Construction of a genetic linkage map for *Camellia sinensis* (tea)

CHRISTINE A. HACKETT*†, FRANCIS N. WACHIRA‡[1], SATISH PAUL‡[2] WAYNE POWELL‡ & ROBBIE WAUGH‡

†*Biomathematics & Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK,* ‡*Department of Cell & Molecular Genetics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK*

Genetic maps are a vital tool in cultivar improvement programmes for woody perennial tree crops such as tea (*Camellia sinensis*). A population thought to be derived from two known, noninbred parents was scored for RAPD and AFLP markers, in order to develop a linkage map. However, a very high proportion of the markers exhibited unexpected segregation ratios in the light of their configurations in the parents, and an exploratory statistical analysis revealed patterns in the marker scores which can most easily be explained by the hypothesis of three male parents contributing pollen to this cross. We discuss the evidence for this and the subsequent analysis required to assemble the markers from the female parent into the first linkage map for tea. The map has 15 linkage groups of three or more markers, agreeing with the haploid chromosome number of tea. The statistical methods that revealed the subpopulations are easy to apply routinely, and may prove a useful diagnostic tool for the analysis of noninbred mapping populations.

Keywords: *Camellia*, cluster analysis, diagnostics, linkage, mapping, outbreeder.

Introduction

Genetic linkage maps provide great potential for increasing the speed and precision of cultivar improvement programmes for woody perennial tree crops. Such crops often have long juvenile periods, large plant size and are predominantly allogamous and highly heterozygous. All of these factors hinder conventional breeding by requiring large investments in time and land for progeny trials, and genetic maps provide a vital tool. However, few such maps exist for tree crops. Unlike the short-lived annual plant species, the development of specific genetic stocks and tester lines containing important traits is time consuming and may be hindered by self-incompatibility. The availability of the multigeneration pedigrees required for genome mapping is therefore limited. Pedigrees available for many outbred tree species, including tea, involve only two parents and their full-sib progeny, or maternal half-sib families.

There has therefore been a need to explore alternative approaches to mapping in such species. The approach now widely used (a 'pseudo test cross') involves the use of existing pedigrees in breeding programmes (Grattapaglia & Sederoff, 1994; Hemmat *et al.*, 1994). Heterozygous plants from noninbred species have been successfully used for mapping, e.g. in potato (Gebhardt *et al.*, 1989; Freyre & Douches, 1994), *Picea abies* (Binelli & Bucci, 1994), apple (Hemmat *et al.*, 1994), eucalyptus (Grattapaglia & Sederoff, 1994) and alfalfa (Echt *et al.*, 1994).

Tea (*Camellia sinensis*) is a beverage tree crop native to SE Asia, and has been introduced into many other countries. Breeding strategies in tea depend on its highly outbred nature, long generation time from seed to flower and its potential for vegetative propagation. The harvestable yield of tea is confined to the terminal two leaves and a bud, which constitute only 10–18% of the total biomass and dry matter produced by the plant (Magambo & Cannell, 1981). Leaf characteristics therefore form the basis of selection for the two most important agronomic traits: yield and quality. These traits require several years to develop, and are not amenable to early selection. Thus selection based on one or more molecular markers linked to a quantitative trait

*Correspondence. E-mail: chacke@scri.sari.ac.uk

[1]Permanent address: Tea Research Foundation of Kenya, PO Box 820, Kericho, Kenya

[2]Permanent address: HP Agricultural University, Palampur 176062, India

locus (QTL) could potentially shorten the breeding cycle. However, until now no genetic map has been available for tea.

In this study, a pseudo test cross approach was adopted to generate a map for tea using RAPD and AFLP markers. The choice of mapping population used was based largely on population size because most tea populations raised from controlled crossing are small. In the course of the linkage analysis, a large proportion of markers were found to have unexpected segregation ratios. An exploratory statistical analysis indicated the presence of three subpopulations, which we hypothesize to have different male parents. These subpopulations were not revealed by standard linkage analysis software, but they have a considerable impact on the linkage map. In this paper we describe the analysis of the molecular marker data, the evidence for the contribution of three male parents to the population and the estimation of a linkage map allowing for this. The statistical methods that revealed the subpopulations are easy to apply routinely, and may prove a useful diagnostic tool for the analysis of noninbred mapping populations.

Materials and methods

Plant material

The mapping population consisted of 90 genotypes, thought to be the F1 generation of a cross between two noninbred diploid clones SFS150 (a Malawian 'assam' type local selection of unknown pedigree) and TN14/3 (a Kenyan 'china' type local selection clone of unknown pedigree). The population was planted out in the field at the Tea Research Foundation of Kenya situated at Kericho, Kenya in 1991.

Molecular techniques

DNA isolation was performed using the modified method of Gawel & Jarret (1991) as described by Orozco-Castillo *et al.* (1994). RAPD and AFLP methods were performed as described by Wachira *et al.* (1995) and Paul *et al.* (1997), respectively. Primers used for RAPD and AFLP analysis are listed in Table 1. The names of individual RAPD and AFLP loci were described using the PCR primer code name followed by the relative marker size, i.e. from the largest to the smallest, using alphabetical letters a,b,c, etc. All loci were scored twice, independently, to minimize scoring errors. Data were collected in the mapping population for all polymorphic segregating marker alleles, whether these were from one parent or both.

