Theor Appl Genet (1997) 95 : 6433654 © Springer-Verlag 1997

G. Bucci · T. L. Kubisiak · W. L. Nance · P. Menozzi

A population 'consensus', partial linkage map of *Picea abies* Karst based on RAPD markers

Received: 4 January 1997 / Accepted: 24 January 1997

Abstract We built a "consensus" partial linkage map based on RAPD markers using 48 sibships of eight megagametophytes each from a natural population of Norway spruce. A RAPD linkage map for a single individual from the same population had previously been constructed. Using 30 random decamers that had yielded 83 RAPD markers in the single-tree map, eight megagametophytes for each of the 48 sibships were screened. The linkage relationship among markers was estimated considering each family of eight megagametophytes as a progeny of a phase-unknown backcross mating between a heterozygous mother and a fictitious 'recessive' father. Markers were assigned to windows using LOD = 2.0 and $\theta = 0.4$ as thresholds, and ordered using a criterion of interval support ≥ 2.0. For eight "windows" of recombination selected on the single-tree map, we investigated the consistency of marker order in the two maps. We adopted restrictive criteria for rejecting co-linearity between the two locus orders. For each window we imposed the most likely locus order obtained from one data set to the other (and vice versa), obtaining two symmetrical log-likelihood differences. We considered the hypothesis of co-linearity rejected when both symmetrical differences were significant (ALOD > 3.0). By bootstrapping a subset of markers for each window (highly informative, 'framework' loci chosen on the previous single-tree map using a matrix correlation method) the sampling variability of the single-tree and population maps was estimated. As expected the population map was affected by a larger variability than the single-tree map. Heterogeneity in pairwise recombination fractions among groups of sibship revealed a (possible) alternative genomic arrangement detected within a single recombination window.

Key words Norway spruce · Genomic co-linearity · Integrated map · Log-likelihood · Genomic arrangements

Communicated by P. M. A. Tigerstedt

G. Bucci [1] (✉) · P. Menozzi
Dipartimento di Scienze Ambientah, Università di Parma, v.le delle Scienze, I-43100 Parma, Italy

T. L. Kubisiak · W. L. Nance
USDA Forest Service, USDA - Southern Institute of Forest Genetics, Harrison Experimental Forest, 23332 Hwy 67. Saucier, MS 39574. USA

Present address:
[1] Istituto Miglioramento Genetico Piante Forestali (IMGPF-CNR), v. Atto Vannucci, 13-50134 Firenze, Italy

Introduction

In recent years, the availability of a large number of DNA markers has made it possible to obtain fine-scale genetic maps from controlled crosses of many crop species (for a review see Tanksley 1993). Most maps are based on a large number of progeny from single experimental crosses (Burr et al. 1988; Tanksley et al. 1992). Interpopulation or interspecific crosses have frequently been used to maximize polymorphism in the progeny (Ellis et al. 1992; Bennetzen and Freeling 1993; Bradshaw and Stettler 1995; Kaga et al. 1996) and the number of segregating markers across the genome (Graner et al. 1991; Durham et al. 1992). Extension of linkage information from single crosses to the whole species is based on the assumption of prevailing intraspecific synteny.

Alignment and integration of single-pedigree maps have been recently attempted for several crop species (Beavis and Grant 1991; Kianian and Quiros 1992; Hauge et al . 1993). Procedures for integrating linkage information from different pedigrees are available in computer packages (Lathrop et al. 1984; Lander and Green 1987; Stam 1993) and have been widely used in mapping DNA polymorphisms in human chromosomes

(Weissebach et al. 1992; NIH/CEPH Coll. Mapping Group, 1992; CHLC et al. 1994). Structural differences between genomes have been reported to be associated with heterogeneity of recombination fractions among different crosses and/or families within populations (Morton 1955; Slocum et al. 1990; Ellis 1994), generating inconsistencies in multilocus linkage maps based on different pedigrees (Ellis et al. 1992). Therefore, single-pedigree linkage maps need validation to be used as a general genetic tool for the species considered, while a species "consensus" map may be less reliable because of the uncertainty due to genomic divergence between parental accessions.

In forest tree species, data collected from several families of megagametophytes have been routinely used to study linkage relationships between allozymic markers (Adams and Joly 1980; O'Malley 1986). To compensate for the limited number of segregating markers within a single individual, a large number of open-pollinated seeds have usually been collected from several highly heterozygous individuals (Strauss and Conkle 1986; Szmidt and Muona 1989) and genetic maps have been constructed pooling linkage information over families of progeny array data (Conkle 1981; Strauss and Conkle 1986; Geburek and von Wuehlisch 1989; Gerber et al. 1993). Nonetheless, genomic regions homozygous in the analyzed crosses (or individuals) could determine gaps between groups of markers in maps built using single-pedigree progenies or haploid megagametophytes from a single tree (Binelli and Bucci 1994). Since linkage information is obtained by pooling segregating markers from many individuals (Lander and Green 1987), mapping markers in natural populations should not encounter this problem.

RAPD markers have been used to build genetic linkage maps (Nelson et al. 1993; Grattapaglia and Sederoff 1994; Kubisiak et al. 1995). On the other hand, due to their instability in different genomic backgrounds, linkage relationships evaluated by RAPD markers have been claimed to have poor reproducibility even in maps built from progenies of trees from the same population (Mitchell-Olds 1995; Plomion et al. 1995 a). Once consistency of results across different genetic backgrounds has been obtained for each marker (Rieseberg 1996), population genetic and evolutionary studies may greatly benefit from the use of RAPD markers (Bradshaw et al. 1995; Rieseberg et al. 1995, 1996). Recently, extensive conservation of mapped RAPD markers between trees from the same population has been reported in ***Eucalyptus*** (Brondani and Grattapaglia 1996 a).

The main goal of the present study is to investigate the feasibility of using routinely collected data in forest-tree population studies (small number of open-pollinated seeds from a large number of seed trees) to identify a set of RAPD loci showing stability in different genetic backgrounds and unambiguous linkage relationships over different genotypes. Such markers might be used as 'landmarks' on the Norway spruce genome suitable for single-tree map merging, QTL analyses, and further population studies. Mapping exercises carried out on such routinely collected data could also provide useful hints of major genomic rearrangements existing in wild populations. We screened 83 RAPD markers obtained by 30 random decamers previously mapped using a large progeny from a single tree (Binelli and Bucci 1994) on eight megagametophytes for each of 48 individuals sampled in a natural population of Norway spruce. Linkage relationships among markers have been estimated and the population linkage map obtained has been compared with the single-tree linkage map. Heterogeneity in the recombination fractions among pairs of markers has been used for identifying groups of individuals within the population carrying (putative) alternative genomic arrangements of the markers analyzed.

