2 Pines

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

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

2.1 Introduction

2.1.1 History of the Genus

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 by 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 basin. 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., \textit{P. ponderosa}, \textit{P. sylvestris}). Pines are adapted to a wide range of ecological conditions: from tropical (e.g., \textit{P. merkusii}, \textit{P. kesiya}, \textit{P. tropicalis}), temperate (e.g., \textit{P. pungens}, \textit{P. thunbergii}), and subalpine (e.g., \textit{P. albicaulis}, \textit{P. cembra}) to boreal (e.g., \textit{P. pumila}) climates (Richardson and Rundel 1998, Burdon 2002). 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., \textit{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 \textit{P. longaeva} (more than 3,000 years).

Botanical Descriptions
The genus is distinguished from other members of the Pinacea family by its needlelike 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 encompassing 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 three resin ducts in the cotyledons and either one or two vascular bundles (Farjon 1984). Primary leaves are single, generally helically arranged, and acicular, and they are produced in most species only during the first growth season but for a longer period in a few species. Cataphylls, the nonchlorophyllous 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 subsulate 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, and 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 major divisions of genus *Pinus*. The trunk is usually single, erect, and columnar. 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 to the formation of scales that eventually fall off.

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 of 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 lie two separate ovules, each consisting of a cell mass protected by an integument. The micropyle through which the pollen tube penetrates is turned toward 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 2, up to even 3, 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*).

**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 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 Diplloxylon, sometimes called, respectively, soft pines and hard pines, and, more recently, Strobus and Pinus) named from type species as recommended by botanical nomenclature code; http://tolweb.org/tree?group=Pinus&contgroup=Pinaeae). The taxonomic position of a singular species with flat secondary leaves *P. kremphi* 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 subsequently been grouped into two sections *Pinus* and *Trifoliae* (Gernandt et al. 2005), the latter being called New World diplloxylon 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. lumholtzi* 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 into 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 *Gerardianae* 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 *Gerardianae* of Southeast 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 *Gerardianae*), and thick cone scales (again as in subsections *Krempfianae* and *Gerardianae*). Subsection *Nelsoniae* shows persistent fascicle sheath not found in other species of subgenus *Strobus*. Most monophyletic groups cannot be identified from unique morphoanatomical traits.

Two of the 11 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*.

**Hybridization**

Interspecific hybridization occurs in pines but is limited mostly to related species within a subsection. Some species such as *P. engelmanii*, *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...
Table 1. Systematics of the genus *Pinus*, according to Gernandt et al. (2005)

<table>
<thead>
<tr>
<th>Subgenus</th>
<th>Section</th>
<th>Subsection</th>
<th>Number of species</th>
<th>Best-known species</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinus</td>
<td>Pinus</td>
<td>Pinus</td>
<td>17</td>
<td><em>P. sylvestris</em>, <em>P. kesiya</em>, <em>P. merkusii</em></td>
<td>Eurasia, North America</td>
</tr>
<tr>
<td></td>
<td>Pinaster</td>
<td></td>
<td>7</td>
<td><em>P. pinaster</em></td>
<td>Mediterranean, Asia</td>
</tr>
<tr>
<td>Trifolii</td>
<td>Contorta</td>
<td></td>
<td>4</td>
<td><em>P. banksiana</em>, <em>P. contorta</em></td>
<td>America</td>
</tr>
<tr>
<td>Australes</td>
<td></td>
<td></td>
<td>26</td>
<td><em>P. elliottii</em>, <em>P. radiata</em>, <em>P. taeda</em></td>
<td>America</td>
</tr>
<tr>
<td>Ponderosa</td>
<td>Parrya</td>
<td>Balfouriana</td>
<td>3</td>
<td><em>P. balfouriana</em></td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Parrya</td>
<td>Cembroides</td>
<td>11</td>
<td><em>P. cembroides</em>, <em>P. culminicola</em></td>
<td>America</td>
</tr>
<tr>
<td>Quinquefoliae</td>
<td>Nelsoniae</td>
<td></td>
<td>1</td>
<td><em>P. nelsonii</em></td>
<td>Central America</td>
</tr>
<tr>
<td></td>
<td>Gerardianna</td>
<td></td>
<td>3</td>
<td><em>P. bungeana</em>, <em>P. gerardianna</em></td>
<td>Asia</td>
</tr>
<tr>
<td></td>
<td>Kremfiana</td>
<td></td>
<td>1</td>
<td><em>P. krempfi</em></td>
<td>Asia</td>
</tr>
<tr>
<td>Strobos</td>
<td>Parrya</td>
<td></td>
<td>21</td>
<td><em>P. cembra</em>, <em>P. lambertiana</em>, <em>P. strobus</em></td>
<td>America, Eurasia</td>
</tr>
</tbody>
</table>

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

2.1.2 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 (meta-centric) chromosomes and one or two pairs of smaller heterobrachial (submeta-centric) 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 A3) and DAPI (4',6-diamidino-2-phenylindole) banding that distinguished all 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 *Strobos* subgenera, respectively. Current work in *P. taeda* (subgenera *Pinus*, section *Pinus*, subsection *Australes*) (Fig. 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 cy-
Fig. 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.

togenetic map for use in physical genome mapping in the subsection *Australes* (Islam-Faridi et al. 2003, Islam-Faridi et al. 2007). Comparison of these results with 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 (Islam-Faridi et al. 2007).

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 ca. 44 pg (*P. banksiana*) to 75 pg (*P. gerardiana*) (1 pg = 960 Mb; Arumuganathan and Earle 1991) based on laser flow cytometry (Grotkopp et al. 2004). For instance, the size of the maritime pine (*P. pinaster*) genome has been estimated to be between 51 and 60 pg/2C (Chagné et al. 2002; Grotkopp et 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, including latitude of range and invasiveness (Grotkopp et al. 2004). An extremely large genome size is common to other gymnosperms (reviewed by Ohri and Khosho 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 to 30% of the pine genome corresponds to low- to single-copy sequences, while 70 to 75% corresponds to highly repeated sequences (Miksche and Hotta 1973; Rake et al. 1980; Kriebel 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 complex-
ity (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 to 30 pg; see above) is similar to Arabidopsis in having about 30,000 expressed genes and an average gene size of 2,000 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 to 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).

2.1.3 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 unfavorable site conditions; they are well suited for reforestation and for simple silviculture (monocultures and clear-felling); their wood is easily processed and utilized 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, notably _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 stabilization, 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. virgiana_ being very widely grown for Christmas trees.

2.1.4 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 favorable 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 fiber 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; Ivkovic 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.

### 2.1.5 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-Upton et al. 2000), *P. tecumhamani* (Hodge and Dvorak 1999), and *P. contorta* (Cahalan 1981). In advanced-generation breeding programs, forward selections and/or combined selection (among- and within-family) are the major sources of genetic gain (Wei et al. 1997; Rosvall 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 m³ 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 to 30 m³ ha⁻¹ yr⁻¹ on rotations of 20 to 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 over 25 years is often 25 to 30 m³ ha⁻¹ yr⁻¹ (Lamprécht 1990). However in New Zealand, growth rates of up to 50 m³ ha⁻¹ yr⁻¹ have been recorded on the best sites and as low as 11 m³ ha⁻¹ yr⁻¹ on very dry sites (Burdon and Miller 1992). The average productivity of *P. pinaster* in southern France is about 10 m³ ha⁻¹ yr⁻¹ but could reach 20 to 25 m³ 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 m³ 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 program 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 that have been clonally tested. Deployment of tested clones by CLF is being increasingly employed with *P. radiata* in New Zealand (Sorensen 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 nonadditive effects, which are not captured via sexual propagation, plus the benefits of greater uni-
formity 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).

2.1.6 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 (*GST*) averaged 0.076 (Ledig 1998). Exceptions are represented by *P. pinea* (G.G. Vendramin et al. personal communication), *Pinus resinosa*, a species that has a vast range across northeastern 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 to 27% for *P. radiata* and 22% for *P. muricata* (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; J. Duminil et al. personal communication) 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 *P. 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 *GST* 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, *GST* is nearly always larger at mitochondrial DNA markers than at chloroplast DNA markers. On the other hand, there is no significant difference between *GST* 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 allow 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 *P. pinaster* (Burban and Petit 2003), *P. sylvestris* (Robledo-Arnuncio et al. 2005), *P. canariensis* (Gómez et al. 2003), and *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; Soranzo et al. 2000; Gugerli et al. 2001), generally provide a clear picture of nonoverlapping areas colonized from different refugia (e.g., in *P. pinaster*; Burban and Petit 2003).

Neutral markers also have allowed investigating spatial genetic structure (SGS) in natural pine popu-
lations. A generally weak within-population structure has been described. *P. pinaster* showed a fine-scale structure at the seedling stage with a patch size of ca. 10 m 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 postdispersal 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 *P. strobos* (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 *P. 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 single nucleotide polymorphisms (SNPs), which are 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 depend 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 of basic population genetic parameters such as the pattern of nucleotide diversity and LD for each particular species and candidate gene set.

First estimates indicate that nucleotide diversity varies considerably between plant species. Interestingly, the pines (e.g., *P. sylvestris*, Dvornyk et al. 2002; García-Gil et al. 2003; *P. taeda*, Brown et al. 2004; González-Martínez et al. 2006a; *P. pinaster*, Pot et al. 2005a) are not among 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 2.7 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 (*Fst*) is higher for the gene-targeted marker with respect to 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 *P. 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 to 2,000 bp in *Pinus taeda* (Brown et al. 2004; González-Martínez et al. 2006a), *Pinus sylvestris* (Dvornyk et al. 2002 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.
2.2 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
- 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.