Results

Segregation analysis

An initial screen of 120 random decamer RAPD primers on 52 progeny of the population revealed that only 44 (36.7%) were polymorphic. Thirteen primers (see Table 1) were screened on the entire population of 90 progeny and the parents, yielding 141 bands, of which 26 were unambiguously polymorphic (2.0 polymorphic bands/polymorphic primer). Forty-one AFLP primer combinations were screened on the mapping population. The number of visible AFLP bands per gel ranged between 32 and 150 with a mean of 84.7 and an average of 10.5 polymorphic bands per primer combination. AFLP fragments amplified ranged in size from 65 bp to 450 bp. *Eco*RI/*Mse*I primer combinations amplified a slightly larger number of total visible bands (a mean of 88.7 per primer combination) compared to the *Pst*I/*Mse*I primer combinations (a mean of 73.3 per primer combination).

In all, 420 segregating bands (RAPDs and AFLPs) were scored for the parents and the offspring population. Of these bands, 116 occurred in both parents, and so would be expected to segregate in a 3:1 ratio in the F1 population. There were 208 bands that occurred in the maternal parent only and 96 in the paternal parent only, and these would be expected to segregate in a 1:1 ratio.

The modules of JOINMAP 2.0 (Stam & Van Ooijen, 1995) were used for linkage analysis. However, the JoinMap single locus analysis module (JMSLA) indicated that a high proportion of the markers had ratios that were significantly different from those expected (Table 2). The markers from the paternal parent had the highest proportion of distortions, and those from the maternal parent had least. When the distorted markers were omitted from the dataset, a linkage map could be estimated (using JOINMAP). However, most of the undistorted markers from the male parent were found to be unlinked to other markers and could not be placed on the map.

Exploratory analysis

Multivariate analyses were used to explore the distorted ratios in the set of markers from the male parent. Cluster analysis (Digby & Kempton, 1987, chapter 5) is one such technique, and may be used in a genetic context to cluster either the markers or the individuals. This requires the calculation of a measure of similarity between all pairs of markers or individuals and a suitable measure for presence/absence data is the simple matching coefficient, defined as the proportion of

Table 1 Sequences of primers* and adaptors used in mapping studies

(a) RAPD primers

Code name	Sequence
SC10–37	5′-GCCAATCCTG-3′
SCRI-71	5′-CTGGCGTAGT-3′
SC10–82	5′-GCCGTGAAGT-3′
SC10–83	5′-AATGACGGGG-3′
SC10–91	5′-CTCGACTAGG-3′
OPE-20	5′-AACGGTGACC-3′
OPS-07	5′-TCCGATGCTG-3′
OPU-20	5′-ACAGCCCCCA-3′
OPV-15	5′-CAGTGCCGGT-3′
OPW-02	5′-ACCCGCCAA-3′
OPW-11	5′-CTGATGCGTG-3′
OPX-09	5′-GGTCTGGTTG-3′
OPY-17	5′-GACGTGGTGA-3′

(b) AFLP primers

Primer	Sequence			
*Eco*RI adaptor	5′-CTCGTAGACTGCGTACC-3′ 3′-CTGACGCATGGTTAA-5′			
*Pst*I adaptor	5′-CTCGTAGACTGCGTACATGCA-3′ 3′-CATCTGACGCATGT-5′			
*Mse*I adaptor	5′-GACGATGAGTCCTGAG-3′ 3′-TACTCAGGACTCAT-5′			
E00 primer	5′-GACTGCGTACCAATTC-3′			
Primers with 3′ selective bases	E01(A) E34(AAT) E32(AAC) E36(ACC) E33(AAG) E37(ACG)	E38(ACT) E39(AGA) E40(AGC)	E41(AGG) E42(AGT) E43(ATA)	E44(ATC) E46(ATT) E88(TGC)
P00 primer	5′-GACTGCGTACATGCAG-3′			
Primers with 3′ selective bases	P01(A) P14(AT) P12(AC) P15(CA) P13(AG) P16(CC)	P17(CG) P25(TG) P31(AAA)	P34(AAT) P40(AGC) P46(ATT)	P55(CGA) P95(AAAA)
M00 primer	5′-GATGAGTCCTGAGTAA-3′			
Primers with 3′ selective bases	M01(A) M34(AAT) M12(AC) M35(ACA) M31(AAA) M38(ACT) M32(AAC) M39(AGA) M33(AAG) M41(AGG)	M42(AGT) M45(ATG) M46(ATT) M47(CAA) M48(CAC)	M49(CAG) M50(CAT) M52(CCC) M54(CCT) M55(CGA)	M62(CTT) M81(TAG) M95(AAAA)

*For AFLP primers, only the selective bases are presented (in parentheses).

