RESEARCH ARTICLE

Marker-assisted optimization of an expert-based strategy for the acquisition of modern lettuce varieties to improve a genebank collection

R. van Treuren · Th. J. L. van Hintum ·
C. C. M. van de Wiel

Abstract To regularly improve the composition of the lettuce collection of the Centre for Genetic Resources, the Netherlands (CGN) with modern varieties, feedback from crop experts is used to select approximately 10% of the new material for incorporation in the collection. In the present study, assessments of six experts were compared to microsatellite data of 414 new varieties and 1408 existing accessions. Based on the microsatellite data, the extent to which the genetic diversity of the collection would be enriched (added value) was calculated for specific sets of new varieties. When individual assessments of experts were evaluated, the total added value of expert-based selections was not significantly higher compared to randomly chosen groups, except for a single expert. Unfamiliarity with new varieties was shown to be a crucial factor in the assessment of crop experts. According to the current acquisition protocol that seeks for consensus among experts, varieties are selected based on recommendations from at least three experts. This protocol also did not perform better than randomly chosen groups of new varieties. However, significantly better results were obtained with alternative protocols. It was concluded that breeding value was a more decisive criterion in the current acquisition protocol than maximal extension of the genetic diversity within the collection. A modified protocol addressing both commercial and diversity aspects was suggested in order to meet the demands of plant breeders as well as conservationists.

Keywords Collection composition · Genetic diversity · *Lactuca* · Lettuce · Microsatellites · PGR management · Plant breeding

Introduction

Crop collections established and maintained by genebanks facilitate the present and future utilization of crop genetic resources. In the early stages of genetic resources conservation, the main priority was to safeguard genetic diversity from erosion. As a result, genebanks included almost all material they had access to, rather than composing collections based on careful decisions about what material to include, and what material not to include in a collection. In many cases, this resulted in collections of considerable size, unbalanced composition and high levels of redundancy. It has been estimated that worldwide more than six million accessions are

R. van Treuren (✉) · Th.J. L. van Hintum
Centre for Genetic Resources, Wageningen University and Research Centre, P.O. Box 16, 6700 AA, Wageningen, The Netherlands
e-mail: Robbert.vanTreuren@WUR.NL

C. C. M. van de Wiel
Plant Research International B.V., Business Unit Biodiversity and Breeding, Wageningen University and Research Centre, P.O. Box 16, 6700 AA, Wageningen, The Netherlands

maintained in over 1,300 collections (FAO 1996). Based on a study of 2.5 million accessions belonging to 37 different crops, only 35% of the stored accessions are supposed to be unique (Plucknett et al. 1987). Therefore, improvement of the efficiency in managing plant genetic resources has gained considerable interest from genebank curators (Engels and Visser 2003). In this context, molecular marker data could potentially play an important role (Bretting and Widrlechner 1995; Brown and Kresovich 1996; Karp et al. 1997; Spooner et al. 2005), but so far their role has been rather limited.

Concerning the composition of genebank collections, it is generally agreed that the main goal is to constitute collections that represent as wide as possible genetic diversity of a crop gene pool with a minimum level of redundancy. To improve collection composition, curators nowadays put considerable effort in gaining information about genebank accessions. Characterization data obtained from morphological and/or molecular analyses are used to determine genetic relationships between accessions and to investigate collection structure (Powell et al. 1996; Milbourne et al. 1997; Russell et al. 1997). Characterization data in combination with passport data are sometimes used to identify or verify redundancies that subsequently can be removed to optimize collection composition (Waycott and Fort 1994; Virk et al. 1995; Phippen et al. 1997; van Treuren et al. 2001; Lund et al. 2003; van Treuren and van Hintum 2003).

Optimization of collection composition also involves decisions about acquisition in case genebanks obtain access to material that potentially may enrich the genetic diversity of a collection. These may include local landraces and crop related wild relatives obtained during collection missions and novel varieties released by the plant breeding industry. However, it is often unclear a priori to what extent the genetic diversity within the existing collection will be extended by the incorporation of new material. Insufficient knowledge about potentially interesting new material and about the genetic diversity of the existing collection hampers the careful selection of new material. In most cases, new material is simply included in the collection and the added value determined a posteriori, if at all.