2.2.1 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, quantitative trait loci analysis, survey of genetic diversity, forensic applications). The purpose of this section is to briefly review the different types of molecular marker techniques that have been developed in pines and used for genetic mapping applications. We will not present the details 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.

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 other proteins published in sequence databases, therefore providing functional markers expressed in the tissues analyzed. Two-dimensional protein markers were developed only in *P. pinaster* (reviewed in Cánovas et al. 2004 and Plomion 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 quantitative trait loci (QTL) detection.

RFLPs
Restriction fragment length polymorphism (RFLP) markers 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).

RAPDs and AFLPs
In the mid-1990s PCR-based multiplex DNA fingerprinting techniques provided very powerful tools to generate dense linkage maps in a short period of time. Random amplified polymorphic DNA (RAPD; Williams et al. 1990) and then amplified fragment length polymorphism (AFLP; 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 Sect. 2.2.2.), these markers tremendously boosted up genetic analysis in most forest tree species including pines (reviewed in Cervera et al. 2000b).

Nuclear Microsatellites
In contrast to other plant species, few polymorphic single-copy nuclear microsatellite markers or simple sequence repeats (SSRs) have been reported in pines (reviewed in Chagné et al. 2004). The genome
<table>
<thead>
<tr>
<th>Feature</th>
<th>Protein markers</th>
<th>DNA markers</th>
<th>PCR-based</th>
<th>Hybridization-based</th>
<th>2D-PAGE</th>
<th>Isozyme</th>
<th>RFLP</th>
<th>PCR-based</th>
<th>AFLP</th>
<th>SSR</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue quantity (g)</td>
<td>0.5–2</td>
<td>0.25–0.5</td>
<td>0.02</td>
<td></td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>DNA quantity (mg)</td>
<td>na</td>
<td>0.02</td>
<td>0.02</td>
<td>0.005–0.02</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>DNA quality</td>
<td>na</td>
<td>1</td>
<td>10</td>
<td></td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Allelism and mode of inheritance</td>
<td>Multiallelic</td>
<td>Biallelic</td>
<td>Biallelic</td>
<td></td>
<td>Biallelic</td>
<td>Biallelic</td>
<td>Biallelic</td>
<td>Biallelic</td>
<td>Biallelic</td>
<td>Biallelic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>or dominant</td>
<td>codominant</td>
<td>codominant</td>
<td></td>
<td>dominant</td>
<td>codominant</td>
<td>dominant</td>
<td>codominant</td>
<td>codominant</td>
<td>codominant</td>
<td></td>
</tr>
<tr>
<td>Number loci analysed per assay</td>
<td>1–5</td>
<td>1–6</td>
<td>2–30</td>
<td>20–100</td>
<td>1–10</td>
<td>1–several hundred</td>
<td>(multiplexing)</td>
<td>(multiplexing)</td>
<td>(multiplexing)</td>
<td>(multiplexing)</td>
<td></td>
</tr>
<tr>
<td>Polymorphism</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Development cost</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Cost per analysis</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Amenable to automation</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Technical demand</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Reproducibility</td>
<td>High&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Medium&lt;sup&gt;b&lt;/sup&gt;</td>
<td>High</td>
<td>Low-high&lt;sup&gt;c&lt;/sup&gt;</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

**Applications**

<table>
<thead>
<tr>
<th></th>
<th>Certification</th>
<th>Diversity</th>
<th>Phylogeny</th>
<th>Mapping</th>
<th>Comparative mapping</th>
<th>QTL analysis</th>
<th>Association studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
</tr>
</tbody>
</table>

<sup>a</sup> Sequence dependent (exon vs UTR and intron)

<sup>b</sup> Under the same environmental conditions, type of tissue, and age

<sup>c</sup> Between laboratories vs single laboratory

<sup>d</sup> To construct dense maps

<sup>e</sup> Used as control loci
structure of these species, characterized by a large physical size, with a large amount of repeated sequence (Sect. 2.1.2), has been the main obstacle to the development of useful markers using classical SSR-enriched library approaches (e.g., Auckland et al. 2002; Guevara et al. 2005a; C.S. Echt and C.D. 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, nonspecific amplification, multibanding patterns, or lack of polymorphism (Echt and Nelson 1997; 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). P. 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, 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).

**EST Polymorphisms**

With the availability of sequence data obtained by random sequencing of pine cDNAs (Sect. 2.5.2), there is now a clear trend toward 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 (denaturing gradient gel electrophoresis, DGGE; Myers et al. 1987), conformation (single-strand conformation polymorphisms, SSCP; Orita et al. 1989) under specific polycrylamide 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 1,400 candidate SNPs in P. 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 LD, 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 P. taeda and P. pinaster expressed sequence tags (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.

**2.2.2 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 (http://linkage.rockefeller.edu/soft/list.html).
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 that is 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.

Diploid- or Full-Sib-Based Mapping Strategy

Different strategies have been followed to construct genetic linkage maps of Pinus: the “pseudotestcross strategy,” the “F2 inbred model,” and the “three-generation outbred model.” The pseudotestcross 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 F1 progeny as in a testcross. Grattapaglia and Sederoff (1994) introduced the term two-way pseudotestcross to define this mapping strategy, where two independent genetic linkage maps are constructed by analyzing the cosegregation 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 outcrossed tree species such as pines. The F2 inbred model is based on a three-generation pedigree for which the grandparents are treated as inbred lines. In the F2 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 pseudotestcross 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). These segregation data are 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, thereby 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).

2.2.3 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. In addition, some maps have been published together with QTL studies and will be found in the references cited in Sect. 2.3.

Maritime Pine

Linkage maps of the Maritime pine (Pinus pinaster Ait.) genome were first constructed by analyzing the cosegregation 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-locus 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 F2) 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 latter 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).
Table 3. Summary of genetic linkage maps of *Pinus* species

<table>
<thead>
<tr>
<th>Species pedigree</th>
<th>Number of LGs (+pairs, triplets)</th>
<th>Total number of linked marker loci</th>
<th>RAPD, AFLP</th>
<th>SSR</th>
<th>EST-based</th>
<th>Others</th>
<th>Total genetic distance (cM)</th>
<th>Genome saturation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus pinaster</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haploid (1 tree)</td>
<td>Not estimated</td>
<td>119</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>119</td>
<td>(2D proteins)</td>
<td>Not estimated</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Haploid (8 trees)</td>
<td>Not estimated</td>
<td>65</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>65</td>
<td>(2D proteins)</td>
<td>530</td>
<td>25%</td>
</tr>
<tr>
<td>Haploid (H12 self)</td>
<td>13(+5)</td>
<td>263</td>
<td>251</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1223</td>
<td>90%</td>
</tr>
<tr>
<td>Haploid (H12 open pollinated)</td>
<td>12</td>
<td>463</td>
<td>436</td>
<td>–</td>
<td>–</td>
<td>27</td>
<td>(2D proteins)</td>
<td>1860</td>
<td>100%</td>
</tr>
<tr>
<td>Haploid (H12 self)</td>
<td>11</td>
<td>94</td>
<td>94</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1169</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Haploid (H12 self)</td>
<td>11</td>
<td>94</td>
<td>94</td>
<td>–</td>
<td>–</td>
<td>36</td>
<td>(2D protein)</td>
<td>1873</td>
<td>93.4%</td>
</tr>
<tr>
<td>F1 - AFOCEL</td>
<td>12</td>
<td>759</td>
<td>738</td>
<td>14</td>
<td>7</td>
<td>–</td>
<td></td>
<td>1994</td>
<td>Not estimated</td>
</tr>
<tr>
<td>9.103.3 × 10.159.3</td>
<td>12</td>
<td>620</td>
<td>620</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1441</td>
<td>Not estimated</td>
</tr>
<tr>
<td>9.103.3 × 10.159.3</td>
<td>12</td>
<td>326</td>
<td>276</td>
<td>–</td>
<td>50</td>
<td>–</td>
<td></td>
<td>1638</td>
<td>Not estimated</td>
</tr>
<tr>
<td><em>Pinus taeda</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Various pedigrees</td>
<td>Not estimated</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>20</td>
<td>isozymes</td>
<td>Not estimated</td>
<td>Not estimated</td>
</tr>
<tr>
<td>Base pedigree</td>
<td>20</td>
<td>75</td>
<td>458</td>
<td>–</td>
<td>–</td>
<td>90</td>
<td>6</td>
<td>1727</td>
<td>97%</td>
</tr>
<tr>
<td>Haploid (Tree 10-5)</td>
<td>16</td>
<td>458</td>
<td>458</td>
<td>–</td>
<td>–</td>
<td>257</td>
<td>12</td>
<td>1300</td>
<td>76.4%</td>
</tr>
<tr>
<td>Base and QTL pedigrees</td>
<td>18</td>
<td>357</td>
<td>67</td>
<td>–</td>
<td>257</td>
<td>–</td>
<td></td>
<td>1741</td>
<td>100%</td>
</tr>
<tr>
<td>Tree 7-56</td>
<td>12</td>
<td>508</td>
<td>508</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1281</td>
<td>75.4%</td>
</tr>
<tr>
<td>Base and QTL pedigrees</td>
<td>20</td>
<td>265</td>
<td>253</td>
<td>12</td>
<td>–</td>
<td>–</td>
<td></td>
<td>795</td>
<td>46.7%</td>
</tr>
<tr>
<td>Base pedigree</td>
<td>15</td>
<td>51</td>
<td>51</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>2638</td>
<td></td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49-2</td>
<td>14</td>
<td>261</td>
<td>261</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>2638</td>
<td></td>
</tr>
<tr>
<td>F1 of P315 × E1101</td>
<td>12(+4)</td>
<td>179</td>
<td>179</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1645</td>
<td>98%</td>
</tr>
<tr>
<td>F1 of AC3065 × Y3038</td>
<td>12</td>
<td>188</td>
<td>188</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>1681</td>
<td></td>
</tr>
<tr>
<td>F1 of E635 × E1101</td>
<td>12(+3)</td>
<td>260</td>
<td>194</td>
<td>4</td>
<td>61</td>
<td>–</td>
<td></td>
<td>1314</td>
<td>66–85%</td>
</tr>
</tbody>
</table>
### Table 3. (continued)

