Plomion C, Chagné D, Pot D, Kumar S, Wilcox PL, Burdon RD, Prat D, Peterson DG, Paiva J, Chaumeil P, Vendramin GG, Sebastiani F, Nelson CD, Echt CS, Savolainen O, Kubisiak TL, Cervera MT, de María N, Islam-Faridi MN (2006) The Pines. In: Genome Mapping and Molecular Breeding in Plants, Vol. 7 Forest Trees, Chitta R. Kole (Ed). Springer, Heidelberg, Berlin, New York, Tokyo (in press)

THE PINES

C. Plomion[1*], D. Chagné[2], D. Pot[3], S. Kumar[4], P.L. Wilcox[5], R.D. Burdon[6], D. Prat[7], D.G. Peterson[8], J. Paiva[1], P. Chaumeil[1], G.G. Vendramin[9], F. Sebastiani[10], C.D. Nelson[11], C. S. Echt[11], O. Savolainen[12], T.L. Kubisiak[11], M. T. Cervera[13], N. de María[13], and **M.N. Islam-Faridi[14]**

[1]INRA, UMR BIOGECO, 69 route d'Arcachon, 33610 Cestas, France
[2]HortResearch, Private Bag 11030, Palmerston North, New Zealand
[3]Coffee Genomic Team, CIRAD, UMR PIA 1096, 34398 Montpellier cedex 5, France
[4]Ensis-Genetics, Ensis, Private Bag 3020, Rotorua, New Zealand
[5]Cellwall Biotechnology Centre, Scion, Private Bag 3020, Rotorua, New Zealand
[6]Ensis Genetics, Ensis, Private Bag 3020, Rotorua, New Zealand
[7]Université Claude Bernard - Lyon 1, EA 3731 Génome et Evolution des Plantes Supérieures, Bât F.A. Forel, 6 rue Raphaël Dubois, 69622 Villeurbanne Cedex, France
[8]Department of Plant and Soil Sciences, 117 Dorman Hall, Box 9555, Mississippi State University, Mississippi State, MS 39762, USA
[9]Plant Genetics Institute, National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Firenze, Italy
[10]Department of Agricultural Biotechnology, University of Florence, Via della Lastruccia 14, 50019 Sesto Fiorentino, Firenze, Italy
[11]USDA- Forest Service, Southern Research Station, Southern Institute of Forest Genetics, 23332 Mississippi 67, Saucier, MS 39574, USA
[12]Department of Biology, PL 3000, FIN-90014 University of Oulu, Finland
[13]Genética y Ecofisiología Forestal & Unidad Mixta de Genómica y Ecofisiología Forestal INIA-UPM. Departamento de Sistemas y Recursos Forestales, CIFOR - INIA, Carretera de La Coruña km 7, 28040 Madrid, Spain
[14]USDA-Forest Service, Southern Institute of Forest Genetics, Forest Tree Molecular Cytogenetics Laboratory, Texas A&M University, College Station, Texas 77843, USA

*Corresponding author: plomion@pierroton.inra.fr

1. Introduction

1.1. History of the Genus

1.1.1. Origin and Distribution

Pinus is the most important genus within the Family *Pinaceae* and also within the Gymnosperms by the number of species (109 species recognized by Farjon 2001) and its contribution to forest ecosystems. All pine species are evergreen trees or shrubs. They are widely distributed in the northern hemisphere, from tropical areas to northern areas in America and Eurasia. Their natural range reaches the equator only in southeast Asia. In Africa, natural occurrences are confined to the Mediterranean bassin. Pines grow at various elevations from sea level (not usual in tropical areas) to highlands. Two main regions of diversity are recorded, the most important one in Central America (43 species found in Mexico), and a secondary one in China. Some species have a very wide natural range (e.g. *P. ponderosa*, *P. sylvestris*). Pines are adapted to a large range of ecological conditions: from tropical (e.g. *P. merkusii*, *P. kesiya*, *P. tropicalis*), temperate (e.g. *P. pungens*, *P. thunbergii*), subalpine (e.g. *P. albicaulis*, *P. cembra*), to boreal (e.g. *P. pumila*) climates (Richardson and Rundel 1998). They can grow in quite pure stands or in mixed forest with other conifers or broadleaved trees. Some species are especially adapted to forest fires, e.g. *P. banksiana*, in which fire is virtually essential for cone opening and seed dispersal. They can grow in arid conditions, on alluvial plain soils, on sandy soils, on rocky soils, or on marsh soils. Trees of some species can have a very long life as in *P. longaeva* (more than 3000 years).

1.1.2. Botanical Descriptions

The genus is distinguished from other members of the Pinaceae family by its needle-like secondary leaves, borne commonly in fascicles of 1 to 8 on dwarf shoots, with a fascicle sheath of bud scales. The leaves of pines are of four types considering the complete plant development: cotyledons, juvenile leaves, scale leaves (cataphylls), and secondary leaves. Cotyledons vary in number from 4, up to 24 in *P. maximartinezii*. There are up to 3 resin ducts in the cotyledons and either one or two vascular bundles (Farjon 1984). Primary leaves are single, generally helically arranged and acicular, they are produced in most species only during the first growth season, but for a longer period in a few species. Cataphylls, the non-chlorophyllous primary leaves produced on shoots, occur in an extension of the helical arrangement of the primary leaves, and subtend all shoot structures, but they are typically small, and subulate or lanceolate. Secondary leaves, the needles, appear by the end of the first growing season, or later in some species. They are the most common pine leaves, permanently green, metabolically active, ranging in length from 2 to 50 cm (generally smaller ones occur in subalpine or aridity- adapted species). They are borne on dwarf shoots axillary to cataphylls, in fascicles of one (*P. monophylla*, with circular section needles) to eight needles, usual numbers per fascicle being, two, three or five. A fascicle is initially bound together by a basal sheath that may then fall off or persist, but actual leaf fall involves the entire fascicle structure. The number of leaves in the fascicles determines the transverse leaf shape (Farjon and Styles 1997). Stomata are arranged in several longitudinal lines along the entire leaf length. There are generally two or more resin ducts in needles. The number of vascular bundles (one or two as in cotyledons) is the major trait for the identification of the main divisions of genus *Pinus*. The trunk is usually single, erect and columar. The branches are grouped into pseudowhorls (often called clusters), at least when young. Bark patterns in pines result mainly from fissuring due to expansion growth, and the formation of scales that eventually fall of.

Pine species are monoecious. Pollen cones are relatively small and soft; these ephemeral structures consist of an axis with many helically arranged microsporophylls. Two microsporangia are attached to the underside each microsporophyll. These cones open in spring, at least for temperate-climate species, and release large quantities of pollen into the air. Pollen is of the bisaccate type giving it great buoyancy in the air. Pine pollen can be blown over long distances. Its morphology is very similar for all pine species. Seed conelets are found in most species at the ends of new twigs, taking the position of a lateral bud. They are usually located on the higher branches. They consist of scales, the megasporophylls, arranged around an axis; on each megasporophyll lays two separate ovules, each consisting of a cell mass protected by an integument. The micropyle through which the pollen tube penetrates is turned towards the axis. After pollination the seed cone closes its scales by expansion. The seed cone then grows rapidly. Pollen germinates and produces the male gametophyte, with two sperm nuclei. Fertilization takes place later, about one year after pollination in temperate pine species. Seed cone maturation requires one growing season after fertilization for many species, and even a third for some species. In tropical climates the cycle can be shortened because of the lack of winter dormancy. Individual seed cones thus persist for two, up to even three years, on the same tree for most pine species. Seeds contain an embryo embedded in the remaining megagametophyte and the seed coat. The seed wings derive from tissue on the adaxial face of the seed scale. Even in species with vestigial or absent wings, remnants of basal wing tissue are present on the seed scale, on the seed, or on both. Seeds are mostly wind dispersed. In some species, birds are important seed dispersal vectors (*P. albicaulis*). Seed cones are serotinous in some species and open only following exposure to fire (*P. banksiana*).

1.1.3. Systematics and Phylogeny

The genus is divided into subgenera, sections and subsections. Various classifications have been proposed in this genus since Linnaeus. Recent ones obtain support from DNA phylogenetics to identify related species. Many phylogenetic studies have been carried out in pines. Some of the first studies involved restriction patterns of the chloroplast genome (Strauss and Doerksen 1990; Govindaraju et al. 1992; Krupkin et al. 1996). More recent classifications, including a large number of species, were established from nuclear sequences (ITS, Liston et al. 1999) and chloroplast sequences (*rbcL*: Gernandt et al. 2005); the chloroplast genome being paternally inherited in the genus *Pinus* (Neale and Sederoff 1989). Some studies have also been focused on subsets of the genus *Pinus*: subgenus *Pinus* (Geada López et al. 2002), section *Parrya* (Gernandt et al. 2003), and Eurasian species (Wang et al. 1999). Comprehensive classifications of the genus *Pinus* were earlier established by Gaussen (1960) and Van der Burgh (1973) using morphological and anatomical traits. Later, Price et al. (1998) and Gernandt et al. (2005) also included molecular data and identified monophyletic subgenera, sections and subsections. Some features are consistent but variations are noticed between the classical and molecular approaches. The main division into two subgenera according to the number (one or two) of leaf vascular bundles has been recognized by these authors with various subgenera names (*Haploxylon* and *Diploxylon,* called sometimes respectively soft pines and hard pines; and more recently as *Strobus* and *Pinus* named from type species as recommended by botanical nomenclature code; http://tolweb.org/tree?group=Pinus&contgroup=Pinaceae). The taxonomic position of a singular species with flat secondary leaves *P. krempfii,* is not fully agreed. It has been considered as a third monospecific subgenus by Gaussen (1960) while molecular data place it as a member of the subgenus *Strobus* (Wang et al. 2000; Gernandt et al. 2005). Most species belong to the subgenus *Pinus*. Subgenus *Pinus* species are characterized by thick seed-cone scales and persistent fascicle sheaths. The numerous sections proposed by Van der Burgh

(1973) for this subgenus have then been grouped into two sections *Pinus* and *Trifoliae* (Gernandt et al. 2005), the latter being called New World diploxylon pines by Price et al. (1998). The section *Trifoliae* consists of American species distributed into subsections *Australes,* (septal, internal or medial needle resin ducts), *Ponderosae* (internal or medial needle resin ducts), and *Contortae* (medial needle resin ducts): most of these species are characterized by three-needled fascicles (Table 1).

Each of these subsections groups two or more previously described subsections. *P. leiophylla* and *P. lumholtzii* are now clustered within subsection *Australes* and are not further differentiated. The section *Pinus* is divided into subsections *Pinus* and *Pinaster*. Species within this section, with few exceptions, grow in Eurasia and northern Africa. Subsection *Pinaster*, characterized by the lack of a spine on the umbo of the cone scale, includes *P. pinaster* as the type. More recently all of the other species included in the section *Pinaster* as defined by Van den Burgh (1973) were found to cluster within the *Australes* subsection of the genus (Gernandt et al. 2005). Species of the subsection *Pinus* including the type species of the genus, *P. sylvestris* were previously grouped into a section called *Sylvestres* by Van der Burgh (1973). The subgenus *Strobus* has been divided in to two sections: *Quinquefoliae* and *Parrya*. They differ from the sections *Strobus* and *Parrya* of Van der Burgh (1973) and Price et al. (1998) by the transfer of the subsections *Krempfianae* and *Geradianae* from the section *Parrya* into the section *Quinquefoliae*, which also includes the subsection *Strobus*. The subsection *Strobus* consists of species with five-needled fascicles, thin cone scales, terminal position of spines on seed cone, and several other features absent in the subsections *Krempfianae* and *Geradianae* of South-East Asia already differentiated by Van der Burgh (1973) and Price et al. (1998). The three subsections of the *Quinquefoliae* section share a deciduous fascicle sheath. The section *Parrya* consists of the subsections *Cembroides*, *Nelsoniae* and *Balfourianae*; they share an American distribution, the external position of resin ducts (as in subsections *Krempfianae* and *Geradianae*), and thick cone scales (again as in subsections *Krempfianae* and *Geradianae*). Subsection *Nelsoniae* shows persistent fascicle sheath not found in other species of subgenus *Strobus*. Most monophyletic groups cannot be identified from unique morpho-anatomical traits.

Two of the eleven subsections consist of American species and Eurasian species. Sections *Pinus* and *Quinquefoliae* have an Asiatic origin according to chloroplast data. The subsection *Strobus* lineage would have then evolved in America before coming back to Eurasia. Few dispersal events to eastern North America have probably occurred to explain the presence there of the limited number of species of the subsection *Pinus*. The development and utilization of low-copy-number nuclear genes (Syring et al. 2005) should provide new insights to solve remaining classification problems. Most ancient pine fossils have been dated to the early Cretaceous (Millar 1998). They have been found in China, North America, and Europe (which was very close to eastern North America at that time). They did not further refine the putative geographic origin of genus *Pinus*.

1.1.4. Hybridization

Interspecific hybridization occurs in pines but limited mostly to related species within a subsection. Some species such as *P. engelmannii*, *P. jeffreyi*, and *P. ponderosa* are compatible in a number of different combinations (Liston et al. 1999). Natural hybridizations are often indicated by the introgression of the paternally inherited chloroplast genome. Barriers occur at different stages, from the failure of pollen germination to failure at embryogenesis (Ledig 1998). A few species are postulated to have been derived from interspecific hybridization. They also exhibit the highly conserved chromosome number in pines (2n = 24). This is the situation for *P. densata* that has been shown to combine nuclear polymorphisms of *P. tabuliformis* and *P. yunnanensis* with the chloroplast genome of the latter (Wang et al. 2001)

and probably several other species. *P. densata* exchanged genes with ancestral populations prior to its isolation with local differenciation (Ma et al. 2006). Combinations of parental traits and selection for adaptation to new conditions favored colonisation of new territories by the hybrid species.

1.2. Cytogenetics, DNA Content and Genome Composition

Sax and Sax (1933), Mergen (1958) and Khoshoo (1961) were the earliest to describe the karyotypes of various conifer species. They found that species of the genus *Pinus* were diploid with 24 chromosomes (2n = 2x = 24). The chromosomal complements generally consist of 10 or 11 pairs of large homobrachial (metacentric) chromosomes and one or two pairs of smaller heterobrachial (submetacentric) chromosomes (Saylor 1961, 1964, 1972, 1983). Several attempts have been made to construct chromosome-specific karyotypes for various pine species using traditional cytogenetics techniques, viz., C-banding, Giemsa, and fluorescent banding (Borzan and Papes 1978; MacPherson and Filion 1981; Drewry 1982; Saylor 1983; Hizume et al. 1989, 1990). More recently fluorescent *in situ* hybridization (FISH) has been utilized in several pine species (Doudrick et al. 1995; Lubaretz et al. 1996; Jacobs et al. 2000; Hizume et al. 2002; Liu et al. 2003; Cai et al 2006). Doudrick et al. (1995) developed a FISH-based karyotype for *P. elliottii* var. *elliottii* using 18S-25S and 5S rDNA probes and CMA (chromomycin A_3) and DAPI (4',6-diamidino-2-phenylindole) banding that distinguished all the 12 homologous pairs of chromosomes. They went further to suggest that the presented karyotype might be useful as a standard or reference karyotype for *Pinus*. Lubaretz et al. (1996) used computer-aided chromosome analysis on the basis of chromosome length, chromosome arm length ratio and the positions of rDNA (18S-28S and 5S) and telomere (*Arabidopsis*-type telomere repeat sequence or *A*-type TRS) detected with FISH to discriminate three chromosomes of *P. sylvestris*. Hizume et al. (2002) used four probes [45S rDNA, 5S rDNA, PCSR (CMA-band specific repeat) and *A*-type TRS] in developing FISH-based karyotypes for four different pine species (*P. densiflora*, *P. thunbergii*, *P. sylvestris*, and *P. nigra*). Liu et al. (2003) and Cai et al. (2006) used FISH to establish rDNA positions in several species of the *Pinus* and *Strobus* subgenera, respectively. Current work in *P. taeda* (subgenera *Pinus*, section *Pinus*, subsection *Australes*) (Figure 1) emphasizes an improved chromosome preparation technique (based on Jewell and Islam-Faridi 1994 and Islam-Faridi and Mujeeb-Kazi 1995) and statistical analyses of chromosome arm lengths and FISH signal positions and intensities to develop a reference karyotype and cytogenetic map for use in physical genome mapping in the subsection *Australes* (Islam-Faridi et al. 2003, 2006). Comparison of these results to those obtained in other subsections suggests that a subsection-specific karyotype may be required for more robust physical mapping across the entire genus as a whole.

Another feature of the pine genome is its large physical genome size expressed in DNA content. Pines exhibit some of the largest DNA contents per diploid cell in the plant kingdom ranging from approximately 44 pg (*P. banksiana*) to 75 pg (*P. gerardiana*) (1pg = 960Mb, Arumuganathan and Earle 1991) based on laser flow cytometry (Grotkopp et al. 2004). For instance, the size of maritime pine (*P. pinaster*) genome has been estimated to be between 51 and 60 pg/2C (Chagné et al. 2002; Grotkopp at al. 2004) which is about seven times the size of the human genome (7 pg/2C; Morton 1991) and 170-fold larger than the genome of model plant *Arabidopsis thaliana* (0.3 pg/2C). The DNA amount varies according to the subgenus and the section. The genome of subgenus *Pinus* is lower (54.0 pg/2C) than that of subgenus *Strobus* (65.6 pg/2C). This variation has been related to seed mass. Relationships of such variation with ecological conditions, such as drought tolerance, have also been investigated (Wakamiya et al. 1996). Various environmental conditions appear to be related to DNA content: latitude of range, invasiveness (Grotkopp et al. 2004). An exteremly

large genome size is common to other gymnosperms (reviewed by Ohri and Khoshoo 1986; Murray 1998; Leitch et al. 2001; Grotkopp at al. 2004).

At the whole-genome level, reassociation kinetics data (i.e., Cot analysis) indicates that 25-30% of the pine genome corresponds to low- to single-copy sequences, while 70-75% corresponds to highly repeated sequences (Miksche and Hotta 1973; Rake et al. 1980; Kriebel et al. 1985; Peterson et al. 2006). The repetitive sequences of pine have not been studied in much detail. However, it appears that the majority of repetitive DNA consists of repeats of low sequence complexity (Schmidt et al. 2000), retrotransposons (Kamm et al. 1996; Kossack and Kinlaw 1999; Friesen et al. 2001) and 18S-5.8S-25S rDNA genes. As shown by molecular, cytological and linkage-mapping studies (Friesen et al. 2001; Scotti et al. 2005), most of these repeat sequences are present at multiple loci and are presumably interspersed among other sequences, although they tend to form loose clusters that surround gene-rich islands. The remainder of the genome is composed of low-copy sequences of which an unknown portion are genes. If the pine genome (1C ~ 25-30 pg - see above) is similar to Arabidopsis in having about 30,000 expressed genes and an average gene size of 2000 bp including introns and UTRs (The *Arabidopsis* Genome Initiative 2000), only 0.5% of its genome is likely to be transcribed. In comparison, 54% of the Arabidopsis sequence and 5-6% of mammalian genomes are transcribed (Rat Genome Sequencing Project Consortium 2004).

Gene duplication and the formation of complex gene families have been widely cited as a potential cause of the abundance of low-copy DNA in the pine genome. Kinlaw and Neale (1997) suggested that levels of multiplication were greater in conifers than in other plant species and transcriptional profiling studies have noted surprising levels of transcriptome complexity (Lorenz and Dean 2002). However, considering the relatively low proportion of a conifer genome represented by expressed sequences, it is unlikely that the evolution of multigene families alone can explain the enormous size of pine nuclear genomes. It is possible that a relatively large proportion of low-copy sequences in pine are pseudogenes and/or highly-diverged repeat sequences, and indeed there is evidence supporting both possibilities (Elsik and Williams 2000; Rabinowicz et al. 2005).

1.3. Economic Importance

Pine species play an especially important role in modern plantation forestry worldwide, and now form a large part of both the annual wood harvest and the immature plantation forests that will provide wood in the future (Burdon 2002). Pines enjoy such great popularity because the large number of species allows choice for widely varying site and environmental conditions; the volume of production of some species can be high to very high, even under unfavourable site conditions; they are well suited for reforestation and for simple silviculture (monocultures and clear-felling); easy processing and utilisation of their wood for a wide variety of end-uses (lumber, pulp and paper, particleboard etc.); and even if species lack naturally durable heartwood, treatment with preservatives is easy.

Various minor forest products come from pines (Burdon 2002). Some pine plantations, especially *P. pinaster* in the Landes of southwest France, were established largely for resin production. Resin production from pines was also a major economic activity in the southeast of the USA, Mediterranean basin, northern India, and east and southeast Asia. Resin products are still recovered from chemical pulping of pines. Foliage, in the form of litterfall, has been used for a range of purposes. The bark is used for a variety of purposes, and that of many species has a high tannin content. For a few species with large seeds, notable *P. pinea*, *P. edulis* and *P. sibirica*, seeds have been a prized item of the human diet. Edible fungi, representing the fruiting bodies of mycorrhizal symbionts, are often collected from pine stands.

Pine species are also used widely for shelter and the interlinked purposes of revegetation, soil stabilisation and soil conservation, as befits their tolerance of exposure and degraded soils in certain species. Most of the widely planted species of pines are used in some degree for one or more of these purposes, often with timber production as a major bonus. Ornamental and festive use is common, with *P. sylvestris* and *P. virginiana* being very widely grown for Christmas trees.

1.4. Classical Breeding Objectives

In most pine breeding programs, the setting of goals was originally done subjectively, based on perceptions of the main traits limiting profitability, of the variability and heritability of the traits concerned, and of the genetic correlations (both favourable and adverse) among traits (Burdon 2004). The setting of breeding goals is crucial to any tree-improvement program and yet most pine tree-breeding programs do not have formally defined breeding objectives. Various reasons for this include: complexity of the forest processing industry; difficulties in determining the relationships between selection traits and end-uses; and long rotation ages that create uncertainty about their use (Apiolaza and Greaves 2001). Earlier attempts of most improvement programs have been to improve growth, form, climatic adaptation, and resistance to certain diseases (Cahalan 1981; Danjon 1995; Paul et al. 1997; Shelbourne et al. 1997).

Most recent efforts have been diverted at improving traits related to pulp and paper and solid-wood products (Shelbourne et al. 1997; Chambers and Borralho 1999; Greaves 1999; Lowe et al. 1999; Pot et al. 2002; Kumar 2004). Wood density and fibre morphology (e.g. tracheid length and coarseness) have been reported to be crucial for pulp yield and quality. Wood properties such as wood density and stiffness are crucial selection traits for the improvement of wood stiffness, while compression wood, spiral grain and microfibril angle are candidate traits for reducing in-service instability (Shelbourne et al. 1997; Ivakovic et al. 2006). Heartwood, resinous defects, and internal checking are also being considered as selection traits to develop germplasm for appearance-grade products (Shelbourne et al. 1997). There are some situations, especially involving exotic species and disease resistance etc., where hybrids between selective species are desirable to make genetic improvement in the breeding-objective traits (Hyun 1976; Byun et al. 1989; Blada 1994; Nikles 2000; Shelbourne 2000). Pine species vary widely in their amenability to various forms of vegetative propagation (Hartmann et al. 1990). Ease of propagation was generally a minor criterion for species selection in breeding programs of pines. Currently with improved technology, selection is based more on silvicultural performance and wood quality.

1.5. Classical Breeding Achievements

Various selection and deployment strategies are being used for different pine species. Species selection followed by provenance and family-within-provenance selection, and establishment of seed orchards are quite common first steps across various species including *P. radiata* (Falkenhagen 1991, Matziris 1995), *P. oocarpa* (Moura et al. 1998), *P. caribaea* (Zheng et al. 1994), *P. strobus* (Beaulieu et al. 1996), *P. sylvestris* (Quencez and Bastien 2001), *P. pinaster* (Alía et al. 1995, 1997; Danjon 1995), *P. taeda* (Jiang et al. 1999; Lopez et al. 2000), *P. tecunumanii* (Hodge and Dvorak 1999) and *P. contorta* (Cahalan 1981). In advanced-generation breeding programs, forwards selections and/or combined selection (among- and within-family) are the major sources of genetic gain (Wei et al. 1997; Rossvall et al. 1998; Lambeth 2000; Alazard 2001; Olsson et al. 2001; Plomion et al. 2001; Burdon and Kumar 2004). In the species that are easy to propagate, the concept of using clonal replication of individuals within families as a means of genetic testing is being implemented in order to increase the efficiency of genetic improvement. Significant genetic gains from such a strategy

have been reported for *P. radiata* (Matheson and Lindgren 1985; Shelbourne 1992) and *P. taeda* (Isik et al. 2004).

Pinus patula, *P. taeda*, and *P. elliottii,* are planted in South Africa in summer rainfall zones and their average productivity is 15 m^3 ha^{-1} yr^{-1} (Du Toit et al. 1998). Brazil, with its humid summers, largely uses *P. caribaea, P. taeda* and *P. elliottii.* Pines in Brazil produce 8-30 m^3 ha^{-1} yr^{-1} on rotations of 20-25 years. *P. radiata* is grown as large plantations in the temperate Southern Hemisphere countries of Australia, Chile, New Zealand and South Africa. The mean annual increment (MAI) over 25 years is often 25-30 m^3 ha^{-1} yr^{-1} (Lamprecht 1990). However in New Zealand, growth rates of up to 50 m^3 ha^{-1} yr^{-1} have been recorded on the best sites and as low as 11 m^3 ha^{-1} yr^{-1} on very dry sites (Burdon and Miller 1992). The average productivity of *P. pinaster* in southern France is about 10 m^3 ha^{-1} yr^{-1}, but could reach 20 to 25 m^3 ha^{-1} yr^{-1} on the best sites. Deployment of genetically improved loblolly pine in the USA has been reported to yield up to about 21 m^3 ha^{-1} yr^{-1} (McKeand et al. 2003). Dhakal et al. (1996) reported a realized gain in volume of about 22% in a slash pine improvement programme in the USA. Wood-quality traits that are currently being included, in addition to growth, form and health traits, in the breeding objectives of various pines species, appear to be under moderate-to-strong genetic control (Burdon and Low 1992; Hannrup et al. 2000; Atwood et al. 2002; Gwaze et al. 2002; Pot et al. 2002; Kumar 2004), and predicted genetic gains from selection appeared to be in the order of about 10% for traits such as wood density and stiffness (Kumar 2004).

Clonal forestry (CLF) represents the large-scale propagation and deployment of selected clones, which have been clonally tested. Deployment of tested clones by clonal forestry is being increasingly employed with *P. radiata* in New Zealand (Sorenson and Shelbourne 2005), *P. taeda* in the USA (Stelzer and Goldfarb 1997), and pine hybrids in Australia (Walker et al. 1996). In principle, CLF offers additional genetic gains from capturing non-additive effects, which are not captured via sexual propagation, plus the benefits of greater uniformity and predictability in performance resulting from a lack of genetic segregation. Genetic gains from CLF have been predicted to be considerably higher than those from family forestry in *P. radiata* (Aimers-Halliday et al. 1997), *P. taeda* (Stelzer and Goldfarb 1997), and *P. strobus* (Park 2002).

1.6. Molecular Diversity

The majority of studies aiming to monitor the level and distribution of genetic diversity in *Pinus* natural populations were based on the use of neutral markers. Indeed, molecular markers, such as microsatellites, provided useful information on historical demography and population evolution.

The use of isozyme electrophoresis significantly increased the amount of data on the genetic structure of populations (Petit et al. 2005). These data revealed that pine species had high genetic diversity within populations and only low levels of differentiation among populations. For 28 north temperate pines, genetic differentiation (G_{ST}) averaged 0.076 (Ledig 1998). Exceptions are represented by *P. pinea* (Vendramin et al. in preparation), *Pinus resinosa*, a species that has a vast range across north-eastern North America (Echt et al. 1998; Walter and Epperson 2001), and *P. torreyana* (Ledig and Conkle 1983; Provan et al. 1999) where near absence of variation was observed, and by Mexican pine species where higher differentiation among populations was observed, probably because their natural distributions are more highly fragmented by physiography than those of species at more northerly latitudes. Other pine species with disjunct populations and restricted gene flow also showed higher differentiation among populations: for example 16-27% for *Pinus radiata* and 22% for *P. muricata* (Millar et al. 1997; Wu et al. 1999). On the other hand, experimental evidence indicates that seeds of some pines (e.g. *P. palustris*; Grace et al. 2004) have the potential to

disperse greater distances than previously reported, which partly contributes to the low levels of genetic differentiation observed in these species.

In general, the typical distribution of the genetic diversity within and among populations of *Pinus* species is correlated with their mating system and life history (pines, for example, are wind-pollinated and tend to be predominately outcrossing) (Hamrick and Godt 1996; Duminil et al. 2006) and biogeographic history (the distributions of many species have been affected by Pleistocene glacial advances), even if, in some cases, human activities also played a relevant role (e.g. *P. pinaster* in Portugal; Ribeiro et al. 2001).

Self-fertilization in pines generally occurs at a low level (Muona and Harju 1989), and a high outcrossing at the mature seed stage is maintained. There is evidence of selection at the embryonic stage so that the number of inbreds is already low at the seedling stage (Kärkkäinen and Savolainen 1993). Selection after the seedling stage is still severe. This has been shown by observing the survival of selfed seedlings in *Pinus sylvestris* (Muona et al. 1987) and in *P. leucodermis* (Morgante et al. 1993).

The mode of inheritance has a major effect on the partitioning of genetic diversity in pines, with studies based on maternally-inherited markers (transmitted by seeds only) having significantly higher G_{ST} values than those based on paternally or biparentally inherited markers for pine (Burban and Petit 2003; Petit et al. 2005). In fact, the chloroplast and mitochondrial genomes are generally paternally and maternally inherited in pines, respectively (Petit and Vendramin 2006). In pines, G_{ST} is nearly always larger at mitochondrial DNA markers than at chloroplast DNA markers. On the other hand, there is no significant difference between G_{ST} at biparentally-inherited markers and at paternally-inherited markers in pines (Petit et al. 2005). This is expected considering that both the cpDNA and half the nuclear genomic complement are dispersed by pollen and by seeds, i.e., they use the same vehicles to achieve gene flow.

Chloroplast and mitochondrial DNA markers allowed describing phylogeographic structure in pines. While chloroplast DNA generally exhibits the highest diversity, phylogeographic inferences from these markers can be blurred by extensive pollen flow. Low population structure due to extensive pollen flow has been inferred in *Pinus pinaster* (Burban and Petit 2003), *P. sylvestris* (Robledo-Arnuncio et al. 2005), *P. canariensis* (Gómez et al. 2003) and in *P. albicaulis* (Richardson et al. 2002). In contrast, the mitochondrial markers, despite their generally lower level of diversity in pines (but this holds for all conifers) (see Soranzo et al. 2000; Gugerli et al. 2001), generally provide a clear picture of non-overlapping areas colonized from different refugia (e.g. in *P. pinaster*; Burban and Petit 2003).

