Molecular characterisation of a lettuce germplasm collection

Theo van Hintum
Centre for Genetic Resources, The Netherlands (CGN), DLO Foundation
PO Box 16, 6700 AA Wageningen, The Netherlands.
Contact: theo.vanhintum@wur.nl, (+31) 317 477078.

Abstract: In the framework of a large EU funded project, AFLPs and microsatellites were used to characterise the entire lettuce collection of the Centre for Genetic Resources, The Netherlands (CGN). Some highlights of the results of the analysis of the AFLP data are reported. 8020 Plants of 2323 accessions of lettuce and related wild species (*Lactuca* spp. plus related genera) were characterised using three AFLP primer combinations. It was clear from the beginning that no suitable software for the analysis was available, a spin-off of the project is another EU funded project for the creation of such software. The genetic diversity within accessions was higher than expected, the comparison between the groups however fully complied with the expectations. The number of cases that plants from one *L. sativa* accession differed exactly one band was one fifth of the cases where the plants were identical, implying some residual variation or possibly methodological 'noise'. It proved difficult to identify 'identical plants' due to this residual variance, but also due to missing values. However an estimated lower bound showed 20% 'redundancy' based only on homogeneous identical accessions implying that the fingerprints used were not able to distinguish all genotypes. The AFLP fingerprints proved very suitable to describe the wild species collections and its individual accessions. Based on the genetic differentiation between accessions, the self-, or cross-pollinating nature of accessions could be revealed, the genetic structure of the species could be analysed, the taxonomic classification and origin data of the accessions could be validated, and redundancies could be identified. The fingerprint also allowed improvement of the balance in the representation of the species and better selection of material for use. Finally, the association amongst markers and between markers and traits was shown to be highly significant, and can be explained by genetic linkages, but also by effects of geographical isolation, founder effects, etc.

Keywords: *Lactuca,* molecular markers, genetic resources, genebank

Introduction

Genebanks conserve *ex situ* genetic diversity of crops and their wild relatives, which would otherwise run the risk of disappearing. Crop genetic diversity is conserved by sampling the diversity in the genepool and storing the samples, the so-called genebank accessions. Both the sampling and the storing needs to be done as efficiently as possible to avoid accidental loss of diversity, and optimally use of the limited resources. This raises questions concerning which samples to include in a germplasm collection, and how to determine possible omissions or redundancies, but also how to multiply samples without loss of diversity due to contamination, unintentional or natural selection or genetic drift.

Apart from conserving, another objective of germplasm management is making the collections accessible to the end-users. Given the number of collections worldwide and the sizes of these collections it is difficult to choose material for utilisation by the end-user; can small subsets be selected which contain the main part of the genetic diversity in the collection (so called core collections, Hintum et al. 2000) ?

To study how molecular markers could play a role in answering these questions, a large EU funded project was carried out in which AFLPs[1] and microsatellites were used to characterise the entire lettuce collection (*Lactuca* spp. plus related genera) of the Centre for Genetic Resources, The Netherlands (CGN). The project ran from October 1997 to November

[1] The AFLP® technology is covered by patents and patent applications owned by Keygene N.V. (Wageningen, The Netherlands)

2000 and involved apart from CGN, five European partners: the International Plant Genetic Resources Institute in Italy, the Botanisk Institut in Denmark, IARC - Long Ashton Research Station in the UK and Keygene N.V. and Perkin-Elmer BV both in the Netherlands.

Some highlights of the results of the AFLP screening will be presented here.

Material and methods

DNA material

The germplasm analysed consisted of all 2323 accessions in the CGN lettuce collection at the moment of tissue sampling, autumn 1997. It was mainly *Lactuca sativa* (64%) but also wild accessions including *L. serriola* (27%), other *Lactuca* species (9% of 17 species) and even related genera (1% of 4 species). The cultivated accessions were sampled with two plants per accession, the wild accessions with five plants. Additional to this several accessions were sampled with 30 plants to determine the diversity within these accessions. The total number of plants analysed was 8020.

The accessions were grown out in the greenhouse, one accession per pot. When the plants had two mature leaves, one and a half square centimetre was sampled from the third immature leave. Most accessions reached this stage two weeks after sowing, but for some accessions this could take up to seven weeks. DNA was extracted from the sampled leave tissue. For the DNA extraction, a protocol based on Beek et al. (1992) was used. Volumes were adapted for high throughput DNA extraction.

AFLP protocol

Based on the number of AFLP markers per lane and the distribution of the AFLP markers in a lane that came out of an AFLP pre-screening using different DNA pools, *Eco* RI/*Mse* I +3/+4 AFLP primer combinations were chosen to characterize the collection with. The first combination (*Eco* RI+ACA / *Mse* I+CTAT) was used for all species, and based on these results, primer combinations for further screening were specifically selected per species. The AFLP protocol followed the procedures described by Vos et al. (1995).