<table>
<thead>
<tr>
<th>Species pedigree</th>
<th>Number of LGs (+pairs, triplets)</th>
<th>Total number of linked marker loci</th>
<th>RAPD, AFLP</th>
<th>SSR</th>
<th>EST-based</th>
<th>Others</th>
<th>Total genetic distance (cM)</th>
<th>Genome saturation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus radiata</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-generation</td>
<td>22</td>
<td>208</td>
<td>41</td>
<td>2</td>
<td>165</td>
<td>–</td>
<td>1382</td>
<td>Not estimated</td>
<td>Devey et al. 1996</td>
</tr>
<tr>
<td>Haploid (full-sib seed)</td>
<td>14</td>
<td>267</td>
<td>267</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1665</td>
<td>93%</td>
<td>Emebiri et al. 1998</td>
</tr>
<tr>
<td>S1</td>
<td>19</td>
<td>172</td>
<td>168</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>1117</td>
<td>56%</td>
<td>Kuang et al. 1999b</td>
</tr>
<tr>
<td>Pseudotestcross</td>
<td>20</td>
<td>235</td>
<td>224</td>
<td>11</td>
<td>–</td>
<td>–</td>
<td>1414</td>
<td>85%</td>
<td>Wilcox et al. 2001a</td>
</tr>
<tr>
<td>(Parent 850.055)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudotestcross</td>
<td>21</td>
<td>194</td>
<td>185</td>
<td>9</td>
<td>–</td>
<td>–</td>
<td>1144</td>
<td>77%</td>
<td>Wilcox et al. 2001a</td>
</tr>
<tr>
<td>(Parent 850.096)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two full-sib families</td>
<td>12</td>
<td>311</td>
<td>–</td>
<td>213</td>
<td>98</td>
<td>–</td>
<td>1352</td>
<td>Not estimated</td>
<td>Wilcox et al. 2004</td>
</tr>
<tr>
<td><em>Pinus elliottii</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree 8-7</td>
<td>13(+9pairs)</td>
<td>73</td>
<td>73</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>782</td>
<td>64–75%</td>
<td>Nelson et al. 1993</td>
</tr>
<tr>
<td>18-62 × 8-7</td>
<td>17(+12pairs)</td>
<td>129</td>
<td>129</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1146</td>
<td>Not estimated</td>
<td>Kubisiak et al. 2000</td>
</tr>
<tr>
<td>D4PC40 × D4PC13</td>
<td>15</td>
<td>154</td>
<td>63</td>
<td>–</td>
<td>41 (RFLPs)</td>
<td>–</td>
<td>1115</td>
<td>Not estimated</td>
<td>Brown et al. 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45</td>
<td>5 isozyme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus palustris</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree 3-356</td>
<td>16(+6pairs)</td>
<td>133</td>
<td>133</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1635</td>
<td>85%</td>
<td>Nelson et al. 1994</td>
</tr>
<tr>
<td><em>P. elliottii and P. palustris hybrids</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 of - 3-356 × H-28</td>
<td>♀: 18(+3 pairs)</td>
<td>122</td>
<td>122</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1368</td>
<td>81%</td>
<td>Kubisiak et al. 1995</td>
</tr>
<tr>
<td></td>
<td>♂: 13(+6 pairs)</td>
<td>91</td>
<td>91</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>953</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>BC1 of 488 × 18-27</td>
<td>♀: 17</td>
<td>133</td>
<td>133</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1338</td>
<td>91%</td>
<td>Weng et al. 2002</td>
</tr>
<tr>
<td></td>
<td>♂: 19</td>
<td>83</td>
<td>83</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>995</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td><em>P. elliottii and P. caribea var. hondurensis hybrids</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 of 2PEE1-102 ×</td>
<td>♀: 24</td>
<td>125</td>
<td>117</td>
<td>8</td>
<td>–</td>
<td>–</td>
<td>1548</td>
<td>82%</td>
<td>Shepherd et al. 2003</td>
</tr>
<tr>
<td>1PCH1-63</td>
<td>♂: 25</td>
<td>155</td>
<td>145</td>
<td>10</td>
<td>–</td>
<td>–</td>
<td>1823</td>
<td>88%</td>
<td></td>
</tr>
</tbody>
</table>
The two-way pseudotrestcross 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 (N. de María and M.T. 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 between male and female recombination was also reported in other pine species (Moran et al. 1983; Groover et al. 1995; Sewell et al. 1999).

**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 (LGs) (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 revealed 20 LGs. Genomic mapping was successfully used by Wilcox et al. (1996) to define a single gene and revealed 20 LGs. 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 developed 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 1,300 cM of genetic distance on 18 LGs. 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 LGs equaling the basic number of chromosomes and about 1,700 cM of genetic distance. Their analysis suggested that this distance saturated the genome, in slight contrast to an earlier estimate of 2,000 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 9 were mapped in both species; however, these codominant markers along with several codominant AFLP markers were useful in defining homeologous LGs 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 on 15 LGs, 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.

**Radiata Pine**

Over the past 15 years, a range of DNA-marker-based linkage maps have been constructed for this species. Wilcox (1997) briefly reviewed mapping studies undertaken up till that date. Although results were summarized 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 LGs and covered 1,382 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 LGs equal to the haploid number of chromosomes and the remainder ranging between 14 and 22 LGs. Total map length estimates were undertaken in three studies ranging between 1,978 and 3,000 cM. Subsequent to these earlier studies a number of maps were published, some of which were included in Wilcox’s 1997 review. Using 222 RAPD markers to genotype 93 megagametophytes, Emeliri et al. (1998) constructed a linkage map that covered 14 LGs and spanned a total distance of 1,665 cM. Kuang et al. (1999) developed a map constructed using megagametophytes of 198 *S1* seeds that had been genotyped with 168 RAPD and four microsatellite markers. The resulting map
consisted of 19 LGs, covering 1,116.7 cM, which was estimated to cover 56% of the genome. Because this was constructed using an S1 family, elevated levels of segregation distortion were observed. Wilcox et al. (2001a) 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 1,414 and 1,334 cM in 20 and 21 LGs, respectively. These maps have subsequently been added to, using over 300 SSR and EST markers, and have been reduced to 12 LGs (Wilcox et al. 2004). Devy et al. (1999) published a comparative map of radiata and loblolly pine based on RFLP, SSR, and RAPD loci and showed that 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).

**Scots Pine**

Early mapping work in *P. 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 had been very low (Karhu et al. 1996). Thus, the next mapping efforts were based on RAPD markers segregating in haploid megagametophytes. Yazdani et al. (1995) mapped 261 markers in 14 LGs in a tree that is part of the breeding program. Hurme et al. (2000) made a low-coverage RAPD map for an F1 tree that was a result of north \( \times \) 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 1,645 and 1,681 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 F1 progeny of a north \( \times \) south cross. This map also contained markers that had been previously developed for *P. 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 LGs in the two species. More markers were later added to this map (Pyhäjärvi et al. unpublished).

**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 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, 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), to 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 F1 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 promising for further comparative analyses across these species (Shepherd et al. 2002a; C.D. Nelson and C.S. Echt, pers. comm.).
Other Pines
Genetic linkage maps have also been 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 on RAPDs and AFLPs.

2.2.4 Genetic vs. Physical Size and Practical Implication

These mapping studies have led to the conclusion that the total genetic distance of the pine genome is around 2,000 cM (Gerber and Rodolphe 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 $13 \times 10^6$ 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 will be discussed in Sects. 2.3 and 2.4, 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 crossings over per chromosome was found to be highly conserved between both genera: 2 to 4 chiasmata per bivalent (1 chiasma = 50 cM; Ott 1991). A comparison of genome lengths among evolutionary divergent pines found *P. pinaster*, *P. palustris*, 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.