Neutral markers also allowed investigating spatial genetic structure (SGS) in natural pine populations. A generally weak within-population structure has been described. *Pinus pinaster* showed a fine-scale structure at the seedling stage with a patch size of approximately 10 meters that seems to be produced by restricted seed flow (González-Martínez et al. 2002). Pines with a heavy seed (differences in seed dispersion capability play an important role), such as *P. pinaster*, are expected to have a short dispersal distance, thus producing a fine-scale structure. However, fine-scale structure often does not persist as stands mature. For example, within-population genetic structure in Mediterranean pines may be affected by post-dispersal events (e.g. mortality due to the severity of the Mediterranean climate and animal-mediated secondary dispersal during the summer period) that may modify the original spatial structure (González-Martínez et al. 2002). Logging can also play a role in decreasing spatial structuring, as observed in *Pinus strobus* (Marquardt and Epperson 2004), suggesting that management practices can alter natural spatial patterns too.

It should be stressed that the presence of fine-scale structure is uncommon in *Pinus* species. Epperson and Allard (1989), studying the spatial pattern of allozyme alleles within *Pinus contorta* ssp. *latifolia* stands, found a lack of structure in the distributions of most

genotypes. Neutral markers by definition do not reflect selective processes, and therefore are not used as an indicator of the population adaptive potential to a changing environment (Morin et al. 2004). Recent and well-established markers to detect functional genetic variation are SNPs, particularly useful for finding genes under selection and studying the dynamics of these genes in natural populations.

The availability of high-density markers, such as SNPs, opens the possibility of studying, by association genetics, the molecular basis of complex quantitative traits in natural populations, taking advantage of the fact that genetic markers in close proximity to causal polymorphisms may be in linkage disequilibrium (LD) to them. The magnitude and distribution of LD determine the choice of association mapping methodology. Extension and distribution of LD depends on many factors including population history (e.g. the presence of population bottlenecks or admixture) and the frequency of recombination.

In order to avoid false associations the optimization of LD mapping requires a detailed knowledge about basic population genetic parameters such as the pattern of nucleotide diversity and linkage disequilibrium for each particular species and candidate gene set.

First estimates indicate that nucleotide diversity varies considerably between plant species. Interestingly, the pines (e.g, *Pinus sylvestris*, Dvornik et al. 2002; García-Gil et al. 2003; *Pinus taeda*, Brown et al. 2004; González-Martínez et al. 2006; *Pinus pinaster*, Pot et al. 2005a) are not amongst the most variable species contradicting expectations from the results obtained using neutral markers and their life history characteristics. First evidences seem to show that broadleaved species (e.g. *Populus*, Ingvarsson 2005; *Quercus*, Table 7 in Pot et al. 2005a) display higher levels of nucleotide diversity than pines.

Markers in specific functional regions of the genome need to be statistically analyzed in order to test for the possibility that these regions might have experienced different selective pressures. In unstructured populations, standard neutrality tests might be applied. When variation is structured in populations, a relatively easy approach is the comparison of genetic differentiation estimates, such as Wright's *F*-statistics, among markers tagging a putative gene under selection and neutral markers or expected distributions computed using coalescence theory (see reviews in van Tienderen et al. 2002; Luikart et al. 2003). If population divergence (F_{ST}) is higher for the gene-targeted marker with respect to a divergence estimates obtained from random markers, this might indicate divergent selection and local adaptation for the tagged gene (van Tienderen et al. 2002). Pot et al. (2005a) found a high differentiation among populations at *Pp1* (*glycine-rich protein* homolog) gene in *Pinus pinaster*, higher than at the neutral level. This result is consistent with diversifying selection acting at this locus in this species. On the other hand, the absence of differentiation observed for the gene *CeA3* (*cellulose synthase*), compared with the significant level observed at neutral markers may indicate balancing selection acting on this gene.

Recent studies on pines reveal a rapid decay in LD with physical distance. LD declines very rapidly within 200-2000 bp in *Pinus taeda* (Brown et al. 2004; González-Martínez et al. 2006), *Pinus sylvestris* (Dvornyk et al. 2003, García-Gil et al. 2003). A rapid decay of LD in pines is consistent with what is expected from outcrossing species with large effective population size.

2. Construction of Genetic Maps

In genetics, mapping is defined as the process of deducing schematic representations of DNA. Three types of DNA maps can be constructed depending on the landmarks on which they are based:

- physical maps, whose highest resolution would be the complete nucleotide sequence of the genome,

- genetic maps, which describe the relative positions of specific DNA markers along the chromosomes, determined on the basis of how often these loci are inherited together,
- and cytogenetic maps, a visual appearance of a chromosome when stained and examined under a microscope.

To provide a first glimpse of the pine genome, high-resolution genetic maps have been established for several pine species using different types of molecular markers and following different strategies that are reviewed in the following sections.

2.1. Development of Molecular Markers in Pines

The construction of a linkage map relies on the availability of enough molecular markers to detect linkage between them. Each type of marker technology has advantages and limitations. Many factors (e.g. polymorphism information content, level of polymorphism exhibited for the mapping progeny, mode of inheritance, genome size) can influence the development of a particular technique and the choice of a marker system for a given purpose (e.g. genetic mapping, QTL analysis, survey of genetic diversity, forensic applications). The purpose of this section is to briefly review the different types of molecular markers techniques that have been developed in pines and used for genetic mapping applications. We will not present the detail of each technique. Both the review by Cervera et al. (2000a) and the references cited in Table 2 will provide the reader with the necessary information for understanding the scientific basis of each technique.

2.1.1. Isozymes and Proteins

The first markers developed for pine were isozymes. Linkage studies were carried out on more than 10 species for about 15 loci (reviewed by Tulsieram et al. 1992). Conkle (1981) located more loci, but still not enough to cover the pine genome. Proteins revealed by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE; O'Farrell 1975) presented the advantage of being multiplexed compared to isozymes. Importantly, proteins can be easily characterized by mass spectrometry (e.g. Gion et al. 2005) and may be recognizable by sequence similarity to others proteins published in sequence databases, therefore providing functional markers expressed in the tissues analyzed. Two-dimensional protein markers were developed only in *Pinus pinaster* (reviewed in Plomion et al. 2004 and Cánovas et al. 2004). Although proteins provided physiologically relevant markers to map the expressed genome, this time consuming technique failed to provide enough markers for genetic application which requires full genome coverage, such as linkage mapping and QTL detection.

2.1.2. RFLPs

RFLPs were developed in the early 1990s for *P. taeda* (Neale and Williams 1991). They offered a sufficient number of markers for high density genome mapping in pine. However, this labor intensive and time-consuming technique was only applied to *P. taeda* (Devey et al. 1994) and *P. radiata* (Devey et al. 1996).

2.1.3. RAPDs and AFLPs

In the mid-nineties PCR-based multiplex DNA fingerprinting techniques provided very powerful tools to generate dense linkage maps in a short period of time. RAPD (random amplified polymorphic DNA, Williams et al. 1990), and then AFLP (amplified fragment length polymorphisms, Vos et al. 1995) became the most popular marker technologies in conifers. Despite their biallelic nature and dominant mode of inheritance (which was actually not an issue for mapping with haploid megagametophyte, see section 2.2.1.), these markers tremendously boosted up genetic analysis in most forest tree species including pines (reviewed in Cervera et al. 2000b).

2.1.4. Nuclear Microsatellites

In contrast to other plant species, few polymorphic single-copy nuclear microsatellite markers or simple sequence repeats (SSR) have been reported in pines (reviewed in Chagné et al. 2004). The genome structure of these species, characterised by a large physical size, with a large amount of repeated sequence (see section 1.2) has been the main obstacle to the development of useful markers using classical simple sequence repeat-enriched library approaches (e.g. Auckland et al. 2002; Guevara et al. 2005a; Echt and Nelson unpublished results). In addition, the ancient divergence time between coniferous species (Price et al. 1998) and the complexity of their genomes means that transferability of single-copy SSRs among genera and even within *Pinus* is generally poor, resulting in a large proportion of amplification failure, non-specific amplification, multi-banding patterns or lack of polymorphism (Echt et al. 1999; Mariette et al. 2001). In an attempt to circumvent these genome-related problems, Elsik and Williams (2001) removed most of the repetitive portion of the genome using a DNA reassociation kinetics-based method, and Zhou et al. (2002) targeted the low-copy portion of the genome using an undermethylated region enrichment method. Both approaches yielded remarkable enrichment for useful SSR markers in *P. taeda*. SSRs made from low copy, undermethylated, and total genomic DNA yielded mappable markers (Nelson et al. 2003; Zhou et al. 2003). *Pinus taeda* SSRs developed by Elsik and Williams (2001) and Zhou et al. (2002) transferred quite well between American hard pines (Shepherd et al. 2002a), but were shown to be less transferable in the phylogenetically divergent Mediterranean hard pines (Chagné et al. 2004; González-Martínez et al. 2004). Interestingly, perfect trinucleotide SSRs transferred from American to Mediterranean pines better than other motifs (Kutil and Williams 2001). A number of nuclear SSR markers have been developed for *P. radiata*, almost all of which are based on the more frequently polymorphic dinucleotide repeat motifs (Smith and Devey 1994; Fisher et al. 1996; Fisher et al. 1998; Devey et al. 2003), and used in a number of applications. Polymorphic chloroplast microsatellite loci have also been identified and applied (Cato and Richardson 1996; Kent and Richardson 1997). More recent SSR discovery efforts have been undertaken in both New Zealand and Australia, and have been most commonly applied to fingerprinting (Kent and Richardson 1997; Bell et al. 2004) and QTL mapping applications (Devey et al. 2004a).

2.1.5. EST-Polymorphisms

With the availability of sequence data obtained by random sequencing of pine cDNAs (see section 5.2.1), there is now a clear trend towards the development of gene-based markers (ESTP: EST-polymorphisms). There are basically two groups of technologies used to detect nucleotide polymorphisms (SNPs and INDELs), either based on the knowledge of nucleotide variants or not. Up to now, techniques based on the detection of differences in the DNA stability (DGGE, denaturating gradient gel electrophoresis**,** Myers et al. 1987), conformation (SSCP, single strand conformation polymorphisms, Orita el al. 1989) under specific polyacrylamide gel conditions, or heteroduplex cleavage (TILLING, targeting induced local lesions in genomes, Colbert et al. 2001), have been successfully applied in pines (Plomion et al. 1999; Temesgen et al. 2001**;** Chagné et al. 2003; Ritland et al. 2006). With the decrease of sequencing costs and the availability of pine cDNA sequences, more targeted and precise approaches are now possible (Pot et al. 2005b). The bioinformatics assembly of ESTs into large contigs (i.e. unigenes) has also made it possible to identify putative SNPs. Le Dantec et al. (2004) identified a set of 1400 candidate SNPs in *Pinus pinaster* contigs containing between 4 and 20 sequence reads. This represents a great resource of molecular markers for this species that can be used to map candidate genes, study linkage disequilibrium, and develop comparative orthologous sequence markers for comparative genome mapping. In

addition to SNPs and INDELs, a large set of microsatellite markers have been developed from *Pinus taeda* and *Pinus pinaster* ESTs (Echt and Burns 1999; Chagné et al. 2004; Echt et al. 2006). These markers present the advantage of being highly polymorphic and located in coding regions.

2.2. Haploid and Diploid-Based Mapping Strategies

The construction of a genetic map requires two components: first a segregating population (mapping pedigree) derived from a cross between parental trees that are heterozygous for many loci, and second a set of molecular markers segregating in the progeny according to Mendelian ratios. Linkage map construction is based on the statistical analysis of polymorphic markers in the mapping population, considering that the distance between two loci is related to the probability of observing a recombination event between them. There is a number of mapping software to facilitate automated analysis (listed at URL http://linkage.rockefeller.edu/soft/list.html).

2.2.1. Haploid or Half-Sib Based Mapping Strategy

In conifers, the haploid megagametophyte constitutes an ideal plant material for genetic mapping. This nutritive tissue surrounding the embryo is derived from the same megaspore that gives rise to the maternal gamete. Therefore, it represents a single meiotic event in the parent tree being genetically equivalent to a maternal gamete. The dominant nature and biallelic mode of inheritance of RAPD and AFLP is not an issue for genetic mapping with haploid megagametophytes. However, quantitative traits can only be measured on half-sib seedlings, limiting the detection of QTLs at the first stages of tree development. Thus, this approach is not applicable to the analysis of QTLs for economically important traits in well-established plantations, unless the megagametophytes were collected and saved, which has generally not been the case.

2.2.2. Diploid or Full-Sib Based Mapping Strategy

Different strategies have been followed to construct genetic linkage maps of *Pinus*: the "pseudo-testcross strategy", the "F_2 inbred model", and the "three-generation outbred model". The pseudo-testcross strategy is mainly based on selection of single-copy polymorphic markers heterozygous in one parent and homozygous null in the other parent and therefore segregating 1:1 in their F_1 progeny as in a testcross. Grattapaglia and Sederoff (1994) introduced the term “two-way pseudo-testcross” to define this mapping strategy, where two independent genetic linkage maps are constructed by analysing the co-segregation of markers in each progenitor. The efficiency of this strategy, as well as for the haploid mapping strategy, depends on finding individual trees that are heterozygous for many loci, which is quite easy using arbitrarily primed PCR assays (RAPD and AFLP) in highly-heterozygous out-crossed tree species such as pines. The F_2 inbred model is based on a three-generation pedigree for which the grandparents are treated as inbred lines. In the F_2 generation, three genotypes occur at any locus: AA, AB, and BB, segregating 1:2:1. The three-generation outbred model (Sewell et al. 1999) is an extension of the pseudo-testcross strategy. Within a single outbred pedigree, any given codominant marker will segregate in one of three different ways. When one parent is heterozygous and the other is homozygous, segregation will be 1:1 (i.e., testcross mating type). When both parents are heterozygous, segregation will be either 1:2:1, if both parents have the same genotype (i.e., intercross mating type), or 1:1:1:1 if they have different genotypes (i.e., fully informative mating type). This segregation data is then subdivided into two independent data sets that separately contain the meiotic segregation data from each parent, and independent maps are constructed for each parent. A sex-average map is then constructed using an outbred mapping program, which uses fully informative and intercross

markers to serve as common anchor-points between each parental data set. Compared to "megagametophyte progeny", full-sibs can be grafted and/or propagated by cuttings, therefore constituting a perpetual population, analogous to recombinant inbred lines in crop plants. The use of such clonally propagated progeny obviously increases the precision of quantitative measurements and therefore enhances the QTL detection power (Bradshaw and Foster 1992).

2.3. Genetic Mapping Initiatives in Pines

In this section and Table 3 we summarize what has been done in terms of linkage map construction in the genus *Pinus* with emphasis made on the most studied species. Besides, some maps have been published together with QTL studies and will be found in the references cited in section 3.

2.3.1. Maritime Pine

Linkage maps of the Maritime pine (*Pinus pinaster* Ait.) genome have been first constructed by analysing the co-segregation of proteins extracted from megagametophytes collected during the germination of the embryo. Bahrman and Damerval (1989) were the first to report a linkage analysis for 119 protein loci using 56 megagametophytes of a single tree. Extending this approach, Gerber et al. (1993) reported a 65 loci linkage map covering one fourth of the pine genome, using 18 maritime pine trees with an average of 12 megagametophytes per tree. A more conventional pedigree (inbred F_2) was used to map 61 proteins using haploid (Plomion et al. 1995a) and diploid (Plomion et al. 1997; Costa et al. 2000) tissues of the same seedlings. In the later case, protein loci were found on each chromosome (Thiellement et al. 2001). As stated above, the advance of PCR-based markers has allowed the construction of saturated linkage maps, in a short period of time, with no prior knowledge of DNA sequence (Plomion et al. 1995a, b; Costa et al. 2000). The two-way pseudo-trest cross mapping strategy was used to construct genetic linkage maps of maritime pine using AFLP markers (Chagné et al. 2002; Ritter et al. 2002) as well as AFLP, SAMPL, SSR and gene-based markers (de María and Cervera unpublished results). Comparing the total map distance of genetic maps constructed based on haploid and diploid progeny from the same Maritime pine tree, a higher rate (28%) of recombination in the pollen parent was found (Plomion and O'Malley 1996). Such a significant difference obtained between male and female recombination was also reported in other pine species (Moran et al. 1983; Groover et al. 1995; Sewell et al. 1999).

2.3.2. Loblolly Pine

Loblolly pine (*Pinus taeda* L.) has been used extensively for genetic mapping including the development of an early map based on 20 isozyme loci that included five linkage groups (Conkle 1981). More recently maps have been constructed in several pedigreed populations using several types of DNA-based markers. Devey et al. (1994) published the first map utilizing 90 RFLP and six isozyme loci. The map was based on a three-generation pedigree with 95 progeny and it revealed 20 linkage groups. Genomic mapping was successfully used by Wilcox et al. (1996) to define a single gene locus for resistance to an isolate of the fungus (*Cronartium quercuum* f. sp. *fusiforme*) that causes fusiform rust disease. O'Malley et al. (1996) described a RAPD-based map of clone 7-56, a top *P. taeda* parent. This map had been constructed several years prior to publication. The first consensus map was produced by Sewell et al. (1999) combining data from two three-generation pedigrees, including the pedigree used by Devey. RFLP, RAPD, and isozyme markers were placed on the integrated or consensus map containing 357 loci and covering about 1300 cM of genetic distance on 18 linkage groups. The first complete genome map was developed by Remington et al. (1999). They utilized haploid megagametophyte samples from an individual mother tree to develop a map based on 508 AFLP markers. This map revealed 12 linkage groups equaling the basic

number of chromosomes and about 1700 cM of genetic distance. Their analysis suggested that this distance saturated the genome in slight contrast to an earlier estimate of 2000 cM based on data from three species of pines (Echt and Nelson 1997). SSR markers developed in radiata pine (*P. radiata*) were used in a comparative mapping project between radiata and loblolly pine (Devey et al. 1999). Of the 20 SSR markers tested only nine were mapped in both species, however these codominant markers along with several codominant RFLP markers were useful in defining homeologous linkage groups between the species. Temesgen et al. (2001) added 56 ESTP markers to the consensus map developed earlier by Sewell. The DGGE method proved quite useful for assaying ESTP markers and suggested a general method for placing genes on the maps since the markers were developed from expressed sequences. Additional ESTP markers were developed and used to identify anchored reference loci based on their sequence similarity between species and their nature to map to conserved locations in more than one species (Brown et al. 2001). Zhou et al. (2003) mapped 51 SSR markers, covering 795 $cM_{(K)}$ on 15 linkage groups, in a three-generation pedigree with 118 progeny. The markers, developed from loblolly pine libraries of three types, were not found to be clustered within the genome further highlighting the value of SSR markers in genome mapping.

2.3.3. Radiata Pine

Over the past fifteen years, a range of DNA marker-based linkage maps have been constructed for this species. Wilcox (1997) briefly reviewed mapping studies undertaken up til that date. Although results were summarised from seven studies involving construction of eight linkage maps, only one map had actually been published by that date, consisting of 208 (mostly) RFLP, SSR and RAPD markers (Devey et al. 1996, 1999). This map consisted of 22 linkage groups and covered 1382 cM. All of the other maps reviewed by Wilcox (1997) were constructed using RAPD markers, either using haploid megagametophytes, or diploid tissues using a pseudotestcross approach. The number of markers used in these studies ranged from 124 to 290, with only one of the maps having linkage groups equal to the haploid number of chromosomes, the remainder ranging between 14 and 22 linkage groups. Total map length estimates were undertaken in three studies ranged between 1978 and 3000 cM. Subsequent to these earlier studies, a number of maps have been published, some of which were included in Wilcox's 1997 review. Using 222 RAPD markers to genotype 93 megametophytes, Emebiri et al. (1998) constructed a linkage map that covered 14 linkage groups and spaned a total distance of 1665 cM. Kuang et al. (1999a) described a map constructed using megagametophytes of 198 S_1 seeds that had been genotyped with 168 RAPD and four microsatellite markers. The resulting map consisted of 19 linkage groups, with covering 1116.7 $cM_{(K)}$, which was estimated to cover 56% of the genome. Because this was constructed using an S_1 family, elevated levels of segregation distortion were observed. Wilcox et al. (2001) published framework maps of both parents of 93 full-sib progenies based on a total of 429 AFLP, RAPD and SSR markers. These parent-specific maps were constructed using a pseudotestcross strategy, and covered 1414 and 1144 $cM_{(K)}$ in 20 and 21 linkage groups, respectively. These maps have subsequently been added to, using over 300 SSR and EST markers, and have been reduced to 12 linkage groups (Wilcox et al. 2004). Devey et al. (1999) published a comparative map of radiata and loblolly pines based on RFLP, SSR and RAPD loci, and showed the highly syntenic nature of *Pinus* applies to these two economically important species. Overall estimates of map length appear to be similar to that of loblolly pine (Wilcox et al. 2001a).

2.3.4. Scots Pine

Early mapping work in *Pinus sylvestris* has been based on isozyme loci (Rudin and Ekberg 1978; Szmidt and Muona 1989). These maps contained no more than 20 loci. The number of RFLP markers developed for Scots pine has been very low (Karhu et al. 1996). Thus, the next mapping efforts were based on RAPD makers segregating in haploid megagametophytes. Yazdani et al. (1995) mapped 261 markers in 14 linkage groups in a tree that is part of the breeding program. Hurme et al. (2000) made a low coverage RAPD map for an F_1 tree that was a result of north x south cross, such that alleles for important quantitative traits were assumed to segregate in the same cross. The map with the best genome coverage, so far, was constructed by Yin et al. (2003). The AFLP mapping in a full-sib family resulted in two maps, for each of the breeding program parents. The map lengths for the two parents, based on about 200 framework markers, were about 1645 and 1681 cM for the male and female trees, respectively, with very high estimated genome coverage. Most recently, Komulainen et al. (2003) mapped about 60 gene based markers in the F_1 progeny of a north x south cross. This map also contained markers that had been previously developed for *Pinus pinaster* (Plomion et al. 1999), and others that had been used for *P. taeda.* Most importantly, the homologous markers allowed defining the correspondence between 12 linkage groups in the two species. More markers have been later added to this map (Pyhäjärvi et al. unpublished).

2.3.5. Longleaf Pine, Slash Pine, Caribbean Pine and their Hybrids

Longleaf pine (*P. palustris* Mill.), slash pine (*P. elliottii* Engelm. var. *elliottii*), and Caribbean pine (*P. caribaea* Morelet.) are hard pines of subsection *Australes* found along the coastal plains of the southeastern United States, eastern Central America, and the Caribbean islands. Although loblolly pine (*Pinus taeda* L.) has been planted on millions of acres that were once typically occupied by these species, in many situations these pines are proving to be the preferred timber species due to their adaptation to the coastal-plain soils and the associated natural disturbances such as frequent fires and hurricanes (Wahlenberg 1946; Shoulders 1984). A number of genetic maps, consisting primarily of random amplified polymorphic DNA (RAPD) markers have been constructed for slash pine, longleaf pine, and their hybrids (Nelson et al. 1993, 1994; Kubisiak et al. 1995) with the main goal being to use these marker maps as a tool for dissecting the inheritance of specific traits of interest and for use in marker-assisted-selection (MAS) strategies within tree improvement programs (Kubisiak et al. 1997, Kubisiak et al. 2000, Weng et al. 2002). Some markers significantly linked to traits of interest have been converted to more easily scorable markers such as sequence characterized amplified region (SCAR) markers to help aid selection efforts (Weng et al. 1998). Brown et al. (2001) assembled a genetic linkage map for slash pine using a variety of markers [RAPDs; expressed sequence tag polymorphisms (ESTPs); restriction fragment length polymorphisms (RFLPs); and isozymes]. An additional genetic map for slash pine and one for Caribbean pine, using amplified fragment length polymorphisms (AFLPs) and microsatellite or SSR markers, were created using an F_1 hybrid population (Shepherd et al. 2003). The number of markers mapped and the genetic distances covered by some of the published maps for these species and their hybrids are summarized in Table 3. Unlike loblolly pine significantly less effort has been focused on comparative mapping across these species. However, studies have shown synteny across slash and longleaf pines using RAPDs (Kubisiak et al. 1995, 1996), slash and loblolly pines using ESTPs as anchored reference loci (Brown et al. 2001), and slash and Caribbean pine using AFLPs (Shepherd et al. 2003). In addition, SSR markers look promissing for further comparative analyses across these species (Shepherd et al. 2002a; Nelson CD and Echt CS, personal communication).

2.3.6. Other Pines

Genetic linkage maps were also constructed in other pine species including *P. brutia* (Kaya and Neale 1995), *P. contorta* (Li and Yeh 2001), *P. edulis* (Travis et al. 1998), *P. massoniana* (Yin et al. 1997), *P. strobus* (Echt and Nelson 1997), *P. thunbergii* (Kondo et al. 2000; Hayashi et al. 2001). These were based mainly RAPD and AFLP.

2.4. Genetic vs. Physical Size and Practical Implication

These mapping studies have led to the conclusion that the total genetic distance of the pine genome was around 2000 cM (Gerber et al. 1994), i.e. about 167 cM per chromosome. Given a physical size of 25 pg/C, one unit of genetic distance (1 cM) would therefore correspond to 13 x 10^6 nucleotides (13 Mb), while it represents 0.23 Mb in the model plant species Arabidopsis! Such high genetic/physical size ratio obviously hampers the characterization of QTLs by fine-mapping and positional-cloning approaches. Hence, as it will be discussed in sections 3 and 4, the only way for understanding the molecular basis underlying quantitative trait variation is the candidate-gene approach in which genes are identified *a priori* as likely candidates for the trait of interest, and their polymorphisms tested against quantitative trait variation. Interestingly, despite the 56-fold difference between *Pinus* and *Arabidopsis* chromosomes, the number of crossing-over per chromosome was found to be highly conserved between both genus: 2 to 4 chiasmata per bivalent (1 chiasma = 50 cM; Ott 1991). A comparison of genome lengths among evolutionary divergent pines found *Pinus pinaster*, *P. palustrus*, and *P. strobus* to have essentially identical rates of recombination (Echt and Nelson 1997). Thus, genetic mapping studies carried out in pines have clearly demonstrated that the mechanism of crossing-over is conserved on a chromosomal basis, and independent of physical map size and the fraction of coding DNA.

2.5. Comparative Mapping: Towards the Construction of a Unified Pine Genetic Map

All pine species have the same number of chromosomes (i.e. n = 12), as well as a similar genome size. Moreover, they are all diploid, which suggests that their genome might be relatively well conserved among species. Approaches for evaluating genome similarity have used cytogenetics and linkage map comparison. Although cytogenetics can provide a direct idea about the conservation between different genomes (Hizume et al. 2002), most of the interspecies comparisons have been carried out using genetic maps. The same tools are also developed for other applications such as gene mapping and QTL detection. Comparative genome mapping aims to measure the conservation of gene content (synteny) and order (collinearity) along the chromosomes and uses orthologous loci as anchor points between maps. Comparative genome mapping has been successfully used in grasses to explain the genome evolution of cereals (Moore et al. 1995), used for choosing a model species for whole genome sequencing (rice, Ware et al. 2002; http://www.cns.fr/externe/English/Projets/Projet_CC/organisme_CC.html), and for transfering of genetic information between related species, such as the position of candidate genes (Schmidt 2002). The lack of genome sequence for a Pinaceae species has made comparative mapping even more important as the primary tool for integrating genetic information across species.

To define the syntenic relationships among phylogenetically related pine species, orthologous markers (i.e. homologous DNA sequences whose divergence follows a speciation event and whose sequence and genome location is conserved between different species), are used. A first example was provided by Devey et al. (1999) who aligned the genetic maps of *P. taeda* and *P. radiata* using RFLPs and SSRs. This first effort was further consolidated in the frame of the Conifer Comparative Genomics Project (http://dendrome.ucdavis.edu/ccgp). Low copy cDNA PCR-based markers were developed in loblolly pine (Harry et al. 1998; Brown et al. 2001; Temesgen et al. 2001; Krutovsky et al. 2004). These markers showed a relatively

good PCR cross-amplification rate between pine species because they target conserved coding regions, showing a relatively high polymorphism rate and a low number of paralogous amplification, when PCR primers were chosen carefully. They were used to study the synteny between species belonging to the Pinaceae family, which included pines, alongside with other important conifers such as spruces and firs. These markers made it possible to assign linkage group homologies for 10 out of 12 chromosomes between *P. elliottii* and *P. taeda* (Brown et al. 2001; http://dendrome.ucdavis.edu/ccgp/compmaps.html), 10 out of 12 between *P. pinaster* and *P. taeda* (Chagne et al. 2003) and 9 out of 12 between *P. sylvestris* and *P. taeda* (Komulainen et al. 2003). About 30 or 40 ESTP markers proved to be useful in demonstrating large areas of synteny between each species pair. While this comparison was only of low density, these pioneer studies suggested that pine genomes did not show any apparent chromosomal re-arrangement. They also provided indication that gene content and gene order is conserved, as it is illustrated for linkage group 6 in Figure 2. Current efforts are being paid to add more markers common to *P. radiata* and *P. taeda* (P Wilcox, personal communication). From an application point of view, these comparisons provide a set of markers that can be used for constructing framework genetic maps of pine species for which maps have not yet been developed.

As more conifer ESTs become available in public databases (329,531 in *Pinus teada*, 132,531 in *Picea glauca*, 27,283 in *Pinus pinaster*, 28,170 in *Picea engelmannii x Picea sitchensis*, 80,789 in *Picea sitchensis*, 6808 in *Pseudotsuga menziesii*, 7639 in *Cryptomeria japonica:* EMBL March 19th, 2006), computational approach could be used for *in silico* development of putative orthologous EST-based markers (Fulton et al. 2002), as it was recently illustrated between loblolly pine and douglas-fir by Krutovsky et al. (2004). Such resources should help to define the precise syntenic relationship across conifers and establish a framework for comparative genomics in Pinaceae.

The alignment of genetic maps of *P. pinaster* and *P. taeda* made it possible to discover putative conserved QTLs between the two species (Chagne et al. 2003). Those are QTLs for wood density and wood chemical composition traits located on linkage groups 3 and 8, respectively. The same observation has been made in *P. radiata* (Telfer et al. 2006). Preliminary studies indicate wood density QTLs are co-locating more frequently than would be expected by chance between these two species (Telfer et al. 2006), and analyses are being extended to other wood property traits. Moreover, candidate genes coding for functions which are linked to wood formation have been mapped in the same regions, which indicates that they may be involved in the molecular control of those traits. These first examples of application of comparative genome mapping in pines show that comparative genome mapping can be used to verify QTLs across species and that the same genes may be involved in the genetic control of the same traits.

