Theor Appl Genet (2002) 106:157–162
DOI 10.1007/s00122-002-1067-1

V. Le Clerc · M. Briard · P. Revollon

Influence of number and map distribution of AFLP markers on similarity estimates in carrot

Received: 14 July 2001 / Accepted: 2 April 2002 / Published online: 1 October 2002

Abstract When genetic diversity among organisms was measured with molecular markers, the question of genome coverage was currently stressed out. In order to check if well-distributed, mapped AFLP markers were more efficient in assessing varietal identification of carrot accessions than randomly chosen markers, nine closely related genotypes were analysed. A software was developed to realise 1,000 random choices of 20 to 70 mapped or unmapped markers, offering numerous genome coverages. We statistically showed that taking into account marker position does not provide a better estimation of genetic distances. Moreover, in the case of carrot, we concluded that 60 AFLP markers offer the best compromise between the level of precision and minimal expense.

Keywords AFLP · Mapped or unmapped markers · Carrot · Varietal identification

Introduction

Molecular markers have been employed for numerous studies such as phylogeny analysis, genetic diversity studies, mapping and varietal identification. However, these analyses are often both time consuming and expensive. Hence, the development of alternative strategies, allowing high precision with minimal cost and time requirements, is explored. For carrot, the model plant of the present study, RAPDs (Grzebelus et al. 1997; Briard et al. 2000b), microsatellites (Niemann et al. 1997; Briard et al. 2000b), ISSRs (Briard et al. 2000b) and AFLPs (Le Clerc et al. 2000; Shim and Jorgensen 2000) have been successfully used for carrot accession identification. In order to analyse a large number of accessions, we proposed to develop analytical procedures giving a maximum amount of information with minimal expense. We focused on the AFLP technique which allows the detection of a high number of markers with few experiments and is known to provide widely distributed markers on the genome. In a previous study (Le Clerc et al. 2000), the efficiency of using a reduced number of mapped markers was compared with the efficiency of using more markers obtained at random. From the results obtained with different sets of markers, offering better or worse distribution of markers across the genome, we showed that mapped markers were not more informative than unmapped markers; their efficiency being dependent only on the number of markers investigated. These results disagreed with Lombard (2000) who showed that a more precise estimation of genetic distances between rapeseed varieties was obtained by taking into account the position of the AFLP markers on the linkage map. In contrast, we agreed with Virk et al. (2000) whose results on 56 rice accessions showed that the marker distribution on the genome had no significant impact on diversity measurement. Futher studies are required to favour either conclusion. Therefore, in order to validate our previous study, we first analysed another set of accessions, more closely related, i.e. more difficult to identify, than in our previous study. Our results, being based on the analysis of only three subsets of markers offering different genome coverages, we then compared a very high number of subsets with both mapped and unmapped markers. Different sizes of data sets were also compared.

Communicated by H.C. Becker

V. Le Clerc (✉) · M. Briard
INH (Institut National d'Horticulture),
GBH (Génie Biologique appliqué à l'Horticulture),
2, Rue Le Nôtre, 49045 Angers cedex 01, France
e-mail: valerie.leclerc@inh.fr
Tel.: +33-241-225493, Fax: +33-241-225515

P. Revollon
INH (Institut National d'Horticulture),
UMR SAGAH (Sciences Agronomiques Appliquées à l'Horticulture),
2, Rue Le Nôtre, 49045 Angers cedex 01, France

Materials and methods

Plant material

Nine accessions were analysed: one three way hybrid 'ACD'; one F1 hybrid, 'AC'; and three inbred lines 'A', 'Cf' and 'D'. As shown in Fig. 1, these accessions are very closely related. Four other inbred lines were included in the study. The accessions '1' and '3' were selected from the same initial population whereas accessions '4' and '5' had different origins. All of the nine accessions were 'Nantes' type.

For each accession, four or five individuals were analysed separately.