2.2.5 Comparative Mapping: Toward 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 have also been 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 (colinearity) among 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), 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 transferring 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 synteny 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 Family Pinaceae, which included pines, along with other important conifers such as spruces and firs. These markers made it possible to assign LG 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 (Chagné et al. 2003), and 9 
out of 12 between P. sylvestris and P. taeda (Komu-
lainen et al. 2003). About 30 or 40 ESTP markers 
were proved to be useful in demonstrating large 
areas of synteny between each species pair. While 
this comparison was only of low density, these 
pioneering studies suggested that pine genomes 
did not show any apparent chromosomal rear-
rangement. They also provided an indication that 
gene content and gene order is conserved, as is 
illustrated for LG 6 in Fig. 2. Current efforts are 
being expended 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 

As more conifer ESTs become available in public 
databases (329,531 in Pinus taeda, 132,531 in Picea 
glauc, 27,283 in Pinus pinaster, 28,170 in Picea engel-
mannii × Picea sitchensis, 80,789 in Picea sitchensis, 
6,808 in Pseudotsuga menziesii, 7,639 in Cryptome-
ria japonica: EMBL 19 March 2006), a computational 
approach could be used for in silico development of 
putative orthologous EST-based markers (Fulton et al. 
2002), as was recently illustrated between loblolly pine 
and Douglas fir by Krutovsky et al. (2004). Such re-
sources should help to define the precise syntenic re-
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 con-
served QTLs between the two species (Chagné et al. 2003). Those are QTLs for wood density and wood chemical composition traits located on LGs 3 and 8, respectively. The same observation has been made in *P. radiata* (Telfer et al. 2006). Preliminary studies indicate wood-density QTLs are colocating 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 that 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 the 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.

### 2.3 Genetic Architecture of Complex Traits

Pine tree improvement is hampered by different inherent characteristics: (1) the time needed to reach sexual maturity, (2) the time lag required to evaluate field performance (e.g., growth; Kremer 1992), and (3) in some cases, the cost of phenotyping (e.g., wood-quality-related traits). This makes breeding of these species a slow process compared to that of 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 to 0.3, and only slightly higher (0.3 to 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, and 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 QTLs (Geldermann 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 in DNA marker-based technologies since the 1980s have now made it possible to genotype hundreds of genetic markers to construct dense genetic linkage maps (Sect. 2.2.3) 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 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, a perspective on the identification of diagnostic markers for pine-tree breeding will be discussed.

#### 2.3.1 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 development 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 the high rate of polymorphism encountered in pines and the relative lack of stability of QTLs in different genetic backgrounds, methods aimed 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 and transcriptome and proteome analysis, together with studies of molecular evolution, will be highlighted.

**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 pseudotestcross 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 is collapsed into only two genotypic classes, thereby reducing the robustness of the analysis. Of course, if codominant 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 that display a divergent phenotypic effect. This three-generation full-sib structure is typically designed for QTL analysis under an outbred model and has been used extensively (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** An extension of the pseudotestcross 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, provided that both parents of the full-sib were 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 experimentally.

**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 with one marker at a time, on the genotypic classes defined by a single marker. This
### Table 4. QTL and marker-trait association detected in pines: pedigree structure and methods

<table>
<thead>
<tr>
<th>Species</th>
<th>Objective</th>
<th>Population (size)</th>
<th>Replications</th>
<th>Clonal replicates</th>
<th>Genetic markers</th>
<th>QTL analysis (analytical model)</th>
<th>Software</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus elliottii</em></td>
<td>Detection</td>
<td>F1OB (186)</td>
<td>1</td>
<td>No</td>
<td>RAPD</td>
<td>SG (54:27 genotypes from each tail), SM, MM</td>
<td>SAS</td>
<td>Kubisiak et al. 2000</td>
</tr>
<tr>
<td><em>Pinus elliottii var elliottii ×</em></td>
<td>Detection</td>
<td>F1OB (89)</td>
<td>2 sites (60,29)</td>
<td>No</td>
<td>AFLP, SSR</td>
<td>SM, CIM</td>
<td>QTLcartographer</td>
<td>Shepherd et al. 2005</td>
</tr>
<tr>
<td><em>Pinus caribaea var hondurensis</em></td>
<td>Detection</td>
<td>F1OB (258)</td>
<td>3 sites (82,83,93)</td>
<td>No</td>
<td>–</td>
<td>SM, IM</td>
<td>SAS, MAPMAKER/QTL</td>
<td>Weng et al. 2002</td>
</tr>
<tr>
<td><em>Pinus palustris Mill ×</em></td>
<td>Detection, stability (age)</td>
<td>F1OB (258)</td>
<td>3 sites (82,83,93)</td>
<td>No</td>
<td>–</td>
<td>SM, IM</td>
<td>SAS, MAPMAKER/QTL</td>
<td>Lerceteau et al. 2000</td>
</tr>
<tr>
<td><em>Pinus elliottii Engl</em> × <em>Pinus elliottii Engl</em></td>
<td>Detection, stability (age)</td>
<td>F1OB (258)</td>
<td>3 sites (82,83,93)</td>
<td>No</td>
<td>–</td>
<td>SM, IM</td>
<td>SAS, MAPMAKER/QTL</td>
<td>Lerceteau et al. 2000</td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td>Detection</td>
<td>F1OB (286)</td>
<td>1</td>
<td>No</td>
<td>AFLP</td>
<td>IM</td>
<td>Qgene</td>
<td>Hurme et al. 2000</td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td>Detection, Stability (time replicate)</td>
<td>F2HS (258)</td>
<td>2</td>
<td>No</td>
<td>RAPD</td>
<td>IM</td>
<td>SAS, MULTIMAPPER/OUTBRED</td>
<td>Lerceteau et al. 2000</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection</td>
<td>F1OB (286)</td>
<td>1</td>
<td>No</td>
<td>RAPD, EST</td>
<td>IM</td>
<td>MAPMAKER/QTL</td>
<td>Plomion et al. 1996a</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection, Stability (Age)</td>
<td>F2S (120)</td>
<td>3 (age)</td>
<td>No</td>
<td>RAPD, EST</td>
<td>IM</td>
<td>QTLcartographer</td>
<td>Costa 1999</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection (age, season)</td>
<td>F2OB (186)</td>
<td>1</td>
<td>No</td>
<td>RAPD, AFLP, EST</td>
<td>IM</td>
<td>MULTIQTL</td>
<td>Brendel et al. 2002</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection</td>
<td>F2OB (186)</td>
<td>1</td>
<td>No</td>
<td>RAPD, EST</td>
<td>IM</td>
<td>MULTIQTL</td>
<td>Pot et al. 2005</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection</td>
<td>F1OB (80)</td>
<td>1</td>
<td>No</td>
<td>RAPD, EST, SSR</td>
<td>IM</td>
<td>SAS</td>
<td>Markussen et al. 2002</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection (age, stability)</td>
<td>F1OB (80)</td>
<td>1</td>
<td>No</td>
<td>RAPD, EST, SSR</td>
<td>IM</td>
<td>MULTIQTL</td>
<td>Pot 2004</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Detection</td>
<td>F2S (120)</td>
<td>1</td>
<td>No</td>
<td>RAPD</td>
<td>IM</td>
<td>MAPMAKER/QTL</td>
<td>Plomion et al. 1996b</td>
</tr>
</tbody>
</table>

---

*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>
<thead>
<tr>
<th>Species</th>
<th>Objective</th>
<th>Population (size)</th>
<th>Replications&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Clonal replicates</th>
<th>Genetic markers</th>
<th>QTL analysis (analytical model)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Software</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus radiata</em></td>
<td>Detection, stability (age)</td>
<td>F1OB (174)</td>
<td>1</td>
<td>No</td>
<td>RAPD</td>
<td>SM, BSA (9 vs 9)</td>
<td>-</td>
<td>Emebiri et al. 1997</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age)</td>
<td>F1OB (80-93)</td>
<td>3 times (age)</td>
<td>No</td>
<td>RAPD, AFLP, SSR</td>
<td>IM, MTM</td>
<td>-</td>
<td>Kumar et al. 2000</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age)</td>
<td>F1OB (93)</td>
<td>2 times (age)</td>
<td>No</td>
<td>RAPD, AFLP, SSR</td>
<td>Bayesian approach</td>
<td>Spluss</td>
<td>Ball 2001</td>
</tr>
<tr>
<td></td>
<td>Detection, validation, stability (genetic background)</td>
<td>CD</td>
<td>1</td>
<td>No</td>
<td>SSR</td>
<td>-</td>
<td>-</td>
<td>Kumar et al. 2004</td>
</tr>
<tr>
<td></td>
<td>Detection, validation, stability (genetic background)</td>
<td>3 unrelated F1OB</td>
<td>2 times (age)</td>
<td>No</td>
<td>RFLP and SSR</td>
<td>SG (Juvenile wood density: 50 vs 50;</td>
<td>GENSTAT</td>
<td>Devey et al. 2004a</td>
</tr>
<tr>
<td></td>
<td>Detection, validation, stability (genetic background)</td>
<td>6 related F2OB</td>
<td>1 site for the</td>
<td>No</td>
<td>RFLP + SSR</td>
<td>SM within family + SM among families</td>
<td>GENSTAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Detection, validation, stability (genetic background)</td>
<td>62 F2OB (202)</td>
<td>detection population</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Govever</td>
<td>Kroover et al. 1994</td>
</tr>
<tr>
<td></td>
<td>Verification: 1 F2OB (400) detection population</td>
<td>6 related F2OB</td>
<td>1 site for the</td>
<td>No</td>
<td>RFLP</td>
<td>SG, SM</td>
<td>Home made</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 F2OB (202) detection population</td>
<td>6 sites (19 to 35 trees per sites)</td>
<td>No</td>
<td>RFLP</td>
<td>IM</td>
<td></td>
<td>Home made</td>
<td>Knott et al. 1997</td>
</tr>
<tr>
<td><em>Pinus taeda</em></td>
<td>Detection, stability (age, genetic background)</td>
<td>2 F2OB (Base : 95, Detection : 172)</td>
<td>2 sites for BASE (48+47), 6 sites for Detection (19 to 35 trees per sites)</td>
<td>No</td>
<td>RFLP, RAPD, Isozymes</td>
<td>IM</td>
<td>Home made</td>
<td>Kaya et al. 1999</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age, genetic background)</td>
<td>F2OB (91)</td>
<td>5 times (age)</td>
<td>No</td>
<td>SSR</td>
<td>IM</td>
<td>Home made</td>
<td>Gwaze et al. 2003</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age, development of statistical methods)</td>
<td>F2OB (Detection population : 172)</td>
<td>6 sites (19 to 35 trees per sites)</td>
<td>No</td>
<td>RFLP</td>
<td>IM</td>
<td>Home made</td>
<td>Sewell et al. 2000</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age, season)</td>
<td>F2OB (Detection population : 172)</td>
<td>6 sites</td>
<td>No</td>
<td>RFLP</td>
<td>IM</td>
<td>Home made</td>
<td>Sewell et al. 2002</td>
</tr>
<tr>
<td></td>
<td>Detection, stability (age, season)</td>
<td>1 F2OB (Detection (D) : 172, validation (V) : 457)</td>
<td>6 sites (19 to 35 trees per sites)</td>
<td>No</td>
<td>RFLP + EST</td>
<td>IM</td>
<td>QTL express</td>
<td>Brown et al. 2003</td>
</tr>
</tbody>
</table>
method suffers from several limitations. First, it does not provide any information regarding the location of the QTLs 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 et al. 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 CIM, developed by Zeng (1993a, b) and Jansen (1993), which combines IM 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 differed from the previous methods through the implementation of multiple-QTL models. The selection of the model (i.e., QTL number, position, effect, and interactions) that 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 reevaluated. MIM also allows the evaluation of QTL epistasis. However, it is important to note that, although this strategy provides significant improvements over CIM, it does not allow the detection of nonsignificant QTLs 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 that 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, 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 bulked 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 QTLs with true effects being left undetected, but in some circumstances (but to a lesser extent) might even lead to the detection of false QTLs. Without the possibility of developing F3 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., Scotti-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 capital investments and technical expertise needed to clonally propagate pines and the wide variability noted among specific genotypes in their ability to produce rooted cuttings.