common presences and absences. If cluster analysis is being used to group markers, the simple matching coefficient is equal to $1 - r$, where r is the recombination fraction between 1:1 markers in a coupling phase and a cluster analysis based on this coefficient assembles the markers into linkage groups. However, a cluster analysis to group the individuals in an F1 population would not be expected to show any particular patterns of common presences and absences of each marker. Figure 1 com-

Table 2 Number of markers showing a significant distortion from the expected segregation ratios

Parental origin	Expected ratio	No. observed	No. significantly distorted	
			$P = 1\%$	$P = 0.1\%$
Paternal	1:1	96	55	47
Maternal	1:1	208	43	31
Both	3:1	116	40	28

Fig. 1 (a) Dendrogram obtained by average linkage cluster analysis of the 90 offspring genotypes, on the basis of their scores from the male parent TN14/3. The three groups of individuals are indicated. The individual at the top of the plot had a very unusual DNA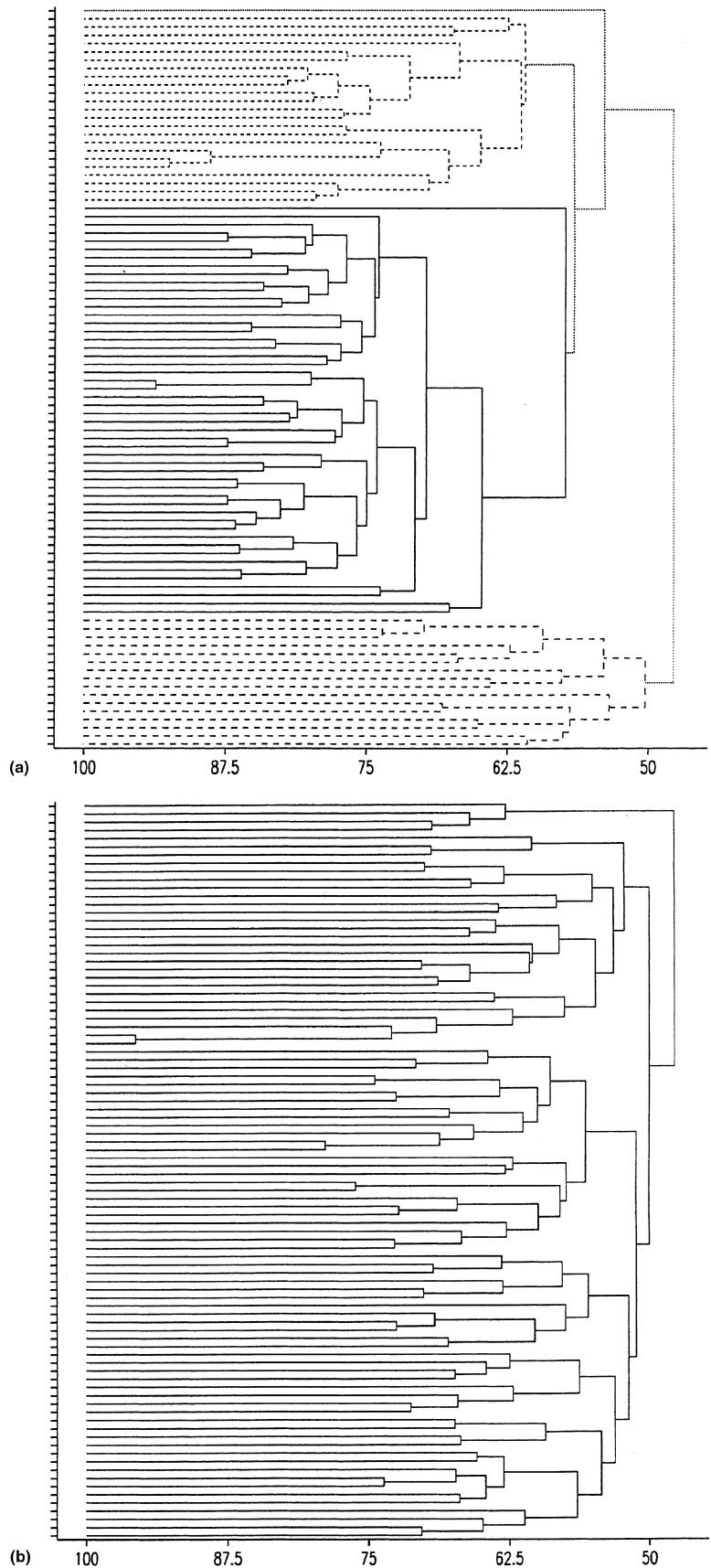 track and was one of those omitted from the subsequent analysis. (b) Dendrogram obtained by average linkage cluster analysis of the 90 offspring genotypes, on the basis of their scores from the female parent SFS150.
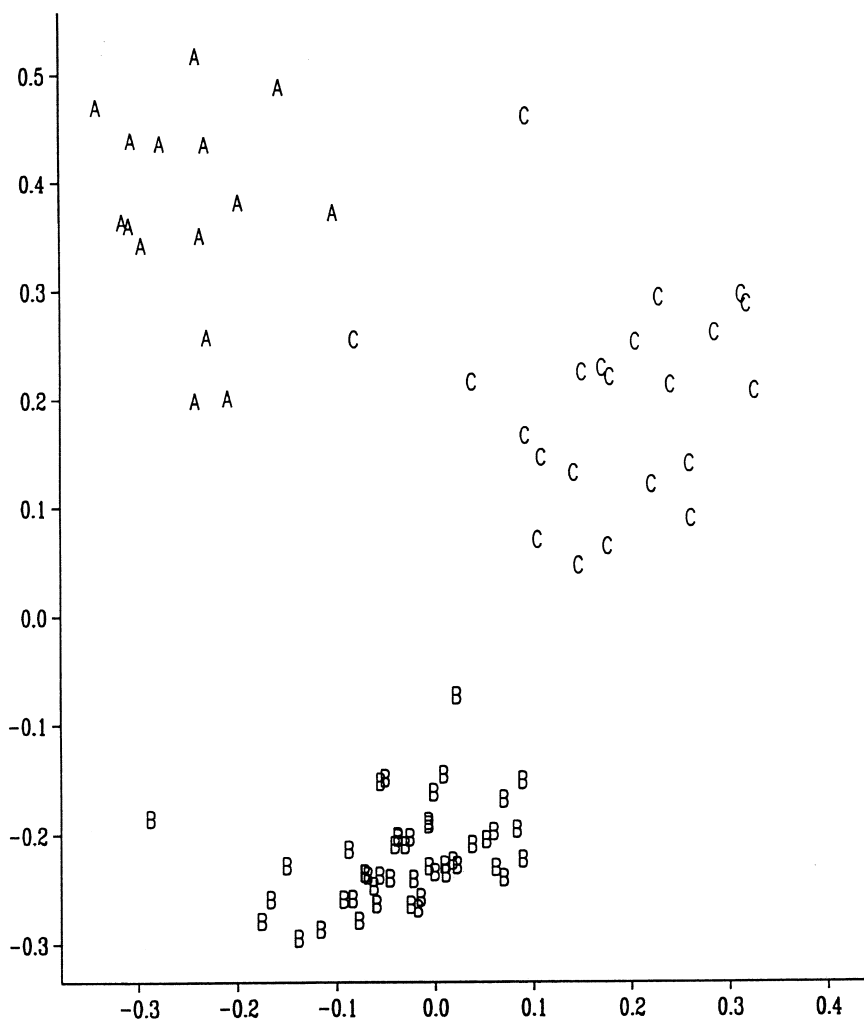
Fig. 2 A plot of the first two principal coordinate scores, indicating the hypothesized division into groups A, B and C.