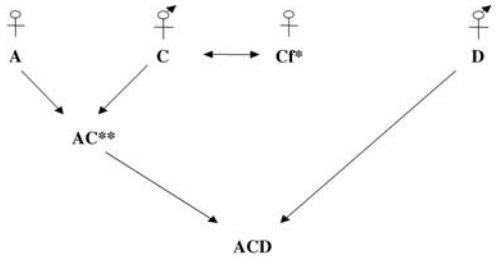

Fig. 1 Genealogy of the 3-way hybrid ACD. * Cf Male sterile version of C, ** AC Female parent of ACD 3 ways hybrid

DNA extraction and AFLP protocol

DNA extraction from freeze-dried leaves and root slices, and the AFLP protocol, were as previously described by Briard et al. (2000a). For selective amplification, 15 primer combinations (*Eco*RI-primer and *Mse*I-primer) were used. After electrophoresis, silver staining was performed for a maximum of 2 h. Acrylamide gels were dried overnight at room temperature. Only non-ambiguous and polymorphic bands were scored as present (1) or absent (0).

Marker sampling

Ninety five markers were added (83 AFLPs, 8 ISSRs and 4 Resistance Gene Analog markers) to the previous map; initially constructed from 188 individuals with 78 markers (Le Clerc et al. 2000). Mapping data was computer-analysed with the 'Mapmaker' program, version 3.0 (Lander et al. 1987). Together with 147 unmapped markers, they resulted in a reference set of 249 scored markers. Some software was developed with Excel in order to generate 1,000 random subsets of 20, 30, 40, 50, 60 and 70 markers from 102 mapped and from 102 unmapped markers. It was also used to calculate IAS values (see below) and the similarity matrices. For each of them, a Sokal and Michener similarity matrix was calculated (Sokal and Michener 1958). In addition to their double presence, this index also takes into account the double absence of a marker between two individuals as recommended for intraspecific analyses (Briard et al. 2000c).

Intravarietal similarity values (IAS)

For each of the 12,000 matrices, the intravarietal similarity values (IAS) of each accession (mean distance between all pairs of one accession) were also calculated. The mean IAS (M_{IAS}) and standard deviation (SD) for 1,000 items were then calculated for each subset size, from 20 markers (M_{IAS20}) to 70 markers (M_{IAS70}). This was carried out with the mapped markers (e.g. $M_{IAS20\mathbf{m}}$) and the unmapped markers (e.g. $M_{IAS20\mathbf{nm}}$).

Dendrograms

UPGMA dendrograms were constructed using Phylip software (Felsenstein 1989). For the reference tree (R) obtained from the 249 markers, the significance of branching nodes was tested by bootstrap re-sampling (100 samples) using the Seqboot program of the Phylip software. The similarity matrix was calculated using an Excel command carried out by Divaret (1999) and analysed using the Phylip's 'Consense' and 'Drawtree' programs.

Ten dendrograms with 70 markers and ten others with 60 markers were designed for both sets of mapped or unmapped markers.

D_f was calculated as the difference between IAS value obtained for the 'ACD' genotype with a set of markers and its reference IAS_{ACD249} calculated from a total of 249 markers. Five other dendrograms were also designed from five sets of 60 mapped markers resulting in different IAS_{ACD} values such as D_f was: $3.20 < D_f < 3.45$; $3.45 < D_f < 3.70$; $3.70 < D_f < 3.95$; $3.95 < D_f < 4.20$ and $D_f > 4.20$. For each of those five dendrograms, the existence of the groupings considered as being solid from bootstrap values (bootstrap >80%) was checked. The six characteristics checked were: (1) all the genotypes were well identified, (2) the genotypes 'Cf', AC', 'A' and '5' were grouped together, (3) individual '1.3' clustered with '1.4', (4) 'D' clustered with 'ACD', (5) individual '3.1' clustered with '3.3' and '3.4', and (6) the genotype '4' was kept apart.

Occurrence of low IAS_{ACD} values

A total of 5,000 samplings of 70 markers was run from the 102 mapped markers. The corresponding 5,000 IAS_{ACD} values were calculated and put into order. The number of occurrences with $3.20 < D_f < 3.45$; $3.45 < D_f < 3.70$; $3.70 < D_f < 3.95$; $3.95 < D_f < 4.20$ and $D_f > 4.20$ was counted. The same was applied to 5,000 samplings of 60 and 50 markers.

Results

By adding 95 markers, we obtained a more-saturated map. Even if this map was not completely saturated, offering a higher number of linkage groups than expected, all the linkage groups were covered. One hundred and seventy three markers were mapped on 15 linkage groups of 5 to 17 markers, two groups of 4 markers, two groups of 3 markers and 4 pairs of loci. Only 102 of these mapped markers were used in the present analysis. It was not always possible to score the other markers and most of them were monomorphic within the present set of genotypes.