Materials and methods

Plant material, DNA extraction and amplification

Seeds were harvested from 48 trees sampled at random from a natural population of Norway spruce (Campolino stand, Apennines, Northern Italy, 44°07′N, 1°15′ W). Megagametophytes were separated from seed coats and embryos under a microscope using a scalpel and forceps. DNA extraction was performed as previously described (Binelli and Bucci 1994). DNA amplifications were set up using a Hamilton Microlab *AT* plus robot in 96-well microtiter plates (Costar) and performed in MJ Research PTC-100/96 thermal cyclers using the conditions described in Binelli and Bucci (1994). The reactions were run in 2% agarose gels in 1 x TAE at 3 V/cm, stained with ethidium bromide, and photographed using a Polaroid camera.

Choice and identification of RAPD markers

Forty eight primers used in the single-tree mapping study (Bucci et al. 1995) were screened on eight megagametophytes from the control tree previously mapped (Ind. #7). The best 30 primers, which produced X3 marker bands in the previous mapping effort, were selected based on the reproducibility and consistency with that obtained in the previous study.

Selected primers were then used to amplify eight megagametophytes from 4X trees of the analyzed population (384 samples). Eight megagametophytes from Ind. #7 amplified in the previous step were included in each amplification round as a reproducibility control (Fig. 1). Marker bands were then scored across the 3X4 megagametophytes by two people (G.B. and T.L.K) independently and the scores then compared by computer programs ("diff" command-UNIX system). Differences in the two data sets were used to assess the reliability of the marker bands across the population. Fourteen marker bands showing more than ten mismatches between the two data bases ($>$ 2.5%) were considered poorly interpretable and were then discarded. The remaining 69 markers were classified as class 'A' (mismatches lower than 1%) and class 'B' (1.0–2.5 % of mismatches), completely re-scored, and then used for further analyses.

Segregation of the markers over the whole population

Single-locus genotypes of individual trees were inferred from progeny array data (families of eight megagametophytes). Not

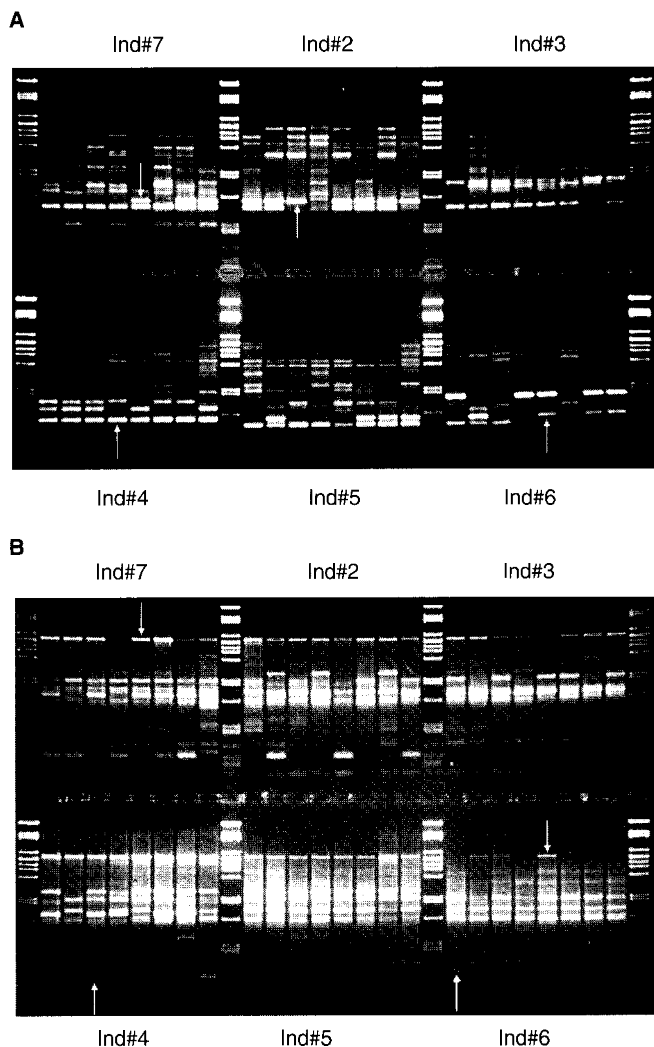

Fig. **1A, B** RAPD amplification of different families of megagametophytes (eight megagametophytes per tree) by two primers. A primer C316 (University of British Columbia series); B Primer OA08 (Operon Technologies, Calif.). Marker bands are indicated by *arrows*. Upper leftmost of each picture: eight megagametophytes from the control tree previously mapped (Ind. # 7, Binelli and Bucci 1994). The molecular-weight marker is Lambda/*Pst* I

knowing the individual genotypes in advance leads to a small ascertainment bias (about 0.019), since the heterozygous families with 8 : 0 or 0: 8 segregation ratios for a given locus were not included in the data set. Out of 3312 (48 trees x 69 markers) single-locus genotypes, 2958 (89.3%) were unambiguously scorable based on at least seven megagametophytes; the others were discarded from further analysis. The average number of individuals per locus was 42.87 ± 0.67, while the average number of single-locus genotypes identified per individual was 61.63 ± 0.99. Overall, 1737 single-locus genotypes inferred from progeny array data were homozygous, and 1221 were heterozygous. Out of 1221 heterozygous single-locus genotypes, 1009 (82.6 %) were based on eight unambiguous scores, while 212 were based on seven unambiguous scores.

Segregation analysis of single markers was performed on the 1009 heterozygote based on eight unambiguous scores as previously described (Bucci and Menozzi 1995). Overall fitting of the segregation ratios within families of megagametophytes was obtained by pooling data across single-locus heterozygous trees and markers using a Kolmogorov-Smirnov test ($\alpha = 0.05$).

Data analysis

Pairwise estimation of linkage between markers over the whole population was carried out using each sibship of eight megagametophytes as the offspring of a phase-unknown double-backcross mating between a double-heterozygous mother ('10/10' or '10/01') and a fictitious double-homozygous father for the recessive allele ('00/00', Gerber et al. 1993). Linkage analyses were restricted to doubly heterozygous trees showing at least seven unambiguous dilocus, haploid scores ($n \geq 7$). Deviation from independent segregation for markers was preliminarily verified by an S test (Gerber and Rodolphe 1994).

The progeny array data-matrix was analyzed using the computer package MAPMAKER V2.0 ("CEPH data type" format, Lander and Green 1987). First, the 36 "class-A" markers were used to build a framework map. Markers were assigned to linkage groups using LOD > 3.0 and $\theta < 0.25$ as thresholds. The order for informative loci within each linkage group was preliminarily established by the matrix-correlation method implemented by MAPMAKER. Locus orders were then tested by three-point analysis using the 'ripple' command. The remaining markers were assigned to linkage groups at LOD threshold >2.0 and $\theta < 0.40$ (Gerber and Rodolphe 1994) and placed at the most likely position using the 'try' command. Marker order within each linkage group was considered significant when all the interval supports were larger than 2.0. Map distances were calculated using the Kosambi mapping function.