pares the dendrograms obtained by an average linkage cluster analysis of the 90 progeny, based on (a) the 96 male parent markers and (b) the 208 female parent markers. Very little structure is observed in the dendrogram based on the female parent markers, but there is noticeable grouping in that based on the male parent markers. At a similarity level of 70%, 47 of the 90 individuals had formed a cluster, and at a similarity level of 60% 23 further individuals had formed a second cluster. We will refer to the cluster of 47 as group B, the cluster of 23 as group C and the remainder as group A.

A principal coordinate analysis (Digby & Kempton, 1987, chapter 3) of the similarities of the individuals, calculated from the 96 markers from the male parent, indicated that the first principal coordinate accounted for 16.6% of the variation, and separated group B from the remainder. The second principal coordinate accounted for 7.8% and separated group C. Figure 2 shows a plot of the first two principal coordinates. A χ^2 analysis showed that there were significant differences among the segre-

Table 3 Number of markers showing significant differences in segregation ratios between groups A, B and C

Parental origin	No. observed	No. significantly distorted	
		$P = 1\%$	$P = 0.1\%$
Paternal	96	65	57
Maternal	208	22	15
Both	116	23	11

gation ratios in groups A, B and C (Table 3). For 53 out of the 65 markers from the male parent where there were significant differences, group A carried the band significantly more often than at least one of groups B and C. Groups A, B and C were defined on the basis of the segregation patterns of markers from the male parent only, so if they are chance groupings then few significant differences would be expected in the segregation ratios of the markers from the female parent only, and the markers found in both parents. However, there were far more

significant differences among the groups A, B and C for these markers than would be expected by chance (Table 3). For 20 of the 22 significant markers from the female parent, group A carried the band significantly less often than at least one of groups B and C, while for 15 of the 23 significant markers from both parents, group A carried the band significantly more often than at least one of groups B and C.

When the markers were being scored, a few bands were observed as present in some offspring, but absent in both parents. These were regarded initially as artefacts and ignored, but 12 such bands were now investigated. One of these bands was of particular interest as it was found in 46 of the 47 individuals designated already as group B. Two of group C had the band and one of group A. A second band segregated in group B, but did not occur in groups A or C. Five bands segregated in group C, but did not occur in groups A or B, and four bands segregated in groups B and C, but not in group A. One band segregated in groups B and C, and also occurred in two individuals of group A. At this stage it was also decided to exclude five individuals from the F1 population as their DNA tracks did not permit accurate scoring. Excluding these individuals, and reexamining the principal coordinate scores, groups A, B and C were defined as in the principal coordinate plot (Fig. 2), with 15, 48 and 22 individuals, respectively.