2.3.2 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* (Kuang et al. 1999a, b; Remington and O’Malley 2000a, b; Williams et al. 2001).

**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* Doug. (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).

**Quantitative Traits**

Growth has been the most studied trait in QTL mapping studies in pines. To date, 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 involving adaptation to the environment have also been studied, including tree response to heavy metal, drought, and 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).

**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) have been detected. This number is in general smaller than in annual crops and has led to a smaller proportion of the phenotypic variation being explained (Table 5). As stated in the chapter on Eucalypts, “the limited power to detect QTL in forest trees compared to crop species may be due to the high environmental and developmental variation in tree plantations, as well as to the small size of the analyzed populations” (mainly between 91 and 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 threefold 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- to 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 differ-
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of trait</th>
<th>Trait (abbreviation used in the article)</th>
<th>#QTLs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>%Variance explained by each QTL&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Note</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus elliottii</em></td>
<td>Adaptative traits</td>
<td>Aluminium tolerance</td>
<td>3&lt;sup&gt;*&lt;/sup&gt;</td>
<td>15.61&lt;sup&gt;all&lt;/sup&gt;</td>
<td></td>
<td>Kubisiak et al. 2000</td>
</tr>
<tr>
<td><em>Pinus elliottii</em></td>
<td>Wood-quality traits and growth traits</td>
<td>Average branch angle (AVBRA)</td>
<td>0</td>
<td>–</td>
<td></td>
<td>Sheperd et al. 2002b</td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var elliottii × Pinus elliottii Engl*</td>
<td></td>
<td>Average number of branch per whorl (AVBRN)</td>
<td>3&lt;sup&gt;***&lt;/sup&gt;</td>
<td>12–18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Average branch diameter (AVBRD)</td>
<td>2&lt;sup&gt;***&lt;/sup&gt;</td>
<td>16–17</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Average whorl spacing (AVWS)</td>
<td>2&lt;sup&gt;***&lt;/sup&gt;</td>
<td>15–19</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Regularity of whorl spacing (CVWS)</td>
<td>1&lt;sup&gt;***&lt;/sup&gt;</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Stem class (SC)</td>
<td>0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Number of distinctively large, steep angled branches observed per tree (RAM)</td>
<td>0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Number of &quot;leaders&quot; (DL)</td>
<td>0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Bark thickness (AVBT)</td>
<td>2&lt;sup&gt;***&lt;/sup&gt;</td>
<td>11–12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Relative bark thickness (RBT)</td>
<td>0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Trunk height (HT)</td>
<td>3&lt;sup&gt;***&lt;/sup&gt;</td>
<td>13–21</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Overbark diameter at breast height (OBDBH)</td>
<td>1&lt;sup&gt;***&lt;/sup&gt;</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Underbark diameter at breast height (UBDBH)</td>
<td>1&lt;sup&gt;***&lt;/sup&gt;</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus caribaea</em> var hondurensis</td>
<td></td>
<td>Basic density (BD)</td>
<td>1&lt;sup&gt;***&lt;/sup&gt;</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(Pinus palustris Mill × Pinus elliottii Engl)</em> × <em>Pinus elliottii Engl</em></td>
<td>Growth traits</td>
<td>Total height (month 7,16,29,41)</td>
<td>5&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>3.6–11</td>
<td></td>
<td>Weng et al. 2002</td>
</tr>
<tr>
<td><em>Pinus elliottii Engl</em></td>
<td></td>
<td>Stem diameter</td>
<td>6&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>4–10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus elliottii Engl</em></td>
<td></td>
<td>Height increments</td>
<td>5&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>4.3–8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus elliottii Engl</em></td>
<td></td>
<td>Diameter increments</td>
<td>5&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>4.2–11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td>Growth, wood-quality, and adaptive traits</td>
<td>Tree height (TH)</td>
<td>3&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>11.5–12.2</td>
<td></td>
<td>Lerceteau et al. 2000</td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td></td>
<td>Trunk diameter at breast height (DBH)</td>
<td>2&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>9.3–15</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pinus sylvestris</em></td>
<td></td>
<td>Trunk diameter 0.5 m from ground (D0.5)</td>
<td>2&lt;sup&gt;lod2&lt;/sup&gt;</td>
<td>12.6–13.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>∗</sup> significant at chromosome-wide level (p < 0.05),
<sup>∗∗∗</sup> significant at genomewide level (p < 0.05),
<sup>lod</sup> lod threshold value provided by authors (IM),
<sup>pv</sup> significance level used for single-marker analysis,
<sup>✓</sup> unique QTLs, defined as subset of QTLs (suggestive or significant) that map within ca. 15 cM of one to another and have the same general profile for their parental and interaction effects (magnitude and direction of effects).
<sup>b</sup> when the % of phenotypic variance explained by each QTL was not provided, the phenotypic variance explained by all QTLs is indicated <sup>all</sup>
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of trait</th>
<th>Trait (abbreviation used in the article)</th>
<th>#QTLs(^a)</th>
<th>%Variance explained by each QTL(^b)</th>
<th>Note</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pinus pinaster</strong></td>
<td>Growth traits</td>
<td>Total height week 15</td>
<td>2**</td>
<td>7–12</td>
<td></td>
<td>Plomion et al. 1996a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total height week 38</td>
<td>3**</td>
<td>6.2–11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total height week 92</td>
<td>1**</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fertile zone length (LF)_ cycle 5</td>
<td>1**</td>
<td>19.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fertile zone length (LF) adjusted to NSU_ cycle 5</td>
<td>1**</td>
<td>11.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of stem units in the fertile zone (NSU)_ cycle 5</td>
<td>1**</td>
<td>20.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean stem unit length (MSUL)_ cycle 5</td>
<td>1**</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Germination date (GERMD)</td>
<td>2**</td>
<td>15–17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypocotyl length (Lhypo)</td>
<td>2**</td>
<td>9.5–12.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Megagametophyte weight (MW)</td>
<td>2**</td>
<td>6–8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adaptive and growth traits</td>
<td>Delta 13C</td>
<td>3** + 5**</td>
<td>4.7–12.4</td>
<td>pop : id to</td>
<td>Brendel et al. 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ring width</td>
<td>1*** + 5**</td>
<td>5.9–18.1</td>
<td></td>
<td>Pot et al 2006 and Pot 2004</td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Total height</td>
<td>0</td>
<td>–</td>
<td>pop : id to</td>
<td>Pot et al. 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean density of all the rings</td>
<td>2**</td>
<td>4.6–4.8</td>
<td></td>
<td>Brendel et al 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Density estimated by Piolyd penetration</td>
<td>2** + 1***</td>
<td>3.7–8.8</td>
<td></td>
<td>and Pot 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Density heterogeneity</td>
<td>3**</td>
<td>5.1–8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin content</td>
<td>7**</td>
<td>5–9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Number of QTLs