These +3/+4 combinations resulted in a relatively low density of bands. For example, in the case of *L. sativa* the three primer combinations yielded 167 polymorphic bands and an average of 69.54 bands per plant.

Genetic diversity and differentiation

Gene diversity (H) is used as a measure of the genetic diversity in accessions, sets of accessions or the collection. If p is the frequency of a band the diversity for that marker $H=1-(p^2+(1-p)^2)=2p(1-p)$. To correct for small sample sizes, an unbiased estimate of H is obtained by multiplying the value by the factor $n/(n-1)$ where n is the number of individuals measured (Nei 1987).

To estimate the average H within accessions of a species or another group of accessions, the definition of H (being the chance that two plants are unequal for a random locus) was used; it was calculated as the average frequency of bands being different between pairs of plants of an accession.

G_{ST} was used as a measure of genetic differentiation between accessions (see Nei 1975 for a discussion). It is the proportion of the total genetic diversity that is not found within the accessions: $(H_T - H_S)/H_T$ in which H_T is the total genetic diversity in a set of accessions and H_S is the average diversity within these accessions.

Cluster analysis
For the clustering an agglomerative approach was used in which the genetic distance between clusters (an individual is a cluster of size one) was calculated using Rogers distance (Rogers 1972, Wright 1978). The clustering algorithm used the cluster size as weight: the absolute frequencies of bands were summed over fused clusters, and all distances of the new cluster with all others were calculated.

Software
All data analyses were performed using custom made software written in MS Visual Basic.

Results and discussion

The first result of this large scale screening of germplasm was the finding that no good software is available for the analysis of this type of data. The currently used software for diversity analysis is not only unable to handle datasets of this size, but more importantly lacks some vital concepts. For example, a question such as 'what is a genebank duplicate in terms of molecular diversity within and between accession' has not been raised before. As a result, the major outcome of the project is a new project called 'Improved use of germplasm collections with the aid of novel methodologies for integration, analysis and presentation of genetic data sets'. This project has received EU funding and is currently running.

Genetic diversity within and between accessions
The genetic diversity within groups of accessions such as 'Butterhead lettuce' was estimated with the 'definition based' method as described above. The resulting estimates (see table 1) were higher than expected. The comparison between the groups however fully complied with the expectations. The statistical properties of these estimates have not yet been determined.

Table 1. Average genetic diversity (H) within accessions of a certain group. (Estimates of H are not comparable over species since different AFLP primer combinations were used.)

Group	H	Group	H
L. sativa	.0075	*L. serriola*	.0107
Butterhead	.0061	origin western Europe	.0054
Cos	.0094	origin eastern Europe	.0206
Crips	.0041	origin southern Europe	.0065
Cutting	.0136	origin middle east	.0111
Latin	.0065		
Stalk	.0163		

The diversity within accessions was obviously much lower as that between accessions, however, as can be seen from Table 1, there is frequently diversity within accessions. Figure 1 shows the distribution of those differences, based on 167 polymorphic bands. It can be seen that the number of cases that the plants differed exactly one band was more than one fifth of the cases where the plants were identical, implying some residual variation or possibly methodological 'noise'. Such noise was also observed in other studies involving potato (McGregor et al 2002) and barley (Treuren and Hintum 2001). If plants were compared over accessions a more or less normal distribution of differences could be observed with an average of 23.3 differences and a standard deviation of 7.8 (based on 3052959 comparisons of all plants without missing values, data not shown).