On the basis of these observations it was hypothesized that the pollen used to make this cross was not from a single plant, but from three plants, possibly related, and therefore that the offspring population actually consists of three half-sib populations. Group A offspring actually have the intended male parent, TN14/3, while groups B and C have different male parents, whose genotypes at the marker loci may only be deduced from the segregations in the offspring. We will consider each set of markers in turn, and consider how this hypothesis fits the observations. The hypothesized male parents for groups B and C will be denoted by MB and MC.

(a) Markers from male parent TN14/3 only We observed that for the majority of markers found in male parent TN14/3, but not in the female parent, there were significant differences between groups A, B and C, and generally group A carried the marker more often than groups B and C. If MB and/or MC do not carry these markers we would not expect to find such markers in groups B and/or C. There were five markers where all or all but one of group A carried a marker, which segregated or was absent in groups B and/or C. This is explained if TN14/3 is homozygous for such a marker, while MB and MC are heterozygous or do not carry it. Ten markers were present for all or all but one of group B, a significantly higher proportion than in group A: we

explain these on the hypothesis that parent MB is homozygous for these markers while TN14/3 is a heterozygote. Our hypothesis runs into difficulties with the segregation patterns of 6 of the 96 markers from TN14/3, where at most one of group A carries the band. There are only 15 individuals in group A so it is possible that no individual carries that marker by chance. Other explanations are more alarming: an artefact in the parental DNA, or a still more complicated parental structure to the population.

(b) Markers from the female parent SFS150 only We observed that a much lower proportion of the markers from the female parent showed differences between groups A, B, and C, although this proportion was higher than would be expected by chance. For 20 out of the 22 significant markers groups B and C carried higher proportions of the marker than group A. This can be explained as those markers carried by MB and/or MC in addition to the female parent, so that the expected segregation ratios change from 1:1 to 3:1 or 1:0 depending on whether the marker is heterozygous or homozygous in MB and/or MC. Both of these ratios were observed in the population.

(c) Markers from SFS150 and TN14/3 116 markers were apparently present in the heterozygous state in both parents, with 23 showing significantly different proportions between groups A, B and C. For 15 of these, group A had significantly higher proportion than groups B and/or C. This may be explained in a similar manner to (a), by the absence of the marker in parents MB and/or MC. For these markers, the proportion present in groups B and C was never significantly lower than a 1:1 ratio, consistent with presence of the marker in just the female parent. Eight of the markers were present for every individual of group A, suggesting that these markers were homozygous in TN14/3 and heterozygous or absent in MB and MC. Seven markers had significantly lower proportions in group A than in groups B and/or C, where they were almost always present: this is consistent with MB and/or MC being homozygous for the marker.