Control megagametophytes from Ind. # 7 were also scored for the 69 markers screened over the whole population and appended to the database previously used for the construction of the single-tree genetic map (Binelli and Bucci 1994). The overall number of megagametophytes analyzed for Ind. # 7 was then 80 (72 in the previous phase plus eight this round). Partial linkage groups, orders, map distances, and likelihood for the above windows in Ind. # 7 were obtained as described above. Linkage groups were then re-drawn using CHROMOSOME DRAWING V 3 0 (L. Gianfranceschi, ETH Zurich, CH).

Comparison of the population and single-tree maps

Co-linearity of the markers within each linkage group for both maps was tested as follows. For the population data set (data set α), the log-likelihood of the most-likely locus order obtained in the single-tree map was calculated. Differences between the most-likely population order obtained (locus order A) and the imposed single-tree order (locus order B) were considered significant when the LOD difference [$\Delta LOD_{(A \text{ vs } B \text{ on } \alpha)}$] was larger than 3.0 (i.e. the best locus order found for a given linkage group in the population data set was more than 1000-fold more likely than the locus order found in the single-tree data set). As for the single-tree data set (data set β), the relative log-likelihood was calculated using the best locus order obtained in the population data set (locus order A). A $\Delta LOD_{(B \text{ vs } A \text{ on } \beta)} > 3.0$ between the most likely single-tree order (locus order B) and the imposed population order (locus order A) was considered significant. Co-linearity between the population and single-tree locus orders was rejected when both comparisons (A vs B on a and B vs A on β) showed ALOD > 3.0.

Numerical re-sampling of the data sets and estimation of locus-ordering error rates

Bootstrapping (Efron 1982; Weir 1990) was used to evaluate the reliability of the locus orders obtained for the single-tree and the

population partial linkage maps. A subset of 27 evenly spaced, highly informative markers ('framework' loci, Table 2) was chosen on the single-tree data set using the matrix correlation method implemented by MAPMAKER 3.0. ('order' command, LOD threshold = 3.0, Lander et al. 1987) with informativeness-criteria parameters of 5.0 cM (minimum distance) and 70 (minimum number of individuals informative for each marker). Megagametophytes were sampled with replacement and a new set of 20 single-tree, bootstrapped data sets (80 megagametophytes genotyped at 27 loci within an individual family) were obtained. Partial linkage groups (windows) were re-constructed for each of the 20 re-sampled data sets by multipoint analysis (using the same procedures previously described). The best locus order within each window for each re-sampled data set was obtained based on log-likelihood. Pairwise differences among the 20 re-sampled locus orders (overall: 20 x 19/2 = 190 map comparisons) were computed based on mismatched positions of the markers, The estimated error rate for each window, obtained considering one difference in a locus order equal to an 0.5 mistake per map (Plomion et al. 1995 b), has been considered representative of the single-tree locus order uncertainty for that window.

In the population, partial linkage map, 20 bootstrapped data sets were obtained by re-sampling with replacement within each family of eight megagametophytes genotyped at the same 27 loci as the single-tree megagametophytes. The locus order for each of the 20 data sets was found by multipoint analysis using the same criteria described above. Pairwise differences among the 20 locus orders and estimated error rates were obtained as described for the single-tree data set. The proportion of mis matched marker positions for each window between all population-bootstrapped locus orders was considered representative of the uncertainty of the population partial linkage map observed for that window.

The proportion of mismatches for each window between single-tree and population locus orders was obtained comparing each re-sampled locus order from the single-tree data set with each of the re-sampled locus orders from the population data set (overall = 20 x 20 locus order comparisons).

Test of homogeneity of recombination and classification of single trees

Homogeneity of recombination between markers over megagametophyte families was verified by the M-test (Morton 1955). To verify the existence within the population of two (or more) subsets of families homogeneous for the recombination fraction between pairs of markers showing overall heterogeneity, we proceeded as follows. Pairwise recombination fractions (henceforth: *rf*) between all the markers in window GE were estimated for each family analyzed (i.e. for each co-informative, double-heterozygous tree): the log-likelihood function for phase-unknown double-backcross progeny was constructed (Ott 1991) and the maximum value within the interval 0.001 and 0.45 *rf* was found by increasing the recombination value by a step of 0.01. The Pearson product-moment correlation coefficient was computed for each pair of individuals based on the pairwise *rf* between their co-informative loci. A correlation matrix between trees based on the pairwise *rf* between markers was obtained. Cluster analysis was performed on the correlation matrix by the UPGMA method using (I – Pearson *r*) as the pairwise distance between single trees.

A multipoint homogeneity test was also computed for the two data subsets obtained as described in Beavis and Grant (1991). This statistic is asymptotically distributed as a χ^2 with $n-1$ degrees of freedom.

Identity of RAPD marker bands among individuals

Identity of RAPD marker bands among individuals was verified as follows. We considered the RAPD bands scored as the same marker in different individuals from heterozygous trees. For each marker, bands from randomly chosen megagametophytes were excised from gels and pooled (maximum of four bands per pool). Three to four pools (plus the band pool from control megagametophytes, Ind # 7) were obtained and re-amplified using the appropriate decamer according to the conditions described above. Restriction analysis of re-amplified bands was carried out using several endonucleases according to conditions described by the manufacturer (Promega, Boehringer). Identity of the marker bands was verified by comparing restriction profiles of random and control pools (Rieseberg 1996).

Results

Segregation analysis over the whole population

The overall number of amplification reactions was 11904 (48 primers x 8 control megagametophytes from Ind #7 + 384 megagametophyte samples from 48 individuals x 30 primers). Segregation analysis of each marker was carried out by pooling segregation ratios within each sibship of megagametophytes from heterozygous trees. Out of 69 markers analyzed, only two (2.90%–C3991986 and C6870450) showed significant differences from binomial expectation, less than what would be expected by chance alone. Altogether, pooling segregation ratios over markers and heterozygous trees, no significant differences from expectation were found by the Kolmogorov-Smirnov test (KS $d = 0.0418$; $p > 0.05$ – Fig. 2).

Construction of a population ("consensus") map

Out of 83 RAPD markers chosen, 70 were placed on 15 major linkage groups on the original single-tree map, five on triplets or doublets, and eight were unlinked. Of the above 70 markers, 47 fell into eight "windows" of

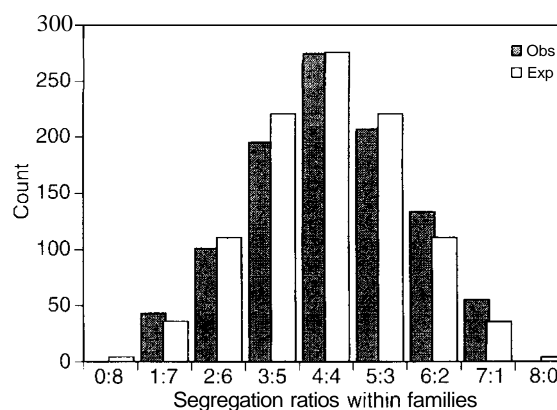

Fig. 2 Fit of the observed segregation ratios within families over the whole population to their binomial expectation (n = 1009; KS $d = 0.0418$; $p > 0.05$). Individuals showing family segregation ratios of 8 : 0 and 0 : 8 were considered homozygous and were not used in further analysis (see Materials and methods)

recombination along eight of the above 15 linkage groups (henceforth named simply "windows"), with an average distance of 12.27 $\pm$ 1.21 cM on the original map. The remaining 36 markers (= 83347) were either evenly spaced within and between linkage groups ($rf > 0.4$) or unlinked on the previously reported linkage map (Binelli and Bucci 1994).

