slash pine) x slash pine BC1 family. Theor Appl Genet 104, 852-859.

WILCOX P.L., RICHARDSON T.E., CORBETT G.E., BALL R.D., LEE J.R., DJOROVIC A., CARSON S.D., 2001. Framework linkage maps of *Pinus radiata* D. Don based on pseudotestcross markers. For Genet 8, 109-117.

YAZDANI R., NILSSON J.E., PLOMION C., MATHUR G., 2003. Marker trait association for autumn cold acclimation and growth rhythm in *Pinus sylvestris*. Scand J Forest Res 18, 29-38.

YAZDANI R., YEH F., RIMSHA J., 1995. Genomic mapping of *Pinus sylvestris* L. using random amplified polymorphic DNA markers. For Genet 2, 109-116.

YIN T.M., WANG X.R., ANDERSSON B., LERCETEAU-KOHLER E., 2003. Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. Theor Appl Genet 106, 1075-1083.

ZHOU Y., GWAZE D.P., REYES-VALDES M.H., BUI T., WILLIAMS C.G., 2003. No clustering for linkage map based on low-copy and undermethylated microsatellites. Genome 46, 809-816.