On the reference dendrogram (R) made with the reference set of 249 markers (Fig. 2), all genotypes were well identified, with one cluster corresponding to one genotype (bootstrap = 98% or 100%). The genotypes 'A', 'Cf', 'AC' and '5' formed one large group (bootstrap = 96%), whereas genotype 'D' clustered with genotype 'ACD' (bootstrap = 100%). The genotype '4' was separate (bootstrap = 84%).

Whatever the genotype, the value of intravarietal similarity obtained with 249 markers (IAS_{249}) was high (Table 1). It varied from 86.7% for the three-way hybrid 'ACD' to 98.78% for the inbred line 'A'. Similar means of IAS values (M_{IAS}) were obtained not only with the

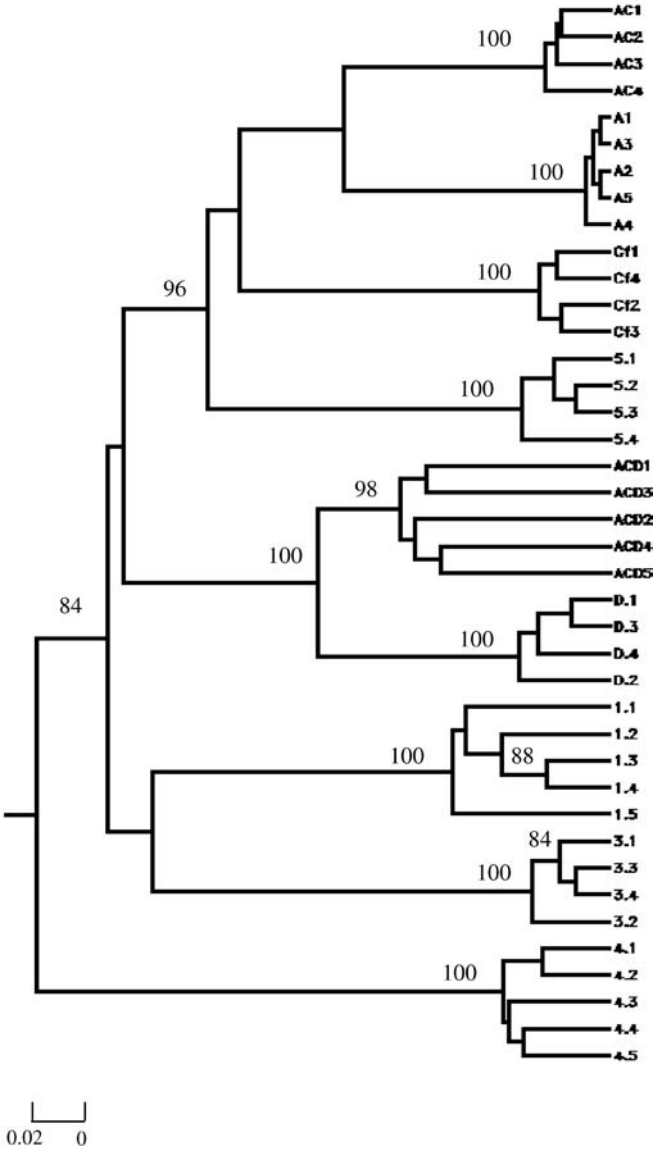

Fig. 2 Dendrogram (R) of nine carrot accessions based on UPGMA analysis of Sokal and Michener's similarity matrix calculated from 249 AFLP markers. The values on the dendrogram gives the stability of nodes estimated with a bootstrap procedure (only values >80%)