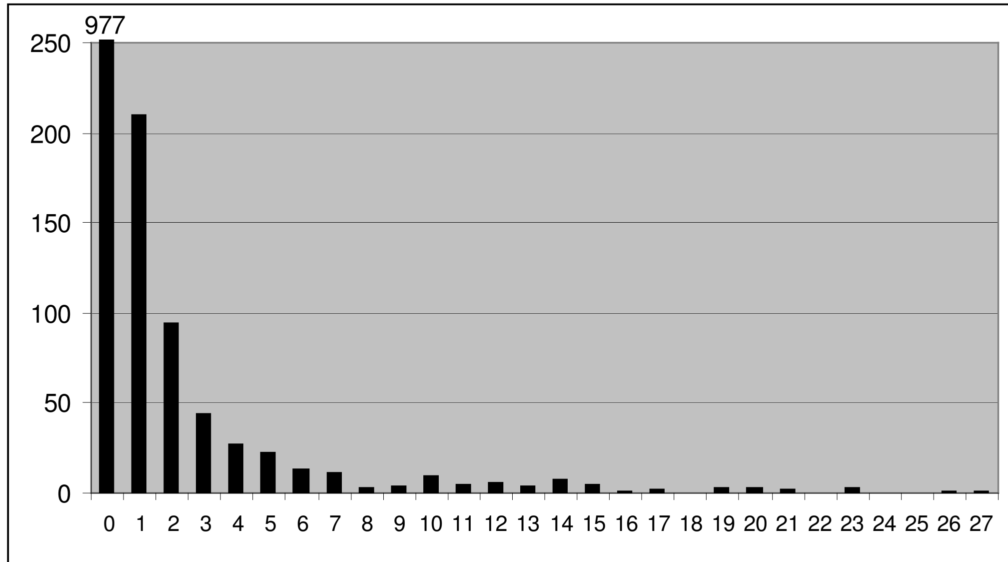

Figure 1. Distribution of absolute frequencies (on y-axis) of differences (on x-axis) in AFLP fingerprint of plants of the same cultivated lettuce accession (*L. sativa*).

Redundancy within the collection

Based on the AFLP fingerprints it should be easy to identify redundant identical accessions. However since more than one plant per accession was characterised and since small differences between the fingerprints of otherwise genetically identical plants occurred (see Figure 1), the analysis proved more complex. If small differences were accepted or as a result of missing data-points, it happened that one plant was identical to second plant, which in turn could be identical to a third, which however was not identical to the first. Parallel to this problems, the heterogeneity within accessions required a definition of duplication accounting for diversity within the accessions.

If only accessions of *L. sativa* characterised with two identical plants without missing values were considered, the number of accessions decreased from 1417 in the entire collection to 794. If these homogeneous accessions were compared, it could be shown that there were 633 different haplotypes, i.e., a 'redundancy' of 20%. These 'redundant' accessions often occurred in groups larger than two accessions (see Table 2). If all 1417 accessions would have been taken into consideration, allowing for one or two differences and taking account of missing values, this frequency could be expected to be much larger. These results imply that either the collection contains over 20% redundant accessions, which is not very probable, or the resolution of the three primer combinations used was not high enough to distinguish between the accessions. However the groups of identical accessions can be considered to be very similar which has implications for selection strategies. If the fingerprints were not able to fully distinguish the duplicates, they were able to identify non-duplicates.

Table 2. Frequency of groups of identical accessions in a group of 794 completely homogeneous accessions without missing values.

group size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	total
frequency	553	52	12	8	2	1	2	0	0	1	0	0	1	0	0	1	633
accessions	553	104	36	32	10	6	14	0	0	10	0	0	13	0	0	16	794

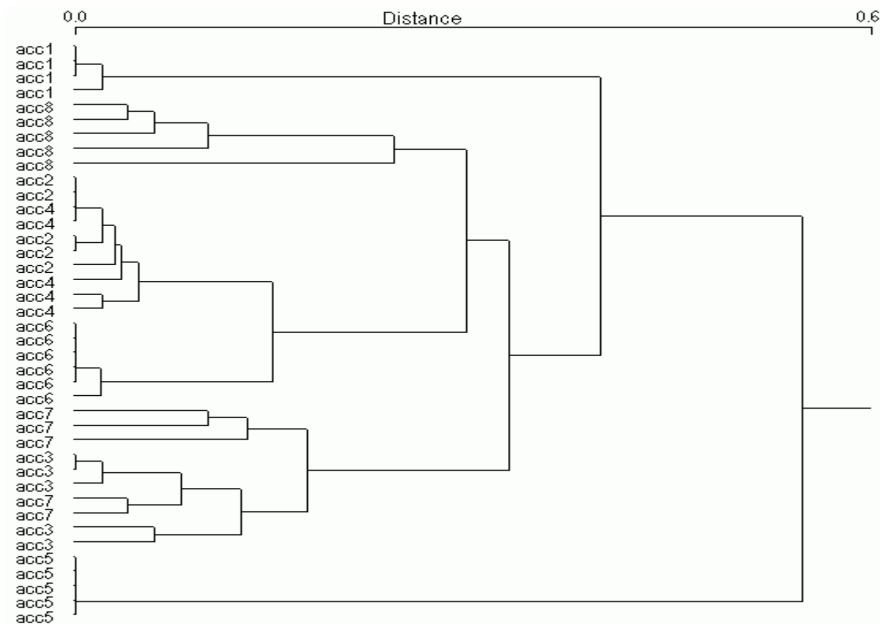

Figure 2. Genetic diversity in the 8 CGN *Mycelis muralis* accessions determined with 50 polymorphic AFLP bands. Accessions 2 and 4, and 3 and 7 are pairwise inter-mingled. Information about the origin of the accessions is not available, apart from the fact that they were received from four different botanical gardens.