3. Genetic Architecture of Complex Traits

Pine tree improvement is hampered by different inherent characteristics: i) the time needed to reach sexual maturity, ii) the time lag required to evaluate field performance (e.g. growth, Kremer 1992), and iii) in some cases, the cost of phenotyping (e.g. wood quality related traits). This makes breeding of these species a slow process compared to most common crop plants. In addition, selection of these traits remains imprecise because environmental effects are rather high for most traits of interest. Heritabilities for height, diameter, volume, branching traits, and bole taper, i.e., straightness, are in the range of 0.1-0.3, and only slightly higher (0.3-0.6) for wood and end-use properties (reviewed by Cornelius 1994). In this context, any tool directed toward selection processes that would improve the evaluation of genetic value and also reduce the generation time would be of considerable value.

Most traits important to forestry, such as biomass production, wood quality, biotic and abiotic stress resistance are complex quantitative traits. In the theory of Quantitative Genetics it is assumed that the heredity of a quantitative trait can be ascribed to the additive effects of a large number of genes with small and similar actions, modulated by environment. This assumption has been questioned since the early experiments of Sax (1923) and Thoday (1961), and it is now well known that a small number of segregating loci are involved in the genetic control of quantitative trait variation. These genes act together to provide a quantitative difference, and are referred to as quantitative trait loci, or QTL (Gelderman 1975). The basic theory of using genetic markers to detect QTLs was introduced by Sax (1923). Initially, the application of this theory was limited by the lack of available segregating markers, however, rapid advances of DNA marker-based technologies since the 1980s have now made it possible to genoptype hundreds of genetic markers to construct dense genetic linkage maps (see section 2.3), and further to carry out a comprehensive search of QTLs along the genome. Sewell and Neale (2000) and Guevara et al. (2005b) recently reviewed the science of QTL mapping in pine trees.

In the first part of this section, devoted to the genetic dissection of agronomically important traits, we will present an update of the studies that were carried out in pines. The the specificity of pines regarding the different types of populations and strategies used to detect QTLs will be presented. Then, the main results of these studies will be discussed. In the second part, the perspective towards the identification of diagnostic markers for pine tree breeding will be discussed.

3.1. Strategy and Methods Used for QTL Detection in Single Family Pedigrees

Pines are characterized by late maturity (longevity), an outbred mating system, and a high genetic variability (Hamrick and Godt 1990). Their outbred mating system and high genetic load have hampered the construction of inbred lines, the material of choice from a QTL mapping perspective. In this context, specific populations and statistical methods were developed specially for forest trees, and pines in particular. Until recently, most of the QTL mapping efforts were focused on single family pedigrees. However, given to the high rate of polymorphism encountered in pines and the relative lack of stability of QTLs in different genetic backgrounds, methods aiming at validating markers linked to the traits of interest in more complex pedigrees or even in unrelated genotypes are emerging. In this section, we will review the type of population, statistical methods, and main results that have been obtained using simple mapping pedigrees. The limitations of this approach will be underlined and the use of complex pedigree designs or unrelated populations to unravel the genetic architecture of complex traits will be presented. Finally, the relevance of integrative approaches combining classical QTL studies, transcriptome and proteome analysis, together with studies of molecular evolution will be highlighted.

3.1.1. Single Family Designs

Two-Generation Full-Sib Design

The advantage of naturally high levels of heterozygosity in outbred forest trees can be utilized in a two-generation population structure, where it may be expected that trees chosen as potential parents will likely be heterozygous for some number of QTLs which will then segregate in the F_1. Typically, “plus trees” are identified and are used as parents of a QTL mapping population. Various studies have taken advantage of this two-generation design to analyze each parent under a pseudo-testcross model (Kumar et al 2000; Lerceteau et al. 2000; Ball 2001; Shepherd et al. 2002b; Weng et al. 2002; Markussen et al. 2003; Yazdani et al. 2003). This model is well suited for dominant markers. However, the main limitation is that

the phenotypic effects inherited from each parent are analyzed individually, even though the genetic contribution of each parent simultaneously contributes to the phenotypic variation in the progeny population. Consequently, the genetic information in the four progeny classes of an outbred pedigree are collapsed into only two genotypic classes, thereby reducing the robustness of the analysis. Of course, if co-dominant markers are used, a consensus map can be built precisely to detect allele effects from both progenitors simultaneously.

Three-Generation Full-Sib Design
In the three-generation outbred population structure, two crosses are made among four unrelated grandparents, where each mating pair is selected among individuals displaying divergent phenotypic values for the trait (e.g., Groover et al. 1994). From each grandparental mating, a single phenotypically intermediate individual is chosen as a parent. Presumably, these intermediate parents are heterozygous for both marker and QTL alleles, and are potentially heterozygous for different allelic pairs which display a divergent phenotypic effect. This three-generation full-sib structure is typically designed for QTL analysis under an outbred model and has been extensively used (Table 4). Alternatively, Plomion et al. (1996a, b) utilized the selfing ability of *P. pinaster* to design experiments that fit an inbred F_2 model. Although RAPD markers were used, genetic information from progeny and corresponding megagametophytes were utilized to overcome the limitations associated with dominant markers using an F_2 model. It was also proposed to use trans dominant linked markers to overcome the problem of dominant markers on the sporophytic phase of F_2 trees (Plomion et al. 1996c).

Two-Generation Half-Sib Design
With this population structure, the effects of two maternal QTL alleles are averaged over a large pollen pool (Hurme et al. 2000). This type of structure allows testing the stability of the effect of the maternal alleles in different genetic backgrounds.

In addition to these classical mapping designs, particular populations (F_1S) were especially developed to analyse inbreeding depression in *P. taeada* and *P. radiata* (Kuang et al. 1999 a, b; Remington and O'Malley 2000a, b; Williams et al. 2001).

Two-Generation Full-Sib and Half-Sib Design
An extension of the pseudo-test cross QTL mapping strategy, in which QTLs are defined in a narrow genetic background, Plomion and Durel (1996) show that a "general" value of a "specific" QTL detected in a full-sib family could be easily evaluated, providing that both parents of the full-sib are involved in maternal half-sib (open-polinated or polycross) families. Such two-generation pedigrees are widely available in most pine breeding programs that involve the simultaneous estimation of specific and general combining abilities of selected trees. However, this strategy has never been tested experimentaly.

3.1.2. Methods Used for QTL Mapping in Single Family Designs
Regardless of the population structure and size (Beavis 1994), several factors must be considered for successful QTL detection. Statistical methodology significantly influences the accuracy of QTL position and effect estimation. Simple statistical methods such as analysis of variance (ANOVA) have opened the way to the development of more powerful QTL detection methods integrating information available at multiple markers: Interval Mapping (IM), Composite Interval Mapping (CIM), and Multiple Interval Mapping (MIM).

The first method called "single marker analysis", proposed by Edwards et al. (1987) is the simplest one. ANOVA is performed one marker at a time, on the genotypic classes defined by a single marker. This method suffers from several limitations. First, it does not

provide any information regarding the location of the QTL in the genome, and furthermore the definition of the QTL effect is largely inaccurate given the inability to separate small effect QTLs with tight linkage from QTLs of large effect but more distant linkage.

In order to improve the efficiency of QTL mapping through the use of genetic map information, Lander and Botstein (1989) developed the IM method. This approach allows QTLs to be detected in the intervals defined by the markers. The IM has been widely used in pines (Groover 1994; Plomion et al. 1996a, b; Knott 1997; Kaya et al. 1999; Costa and Plomion 1999; Costa et al. 2000; Sewell et al. 2000, 2002; Brown et al. 2003). This method is however rather limited as QTL detection is made in a linear way, i.e., the same test is applied at each point of the interval without taking into account the results of successive tests.

More recently, statistical approaches have been developed to increase the statistical power of QTL detection and have received increasing attention in pines (Costa 1999; Brendel et al. 2002; Pot et al. 2005b). One of these approaches is Composite Interval Mapping (CIM) developed by Zeng (1993a, b) and Jansen (1993), which combines interval mapping with multiple regression. Like IM, this method evaluates the presence of a QTL at multiple analysis points across each interlocus interval. However, at each point it also includes in the analysis the effect of one or more markers elsewhere in the genome. However, although IM and CIM brought significant improvements, both methods lack dimensionality: i.e. only a single QTL is being searched at a time.

The method proposed by Kao et al. (1999) and Zeng et al. (2000) called MIM (Multiple Interval Mapping) differed from the previous methods through the implementation of multi-QTL models. The selection of the best model (i.e. QTL number, position, effect, and interactions), which best fits the data follows an iterative process. After identification of QTL number and position by CIM, the MIM strategy consists of looking for additional QTLs through forward-backward selection cycles while integrating interaction information between the different QTLs. Within each iterative cycle, QTL position and effect are re-evaluated. MIM also allows the evaluation of QTL epistasis. However, it is important to note that, although this strategy provides significant improvements compared to CIM, it does not allow detection of non-significant QTL at the individual level. Therefore, even if the concept aimed at detecting multiple QTL effects is simple, its implementation is relatively complex given the number of potential QTLs and the resulting number of possible interactions. Thus, the problem is no longer genetic, but resides in the ability to test all the possible genetic models and select the one which best fits the observed data (*model selection*). Until now, the MIM algorithm implemented in QTL Cartographer developed by Kao et al. (1999) and Zeng et al. (2000) has not been directly used in pines, but Bayesian approaches, also based on model selection, have been applied in these species (Hurme et al. 2000; Ball 2001).

The simultaneous analysis of multiple correlated traits has also been incorporated into algorithms in order to improve QTL detection efficiency (Jiang and Zeng 1995; Korol et al. 1995; Korol et al. 1998). Although these algorithms have not been specifically used up till now, there is much interest in such algorithms, especially for such traits as annual growth or wood density. Additionally, application of specific genotyping strategies has further maximized the efficiency of QTL mapping. One such strategy called selective genotyping has been extensively used in pines (Groover et al. 1994; Hurme et al. 2000; Kubisiak et al. 2000; Devey et al. 2004a). Another strategy termed bulk segregant analysis (BSA), commonly used to study qualitative traits, has also been applied to analyze quantitative traits (Emebiri et al. 1997).

Besides statistical procedures, it has been clearly shown that the power of QTL detection largely depends upon the quality of the phenotypic assessment. Poor phenotypic assessments, i.e., imprecise measurements, generally result in QTL with true effects being left undected, but in some circumstances (but to a lesser extent), might even lead to the detection

of false QTL. Without the possibility of developing F_3 populations or recombinant inbred lines to precisely estimate the value of the traits, clonally propagated material has become the material of choice for forest tree geneticists (e.g. Scottti-Saintagne et al. 2004). However, this type of material has only been rarely used in pines (Devey et al. 2004b). This is most likely due to the large captial investments and technical expertise needed to clonally propagate pines and the wide variability noted among specific genotypes in their ability to produce rooted cuttings.

3.2. QTL Discovery in Single Family Pedigree Designs

Twenty six QTL studies aimed at detecting associations between molecular markers and trait variation have been performed in seven pine species. However, most of the efforts have been concentrated in three species, *P. pinaster*, *P. taeda*, and *P. radiata* (Tables 4 and 5). In addition studies on inbreeding depression have been carried out in *P. radiata* and *P. taeda* (Remington and O'Malley 2000a, b; Williams et al. 2001; Kuang et al. 1999a, b).

3.2.1. Qualitative Traits

Associations between qualitative traits and molecular markers have led to the identification of markers linked to resistance to different rust diseases: pine needle gall midge in *P. thunbergii* (Hayashi et al. 2004), white pine blister rust in *P. lambertiana* Dougl (Devey et al. 1995; Harkins et al. 1998), and fusiform rust disease in *P. taeda* (Wilcox et al. 1996). A major gene controlling the biosynthesis of δ-3 carene has also been mapped in *P. pinaster* (Plomion et al. 1996b). Most of these studies have been based on intraspecific mapping populations. Additionally, two interspecific crosses were used to increase the level of polymorphism and the range of variation for branch architecture in the mapping population (Shepherd et al. 2002b).

3.2.2. Quantitative Traits

Growth has been the most studied trait in QTL mapping studies in pines. Nowadays, a comparable number of studies have been achieved for wood and end-use properties. These two classes of traits have been studied either globally (e.g. height growth, specific gravity), or after decomposition into simpler components (e.g. growth unit, ring density). Traits involved in adaptation to the environment have also been studied, including tree response to heavy metal, drought, cold stress, as well as bud phenology (Costa 1999; Hurme et al. 2000; Kubisiak et al. 2000; Lerceteau et al. 2000; Brendel et al. 2002; Yazdani et al. 2003).

3.2.3. QTL Results

The diversity of population types, population sizes, marker types, QTL detection methods, detection thresholds, and variation of phenotypic trait measurements have made the comparison between QTL experiments, a difficult task. However, general observations can be made regarding QTL number, position, phenotypic effect and stability.

QTL Number and Effect

Most QTLs have been detected at a single maturation stage and in a single environment. Therefore, although all the analyzed traits show continuous variation, suggesting polygenic control of the traits, only a limited number of QTLs per trait (Table 4: between 0 and 8) were detected. This number is in general smaller than in annual crops and has lead to a smaller proportion of the phenotypic variation being explained (Table 5). As stated in the Eucalyptus chapter: “the limited power to detect QTL in forest trees compared to crop species may be due to the high environmental and developemental variation in tree plantations, as well as to the small size of the analyzed populations” (mainly between 91-200 genotypes, Table 4).

However, QTL analyses carried out by Brown et al. (2003) and Devey et al. (2004a), using large mapping populations (> 400 individuals) do not support the latest hypothesis. Brown et al. (2003) reported QTL effects that were two- to three-fold smaller than those reported by Sewell et al. (2000, 2002) for the same traits. Such divergence likely represents more accurate estimates of the QTL effects owing to the larger segregating population analyzed. These results suggest that most QTL studies performed in pines, with the exception of the analyses carried out by Devey et al. (2004a, b) and Brown et al. (2003), have yielded an overestimation of the QTL effects, which was also suggested by Beavis et al. (1994). Consequently, as underlined by Wilcox et al. (1997) the genetic determinism of most target traits for pine breeding is likely to be explained by small effect genes, rather than any moderate-large effect genes. However, a larger number of QTL experiments with larger pedigree sizes will be required to validate this hypothesis.

Compared to the small number of QTLs detected at any one maturation stage, the analysis of QTL stability along a cambial age or seasonal gradient (Table 4) revealed a significant increase in the number of QTLs detected. For instance, the simultaneous analysis of different maturation stages combined with seasonal variation allowed for the detection of 7 to 23 QTLs for wood specific gravity, the percentage of latewood, and microfibrilar angle (Sewell et al. 2002). Similar results were reported by Pot (2004), who observed a total of 30 QTLs for wood density, 42 for wood heterogeneity, and 33 for radial growth, when measured over several cambial ages.

QTL Stability

In their review, Sewell and Neale (2000) pointed out that "before a commitment to marker-aided selection or breeding (MAS/MAB) can be made in tree breeding, QTL that have been detected must be verified in different experiments, as well as in different genetic and environmental backgrounds". Indeed, estimation of QTL stability (position and phenotypic variation explained) is one of the most critical factor if QTL analysis is to be performed for application in MAS. This question has been addressed at different levels, including the stability of QTL across different developmental stages (ontogenic or cambial age effect), time points along the growing season (seasonal effect), environments, and among diverse genetic backgrounds.

Ontogenic effect: Given the long-lived characteristic of forest trees, it has been questioned wether or not the same genomic regions would control quantitative traits (e.g. annual growth, density) as trees mature. Several experiments have been conducted to answer this question (Plomion et al. 1996a; Emebiri et al. 1997; Kaya et al. 1999; Sewell et al. 2000, 2002; Weng et al. 2002; Brown et al. 2003; Gwaze et al. 2003; Pot 2004). In general, a rather low QTL stability has been noted across different maturation stages for growth (Brown et al. 2003; Pot 2004). In only one case did the authors report on the detection of the same QTL regions at different maturation stages (Gwaze et al. 2003). Conversely, Kaya et al. (1999) did not find a single "common" QTL controlling growth rate through tree development. The same trend – low QTL stability across maturation stages – was observed for wood density, a trait that presents a higher juvenile-mature correlation than growth (e.g. Williams and Megraw 1993, Hannrup et al. 1998). This unexpected result suggests that some of the wood density QTLs probably reflect the genotypic response to annual climatic variation. Recently, Rozenberg et al. (2002) reported on the alteration of wood density profiles in response to drought. More interestingly they showed that the alteration of annual wood density profiles (the presence of a false late wood ring in the early wood zone) in response to drought was genetically controlled.

Seasonal effect: Sewell et al. (2000, 2002) and Pot (2004) analysed the seasonal stability of growth and wood quality QTLs for traits measured in spring (early wood) or summer (late wood). Overall, half of the detected QTLs were specific to one type of wood (early vs. late wood). These results agree with recent transcriptome studies that reveal that different sets of genes are regulated throughout the growing season (Le Provost et al. 2003; Egertsdotter et al. 2004).

Environmental effect: Although several studies have been performed in multi-site trials (Groover et al. 1994; Knott et al. 1997; Kaya et al. 1999; Sewell et al. 2000, 2002; Shepherd et al. 2002b; Weng et al. 2002; Brown et al. 2003; Devey et al. 2004a, b), in only one study did the authors analyze QTL stability at different sites (Groover et al. 1994).

Genetic effect: A complete understanding of the genetic variability of the traits of interest will rely on the analysis of multiple populations as all the major genes involved in the genetic control of a given trait are unlikely to be polymorphic in a single family. On a more practical side, QTL stability across different genetic backgrounds is a prerequesite to marker- assisted breeding (MAB) in multi-parental tree breeding programs. As underlined by Brown et al. (2003), MAB will reveal its full potential under two scenarios: i) if a genetic marker in full linkage disequilibrium with molecular polymorphism causing trait variation at the population level is discovered, or ii) if the gene (polymorphism) underlying the QTL is identified. Given their allogamous reproductive system and their recent domestication, pines are characterized by high levels of genetic diversity and low levels of linkage disequilibrium (reviewed in Gonzalez-Martinez et al. 2006). The combination of these two factors (high diversity and low linkage disequilibrium), together with their perennial characteristics (maturation, environmental heterogeneity), is likely to contribute to QTL instability across genotypes. Yet, few studies have addressed this important issue in pines. Kaya et al. (1999) did not find any QTL shared between pedigrees, while Brown et al. (2003) and Devey et al. (2004a, b) found only a small fraction of the detected QTLs to be common across different genetic backgrounds.

MultipleTrait QTL Analysis

As stated ealier, multiple-trait QTL analysis will likely become very important for breeding purposes since pine breeding is a multi-trait process. In several QTL experiments, more than one trait (e.g. growth, wood quality) has been studied. In most cases, co-localisations between QTLs for different traits have been observed which might be expected for highly correlated traits. Possibly more important, was the occurrence of multiple co-localizations in the genome suggesting the effect of pleiotropic genes rather than the existence of physically linked genes controlling different traits. It should be noted that QTL clusters were also observed for traits that were not phenotypically correlated (Brown et al. 2003; Pot et al. 2005b), suggesting strong environmental/developmental effects masking genetic correlation.

3.3. Future Direction on QTL Mapping

As underlined in the previous section, the most reliable QTLs - in a breeding perspective - are those that have been consistently detected at different developmental stages, environments, and in diverse genetic backgrounds. However, it is important to remind the reader that most of the studies performed so far in pines were based on single pedigree analysis, and that only a handfull of experiments have attempted to validate QTLs across different genetic backgrounds (Brown et al. 2003; Devey et al. 2004a, b). Considering that pine improvement involved the deployment of many families/clones, the genetic stability of marker-trait association is a pre-requisite before any extended use of molecular markers is considered in

operational breeding programs. There have been two major attempts to identify diagnostic markers for MAB either using neutral markers spanning the genome or selecting candidate genes based on their coincidence with QTLs.

3.3.1. Complex Designs for Detecting Marker-Trait Association: LD Mapping

As underlined in Guevara et al. (2005b), while QTL identification is based on physical linkage disequilibrium generated in one or a few generations of crossing, association mapping or LD mapping takes advantage of events that created association in the past to find statistical association between molecular markers and a phenotype on a much finer scale. As presented in section 1.6, five studies have provided valuable information regarding the extent of linkage disequilibrium in pines. In all cases, it was found that LD extended only over short distances. In *P. taeda*, Brown et al. (2004) and González-Martínez et al. (2006) revealed a rapid decay of LD within 800-2000 bp in candidate genes for wood quality and drought stress response. In *P. sylvestris* a rapid decay of LD was also detected by Dvornyk et al. (2003) and García-Gil et al. (2003). The genotyping of a breeding population of *P. radiata* with microsatellites (Kumar et al. 2004) also yielded the same result, i.e. no significant LD was observed between pairs of genetically linked markers, suggesting that LD decreases rapidly with physical distance.

LD mapping is only in its infancy in pines. To our knowledge only one study reported the use of molecular markers (SSRs) in unrelated trees (Kumar et al. 2004). Marker-trait associations were analyzed in a *P. radiata* trial consisting of 45 parents (40 males and 5 females). Parental trees were genotyped and association between parental genotypes and performance of 200 full-sib generated families were analysed according to the strategy used in larch and eucalyptus by Arcade et al. (1996) and Verhaegen et al. (1998), respectively. This analysis allowed the identification of several significant associations between markers and traits. However, it is important to note that none of the marker-trait associations found using full-sib family performance were identified when the parental general combining ability was regressed on the allelic frequencies of the marker. This result suggests that the first associations detected were probably biased due to population structure (only five females), as in larch and eucalyptus.

Pan-genomic LD mapping in pines will require an extremely high marker density, given the low extent of LD as estimated so far within the few genes analyzed. Alternatively, marker-trait association can be performed on selected candidate genes.

3.3.2. The Candidate-Gene Approach

Candidate genes (CGs) can be proposed based on the coincidence between QTLs and known functional genes putatively involved in the genetic control of the trait. Such positional CGs have been described in *P. taeda* (Brown et al. 2003) and *P. pinaster* (Chagné et al. 2003; Pot et al. 2005b) for wood quality related traits. In *P. taeda*, co-localizations between genes involved in mono-lignol biosynthesis (4CL, C4H, C3H and CcOAOMT) and QTLs for wood density were observed (Brown et al. 2003). In *P. pinaster*, a single candidate gene-QTL co-localization was found between KORRIGAN, a gene involved in the hemicellusose/cellulose biosynthesis, and QTLs for hemicellulose and fibre characteristics.

Other types of analysis can be used to select candidate genes related to plant adaptation. As underlined by González-Martínez et al. (2006) “Standard neutrality tests applied to DNA sequence variation data can be used to select candidate genes or amino acid sites that are putatively under selection for association mapping.”. Unusual patterns of nucleotide diversity and/or population differentiation have been detected in *P. taeda* (González-Martínez et al. 2006), *P. pinaster* and *P. radiata* (Pot et al. 2005a). Pot et al. (2005a) observed singular patterns of nucleotide diversity in three genes: a glycin rich protein homolog that was found to be up-regulated in late wood-forming tissue (Le Provost et al.

2003); CesA3, a cellulose synthase gene, and KORRIGAN, a membrane-bound endo-1,4-beta-glucanase involved in cellulose/cellulose biosynthesis. As mentioned before, co-localisation between KORRIGAN and wood quality QTLs was reported (Pot et al. 2005b). In *P. taeda*, although the action of neutral process can not be completely ruled out to explain the patterns of nucleotide diversity observed for CcoA-OMT1, several characteristics of its nucleotide diversity seems to indicate the action of natural selection on this gene. The colocalization of this gene with a QTL of water use efficiency (Brendel et al. 2002, Pot et al. 2005b), and the differential expression of this protein under different watering regimes (Costa et al. 1998), clearly emphasizes the putative role of this gene in wood trait variation. Association studies should now be used to validate this hypothesis.

4. Marker-Assisted Breeding

This section discusses the application of information from DNA polymorphisms in conifer-type tree improvement programs. Note here that such information includes not only purposefully designed markers, but also information obtained directly (re)sequencing, for which specific markers may not have been designed. Applications of information from DNA polymorphisms fall into four generic areas: audit and quality control, elucidation of genetic phenomena, population management, and selection and breeding. We discuss each of these aspects below, covering both existing and potential applications. Furthermore, we provide examples where appropriate, as well as comments regarding the current status of each of these applications.

4.1. Quality Control and Audit

Development of an array of DNA marker systems over the past two decades has provided tree breeding programs with a range of tools to achieve basic aspects of quality control that hitherto could only be addressed with difficulty, if at all. Seedlot and/or clonal fidelity have been the key concerns for both commercial and research applications. Indeed, this area was the earliest commercial application of DNA markers in tree breeding programs, and is still the most widespread, at least in coniferous species.

From a commercial perspective, the key objective is assurance of genetic gain by ensuring that seedlots and/or clones are true to intention. The applicability and efficacy of various DNA marker systems is dependent, therefore, on the way in which genetic gain is delivered. In conifers, such delivery can come in a number of different forms (and costs), even within the same breeding program.

4.1.1. Open-pollinated seed

Open-pollinated (OP) seed from a mixture of selected seed orchard material, generally consists of a few tens of maternal parents, with little if any control of the pollen source. This is a common means of seed production, particularly in commercial *Pinus* species in the southeastern United States. In such cases, there may be a need to ensure no contributions from unwanted maternal parents, which can be achieved by genotyping megagametophytes with sufficient markers, assuming maternal parentage information is maintained for all seed. Furthermore, markers could be used to quantify relative contributions of each maternal parent to individual seedlots, if seed counts are not available. A variation on this method of seed production is used in mass pollination techniques, such as supplemental mass pollination or liquid pollination. In these approaches, receptive conelets are pollinated using pollens from selected genotypes by various means, but the conelets themselves are not actually covered to prevent fertilization by unwanted pollens.

4.1.2. Control-pollinated seed

Control-pollinated (CP) seed, where pollens from selected parents are used to fertilize selected seed parents, and in the process, excluding pollen from other sources either using bags to cover receptive cones, and/or undertaking pollinations in contained greenhouse facilities. This is usually undertaken to produce full-sib families, or individual half-sib families with known selected pollen parents. Such methods are used in both breeding and commercial production for species such as *Pinus radiata*. Vegetative propagation, sometimes involving in vitro technologies, can be used to amplify genotypes particularly for commercial production, largely because of shortages of seed and/or cost of seed production. DNA markers – particularly codominant multiallelic marker systems such as microsatellites – have been developed for such purposes (e.g. Devey et al. 2003). In addition, paternal parentage can sometimes be evaluated via paternally inherited chloroplast markers. Such markers have been developed and applied in *P. radiata* for checking paternal inheritance (Kent and Richardson 1997), although judiciously chosen nuclear DNA markers may also suffice. In general, markers have shown misidentification is common in breeding programs; for example, Bell et al. (2004) estimated that 2.6% of parents were misidentified in a sample of an Australian breeding population of *P. radiata*, and that 8.4% of offspring of 10 families were not consistent with expected parentage. These results also indicate that a proportion of misidentification of open-pollinated seedlots is possible. Furthermore, our experience at Scion with putative full-sib families used for gene mapping experiments has revealed very few such families – produced either commercially or by research groups – are completely consistent with expected parentage (unpublished data). DNA markers, particularly codominant marker systems, have therefore been useful in checking and assuring parentage.

4.1.3. Clonal production

Clonal production is used either for experimental purposes (such as clonal tests) or for mass-propagation of tested clones for clonal forestry. In these cases, genotyping is undertaken to ensure that ramets do represent the desired genotype(s). Similarly, ramets deployed in seed orchards addresses are subjected to the same question, although sometimes misidentification of parents is detected via parentage testing of the seed obtained via methods described above. For applications where genotype fidelity is needed, 'profiling' marker systems such as RAPDs and AFLPs is useful, in that they are generally cheaper to both develop and use for this specific application, particularly as the high level of polymorphism revealed overcomes issues associated with dominance. However, in more recent years, fingerprinting kits have tended to utilise codominant markers (see references above), as these have been developed for other purposes (above) and are generally adequate for clonal fingerprinting, particularly if sufficient marker loci are used (e.g. Kirst et al. 2005b).

DNA markers have, therefore, been developed and utilised extensively in tree breeding programs, for both verifying commercially deployed materials and ensuring that experimental materials are true to the requirements. Even so, some programs still do not universally implement or rely on such genotyping, largely due to expenses involved.

4.2. Elucidation of Genetic Phenomena

All genetic gain ultimately depends on DNA polymorphisms. Knowledge on the nature and the effects of the polymorphisms has the potential to generate far more gain than the use of purely phenotypic data — on the parents and/or their progeny. Information from DNA polymorphisms enables the elucidation of phenomena, such as understanding of the genetic architecture of trait variation, revealation of population structure and history, and detecting fingerprints of selection, all of which can have direct or indirect applications in tree breeding.

The genetic architecture of trait variation can be defined as the frequencies, location, magnitude and mode(s) of action of quantitative trait loci/nucleotide (QTL/N) effects

underpinning quantitative traits. While QTL mapping has been very informative in this regard, the results are relevant only to the pedigree(s) used, rather than to whole populations. Association genetics may, therefore, be more relevant for understanding the genetic landscape of trait variation in forest trees.

4.2.1. Estimation of Genetic Variance Structures and Heritabilities

DNA markers offer new ways of obtaining some key knowledge which is fundamental to tree improvement; regarding genetic parameters, in particular genetic variances, heritabilities and genetic correlations among traits. Such knowledge informs the breeder not only about the feasibility of breeding for certain traits or combinations thereof, but also how it might be efficiently undertaken.

For obtaining genetic gain, a trait must be both variable and heritable. Genetic correlations *inter se* economic traits can be a major constraint if they are adverse, but they can provide the breeder with great opportunities, if they are favourable. Even genetic correlations between non-economic, indicator traits and economic ones can be used to a great advantage, if the correlations are strong and the indicator traits highly heritable. For genetic variances and heritabilities, if they have to be estimated from seed collected in natural stands, the coefficients of relationship within seed-parent families need to be known. Indeed, the relative contributions of inbreeding as such and finite effective numbers of [unrelated] pollen parents per seed parent represent important information. While isozymes were a valuable tool for providing such information, DNA markers can be much more powerful for the purpose, and with a greater range of applications. This issue of non-random components in the mating system can be important with native stands of conifers (Burdon et al. 1992), despite their wind pollination, even if it is less acute than in insect-pollinated species like eucalypts (Hodge et al. 1996). The wind-pollination of conifers tends to reduce the non-randomness of mating, while the conifers' mechanism of archegonial polyembryony probably reduces the inbreeding component still further.

For estimating genetic correlations between traits, knowledge of coefficients of relationship within such families is likely to be much less crucial.

As we will mention later, *a priori* family information as such may not be essential for this purpose, but can be supplemented with information from markers.

4.2.2. Population Structuring and History

The structuring and history of populations influence both the availability of genetic variation for the breeder to exploit and the potential for association genetics to contribute. While large, essentially panmictic populations cannot be expected to have appreciable across-family linkage disequilibrium (LD), cryptic structuring may exist which generates significant disequilibrium that could be useful for breeders. For example, localised population bottlenecks, followed by coalescences, could easily cause this. Such LD could provide valuable clues to 'metapopulation' history. Despite wind pollination, various factors can generate population structure in conifers (Mitton 1992). Interesting possibilities of structure exists in populations derived from recent admixture. In *P. radiata*, the exotic, domesticated 'land-races' still have large elements of the wild state. Interestingly, they evidently represent a genetically recent fusion of two of the native populations, Año Nuevo and Monterey (Burdon et al. 1998), which may provide a basis for some admixture disequilibrium.