The lettuce collection of the Centre for Genetic Resources, the Netherlands (CGN) comprises nearly

2,600 accessions, consisting of approximately 65% cultivated material and 35% crop related wild relatives. The collection is well-documented and has been characterized entirely with AFLPs and microsatellites (van Hintum 2003). These data are being used in combination with passport and other characterization data to rationalize the collection. Breeding efforts in lettuce are mainly directed to resistance to downy mildew (*Bremia lactucae*) (Crute 1992). For this purpose, the crop related wild gene pool is considered a rich source of novel resistance characters that may be introduced into a cultivated background (e.g. Bonnier et al. 1992; Sicard et al. 1999).

To enrich the genetic diversity of CGN's lettuce collection with cultivated material, once every five years an overview is made of novel varieties that were released by the plant breeding industry in the preceding period. Based on recommendations of a panel of crop experts, the top 10% of these varieties is then added to the collection. The present paper describes the results of a comparative study with microsatellites in order to relate information from crop experts with molecular marker data and to evaluate the decision protocol used by CGN to select new varieties based on the assessments of crop experts.

Material and methods

Study material

Investigation of the Dutch section of the common European variety list (1997–2001) resulted in 439 registered varieties. Seeds from 414 of these varieties, comprising 15 different plant breeding companies were obtained from the Centre for Variety Research, the Netherlands, which at the time of initiation of the study formed part of CGN. Seedlings were raised in a greenhouse and for each variety approximately 100 mg of leaf material was harvested from a single randomly chosen three-weeks-old plant.

Molecular analyses

Total genomic DNA was extracted from freeze-dried leafs using a combination of the methods described by Fulton et al. (1995) and the DNeasy 96 Plant Kit (Qiagen, Westburg, The Netherlands). Eleven lettuce microsatellites (van de Wiel et al. 1999) were

amplified by multiplex-PCR in four sets (Table 1). PCR was performed using a MJ PTC200 thermocycler and carried out in 20 μl reaction volumes containing 12 μl of PCR mix and 8 μl DNA template (2 ng/μl). The PCR mix consisted of an optimized concentration of primer pairs (Table 1), 100 μM of each dNTP, 10 mM Tris–HCl pH 9.0, 20 mM ($NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$ and 0.4 U Goldstar Taq DNA polymerase (Eurogentec, the Netherlands). The amplification profile consisted of an initial cycle of 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 45 s, and a final extension cycle at 72°C for 10 min. PCR products were purified on a 96-well Millipore multiscreen filter plate containing Sephadex G50 fine (Sigma), and 1.5 μl of purified product was mixed with 2.5 μl ultrapure formamide, 0.5 μl loading buffer (PE Biosystems) and 0.5 μl of ROX-labeled size standard. Samples were denatured for 2 min at 90°C, directly put on ice and 1.25 μl was loaded on an ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, Calif.). Fragment sizes and peak areas were determined automatically using the GENE-SCAN analysis software (release 1.1 3700 software, Applied Biosystems) and further processed with the software package Genotyper, version 3.5 NT (Perkin–Elmer).

In addition to the 414 samples from the new varieties, replicate samples from six different varieties were included in the analyses in order to verify reproducibility of the marker data. To enable comparison of the SSR data from the 414 new varieties with those of CGN's lettuce collection, stored DNA's from 20 reference samples were also included in the study. These samples, together with the remainder of CGN's lettuce collection, had been characterized before with AFLPs and SSRs (van Hintum 2003).

Crop expert information

The list of 439 new varieties was presented to six Dutch crop experts involved in lettuce breeding or plant variety registration. The crop experts were asked to independently qualify the new varieties according to the importance for inclusion in CGN's lettuce collection, using the categories 5 (very high), 4 (high), 3 (intermediate), 2 (low) and 1 (very low). Material was classified as 0 (unknown) in case crop experts were unfamiliar with a new variety. The standard protocol that is used by CGN to select varieties for accession is based on the highest qualifications from a combination of three experts. In the present study this procedure resulted in a preliminary selection of 32 new varieties that received either three times the qualification 5, or twice a 5 and once a 4 or 3, or once a 5 and twice a 4.