In this investigation, 69 (class of polymorphism "A" and "B") of the above 83 markers were used to establish the reliability of RAPD markers in population studies. The overall number of pairwise co-informative meioses analyzed in this study was 135 020, while the number of co-informative triplets (number of scores informative for three loci at the same time) was 80477. Average numbers of doubly and triply co-informative meioses for each window are reported in Table 1.

Out of 47 RAPD markers falling within the eight "windows" described above, six (markers: C1671450, window A; OG100470, C6870537, C1691180, window C; 0F140750, C2680280, window GE) were unlinked at a LOD threshold of 2.0. The low number of co-informative meioses for most of the pairwise combinations for these loci (i.e. the low number of doubly and triply heterozygous trees found in the population) seems a reasonable explanation for this observation.

The remaining 41 markers fell within the eight windows on the linkage groups, as expected (Fig. 3). None of the other 22 markers (unlinked or located elsewhere on the previous map) showed any significant linkage relationship at a LOD threshold equal to 2.0. The overall coverage obtained by mapping the 41 markers within the eight windows in the population map was 474.6 cM, that is about one-fifth (18.67%) of the genetic distance covered by the 185 markers in the single-tree genetic map previously reported (Binelli and Bucci 1994). No differences were found between the distribution of pairwise map distances in the two data sets (Kolmogorov-Smirnov test: KS $d = 0.052$; $p > 0.05$).

Comparison between the single-tree and population maps

Comparison between the population and single-tree maps was carried out by imposing the best locus order for each window for one map on the other (Table 1). After imposing the single-tree best orders on the population data set, log-likelihood differences were significant for window F (ALOD = 3.68) and window GE (ALOD = 21.54). On the other hand, the log-likelihood for each window obtained by imposing the population best orders to the single-tree data set showed significant differences for window A (ALOD = 12.46) window B (ALOD = 10.95) and window GE (ALOD = 3.06). Therefore, the only window showing consistently incompatible locus orders between single-tree and population maps (both ALOD > 3.0) was GE.

Numerical re-sampling of the data sets and estimation of locus-ordering error rates

'Bootstrap' analysis was carried out on the two data sets to assess the reliability of locus orders obtained for the two partial linkage maps. The number of re-sampled data sets was limited to 20, due to the laboriousness of the framework map-construction methods (Plomion et al. 1995 b), and to a subset of loci, in order to reduce the number of computations that rapidly increases with the number of markers. The loci to be compared were framework loci chosen in the single-tree data set to avoid the possible bias due to the variable amount of linkage information of different markers in the population data set.

Table 2 shows the results of bootstrapping of the two data sets. In general, locus ordering in the population data set seems to be affected by a larger (though uneven) error rate than in the single-tree dataset, as expected due to the choice of framework markers based on informativeness in the single-tree map. The robustness of the population locus ordering was fairly satisfying (error rate consistent to that expected by chance) for windows A, C, E, F, GE and GD. The low number of families where three-point linkage information could be obtained seems to account for the uncertainty of the window B and D locus orders in the population map (Table 1).

The co-linearity for each window between population and single-tree maps was estimated by computing the number of mismatches in locus orders between each population-bootstrapped order with each single-tree-bootstrapped order. The proportion of mismatches was fairly large for windows B, D, and GE. As for the two former windows, the inconsistency of the single-tree and population-map locus orders seems to be due to the large estimated error rate obtained for the population partial linkage map (Table 2). On the other hand, the (population) estimated error rate for window GE alone does not seem to account for the large proportion of mismatches found.

Heterogeneity of the recombination fractions between individuals

Heterogeneity of the pairwise recombination fractions of markers over the population was detected for 37 out of 2436 pairs of loci (1.57%) a proportion smaller than what would be expected by chance. Significant cases, primarily involving markers 0B05 1150 (4 cases), 0B051020 (3), Cl690270 (5), C5030350 (2), 0A080302 (1), OA080340 (4) and C2662000 (2), all belonging to window GE, did not appear to be randomly distributed.

To verify the identity among individuals of the marker bands belonging to window GE, restriction analysis of gel-isolated fragments was carried out. No

Table 1 Comparison between marker orders in the population and single-tree data sets. Window labels are the same as in Fig. 3 (in parentheses, the number of loci considered). Number of co-informative meioses (megagametophytes informative for two and three loci) for each window was calculated between all loci belonging to the same window only. Support for the best orders shown in Fig. 3 is reported for each window (columns 4 and 9). LOD differences (columns 6 and 11) were obtained as the difference between the best-order support for the single-tree order (column 4) and the imposed population order (column 5) and vice versa (columns 9 and 10). For further details see Materials and methods

Window (no. markers)	Single-tree data set					Population data set					Co-linearity
	Av. number of co-info meioses (± SE)	Av. number of co-info triplets (± SE)	Best order log-like	Population order log-like	LOD difference	Av. number of co-info meioses (± SE)	Av. number of co-info triplets (± SE)	Best order log-like	Single-tree order log-like	LOD difference	
Window A (7)	75.13 (0.67)	66.60 (0.51)	– 76.29	– 88.75	12.46[b]	62.89 (7.54)	42.43 (5.05)	– 269.86	– 270.55	0.69	Not rejected
Window B (5)	76.10 (0.3 1)	66.17 (0.31)	– 59.41	– 70.36	10.95[b]	49.66 (7.39)	22.47 (1.54)	– 154.61	– 157.05	2.44	Not rejected
Window C (3)	62.07 (2.96)	59.00 (–)	– 30.24	– 30.24	0.00	51.66 (13.86)	46.00 (–)	– 97.35	– 97.35	0.00	Not rejected
Window D (5)	71.80 (0.53)	59.67 (0.67)	– 72.03	– 72.90	0.87	79.80 (6.16)	38.00 (3.54)	– 254.29	– 255.64	1.35	Not rejected
Window E (3)	73.17 (0.83)	62.00 (-)	– 42.83	– 42.83	0.00	65.67 (21.72)	42.00 (–)	– 117.72	– 117.72	0.00	Not rejected
Window F (6)	66.73 (2.82)	55.60 (3.55)	– 72.71	– 73.55	0.84	56.60 (5.09)	35.82 (4.33)	– 257.12	– 260.80	3.68[b]	Not rejected
Window GE (8)	71.33 (0.57)	59.97 (0.61)	– 74.03	– 77.09	3.06[b]	67.80 (4.99)	45.20 (2.31)	– 342.83	– 364.37	21.54[b]	Rejected
Window GD (4)	72.66 (0.62)	62.33 (1.20)	– 61.30	– 61.30	0.00	73.90 (6.73)	42.43 (4.48)	– 175.74	– 175.74	0.00	Not rejected
Mean (69)	70.83 (0.15)	61.05 (0.17)				57.55 0.61	42.29 1.72				
Overall	166 176	139066				135020	80477				