Polymorphisms revealed by DNA sequence data derived from both genic and non-genic regions can reveal much about genetic history of those regions. Departures from Hardy-Weinberg equilibrium could reveal the presence of previously undetected genetic phenomena, such as presence/absence of inbreeding. Indeed, genetic variance (and gain) estimates are based on assumptions regarding relatedness of parents used in genetic tests. Such data can be

used to check these assumptions and provide empirical data for more accurate estimates. Similarly, sequence data from genic regions can reveal evidence of selection: for example balancing selection was detected by Cato et al. (2006b) in a gene associated with wood density and growth rate in *P. radiata*. Krutovsky and Neale (2005) found evidence for selection in three of 18 genes in douglas-fir. Such evidence – which can be generated on a relatively small subset of genotypes – could be an effective pre-screen for genes more likely to be associated with trait variations, although some caveats apply regarding power to detect effects of selection (Wright and Gaut 2005).

4.2.3. Assignment of Gene Function

Knowledge of gene function, if acquired, offers the greatest long-term opportunities to capture genetic gain, using either endogenous variation or genes introduced by conventional breeding or genetic engineering.

Linkage disequilibrium mapping and association genetics can provide clues in the search for QTN regarding which genes have functional roles in trait variation. While relatively short stretches of disequilibrium (usually hundreds to low thousands of base pairs) represent constraints for breeding applications, a key advantage is the potential for assignment of function – even identification of individual QTN. However, because of size of conifer genomes and lack of genomic sequence, identifying candidate genes will be crucial. Possible approaches are described in the following section, and in more detail by Wilcox et al. (2006).

In carefully selected cases, genetic transformation can be used to verify the role of a QTN in generating phenotypic variation. The costs of achieving a transformation, and often the regulatory issues, will demand highly selective application of this approach. On the other hand, knowledge of gene function may be useful for identifying genes to target for genetic transformation, to create new variation of use for breeders, and provide commercial cultivars. There are no reports yet of cloning QTL from tree species, partly due to the large number of candidates within QTL confidence intervals, but also because of the time required for trait expression of transformants arising from complementation studies. Nonetheless, association genetics will be a key step in increasing resolution, and in some cases, identifying putative QTN for further analyses.

4.3. Population Management

4.3.1. Pedigree Reconstruction and Detection of Genetic Contamination

Breeding populations represent the 'engine room' for capturing the additive gene effects that allow cumulative genetic gain over successive generations, through recurrent cyles of selection, intermating, evaluation, selection, and so on. Maintaining full pedigree has been favoured on the grounds that it helps preclude inbreeding and maintains effective population size. However, the expense of maintaining full pedigree can limit the size of the breeding population that can be handled, raising the question of whether larger populations can be handled if pair-crossing is not mandatory. Moreover, there may be situations where the breeder has reason to resort to material, e.g. commercial stands, in which pedigree has been sacrificed but which has the advantage of huge numbers (Burdon 1997).

Modern marker technology has major potential for pedigree reconstruction (Lambeth et al. 2001; Kumar et al. 2006), at least in open-pollinated families of known seed parentage. Complete reconstruction of predigree will be more challenging, especially with the wind pollination that characterises conifers. In maintaining gene resources that underpin breeding populations, controlled crossing is typically far too expensive, yet there may be a call for quantifying and even detecting individual cases of contamination (Burdon and Kumar 2003). Different types of markers may be required for this purpose, and the task may be challenging, but it appears inherently feasible.

Pedigree reconstruction is discussed later, in connection with selection.

4.3.2. Tracking and Maintenance of Genetic Diversity

A key element of population management is maintenance of genetic diversity, to give the breeder flexibility in both the short- and long-terms, and to safeguard continuing long-term genetic gain. While pedigree information is an indicator of genetic diversity, it does not provide definitive information in itself. Achieving that poses significant challenges.

The 'gold standard' for functional genetic diversity will usually be performance in well-designed and properly located common-garden genetic experiments. However, such experiments are expensive and often slow to deliver results. The use of DNA polymorphisms is clearly much quicker, but such marker diversity will need to be cross-referenced with the functional diversity, since the two classes of diversity are not necessarily closely coupled, at least among species (Morgante and Salamini 2003; Paran and Zamir 2003). Components of DNA diversity obviously include percentage of polymorphic genes (in either the coding or regulatory regions), percentage of base pairs that are polymorphic, particularly for QTN that must exist in coding or regulatory regions), and allele frequencies for the polymorphisms. Such information, in conjunction with knowledge of magnitudes of QTN effects, provides a benchmark for monitoring changes in diversity, for any forest trees.

Loss of low-frequency alleles is an obvious manifestation of a run-down of genetic diversity. Paradoxically, abrupt increases in the frequencies of such alleles, as can occur through genetic drift, can be a manifestation of the same phenomenon. While the inherently outbreeding genetic systems of forest trees may be able to cope with significant inbreeding in the wild, through selection for balanced heterozygotes, such mechanisms may be impeded under conditions of artificial breeding. In principle, almost any sort of genetic marker should be able to manifest the losses of alleles or sharp fluctuations in their frequencies. Nevertheless, it seems preferable to know what genes are of particular current or contingent importance, and monitor their frequencies. This, however, will depend on knowing the functions of the genes.

4.3.3. Provision for Biotic Crises

While it appears QTN typically exert minor individual effects in conifers, disease-resistance genes can represent a notable exception (Burdon 2001; Burdon and Wilcox 2006). Such genes can both have large effects (despite outward appearances of classical quantitative inheritance) (e.g. Kinloch et al. 1970; Wilcox et al. 1996), and be present at low frequencies (Burdon and Wilcox 2006). These genes can also have the feature of gene-for-gene specificities between host genotypes and fungal strains (pathotypes), which has important implications for ensuring durability of disease resistance against pathogen mutations (Burdon 2001). Establishing the nature of such genes can require carefully planned mating between parents, and inoculation studies based on single-spore isolates, preferably backed up with genomic studies. Here, as in other areas, comparative genomics can have a major role, at least in studying resistance genes of lower specificity.

The implications of such patterns of genetic variation within both hosts and pathogens, which are called pathosystems, can be major for both population management and selection. Disease resistance can be very important in plantation forestry, especially with exotics that are being grown in the absence of natural or other pathogens. Conserving low-frequency genes of large effect, which are sometimes important in conifers, can pose a three-fold challenge. A large population may be needed to find trees with such resistance. The requisite population size can be increased by the desirability of obtaining resistance genes in unrelated pedigrees. It may be further increased by the desirability of combining ('pyramiding') genes in the genetic material that represent diverse mechanisms of resistance to the pathogen, towards

ensuring durability of resistance against mutation or genetic shifts in the pathogen population. In implementing this approach, identification of resistance genes will be of enormous help.

The information on forest-tree pathosystems, in general, is still very sketchy, although there are resistance genes identified for some co-evolved forest pathosystems (e.g. Wilcox et al. 1996). Nevertheless, what is known indicates that preparations for biotic crises in the form of new fungal diseases should include having very large population resources available. In at least some cases, it is very doubtful whether breeding programs effectively contain such provision (Burdon and Gea 2006).

4.4. Selection and Breeding

4.4.1. Pedigree Reconstruction as an Alternative to Maintaining Pedigree Records

A proposed application (Lambeth et al. 2001) is to use markers in genetic tests to reconstruct pedigrees retrospectively, as opposed to maintaining pedigree information throughout the life of a genetic test. This involves applying pollen mixes of known composition to a range of seed parents, planting the resultant offspring in designed experiments with limited or no maintenance of family information (Kumar et al. 2006). Upon subsequent measurement, pedigree identity is ascertained via DNA markers. Benefits of such an approach include logistical simplicity, cost reductions for breeding and testing, potentially better estimates of genetic parameters and increased gains, and assurance of parentage (see above). Furthermore, less financial costs are incurred if experiments are abandoned or lost prior to remeasurement – which does happen in tree improvement programs. Disadvantages include under-representation of some specific families due to either chance or factors affecting fertilization, prevalence of inbreeding in some families distorting breeding-value estimates of those families, difficulties in detecting full-sib specific combining ability in full-sib families, and genotyping costs. Most of these disadvantages could be overcome, although further research is needed to fully evaluate various possible application scenarios. To date, this approach is currently being evaluated in the context of operational breeding in New Zealand with *Eucalytpus* and *Pinus radiata*.

4.4.2. Estimation of Heritability based on Marker-Ascertained Relatedness

Because molecular markers have the potential to estimate relatedness between genotypes, methods to estimate heritability of phenotypic characteristics have been proposed and evaluated (Andrew et al. 2005; Kumar and Richardson 2005) without need to have any prior knowledge of genetic relationships (see above). Such methods have the potential benefit of obtaining information from existing forests (e.g. natural forests or commercial plantations) thereby obviating the need for genetic testing involving progenies, thus speeding up the generation of information, particularly for species where little *a priori* information is available. Some disadvantages involve the reliability of marker-based estimates, particularly as assumptions need to be either made or checked regarding genetic structure and prior levels of relatedness, as well as reliability of phenotypic information, particularly if sourced from forests rather than purpose-designed common-garden tests with appropriate controls. These approaches have not been extensively investigated but may have a role. In part, this may be due to the fact that existing approaches are well established and successful, as well as the existence of infrastructure such as extensive testing involving progenies. Kumar and Richardson (2005) compared phenotype- with marker-based heritability estimates for wood density in *P. radiata* and found very little difference. Similarly, Andrew et al. (2005) reported non-zero estimates of heritability for a number of foliar defence-related chemicals in *Eucalyptus melliodora*, some of which were consistent with independent estimates based on phenotype (see Andrew et al. 2005 and references therein). Such approaches may be of use

also for association genetics, particularly in revealing cryptic population structuring in natural forests, as well as for selecting maximally unrelated individuals for estimating LD.

4.4.3. Within-family Selection Based on Marker-trait Associations Derived from Pedigreed QTL Mapping Populations

Such selection is based upon selecting individuals within known full- or half-sib families where marker-trait associations have been previously determined. Genotypes with the desired multi-locus genotypic composition are selected from within families. In theory, this can be applied within families in breeding populations and/or production populations. Generic benefits include earlier selection, increased selection intensity and potentially cheaper selection (Stromberg et al. 1994). In conifer improvement, breeding populations are usually based on a few hundreds of parent genotypes, thus it is largely impractical and very expensive to detect associations for each genotype independently (Strauss et al. 1992; Johnson et al. 2000). Rather, application of markers is more likely to be restricted to populations where specific families are advanced, as for example, in elite populations, or in small nucleus- or mainline breeding populations. A number of situations have been evaluated either for species-specific scenarios, or from a wider theoretical perspective. One of the earliest such studies was that by Strauss et al. (1992) who concluded that within-family marker-assisted selection (MAS) was of limited or no value unless a high proportion of additive genetic variation could be explained by markers for traits of low heritability but high value, and where selection intensities within families were high compared to among families. They also concluded that difficult- or expensive-to-measure traits such as wood quality or resistance to certain diseases showed the most potential for MAS. Williams and Neale (1992) evaluated the relative efficiency (RE) of MAS and also concluded RE was greatest where (a) traits were not expressed sufficiently early to enable a reduction in generation length, as well as (b) lower-heritability traits where markers explain a high proportion of additive genetic variability for the traits. However, results from many QTL mapping experiments have indicated the genetic architecture of quantitatively inherited traits is dominated by genes of small effect (Wilcox et al. 1997; Sewell and Neale 2000; Brown et al. 2003; Devey et al. 2004a, b), thus limiting the opportunity to explain sufficient variation with markers. Wu (2002) evaluated the trade-off between proportion of variance explained by markers and shortening of selection interval. For traits of moderate heritability (0.2-0.4) marker-assisted early selection (MAES) was more efficient than phenotypic selection, if they explained greater than 5-10% of additive genetic variance and allowed selection interval to be reduced by half. For later-expressed traits, even less variance explained by markers was necessary for markers to be more efficient than phenotypic selection. Wu (2002) also evaluated a number of other scenarios, including combining phenotypic selection with markers, and found that MAES is only marginally more efficient than phenotypic selection except in cases of relatively low juvenile-mature phenotypic correlations. Given genetic architectures of QTL explaining a few percent each (Wilcox et al. 1997; Brown et al. 2003; Devey et al. 2004a; b), and limitations for selection of up to 10 unlinked markers per family linked to such QTL, the effectiveness of early selection is limited for traits with low-moderate age-age phenotypic correlations, particularly once costs of QTL detection are taken into account. Time savings for traits of delayed expression or low juvenile-mature correlation will, however, depend on QTL/phenotype associations having already been established.

All the above studies investigating application of markers for breeding population advancement have generically shown that while there are scenarios where markers could be used to generate genetic gains, the financial gains could be quite limited. This is in part due to cost of genotyping, as well as the length of time, additional costs need to be compounded. A further limitation stems from the fact that most breeding objectives involve multiple

quantitatively inherited traits whose genetic architectures are likely to involve predominantly small-effect genes (see references above). This means limited genetic gains from individual markers. Furthermore, relatively few markers can be used to select within any particular family: for a simple 2-genotype-per-marker scenario, it would be necessary to generate and genotype thousands of offspring per family to have sufficient power to generate the optimal 10-marker-locus genotype. If such selection were restricted to a limited range of traits, there would also be the need to augment marker information with phenotypic records to effectively address the breeding objective, therefore the opportunity for earlier selection could be lost. On the other hand, few studies have investigated the efficacy of marker-based or marker-assisted selection over multiple generations. Application over generations is likely to reduce costs, although not in a manner linearly proportional to gains: This is owing to the fact that additional revenues in future generations are discounted relative to cost, as costs are effectively incurred early (e.g. Johnson et al. 2000).

An exception to the limitations outlined above is selection for genes of large effect such as major-gene disease resistance (see references above). In breeding populations, frequencies of resistance genes could be increased much faster via marker-based selection than phenotypic selection; markers could obviate most of the need for field and/or greenhouse screening, by allowing only those individuals to be deployed in field tests (for screening for other traits) that have the favourable marker phenotypes. Where gene pyramiding is crucial for durable resistance this marker-based approach could be especially valuable. As with all marker-assisted (or based) selection scenarios, such genes need to be detected, which incurs costs early in the breeding cycle. However, such costs are likely to be less in that such large-effect genes are less financially burdensome to detect, and in cases where resistance is valuable, financial gains can be considerable, particularly if undertaking development disease-resistant (or tolerant) breeds.

Opportunities for MAS have also been evaluated for production populations in species such as douglas-fir (*Pseudotsuga menziesii)* (Johnson et al. 2000) and *Pinus radiata* (Wilcox et al. 2001b), and for clonal deployment in *P. radiata* (Kumar and Garrick 2001). Results in general indicate that gain from MAS is possible in *P. radiata* for a range of options, but marginal for douglas-fir, implying rotation length and product value are both important. For example, Wilcox et al. (2001) showed that even modest gains in physical traits of 3.0-3.4% resulted in product value gains in excess of 9% and internal rates of return ranging from 9.1 to 21%. However, a key condition here was the need to achieve adequate multiplication rates via vegetative propagation. Propagation technologies are key to MAS being cost-effective. Johnson et al. (2000) showed that modest genetic and financial gains were possible for MAS in douglas-fir for production population applications, although results were highly dependent upon assumed genetic architectures, and relatively large areas of plantations were needed to justify investment in extra costs associated with MAS.

Given all of the above, it is not surprising that relatively few breeding programs are actively pursuing MAS, particularly for quantitatively inherited characteristics. Some private companies in the US and New Zealand have undertaken or are still undertaking MAS on a limited scale, but there is yet to be widespread uptake either for breeding-population advancement, or as a tool for more immediate genetic gains. Nonetheless there is still some potential, as some of the above studies have indicated, as well as other possible areas of application yet to be explored. It may be however that other technologies (e.g. GAS, below, or genetic modification) may supersede MAS.

4.4.4. Combined Among- and Within-family Selection Based Upon Results from Association Genetics Experiments

Advances in genomics technologies over the past 5-10 years have made possible almost unrestricted access to any region of tree genomes, particularly sequences within and associated with expressed genes. Variation in genic regions and the associated regulatory regions form the basis of phenotypic variation. LD mapping and association genetics are key tools in correlating such sequence variation with observed trait variation. Unlike marker-trait associations from pedigreed QTL detection populations (i.e., families), associations derived from association genetics have applicability for both within- and among-family selection. The term gene-assisted selection (GAS) has been used to describe this method of selection (Wilcox et al. 2003, 2006**;** Wilcox and Burdon 2006), and differentiates the manner in which markers are found (i.e., via association genetics) and utilised, from MAS (above).

Because association genetics is relatively new to forest trees (Neale and Savolainen 2004; Wilcox and Burdon 2006; Wilcox et al. 2006) – and many other plant species (Flint-Garcia et al. 2003) – very little quantitative analyses have been undertaken to date that involving detailed specific strategies for incorporation into tree breeding programs. Wilcox et al. (2006) described a range of applications within tree breeding programs, which are summarised below.

Conifer breeding programs can be generically characterised as consisting of highly heterozygous genotypes in hierarchically-arranged (and managed) populations, with multi-trait breeding objectives. To be effective, marker-trait associations derived from association genetics must integrate within this scheme, as with MAS. A key generic benefit of GAS in this regard is application to *both* within- and among-family selection, in contrast to MAS, which is restricted to those families in which the associations have been detected. Association genetics, therefore, has application to all levels of population hierarchies, from essentially unimproved germplasm, through to advanced-generation lines developed for clonal deployment.

A further benefit would be selection at the seedling stage, much sooner than full trait expression – as with MAS. For most commercially important pine species, phenotypic selection is applied at 6-12 years of age, with the onset of reproductive competency occurring slightly earlier. Opportunity exists for early selection to either increase selection intensities via a multi-stage approach as with MAS (Wu 2002), or undertake an additional round of selection prior to finalising genotypes for deployment (Wilcox et al. 2001b). Where onset of reproductive maturity precedes trait expression, there is also opportunity to reduce breeding-cycle length, although this would require known marker-trait associations for at least the majority of traits in the breeding objective(s). The opportunity for selection among- as well as within families provides even greater opportunity to shorten breeding cycles compared to MAS.

An additional benefit is cost reduction relative to phenotypic selection. Once marker-trait associations are detected, genotyping costs are generally somewhat less as compared to phenotypic evaluation. Such evaluations, however, are considerably more expensive than obtaining sufficient markers for genome coverage for linkage and QTL mapping, and MAS purposes, as virtually all polymorphisms within and associated with expressed genes need to be detected – which, on a whole-transcriptome basis, entails major effort. Such information nonetheless has other potential applications (e.g. pre-screening genes for genetic modification experiments), so the costs could be spread over several funding sources and applications.

Because of the breadth of potential selection applications within any one breeding program, as well as the fact that conifer breeding programs all differ in some regard from one another, the application of association genetics will need case-by-case evaluation – as has been the case with MAS - ultimately needing numerical simulation to evaluate different

scenarios. However, although a number of breeding programs are involved in some of the basic research required for detecting marker-trait associations, as yet there are no published reports of evaluations of strategies that incorporate association genetics within conifer breeding programs. We expect this will change as results from association genetics and linkage disequilibrium experiments become available.

Indeed, the outlook from results to date is cautiously favourable, as marker-trait associations have been reported for a number of tree species, including *P. taeda* (Brown et al. 2004), *P. radiata* (Cato et al. 2006b), and *Eucalyptus nitens* (Thumma et al. 2005), - although all of these need to be independently verified. Results from a range of other plant species are also emerging (see Gupta et al. 2005, for a review), also with encouraging messages.

Given the size of gymnosperm genomes, and the lack of widespread genomic DNA sequence for most hardwood species, it appears that sequences within and associated with expressed genes are likely to be investigated. But how should such sequence be chosen? Possible approaches are:

- Identifying gene sequences using EST libraries from both angiosperms and conifers, and complete genomic sequences from model angiosperm species to select numerous genes for high-throughput (re)sequencing. This approach is being used in *P. taeda* (see http://dendrome.ucdavis.edu/adept2).
- Expression information at the mRNA level — via *in silico* expression profiling, or application of a number of gene expression profiling technologies. Protein expression information is also available for some species.
- Selecting genes on the basis of biochemical role(s) and/or known sequence motifs.
- Information from co-localisation of expressed genes and QTL from pedigreed QTL mapping populations, within either the species of interest or related species.
- Combinations of the above, for example gene- or protein-expression information combined with QTL mapping (Kirst et al. 2004).
- Genes shown to be involved in trait variation via genetic modification experiments.

Technical advances in 'omics' technologies may also offer a number of other approaches.

The undoubted importance of comparative genomics, especially for establishing gene function, means that genomic information is needed on other species, in addition to the classical model species like *Arabidopsis*. This need is actually being met, to some degree at least. In conifers, parallel research is proceeding in *P. radiata* in Australasia, *P. taeda* in southeastern USA, and *P. pinaster* in Europe. The evidently close synteny among these species promises great opportunities to apply genomic information on any one species to the others. Using these as model species within conifers should allow one to capitalise on the most favourable model features of all the species at once. Similarly, whole-genome sequences of *Populus* and *Eucalyptus* will allow for even more rapid advances in these species.

Should association genetics be utilised by tree breeders, then a number of other requirements need to be addressed. These include:

- populations and analytical methods for adequate detection of marker-trait association (see Ball 2005),
- suitable laboratory and bioinformatics capabilities functionally integrated into tree-improvement programs, and
- development of strategic alliances among appropriate entities to effectively resource association genetics programs in a manner that allows operational implementation.

The above issues are discussed in more detail by Wilcox et al. (2006).

4.5. Summary of Applications of DNA Polymorphism in Conifer Breeding

Information from DNA polymorphisms has a wide range of applications for tree improvement, including quality control, pedigree reconstruction, elucidation of genetic phenomenon, monitoring and maintenance of genetic diversity, and selection and breeding based upon polymorphisms associated with trait variation. To date, only some of these applications have been implemented in operational breeding programs – largely those associated with quality control. Some potential applications – such as association genetics and pedigree reconstruction as part of operational testing, are largely in the research and development phase, while others such as within-family MAS have been sufficiently evaluated to identify at least some strategies for implementation. The lack of uptake across the spectrum of potential applications is likely due to cost, which is particularly important in tree improvement programs which usually take years to recoup such costs. Nonetheless, technological advances will increase the scope of applications for tree improvement.

5. Genomics Resources For The Genus *Pinus*

5.1. Efforts Towards Complete Genome Sequencing

5.1.1. BAC Libraries

Large-insert genomic DNA libraries in which each DNA clone is stored and archived individually (i.e., ordered libraries) are a fundamental tool in modern genomics research. The most popular large-insert vector is the bacterial artificial chromosome (BAC), which, despite its name, is not a chromosome but rather a modified bacterial plasmid (F-factor) that can stably carry large inserts (ca. 50-400 kb) (Zhang and Wu 2001). BAC libraries have been made for a host of taxa and employed in a variety of applications (reviewed by Zhang and Wu 2001).

Despite the importance of pine and other conifers and the value of BAC libraries to genomic research, BAC resources are extremely limited for gymnosperms. To our knowledge, there are three gymnosperm BAC libraries in existence - all of these are for pine. The first pine BAC library, constructed by Islam-Faridi et al. (1998), affords 0.05× coverage of the *Pinus taeda* genome. A second library developed by Claros et al. (2004) for *P. pinaster* provides considerably higher genome coverage (0.32×), although the probability of finding a gene of interest is still relatively low (27%). However, the *P. pinaster* library is reportedly being expanded to 3× (Claros et al. 2004) which would provide 95% probability of finding a given genome sequence (see Figure 3). In September 2004, the US National Science Foundation funded a three-year project that included construction of a 10× BAC library from the *P. taeda* genotype "7-56" (Peterson et al. 2005) as a primary objective (Figure 4). When completed, this library will be, to our knowledge, the single largest BAC library ever made with roughly 1.7 – 2 million individual clones. As of this writing, the 7-56 BAC library affords 2× coverage of the pine genome. When the 7-56 library reaches 3× coverage (ca. June 2006) it will be gridded onto BAC macroarrays and screened with STS and molecular genetic markers including those associated with genes of economic importance. Resulting information will be utilized to isolate and sequence intact pine genes. Additionally, marker localization will represent the start of STS-based physical mapping in pine. Comparison between the *P. pinaster* and *P. taeda* libraries should provide considerable insight into genome evolution within *Pinus*.

5.1.2. Need for Pine Genomic Sequence

Although EST sequencing in pine is relatively advanced (see section 5.2), there is little genomic sequence data available for pine or any other gymnosperm. As of March 6, 2006, the longest continuous gymnosperm/conifer genomic sequence in GenBank was only 6884 bp. While ESTs provide the coding regions of many expressed genes, an understanding of gene function and regulation requires knowledge of those non-coding sequences that coordinate

gene expression in response to biotic and abiotic cues (e.g. promoters, enhancers, silencers). Such information necessarily comes from sequencing of genomic DNA.

As the pine genome is 10-times larger than that of maize (the largest plant genome to be the target of a full-scale genome sequencing effort; NSF 2005), it may be a while before whole-genome sequencing of pine becomes a reality. Nonetheless, pine genomics can be greatly advanced by even relatively modest sequencing of genomic DNA, especially sequencing of large continuous pieces of DNA such as BAC inserts. The following are just a few areas of pine genome research that will benefit from more extensive sequencing of genomic DNA:

- *Genome structure*: Very little is known about the structure of any conifer genome, and relationships between genes, repeats, and pseudogenes have only been tangentially explored (e.g. Elsik and Williams 2000, Rabinowicz et al. 2005, Peterson et al. 2006). BAC and shotgun sequencing of pine DNA will provide a more detailed understanding of the sequence structure of the pine genome. Such information will be essential in the development of an efficient strategy for sequencing the *gene space* of pine.
- *Differential gene expression*: The lack of genomic sequence for pine has severely limited study of the non-coding regulatory regions of its genes. Information on these regions is essential if we are to understand differential gene expression in pine and manipulate conifer genes in a useful and controllable manner (No et al. 2000). To determine the general tissue/development/environment specificity of the regulatory sequences for a given gene, the expression profiles of that gene can be examined – since there is already a considerable amount of expression profile data available (e.g. Egertsdotter et al. 2004; Lorenz et al. 2006), associating regulatory sequences with specific tissues and/or developmental events could be initiated immediately after sequencing. Comparison of regulatory sequences from genes with similar expression patterns can be used to gain insight into the activation/repression mechanisms underlying major developmental events.
- *Comparative genomics of conifers*: Pine BAC/shotgun sequences and comparative genomic approaches can be utilized to advance study of other gymnosperm genomes, which will afford considerable insight into the evolution of this important group of organisms. Comparison of orthologous regions of conifer genomes (e.g. through sequencing of BACs and/or BAC contigs recognized by common markers) will provide high resolution means of investigating conifer sequence evolution (see Stirling et al. 2003 and Yan et al. 2004 for angiosperm examples).
- *Gymnosperm/angiosperm comparisons*: Information garnered from pine genomic sequence should provide nearly limitless opportunities to explore comparative genome evolution between gymnosperms and angiosperms. BAC sequences from pines can be compared with existing plant genome sequences as a mean of evaluating macro- and microsynteny as well as sequence conservation/divergence between pines and angiosperms. As pine has been widely used as an outgroup in angiosperm comparative research, pine genomic sequences should enable investigations by many angiosperm research groups throughout the world.
- *Polymorphism discovery and characterization*: EST sequencing has afforded tremendous insight into polymorphisms within the coding sequences of pine genes (e.g. Krutovsky et al. 2004; Le Dantec et al. 2004). However, sequence changes in the non-coding regulatory regions of pine genes may provide as much, if not more, information on phenotypic diversity than coding sequence.
- *Association mapping*: Relatively long DNA stretches produced by BAC sequencing will facilitate genome-wide studies of linkage disequilibrium, which, in turn, will fuel association mapping projects (e.g. Brown et al. 2004; Krutovsky and Neale 2005).

- *Sequencing adaptive and economic genes of pine*: There is considerable interest in sequencing intact genes of economic importance and/or particular value to conifer/gymnosperm research. Potential target genes include those associated with outcrossing (Williams et al. 2001), embryogenesis (Ciavatta et al. 2001), disease resistance/susceptibility (e.g. Devey et al. 1995; Morse et al. 2004; Kayihan et al. 2005), abiotic stress resistance (e.g. Chang et al. 1996 ; Dubos and Plomion 2003 ; Dubos et al. 2003 ; Krutovsky and Neale 2005 ; Lorenz et al. 2006), and wood properties (Brown et al. 2003 ; Devey et al. 2004a ; Krutovsky and Neale 2005 ; Pot et al. 2005b). Complete gene sequences can be utilized to explore pine/conifer evolution and facilitate pine improvement through traditional breeding, marker-aided selection, and genetic engineering. Of note, a BAC containing a target gene may well contain other genes of interest. Study of these "bonus" genes will likely be as important as studying target genes themselves.
- *Exploring the evolution of repeat sequences*: Recently a 454 Life Sciences Genome Sequencer 20 (Margulies et al. 2005) was used to generate 100 Mb of sequence from the *P. taeda* genotype 7-56 (Peterson et al. 2006). This sequence should provide considerable insight into pine genome structure although the short length of 454 reads (ca. 100 bp each) limits the utility of the sequence. While pseudomolecule contigs have been assembled, 100 Mb affords only 0.046× coverage of the pine genome and consequently it is likely that any 454 contig of significant length is a conglomeration of sequences in a repeat sequence family rather than an individual element found in native DNA. However, repeat sequences obtained from sequenced BACs and/or contigs assembled from Sanger/capillary shotgun reads can presumably be used as scaffolds on which 454 reads can be aligned. The depth and shape of the resulting alignments should permit detailed analysis of the evolution of major repeat-sequence families in pine.
- *Genetic engineering*: Regulatory sequences obtained from genomic sequences can be attached to reporter genes and used to study promoter/regulator specificities (e.g. No et al. 2000). Such testing will provide insight into pine gene regulation, which will eventually enable insertion of constructs containing pine promoters and genes/alleles of value into conifers. In addition, the genomic sequences can serve as sources of molecular markers that can be utilized in tree improvement through marker-aided selection.

5.1.3. Physical Mapping

Given the enormous size of the pine genome and the time and financial resources required to conduct physical mapping research, it is likely that complete physical mapping of the pine genome will be too costly to pursue, at least with present technology. Because DNA sequencing technologies are advancing rapidly (e.g. Margulies et al. 2005), it is probable that complete shotgun sequencing of the pine genome will become affordable before complete physical mapping does. However, physical mapping of gene-rich genomic regions has been utilized to advance understanding of important chromosomal regions in many species (e.g. Folkertsma et al. 1999; Sanchez et al. 1999; Dilbirligi et al. 2004; Barker et al. 2005), and indeed, physical mapping of gene-rich regions is a goal of the research group constructing the *P. taeda* 7-56 BAC library.