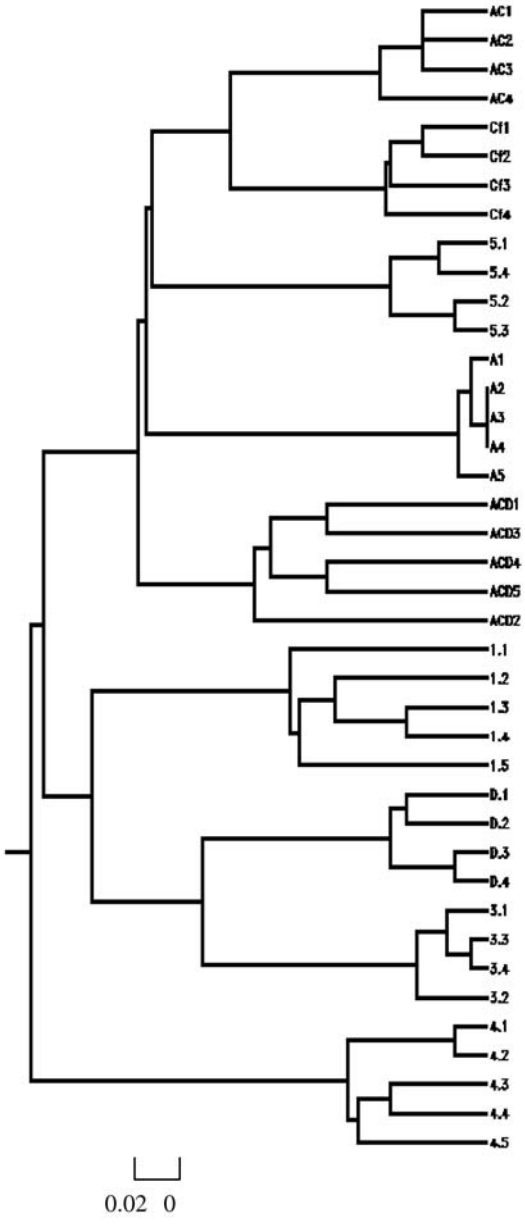

Fig. 3 Dendrogram obtained from 60 AFLP markers with $D_f > 4.20$ (D_f: IAS_{ACD249}-IAS_{ACD60})

1,000 subsets of 70 markers but also with the 1,000 subsets of 20 markers. The standard deviations (SDs) of 20 marker subsets were 3-times greater than the SDs calculated with 70 marker subsets. SD levels were very low within the 1,000 subsets of 70 mapped or unmapped markers. Whatever the number of markers analysed, the lowest minimum IAS value (Min IAS) was obtained for the genotype 'ACD' whereas the highest was obtained for genotype 'A'. For unmapped markers, the means of IAS values (M_{IAS}) were always greater than the M_{IAS} calculated with mapped markers and greater than IAS_{249}.

A better precision of IAS values was reached with a higher number of markers, meaning that more markers are analysed and the more reliable the measurement is. This raises the question of optimal number. Are 70, 60 or 50 markers enough? The fact that all the 20 dendrograms obtained with 70 markers were in accordance with the reference dendrogram R suggests that 70 are enough. Sometimes, some minor modifications might be observed, only for nodes with bootstrap values inferior to 80%. With 60 markers, most of the 20 dendrograms were also similar to the reference dendrogram. As expected from its IAS_{ACD} values (lowest IAS values), and genetic data (three-way hybrids compared to inbred lines or F1 hybrids), the mistmatchings concerned the genotype 'ACD'. Therefore, we focused our investigation on

Table 1 Mean (M IAS), standard deviation (SD IAS) and minimum values (Min IAS) of IAS (Intravarietal Similarity) obtained with 1,000 samplings of 20 to 70 mapped (m) or unmapped (nm) markers for the nine accessions 'A', 'Cf', 'AC', 'ACD', '1', 'D', '3', '4' and '5'