Data analysis

New varieties

The majority of new varieties could be scored for all investigated microsatellites. Four missing values were recorded for a single variety, while a single missing value was observed for eleven varieties. All

Table 1 Composition of the four multiplexes, PCR details and variability of the microsatellites used to screen the 414 new varieties

Multiplex	SSR locus	Fluorescent label	Optimized end- concentration of the primer pairs (pMol/reaction)	Number of observed alleles	PIC value
A	LsA001	NED	1.6	11	0.679
A	LsA004	FAM	1.2	9	0.587
A	LsD106	HEX	0.6	4	0.374
B	LsB101	NED	1.6	10	0.715
B	LsB104	FAM	1.6	8	0.531
B	LsD103	FAM	2.0	4	0.504
B	LsE003	HEX	4.0	4	0.169
D	LsD108	FAM	1.0	15	0.712
D	LsD109	HEX	1.0	12	0.742
E	LsE011	HEX	2.0	3	0.225
E	LsE018	NED	1.0	4	0.096

414 varieties were included in the data analyses. To obtain estimates of the level of information contained within the marker data, PIC values (Polymorphic Information Content) were calculated for each locus according to the methods of Botstein et al. (1980). In addition, the probability of finding identical multi-locus microsatellite genotypes by chance was esti-mated. For each locus, absence of null alleles was assumed to estimate allele frequencies, which were subsequently used to calculate the probability of finding two identical genotypes. Single-locus proba-bilities were then multiplied to obtain an estimate of the multi-locus probability. This estimation method relies on the assumption that loci segregate independ-ently. Mapping information about the microsatel-lites was available from previous research and the set of markers used in the present study was selected based on a wide genomic distribution, including coverage of all chromosomes (van de Wiel, unpub-lished data). Therefore, the probability that the estimation procedure was biased because of co-segregating loci is highly unlikely. To investigate the ability of microsatellites to discriminate between the different lettuce crop types, genetic relationships between samples were visualized by Principal Coor-dinates (PCO) using the software package Genstat (release 8.11). For this purpose, similarity values were calculated based on Jaccard's coefficient after transforming the microsatellite data to binary scores for allele absence and presence.

CGN collection

Apart from locus LsE018, data from the same microsatellites (Table 1) were available for CGN's entire lettuce collection sampled in autumn 1997 (van Hintum 2003). Because the present study focused on cultivated material, wild crop relatives were removed from the dataset. Apart from a few cases, the majority of accessions were represented by two samples. The data set was further reduced by maintaining only one sample per accession in case of identical microsat-ellite profiles and by excluding samples with more than two missing values. The final data set used in the present study comprised a total of 1,688 samples from 1,408 accessions, representing 92% of the cultivated material within CGN's current collection. A summary of all material included in the present study is given in Table 2.

Table 2 Number of accessions of CGN's lettuce collection and number of new varieties included in the microsatellite analysis

Crop type	CGN collection	New varieties
Butterhead	718	191
Crisp	219	90
Cos	217	29
Cutting	152	73
Latin	56	14
Stalk	32	–
Oilseed	8	–
Unknown type	6	17
Total	1,408	414

Data are presented for each of the main lettuce crop types

Added value of new varieties to the CGN collection

To determine the extent to which the genetic diversity of the CGN collection would be enriched by includ-ing new varieties, a parameter called "added value" was used. The added value of new varieties was estimated from the two microsatellite data sets using the following steps. (1) Each new variety was compared with each of the existing CGN accessions, and for each comparison it was recorded at how many loci a different microsatellite profile was observed. (2) For each new variety the added value to the collection was calculated as the mean number of differences with the five genetically most similar accessions, i.e. the five accessions with the lowest number of differences. (3) The new variety with the highest added value was virtually added to the collection. (4) For each of the remaining new varieties the whole procedure was repeated, each time with a reduced set of new varieties and an expanded collection. By this sequential procedure the order of adding new entries to the collection was determined based on maximization of the added value in each step, taking into account the variation both within the collection and among the new varieties.