5.1.4. Sequencing the Pine Genome

C. Plomion: Will the pine genome be sequenced?
D.G. Peterson: Yes.
C. Plomion: When?
D.G. Peterson: Not this week.

The complete sequencing and assembly of the *Arabidopsis* and rice genomes have afforded considerable insight into the evolution of higher plants including pines. For example, Kirst et al. (2003) compared the genome sequence of *Arabidopsis* with pine ESTs and showed that gymnosperms and angiosperms possess highly similar gene complements. Other plant genomes, most notably poplar, sorghum (*Sorghum bicolor* L.), and maize (*Zea mays* L.), are current subjects of full-scale genome sequence efforts, and the genome sequences of these plants will further facilitate understanding of plant genome evolution and function (Paterson et al. 2005). The poplar sequence, in particular, may provide information that will help advance understanding of wood formation in trees (Tuskan et al. 2004). However, angiosperms and gymnosperms diverged from a common ancestor more than 300 million years ago (Bowe et al. 2000), and consequently the utility of angiosperm sequences in the study of pine and other gymnosperms will be limited. Ultimately, the best means of advancing pine and conifer genomics is complete sequencing of a conifer genome. While it is now theoretically feasible to sequence the genome of any organism, the large, repetitive nature of conifer genomes will likely prevent them from being targets of full-scale genome sequencing for at least a few years. However, the likelihood that the pine genome will be sequenced is high based upon past and present investments made by the NSF (including construction of the *P. taeda* 7-56 BAC library) and other granting agencies and growing worldwide interest in pine as a biofuel/carbon sequestration crop (Jackson and Schlesinger 2004; Perlack et al. 2005).

While whole-genome sequencing may have to wait, sequencing of gene-rich BACs is likely to begin in relatively short order. Additionally, reduced-representation sequencing (RRS) methods (see Peterson 2005 for review) are being used to investigate sequence subsets of the pine genome, and are likely to play a role in eventual whole-genome sequencing as well. Two RRS techniques that have been successfully utilized in maize genome exploration are methylation filtration (MF) and Cot filtration (CF)[1] (Whitelaw et al. 2003; Springer et al. 2004).

- MF is based on the preferential hypermethylation of retroelements and hypomethylation of genes observed in some plants (Rabinowicz et al. 1999). In MF, genomic DNA fragments are ligated into a plasmid containing an antibiotic-resistance gene and the resulting recombinant molecules are used to transform a bacterial strain containing enzymes that preferentially cleave hypermethylated DNA. Linearization of a hypermethylated recombinant molecule results in its loss and makes its host cell susceptible to an antibiotic in the growth medium. Consequently, only hypomethylated sequences are successfully cloned. While a useful gene-enrichment tool in maize and sorghum (Rabinowicz et al. 1999; Bedell et al. 2005), MF has proven considerably less effective when applied to the "mega-genomes" of pine and wheat (Rabinowicz et al. 2005).
- CF is rooted in the principles of Cot analysis, the study of DNA reassociation in solution (Peterson et al. 2002). When sheared DNA is heated to 100ºC, the two complementary strands of each double helix come apart in a process known as denaturation. If the denatured DNA is slowly cooled, complementary DNA strands find each other and form double helices (duplexes). The rate at which a particular DNA sequence finds a complementary sequence with which to pair is proportional to the number of times that a sequence (and hence its complementary sequence) is found within the genome. In other

[1] Cot filtration was originally called "Cot-Based Cloning and Sequencing" (CBCS; Peterson et al. 2002) and later "high Cot" (HC; Yuan et al. 2003) sequencing . However, the term "Cot filtration" and the acronym CF are becoming more widely used because (a) the acronym "CBCS" can also stand for "clone-by-clone sequencing," (b) "high Cot" is frequently confused with "high copy," and (c) Cot components may be sequenced without prior cloning (Peterson et al. 2006).

words, repetitive sequences renature more quickly than low-copy sequences. In CF, sheared genomic DNA is denatured and allowed to reassociate for a period of time in which only repetitive DNA sequences are likely to form duplexes. The double-stranded repetitive DNA is then separated from single-stranded low-copy DNA using hydroxyapatite chromatography. The low-copy DNA and/or the repetitive DNA can then be cloned and sequenced (Peterson et al. 2002; Yuan et al. 2003; Lamoureux et al. 2005) or used directly as a pyrosequencing substrate. With regard to the latter option, Peterson et al. (2006) recently sequenced Cot-filtered *P. taeda* 7-56 DNA including isolated highly repetitive, moderately repetitive, and low-copy sequences using the new 454 Life Sciences Genome Sequencer 20 (see Margulies et al. 2005 for details on this instrument). These data are currently being evaluated, although initial results look promising. CF has proven highly effective in wheat where it provides 19.5-fold enrichment for low-copy DNA (Lamoureux et al. 2005).

5.2. Genomic Tools to Identify Genes of Economical and Ecological Interest

5.2.1. Characterisation of the Pine Gene-Space

A large scale Expressed Sequence Tag (EST) sequencing project can generate partial sequence for a large proportion of genes from a given organism (Adams et al. 1991). As only genic sequences are analyzed regardless of genome size, this approach is considered as the most cost-efficient strategy of gene discovery for organisms whose genome sequences are not yet available (Rudd 2003), especially for species such as pines with exceedingly large genome primarily composed of non-genic repetitive elements. For this reason, such projects have been initiated for many organisms including the numerous agronomically important plant species (Rounsley et al. 1998). To rapidly scan for all the protein coding genes and to provide a sequence tag for each gene of the pine genome, large collections of expressed sequence tags (ESTs) has been developed in *P. taeda* and to a lesser extend in *P. pinaster*.

Loblolly pine ESTs

Three major EST sequencing projects, mainly funded by the NSF have been conducted so far in *P. taeda*. A first project "Genomics of wood formation in Loblolly pine", was concerned with gene discovery in wood formation (http://pinetree.ccgb.umn.edu/). It resulted in about 60,000 ESTs placed in the public domain. A second project entitled "Transcriptome responses to environmental conditions in Loblolly pine roots" aimed at completing this information by targeting especially roots undergoing a variety of biotic and abiotic stresses (http://fungen.org/Projects/Pine/Pine.htm). It resulted in 140,000 additional ESTs. A third project "Genomics of Loblolly pine embryogenesis" seek to identify and characterize an additional 85,000 ESTs derived somatic and zygotic embryos (http://www.tigr.org/tdb/e2k1/pine/index.shtml).

Maritime pine ESTs

Early molecular biology studies paved the way for the construction of cDNA libraries from different woody tissues (Cantón et al. 1993; García-Gutiérrez et al. 1995), roots (Dubos and Plomion 2003) and buds (Collada, Guevara and Cervera, unpublished results) of maritime pine. Funded by the European Union (GEMINI project), INRA (lignome project), and CNS (ForEST project), 28,000 high-quality sequences have been generated in about equal number from wood-forming tissues, roots and buds. These ESTs have been made available at the EMBL database and at the following URL http://cbi.labri.fr/outils/SAM/COMPLETE/. The *P. pinaster* cDNA clones are available for the scientific community, at Platform for Integrated Clone Management PICME (http://www.picme.at/). A web-based environment for assembly

and annotation of collections of ESTs (Le Provost et al. in prep) was used to analyse the *P. pinaster* ESTs.

A comprehensive unigene assembled for the user community
The number of pine ESTs in public databases has increased dramatically during the past decade (Figure 5), especially for *Pinus taeda*, ranking at the 19th position in dbEST with 329,469 ESTs, and to a lesser extent for *P. pinaster* (about 28,000 ESTs). The other pine species, *P. elliottii, P. banksiana*, *P. patula* and *P. sylvestris* represent less than 2000 sequences. In 2005, the *Pinus* genus was introduced in the TIGR gene index (Quackenbush et al. 2001) where a comprehensive collection of all publically available pine ESTs has been clustered into a tentative unigene set of 45,500 genes (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=pinus), representing probably a major fraction of the coding genome of pine. This resource has laid the foundation for future identification, characterization and cloning of economically important genes in pines, and should provide breeders with a wealth of information to significantly enhance and speed up the breeding process.

Besides this resource, a significant number of ESTs have not been released yet in the public domain, and are mainly held by private companies. For instance, Strabala (2004) reported that in August 2001, 344,279 ESTs from *P. radiata* were sequenced by AgriGenesis Biosciences Ltd. based in New Zealand.

5.2.2. Microarrays: New Nools to Study the Functioning of the Pine Genome

ESTs represent snapshots of the genes expressed in a given tissue or at a given developmental stage, and provide a good representation of the expressed regions of the genome if a sufficient number of different libraries from a range of tissues, developmental stages and after various environmental challenges are analyzed. To this end, several cDNA libraries were produced from pine roots, needles, differentiating xylem, and buds (see previous section). These genes have enabled the development of microarrays to identify genes whose expression varies in response to various environmental and developmental cues (mainly drought-stress response and wood formation). These studies, which will likely increase in the future, have been listed in Table 6 with others that have used other transcriptomic (SAGE, SSH, cDNA-AFLP) and proteomic (2D PAGE MS/MS) approaches to study the functioning of the pine genome.

6. Future Perspective: Challenges for Molecular Breeding of Pines

Pines offer a great opportunity for molecular breeding given their large size and long generation interval. The large size causes inefficiencies in progeny testing and concomitant decreases in heritability of traits. Molecular marker information can reduce these inefficiencies and effectively increase heritability, leading to more accurate selections and commensurate increases in genetic gain (Fernando and Grossman 1989; Lande and Thompson 1990; Hospital et al. 1997). The long generation intervals cause delays in obtaining genetic gain. Savings of many years in the breeding cycle can easily be envisioned if selections can be based on DNA marker information collected at a very early age as opposed to phenotypes measured much later (Williams and Neale 1992). However, these gains can only be achieved if breeding can commence very soon after the selections are made. This has traditionally been known as accelerated breeding and it is a critical, complementary technology (e.g. Bramlett and Burris 1995).

While the opportunity for increased genetic gain per year is great, major challenges to implement molecular breeding include pines' highly heterozygous, outbreeding nature and associated presence of low levels of linkage disequilibrium (LD) in natural populations

(Strauss et al. 1992; Brown et al. 2004). These factors each limit the opportunity for finding QTL by association (Kruglyak 1999; Hirschhorn and Daly 2005) or candidate-gene (Pflieger et al. 2001; Neale and Savolainen 2004) mapping approaches, and for extending QTL results from one pedigree to another within a breeding population. However, given very dense genetic maps or good leads about possible candidate genes, these low LD levels allow for the possibility of discovering tight linkages to important genes. These close linkages should be generally useful across the population for predicting genetic value and implementing molecular breeding.

The challenge in moving forward lies in the development of powerful genetic markers and maps and populations providing genetic information on traits of interest (e.g. Georges et al. 1995; Farnir et al. 2002; Laurie et al. 2004). These populations can include both pedigreed and non-pedigreed populations and both are valid for application of molecular breeding. Pedigreed populations place and maintain large regions of the genome in LD allowing for the tracking of most regions with mapped genetic markers (Lander and Botstein 1989). Once certain regions are identified as causing (or at least being predictive of) significant variation in the phenotype, marker alleles for these regions can be selected for two or more successive generations thereby increasing the frequency of favorable alleles in the population (Edwards and Page 1994; Hospital et al. 2000). One such view of this for pine breeding is termed marker-directed population improvement (MDPI) (Nelson and Echt 2003, 2004).

Non-pedigreed populations provide an opportunity for tracking very small regions across the genome (Hirschhorn and Daly 2005). These are the regions that have remained in LD for very long times, possibly since the creation of the DNA polymorphism used as the genetic marker. Extremely dense genetic maps are required for mapping QTL in non-pedigreed populations through a genome-scan analysis (Kruglyak 1999). Testing specific genes known as candidate genes for association to phenotype does not require a genetic map (Pflieger et al. 2001; Neale and Savolainen 2004). Instead, a hypothesis about what gene(s) might affect a trait's phenotype is needed as well as markers that allow detection of alternative alleles of the candidate gene(s) in the population (Krutovsky and Neale 2005).

The greatest opportunities for implementing molecular breeding exist in breeding programs that utilize best testing practices, including good experimental design, careful field site selection and maintenance, and high-quality measurements on all important traits (e.g. Stuber et al. 1999; Tanksley and Nelson 1996). Most of these programs will contain populations with pedigree structure, allowing for moderately dense genetic maps (10-20 cM spacing) to be used in various linkage-mapping approaches (Darvasi et al. 1993) including MDPI. In addition, all trees are drawn from a source population where tight linkages will be maintained over short recombination intervals allowing for mapping by association through a genome-scan or candidate-gene approach (Wu et al. 2002; Lund et al. 2003).

Where no breeding program exists, large random mating populations could be established specifically for association mapping and subsequent marker-assisted selection and breeding. These would most likely use candidate-gene methods, as candidate genes could be selected from first principles or knowledge of related species and then tested for association with phenotype in the target species and population. It would seem that genome-scan approaches will only be viable in very intensively studied species where dense genetic maps will be developed. In all cases, highly efficient DNA isolation and genotyping methods (e.g. Darvasi and Soller 1994; Mosig et al. 2001) and effective database management and bioinformatic tools will be required for cost effective implementation (Nelson 1997). Expected results include increased genetic gain per generation, due to increased effective heritability and with accelerated breeding, increasing number of generations per unit of time. In addition, traits that are difficult to measure can be improved more readily assuming they are measured and mapped in an early generation.

Acknowledgements: This work was funded, in part, by (A) ANR awards GENOQB (GNP05013) and DIGENFOR (ANR-05-GPLA-006-01) to C Plomion, (B) NSF award DBI-0421717 to DG Peterson, CD Nelson, and MN Islam-Faridi, (C) award RTA03-213 from the Programa Nacional de Recursos y Tecnologías Agroalimentarias (Ministerio de Educación y Ciencia, MEC) to MT Cervera, (D) National Research Council of Italy (Commessa: "Evoluzione e analisi della diversità genetica in piante forestali") to GG Vendramin, (E) New Zealand Foundation for Research Science and Technology (FRST) contract CO4X0207 to Scion (PL Wilcox and RD Burdon), (F) the Finnish Funding Agency for Technology and Innovation and by the Biosciences and Environment Research Council of Finland, to O Savolainen.

References

1. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651-1656
2. Aimers-Halliday J, Shelbourne CJA, Hong SO (1997) Issues in developing clonal forestry with *P. radiata*. In: Burdon RD, Moore JM (eds) "IUFRO '97 Genetics of Radiata Pine", Proc IUFRO Conf, 1-4 Dec and Workshop, 5 Dec, Rotorua, New Zealand. FRI Bulletin No 203, pp 264-272
3. Alazard P (2001) Genetic gain in seeds from first generation seed orchards of maritime pine. Informations Forests, Afocel, France, No 628, 6 p
4. Alía R, Gil L, Pardos JA (1995) Performance of 43 *Pinus pinaster* provenances on 5 locations in Central Spain. Silvae Genet 44: 75-81
5. Alía R, Moro J, Denis JB (1997) Performance of *Pinus pinaster* Ait. provenances in Spain: interpretation of the genotype-environment interaction. Can J For Res 27: 1548-1559
6. Allona I, Quinn M, Shoop I, Swope K, ST-Cyr S, Carlis J, Riield J, Retzel E, Campbell M, Sederoff R, Whetten RW (1998) Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci USA 95: 9693-9698
7. Andrew RL, Peakall R, Wallis IR, Wood JT, Knight EJ, Foley WJ (2005) Marker-based quantitative genetics in the wild?: The heritability and genetic correlation of chemical defenses in Eucalyptus. Genetics 171: 1989-1998
8. Apiolaza LA, Greaves BL (2001) Why are most breeders not using economic breeding objectives. In: IUFRO Symposium "Developing the Eucalypt for the Future". 10-14 Sept, 2001, Valdivia, Chile
9. Arcade A, Faivre-Rampant P, Le Guerroue B, Paques LE, Prat D (1996) Quantitative traits and genetic markers: analysis of a factorial mating design in larch. In: Ahuja MR, Boerjan W, Neale D (eds) Somatic cell genetics and molecular genetics of trees. Kluwer Academic Publ, Doedrecht, The Netherlands, pp 211-216
10. Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218
11. Atwood RA, White TL, Huber DA (2002) Genetic parameters and gains for growth and wood properties in Florida source loblolly pine in the southeastern United States. Can J For Res 32: 1025-1038
12. Auckland L, Bui T, Zhou Y, Shepherd M, Williams CG (2002) Conifer Microsatellite Handbook. Corporate Press, Texas A&M University, College Station, Texas, USA, 57 p
13. Avila C, Muñoz-Chapuli R, Plomion C, Frigerio J-M, Cánovas FM (2000) Two genes encoding distinct cytosolic glutamine synthetases are closely located in the pine genome. FEBS Lett 477: 237-243
14. Bahrman N, Damerval C (1989) Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait). Heredity 63: 267-274
15. Ball RD (2001) Bayesian methods for quantitative trait loci mapping based on model selection: Approximate analysis using the Bayesian Information Criterion. Genetics 159: 1351-1364
16. Ball RD (2005) Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. Genetics 170: 859-873
17. Baradat P, Pastuszka P (1992) Le pin maritime. In: Gallais A, Bannerot H (eds), Amélioration des espèces végétales cultivées, INRA Editions, Paris, pp 695-709
18. Barker CL, Donald T, Pauquet J, Ratnaparkhe MB, Bouquet A, Adam-Blondon AF, Thomas MR, Dry I (2005) Genetic and physical mapping of the grapevine powdery mildew resistance gene, Run1, using a bacterial artificial chromosome library. Theor Appl Genet 111: 370-377

19. Beaulieu J, Plourde A, Daoust G, Lamontagne L (1996) Genetic variation in juvenile growth of *Pinus strobus* in replicated Quebec provenance-progeny tests. For Genet 3: 103-112
20. Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Proc 49th Annu Corn and Sorghum Industry Res Conf, ASTA, Washington DC, USA, pp 250-266
21. Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rholfing T, Fries J, Bradford K, McMenamy J, Smith M, Holeman H, Roe BA, Wiley G, Korf IF, Rabinowicz PD, Lakey N, McCombie WR, Jeddeloh JA, Martienssen RA (2005) Sorghum genome sequencing by methylation filtration. PLoS Biol 3: 13
22. Bell JC, Powell M, Devey ME, Moran GF (2004) DNA profiling, pedigree lineage analysis and monitoring in the Australian breeding program of radiata pine. Silvae Genet 53: 130-134
23. Blada I (1994) Inter-specific hybridization of Swiss Stone Pine *Pinus cembra* L. Silvae Genet 43: 14-20
24. Borzan Z, Papes D (1978) Karyotype analysis in *Pinus*: contribution to the standardization of the karyotype analysis and review of some applied techniques. Silvae Genet 27:144-150
25. Bowe LM, Coat G, dePamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc Natl Acad Sci USA 97: 4092-4097
26. Bradshaw HD, Foster GS (1992) Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. Can J For Res 22: 1044-1049
27. Bramlett DL, Burris LC (1995) Topworking young scions into reproductively-mature loblolly pine. In: Weir RJ, Hatcher AV (eds) Proc 23rd South For Tree Improv Conf, Asheville, North Carolina, USA, pp 234-241
28. Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM (2002) Genetic parameters and QTL analysis of δ13C and ring width in maritime pine. Plant Cell Environ 25: 945-953
29. Brettschneider R (1998) RFLP analysis. In: Karp A, Isaac PG, Ingram DS (eds) Molecular Tools for Screening Biodiversity Plants and Animals. Chapman & Hall, London, UK, pp 83-95
30. Brinker M, van Zyl L, Liu WB, Craig D, Sederoff RR, Clapham DH, von Arnold S (2004) Microarray analyses of gene expression during adventitious root development in *Pinus contorta*. Plant Physiol 135(3): 1526-1539
31. Brown GR, Kadel EE, Bassoni DL, Kiehna KL, Temesgen B, Van Buijtenen JP, Sewell MM, Marshall KA, Neale DB (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159: 799-809
32. Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB (2003) Identification of quantitative trait loci influencing wood property traits in loblolly pine. III. QTL verification and candidate gene mapping. Genetics 164: 1537-1546
33. Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101: 15255-15260
34. Brown GR, Kadel III EE, Bassoni DL, Kiehne KL, Temesgen B, van Buijtenen JP, Sewell MM, Marshall KA, Neale DB (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159: 799-809
35. Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. Mol Ecol 12: 1487-1495
36. Burdon RD (1997) Genetic diversity for the future: conservation or creation and capture? Rotorua, New Zealand, pp 237-264
37. Burdon RD (2001) Genetic diversity and disease resistance: some considerations for research, breeding and deployment. Can J For Res 32: 596-606
38. Burdon RD (2002) An introduction to pines. In: Pines of Silvicultural Importance. CABI Publ, Wallingford, UK
39. Burdon RD (2004) Breeding goals: Issues of goal-setting and applications. In: Walter C, Carson M (eds) Plantation Forest Biotechnology for the 21st Century, Research Signpost, Kerala, India, pp 101-118
40. Burdon RD, Bannister MH, Low CB (1992) Genetic Survey of *Pinus radiata* 4: Variance structures and heritabilities in juvenile clones. NZ J For Sci 22: 187-210
41. Burdon RD, Firth A, Low CB, Miller MA (1998) Multi-site provenance trials of *Pinus radiata* in New Zealand. FAO, Rome, Italy
42. Burdon RD, Gea LD (2006) Pursuit of genetic gain and biotic risk management for *Pinus radiata* in New Zealand. 13th Australasian Plant Breed Conf, Christchurch, New Zealand
43. Burdon RD, Kumar S (2003) Stochastic modelling of the impact of four generations of pollen contamination in unpedigreed gene resources. Silvae Genet 52: 1-7
44. Burdon RD, Kumar S (2004) Forwards versus backwards selection: trade-offs between expected genetic gain and risk avoidance. NZ J For Sci 34: 3-21

45. Burdon RD, Low CB (1992) Genetic survey of *Pinus radiata*. 6: Wood properties: variation, heritabilites, and interrelationship with other traits. NZ J For Sci 22: 228-245
46. Burdon RD, Miller JT (1992) Introduced forest trees of New Zealand: Recognition, role and seed source. 12. Radiata pine (*Pinus radiata* D. Don). NZ For Res Inst Bull 124: 59
47. Burdon RD, Wilcox PL (2006) Population management: potential roles for genomics. Tree Genet Genom (submitted))
48. Byun KO, Kim MZS, Shim SY, Hong SH, Sohn SI (1989) Review of pitch- loblolly hybrid pine (*Pinus rigida* and *P. taeda*) breeding researches in Korea and future strategy. Res Rep, For Genet Res Inst, Korea 25: 204-211
49. Cahalan CM (1981) Provenance and clonal variation in growth, branching and phenology in *Picea sitchensis* and *Pinus contorta*. Silvae Genet 30: 40-46
50. Cai Q, Zhang D, Liu ZL, Wang XR (2006) Chromosomal localization of 5S and 18S rDNA in five species of subgenus *Strobus* and their implications for genome evolution of *Pinus*. Ann Bot (in press)
51. Cánovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Hans-Peter Mock HP, Rossignol M (2004) Plant proteome analysis. Proteomics 4: 285-298
52. Cantón FR, García-Gutiérrez A, Gallardo F, de Vicente A, Cánovas FM (1993) Molecular characterization of a cDNA clone encoding glutamine synthetase from a gymnosperm: *Pinus sylvestris.* Plant Mol Biol 22: 819-828
53. Cantón FR, Le Provost G, Garcia V, Barré A, Frigerio J-M, Paiva J, Fevereiro P, Avila C, Mouret J-F, de Daruvar A, Cánovas FM, Plomion C (2003) Transcriptome analysis of wood formation in maritime pine In: Espinel S, Barredo Y, Ritter E (eds) Sustainable Forestry, Wood products and Biotechnology DFA-AFA Press, Vitoria-Gasteiz, Spain, pp 333-347
54. Cato S, Mc Millan L, Donaldson L, Richardson T, Echt C, Gardner R (2006a) Wood formation from the base to the crown in *Pinus Radiata*: gradients of tracheid wall thickness, wood density, radial growth rate and gene expression. Plant Mol Biol 60:565-581
55. Cato SA, Pot D, Kumar S, Douglas J, Gardner RC, Wilcox PL (2006b) Balancing selection in a dehydrin gene associated with increased wood density and decreased radial growth in *Pinus radiata* (Abstr). In: Plant and Animal Genome XIV Conf, San Diego, CA, USA
56. Cato SA, Richardson TE (1996) Inter- and intra-specific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. Theor Appl Genet 93: 587-592
57. Cervera MT, Plomion C, Malpica CA (2000a) Molecular markers and genome mapping in woody plants. In: Jain SM, Minocha SC (eds) Molecular Biology of Woody Plants. Kluwer Academic Publ, Dordrecht, The Netherlands, pp 375 -394
58. Cervera MT, Remington D, Frigerio JM, Storme V, Ivens B, Boerjan W, Plomion C (2000b) Improved AFLP analysis of tree species. Can J For Res 30: 1608-1616
59. Chagne Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale DB, Plomion C (2003) Comparative genome and QTL mapping between maritime and loblolly pines. Mol Breed 12:185-195
60. Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera M-T, Vendramin GG, Garcia V, Frigerio JM, Echt C, Richardson T, Plomion C (2004) Cross species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109: 1204-12
61. Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigério J-M, Echt C, Richardson T, Plomion C (2004) Cross species transferability of genomic and cDNA SSRs in pines. Theor Appl Genet 109: 1204-1214
62. Chagné D, Lalanne C, Madur D, Kumar S, Frigerio JM, 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
63. Chambers PGS, Borralho NMG (1999) A simple model to examine the impact of changes in wood traits on the cost of thermomechanical pulping and high-brightness newsprint production with radiata pine. Can J For Res 29: 1615-1626
64. Chang S, Puryear JD, Dias MADL, Funkhouser EA, Newton RJ, Cairney J (1996) Gene expression under water deficit in loblolly pine (*Pinus taeda*): isolation and characterization of cDNA clones. Physiol Planta 97: 139-148
65. Chaumeil P (2006) Plasticité moléculaire de deux écotypes de pin maritime soumis à un stress osmotique. Thèse de Doctorat. Université Henri Poincaré, Nancy I, Nancy, 282 p
66. Ciavatta VT, Morillon R, Pullman GS, Chrispeels MJ, Cairney J (2001) An aquaglyceroporin is abundantly expressed early in the development of the suspensor and the embryo proper of loblolly pine. Plant Physiol 127: 1556-1567
67. Claros MG, Cantón FR, Cánovas FM (2004) Construction of a pine BAC genomic library and bioinformatics platform for automated high-throughput sequence analysis (Abstr). Second Annual

Meeting of the Spanish Forest Functional Genomics Network. Pontevedra, Spain (11-29-2004), http://www.difo.uah.es/forestgenomics/pontevedra_04/Gonzalo%20CLAROS.pdf
68. Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-Throughput Screening for Induced Point Mutations. Plant Physiol 126: 480-484
69. Conkle MT (1981) Isozyme variation and linkage in six conifer species. In: Proc Isozymes of North American Forest Trees and Insects, USD Forest Service, General Technical Report PSW-48, pp 11-17
70. Costa P (1999) Réponse moléculaire, physiologique et génétique du pin maritime à une contrainte hydrique. PhD thesis. Univ of Nancy, 116 p
71. Costa P, Plomion C (1999) Genetic analysis of needle protein in Maritime pine. 2. Quantitative variation of protein accumulation. Silvae Genet 48: 146-150
72. Costa P, Pot D, Dubos C, Frigerio JM, Pionneau C, Bodenes C, Bertocchi E, Cervera MT, Remington DL, Plomion C (2000) A genetic map of Maritime pine based on AFLP, RAPD and protein markers. Theor Appl Genet 100: 39-48
73. Costa, P, Bahrman, N, Frigerio, J-M, Kremer A, Plomion C (1998) Water-deficit-responsive proteins in maritime pine. Plant Mol Biol 38: 587-596
74. Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. Can J For Res 24: 372–379
75. Danjon F (1995) Observed selection effects on height growth, diameter and stem form in maritime pine. Silvae Genet 44(1): 10-19
76. Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138: 1365-1373
77. Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943-951
78. Devey ME, Bell JC, Smith DN, Neale DB, Moran GF (1996) A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theor Appl Genet 6: 673- 679
79. Devey ME, Bell JC, Uren TL, Moran GF (2003) A set of microsatellite markers for fingerprinting and breeding applications in *Pinus radiata*. Genome 45: 984-989
80. Devey ME, Bell JC, Uren TL, Moran GF (2003) A set of microsatellite markers for fingerprinting and breeding applications in *Pinus radiata*. Genome 45: 984-989
81. Devey ME, Carson SD, Nolan MF, Matheson AC, Te Riini C, Hohepa J (2004a) QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection. Theor Appl Genet 108: 516-524
82. Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale DB (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. Theor Appl Genet 88: 273-278
83. Devey ME, Fino-Mix A, Kinloch BB Jr, Neale DB (1995) Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white pine blister rust in sugar pine. Proc Natl Acad Sci USA 92: 2066-2070
84. Devey ME, Groom KA, Nolan MF, Bell JC, Dudzinski MJ, Old KM, Matheson AC, Moran GF (2004b) Detection and verification of quantitative trait loci for resistance to Dothistroma needle blight in *Pinus radiata*. Theor Appl Genet 108: 1056-1063
85. Devey ME, Sewell MM, Uren TL, Neale DB (1999) Comparative mapping in loblolly and radiate pine using RFLP and microsatellite markers. Theor Appl Genet 99: 656-662
86. Dhakal LP, White TL and Hodge GR (1996) Realized genetic gains from slash pine (*Pinus elliottii*) tree improvement. Silvae Genet 45: 190-197
87. Dilbirligi M, Erayman M, Sandhu D, Sidhu D, Gill KS (2004) Identification of wheat chromosomal regions containing expressed resistance genes. Genetics 166: 461-481
88. Doudrick RL, Heslop-Harrison JS, Nelson CD, Schmidt T, Nance WL, 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
89. Drewry A (1982) G-banded chromosomes in *Pinus resinosa*. J Hered 73: 305-306
90. Du Toit B, Smith C, Carlson C, Esprey L, Allen R, Little K (1998) Eucalypt and pine plantations in South Africa. In: Nambiar EKS, Cossalter C, Triarks A (eds) Workshop Proceedings. CIFOR, 16-20 Feb 1998, South Africa, pp 23-30
91. Dubos C, Plomion C (2003) Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. Plant Mol Biol 51(2): 249-262
92. Dubos C, Provost G, Pot D, Salin F, Lalane C, Madur D, Frigerio JM, Plomion C (2003b) Identification and characterization of water-stress-responsive genes in hydroponically grown maritime pine (*Pinus pinaster*) seedlings. Tree Physiol 23: 169-179
93. Duminil J, Fineschi S, Hampe A, Jordano P, Salvini D, **V**endramin GG, Petit RJ (2006) Can population genetic structure be predicted from life history traits? Am Naturalist (submitted)