[a] Number of markers within each window of recombination considered
[b] Locus orders 1000-fold more likely than the alternative order. For additional details see text

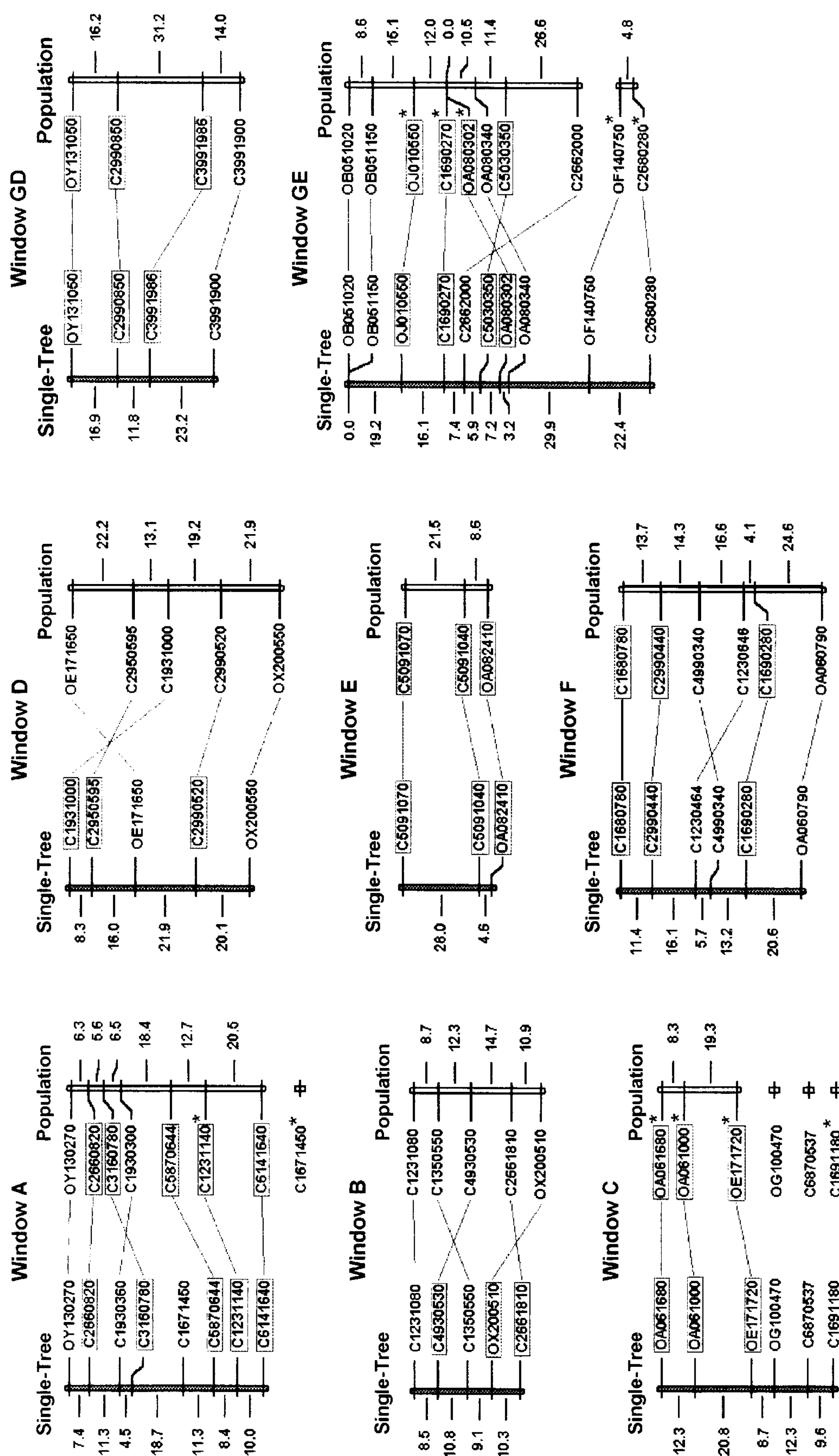

Fig. 3 Schematic representation of the eight windows obtained by analyzing the population and single-tree data sets. Distances are shown in centimorgans. *Boxed markers* on the single-tree map identified by three-point linkage analysis have been considered as framework loci. *Boxed markers* on the population map are RAPD loci showing stability in different genetic backgrounds (unambiguous linkage relationships over different genotypes). Markers Cl671450 (window A), OG100470, C6870537, Cl691 180 (window C), and 0F140750, C2680280 (window GE) turned out to be unlinked at a LOD threshold of 2.0 and were therefore discarded from further analyses. Independence for pairs of markers with *asterisks* was rejected by the S test (Gerber and Rodolphe 1993)

Table 2 Estimation of locus-order error rates by re-sampling ('bootstrap') of megagametophytes within families in the two data sets. Column 2 shows the framework loci chosen from the single-tree data set using a matrix correlation method implemented by MAPMAKER (for additional details see text). Twenty data sets (columns 3 and 5) were generated, re-sampling megagametophytes within families. For the single-tree data set, sampling with replacement of 80 megagametophytes genotyped at 27 framework loci was carried out. For the population data set, re-sampling was carried out within each sibship of eight megagametophytes genotyped at the same 27 loci. Estimated error rates (columns 4 and 6) were calculated following Plomion et al. (1995 b) by a pairwise comparison of re-sampled locus orders, counting 0.5 mistakes per map for each mismatched position of the markers. The proportion of mismatches between single-tree and population locus orders (column 6) were obtained as described in Materials and methods comparing each re-sampled locus order obtained from the single-tree data set with each of the re-sampled locus orders obtained from the population data set

Linkage group	Single-tree map			Population map		Comparison % mismatches
	Framework loci	n	Estimated error rate	n	Estimated error rate	
Window A	C2660820 C3160780 C5870644 Cl231140 C6141640	20	2.0% (38/950)	20	5.4% (1021950)	3.7% (148/2000)
Window B	C4930530 C2661810 OX2005 10	20	0.0% (0/570)	20	16.0% (182/570)	21.7% (520/1200)
Window C	OA061680 0A061000 OE171720	20	0.0% (0/570)	20	6.3% (72,570)	3.3% (80/1200)
Window D	C2950595 C1931000 C2990520	20	3.3% (38/570)	20	17.4% (198 570)	15.2% (364/1200)
Window E	C5091070 C5091040 OA082410	20	3.3% (38,570)	20	3.3% (38/570)	3.2% (76/1200)
Window F	Cl680780 C2990440 Cl690280	20	0.0% (0,570)	20	0.0% (0/570)	0.0% (0/1200)
Window GE	05010550 Cl690270 OA080302 C5030530	20	0.0% (0/760)	20	6.7% (1021760)	26.9% (860 1600)
Window GD	OY131050 C2990850 C3991986	20	0.0% (0 '570)	20	0.0% (0/570)	0.0% (0/ 1200)