The same approach using the marker data was carried out for varieties that were considered of very high and high importance for inclusion in the collection by each of the six crop experts individually and for the preliminary selection according to CGN's standard protocol based on the combined assessments

of at least three experts. Permutation or randomization tests (e.g. Sokal and Rohlf 1981) were then used for significance testing of candidate selections. The total added value of a randomly selected set of new varieties was calculated, and based on 1,000 permutations the proportion of equal or higher added values than obtained for the candidate selection was recorded. This proportion, denoted as P, was used to test whether candidate selections performed significantly better than random selections. The sample size used for the permutation tests could vary depending on the size of the candidate selection. To investigate the effect of unfamiliarity with new material, similar analyses were also carried out using only the varieties that were known to the individual crop expert. These reductions of the group from which varieties are chosen have no effect on experts-based selections but may influence the marker-based and random selections, depending on the distribution of genetically distinct material between the groups of known and unknown varieties. The marker data were also used to evaluate the added value of specific breeding companies by comparing the added value of their varieties with those of randomly chosen sets of the same size.

To determine the optimal protocol for combining expert-based assessments in terms of added value, various alternative protocols were examined selecting a fixed number of 40 new varieties in each case. The following protocols were compared:

(1) CGN's standard protocol denoted by "CGN standard", using the average of the three highest scores received from experts as ranking order.

(2) The highest scores per individual expert denoted by "Expert 1" to "Expert 6", using the ranking order: "very high"—"high" — "unknown".

(3) The single highest score denoted by "High 1", selecting varieties that received the qualification "very high" from at least a single expert.

(4) High scores from at least two experts denoted by "High 2", using the ranking order: twice "very high"—once "very high" and others "unknown"—once "very high" and once "high".

(5) High scores from at least three experts denoted by "High 3", using the ranking order: three times "very high"—twice "very high" and others "unknown"—once "very high" and others "unknown"—once "very high", once "high" and others "unknown" —average of the three highest scores.

(6) The highest average score denoted by "Average", using the average of all scores per variety excluding the category "unknown".

(7) Unknown material supplemented with highest average denoted by "Unknown + average", using the ranking order: six times "unknown"—highest average score excluding the category "unknown".

(8) The single highest score after standardization to average 0 and standard deviation 1 for each expert, denoted by "Standardized high"

(9) The average score after standardization to average 0 and standard deviation 1 for each expert, denoted by "Standardized average".

Because of the fixed sample size, in some cases varieties needed to be chosen randomly from a group of varieties with the same score. In those cases, five random selections were made and the average results were used for comparison. All analyses were carried out using tailor-made computer programs written in Turbo Pascal.

Results

New varieties

Among the 414 new varieties, a total number of 84 alleles were observed for the 11 microsatellites, the number of alleles per locus ranging from 3 to 15. The PIC values of the markers ranged from 0.096 to 0.742, depending on the number of alleles and their frequencies among the investigated sample (Table 1). Among the new varieties, a total number of 281 multi-locus microsatellite profiles were observed. Unique microsatellite profiles were found for 231 varieties, whereas a total of 50 different genotypes were observed for the remaining 183 varieties. Identical microsatellite genotypes were found both within and between different plant breeding companies and always involved the same lettuce crop type, except in two cases. Using the microsatellite allele frequencies among the total sample, the probability of finding identical multi-locus genotypes purely by chance was estimated to be

1.04×10^{-7}. Therefore, shared microsatellite profiles observed between varieties seem to point towards common ancestry. Microsatellite allele frequencies varied markedly among the lettuce crop types, which were clearly distributed non-randomly in a principal coordinate plot of the investigated varieties (Fig. 1). Particularly between butterhead, crisp and cutting lettuce, the majority of varieties occupied different positions in the plot. However, also overlap between different crop types was observed, suggesting the occurrence of intermediate types.