Item	No. of markers	A	Cf	AC	ACD	1	D	3	4	5
IAS_{249}	249	98.78	96.03	96.59	86.70	91.07	95.28	96.18	93.38	95.73
M IAS%	20nm	99.39	96.86	98.70	87.59	91.79	95.64	97.01	94.93	96.51
	20m	98.75	94.18	95.00	87.55	89.09	95.30	95.85	91.73	95.05
	30nm	99.37	96.97	98.80	87.62	91.58	95.74	97.06	95.04	96.41
	30m	98.72	94.12	94.81	87.44	88.89	95.30	95.79	91.75	95.02
	40nm	99.38	96.99	98.82	87.48	91.73	95.71	97.09	94.97	96.50
	40m	98.77	94.05	94.91	87.41	89.06	95.40	95.94	91.87	94.99
	50nm	99.34	96.93	98.77	87.63	91.70	95.75	97.11	94.92	96.41
	50m	98.77	94.23	94.85	87.40	88.96	95.38	95.83	91.89	95.05
	60nm	99.39	96.98	98.80	87.60	91.67	95.71	97.07	95.03	96.43
	60m	98.75	94.11	94.82	87.41	89.02	95.36	95.81	91.83	94.94
	70nm	99.38	96.90	98.77	87.60	91.75	95.72	97.06	94.97	96.45
	70m	98.73	94.07	94.84	87.47	89.04	95.35	95.80	91.75	94.92
SD IAS%	20nm	0.88	2.41	1.30	3.71	3.18	2.51	2.07	2.78	2.27
	20m	1.46	2.79	2.96	3.95	3.74	2.71	2.42	3.47	2.84
	30nm	0.69	1.76	0.98	2.93	2.49	1.85	1.52	2.05	1.76
	30m	1.12	2.06	2.26	2.98	2.86	2.15	1.88	2.69	2.02
	40nm	0.55	1.39	0.81	2.31	2.04	1.51	1.27	1.62	1.41
	40m	0.92	1.83	1.82	2.43	2.28	1.63	1.49	1.99	1.71
	50nm	0.46	1.14	0.66	1.88	1.61	1.26	1.03	1.35	1.16
	50m	0.77	1.46	1.46	1.93	1.92	1.35	1.22	1.81	1.37
	60nm	0.38	0.97	0.55	1.60	1.29	1.05	0.85	1.14	0.97
	60m	0.61	1.17	1.23	1.64	1.53	1.13	1.03	1.45	1.16
	70nm	0.30	0.81	0.43	1.30	1.05	0.83	0.68	0.90	0.78
	70m	0.49	0.94	1.00	1.31	1.25	0.93	0.83	1.13	0.96
Min IAS%	20nm	96.80	88.13	93.75	75.20	81.60	85.63	90.63	82.40	88.13
	20m	93.60	84.38	85.00	76.00	76.80	86.88	88.13	78.40	85.00
	30nm	97.87	90.83	95.83	78.13	83.47	89.17	91.25	88.27	90.83
	30m	95.73	87.08	84.17	76.00	79.20	88.75	89.58	81.87	88.75
	40nm	98.40	92.19	96.88	78.40	86.00	90.94	93.44	90.00	91.88
	40m	96.80	88.75	89.38	77.20	82.40	90.31	91.25	86.00	89.38
	50nm	98.72	93.75	97.50	80.80	86.88	92.00	94.00	90.72	92.75
	50m	97.44	89.50	90.00	79.52	83.04	91.25	91.50	85.60	89.75
	60nm	98.93	94.79	97.92	82.93	87.47	92.71	95.00	92.00	93.96
	60m	97.87	90.63	91.67	82.40	84.00	92.08	92.92	86.67	91.38
	70nm	99.09	95.54	98.21	83.09	89.26	93.75	95.71	92.69	94.82
	70m	98.17	91.43	92.50	83.09	85.14	93.21	93.93	88.57	92.68

Table 2 Number of IAS_{ACD} values contained in different classes of D_f. D_f is the difference between IAS_{ACD} calculated with 50, 60 or 70 markers and IAS_{ACD249} calculated with 249 markers

D_f (%)	3.20–3.44	3.45–3.69	3.70–3.94	3.95–4.20	>4.20
50 markers	31	21	0	11	24
60 markers	12	7	5	5	2
70 markers	1	2	0	0	0

IAS_{ACD} values in the different subsets to determine the threshold IAS value for a reliable identification. Indeed, carrot cultivar identification seems easier when genotypes have high IAS values (Briard et al. 2000b) and, if reliable identification is obtained for a genotype with a low IAS, it should be true for other genotypes with higher IAS values.

In the situation of five data sets, resulting in a range of D_f (the difference between IAS_{ACD} and its reference IAS_{ACD249}), dendrograms were not always in accord with the reference dendrogram R. When D_f was lower than 3.70, dendrograms were in complete accord with the reference dendrogram. When D_f was higher than 3.70 but lower than 3.95, only the 'ACD' genotype became poorly identified. For higher D_f values, the dendrograms were quite different from the reference dendrogram (e.g. Fig. 3, with 'ACD' separated from 'D').

Through 5,000 draws, the number of D_f values higher than 3.70% is 0 with 70 markers, 12 with 60 markers and 35 with 50 markers (Table 2). A D_f difference higher than 4.20% occurred twice and 24-times with 60 and 50 markers respectively. We never got a $D_f > 4.2\%$ with 70 markers.