Structure of wild species collections

The AFLP fingerprints provided excellent information on the genetic structure of the wild species collections. Based on the genetic differentiation between accessions, the self-, or cross-pollinating nature of accessions could be revealed, important information not always available for species in a genebank. For example, the G_{ST} of the known cross pollinator *L. perennis* was .37 and for the expected cross pollinator *L. tatarica* .51, low values as compared to the other species that varied between .82 and 1.00. Also the genetic structure of the species could be analysed. For example, the species *L. virosa* was found to consist of two distinct groups of accessions, a result also observed earlier on the basis of other marker systems (Koopman et al. 1998). It also proved simple to validate the taxonomic classification and the origin data of the accessions. For example, an accession assumed to have originated from Chile, appearing in a French cluster, pointed at an error in the data. Furthermore, it was possible to identify redundancies. For example, some accessions with identical fingerprints were received from different botanical gardens without proper accompanying documentation (see Figure 2). The botanical gardens probably had received it from each other or from a common source. The AFLP diversity also showed how the balance in the representation within the species in the collection could be improved, i.e., which groups were relatively over- and which were underrepresented. Obviously it could not show the balance between species since unfortunately different primer combinations had to be used for the different species. Finally the selection of material for use could be improved based on these data, by selecting material with maximally different AFLP fingerprints.

Association between markers and traits

Given the potential role of marker-assisted selection in commercial plant breeding, and the need for markers close to the target gene or QTL both for marker-assisted backcross (MABC) and marker-assisted recurrent selection (MARS) (Lande and Thompson, 1990; Melchinger,

1990) a preliminary search for associations between phenotypes and markers was made. The correlation between all pairs of AFLP bands and between the AFLP bands and the binarised phenotypic scores of a number of traits were calculated with a simple Chi-square with one degree of freedom. This resulted in many highly significant values, clearly showing association between the markers and the traits, even if the number of correlations was taken into account. The highest Chi-square value was 1219, between an AFLP band and seed color.

If the correlated bands were analysed further, clear blocks of markers associated closely together could be observed. A similar pattern of block appeares when the association between resistance to the different races of Bremia and the presence of AFLP bands was analysed.

The interpretation of these correlations and blocks of correlated AFLP bands or *Bremia* resistances can be a simple genetic linkage of the desired type, but also the result of geographical isolation, founder effects, etc. A precise interpretation will requires additional research.

Acknowledgements

This research was part of a project called 'Molecular markers for genebanks: Application of marker technology for the improvement of *ex situ* germplasm conservation methodology', which was supported by a research grant from the European Union in Framework 4 Biotech program (Bio4 CT 9 PL96.2062). The author would like to thank all partners involved in this project.

References

Beek, J.G. van der, Verkerk R., Zabel P. and Lindhout P. 1992. Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. Theor Appl Genet 84: 106-112.

Hintum, Th.J.L. van, Brown A.H.D., Spillane C. and Hodgkin T. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No.3. International Plant Genetic Resources Institute, Rome, Italy. 48p.

Koopman, W.J.M., Guetta E., Wiel C.C.M. van de, Vosman B. and Berg R.G. van den 1998. Phylogenetic relationships among *Lactuca* (Asteraceae) species and related genera based on ITS-1 DNA sequences. American Journal of Botany 85:1517-1530.

Lande, R. and Thompson R. 1990. Edfficiency of marker-selection in the improvement of quantitative traits. Genetics 124: 743-756.

McGregor, C.E., Treuren R. van, Hoekstra R. and Hintum Th.J.L. van 2002. Analysis of the wild potato germplasm of the series Acaulia with AFLPs: implications for ex situ conservation. Theoretical and Applied Genetics 104: 146-156.

Melchinger, A.E. 1990. Use of molecular markers in breeding for oligogenic disease resistance. Plant Breeding 104: 1-19.

Nei, M. 1975. Molecular Population Genetics and Evolution. American Elsevier, New York, 356p

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, USA

Rogers, J.S. 1972. Measures of genetic similarity and genetic distance. In: Studies in Genetics VII. University of Texas Publication 7213, Austin, USA

Treuren, R. van, and Hintum Th.J.L. van, 2001. Identification of intra-accession genetic diversity in selfing crops using AFLP markers: implications for collection management. Genetic Resources and Crop Evolution 48: 287-295.

Vos, P., Hogers R., Bleeker M., Reijans M., Lee T. van de, Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. and Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-4414.

Wright, S. 1978. Evolution and the Genetics of Populations, Vol. 4: Variability Within and Among Natural Populations. University of Chicago Press, Chicago, USA.