94. Dvornyk V, Sirviö A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19: 179-188
95. Echt CS, DeVerno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol Ecol 7: 307-317
96. Echt CS, Burns R (1999) PCR primer pair sequences and SSRs for various *P. taeda* ESTs. Available on the dendrome web site at http://dendrome.ucdavis.edu/Gen_res.htm, accessed 20 March 2006
97. Echt CS, Nelson CD (1997) Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). Theor Appl Genet 94: 1031-1037
98. Echt CS, Saha S, Deemer D, Nelson CD (2006) Evaluation of pine EST-SSR markers. In: Plant and Animal Genome XIV Conf, San Diego, CA, USA, W-117 (http://www.intl-pag.org/14/abstracts/PAG14_W117.html, accessed 20 March 2006)
99. Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. Theor Appl Genet 88: 376-382
100. Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113-125
101. Egertsdotter U, van Zyl LM, Mackay J, Peter G, Kirst M, Clark C, Whetten R, Sederoff R (2004) Gene expression during formation of earlywood and latewood in loblolly pine: expression profiles of 350 genes. Plant Biol 6: 654-663
102. Elsik CG, Williams CG (2000) Retroelements contribute to the excess low-copy-number DNA in pine. Mol Gen Genet 264: 47-55
103. Elsik CG, Williams CG (2001) Low copy microsatellite recovery form a conifergenome. Theor Appl Genet 103: 1189-1195
104. Emebiri LC, Devey ME, Matheson AC, Slee MU (1997) Linkage of RAPD markers to NESTUR, a stem growth index in radiata pine seedlings. Theor Appl Genet 95: 119-124
105. Emebiri LC, Devey ME, Matheson AC, Slee MU (1998) Age-related changes in the expression of QTLs for growth in radiata pine seedlings. Theor Appl Genet 97: 1053-1061
106. Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within population of lodgepole pine. Genetics 121: 369-377
107. Falkenhagen E (1991) Provenanace variation in *Pinus radiata* at six sites in South Africa. Silvae Genet 40: 41-50
108. Farjon A (1984) Pines. Drawings and description of the genus *Pinus*. EJ Brill, Leiden
109. Farjon A (2001) World checklist and bibliography of conifers, 2nd edn. The Royal Botanic Gardens, Kew, UK
110. Farjon A, Styles BT (1997) *Pinus* (Pinaceae). Monograph 75, Flora Neotropica. New York Botanical Gardens, USA
111. Farnir F, Grisart B, Coppieters W, Riquet J, Berzi P, Cambisano N, Karim L, Mni M, Moisio S, Simon P, Wagenaar D, Vilkki J, Georges M (2002) Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. Genetics 161: 275-287
112. Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 21: 246-477
113. Fisher PJ, Gardner RC, Richardson TE (1996) Single locus microsatellites isolated using 5' anchored PCR. Nucl Acids Res 24: 4369-4371
114. Fisher PJ, Gardner RC, Richardson TE (1998) Characteristics of single and multi-copy microsatellities from *Pinus radiata*. Theor Appl Genet 96: 969-979
115. Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54: 357-374
116. Folkertsma RT, Spassova MI, Prins M, Stevens MR, Hille J, Goldbach RW (1999) Construction of a bacterial artificial chromosome (BAC) library of *Lycopersicon esculentum* cv. Stevens and its application to physically map the *Sw-5* locus. Mol Breed 5: 197-207
117. Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. Mol Biol Evol 18: 1176-1188
118. Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457-1467
119. García-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12: 1195-1206
120. García-Gutiérrez A, Cantón FR, Gallardo F, Sánchez-Jiménez F, Cánovas FM (1995) Expression of ferredoxin-dependent glutamate synthase in dark-grown pine seedlings. Plant Mol Biol 27: 115-128

121. Gaussen H (1960) Les Gymnospermes actuelles et fossils. Fascicule VI. Généralités, genre *Pinus*; Faculté des Sciences, Toulouse
122. Geada López G, Kamiya K, Harada K (2002) Phylogenetic relationships of *Diploxylon* pines (subgenus *Pinus*) based on plastid sequence data. Intl J Plant Sci 163: 737-747
123. Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor Appl Genet 46: 319-330
124. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920
125. Gerber S, Rodolphe F (1994) An estimation of the genome length of maritime pine (*Pinus pinaster* Ait). Theor Appl Genet 88: 289-292
126. 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
127. Gernandt DS, Geada Lopez G, Ortiz Garcia S, Liston A (2005) Phylogeny and classification of *Pinus*. Taxon 54: 29-42
128. Gernandt DS, Liston A, Pinero D (2003) Phylogenetics of *Pinus* subsections *Cembroides* and *Nelsoniae* inferred from cpDNA sequences. Syst Bot 28: 657-673
129. Gion JM, Lalanne C, Le Provost G, Ferry-Dumazet H, Paiva J, Chaumeil P, Frigerio JM, Brach J, Barré A, de Daruvar A, Claverol S, Bonneu M, Sommerer N, Negroni L, Plomion C, Gion J-M, Lalanne C, et al. (2005) The proteome of maritime pine wood forming tissue. Proteomics 5: 3731-3751
130. Gómez MA, González-Martínez SC, Collada C, Climent J, Gil L (2003) Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. Theor Appl Genet 107: 1123-1131
131. González-Martínez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. New Phytol 170: 227-238
132. González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda*. Genetics (in press, doi:10.1534/genetics.105.047126)
133. González-Martínez SC, Gerber S, Cervera MT, Martínez-Zapater J, Gil L, Alía R (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. Theor App Genet 104: 1290-1297
134. González-Martínez SC, Robledo-Arnuncio JJ, Collada C, Díaz A, Williams CG, Alía R, Cervera MT (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. Theor Appl Genet 109: 103-11
135. Govindaraju D, Lewis P, Cullis C (1992) Phylogenetic analysis of pines using ribosomal DNA restriction fragment length polymorphisms. Plant Syst Evol 179: 141-153
136. Grace SL, Hamrick JL, Platt WJ (2004) Estimation of seed dispersal in an old-growth population of Longleaf Pine (*Pinus palustris*) using maternity exclusion analysis. Castanea 69: 207-215
137. Grattapaglia D, 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
138. Greaves BL (1999) The value of tree improvement: A case study in radiata pine grown for structural sawn timber and liner-board. In: Nepveu G (ed) Connection between silviculture and wood quality through modelling approaches and simulation software. Proc of IUFRO Workshop, 5-12 Dec 1999, France, pp 448-459
139. Groover AT, Devey ME, Lee JM, Megraw R, Mitchell-Olds T, Sherman B, Vujcic S, Williams C, Neale DB (1994) Identificationof quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138: 1293-1300
140. Groover AT, Williams CG, Devey ME, Lee JM, Neale DB (1995) Sex-related differences in meiotic recombination frequency in *Pinus taeda.* J Hered 86: 157-158
141. Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL (2004) Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. Evolution 58: 1705-0729
142. Guevara MA, Chagné D, Almeida MH, Byrne M, Collada C, Favre JM, Harvengt L, Jeandroz S, Orazio C, Plomion C, Ramboer A, Rocheta M, Sebastiani F, Soto A, Vendramin GG, Cervera MT (2005a) Isolation and characterization of nuclear microsatellite loci in *P. pinaster.* Mol Ecol Notes 5: 57- 59
143. Guevara MA, Soto A, Collada C, Plomion C, Savolainen O, Neale DB, González-Martínez SC, Cerevera MT (2005b) Genomics applied to the study of adaptation in pine species. Invest Agrar Sist Recur For 14: 292-306
144. Gugerli F, Senn J, Anzidei M, Madaghiele A, Buchler U, Sperisen C, Vendramin GG (2001) Chloroplast microsatellites and mitochondrial *nad1* intron 2 sequences indicate congruent phylogenetic

relationships of Swiss stone pine (*Pinus cembra*), Siberian stone pine (*P. sibirica*) and Siberian dwarf pine (*P. pumila*). Mol Ecol 10: 1489-1497
145. Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol Biol 57: 461-485
146. Gwaze DP, Harding KJ, Purnell RC, Bridgwater FE (2002) Optimum selection age for wood density in loblolly Pine. Can J For Res 32: 1393-1399
147. Gwaze DP, Zhou Y, Reyes-Valdés MH, Al-Rababah MA, Williams CG (2003) Haplotypic QTL mapping in an outbred pedigree. Genet Res 81: 43-50
148. Hamrick JL, Godt JW (1990) Allozyme diversity in plant species. In: Brown A, Clegg MT, Kahler AL, Weir BS (eds) Plant Population Genetics, Breeding and Genetic resources Sinauer Associates, Sunderland, MA, pp 43-63
149. Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Phil Trans R Soc Lond B Biol Sci 351:1291-1298
150. Hannrup B, Ekberg I (1998) Age-age correlations for tracheid length and wood density in *Pinus sylvestris*. Can J For Res 28: 1373-1379
151. Hannrup B, Ekberg I, Persson A (2000) Genetic correlations among wood, growth capacity and stem traits in *Pinus sylvestris*. Can J For Res 15: 161-170
152. Harkins DM, Johnson GN, Skaggs PA, Mix AD, Dupper GE, Devey ME, Kinloch Jr BB, Neale DB (1998) Saturation mapping of a major gene for resistance to white pine blister rust in sugar pine. Theor Appl Genet 97: 1355-1360
153. Harry DE, Temesgen B, Neale DB (1998) Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. Theor Appl Genet 97: 327-336
154. Hartmann HT, Kester DE, Davies FT Jr (1990) Plant Propagation Principles and Practices. 5th edn, Prentice Hall, New Jersey, USA, 647 p
155. 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
156. Hayashi E, Kondo T, Terada K, Kuramoto N, Kawasaki S (2004) Identification of AFLP markers linked to a resistance gene against pine needle gall midge in Japanese balck pine. Theor Appl Genet 108: 1177-1181
157. Heath LS, Ramakrishnan N, Sederoff RR, Whetten RW, Chevone BI, Struble CA, Jouenne VY, Chen D, van Zyl LM, Grene R (2002) Studying the functional genomics of stress responses in loblolly pine with the "Expresso microarray experiment management system". Comp Funct Genom 3: 226-243
158. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95-108
159. Hizume M, Arai M, Tanaka A (1990) Chromosome banding in the genus *Pinus*. III. Fluorescent banding pattern of *P. luchuensis* and its relationships among the Japanese diploxylon pines. Bot Mag Tokyo 103: 103-111
160. Hizume M, Ohgiku A, Tanaka A (1989) Chromosome banding in the genus *Pinus*. II. Interspecific variation of fluorescent banding patterns in *P. densiflora* and *P. thunbergii*. Bot Mag Tokyo 102:25-36
161. 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
162. Hodge GR, Dvorak WS (1999) Genetic parameters and provenance variation of *Pinus tecunumanii* in 78 international trials. For Genet 6: 157-180
163. Hodge GR, Volker PW, Potts BM, Owen JV (1996) A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two Eucalypt species. Theor Appl Genet 92: 53-63
164. Hospital F, Goldringer I, Openshaw S (2000) Efficient marker-based recurrent selection for multiple quantitative trait loci. Genet Res 75: 357-368
165. Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181-1189
166. Hurme P, Sillanpää M, Repo T, Arjas E, Savolainen O (2000) Genetic basis of climatic adaptation in Scots pine by Bayesian QTL analysis. Genetics 156: 1309-1322
167. Hyun SK (1976) Inter-specific hybridization in pines with the special reference to *Pinus rigida* and *P. taeda*. Silvae Genet 25: 188-191
168. Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). Genetics 169: 945-953
169. Isik F, Li B, Frampton J, Goldfarb B (2004) Efficiency of seedlings and rooted cuttings for testing and selection in *Pinus taeda*. For Sci 50:44-53

170. Islam-Faridi MN, Chang Y-L, Zhang HB, Kinlaw C, Doudrick RL, Neale DB, Echt CS, Price HJ, Stelly DM (1998) Construction of a pine BAC library. In: Plant and Animal Genome VI Conf, San Diego, CA, USA
171. Islam-Faridi MN, Mujeeb-Kazi A (1995) Visualization of *Secale cereale* DNA in wheat germplasm by FISH. Theor Appl Genet 90: 595-600
172. Islam-Faridi MN, Nelson CD, Kubisiak TL (2006) Standard karyotype and cytogenetic map for loblolly pine (*Pinus taeda*) using fluorescent *in situ* hybridization. Proc Natl Acad Sci USA (submitted))
173. Islam-Faridi MN, Nelson CD, Kubisiak TL, Gullirmo MV, McNamara VH, Price HJ, Stelly DM (2003) Loblolly pine karyotype using FISH and DAPI positive banding. In: Proc 27th Southern For Tree Improv Conf, June 25-27, 2003, Stillwater, OK, USA, pp 184-188
174. Ivakovic M, Kumar S, Wu H (2006) Adding value to the end-products of radiata pine: A review of breeding for structural timber production. In: Proc 13th Australasian Plant Breed Conf, 18-22 April 2006, Christchurch, NZ (in press)
175. Jackson RB, Schlesinger WH (2004) Curbing the U.S. carbon deficit. Proc Natl Acad Sci USA 101: 15827-15829
176. Jacobs MD, Gardner RC, Murray BG (2000) Cytological characterization of heterochromatin and rDNA in *Pinus radiata* and *P. taeda*. Plant Sys Evo1 223: 71-79
177. Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135: 205-211
178. Jewell DC, Islam-Faridi MN (1994) Details of a technique for somatic chromosome preparation and C-banding of maize. In: Freeling M, Walbot V (eds) The Maize Hand Book. Springer-Verlag, New York, USA. pp 484-493
179. Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140: 1111-1127
180. Jiang JM, Liu ZX, Lu BS, Jiang JM, Liu ZX, Lu BS (1999) Provenance genetic variation analysis of loblolly pine and determination of suitable provenances (areas). For Res, Beijing 12: 485-492
181. Johnson GR, Wheeler NC, Strauss SH (2000) Financial feasibility of marker-aided selection in Douglas-fir. Can J For Res 30: 1942-1952
182. Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of *Ty*1-copia-like sequences as a component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. Proc Natl Acad Sci USA 93: 2708-2713
183. Kao C-H, Zeng J, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152: 1203-1216
184. Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K, Neal DB, Savolainen O (1996) Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? Theor Appl Genet 93:215-221
185. Kärkkäinen K, Savolainen O (1993) The degree of early inbreeding depression determines the selfing rate at the seed stage: model and results from *Pinus sylvestris* (Scots pine). Heredity 71: 160-166
186. Kaya Z, Neale DB (1995) Utility of random amplified polymorphic DNA (RAPD) markers for linkage mapping in Turkish red pine (*Pinus brutia* Ten). Silvae Genet 44: 110-116
187. Kaya Z, Sewell MM, Neale DB (1999) Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 98: 586-592
188. Kayihan GC, Huber DA, Morse AM, White TL, Davis JM (2005) Genetic dissection of fusiform rust and pitch canker disease traits in loblolly pine. Theor Appl Genet 110: 948-958
189. Kent J, Richardson TE (1997) Fluorescently labelled, multiplexed chloroplast microsatellites for high-throughput paternity analysis in *Pinus radiata*. NZ J For Sci 27: 305-312
190. Khoshoo TN (1961) Chromosome numbers in gymnosperms. Silvae Genet 10: 1-9
191. Kinlaw CS, Neale DB (1997) Complex gene families in pine genomes. Trends Plant Sci 2: 356-359
192. Kinloch BB, Parks GK, Flower CW (1970) White pine blister rust: Simply inherited resistance in sugar pine. Science 167: 193-195
193. Kirst M, Basten CJ, Myburg A, Zeng Z-B, Sederoff R (2005a) Genetic architecture of transcript level variation in differentiating xylem of *Eucalyptus* hybrids. Genetics 169: 2295-2303
194. Kirst M, Cordeiro CM, Rezende GDSP, Grattapaglia D (2005b) Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. J Hered 96: 161-166
195. Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R (2003) Apparent homology of expressed genes from wood forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100: 7383-7388
196. Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. Plant Physiol 135: 2368-2378

197. Kirst M, Myers RM, De León JPG, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. Plant Physiol 135: 2368-2378
198. Knott S, Neale DN, Sewell MM, Haley C (1997) Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. Theor Appl Genet 94: 810-820
199. Komulainen P, Brown GR, Mikkonen M, Karhu A, Garcia-Gil MR, O'Malley DM, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. Theor Appl Genet 107: 667-678
200. Kondo T, Terada K, Hayashi E, Kuramoto N, Okamura M, Kawasaki H (2000) RAPD markers linked to a gene for resistance to pine needle gall midge in Japanese black pine (*Pinus thunbergii*). Theor Appl Genet 100: 391-395
201. Korol AB, Ronin YI, Kirzhner VM (1995) Interval mapping of quantitative trait loci employing correlated traits complexes. Genetics 140: 1137-1147
202. Korol AB, Ronin YI, Nevo E, Hayes PM (1998) Multi-interval mapping of correlated trait complexes. Heredity 80: 273-284
203. Kossack DS, Kinlaw CS (1999) IFG, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39: 417-426
204. Kremer A (1992) Predictions of age-age correlations of total height based on serial correlations between height increments in Maritime pine (*Pinus pinaster* Ait.). Theor Appl Genet 85: 152-158
205. Kriebel HB (1985) DNA sequence components of the *Pinus strobus* nuclear genome. Can J For Res 15: 1-4
206. Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139-144
207. Krupkin AB, Liston A, Strauss SH (1996) Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot 83: 489-498
208. Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold hardiness and wood quality related candidate genes in Douglas-fir. Genetics 171: 2029-2041
209. Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the Pinaceae. Genetics 168: 447-461
210. Kuang H, Richardson TE, Carson SD, Bongarten B (1999b) Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. II. Genetics of viability genes. Theor Appl Genet 99: 140-146
211. Kuang H, Richardson TE, Carson SD, Wilcox PL, Bongarten B (1999a) Genetic analysis of inbreeding depression in plus tree 850.055 of *Pinus radiata* D. Don. 1. Genetic map with distorted markers. Theor Appl Genet 98: 697-703
212. Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine x slash pine F_1 family. Theor Appl Genet 90: 1119-1127
213. Kubisiak TL, Nelson CD, Nance WL, Stine M (1996) Comparison of RAPD linkage maps constructed for a single longleaf pine from both haploid and diploid mapping populations. For Genet 3: 230-211
214. Kubisiak TL, Nelson CD, Nowak J, Friend AL (2000) Genetic linkage mapping of genomic regions conferring tolerance to high aluminum in slash pine. J Sust For 10: 69-78
215. Kubisiak TL, Nelson, CD, Stine M (1997) RAPD mapping of genomic regions influencing early height growth in longleaf pine x slash pine F_1 hybrids. Proc 24th Southern For Tree Improv Conf, Orlando, Florida, USA, June 9-12, 1997, pp 198-206
216. Kumar S (2004) Genetic parameter estimates for wood stiffness, strength, internal checking and resin bleeding for radiata pine. Can J For Res 34: 2601-2610
217. Kumar S, Echt C, Wilcox PL, Richardson TE (2003) Testing for linkage disequilibrium in the New Zealand radiata pine breeding population. Theor Appl Genet 108: 292-298
218. Kumar S, Garrick DJ (2001) Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes. Can J For Res 31: 779-785
219. Kumar S, Gerber S, Richardson TE (2006) Pedigree reconstruction using SSR markers in a radiata pine breeding programme. 13th Australasian Plant Breed Conf, 17-21 April, Christchurch, NZ
220. Kumar S, Richardson TE (2005) Inferring heritability and relatedness using molecular markers in radiata pine. Mol Breed 15: 55-64
221. Kumar S, Spelman R, Garrick D, Richardson TE, Wilcox PL (2000) Multiple marker mapping of quantitative trait loci on chromosome *three* in an outbred pedigree of radiata pine. Theor Appl Genet 100: 926-933
222. Kutil BL, Williams CG (2001) Triplet-repeat microsatellites shared among hard and soft pines. J Hered 92: 327-32

223. Lambeth C (2000) Realized genetic gains for first generation improved loblolly pine in 45 tests in coastal North Carolina. J Appl For 24: 140-144
224. Lambeth C, Lee BC, O'Malley D, Wheeler N (2001a) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. Theor Appl Genet 103: 930-943
225. Lamoureux D, Peterson DG, Li W, Fellers JP, Gill BS (2005) The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.). Genome 48: 1120-1126
226. Lamprecht H (1990) Silviculture in the Tropics: tropical forest ecosystems and their tree species-possibilities and methods for their long-term utilization. GTZ, Eschborn, Germany, 296 p
227. Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756
228. Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199
229. Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168: 2141-2155
230. Le Dantec L, Chagné D, Pot D, Cantin O, Garnier-Géré P, Bedon F, Frigerio JM, Chaumeil P, Léger P, García 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
231. Le Provost G, Paiva J, Pot D, Brach J, Plomion C (2003) Seasonal variation in transcript accumulation in wood forming tissues of maritime pine (*Pinus pinaster* Ait.) with emphasis on a cell wall Glycine Rich Protein. Planta 217: 820-830
232. Ledig FT (1998) Genetic variation in *Pinus*. In: Richardson DM (ed) Ecology and Biogeography of *Pinus*. Cambridge Univ Press, Cambridge, UK, pp 251-280
233. Ledig FT, Conkle MT (1983) Gene diversity and genetic structure in a narrow endemic Torrey pine (*Pinus torreyana* Parry ex Carr). Evolution 37: 79-85
234. Leitch IJ, Hanson L, Winfield M, Parker J, Bennett MD (2001) Nuclear DNA C-values complete familial representation in gymnosperms. Ann Bot 88: 843-849
235. Lerceteau EC, 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 Breed 6: 451-458
236. Li C, Yeh FC (2001) Construction of a framework map in *Pinus contorta* subsp. *latifolia* using random amplified polymorphic DNA markers. Genome 44: 147-153
237. Liston A, Robinson WA, Pinero D, Alvarez-Buylla ER (1999) Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenet Evol 11: 95-109
238. Liu ZL, Zhang D, Hong DY, Wang XR (2003) Chromosomal localization of 5S and 18S-5.8S-25S ribosomal DNA sites in five Asian pines using fluorescence *in situ* hybridization. Theor Appl Genet 106: 198-204
239. Lopez-Upton J, White TL, Huber DA (2000) Species differences in early growth and rust incidence of loblolly and slash pine. For Ecol Mng 132: 211-222
240. Lorenz WW, Dean JFD (2002) SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). Tree Physiol 22: 301-310
241. Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JFD (2005) Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. Tree Physiol 26: 1-16
242. Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JF (2006) Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. Tree Physiol 26: 1-16
243. Lowe WJ, Byram TD, Bridgwater FE (1999) Selecting loblolly pine parents for seed orchards to minimise the costs of producing pulp. For Sci 45: 213-216
244. Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I (1996) Karyotyping of three *Pinaceae* species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. Theor Appl Genet 92:411-416
245. Luikart G, England P, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4: 981-994
246. Lund MS, Sorensen P, Guldbrantsen B, Sorensen DA (2003) Multitrait fine mapping of quantitative trait loci using combined linkage disequilibrium and linkage analysis. Genetics 163: 405-410
247. Ma XF, Szmidt A, Wang WR (2006) Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol (in press)

248. MacPherson P, Filion WG (1981) Karyotype analysis and the distribution of constitutive heterochromatin in five species of *Pinus.* J Hered 72: 193-198
249. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380
250. Mariette S, Chagné D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C (2001) Microsatellite markers for *Pinus pinaster* Ait. Ann For Sci 58: 203-206
251. Markussen T, Fladung M, Achere V, Favre JM, Faivre-Rampant P, Aragones A, Da Silva Perez, 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
252. Marquardt PE, Eppreson BK (2004) Spatial and population genetic structure of microsatellites in white pine. Mol Ecol 13: 3305-3315
253. Matheson AC, Lindgren D (1985) Gain from the clonal and the clonal seed orchard options compared for tree breeding programs. Theor Appl Genet 71:242-249
254. Matziris DI (1995) Provenance variation of *Pinus radiata* grown in Greece. Silvae Genet 44: 88-96
255. McKeand S, Mullin T, Byram T, White TL (2003) Deployment of genetically improved loblolly and slash pines in the South USA. J For 101: 32-37
256. Mergen F (1958) Natural polyploidy in slash pine. For Sci 4: 283-295
257. Miksche JP, Hotta Y (19931973) DNA base composition and repetitious DNA in several conifers. Chromosoma 41: 29-36
258. Millar CI (1998) Early evolution of pines. In: Richardson DM (ed) Ecology and Biogeography of *Pinus*. Cambridge Univ Press, Cambridge, UK, pp 69-91
259. Mitton JB (1992) The dynamic mating system of conifers. New For 6: 187-216
260. Moore G, Devos KM, Wang Z, Gale MD (1995) Grasses, line up and form a circle. Curr Biol 5 (7): 737-739
261. Moran GF, Bell JC, Hilliker AJ (1983) Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. J Hered 74: 62
262. Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14: 214-219
263. Morgante M, Vendramin GG, Rossi P, Olivieri AM (1993) Selection against inbreds in early life-cycle phases in *Pinus leucodermis*. Heredity 70: 622-627
264. Morin PA, Luikart G, Wayne RK and the SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19: 208-2169
265. Morse AM, Nelson CD, Covert SF, Holliday AG, Smith KE, Davis JM (2004) Pine genes regulated by the necrotrophic pathogen Fusarium circinatum. Theor Appl Genet 109: 922-932
266. Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci USA 88: 7474-7476
267. Mosig MO, Lipkin E, Khutoreskaya G, Tchourzyna E, Soller M, Friedmann A (2001) A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a Daughter design, using an adjusted false discovery rate criterion. Genetics 157: 1683-1698
268. Moura VPG, Dvorak WS, Hodge GR (1998) Provenance and family variation of Pinus oocarpa grown in the Brazilian cerrado. For Ecol Mng 109: 315-322
269. Müller-Starck G (1998) Isozymes. In: Karp A, Isaac PG, Ingram DS (eds) Molecular Tools for Screening Biodiversity: Plants and Animals. Chapman and Hall, London, UK, pp 75 - 81
270. Muona O, Harju A (1989) Effective population sizes, genetic variability and mating system in natural stands and seed orchards of *Pinus sylvestris*. Silvae Genet 38: 221-228
271. Muona O, Yazdani R, Rudin R (1987) Genetic change between life stages in *Pinus sylvestris*: allozymes variation in seeds and planted seedlings. Silvae Genet 35: 39-42
272. Murray B (1998) Nuclear DNA amounts in gymnosperms. Ann Bot 82 (Suppl A): 3-15
273. Myers RM, Maniatis T, Lerman LS (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. Meth Enzymol 155: 501-527
274. Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:325-330
275. Neale DB, Sederoff RR (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet 77: 212-216