Comparison of new varieties to the CGN collection

All microsatellite alleles observed among the new varieties could also be found in accessions already included in the CGN collection, except in two cases. For locus LsB101 a novel allele of ∼217 bp and for locus LsB104 a novel allele of ∼205 bp was observed. However, based on the multi-locus microsatellite profiles, 247 genotypes were observed among the 414 new varieties that could not be matched with available genotypes from CGN's existing collection. The number of different scores with the genetically most similar CGN sample ranged from 0 to 4 loci.

Qualifications given to new varieties for inclusion in CGN's collection differed substantially among the six crop experts (Table 3A). The number of unknown varieties varied markedly between the experts ranging from 90 (22%) to 380 (92%). Experts 1, 2 and 3 were unfamiliar with the majority of new varieties. A relatively high proportion of the known varieties was designated very highly important by expert 3 (27%), while expert 4 considered the majority of known varieties being of very low importance (59%). With the exception of experts 1 and 2, the average added value calculated for individual varieties declined from material considered of very high importance to material considered of medium importance (Table 3B). In general, the decline did not continue to material considered of very low importance. For the category of unknown material, relatively high added values were observed, suggesting the presence of potentially interesting varieties within this group.

When tested against randomly chosen sets of identical size using a significance level of 5%, it appeared that a significantly better performance was achieved only by expert 6 (Table 4A). Also the preliminary selection based on CGN's standard protocol was not significantly better than a randomly chosen group of the same size. Except in the case of expert 2, the results improved considerably when the analyses were performed without the varieties unknown to the crop expert (Table 4B). Also expert 5 now performed significantly better than a random selection, while the test results for expert 3 revealed a *P*-value equal to the significance level. These results indicated their ability to make good choices from the material that they are familiar with. However, the preliminary selection based on CGN's standard protocol still did not perform any better than a randomly chosen group.

In order to investigate how CGN's standard protocol could be improved, several alternative protocols were investigated (Table 5). Out of the investigated alternatives, a protocol solely based on the recommendations of expert 6 resulted in the lowest *P*-value, and thus the best selection. However, also a protocol based on the highest average score over the six experts performed significantly better than randomly chosen groups. A nearly identical performance was observed for a protocol based on unfamiliarity by all experts (14 varieties) supplemented with highest average score. All other investigated protocols resulted in non-significant differences compared to randomly chosen varieties. The relative performance of CGN's standard protocol in comparison with alternative protocols and a protocol solely based on the SSR data is graphically represented in Fig. 2.

Concerning the origin of the varieties, the analyses revealed only a single breeding company that appeared to release varieties with a significantly higher added value to the collection than randomly selected groups ($P = 0.013$). Interestingly, expert 6 is associated with this plant breeding company which may explain the expert's good performance in recommending varieties with a large added value to CGN's collection.