\(^b\) %Variance explained by each QTL
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of trait</th>
<th>Trait (abbreviation used in the article)</th>
<th>#QTLs(^a)</th>
<th>%Variance explained by each QTL(^b)</th>
<th>Note</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Wood density 1–5 years old</td>
<td>1^{***}</td>
<td>–</td>
<td>Pop identical</td>
<td>Kumar et al. 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wood density 6–10 years old</td>
<td>0</td>
<td>–</td>
<td>to Ball 2001,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wood density 14 years old</td>
<td>0</td>
<td>–</td>
<td>LG3 only, multitrait model analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Wood density 1–5 year-old</td>
<td>1^{***}</td>
<td>–</td>
<td>Pop identical</td>
<td>Ball 2001</td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Water extractives content</td>
<td>2(^{<strong>} + 1^{</strong>*})</td>
<td>5.8–8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Acetone extractive content</td>
<td>0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Alpha cellulose content</td>
<td>4^{**}</td>
<td>3.7–7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Hemicellulose content</td>
<td>3^{<strong>} + 1^{</strong>*}</td>
<td>4.5–8.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Lignin composition</td>
<td>3^{**}</td>
<td>5.1–9.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Kraft pulping yield adjusted to Kappa number</td>
<td>1^{***}</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Alpha cellulose content</td>
<td>3^{**}</td>
<td>4.7–6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Arithmetic fiber length</td>
<td>2(^{<strong>} + 1^{</strong>*})</td>
<td>4.9–12.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Weighted fiber length</td>
<td>2(^{<strong>} + 1^{</strong>*})</td>
<td>6.1–11.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Fibre width</td>
<td>2(^{<strong>} + 1^{</strong>*})</td>
<td>5.4–8.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Coarseness</td>
<td>2(^{<strong>} + 1^{</strong>*})</td>
<td>4.4–11.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Curl</td>
<td>1(^{<strong>} + 1^{</strong>*})</td>
<td>6–10.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality trait</td>
<td>Zero span tensile value</td>
<td>5(^{<strong>} + 2^{</strong>*})</td>
<td>4.9–7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Alpha cellulose content</td>
<td>5^{**}</td>
<td>5.98–12.24</td>
<td>Markussen et al. 2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Lignin content</td>
<td>6^{**}</td>
<td>5.98–13.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Pulp yield</td>
<td>7^{**}</td>
<td>9.65–18.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Extractives content</td>
<td>2^{**}</td>
<td>4.26–13.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>CIE brightness</td>
<td>3^{**}</td>
<td>7.76–13.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Mean wood density</td>
<td>8^{**}</td>
<td>6.17–12.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Minimum wood density</td>
<td>3^{**}</td>
<td>5.09–8.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Maximum wood density</td>
<td>6^{**}</td>
<td>7.50–14.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Diameter</td>
<td>7^{**}</td>
<td>5.04–14.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Height growth</td>
<td>3^{**}</td>
<td>4.93–9.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wood-quality and growth traits</td>
<td>Wood density 1–5 years old</td>
<td>1^{***}</td>
<td>–</td>
<td>Pop identical</td>
<td>Ball 2001</td>
</tr>
</tbody>
</table>

\(^a\) Number of QTLs identified

\(^b\) Percentage of variance explained by each QTL
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of trait</th>
<th>Trait (abbreviation used in the article)</th>
<th>#QTls(^a)</th>
<th>%Variance explained by each QTL(^b)</th>
<th>Note</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood density</td>
<td>14 year-old</td>
<td>0</td>
<td>–</td>
<td>to Kumar et al. 2000, LG1 and 3 only</td>
<td>Kumar et al. 2004</td>
<td></td>
</tr>
<tr>
<td><strong>Growth and wood-quality traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter at breast height</td>
<td>2** 0</td>
<td>2** 0</td>
<td>–</td>
<td></td>
<td>Kumar et al. 2004</td>
<td></td>
</tr>
<tr>
<td>Stem straightness</td>
<td>2** 0</td>
<td>2** 0</td>
<td>–</td>
<td></td>
<td>Kumar et al. 2004</td>
<td></td>
</tr>
<tr>
<td>Branching cluster frequency</td>
<td>2** 0</td>
<td>2** 0</td>
<td>–</td>
<td></td>
<td>Kumar et al. 2004</td>
<td></td>
</tr>
<tr>
<td>Wood density</td>
<td>1** 0</td>
<td>1** 0</td>
<td>–</td>
<td></td>
<td>Kumar et al. 2004</td>
<td></td>
</tr>
<tr>
<td><strong>Growth and wood-quality traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter at breast height</td>
<td>Verification: 2** 0</td>
<td>Verification: 2, 2 all</td>
<td></td>
<td></td>
<td>Devey et al. 2004a</td>
<td></td>
</tr>
<tr>
<td>Juvenile wood density</td>
<td>Verification: 8** 0</td>
<td>Verification: 14, 1 all</td>
<td></td>
<td></td>
<td>Devey et al. 2004b</td>
<td></td>
</tr>
<tr>
<td><strong>Biotic stress tolerance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistance to Dothistroma needle blight</td>
<td>4** 0 in the detection population, 4** 0 in the detection population</td>
<td>detection: not estimated, validation: 1.76–4.8</td>
<td></td>
<td></td>
<td>Devey et al. 2004b</td>
<td></td>
</tr>
<tr>
<td><strong>Pinus taeda</strong></td>
<td>Wood-quality trait</td>
<td>Wood specific gravity</td>
<td>5** 0.05</td>
<td>23 all</td>
<td>Groover et al 1994</td>
<td></td>
</tr>
<tr>
<td><strong>Growth traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter increment</td>
<td>Detection: 8</td>
<td>Detection: 12.5–59.5 all</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Growth traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total height 2 years</td>
<td>2**</td>
<td>7.9–10.8</td>
<td></td>
<td></td>
<td>Gwaze et al. 2003</td>
<td></td>
</tr>
<tr>
<td>Total height 3 years</td>
<td>1***</td>
<td>10.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total height 4 years</td>
<td>1***</td>
<td>12.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total height 5 years</td>
<td>1***</td>
<td>12.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>Type of trait</td>
<td>Trait (abbreviation used in the article)</td>
<td>#QTLs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>% Variance explained by each QTL&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Note</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>------------------------------------------</td>
<td>-----------------</td>
<td>---------------------------------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total height 10 years</td>
<td>1&lt;sup&gt;**&lt;/sup&gt;</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Growth rate</td>
<td>1&lt;sup&gt;***&lt;/sup&gt;</td>
<td>11.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood-quality trait</td>
<td>Wood specific gravity (individual rings and composite rings, measured for early and latewood)</td>
<td>23&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.4–15.7</td>
<td>Same population as Groover et al. 1994 and Knott et al. 1997</td>
<td>Sewell et al. 2000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage of latewood (individual rings and composite rings)</td>
<td>16&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.5–12.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average microfibrilar angle (individual rings and composite rings, measured for early and latewood)</td>
<td>7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.4–11.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood-quality trait</td>
<td>Cell wall content</td>
<td></td>
<td>8&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.3–12.7</td>
<td>Same population as Groover et al. 1994 and Knott et al. 1997</td>
<td>Sewell et al. 2002</td>
</tr>
<tr>
<td>Wood-quality traits</td>
<td>Wood-specific gravity</td>
<td></td>
<td>V : 18&lt;sup&gt;b&lt;/sup&gt;, U : 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>V : 1.7–5.7, U : 1.8–4.4</td>
<td>Detection population: same as Groover et al. 1994, Knott et al. 1997</td>
<td>Brown et al. 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percentage of latewood (individual rings and composite rings)</td>
<td>V : 12&lt;sup&gt;b&lt;/sup&gt;, U : 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>V : 1.7–5.7, U : 1.8–4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average microfibrilar angle (individual rings and composite rings, measured for early and latewood)</td>
<td>V : 4&lt;sup&gt;b&lt;/sup&gt;, U : 2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>V : 1.7–5.7, U : 1.8–4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell wall content</td>
<td>V : 10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>V : 1.7–5.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ent genetic and environmental backgrounds.” Indeed, estimation of QTL stability (position and phenotypic variation explained) is one of the most critical factors if QTL analysis is to be performed for application in MAS. This question has been addressed at different levels, including the stability of QTLs across different developmental stages (ontogenic or cambial age effect), time points during 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 whether 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 and Ekberg 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) analyzed 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 multisite 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 prerequisite to marker-assisted breeding (MAB) in multiparental tree breeding programs. As underlined by Brown et al. (2003), MAB will reveal its full potential under two scenarios: (1) if a genetic marker in full LD with molecular polymorphism causing trait variation at the population level is discovered or (2) if the gene (polymorphism) underlying a 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 LD (reviewed in González-Martínez et al. 2006b). The combination of these two factors (high diversity and low LD), 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.

Multiple-Trait QTL Analysis As stated earlier, multiple-trait QTL analysis will likely become very important for breeding purposes since pine breeding is a multitrait process. In several QTL experiments, more than one trait (e.g., growth, wood quality) was studied. In most cases, colocalizations between QTLs for different traits were observed, which might be expected for highly correlated traits. Possibly more important was the occurrence of multiple colocalizations 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.
2.3.3 Future Direction on QTL Mapping

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

Complex Designs for Detecting Marker-Trait Associations: LD Mapping

As underlined in Guevara et al. (2005b), while QTL identification is based on physical LD generated in one or a few generations of crossing, association mapping or LD mapping takes advantage of events that created associations in the past to find a statistical association between molecular markers and a phenotype. Association mapping or LD mapping takes advantage of events that created associations in the past to find a statistical association between molecular markers and a phenotype, making it an attractive alternative when few pedigrees are available for genetic analysis. As presented in Sect. 2.1.6., five studies have provided valuable information regarding the extent of LD 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. (2006a) revealed a rapid decay of LD within 800 to 2,000 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 *R. 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 has 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 the association between parental genotypes and the performance of 200 full-sib-generated families were analyzed 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.