276. Neale DB, Williams CG (1991) Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can J For Res 21: 545-554
277. Nelson CD, Echt CS (2003) New models for marker-assisted selection in tree breeding. In: McKinley C (ed) Proc 27th South For Tree Improv Conf, Stillwater, Oklahoma, USA, p 114
278. Nelson CD, Echt CS (2004) Marker-directed population improvement. In: Li B, McKeand S (eds) Proc IUFRO Joint Conf of Div 2, For Genet Tree Breed in the Age of Genomics: Progress and Future, Charleston, South Carolina, USA, p 255
279. Nelson CD, Kubisiak TL, Johnson G, Burdine C, Bridgwater FE (2003) Microsatellite analysis of loblolly pine. In: Plant and Animal Genome XI Conf, San Diego, CA, USA, P-560 (http://www.intl-pag.org/11/abstracts/P5i_P560_XI.html, accessed 20 March 2006)
280. Nelson CD, Kubisiak, TL, 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
281. Nelson CD, Nance WL, Doudrick RL (1993) A partial genetic linkage map of slash pine (*Pinus elliottii* Engelm. var. *elliottii*) based on random amplified polymorphic DNAs. Theor Appl Genet 8: 145-151
282. Nelson JC (1997) QGENE: software for marker-based genomic analysis and breeding. Mol Breed 3: 239-245
283. Nikles DG (2000) Experience with some Pinus hybrids in Queensland, Australia. In: Dungey HS, Dieters MJ, Nikles DG (Compl) Hybrid Breeding and Genetics of Forest Trees. Proc of QFRI/CRC-SPF Symp, 9-14 April 2000, Noosa, Queensland, Australia. Dept of Pri Indus, Brisbane, pp 14- 26
284. No E, Zhou Y, Loopstra CA (2000) Sequences upstream and downstream of two xylem-specific pine genes influence their expression. Plant Sci 160: 77-86
285. NSF (2005) NSF, USDA and DOE Award $32 Million to Sequence Corn Genome. Press Release 05-197. National Science Foundation. 11-15-2005. http://www.nsf.gov/news/news_summ.jsp?cntn_id=104608
286. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins J Biol Chem 250: 4007-4021
287. O'Malley D, Grattaplagia D. Chaparro JX, Wilcox PL, Amerson HV, Liu B-H, Whetton R, McKeand S, Kuhlman EG, McCord S, Crane B, Sederoff RR (1996) Molecular markers, forest genetics and tree breeding. In: Gustafson JP, Flavell RB (eds) Genomes of Plants and Animals: 21st Stadler Genet Symp. Plenum Press, New York, USA
288. Ohri D, Khoshoo TN (1986) Genome size in gymnosperms. Plant Syst Evol 153: 119-131
289. OlssonT, Lindgren D, Li B (2001) Balancing genetic gain and relatedness in seed orchards. Silvae Genet 50: 222-227
290. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86: 2766-2770
291. Ott J (1991) Analysis of Human Genetic Linkage. Johns Hopkins Univ Press, Baltimore, USA
292. Pacheco D, Díaz S, Osuna D, Bautista R, Claros MG, Cánovas FM, Cantón FR (2005) Identification of candidate genes conferring specific properties to distinct types of *Pinus pinaster* Ait. wood. IUFRO Tree Biotechnol, Pretoria, South Africa
293. Paiva J (2006) Phenotypic and molecular plasticity of wood forming tissues in maritime pine (*Pinus pinaster* Ait.). PhD thesis of Universidad Nova de Lisboa and Bordeaux 1
294. Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. Trends Genet 19: 303-306
295. Park YS (2002) Implementation of conifer somatic embryogenesis in clonal forestry: technical requirements and deployment considerations. Ann For Sci 59: 651-656
296. Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: genomics of model cereals. Genome Res 15: 1643-1650
297. Paul AD, Foster GS, Caldwell T, McRae J (1997) Parameters for height, diameter, and volume in a multilocation clonal study with loblolly pine. For Sci 43: 87-98
298. Pavy N, Laroche J, Bousquet J, MacKay J (2005) Large-scale analysis of secondary xylem ESTs in pine. Plant Mol Biol 57: 203-224
299. Perlack RD, Wright LL, Turhollow AF, Graham RL, Stokes BJ, Erbach DC (2005) Biomass as a feedstock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply. Report No DOE/GO-102005-2135. US Dept of Energy/US Dept of Agri. http://www1.eere.energy.gov/biomass/pdfs/final_billionton_vision_report2.pdf
300. Peterson DG (2005) Reduced representation strategies and their application to plant genomes. In: Meksem K, Kahl G (eds) The Handbook of Genome Mapping: Genetic and Physical Mapping. Wiley-VCH Verlag GmbH, KGaA, Weinheim, pp 307-335
301. Peterson DG, Chouvarine P, Thummasuwan S, Saha S, Mukherjee D, Carlson JE (2006) Exploring the pine genome using Cot filtration and 454 Life Sciences massively parallel shotgun sequencing. In: Plant

and Animal Genomes XIV Conf, San Diego, CA, USA: http://www.intl-pag.org/pag/14/abstracts/PAG14_W376.html
302. Peterson DG, Nelson CD, Islam-Faridi MN, Main DS, Tomkins JP (2005) Accelerating pine genomics through development and utilization of molecular and cytogenetic resources. In: Plant and Animal Genomes XIII Conf, San Diego, CA, USA: http://www.intl-pag.org/pag/13/abstracts/PAG13_P515.html
303. Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12: 795-807
304. Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organisation of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14: 689-701
305. Petit RJ, Vendramin G G (2006) Phylogeography of organelle DNA in plants: an introduction. In: Weiss S, Ferrand N (eds) Phylogeography of Southern European Refugia. Kluwer, Amsterdam, The Netherlands (in press)
306. Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: A review. Mol Breed 7: 275-291
307. Plomion C, Bahrman N, Costa P, Dubos C, Frigério J-M, Gerber S, Gion J-M, Lalanne C, Madur D, Pionneau C (2004) Proteomics for genetics and physiological studies in forest trees: application in maritime pine. In: Kumar S, Fladung M (eds) Molecular Genetics and Breeding of Forest Trees. Haworth Press, Bringhamton, NY, USA, pp 53-80
308. Plomion C, Bahrman N, Durel CE, O'Malley DM (1995a) Genomic analysis in *Pinus pinaster* (Maritime pine) using RAPD and protein markers. Heredity 74: 661-668
309. Plomion C, Costa P, Bahrman N, Frigerio JM (1997) Genetic Analysis of needle proteins 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
310. Plomion C, Durel C-E (1996) Estimation of the average effects of specific alleles detected by the pseudo-testcross QTL mapping strategy. Gen Sel Evol 28: 223-235
311. Plomion C, Durel C-E, O'Malley D (1996a) Genetic dissection of height in maritime pine seedlings raised under accelerated growth condition. Theor Appl Genet 93: 849-858
312. Plomion C, Hurme P, Frigerio J-M, Ridolphi M, Pot D, Pionneau C, Avila C, Gallardo F, David H, Neutlings G, Campbell M, Canovas FM, Savolainen O, Bodénès C, Kremer A (1999) Developing SSCP markers in two *Pinus* species. Mol Breed 5: 21-31
313. Plomion C, LeProvost G, Pot D, Vendramin G, Gerber S, Decroocq S, Brach J, Raffin A, Pastuszka P (2001) Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites. Can J For Res 31: 1816-1825
314. Plomion C, Liu BH, O'Malley DM (1996c) Genetic analysis using trans dominant linked markers in an F_2 family. Theor Appl Genet 93: 1083-1089
315. Plomion C, O'Malley DM (1996) Recombination rate differences for pollen parents and seed parents in pine. Heredity 77: 341-350
316. Plomion C, O'Malley DM, Durel CE (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
317. Plomion C, Pionneau C, Brach J, Costa P, Baillères H (2000) Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol 123: 959-969
318. Plomion C, Yani A, Marpeau A (1996b). Genetic determinism of δ3-carene in maritime pine using random amplified polymorphic DNA (RAPD) markers. Genome 39: 1123-1127
319. Pot D (2004) Déterminisme génétique de la qualité du bois chez le pin maritime: du phénotype aux gènes. Thèse Université de Rennes I, ENSAR
320. Pot D, Chantre G, Rozenberg P, Rodrigues JC, Jones GL, Pereira H, Hannrup B, Cahalan C, Plomion C (2002) Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait.). Ann For Sci 59: 563-575
321. Pot D, McMillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C (2005a) Nucleotide variation in genes involved in wood formation in two pine species. New Phytol 167: 101-112
322. Pot D, Rodrigues J-C, Rozenberg P, Chantre G, Tibbits J, Cahalan C, Pichavant F, Plomion C (2005b) QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.). Tree Genet Gen 2:10-26
323. Price RA, Liston A, Strauss SH (1998) Phylogeny and systematics of *Pinus*. In: Richardson DM (ed) Ecology and Biogeography of *Pinus*. Cambridge Univ Press, Cambridge, UK, pp 49-68
324. Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation rate for chloroplast microsatellites. Genetics 153: 943-947

325. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J (2001) The TIGR Gene Indices: Analysis of gene transcript sequences in highly sampled eukaryotic species. Nucl Acids Res 29: 159-164
326. Quencez C, Bastien C (2001) Genetic variation within and between populations of *Pinus sylvestris* L. (Scots pine) for susceptibility to *Melampsora pinitorqua* Rostr. (pine twist rust). Heredity 86: 36-44
327. Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA (2005) Differential methylation of genes and repeats in land plants. Genome Res 15: 1431-1440
328. Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie WR, Martienssen RA (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23: 305-308
329. Rafalski A (1998) Randomly amplified polymorphic DNA (RAPD) analysis. In: Caetano-Anollés G, Gresshoff PM (eds) DNA Markers Protocols, Applications, and Overviews. Wiley-Liss, New York, USA, pp 75-83
330. Rake AV, Miksche JP, Hall RB, Hansen KM (1980) DNA reassociation kinetics of four conifers. Can J Genet Cytol 22: 69-79
331. Remington DL, O'Malley DM (2000a) Evaluation of major genetic loci contributing to inbreeding depression for survival and early growth in a selfed family of *Pinus taeda*. Evolution 54: 1580-1589
332. Remington DL, O'Malley DM (2000b) Whole genome characterization of embryonic state inbreeding depression in a selfed loblolly pine family. Genetics 155: 337-348
333. Remington DL, Whetten RW, Liu BH, O'Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. Theor Appl Genet 98: 1279-1292
334. Riberio MM, Plomion C, Petit RJ, Vendramin GG, Szmidt AE (2001) Variation of chloroplast single-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). Theor Appl Genet 102: 97-103
335. Richardson BA., Brunsfeld SJ, Klopfenstein NB (2002) DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization of whitebark pine (*Pinus albicaulis*). Mol Ecol 11: 215-227
336. Richardson DM, Rundel PW (1998) Ecology and biogeography of *Pinus*: an introduction. In: Richardson DM (ed) Ecology and Biogeography of *Pinus*. Cambridge Univ Press, Cambridge, UK, pp 3-46
337. Ritland K, Zhuang J, Ralph S, Ritland C, Bohlmann J (2006) Comparative conifer genomics and the new "Treenomix" program. In: Plant & Animal Genomes XIV Conf, San Diego, CA, USA, W124
338. Ritter E, Aragonés A, Markussen T, Acheré V, Espinel S, Fladung M, Wrobel S, Faivre-Rampant P, Jeandroz S, Favre JM (2002) Toward construction of an ultra high density linkage map for *Pinus pinaster*. Ann For Sci 59: 637-643
339. Robledo-Arnuncio JJ, Collada C, Alía R, Gil L (2005) Genetic structure of montane isolates of *Pinus* sylvestris L. in a Mediterranean refugial area. J Biogeo 32: 595-605
340. Rosvall O, Lindgren D, Mullen TJ (1998) Sustainability, robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. Silvae Genet 47: 307- 321
341. Rounsley S, Xiaoying L, Ketchum KA (1998) Large-scale sequencing of plant genomes. Curr Opin Plant Biol 1: 136-141
342. Rozenberg P, Van Loo J, Hannrup B, Grabner M. 2002. Clonal variation of wood density record of cambium reaction to water deficit in *Picea abies* (L.) Karst. Ann For Sci 59: 533–540
343. Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8: 321-329
344. Rudin D, Ekberg I (1978) Linkage studies in *Pinus sylvestris* L. using macrogametophyte allozymes. Silvae Genet 27: 1-12
345. Sanchez AC, Ilag LL, Yang D, Brar DS, Ausubel F, Khush GS, Yano M, Sasaki T, Li Z, Huang N (1999) Genetic and physical mapping of *xa13*, a recessive bacterial blight resistance gene in rice. Theor Appl Genet 98: 1022-1028
346. Sánchez C, Vielva MJ, Vieitez AM, de Mier B, Abarca D, Díaz-Sala C (2005) Identification of genes related to adventitious rooting capacity in pine and chestnut. IUFRO Tree Biotechnol, Pretoria, South Africa
347. Sathyan P (2004) Identification of drought responsive genes in Aleppo pine (*Pinus halepensis*) and loblolly pine (*Pinus taeda* L.). Texas A&M Univ, Texas, USA, 74 p
348. Sathyan P, Newton RJ, Loopstra CA (2005) Genes induced by WDS are differentially expressed in two populations of aleppo pine (*Pinus halepensis*). Tree Genet Genom 1: 166-173
349. Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8: 552-560

350. Sax K, Sax HJ (1933) Chromosome numbers and morphology in the conifers. J Arnold Arbor 14: 356-375
351. Saylor LC (1961) A karyotype analysis of selected species of *Pinus*. Silvae Genet 10: 77-85
352. Saylor LC (1964) Karyotype analysis of *Pinus* –group *Lariciones*. Silvae Genet 13: 165-170
353. Saylor LC (1972) Karyotype analysis of *Pinus* –subgenus *Pinus*. Silvae Genet 21: 155-163
354. Saylor LC (1983) Karyotype analysis of the genus *Pinus* –subgenus *Strobus*. Silvae Genet 32: 119-121
355. Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T (2000) The contribution of short repeats of low sequence complexity to large conifer genomes. Theor Appl Genet 101: 7-14
356. Schmidt R (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol Biol 48: 21-37
357. Scotti I, Burelli A, Cattonaro F, Chagné D, Fuller J, Hedley PE, Jansson G, Lalanne C, Madur D, Neale D, Plomion C, Powell W, Troggio M, Morgante M (2005) Analysis of the distribution of marker classes in a genetic linkage map: a case study in Norway spruce (*P. abies* karst). Tree Genet Genom 1:93-102
358. Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004) Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. Theor Appl Genet 109: 1648-1659
359. Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000): Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. Theor Appl Genet 101: 1273-1281
360. Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB (2002). Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. Theor Appl Genet 104: 214-222
361. Sewell MM, Neale DB (2000) Mapping quantitative traits in forest trees. In: Jain SM, Minocha SC (eds) Molecular Biology of Woody Plants. Kluwer Acad Publ, The Netherlands, pp 407-433
362. Sewell MM, Sherman BK, Neale DB (1999) A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and inetegration if individual linkage maps from two outbred three-generation pedigrees. Genetics 151: 321-330
363. Shepherd M, Cross M, Dieters MJ, Henry R (2003) Genetic maps for *Pinus elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. Theor App Genet 106: 1409-1419
364. Shelbourne CJA (1992) Genetic gains from different kinds of breeding population and seed or plant production populations. S Afr For J 160: 49-65
365. Shelbourne CJA (2000) Some insights on hybrids in forest tree improvement. In: Dungey HS, Dieters MJ, Nikles DG (Compl) Hybrid Breeding and Genetics of Forest Trees. Proc of QFRI/CRC-SPF Symp, 9-14 April 2000, Noosa, Queensland, Australia. Dept of Pri Indus, Brisbane, pp 53-62
366. Shelbourne CJA, Apiolaza LA, Jayawickrama KJS, Sorensson CT (1997) Developing breeding objectives for radiata pine in New Zealand. In: Burdon R, Moore JM (eds) Proc of IUFRO'97 Genetics of Radiata Pine, Rotorua, NZ, 1-4 Dec, 1997. FRI Bull No 203, pp 160-168
367. Shepherd M, Cross M, Dieters MJ, Henry R (2002b) Branch architecture QTL for *Pinus elliottii* var. elliottii x *Pinus caribea* var. hondurensis hybrids. Ann For Sci 59: 617-625
368. Shepherd M, Cross M, Maguire TL, Dieters MJ, Williams CG, Henry RJ (2002a) Transpecific microsatellites for hard pines. Theor Appl Genet 104: 819-827
369. Shoulders E (1984) The case for planting longleaf pine. Proc of the Southern Silvicultural Res Conf, Nov 7-8, 1984, Atlanta, Georgia, USA, pp 255-260
370. Smith DN, Devey M (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37: 977-983
371. Soranzo N, Alia R, Provan J, Powell W (2000) Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol Ecol 9: 1205-1211
372. Sorensson CT, Shelbourne CJA (2005) Clonal forestry. In: Colley M (ed) Forestry Handbook. New Zealand Inst of Foresters, pp 92-96
373. Springer NM, Xu X, Barbazuk WB (2004) Utility of different gene enrichment approaches toward identifying and sequencing the maize gene space. Plant Physiol 136: 3023-3033
374. Stelzer HE, Goldfarb B (1997) Implementing clonal forestry in the southeastern United States: SRIEG satellite workshop summary. Can J For Res 27: 442-446
375. Stirling B, Yang ZK, Gunter LE, Tuskan GA, Bradshaw HD (2003) Comparative sequence analysis between orthologous regions of the *Arabidopsis* and *Populus* genomes reveals substantial synteny and microcollinearity. Can J For Res 33: 2245-2251
376. Strabala TJ (2004) Expressed sequence tag databases from forestry tree species. In: Kumar S, Fladung M (eds) Molecular Genetics and Breeding of Forest Trees. Haworth Press, Binghamton, NY, USA, pp 19-52

377. Strauss SH, Doerksen AH (1990) Restriction fragment analysis of pine phylogeny. Evolution 44: 1081-1096
378. Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular marker-aided selection in forest tree breeding. Can J For Res 22: 1050-1061
379. Stromberg LD, Dudley JD, Rufener GK (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. Crop Sci 34: 1221-1225
380. Stuber CW, Polacco M, Senior ML (1999) Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield. Crop Sci 39:1571-1583
381. Syring J, Willyard A, Cronn R, Liston A (2005) Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot 92: 2086-2100
382. Syvanen AC (2001) Accessing genetic variation. Genotyping single nucleotide polymorphisms. Nat Rev Genet 2: 930-942
383. Szmidt AE, Muona O (1989) Linkage relationships of allozyme loci in *Pinus sylvestris*. Heredity 111: 91-97
384. Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191-203
385. Telfer EJ, Echt CS, Nelson CD, Wilcox PL (2006) Comparative mapping in *Pinus radiata* and *P. taeda* reveals co-location of wood density-related QTL. In: Plant and Animal Genome XIV Conf, San Diego CA, USA
386. Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 102: 664-675
387. Thiellement H, Plomion C, Zivy M (2001) Proteomics as a tool for plant genetics and breeding. In: Dunn MJ, Pennington S (eds) Proteomics: from protein sequence to function. BIOS Scientific Publ, Oxford, UK, pp 289-309
388. Thoday JM (1961) Location of polygenes. Nature 191:368-370
389. Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in *Cinnamoyl CoA Reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics 171: 1257-1265
390. Travis SE, Ritland K, Whitham TG, Keim P (1998) A genetic linkage map of pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. Theor Appl Genet 97: 871-880
391. Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE (1992) Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Bio/Technology 10: 686-690
392. Tuskan GA, DiFazio SP, Teichmann T (2004) Poplar genomics is getting popular: the impact of the poplar genome project on tree research. Plant Biol (Stuttg) 6: 2-4
393. Van der Burgh J (1973) Hölzer der nierderrheinischen Braunkohlenformation. 2. Hölzer der Braunkohlengruben 'Maria Theresia' zu Herzogenrath, 'Zukunft West' zu Eschweiler und 'Victor' (Zülpich Mitte) zu Zülpich. Nebst einer systematisch-anatomischen Bearbeitung der Gattung *Pinus* L. Rev Palaeobot Palynol 15: 73-275
394. van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. Trends Ecol Evol 17: 577-582
395. van Zyl L, von Arnold S, Bozhkov P, Chen Y, Egertsdotter U, MacKay J, Sederoff R, Weir B, Shen J, Sun Y-H, Whetten R, 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
396. Verhaegen D, Plomion C, Poitel M, Costa P, Kremer A (1998) Quantitative trait disscetion analysis in *Eucalyptus* using RAPD markers: 2. linkage disequilibrium in a factorial design between *E. urophylla* and *E. Grandis*. For Genet 5: 61-69
397. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23: 4407-4414
398. Wahlenberg WG (1946) Longleaf Pine. Charles Lathrop Pack Forestry Foundation, Washington DC, USA
399. Wakamiya I, Price HJ, Messina MG, Newton RJ (1996) Pine genome size diversity and water relations. Physiol Planta 96: 13-20
400. Walker S, Haines R, Dieters M (1996) Beyond 2000. In: Dieters MJ, Matheson AC, Nikles DG, Harwood CE (eds) Clonal forestry in Queensland. Queensland Forestry Research Institute. Tree improvement for sustainable tropical forestry. QFRI-IUFRO Conf, Caloundra, Queensland, Australia, 27 Oct-1 Nov 1996. Vol 2. 1996, pp 351-354
401. Walter R, Epperson BK (2001) Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. Mol Ecol 10: 103-111

402. Wang XR, Szmidt AE, Nguyen HN (2000) The phylogenetic position of the endemic flat-needle pine *Pinus krempfii* (Pinaceae) from Vietnam based on PCR-RFLP analysis of chloroplast DNA. Plant Syst Evol 220: 21-36
403. Wang XR, Szmidt AE, Savolainen O (2001) Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan plateau. Genetics 159: 337-346
404. Wang XT, Tsumara Y, Yoshimaru H, Nagasaka K, Szmidt A (1999) Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK, rpl20-rps18* spacer, and *trnV* intron sequences. Am J Bot 86: 1742-1753
405. Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L (2002) Gramene: a resource for comparative grass genomics Nucl Acids Res. 30: 103-105
406. Warren JM, Covert SF (2004) Differential expression of pine and *Cronartium quercuum* f. sp. fusiforme genes in fusiform rust galls. Appl Environ Microbiol 70: 441-451
407. Watkinson JI, Sioson AA, Vasquez-Robinet C, Shukla M, Kumar D, Ellis M, Heath LS, Ramakrishnan N, Chevone B, Watson LT, van Zyl L, Egertsdotter U, Sederoff RR, Grene RWatkinson JI, Sioson AA, et al. (2003) Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. Plant Physiol 133: 1702-1716
408. Wei RP, Lindgren D, Yeh FC (1997) Expected gain and status number following restricted individual and combined-index selection. Genome 40: 1-8
409. Weng C, Kubisiak TL, Nelson CD, 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
410. Weng C, Kubisiak TL, Stine M (1998) SCAR markers in a longleaf pine x slash pine F1 family. For Genet 5: 239-247
411. Whetten R, Sun Y, Zhang Y Sederoff RR (2001) Functional genomics and cell wall biosynthesis in loblolly pine. Plant Mol Biol 47: 275-291
412. Whitelaw CA, Barbazuk WB, Pertea G, Chan AP, Cheung F, Lee Y, van Heeringen S, Karamycheva S, Bennetzen JL, SanMiguel P, Lakey N, Bedford J, Yuan Y, Budiman MA, Resnick A, van Aken S, Utterback T, Riedmuller S, Williams SM, Feldblyum T, Schubert K, Beachy R, Fraser CM, Quackenbush J (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302: 2118-2120
413. Wilcox P L, Richardson TE, Carson SD (1997) Nature of quantitative trait variation in *Pinus radiata*: insights from QTL detection experiments. In: Burdon RD, Moore JM (eds) Proc IUFRO '97: Genetics of Radiata Pine. 1-12-0097. New Zealand For Res Inst. Sala St, Rotorua, pp 304-312
414. Wilcox PL (1995) Genetic dissection of fusiform rust resistance. PhD Thesis, Dept of Forestry, North Carolina State Univ, Raleigh, NC, USA, 125 p
415. Wilcox PL (1997) Linkage groups, map length, and recombination in *Pinus radiata*. In: Burdon RD, Moore JM (eds) Proc IUFRO '97: Genetics of Radiata Pine. 1-12-0097. New Zealand For Res Inst, Sala St, Rotorua
416. Wilcox PL, Amerson HV, Kuhlman EG, Liu B-H, O'Malley DM, Sederoff RR (1996) Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. Proc Natl Acad Sci USA 93: 3859-3864
417. Wilcox PL, Burdon RD (2006) Application of association genetics to coniferous forest trees. 13th Australasian Plant Breed Conf, Christchurch, New Zealand
418. Wilcox PL, Carson SD, Richardson TE, Ball RD, Horgan GP, Carter P (2001b) Benefit-cost analysis of DNA marker-based selection in progenies of *Pinus radiata* seed orchard parents. Can J For Res 31: 2213-2224
419. Wilcox PL, Cato S, McMillan L, Power M, Ball RD, Burdon RD, Echt CS (2004) Patterns of linkage disequilibrium in *Pinus radiata*. In: Plant and Animal Genome XII Conf, San Diego, CA, USA, W89: URL: http://www.intl-pag.org/12/abstracts/W22_PAG12_89.html
420. Wilcox PL, Echt CS, Burdon RD (2006) Gene-assisted selection: applications of association genetics for forest tree breeding. In: Oraguzie N (ed) Association Mapping in Plants. Springer Verlag, New York, USA (in press)
421. Wilcox PL, Echt CS, Cato SA, McMillan LK, Kumar S, Ball RD, Burdon RD, Pot D (2003) Gene assisted selection- a new paradigm for selection in forest tree species? In: Plant and Animal Genomes XI Conf, San Diego, CA, USA
422. Wilcox PL, Richardson TE, Carson SD (1997) Nature of quantitative trait variation in *Pinus radiata*: insights from QTL detection experiments. In: Burdon RD, Moore JM (eds) Proc IUFRO '97: Genetics of Radiata Pine. 1-12-0097. New Zealand For Res Inst, Sala St, Rotorua, pp 304-312
423. Wilcox PL, Richardson TE, Corbet GE, Ball RD, Lee JR, Djorovic A, Carson SD (2001a) Framework Linkage Maps of *Pinus radiata* D. Don based on pseudotestcross markers. For Genet 8:109-117

424. Williams CG, Neale DB (1992) Conifer wood quality and marker-aided selection: a case study. Can J For Res 22: 1009-1017
425. Williams CG, Zhou Y, Hall SE (2001) A chromosomal region promoting outcrossing in a conifer. Genetics 159: 1283-1289
426. Williams C, Megraw RA (1993). Juvenile-mature relationships for wood density in *Pinus taeda*. Can J For Res 24: 714-722
427. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18: 6531-6535
428. Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22: 506-519
429. Wu H (2002) Study of early selection in tree breeding. 4. Efficiency of marker aided early selection (MAES). Silvae Genet 51: 261-269
430. Wu J, Krutovskii KV, Strauss SH (1999) Nuclear DNA diversity, population differentiation, and phylogenetic relationship in the Californian closed-cone pines based on RAPD and allozyme markers. Genome 42: 893-908
431. Wu R, Ma C-X, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160: 779-792.
432. Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND (2004) Comparative physical mapping reveals features of microsynteny between Glycine max, *Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47: 141-155
433. Yang SH, Loopstra CA (2005) Seasonal variation in gene expression for loblolly pines (*Pinus taeda*) from different geographical regions. Tree Physiol 25: 1063-107325(8):
434. Yazdani R, Nilsson JE, Plomion C, Mathur G (2003) Marker trait association for autumn cold acclimatation and growth rhythm in *Pinus sylvestris*. Scand J For Res 18: 29-38
435. Yazdani R, Yeh FC, Rimsha J (1995) Genomic mapping of *Pinus sylvestris* (L.) using random amplified polymorphic DNA markers. For Genet 2: 109-116
436. Yin TM, Huang MR, Wang MX, Zhu LH, Zhai WX (1997) Construction of molecular linkage map in masson pine using RAPD markers and megagametophytes from a single tree. Acta Bot Sin 39: 607-612
437. Yin TM, Wang XR, Andersson B, Lerceteau-Köhler E (2003) Nearly complete maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. Theor Appl Genet 106: 1075-1083
438. Yuan Y, SanMiguel PJ, Bennetzen JL (2003) High-Cot sequence analysis of the maize genome. Plant J 34: 249-255
439. Zeng ZB (1993a) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Pro Natl Acad Sci USA 90: 10972-10976
440. Zeng ZB (1993b) Precision mapping of quantitative trait loci. Genetics 136: 1457-1468
441. Zeng ZB, Liu J, Stam LF, Kao CH, Mercer JM, Cathy CL (2000). Genetic architecture of a morphological shape difference between two *drosophila* species. Genetics 154: 299-310
442. Zhang H-B, Wu C (2001) BACs as tools for genome sequencing. Plant Physiol Biochem 39: 195-209
443. Zheng Y, Ennos R, Wang HR, Zheng YQ, Wang HR (1994) Provenance variation and genetic parameters in a trial of *Pinus caribaea* Morelet var. *bahamensis* Barr and Golf. For Genet 1: 165-174
444. Zhou Y, Bui T, Auckland L, Williams CG (2002) Undermethylation as a Source of microsatellites in large plant genomes. Genome 34:91-99.
445. Zhou Y, Gwaze DP, Reyes-Valdes MH, Biu T, Williams CG (2003) No clustering for linkage map based on low-copy and undermethylated microsatellites. Genome 46:809-816

Table 1. Systematics of the genus *Pinus*, according to Gernandt et al. (2005).

Subgenus	Section	Subsection	Number of species	Best-known species	Distribution
Pinus	*Pinus*	*Pinus*	17	*P. sylvestris, P. kesiya, P. merkusii*	Eurasia, North America
		Pinaster	7	*P. pinaster*	Mediterranean, Asia
	Trifoliae	*Contortae*	4	*P. banksiana, P. contorta*	America
		Australes	26	*P. elliottii, P. radiata, P. taeda*	America
		Ponderosae	17	*P. jeffreyi, P. ponderosa*	America
Strobus	*Parrya*	*Balfourianae*	3	*P. balfouriana*	America
		Cembroides	11	*P. cembroides, P. culminicola*	America
		Nelsoniae	1	*P. nelsonii*	Central America
	Quinquefoliae	*Gerardianae*	3	*P. bungeana P. gerardiana*	Asia
		Krempfianae	1	*P. krempfii*	Asia
		Strobus	21	*P. cembra, P. lambertiana, P. strobus*	America, Eurasia

Table 2. Characteristics and applications of molecular markers.

Feature	Protein markers		DNA markers				
			Hybridization-based	PCR-based			
	Isozyme	**2D-PAGE**	**RFLP**	**RAPD**	**AFLP**	**SSR**	**SNP**
Tissue quantity (g)	0.5-2	1	na	na	na	na	na
DNA quantity (μg)	na	na	10	0.02	0.25-0.5	0.02	0.005-0.02
DNA quality	na	na	high	medium	medium	medium	medium
Allelism and mode of inheritance	multiallelic codominant	biallelic, codominant or dominant	biallelic codominant	biallelic dominant	biallelic dominant	multiallelic codominant	biallelic codominant
Number loci analysed per assay	1-5	several tens	1-6	2-30	20-100	1-10 (multiplexing)	1-several hundreds (multiplexing)
Polymorphism	Low	low	medium	medium	medium	high	medium[a]
Development cost	Low	low	high	low	medium	high	high
Cost per analysis	Low	medium	high	low	medium	low	low
Amenable to automation	Low	low	low	medium	medium	high	high
Technical demand	Low	high	high	low	medium	low	medium
Reproducibility	high[b]	medium[b]	high	low-high[c]	high	high	high
Review	(Müller-Starck 1998)	(Plomion et al. 2004)	(Brettschneider 1998)	(Rafalski 1998)	(Cervera et al. 2000b)	(Echt and Burns 1999)	(Syvanen 2001)
Applications							
Certification	-	-	-	-	+	+++	++
Diversity	+	+	++	+	+	+++	+++
Phylogeny	++	+	++	++	++	++	+++
Mapping	-	-	++	++[d]	++[d]	+++	+++
Comparative mapping	+	+	++	-	-	+++	+++
QTL analysis	+	+	++	+	+	+++	+++
Association studies	+	+	+	-	-	++[e]	+++

[a]Sequence dependent (exon *vs* UTR and intron), [b]Under the same environmental conditions, type of tissue and age, [c]Between laboratories *vs* single laboratory, [d]To construct dense maps, [e]Used as control loci

Table 3. Summary of genetic linkage maps of *Pinus* species.

Species Pedigree	Number of linkage groups (+pairs, triplets)	Total number of linked marker loci	RAPD, AFLP	SSR	EST-based	others	Total genetic distance (cM)	Genome saturation	References
Pinus pinaster									
Haploid (1 tree)	Not estimated	119	-	-	-	119 (2D proteins)	Not estimated	Not estimated	Bahrman and Damerval 1989
Haploid (8 trees)	Not estimated	65	-	-	-	65 (2D proteins)	530	25%	Gerber et al. 1993
Haploid (H12 self) Haploid (H12 open pollinated)	13(+5)	263	251	-	-	-	1223 1236	90% 90%	Plomion et al. 1995b
Haploid (H12 self)	12	463	436	-	-	27 (2D proteins)	1860	100%	Plomion et al. 1995a
Haploid (H12self) Doiploid (H12 self)	11 11	94 94	94 94	- -	- -	- -	1169 1354	Not estimated Not estimated	Plomion and O'Malley 1996
Haploid (H12 self)	13	398	235+127	-	-	36 (2D protein)	1873	93.4%	Costa et al. 2000
F_1 - AFOCEL	12	759	738	14	7	-	1994	Not estimated	Ritter et al. 2002
9.103.3 x 10.159.3 (consensus)	12	620	620	-	-	-	1441	Not estimated	Chagné et al. 2002
9.103.3 x 10.159.3 (consensus)	12	326	276	-	50	-	1638	Not estimated	Chagné et al. 2003
Pinus taeda									
various pedigrees	Not estimated	-	-	-	-	20 isozymes	Not estimated	Not estimated	Conkle et al. 1981
base pedigree	20	75	-	-	90	6	Not estimated	Not estimated	Devey et al. 1994
Haploid (Tree 10-5)	16	458	458	-	-	-	1727	97%	Wilcox 1995
base and qtl pedigrees	18	357	67	-	257	12	1300	76.4%	Sewell et al. 1998

Tree 7-56	12	508	508	-	-	-	1741	100%	Remington et al. 1999
base and qtl pedigrees	20	265	253	12	-	-	1281	75.4%	Devey et al. 1999
base pedigree	15	51	-	51	-	-	795	46.7%	Zhou et al. 2003
Pinus sylvestris									
49-2	14	261	261	-	-	-	2638		Yazdani et al. 2003
F_1 of P315 x E1101	12(+4)	179	179	-	-	-	1000	50%	Hurme et al. 2000
F_1 of AC3065 x Y3038	♀: 12 ♂: 12	188 245	188 245	- -	- -	- -	1645 1681	98%	Yin et al. 2003
F_1 of E635 x E1101	12(+3)	260	194	4	61	-	1314	66-85%	Komulainen et al. 2003
Pinus radiata									
3-generation	22	208	41	2	165	-	1382	Not estimated	Devey et al. 1996
Haploid (full-sib seed)	14	267	267	-	-	-	1665	93%	Emebiri et al. 1998
S_1	19	172	168	4	-	-	1117	56%	Kuang et al. 1999b
Pseudotest-cross (Parent 850.055)	20	235	224	11	-	-	1414	85%	Wilcox et al. 2001a
Pseudotest-cross (Parent 850.096)	21	194	185	9	-	-	1144	77%	Wilcox et al. 2001a
Two full-sib families	12	311	-	213	98	-	1352	Not estimated	Wilcox et al. 2004
Pinus elliottii									
8-7	13(+9pairs)	73	73	-	-	-	782	64-75%	Nelson et al. 1993
18-62 x 8-7	17(+12pairs)	129	129	-	-	-	1146	Not estimated	Kubisiak et al. 2000
D4PC40 x D4PC13	15	154	63	-	45	41 (RFLPs) 5 isozyme	1,115	Not estimated	Brown et al. 2001
Pinus palustris									
3-356	16(+6pairs)	133	133	-	-	-	1635	85%	Nelson et al. 1994
P. elliotti and P palustris **hybrids**									
F_1 of - 3-356 x H-28	♀: 18(+3 pairs) ♂: 13(+6	122 91	122 91	- -	- -	- -	1368 953	81% 62%	Kubisiak et al. 1995

	pairs)								
BC1 of 488 x 18-27	♀: 17 ♂:19	133 83	133 83	- -	- -	- -	1338 995	91% 81%	Weng et al. 2002
***P. elliotti and P. caribea* var. *hondurensis* hybrids**									
F_1 of 2PEE1-102 x 1PCH1-63	♀: 24 ♂: 25	125 155	117 145	8 10	- -	- -	1548 1823	82% 88%	Shepherd et al. 2003

Table 4. QTL and marker-trait association detected in pines: pedigree structure and methods.