Developing a linkage map

The hypothesis of three male parents explains a large proportion of the features observed in this dataset. However, there are a few markers whose segregations are difficult to explain by this hypothesis. It is also not possible to state categorically that there are no more than three male parents involved, to be certain of the male parent associated with every offspring, or to be certain of the genotype of the hypothesized parents MB
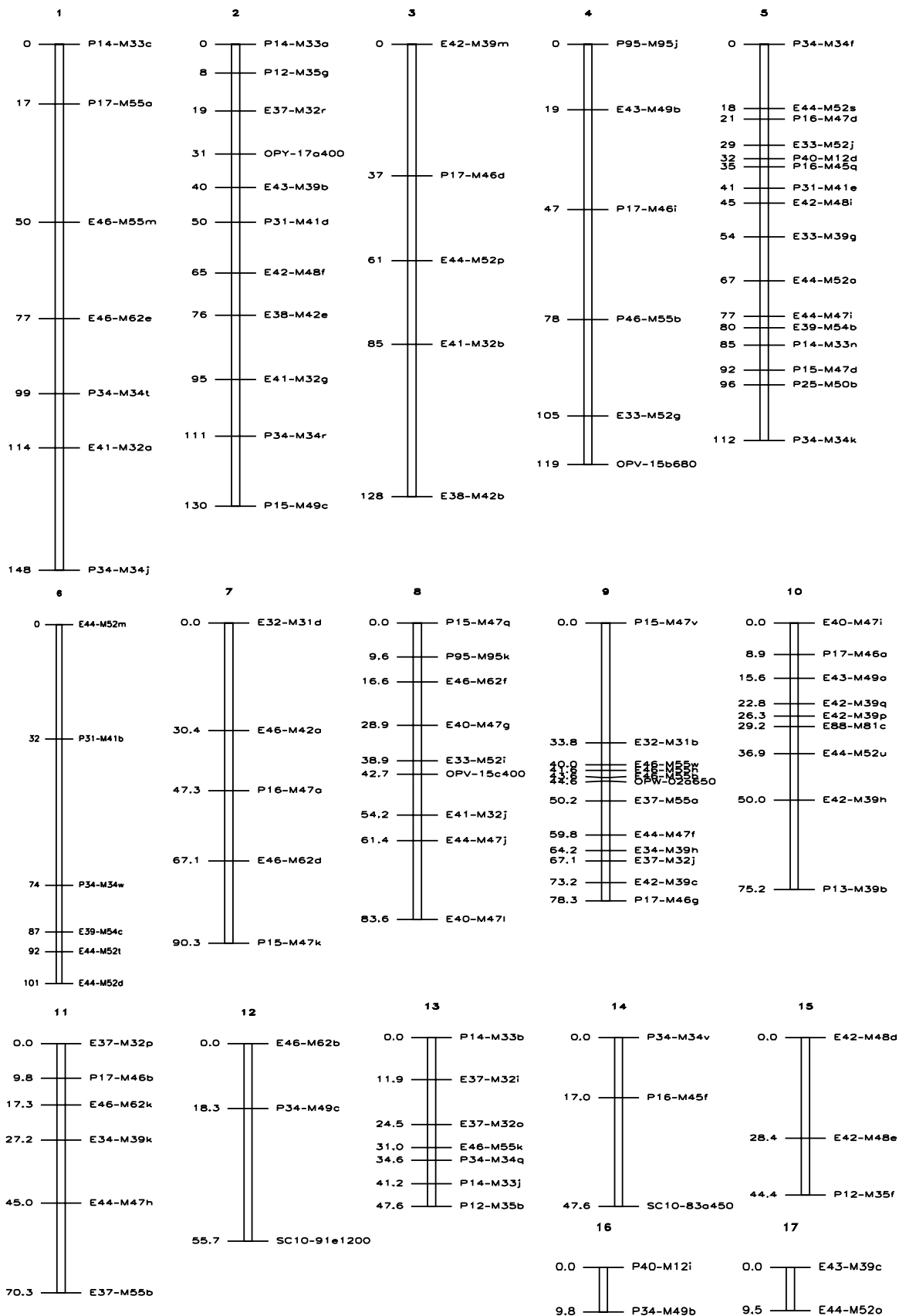
1

0	P14-M33c
17	P17-M55o
50	E46-M55m
77	E46-M62e
99	P34-M34t
114	E41-M32o
148	P34-M34j

2

0	P14-M33o
8	P12-M35g
19	E37-M32r
31	OPY-17o400
40	E43-M39b
50	P31-M41d
65	E42-M48f
76	E38-M42e
95	E41-M32g
111	P34-M34r
130	P15-M49c

3

0	E42-M39m
37	P17-M46d
61	E44-M52p
85	E41-M32b
128	E38-M42b

4

0	P95-M95j
19	E43-M49b
47	P17-M46i
78	P46-M55b
105	E33-M52g
119	OPV-15b680

5

0	P34-M34f
18	E44-M52s
21	P16-M47d
29	E33-M52j
32	P40-M12d
35	P16-M45q
41	P31-M41e
45	E42-M48i
54	E33-M39g
67	E44-M52o
77	E44-M47i
80	E39-M54b
85	P14-M33n
92	P15-M47d
96	P25-M50b
112	P34-M34k

6

0	E44-M52m
32	P31-M41b
74	P34-M34w
87	E39-M54c
92	E44-M52t
101	E44-M52d

7

0.0	E32-M31d
30.4	E46-M42o
47.3	P16-M47o
67.1	E46-M62d
90.3	P15-M47k

8

0.0	P15-M47q
9.6	P95-M95k
16.6	E46-M62f
28.9	E40-M47g
38.9	E33-M52i
42.7	OPV-15c400
54.2	E41-M32j
61.4	E44-M47j
83.6	E40-M47l

9

0.0	P15-M47v
33.8	E32-M31b
40.0	E46-M55w
41.8	E46-M55
43.2	E46-M55
44.6	OPW-02o650
50.2	E37-M55o
59.8	E44-M47f
64.2	E34-M39h
67.1	E37-M32j
73.2	E42-M39c
78.3	P17-M46g

10

0.0	E40-M47i
8.9	P17-M46o
15.6	E43-M49o
22.8	E42-M39q
26.3	E42-M39p
29.2	E88-M81c
36.9	E44-M52u
50.0	E42-M39h
75.2	P13-M39b

11

0.0	E37-M32p
9.8	P17-M46b
17.3	E46-M62k
27.2	E34-M39k
45.0	E44-M47h
70.3	E37-M55b

12

0.0	E46-M62b
18.3	P34-M49c
55.7	SC10-91e1200

13

0.0	P14-M33b
11.9	E37-M32i
24.5	E37-M32o
31.0	E46-M55k
34.6	P34-M34q
41.2	P14-M33j
47.6	P12-M35b

14

0.0	P34-M34v
17.0	P16-M45f
47.6	SC10-83o450

15

0.0	E42-M48d
28.4	E42-M48e
44.4	P12-M35f

16

0.0	P40-M12i
9.8	P34-M49b

17

0.0	E43-M39c
9.5	E44-M52o

Fig. 3 A linkage map of tea, developed from the markers present in the female parent only, and segregating in a 1:1 ratio in each of the offspring groups.

and MC at each locus. Given this, the only markers that may reasonably be mapped are those present in the female parent only, and segregating in a 1:1 ratio in all three groups of offspring. Any marker with a segregation ratio significantly different from 1:1 at the 5% level for any of the offspring groups was excluded. A total of 126 markers was suitable for mapping by this criterion.