Discussion

The breeders and curators proper genetic resource management needs a reliable genotype identification. Precise characterisation of all the accessions is required for both evaluation of the genetic variability of the collection and detection of the duplicates. For any molecular identification technique, precision can be improved by increasing the number of analysed markers. For genetic distance estimations, various studies have related the number of markers to the level of precision (Tivang et al. 1994; Barbosa-Neto et al. 1997; Guérardi et al. 1998; Divaret 1999). No optimal number suitable for all the species can be proposed. This number has to be determined on a species basis. Numerous authors considered that this number may also be related to genome coverage. It has been claimed that genetic distance precision may increase by working with well-distributed markers on the genome. However, very few studies were carried out on the topic (Tivang et al. 1994; Karp et al. 1996; Laurie et al. 1997).

In a previous paper, we suggested that well-distributed mapped markers were no more efficient than randomly chosen ones (Le Clerc et al. 2000) in varietal identification of carrot genotypes. This analysis was performed with very few marker subsets. In the present study, using a new more-saturated map and excell software, we analysed 12.000 genome coverages, with markers non-equally distributed on the linkage groups and, therefore, on the genome. We stated that, in this situation, if bad distributions occur, for example the existence of fully uncovered linkage groups, we should obtain them randomly.

The similarity between all the IAS mean values (from M_{IAS20} to M_{IAS70}) with their IAS_{249} reference can be explained by the high number of repetitions (1,000). To assess the mapping efficiency, the standard deviation has to be considered. Whatever the subset, the similarity between mapped and unmapped marker confirm that markers distribution does not affect precision. Moreover, the complete concordance of the reference dendrogram with the 20 dendrograms constructed with either 70 mapped or unmapped markers validates this conclusion. Therefore, we confirm that in the case of the carrot genotype analysis, mapped markers do not provide a more-accurate estimation of genetic distances. Though, once more, our results are respectively in accord with the results obtained in rice by Virk et al. (2000) and in opposition with those in rapeseed (Lombard 2000). Even if some uncovered groups exist (as observed in our previous study) or if some chromosomal fragments (Virk et al. 2000) are missed, reliable identification is obtained.

Moreover, the efficiency of unmapped markers for varietal identification of carrot cultivars is not only similar to the efficiency of mapped markers but maybe even higher (M_{IASm} always smaller to M_{IASnm}). The IAS_{249}, being calculated from mapped and unmapped markers, is obviously between both values. Virk et al. (2000) also found that, depending on the degree of genetic relationship between the parents of the initial cross used to generate the map, using mapped markers could also result in a misleading pattern of diversity. However, in their data sets another factor may interact: i.e. the number of polymorphic markers analysed, which logically decrease with the distance separating the parents (data not given). While this does not in any way invalidate their conclusion, it would be interesting to clearly identify the role of each factor (distance between the parents and the number of markers) by using the same number of markers. In the present study, the number of mapped and unmapped markers is equal, and the distance between the parents is relatively high. However, the two parents were even more distant from the nine genotypes analysed as, respectively, of the Imperator type from the American continent and of the Nantes type from France. This could result in a slightly less-adapted set of markers. It would be interesting to compare the present results with the efficiency of the same mapped markers for assessing the diversity of other American imperator genotypes.

We previously showed that the quality of the markers (mapped or unmapped) is not crucial, as opposed to their number. Indeed, whether 70 markers or 60 markers were used, Df was lower than 3.7% or occasionally higher than 4.2%, respectively. Therefore, the threshold for mismatching of the dendrograms lies between the two values. To determine this threshold, we studied five dendrograms with different IAS_{ACD} values in order to obtain different D_f values in this interval. To do so, we used mapped markers, since, as previously shown, they provide a higher D_f. When D_f was lower than 3.7%, the results were fully reliable. When D_f was higher than 3.95%, mismatchings were too important to consider the results as accurate. Therefore, these two values can be considered as two important thresholds. Only an applicant will decide which one is the more adapted to his current study, as explained below. Our last consideration concerns the chance of obtaining a D_f superior to the above-mentioned threshold when working with 5,000 draws?