Species	Objective	Population (size) [a]	Replications [b]	clonal replicates	Genetic markers	QTL analysis (analytical model) [c]	Software	Reference
Pinus elliottii	Detection	F1OB (186)	1	no	RAPD	SG (54 : 27 genotypes from each tail), SM, MM	SAS	Kubisiak et al. 2000
Pinus elliottii var elliottii x *Pinus caribaea var hondurensis*	Detection	F1OB (89)	2 sites (60,29)	no	AFLP, SSR	SM, CIM	QTLcartographer	Sheperd et al. 2005
(*Pinus palustris Mill* x *Pinus elliottii Engl*) x *Pinus elliottii Engl*	Detection, Stability (age)	F1OB (258)	3 sites (82,83,93)	no	-	SM, IM	SAS, MapmakerQTL	Weng et al. 2002
Pinus sylvestris	Detection	F1OB	1	no	AFLP	IM	Qgene	Lerceteau et al. 2000
	Detection, Stability (time replicate)	F2HS	2	no	RAPD	SG (1994 : 48 vs 48, 1996 48 vs 48), SM, Bayesian QTL analysis	SAS, Multimapper/OUTBRED	Hurme et al. 2000
	Detection	F1OB	1	no	RAPD	SM	SAS	Yazdani et al. 2003
Pinus pinaster	Detection, Stability (Age)	F2S (120)	3 (age)	no	RAPD	IM	MAPMAKERQTL	Plomion et al. 1996a
	Detection	F2S	1	no	RAPD, AFLP, proteins	CIM	QTLcartographer	Costa 1999
	Detection	F2OB (186)	1	no	AFLP	CIM	MultiQTL	Brendel et al. 2002
	Detection	F20B (186)	1	no	AFLP, EST	CIM	MultiQTL	Pot et al. 2005
	Detection	F1OB (80)	1	no	AFLP,SSR, EST	IM	SAS	Markussen et al. 2002

	Detection, stability (age, season)	F20B (186)	1	no	AFLP, EST	CIM	MultiQTL	Pot 2004
	Detection	F2S (120)	1	no	RAPD	IM	MAPMAKER-QTL	Plomion et al. 1996b
Pinus radiata	Detection, Stability (age)	F1OB (174)	1	no	RAPD	SM, BSA (9 vs 9)	-	Emebiri et al. 1997
	Detection, Stability (age)	F1OB (80-93)	3 times (age)	no	RAPD, AFLP, SSR	IM, MTM	-	Kumar et al. 2000
	Detection, Stability (age)	F1OB (93)	2 times (age)	no	RAPD, AFLP, SSR	Bayesian approach	Splus	Ball 2001
	Detection	CD	1	no	SSR	-	-	Kumar et al. 2004
	Detection, Validation, Stability (genetic background)	3 unrelated F1OB (400 each)	2 sites	no	RFLP and SSR	SG (Juvenile wood density : 50 vs 50; Diameter at breast height : 100 vs 100), SM	GENSTAT	Devey et al. 2004a
	Detection, Validation	Detection: 6 related F2OB (202)	1 site for the detection population	no	RFLP + SSR	SM within family + SM among families	GENSTAT	Devey et al. 2004b
		Verification: 1 F2OB (400)	2 sites for the verification population	yes				
Pinus taeda	Detection	F20B (Detection population : 172)	6 sites (19 to 35 trees per sites)	no	RFLP	SG, SM	home made	Groover et al. 1994
	Detection	F20B (Detection population : 172)	6 sites (19 to 35 trees per sites)	no	RFLP	IM	home made	Knott et al. 1997

Detection, Stability (age, genetic background)	2 F2OB (Base : 95, Detection : 172)	2 sites for BASE (48+47), 6 sites for Detection (19 to 35 trees per sites)	no	RFLP, RAPD, Isozymes	IM	home made	Kaya et al. 1999
Detection, Stability (age), Developement of statistical methods	F2OB (91)	5 times (age)	no	SSR	IM	home made	Gwaze et al. 2003
Detection, Stability (age, season)	F20B (Detection population : 172)	6 sites (19 to 35 trees per sites)	no	RFLP	IM	home made	Sewell et al. 2000
Detection, Stability (age, season)	F20B (Detection population : 172)	6 sites (19 to 35 trees per sites)	no	RFLP	IM	home made	Sewell et al. 2002
Detection, Validation, Stability (genetic background)	1 F2OB Detection (D) : 172, Validation (V) : 457	D : 6 sites (19 to 35 trees per sites),	no	RFLP + EST	IM	QTL express	Brown et al. 2003

[a] F1OB: Two-generation outbred design, F2OB: Three-generation outbred design, F2S: Three-generation inbred design, F2HS : Two-generation half-sib design, CD: complex design (in this case more information is provided)
[b] When the traits were measured on one site and at one given time, 1 is indicated, otherwise more details are provided
[c] SM: single marker analysis, MM: multiple marker analysis (without map information), IM: Interval Mapping, CIM: Composite Interval Mapping, SG: Selective Genotyping, BSA : Bulk Segregant Analysis

Table 5. QTLs and marker-trait associations detected in pines: number of QTLs and phenotypic variance explained.

Species	Type of traits	Traits (abbreviation used in the article)	#QTLs [a]	%Variance explained by each QTL [b]	Note	Reference
Pinus elliottii	Adaptative traits	Aluminium tolerance	3*	15.6[all]		Kubisiak et al. 2000
Pinus elliottii var elliottii x Pinus caribaea var hondurensis	Wood quality traits and growth traits	Average branch angle (AVBRA)	0	-		Sheperd et al. 2002b
		Average number of branch per whorl (AVBRN)	3***	12-18		
		Average branch diameter (AVBRD)	2***	16-17		
		Average whorl spacing (AVWS)	2***	15-19		
		Regularity of whorl spacing (CVWS)	1***	17		
		Stem class (SC)	0	-		
		Number of distinctively large, steep angled branches observed per tree (RAM)	0	-		
		Number of "leaders" (DL)	0	-		
		Bark thickness (AVBT)	2***	11-12		
		Relative bark thickness (RBT)	0	-		
		Trunk height (HT)	3***	13-21		
		Overbark diameter at breast height (OBDBH)	1***	16		
		Underbark diameter at breast height (UBDBH)	1***	17		
		Basic density (BD)	1***	14		
(Pinus palustris Mill x Pinus elliottii Engl) x Pinus elliottii Engl	Growth traits	Total height (month 7,16,29,41)	5[lod2]	3.6-11		Weng et al. 2002
		Stem diameter	6[lod2]	4-10		

		Height increments	5^{lod2}	4.3-8.5	
		Diameter increments	5^{lod2}	4.2-11	
Pinus sylvestris	Growth, wood quality and adaptive traits	Tree height (TH)	3^{lod2}	11.5-12.2	Lerceteau et al. 2000
		Trunk diameter at breat height (DBH)	2^{lod2}	9.3-15	
		Trunk diameter 0.5m from the ground (D0.5)	2^{lod2}	12.6-13.3	
		Branch diameter of the average branch at the fourt branch level from the terminal bud (BDA)	0	-	
		Branch angle (BA)	1^{lod2}	15.9	
		Basic wood density measured with Pilodyn (PIL)	0	-	
		Frost hardiness, Critical temperature giving mean injury (CTm)	2^{lod2}	12.1-21.1	
		Frost hardiness, Estimated temperature causing slight injury to 50% of the needles (CT50)	2^{lod2}	11.3-22.7	
		Tree volume calculated with DBH (VOL)	2^{lod2}	10.7-16.7	
		Tree volume calculated with D0.5 (VOL0.5)	2^{lod2}	11-14.6	
	Adaptive traits	Bud set	3	2.1-12.7	Hurme et al. 2000
		Frost hardiness	7	3-11.1	
	Adaptive traits	Cold acclimation	$16^{p \leq 0.01}$	$9\text{-}19.5^{all}$	Yazdani et al. 2003
	Growth traits	Height growth	$11^{p \leq 0.01}$	$10.8\text{-}30.8^{all}$	
Pinus pinaster	Growth traits	Total height week 15	2**	7-12	Plomion et al. 1996a
		Total height week 38	3**	6.2-11.5	
		Total height week 92	1**	10	
		Fertile zone length (LF)_ cycle 5	1**	19.6	
		Fertile zone length (LF) adjusted to NSU_ cycle 5	1**	11.8	

	Number of stem units in the fertile zone (NSU)_ cycle 5	1**	20.4		
	Mean stem unit length (MSUL)_ cycle 5	1**	6		
	Germination date (GERMD)	2**	15-17		
	Hypocotyl length (Lhypo)	2**	9.5-12.7		
	Megagametophyte weight (MW)	2**	6-8.6		
Adaptive and growth traits	Delta ^{13}C	3*** + 5**	4.7-12.4	pop : id to Pot et al 2006 and Pot 2004	Brendel et al. 2002
	Ring width	1*** + 5**	5.9-18.1		
Wood quality and growth traits	Total height	0	-	pop : id to Brendel et al 2002 and Pot 2004	Pot et al. 2006
	Mean density of all the rings	2**	4.6-4.8		
	Density estimated by Pilodyn penetration	2** + 1***	3.7-8.8		
	Density heterogenity	3**	5.1-8		
	Lignin content	7**	5-9		
	Water extractives content	2** +1***	5.8-8.8		
	Acetone extractive content	0	-		
	Alpha cellulose content	4**	3.7-7.8		
	Hemicellulose content	3** + 1***	4.5-8.1		
	Lignin composition	3**	5.1-9.4		
	Kraft pulping yield adjusted to Kappa number	1***	8.8		
	Kappa index	3**	4.7-6.4		
	Arithmetic fibre length	2** + 1****	4.9-12.3		
	Weighted fibre length	2**+ 1***	6.1-11.9		
	Fibre width	2** + 1***	5.4-8.2		
	Coarseness	2** + 1***	4.4-11.9		
	Curl	1** + 1***	6-10.2		
	Zero span tensile value	5** + 2****	4.9-7.3		

	Wood quality and growth traits	Alpha cellulose content	5**	5.98-12.24		Markussen et al. 2002
		Lignin content	6**	5.98-13.17		
		Pulp yield	7**	9.65-18.43		
		Extractives content	2**	4.26-13.36		
		CIE Brightness	3**	7.76-13.74		
		Mean wood density	8**	6.17-12.95		
		Minimum wood density	3**	5.09-8.58		
		Maximum wood density	6**	7.50-14.51		
		Diameter	7**	5.04-14.51		
		Height growth	3**	4.93-9.78		
	Secondary metabolism	δ3Carene content	1***	26.5		Plomion et al. 1996b
Pinus radiata	Growth traits	Stem growth index	2 $^{p \leq 0.01}$	9.23-10.56		Emebiri et al. 1997
	Wood quality trait	Wood density 1-5 years old	1***	-	pop identical to Ball 2001, LG3 only, multitrait model analysis	Kumar et al. 2000
		Wood density 6-10 year-old	0	-		
		Wood density 14 year-old	0	-		
	Wood quality trait	Wood density 1-5 year-old	1***	-	pop identical to Kumar et al. 2000, LG1 and 3 only	Ball 2001
		Wood density 14 year-old	0	-		
	Growth and wood quality traits	Diameter at breast height	2***_ 0	-		Kumar et al. 2004
		Stem straightness	2***_0	-		
		Branching cluster frequency	2***_0	-		
		Wood density	1***_0	-		
	Growth and wood quality traits	Diameter at breast height	Verification: 2 $^{p \leq 0.05}$ Bridging pop : 0	Verification: 2.2 ^{all}		Devey et al. 2004a

		Juvenile wood density	Verification : 8 $^{p \leq 0.05}$ Bridging pop : 4 $^{p \leq 0.05}$	Verification: 14.1 ^{all}		
	Biotic stress tolerance	Resistance to Dothistroma needle blight	4 $^{p \leq 0.05}$ in the detection population, 4 $^{p<0.05}$ in the validation population	detection: not estimated, validation: 1.76-4.8		Devey et al. 2004b
Pinus taeda	Wood quality trait	Wood specific gravity	5 $^{p \leq 0.05}$	23 ^{all}		Groover et al 1994
	Wood quality trait	Wood specific gravity	4**, 1***	-	re-analysis of Groover et al. 1994	Knott et al. 1997
	Growth traits	Height increment	Base: 6***, Detection: 7***	Base: 23.1-30.5^{all}, detection: 7.3-11.7^{all}	Detection population : same as Groover et al 1994, Knott 1997, Sewell 2000, 2002 and Brown et al 2003.	Kaya et al. 1999
		Diameter increment	Detection : 8	Detection : 12.5-59.5^{all}		
	Growth traits	Total height 2 years	2**	7.9-10.8		Gwaze et al. 2003
		Total height 3 years	1***	10.3		
		Total height 4 years	1***	12.2		
		Total height 5 years	1***	12.2		
		Total height 10 years	1***	10.5		
		Growth rate	1***	11.3		

Wood quality trait	Wood specific gravity (individual rings and composite rings, measured for early and latewood)	23†	5.4-15.7	same population as Groover et al. 1994 and Knott et al. 1997	Sewell et al. 2000
	Percentage of latewood (individual rings and composite rings)	16†	5.5-12.3		
	Average microfibrilar angle (individual rings and composite rings, measured for early and latewood)	7†	5.4-11.9		
Wood quality trait	Cell wall content	8†	5.3-12.7	same population as Groover et al. 1994 and Knott et al. 1997	Sewell et al. 2002
Wood quality traits	Wood specific gravity	V : 18†, U : 5†	V : 1.7-5.7, U : 1.8-4.4	Detection population : same as Groover et al 1994, Knott et al. 1997, Sewell et al. 2000, 2002	Brown et al. 2003
	Percentage of latewood (individual rings and composite rings)	V : 12†, U : 5†	V : 1.7-5.7, U : 1.8-4.4		
	Average microfibrilar angle (individual rings and composite rings, measured for early and latewood)	V : 4†, , U : 2†	V : 1.7-5.7, U : 1.8-4.4		
	Cell wall content	V : 10†	V : 1.7-5.7		

[a] ** significant at the chromosome-wise level ($p<0.05$), *** significant at the genome-wise level ($p<0.05$), [lodx] lod threshold value provided by the authors (IM), [pv ≤ xx] Significance level used for the single marker analysis, †: Unique QTLs, defined as the subset of QTLs (suggestive or significant) that map within approximately 15cM of one to another and have the same general profile for their parental and interaction effects (magnitude and direction of effects)
[b] when the % of phenotypic variance explained by each QTL was not provided, the phenotypic variance explained by all the QTLs is indicated [all]

Table 6. Transcriptome and proteome analysis in pines.

Species	***Methodology***	***Study***	***Tissue***	***Authors***
P. contorta	microarrays	Adventitious root development	roots	Brinker et al. 2004
P. halepensis, P. taeda	cDNA library, Northern blot, RT-PCR	Drought response in two pine species	Roots, needles	Sathyan 2005
P. pinaster	cDNA-AFLP, reverse Northern	Drought response	roots	Dubos and Plomion 2003
P. pinaster	cDNA-AFLP, reverse Northern	Drought response	needles	Dubos et al. 2003
P. pinaster	cDNA library	Bud phenology	buds at different developmental stages	Collada, Guevara and Cerera, unpublished results
P. pinaster	Six SSH cDNA libraries	Wood formation	early vs. late, juvenile vs. mature wood forming tissues.	Pacheco et al. 2005
P. pinaster	cDNA-AFLP; reverse Northern; RT-PCR	Wood formation	early vs. late, opposite vs. compression wood forming tissues	Le Provost et al. 2003
P. pinaster	cDNA library	Wood formation	composite xylem tissues	Cantón et al. 2003
P. pinaster	EST sequencing, microarrays, qPCR	Wood formation	early vs. late, juvenile vs. mature wood forming tissues and 8 different organs	Paiva 2006
P. pinaster	EST sequencing, macroarrays, qPCR, 2DE MS/MS	Drought response in two ecotypes	roots	Chaumeil 2006
P. pinaster	2DE, tandem MS	Wood formation	wood forming tissues	Gion et al. 2005
P. pinaster	2DE, micro-sequencing	Drought response	needles	Costa et al. 1998
P. pinaster	2DE, micro-sequencing	Wood formation	compression and opposite wood forming tissues	Plomion et al. 2000
P. pinea	Two SSH cDNA libraries	Cytokinine induction of adventitious buds	cotyledons	Alonso, Cantón, Ordás, unpublished results
P. sylvestris	cDNA library	Nitrogen and carbon assimilation	cotyledons	Avila et al. 2000
P. radiata	CAGE	Wood formation	wood forming tissues	Cato et al. 2006a
P. radiata	Three SSH cDNA libraries	Adventitious rooting capacity	rooting competent tissues	Sánchez et al. 2005
P. taeda	differential screening, northern blot	Drought response	needles, stems, roots	Chang et al. 1996
P. taeda	microarrays	Wood formation	Wood forming tissues	Yang and Loopstra 2005
P. taeda	microarrays	Drought response and recovery	needles	Watkinson et al. 2003
P. taeda	microarrays	Drought response	needles, rooted cuttings, 2 unrelated genotypes	Heath et al. 2002
P. taeda	differential-display	Response to galled tissues	healthy and galled stem	Warren, and Covert 2004
P. taeda	0.35k microarrays; qRT-PCR	Wood formation	unlignified early and latewood xylem: cambial region and developing secondary non-lignified xylem	Egertsdotter et al. 2004
P. taeda	cDNA and subtractive libraries; EST sequencing; Northern blot	Wood formation	compression and opposite wood	Allona et al. 1998
P. taeda	EST sequencing, microarray	Wood formation	different tissues and organs, including several types of differentiating xylem samples	Whetten et al. 2001
P. taeda	SAGE profiling	Wood formation	Juvenile and mature wood formatting tissues	Lorenz and Dean 2002
P. taeda	cDNA library; EST sequencing, comparative sequence analysis	Wood formation	six types of wood forming tissues	Kirst et al. 2003
P. taeda	microarray, qRT-PCR	Identification of genes preferentially expressed in *P. taeda*	differentiating xylem, needles, megagametophytes, embryo (growing radicule)	Yang et al. 2004
P. taeda	comparative sequence analysis, digital profiling	Computational analysis of transcript accumulation in xylem.	differentt typoes of wood forming tissues	Pavy et al. 2005
P. taeda	EST sequencing, in silico detection of differentially expressed genes	Drought response	roots	Lorenz et al. 2005
P. taeda, P. sylvestris, P. abies	macroarrays	Heterologous arrays	needles	van Zyl et al. 2002

Figure 1. A fluorescent *in situ* hybridization (FISH) photomicrograph of *Pinus taeda* metaphase chromosomes showing 23 of the 24 chromosomes (bar = 10 μm). The red signals detect the 18S-28S rDNA sites and the green signals detect the *Arabidopsis*-type telomere repeat sequence (*A*-type TRS) sites.

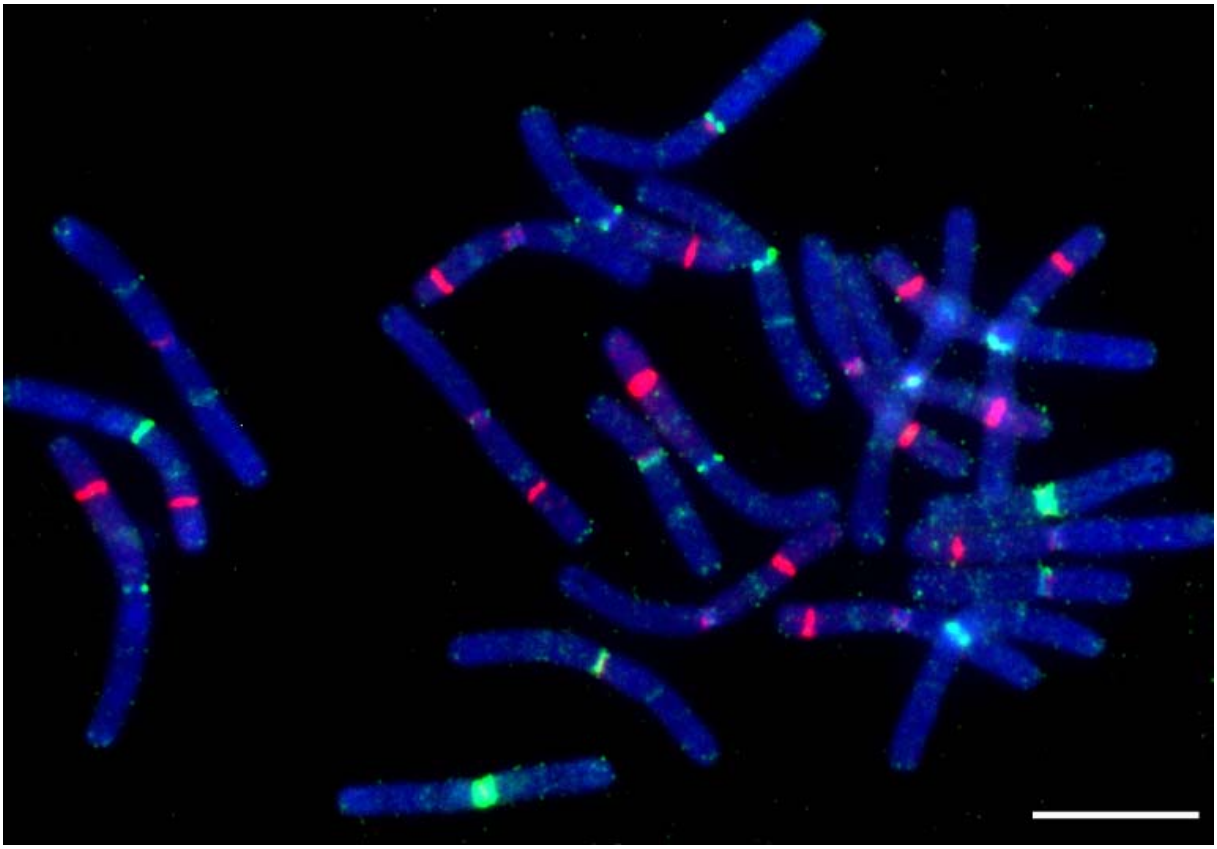

Figure 2. Synteny in the *Pinus* genus (linkage group 6): alignment of the genetic maps of five pine species.

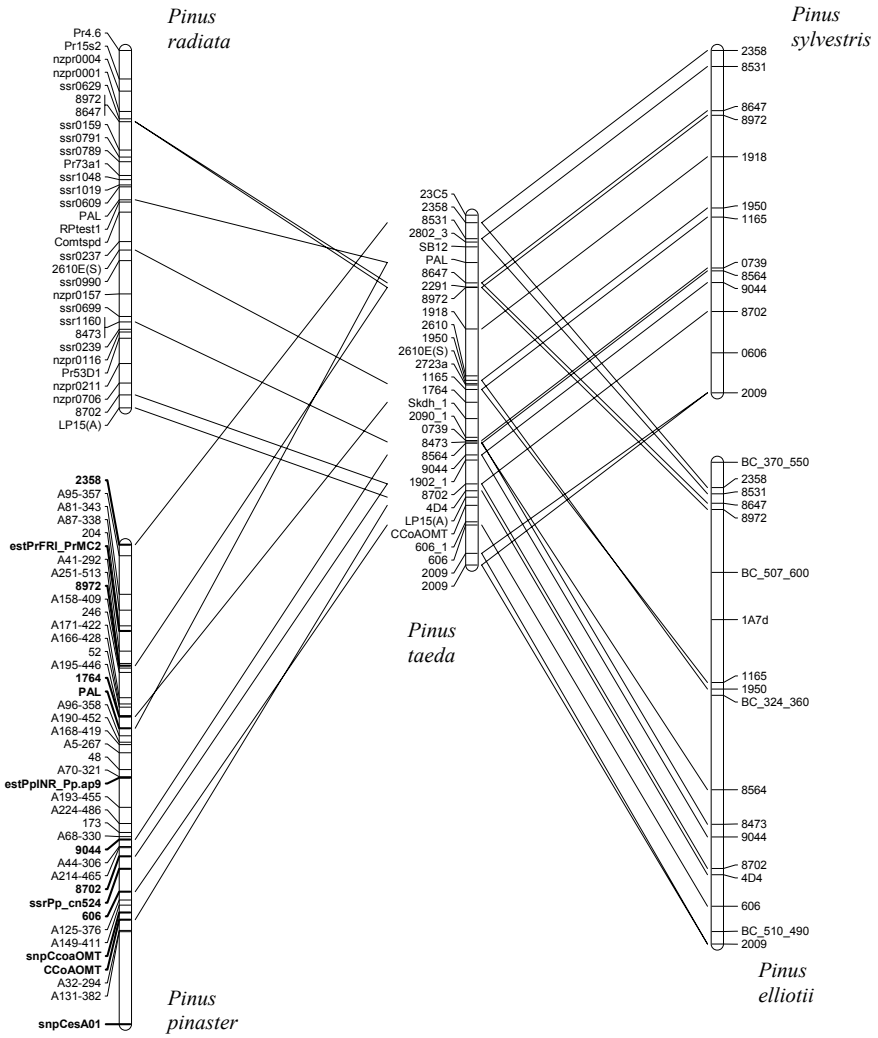

Figure 3. The relationship between genome coverage (×) provided by a BAC library and the probability (*p*) of finding a particular sequence of interest in that library. A 0.05× BAC library, like the *P. taeda* BAC library generated by Islam-Faridi et al. (1998), affords only a 4.8% chance ($p = 0.048$) of finding a sequence of interest (triangle), while a 0.32× library, like the current *P. pinaster* library (Claros et al. 2004), affords a considerably better chance ($p = 0.27$; circle). The *P. taeda* 7-56 library (Peterson et al. 2006) currently has a size of 2× ($p = 0.86$; four-point star) and will be used to isolate genes of interest once it reaches 3× coverage ($p = 0.95$; five-point star). At 10× coverage ($p = 0.9999$; eight-point star), the *P. taeda* 7-56 BAC library will be of adequate size for essentially any use including genome-wide physical mapping and genome sequencing.

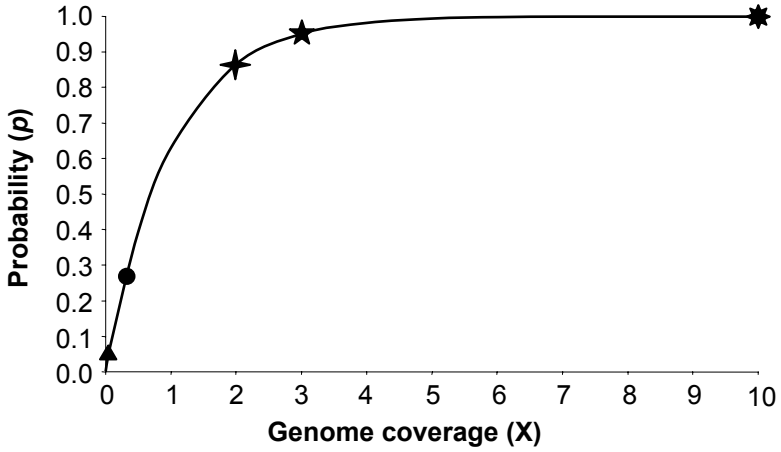

Figure 4. Constructing the *Pinus taeda* 7-56 BAC library. A Genetix QPixII robot picks 7-56 BAC colonies off of agar in a "Q-Tray." Picked clones are used to inoculate media in 384-well microtiter plates (left side of image).

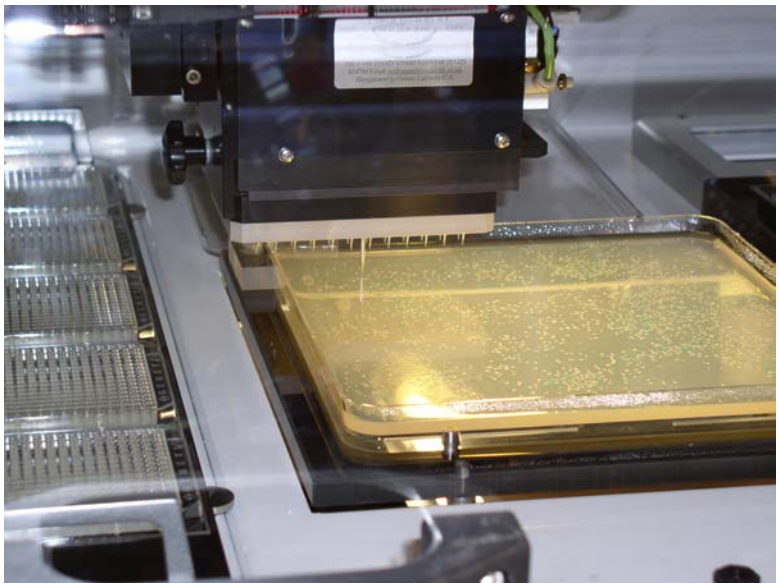

Figure 5. Pine sequence resources available at EMBL (www.ebi.ac.uk/srs), on February 28th 2006.