Separate matrices containing the recombination fractions between each pair of the 126 suitable markers were calculated for each of groups A, B and C, and Mantel's test (Mantel, 1967) was used to test for consistency. This is a test based on the correlation coefficient between pairs of matrices, so the matrices of recombination fractions from groups A and C were compared in turn with that of group B, as this is the largest group and so provides the most precise estimates. A permutation test with 200 permutations was used to calculate the significance. The correlations between the recombination fractions for groups A and B, and C and B, were 0.32 and 0.31, respectively, and the 95% points from the permutations were 0.022 and 0.020. This confirms that there is a close relationship between the estimates from the three groups.

Given this relationship, the individuals from the three groups were combined, and linkage groups were estimated using JOINMAP 2.0 (Stam & Van Ooijen, 1995). At a lod of 4.0 there were 15 groups of three or more markers, three pairs and 13 markers unallocated. If the lod score was reduced to 3.5 there were 15 groups of three or more markers, two pairs and 10 markers unallocated. The markers were ordered within their linkage groups using JMMAP. The mean χ^2 statistic for

approximate goodness-of-fit indicated just one marker as problematical, with a jump of 13.3 in the mean χ^2 statistic. This marker was excluded from the map. The resulting map, at a lod score of 3.5, is shown in Fig. 3. This was drawn using DRAWMAP (Van Ooijen, 1994). The map covered 1349.7 cM, with an average distance of 11.7 cM between loci. Tea is known to have a haploid chromosome number of 15, which agrees with the number of groups with three or more markers.

As a final check on the map, matrices of recombination fractions were calculated from each of the three groups, for each group of five or more markers separately, and Mantel's test was again used as a test of consistency. Of the 12 groups of this size, the correlations between the recombination fractions from groups A and B, and groups B and C, were both larger than the 95% point from the permutation test for nine groups. For the other three groups, one of the correlations was smaller than the 95% point from the simulation test, and in each case there was a single marker involved in all the discrepant linkages. One such group was linkage group 7, where linkages with marker E46-M62d differed between groups A and B. Table 4 shows the joint segregations of this marker and its neighbour, P16-M47a. Marker E46-M62d segregates in a 11:4 ratio in group A, which is compatible with either a 1:1 or a 3:1 ratio. If both markers are regarded as segregating in a 1:1 ratio the maximum likelihood estimate of the recombination fraction between them is 0.40, but if E46-M62d is actually present in both parents for group A, and thus segregating in an expected 3:1 ratio, then the maximum likelihood estimate of the recombination fraction is 0.27, which is more consistent with the estimates from groups B and C. The χ^2 goodness of fit statistic is 3.67 for the 1:1 model and 0.08 for the 3:1 model, both with 1 d.f. Similar investigations revealed the same patterns affecting the segregation of marker P14-M33j (linkage group 3) in group A, and E42-M39m (linkage group 3) in group C. In every case the recombination fraction is smaller if the segregation is 3:1, and so the markers should be slightly more closely linked than shown in Fig. 3. Apart from these three markers the maps are consistent in the three offspring groups, and provide a first linkage map in tea.

Discussion

In this paper we have developed a linkage map of tea, based on AFLP and RAPD markers segregating in the female parent SFS150. In the early stages of this study, all markers segregating with a 1:1 ratio were assembled into a map using JOINMAP, and only the large number of distorted segregation ratios indicated that this analysis was unsatisfactory. However, a multivariate analysis

Table 4 Joint segregations of markers E46-M62d and P16-M47a in the three offspring groups A, B and C. r is the recombination fraction calculated assuming that the markers are both present in just the female parent, and linked in a repulsion phase

Offspring group	E46-M62d	P16-M47a Present	P16-M47a Absent	r
A	Present	5	6	0.40
	Absent	3	1	
B	Present	4	21	0.21
	Absent	16	6	
C	Present	4	5	0.30
	Absent	9	2	

revealed groupings of the individuals, and a high proportion of markers showing different segregation ratios among the groups. These unexpected features are explained most easily by the hypothesis of three male parents contributing to the population, although the hypothesis of more than three male parents cannot be excluded.