As shown in Table 2, the number of D_f values higher than 3.7% was inversely proportional to the number of markers involved. With 60 markers, only 12 D_f values were higher than 3.7%, meaning that a 0.24% risk of not obtaining a dendrogram fully similar to the reference dendrogram exists. More precisely, we showed that for 4,988 dendrograms made from 5,000 draws, complete accord with the reference dendrogram will be achieved; for ten draws, we will get a dendrogram with minor variations and for two draws (0.04% risk), we will obtain a dendrogram with more obvious differences (e.g. Fig. 3). If we consider 50 markers, we take a 0.7% risk of being superior to 3.7%. According to these results, in most of the studies a total of 60 markers should be ample to reach a high level of precision at minimal expense. However, the number of markers to be analysed will depend on the level of precision required. In order to assess the extent and level of genetic diversity in a collection, we would recommend the use of 60 markers for a first anal-

ysis. Then, for further analysis, the number of markers should be increased if some genotypes are not well identified or if greater precision is required, such as the identification of duplicates. If total security is needed, as for essential derivation evaluation, 70 markers will be used.

In conclusion, AFLP was highly efficient in assessing the identification of closely related genotypes. With an accurate estimation of intravarietal similarity values, we should be able to evaluate genetic distances and relationships between genotypes. Subsets of mapped or unmapped markers led to similar conclusions. In other words, there is no advantage in using mapped markers which require time-consuming and expensive preliminary work when no linkage map is available. Therefore our future analyses will be conducted with unmapped markers. A quantity of 60 markers is regarded as the optimal number for carrot-genotype identification, and from this optimal number we will develop a procedure to determine for each genotype the optimal number of individuals to be analysed, depending on its genetic homogeneity.

Acknowledgements We are grateful to Dr. P.W. Simon and L. S. Boiteux for providing the F2 population, and to 'Vilmorin Seeds' for providing accessions. This research was funded by the town of Angers. I declare that these experiments comply with the current laws of France.

References

Barbosa-Neto JF, Hernandez CM, O'Donoughue LS, Sorrells ME (1997) Precision of genetic relationship estimates based on molecular markers. Euphytica 98:59–67

Briard M, Le Clerc V, Grzebelus D, Senalik D, Simon PW (2000a) Modified protocols for rapid carrot genomic DNA extraction and AFLP using silver stain or radioisotopes. Plant Mol Biol Rep 18:235–241

Briard M, Le Clerc V, Mausset AE, Veret A (2000b) A comparative study on the use of ISSR, microsatellites and RAPD markers for varietal identification of carrot genotypes. Acta Hort 546:377–385

Briard M, Le Clerc V, Peltier D, Veret A, Péron JY (2000c) Molecular analysis of the genetic variability within the *Apiaceae* family. Acta Hort 521:155–163

Divaret I (1999) Elaboration des bases de la gestion des ressources génétiques d'une collection de choux cultivés (*Brassica oleracea* L.). Rennes, INRA-Laboratoire des crucifères légumières-Station d'amélioration des plantes. PhD thesis

Felsenstein J (1989) Phylip Phylogeny Inference Package. Cladistics 5:164–166

Grzebelus D, Szklarczyk M, Michalik B (1997) The use of RAPD markers for genotype identification of carrot lines and F1 hybrids. J Appl Genet 38A:33–41

Guérardi M, Mangin B, Goffinet B, Bonnet D, Huguet T (1998) A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD molecular markers. Theor Appl Genet 96:406–412

Karp A, Seberg O, Buiatti M (1996) Molecular techniques in the assessment of botanical diversity. 78:143–149

Lander ES, Green P, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Laurie DA, Bryan GJ, Snape JW (1997) Genomic relationships, conserved synteny and wide-hybrids. In: Callow JA, Ford-Lloyd BV, Newbury HJ (eds) Biotechnology and plant genetic resources- conservation and use, pp 77–101

Le Clerc V, Briard M, Peltier D (2000) Carrot genetic substructure evaluation: comparison of the efficiency of mapped molecular markers and randomly choosen markers. Acta Hort 546:127–134

Lombard V (2000) Estimation de la proximité génétique de variétés de colza sur la base de marqueurs moléculaires: conséquences pour l'inscription et la protection variétale. Paris, Institut national agronomique Paris-Grignon. PhD thesis

Niemann M, Westphal L, Wricke G (1997) Analysis of microsatellite markers in carrot (*Daucus carota* L. *sativus*). J Appl Genet 38A:20–27

Shim SI, Jorgensen RB (2000) Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. Theor Appl Genet 101:227–233

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using bootstrap procedure. Theor Appl Genet 89:259–264

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (2000) Are mapped markers more useful for assessing genetic diversity? Theor Appl Genet 100:607–613