differences were found between restriction profiles of randomly pooled and control bands for any of the following markers: OB051150, OB051020, 05010550, C 1690270, OA080302, OA080340, C5030350, C2662000 (data not show). Cluster analysis of pairwise MLEs of *rf* between markers in window GE allowed us to group sibships (families of megagametophytes from a single tree) in two sets of 30 (subset GE-A) and 15 (subset ***GE-B)*** sibships. The two subsets were used for the construction of two maps of window GE (Fig. 4), according to the procedures described above. Markers OB051020, 0B051150, 05010550 and Cl690270 were in the same order in the two maps, while markers C5030350, OA080302, 0A080340 and C2662000 were arranged differently. The recombination value between C5030350 and C1690270/OA080302 was 19.2% in subset A and 0.0% in subset B. All the interval supports for both maps were larger than 3.0, except for interval OA080340/C2662000 in window ***GE-B*** (ALOD = 1.92). Therefore, linkage was rejected and marker C2662000 was excluded from this analysis. Log-likelihoods were calculated for the two marker orders ***(GE-A*** and ***GE-B);*** significant differences were found ($\Delta LOD_{(A\ vs\ B)} = 6.3\ 1$ and $\Delta LOD_{(B\ vs\ A)}$ 3.03, respectively).

A multipoint homogeneity test detected no significant heterogeneity for intervals 1, 2 and 3 (markers 0B051020, OB051150, 05010550, C1690270), while homogeneity was rejected when intervals 4, 5 and 6 were considered (markers C1690270, OA080302, 0A080340 and C5030350, Table 3a). Single-interval heterogeneity was also evaluated using the log-likelihood estimated by MAPMAKER, showing that the two orders differ only for the position of marker C5030350 (Table 3b).

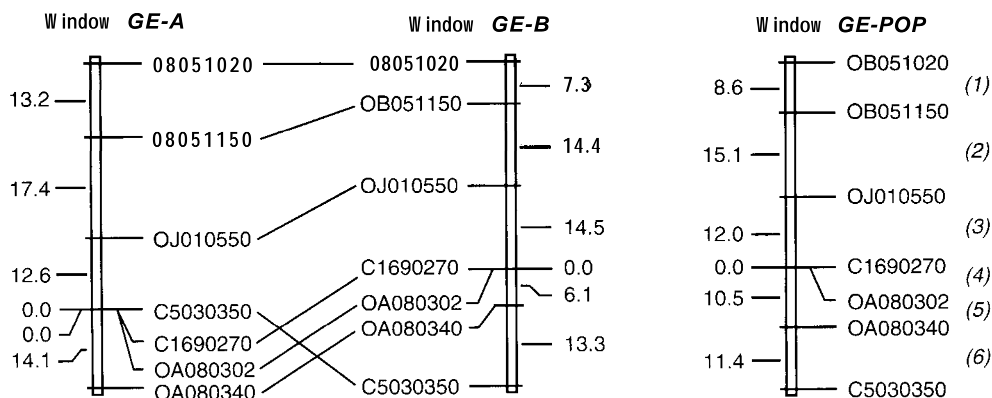

Fig. 4 Maps of the window GE obtained from the two data subsets (GE-A and *GE-B)* identified by cluster analysis. The best supported map obtained by pooling data from subset A and B *(GE-POP)* is shown. The number of families of megagametophytes analyzed were 30, 15 and 45 for the subsets *GE-A*, *GE-B* and *GE-POP*, respectively. Different arrangements between subsets *GE-A* and *GE-B* were mainly due to the position of marker C5030350 (see also Table 3). Numbers in parentheses in the pooled map refer to intervals. For further details see text

Discussion

To verify linkage relationships between RAPD markers in the whole population and to test their reliability in population genetic studies, we selected eight "windows" of recombination in eight different linkage groups on the previously reported single-tree RAPD map (Binelli and Bucci 1994). The high level of polymorphism detected by the RAPD technique (Fritsch and Rieseberg 1992) and the large number of markers analyzed (both increasing the overall number of co-informative meioses) allowed most of the multipoint linkage relationships to be studied, despite the low number of megagametophytes per tree considered in this investigation and the phase-unknown allelic configuration that was assumed for individual parent trees (Gerber et al. 1993). The low frequency in the screened individuals for unlinked markers of window C limited the number of co-informative meioses available for the multilocus analyses, hence lowering the total amount of linkage information in the data set.

Perfect co-linearity between single-tree and population maps was detected for windows C, E and GD, while some inconsistencies between the most-likely locus orders obtained by maximum likelihood methods were observed for windows A, B, D, F and GE. For each window we imposed the most-likely locus order obtained from one data set on to the other (and vice versa), generating two symmetrical log-likelihood differences (Table 1). For rejecting a co-linearity of locus orders between the two maps, we adopted a restrictive criterion based on the significance (ALOD > 3.0) of both symmetrical log-likelihood differences between the observed and imposed locus order on the two data sets. The hypothesis of co-linearity was rejected for window GE only, while most of the differences in locus order observed for the other windows may be attributable to sampling error and/or limitations of the mapping method.

To our knowledge, bootstrapping analysis has been rarely used to assess the reliability of the locus ordering found by maximum-likelihood methods. Numerical re-sampling of family data in humans has been suggested for constructing confidence intervals of marker locations (Suther and Wilson 1990; Suther 1991). Recently, Liu (1996) has proposed a non-parametric approach for quantifying the confidence of gene orders using a combination of bootstrap and jacknife techniques on single-pedigree data sets. Plomion et al. (1995 b) compared the framework loci of two maps from selfed and open-pollinated seed megagametophytes of the same individual of *Pinus pinaster*, obtaining an estimated error rate (2.71%) consistent with that obtained by re-sampling (n = 2) the two data sets. In the present investigation, the inherent error rate for the population map was estimated on framework loci for each partial linkage group (windows) by a pairwise comparison of 20 bootstrapped locus orders. The analysis revealed an average error rate estimated for the population data set (6.76%, though uneven among windows, ranging from 0.0 to 17.4%) larger than the average error affecting the single-tree dataset (1.11%, ranging from 0.0 to 3.3%).