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*, colocalizations between genes involved in monolignol 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 colocalization was found between KORRIGAN, a gene involved in the hemicellulose/cellulose biosynthesis, and QTLs for hemicellulose and fiber characteristics.

Other types of analysis can be used to select CGs related to plant adaptation. As underlined by González-Martínez et al. (2006b), “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. 2006a), *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 *KORRI-
GAN, a membrane-bound endo-1,4-beta-glucanase involved in cellulose/cellulose biosynthesis. As mentioned before, colocalization between KORRIGAN and wood-quality QTLs has been reported (Pot et al. 2005b). In P. taeda, although the action of neutral processes cannot be completely ruled out to explain the patterns of nucleotide diversity observed for CcoA-OMT1, several characteristics of its nucleotide diversity seem 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 emphasize the putative role of this gene in wood-trait variation. Association studies should now be used to validate this hypothesis.

2.4 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 from (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.

2.4.1 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.

Open-Pollinated Seed

Open-pollinated seed from a mixture of selected seed orchard material generally consists of a few dozen 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 the 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.

Control-Pollinated Seed

Control-pollinated seed, where pollens from selected parents are used to fertilize selected seed parents, in the process excluding pollen from other sources 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 P. 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 mark-
ers 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 that 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 ten 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.

### 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 raise the same issue, 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 utilize codominant markers (see references above), as these have been developed for other purposes (above) and are generally adequate for clonal fingerprinting, particularly if enough marker loci are used (e.g., Kirst et al. 2005b).

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

### 2.4.2 Elucidation of Genetic Phenomena

All genetic gain ultimately depends on DNA polymorphisms. Knowledge of the nature and 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, revelation of population structure and history, and detection of selection fingerprints, 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, not to whole populations. Association genetics may, therefore, be more relevant for understanding the genetic landscape of trait variation in forest trees.

### Estimation of Genetic Variance Structures and Heritabilities

DNA markers offer new ways of obtaining some key knowledge that is fundamental to tree improvement regarding genetic parameters, in particular genetic variances, heritabilities, and 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 between economic traits can be a major constraint if they are adverse, but they can provide the breeder with great opportunities if they are favorable. Even genetic correlations between noneconomic indicator traits and economic ones can be used to 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, then 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 in the past 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 nonrandom 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 nonrandomness 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 it can be supplemented with information from markers.

**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 that generates significant disequilibrium that could be useful for breeders. For example, localized 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 exist in populations derived from recent admixtures. In *P. radiata*, the exotic, domesticated “landraces” 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 nongenic regions can reveal much about the genetic history of those regions. Departures from Hardy-Weinberg equilibrium could reveal the presence of previously undetected genetic phenomena such as the presence/absence of inbreeding. Indeed, genetic variance (and gain) estimates are based on assumptions regarding the 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 prescreen for genes more likely to be associated with trait variations, although some caveats apply regarding the power to detect the effects of selection (Wright and Gaut 2005).

**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. LD mapping and association genetics can provide clues in the search for quantitative trait nucleotides (QTNs) 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 QTNs. However, because of the size of conifer genomes and the 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 a 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 variations of use for breeders, and to provide commercial cultivars. There are no reports yet of cloning QTLs 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.
2.4.3 Population Management

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 cycles of selection, intermingling, evaluation, selection, and so on. Maintaining full pedigree has been favored 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 pedigree will be more challenging, especially with the wind pollination that characterizes 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.

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 decrease in 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.

 Provision for Biotic Crises

While it appears QTNs typically exert minor individual effects in conifers, disease-resistance genes can represent a notable exception (Burdon 2001). 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 (R.D. Burdon and P.L. Wilcox personal communication). 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 threefold 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, toward 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 coevolved 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 a provision (Burdon and Gea 2006).

### 2.4.4 Selection and Breeding

#### 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, fewer financial costs are incurred if experiments are abandoned or lost prior to remeasurement – which does happen in tree-improvement programs. Disadvantages include underrepresentation 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 *Eucalyptus* and *P. radiata*.

#### Estimation of Heritability Based on Marker-Ascertained Relatedness

Because molecular markers have the potential to estimate relatedness between genotypes, methods to estimate the heritability of phenotypic characteristics have been proposed and evaluated (Andrew et al. 2005; Kumar and Richardson 2005) without the need to have any prior knowledge of genetic relationships (see above). Such methods have the potential benefit of allowing one to obtain 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 the 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 to the existence of infrastructure such as extensive testing involving progenies. Kumar and Richardson (2005) compared phenotype-based with marker-based heritability estimates for wood density in *P. radiata* and found very little difference. Similarly, Andrew et al. (2005) reported nonzero 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.

**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 multilocus 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 hundred 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 that 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) with 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 that 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 tradeoff between proportion of variance explained by markers and shortening of selection interval. For traits of moderate heritability (0.2 to 0.4) marker-assisted early selection (MAES) was more efficient than phenotypic selection, if they explained more than 5 to 10% of additive genetic variance and allowed the 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 QTLs explaining a few percent each (Wilcox et al. 1997; Brown et al. 2003; Devey et al. 2004a, b), and limitations for the selection of up to ten unlinked markers per family linked to such QTLs, 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 the prior establishment of QTL/phenotype associations.

All the above studies investigating the 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 the cost of genotyping as well as the length of time, meaning that additional costs will be incurred. 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 two-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 ten-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 the 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 favorable 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 lower 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 of 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. (2001b) showed that even modest gains in physical traits of 3.0 to 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 the 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 USA and New Zealand have undertaken or are still undertaking MAS on a limited scale, but there is as yet no 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 (see below) or genetic modification, may supersede MAS.

**Combined Among- and Within-family Selection Based upon Results from Association Genetics Experiments**

Advances in genomics technologies over the past 5 to 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); it differentiates the manner in which markers are found (i.e., via association genetics) and utilized 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 few quantitative analyses have been undertaken to date that involve detailed specific strategies for incorporation into tree breeding programs. Wilcox et al. (2006) described a range of applications within tree breeding programs, which are summarized below.
stricted 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 to 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 multistage approach as with MAS (Wu 2002) or undertake an additional round of selection prior to finalizing 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 lower as compared to phenotypic evaluation. Such evaluations, however, are considerably more expensive than obtaining sufficient markers for genome coverage for linkage and QTL mapping as well as for 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., prescreening 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 LD experiments become available.

Indeed, the outlook from results to date is cautiously favorable, 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 sequences be chosen? Possible approaches are as follows.

- 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 colocalization of expressed genes and QTL from pedigree 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 the
Chapter 2

Pines

69

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 capitalize on the most favorable 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 utilized 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 manage resource association genetics programs in a manner that allows operational implementation.

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

2.4.5 Summary of Applications of DNA Polymorphisms 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.

2.5 Genomics Resources for the Genus *Pinus*

2.5.1 Efforts Toward Complete Genome Sequencing

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 to 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 them for pine. The first pine BAC library, constructed by Islam-Faridi et al. (1998), affords 0.05× coverage of the *P. 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 (Fig. 3). In September 2004, the US National Science Foundation funded a 3-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 (Fig. 4). When completed, this library will be, to our knowledge, the single largest BAC library ever made with roughly 1.7 to 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. The 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.

Need for Pine Genomic Sequence
Although EST sequencing in pine is relatively advanced (Sect. 2.5.2), there are few genomic sequence data available for pine or any other gymnosperm. As of 6 March 2006, the longest continuous gymnosperm/conifer genomic sequence in GenBank was only 6,884 bp. While ESTs provide the coding regions of many expressed genes, an understanding of gene function and regulation requires knowledge of those noncoding sequences that coordinate gene expression in response to biotic and abiotic cues (e.g., promoters, enhancers, silencers). Such information necessarily comes from sequencing genomic DNA.

As the pine genome is ten 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 the study of the noncoding 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., Egerstdotter 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 ac-
Fig. 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).

tivation/repression mechanisms underlying major developmental events.

- **Comparative genomics of conifers**: Pine BAC/shotgun sequences and comparative genomic approaches can be utilized to advance the 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 means 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 genomewide studies of LD, 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-aide selection, and genetic engineering. Of note is the fact that a BAC containing a target gene may well contain other genes of interest. The 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 MAS.

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.

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 100 °C, 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 sequence (and hence its complementary sequence) is found within the genome. In other 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). CF has proven highly effective in wheat, where it provides 19.5-fold enrichment for low-copy DNA (Lamoureux et al. 2005).

\(^{1}\) 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).

2.5.2 Genomic Tools to Identify Genes of Economic and Ecological Interest

Characterization of the Pine Gene Space

A large-scale expressed sequence tag (EST) sequencing project can generate a 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 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 genomes primarily composed of nongenic 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 ESTs has been developed in \(P. \text{taeda}\) and, to a lesser extent, in \(P. \text{pinaster}\).