A fundamentally different interpretation could hypothesize that the male parent groupings were the result of methylation changes in groups of progeny from a single male parent. This is possible as both *Eco*RI and *Pst*I restriction enzymes are sensitive to DNA template methylation to different degrees (the latter generally being more severely affected). However, there were no significant differences between the degree of distortion between *Eco*RI and *Pst*I derived AFLPs (data not shown). Nevertheless, changes in the pattern of DNA templates used for AFLP assays are known to generate a significant portion of the polymorphisms in certain species and some effect of this cannot be ruled out (N. Ellis, pers. comm.). While we consider this to be an unlikely cause for the designation of the three male parental groupings, changes in methylation would offer a convenient explanation for the appearance or disappearance of certain low frequency markers in the population, e.g. the markers from TN14/3 which are present in only one of the 15 group A individuals.

These observations on the male parent aside, a linkage map has been calculated from the markers from the female parent only and segregating in a 1:1 ratio in the three groups of progeny. Groups A and C are small, so we cannot eliminate the possibility that some markers are distorted 3:1 markers. The test for independent segregation used by JOINMAP to group markers into linkage groups is robust to segregation distortion (Stam & Van Ooijen, 1995), so we can have confidence in the identification of linkage groups. It can be shown that we will overestimate the recombination frequency by assuming we have two 1:1 markers when the true situation is a 1:1 marker linked to a 3:1 marker, and the bias is greatest for very closely linked markers. The Mantel test was used to compare recombination frequencies from groups A and C with those from group B, as this group is sufficiently large to have confidence that 3:1 markers have been excluded. There were only three significant discrepancies. We can have confidence that this map of tea is substantially correct.

The findings described here do not seem to be uncommon. A similar analysis revealed mixed parentage in a population of Norway spruce (P. Hedley, pers. comm.), where the grouping by cluster analysis was subsequently confirmed by analysis of microsatellite markers. An unusual group of individuals was also detected in a mapping population of tetraploid potato (Meyer *et al.*, 1998). Such problems would be expected to be uncommon in crosses based on inbred lines, but cluster analysis has indicated individuals with identical genotypes in a doubled haploid population of spring barley, probably as a consequence of a problem in tissue culture. Situations of mixed parentage will be revealed easily by codominant markers such as microsatellites, but these are often scored at a later stage in the mapping analysis than dominant markers such as AFLPs or RAPDs. The multivariate methods used here are readily available in most standard statistical packages, and as a consequence of this study we strongly recommend them as a screening tool at an early stage in any linkage mapping programme, especially where noninbred parents are involved.

Acknowledgement

References

BINELLI, G. AND 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.

DIGBY, P. G. N. AND KEMPTON, R. A. 1987. *Multivariate Analysis of Ecological Communities*. Chapman & Hall, London.

ECHT, C. S., KIDWELL, K. K., KNAPP, S. J., OSBORN, T. C. AND McCOY, T. J. 1994. Linkage mapping in diploid alfalfa (*Medicago sativa*). *Genome*, **37**, 61–71.

FREYRE, R. AND DOUCHES, D. S. 1994. Isoenzymatic identification of quantitative traits in crosses between heterozygous parents: mapping tuber traits in diploid potato (*Solanum* spp.). *Theor. Appl. Genet.*, **87**, 764–772.

GAWEL, N. J. AND JARRET, R. L. 1991. A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Mol. Biol. Rep.*, **9**, 262–266.

GEBHARDT, C., RITTER, E., DEBENER, T., SCHACHTSCHABEL, U., WALKEMEIER, B., UHRIG, H. ET AL. 1989. RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor. Appl. Genet.*, **78**, 65–75.

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

HEMMAT, M., WEEDEN, N. F., MANGANARIS, A. G. AND LAWSON, D. M. 1994. Molecular marker linkage map for apple. *J. Hered.*, **85**, 4–11.

MAGAMBO, M. J. S. AND CANNELL, M. G. R. 1981. Dry matter production and partition in relation to yield of tea. *Exp. Agric.*, **17**, 33–38.

MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.*, **27**, 209–220.

MEYER, R. C., MILBOURNE, D., HACKETT, C. A., BRADSHAW, J. E., MCNICOL, J. W. AND WAUGH, R. 1998. Linkage analysis in tetraploid potato and associations of markers with quantitative resistance to late blight (*Phytophthora infestans*). *Mol. Gen. Genet.*, **259**, 150–160.

OROZCO-CASTILLO, C., CHALMERS, K. J., WAUGH, R. AND POWELL, W. 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theor. Appl. Genet.*, **87**, 934–940.

PAUL, S., WACHIRA, F. N., POWELL, W. AND WAUGH, R. 1997. Diversity and genetic differentiation among populations of

Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. *Theor. Appl. Genet.*, **94**, 255–263.

STAM, P. AND VAN OOIJEN, J. W. 1995. *JoinMap*, Version 2.0: *Software for the Calculation of Genetic Linkage Maps.* CPRO-DLO, Wageningen.

VAN OOIJEN, J. W. 1994. Drawmap: a computer programme for drawing genetic linkage maps. *J. Hered.*, **85**, 66.

WACHIRA, F. N., WAUGH, R., HACKETT, C. A. AND POWELL, W. 1995. Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. *Genome*, **38**, 201–210.