Several reasons might explain the larger estimated error rate found for the population dataset: (1) sampling error due to the lower number of megagametophytes per family considered (eight instead of 80); (2) lowered linkage information in the population data set after re-sampling within families due to the classification of heterozygous families with an unbalanced segregation ratio (e.g. 7 : 1/1 : 7 or 6: 2/2: 6) as homozygous; (3) an uneven number of co-informative meioses for pairs or triplets of markers, which could increase LOD scores for loosely linked marker pairs and/or decrease LOD for tightly linked loci, so affecting the procedures to group up markers and build up linkage maps. Indeed, the integration of the linkage

Table 3 Multipoint homogeneity test for window GE applied to data subsets identified in the population. (a) A multipoint homogeneity test applied to data subsets GE-A and *GE-B* obtained by cluster analysis (see Materials and methods). Orders 1 and 2 refer to the (putative) alternative arrangements found for markers C1690270, 0A080302, OA080340 and C5030350. Multipoint log-likelihoods were calculated using the MAPMAKER v 2 0 computer packages. ALOD refers to the log-likelihood differences between the most-likely order for the subset *GE-A* and the same order imposed on the subset *GE-B*, and vice versa. Intervals considered in the multipoint analysis are shown in Fig. 4. (b) Pairwise homogeneity test (as described before) for the above markers

(a)

Groups	Order 1 (C1690270/OA080302/ /OA080340/C5030530)				Order 2 (C5030350/C1690270/ /OA080302/OA080340)			
	(A)	(B)	(A + B)	$\chi^2[1]$	(A)	(B)	(A + B)	$\chi^2[1]$
Intervals 1, 2, 3, 4, 5, 6								
Log-likelihood	− 168.95	− 106.17	− 275.51	1.847^{ns}	− 175.27	− 103.14	− 279.85	6.624*
ALOD	6.31					3.03		
Intervals 4, 5, 6								
Log-likelihood	− 71.92	− 42.48	− 115.33	4.253*	− 73.27	− 39.53	− 116.94	19.023***
	1.35					2.95		ALOD
Intervals *1*, 2, 3								
Log-likelihood	− 136.98	− 92.60	− 229.69	0.449^{ns}				
n	30	15	45		30	15	45	

(b)

Marker	OB051020	OB051150	05010550	C1690270	0A080302	OA080340	C5030350
OB05 1020							
OB051150	2.680^{ns}						
05010550	0.039^{ns}	0.162^{ns}					
C1690270	1.201^{ns}	0.702^{ns}	0.014^{ns}	–			
OA080302	2.064^{ns}	0.088^{ns}	0.009^{ns}	0.842^{ns}			
OA080340	1.858^{ns}	1.338^{ns}	1.437^{ns}	1.412^{ns}	0.526^{ns}		
C5030350	1.965^{ns}	5.739*	9.391**	8.791**	7.436**	0.701^{ns}	

*, ** Significance at ≤ 0.05 and ≤ 0.01 levels, respectively; ^{ns} = non significant

information over a large number of individuals is dependent on the combined effect of the markers' frequency in the population and the distribution of heterozygous loci among individuals. Nonetheless, the strategy of studying linkage relationships between markers adopted in this investigation (building maps using ML methods and testing the error rate by numerical re-sampling) has produced results acceptably similar to a single-tree map for most of the windows analyzed and therefore may be considered a useful tool for the construction of preliminary population 'consensus' linkage maps, as well as for the identification of loci showing stability in different genetic backgrounds and unambiguous linkage relationships over different genotypes (Fig. 3). We found 21 markers, spread over the Norway spruce genome and matching the above criteria, that can be useful for merging single-tree maps as well as verifying QTL stability in different genomic backgrounds (Brondani and Grattapaglia 1996 b). For windows showing fairly high uncertainty (B and D), further analysis using an increased number of megagametophytes per tree and/or the number of heterozygous trees for the markers involved is needed.

Linkage data of allozymic markers from many highly heterozygous individuals have been produced for forest tree species. Co-linearity of allozymic markers has been reported in different conifer species (see Conkle 1981), although significant heterogeneity of recombination rates between markers among individuals has been extensively detected (Rudin and Eckberg 1978; King and Dancik 1983; Furnier et al. 1986; O'Malley et al. 1986; Barrett et al. 1987; Szmidt and Muona 1989; Plomion and O'Malley 1996).

Population linkage maps constructed using routinely collected data by tree population geneticists may have several helpful advantages. Building a "consensus" map using data from many individuals can test the reliability of markers in different genomic backgrounds (Lu et al. 1995; Brondani and Grattapaglia 1996a). Such markers may be used as consensus markers ('anchor' loci) for QTL analyses and/or may be useful for further population genetic and evolutionary studies. Furthermore, heterogeneity of pairwise recombination fractions of markers between different families of megagametophytes (i.e. trees) could suggest the existence of different genomic rearrangements maintained within

the population ("non-allelic heterogeneity", Ott 1991). We used cluster analysis of recombination values between markers and showed the existence of two subpopulations carrying alternative marker orders. Trees hypothesized as carrying alternative genomic arrangements will be focused for targeted mapping of the region(s) of interest (Reiter et al. 1992), with the goal of establishing a detailed map of these chromosomal segments.

Integration of single-tree linkage information could also lead to a 'species consensus map' by using highly polymorphic markers (RAPD, AFLP, SSR, etc.) showing stability over trees from several provenances all across the species' natural range. Such markers with a well-known genetic basis could be used as 'general' markers in a large survey of genetic variability, as well as in investigations aimed to verify the consistency of QTL location in different genomic backgrounds.

Acknowledgments The authors thank Dr. Elena Maestri for critical reading of the manuscript, Glen N. Johnson for skillful technical assistance, and Prof. Raffaello Giannini for having introduced us to the study of Norway spruce. We also thank Dr. O'Malley for his thorough review of an earlier version of the manuscript and for suggesting a validation of our map comparison by bootstrap. Research supported in part by the Italian National Council of Research (C.N.R., Comitato 13), the Italian Ministry of University and Scientific Research (MURST 40%), and in part by USDA Forest Service, Southern Institute of Forest Genetics.

References

Adams WT, Joly RJ (1980) Linkage relationships among twelve allozyme loci in loblolly pine. J Hered 71: 199–202

Barrett JW, Cheliak WM, Knowles PH (1987) Variation in the PGI/AAT linkage group between populations of black spruce. Can J For Res 17: 756–758

Beavis WD, Grant D (1991) A linkage map based on information from four F_2 populations of maize (Zea *mays* L.). Theor Appl Genet 82 : 636–644

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, co-linearity and compatibility. Trends Genet 9 : 259-261

Binelli G, Bucci G (1994) A genetic linkage map of Picea *abies* Karst. based on RAPD markers as a tool in population genetics. Theor Appl Genet 88 : 283–288

Bradshaw HD Jr, Stettler RF (1995) Molecular genetics of growth and development in *Populus* IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139 : 963–973

Bradshaw HD Jr, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). Nature 376: 762–765

Brondani RPV, Grattapaglia D (1996a) Comparative mapping in Eucalyptus reveals significant RAPD locus linkage and order conservation between trees and greater meiotic recombination in females. Plant Genome IV Meeting, S.Diego, January 11–18, Poster # 235

Brondani RPV, Grattapaglia D (1996 b) Mapped RAPD markers are transferable among *Eucalyptus* trees from the same population. Plant Genome IV Meeting, S.Diego, January 11-18, Poster # 236

Bucci G, Binelli G, Menozzi P (1995) Identification of a new set of molecular markers in *Picea abies* Karst. as revealed by random amplification techniques. In: Baradat P, Muller-Stark G, Adams WT (eds) Population genetics and gene conservation of forest trees. SPB Academic Publishing bv, The Hague, The Netherlands, pp 121-127