Loblolly pine ESTs

Three major EST sequencing projects, mainly funded by the NSF, have been conducted so far in \(P. \text{taeda}\). 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 ca. 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://funegen.org/Projects/Pine/Pine.htm). It resulted in ca. 60,000 additional ESTs. A third project, “Genomics of Loblolly pine embryogenesis,” seeks to identify and characterize
an additional 85,000 ESTs derived from 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 (C. Collada, M.A. Guevara, and M.T. 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 the Platform for Integrated Clone Management PICME Web site (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 analyze 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 (Fig. 5), especially for *P. 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 2,000 sequences. In 2005, the *Pinus* genus was introduced in the TIGR gene index (Quackenbush et al. 2001), where a comprehensive collection of all publicly 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.

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 and developmental stages and after various environmental challenges are analyzed. To this end, several cDNA libraries have been 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, are 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.
### Table 6. Transcriptome and proteome analysis in pines

<table>
<thead>
<tr>
<th>Species</th>
<th>Methodology</th>
<th>Study</th>
<th>Tissue</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. contorta</em></td>
<td>Microarrays</td>
<td>Adventitious root development</td>
<td>Roots</td>
<td>Brinker et al. 2004</td>
</tr>
<tr>
<td><em>P. halepensis, P. taeda</em></td>
<td>cDNA library, Northern blot, RT-PCR</td>
<td>Drought response in two pine species</td>
<td>Roots, needles</td>
<td>Sathyan 2005</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>cDNA-AFLP, reverse Northern</td>
<td>Drought response</td>
<td>Roots</td>
<td>Dubos and Plomion 2003</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>cDNA-AFLP, reverse Northern</td>
<td>Drought response</td>
<td>Needles</td>
<td>Dubos et al. 2003</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>cDNA library</td>
<td>Bud phenology</td>
<td>Buds at different developmental stages</td>
<td>Collada, Guevara, and Cerera, unpublished results</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>Six SSH cDNA libraries</td>
<td>Wood formation</td>
<td>Early vs. late, juvenile vs. mature wood-forming tissues</td>
<td>Pacheco et al. 2005</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>cDNA-AFLP; reverse Northern; RT-PCR</td>
<td>Wood formation</td>
<td>Early vs. late, opposite vs. compression wood-forming tissues</td>
<td>Le Provost et al. 2003</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>cDNA library</td>
<td>Wood formation</td>
<td>Composite xylem tissues</td>
<td>Cantón et al. 2003</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>EST sequencing, microarrays, qPCR</td>
<td>Wood formation</td>
<td>Early vs. late, juvenile vs. mature wood-forming tissues and 8 different organs</td>
<td>Paiva 2006</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>EST sequencing, macroarrays, qPCR, 2DE MS/MS</td>
<td>Drought response in two ecotypes</td>
<td>Roots</td>
<td>Chaumeil 2006</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>2DE, tandem MS</td>
<td>Wood formation</td>
<td>Wood-forming tissues</td>
<td>Gion et al. 2005</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>2DE, micro-sequencing</td>
<td>Drought response</td>
<td>Needles</td>
<td>Costa et al. 1998</td>
</tr>
<tr>
<td><em>P. pinaster</em></td>
<td>2DE, micro-sequencing</td>
<td>Wood formation</td>
<td>compression and opposite wood forming tissues</td>
<td>Plomion et al. 2000</td>
</tr>
<tr>
<td><em>P. pinea</em></td>
<td>Two SSH cDNA libraries</td>
<td>Cytokinine induction of adventitious buds</td>
<td>Cotyledons</td>
<td>Alonso, Cantón, Ordás, unpublished results</td>
</tr>
<tr>
<td><em>P. sylvestris</em></td>
<td>cDNA library</td>
<td>Nitrogen and carbon assimilation</td>
<td>Cotyledons</td>
<td>Avila et al. 2000</td>
</tr>
<tr>
<td><em>P. radiata</em></td>
<td>CAGE</td>
<td>Wood formation</td>
<td>Wood-forming tissues</td>
<td>Cato et al. 2006a</td>
</tr>
<tr>
<td><em>P. radiata</em></td>
<td>Three SSH cDNA libraries</td>
<td>Adventitious rooting capacity</td>
<td>Rooting competent tissues</td>
<td>Sánchez et al. 2005</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Differential screening, northern blot</td>
<td>Drought response</td>
<td>Needles, stems, roots</td>
<td>Chang et al. 1996</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Microarrays</td>
<td>Wood formation</td>
<td>Wood-forming tissues</td>
<td>Yang and Loopstra 2005</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Microarrays</td>
<td>Drought response and recovery</td>
<td>Needles</td>
<td>Watkinson et al. 2003</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Microarrays</td>
<td>Drought response</td>
<td>Needles, rooted cuttings, 2 unrelated genotypes</td>
<td>Heath et al. 2002</td>
</tr>
<tr>
<td>Species</td>
<td>Methodology</td>
<td>Study</td>
<td>Tissue</td>
<td>Authors</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Differential display</td>
<td>Response to galled tissues</td>
<td>Healthy and galled stem</td>
<td>Warren and Covert 2004</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>0.35k microarrays; qRT-PCR</td>
<td>Wood formation</td>
<td>Unlignified early and latewood xylem: cambial region and developing secondary nonlignified xylem</td>
<td>Egertsdotter et al. 2004</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>cDNA and subtractive libraries; EST sequencing; Northern blot</td>
<td>Wood formation</td>
<td>Compression and opposite wood</td>
<td>Allona et al. 1998</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>EST sequencing, microarray</td>
<td>Wood formation</td>
<td>Different tissues and organs, including several types of differentiating xylem samples</td>
<td>Whetten et al. 2001</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>SAGE profiling</td>
<td>Wood formation</td>
<td>Juvenile and mature wood forming tissues</td>
<td>Lorenz and Dean 2002</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>cDNA library; EST sequencing, comparative sequence analysis</td>
<td>Wood formation</td>
<td>Six types of wood forming tissues</td>
<td>Kirst et al. 2003</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Microarray, qRT-PCR</td>
<td>Identification of genes preferentially expressed in <em>P. taeda</em></td>
<td>Differentiating xylem, needles, megagametophytes, embryo (growing radicule)</td>
<td>Yang et al. 2004</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Comparative sequence analysis, digital profiling</td>
<td>Computational analysis of transcript accumulation in xylem.</td>
<td>Different types of wood-forming tissues</td>
<td>Pavy et al. 2005</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>EST sequencing, in silico detection of differentially expressed genes</td>
<td>Drought response</td>
<td>Roots</td>
<td>Lorenz et al. 2005</td>
</tr>
<tr>
<td><em>P. taeda</em>, <em>P. sylvestris</em>, <em>P. abies</em></td>
<td>Microarrays</td>
<td>Heterologous arrays</td>
<td>Needles</td>
<td>van Zyl et al. 2002</td>
</tr>
<tr>
<td><em>P. taeda</em></td>
<td>Microarrays</td>
<td>Disease resistance</td>
<td>Stem (infection court)</td>
<td>Myburg et al. 2006</td>
</tr>
<tr>
<td><em>P. elliottii</em> var. <em>elliottii</em></td>
<td>Microarrays</td>
<td>Disease resistance</td>
<td>Stem (infection court)</td>
<td>Morse et al. 2004</td>
</tr>
</tbody>
</table>
2.6 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 the presence of low levels of LD in natural populations (Strauss et al. 1992; Brown et al. 2004). Each of these factors limits the opportunity for finding QTLs 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 nonpedigreed populations, and both are valid for the 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).

Nonpedigreed 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 a very long time, possibly since the creation of the DNA polymorphism used as the genetic marker or the gene mutation itself. Extremely dense genetic maps are required for mapping QTLs in nonpedigreed 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 to 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 the 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.

Acknowledgement. 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 (P.L. Wilcox and R.D. Burdon), (F) the Finnish Funding Agency for Technology and Innovation and by the Biosciences and Environment Research Council of Finland, to O Savolainen.

References


Ball RD (2005) Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. Genetics 170:859–873


Cato S, Mc Millan L, Donaldson L, Richardson T, Echt C, Gardner R (2006a) Wood formation from the base to the crown in
C. Plomion et al.


Fulton TM, Van der Hoeven R, Eannetta NT, Tanskley SD (2002) Identification, analysis, and utilization of conserved or-
tholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467


Hyun SK (1976) Inter-specific hybridization in pines with the special reference to Pinus rigida and P. taeda. Silvae Genet 25:188–191


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


Kriebel HB (1985) DNA sequence components of the Pinus strobus nuclear genome. Can J For Res 15:1–4


Kuang H, Richardson TE, Carson SD, Wilcox PL, Bongarten B (1999a) Genetic analysis of inbreeding depres-


Lambeth C (1990) Silviculture in the Tropics: tropical forest ecosystems and their tree species-possibilities and methods for their long-term utilization. GTZ, Eschborn, Germany


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


identified by analyses of expressed sequence tag libraries. Tree Physiol 26:1–16


Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8:552–560


etegation if individual linkage maps from two outbred three-generation pedigrees. Genetics 151:321–330

Shelbourne CJA (1992) Genetic gains from different kinds of breeding population and seed or plant production populations. S Afr For J 160:49–65


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

Van der Burgh J (1973) Hölzer der nierderrheinischen Braunkohlenformation. 2. Hölzer der Braunkohlengruben


Wei RP, Lindgren D, Yeh FC (1997) Expected gain and status number following restricted individual and combined-index selection. Genome 40:1–8


Zeng ZB (1993a) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976