Bucci G, Menozzi P (1995) Genetic variation of RAPD markers in a *Picea abies* Karst. population. Heredity 75: 188-197

Burr B, Burr FA, Thompson KH, Albertsen MC, Stuber C (1988) Gene mapping with recombinant inbreds in maize. Genetics 118 : 519–526

Conkle MT (1981) Isozyme variation and linkage in six conifer species. In: Conkle MT (ed) Proceedings of the Symposium on Isozyme of North American Forest Trees and Forest Insects. USDA For Serv Gen Tech Rep, PSW-48, pp 11-17

Cooperative Human Linkage Center – CHLC (1994) A comprehensive human linkage map with centimorgan density. Science 265 : 2049–2054

Duhram RE, Liou PC, Gmitter FG Jr, Moore GA (1992) Linkage of restriction fragment length polymorphisms and isozymes in *Citrus* Theor Appl Genet 84: 39-48

Efron B (1982) The jacknife, the bootstrap and other re-sampling plans. In: Applied Mathematics, # 38, CBMS-NSF Regional Conference Series. Society for Industrial and Applied Mathematics, Philadelphia, USA

Ellis THN (1994) Estimating map distances. Trends Genet 10: 113-114

Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR (1992) Linkage maps in pea. Genetics 130 : 649–663

Fritsch P, Rieseberg LH (1992) High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Dactylis glomerata* Nature 359: 633–636

Furnier GR, Knowles P, Aleksiuk MA, Dancik BP (1986) Inheritance and linkage of allozymes in seed tissue of whitebark pine. Can J Genet Cytol 28 : 601–604

Geburek T, von Wuehlisch G (1989) Linkage analysis of isozyme gene loci in *Piceu abies* Karst. Heredity 62: 185-191

Gerber S, Rodolphe F (1994) Estimation and test for linkage between markers: a comparison of lod score and χ^2 tests in a linkage study of maritime pine (*Pinus pinaster* Ait.). Theor Appl Genet 88 : 293–297

Gerber S, Rodolphe F, Bahrman N, Baradat P (1993) Seed-protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis: genetic determinism and construction of a linkage map. Theor Appl Genet 85 : 521–528

Graner A, Jahoor A, Schoendelmaier J, Siedler H, Pillen H, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an RFLP map in barley. Theor Appl Genet 83 : 250–256

Grattapaglia D, Sederoff R (1994) Genetic linkage map of *Eucalyptus grundis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137 : 1121–1137

Hauge BM, Hanley SM, Cartinhour S, Cherry JM, Goodman HM, Koorrneef M, Stam P, Chang C, Kempin S, Medrano L, Meyerowitz EM (1993) An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J 3 : 745-754

Kaga A, Ohnishi M, Ishii T, Kamijima 0 (1996) A genetic linkage map of azuki bean constructed with molecular and morphological markers using an interspecific population (*Vigna unguicularis x v. nakashimae*). Theor Avvl Genet 93 : 658–663

Kianian SF, Quiros CF (1992) Generation of a *Brassica oleruceu* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor Appl Genet 84: 544–554

King JN, Dancik BP (1983) Inheritance and linkage of isozymes in white spruce. Can J Genet Cytol 25 : 430–436

Kubisiak TL, Nelson CD, Nelson WL, Stine M (1995) RAPD linkage mapping in a longleaf pine x slash pine F_1 family. Theor Appl Genet 90: 1119–1127

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84: 2363–2367

Lander ES, Green P, Abrahamson J, A. Barlow, M. Daly, S. Lincoln L. Newburg (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181

Lathrop GM, Lalouel M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81 : 3443–3446

Liu B (1996) Framework map and confidence of gene order. SRIEG Conference: Novel Applications of Molecular Markers, Texas A&M University, IBT, Houston, June 23–26

Lu M-Z, Szmidt AE, Wang X-R (1995) Inheritance of RAPD fragments in haploid and diploid tissues of *Pinus sylvestris* (L.). Heredity 74 : 582–589

Mitchell-Olds T (1995) The molecular basis of quantitative variation in natural populations. Trends Ecol Evol 10: 324–328

Morton NE (1955). Sequential tests for the detection of linkage. Am J Hum Genet 7 : 277–318

Nelson CD, Nance WL, Doudrick RL (1993) A partial genetic linkage map of slash pine (*Pinus elliottii* Englm. var. *elliottii*) based on random amplified polymorphic DNAs. Theor Appl Genet 87: 145–151

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Nature 258 : 67–86

O'Malley DM, Guries RP, Nordheim EV (1986) Linkage analysis for 18 enzyme loci in *Pinus rigida* Mill. Theor Appl Genet 72: 530–535

Ott J (1991) Analysis of human genetic linkage (revised edition). John Hopkins University Press, Baltimore

Plomion C, O'Malley DM (1996) Recombination rate differences for pollen and seed parents in *Pinus pinaster* Heredity 77 : 341-350

Plomion C, Bahrman N, Durel C-E, O'Malley DM (1995a) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. Heredity 74 : 661–668

Plomion C, O'Malley DM, Durel C-E (1995 b) Genomic analysis in Maritime pine (*Pinus pinaster*). Comparison of two RAPD maps using selfed and open-pollinated seeds from the same individual. Theor Appl Genet 90: 102881034

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992). Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89 : 1477–1481

Rieseberg LH (1996) Homology among RAPD fragments in interspecific comparisons. Mol Ecol 5 : 99-105

Rieseberg LH, Van Fossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375: 313-316

Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM (1996) Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. Science 272 : 741–745

Rudin D, Eckberg I (1978) Linkage studies in *Pinus sylvestris* L. using macrogametophyte allozymes. Silvae Genet 27 : 1-12

Slocum MK, Figdor SS, Kennard WC, Suzuki JI, Osborn TC (1990) Linkage arrangements of restriction length polymorphism loci in *Brassica oleracea* Theor Appl Genet 80: 57–64

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3 : 739–744

Strauss SH, Conkle MT (1986) Segregation, linkage and diversity of allozymes in knobcone pine. Theor Appl Genet 72 : 4833493

Suther GK (1991) Estimation of an approximate confidence interval for FRAXA location by using linkage data from many pedigrees. Am J Hum Genet 49 : 462–464

Suther GK, Wilson SR (1990) Genetic counseling in rare syndromes: a re-sampling method for determining an approximate confidence interval for gene location with linkage data from single pedigree. Am J Hum Genet 47 : 53-61

Szmidt AE, Muona 0 (1989) Linkage relationships of allozyme loci in *Pinus syloestris.* Hereditas 111 : 91–97

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27 : 205–233

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ et al. (1992) High-density molecular linkage maps of the tomato and potato genome. Genetics 132 : 1141–1 160

Weir BS (1990) Genetic Data Analysis. Sinauer Associates, Inc. Publishers, Sunderland, Mississippi

Weissebach J, Gyapay G, Dib C, Vignal A, Morisette J, Millasseau P, Vaysseix G, Lathrop GM (1992) A second generation linkage map of the human genome. Nature 359 : 7944801