Discussion

Contributing to the conservation of genetic resources for present and future use, genebanks aim to constitute collections that represent as wide as possible
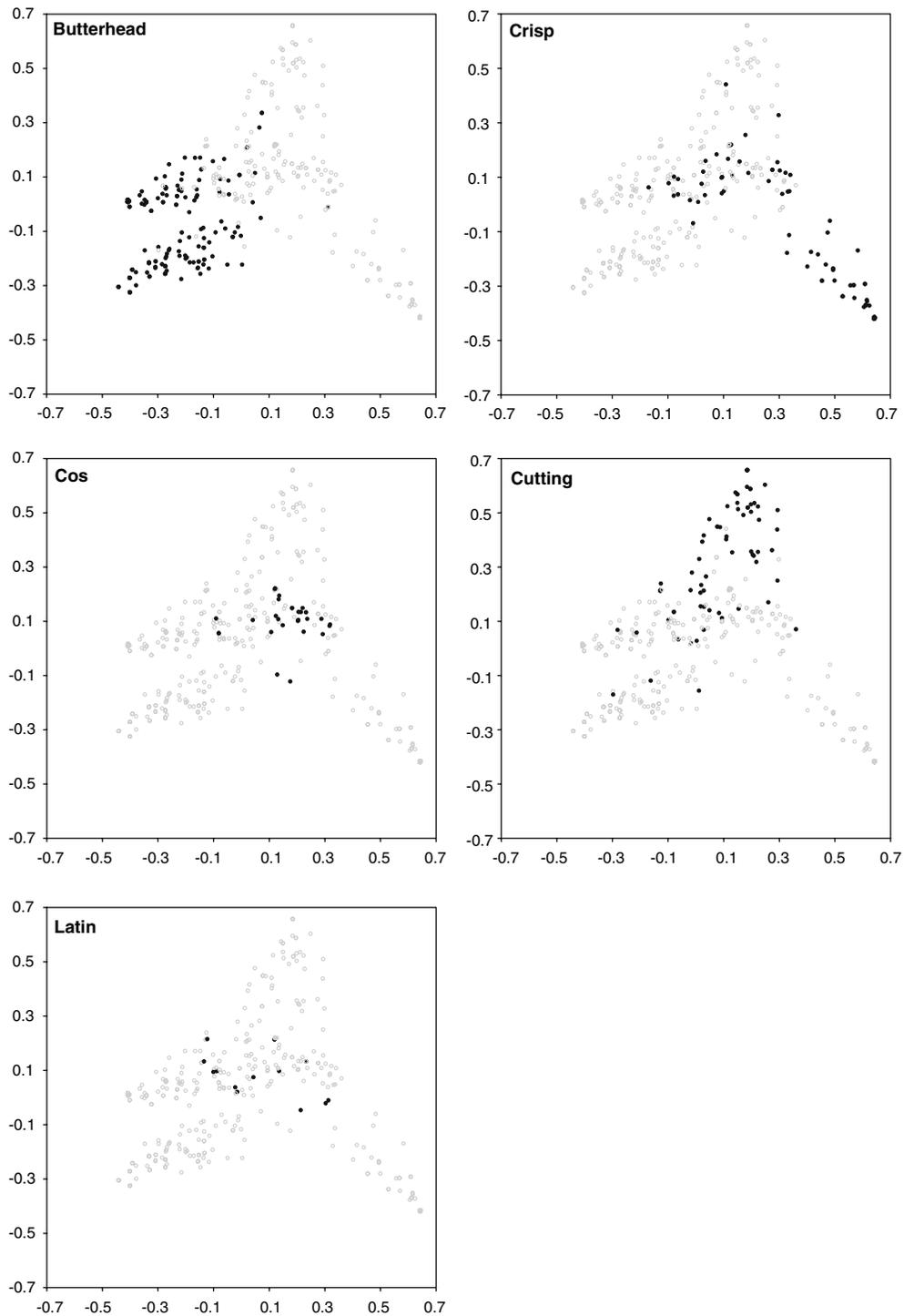
Fig. 1 Principal coordinate plot of the 414 new varieties, presented in separate graphs for each of the five lettuce crop types. The first two principal axes explained 14.7% and 9.7%, respectively, of the total observed variation

Table 3 Distribution of scores assigned by the six experts to new varieties (A), and the corresponding average individual added value of the varieties (B)

Expert	Very high	High	Medium	Low	Very low	Unknown
A. Number of varieties						
1	5	3	9	12	5	380
2	7	8	14	22	5	358
3	29	8	34	12	26	305
4	5	19	27	35	126	202
5	20	57	43	60	53	181
6	44	51	78	123	28	90
B. Average individual added values						
1	1.24	1.27	1.47	1.07	0.52	1.24
2	1.69	1.13	1.73	0.90	1.80	1.21
3	1.19	1.15	0.92	0.60	1.02	1.31
4	1.48	1.16	0.84	0.91	1.25	1.32
5	1.26	1.16	0.96	0.97	0.82	1.51
6	1.65	1.21	1.07	1.20	0.94	1.30
Weighted mean	1.43	1.18	1.05	1.05	1.09	1.29

Classifications ranged from very high to very low importance for inclusion in CGN's lettuce collection. In the last column the number of varieties unknown to the expert is presented

Table 4 Total added value of new varieties to the diversity of CGN's collection based on the recommendations by crop experts (varieties that received the score "very high" or "high" importance) compared to a random selection

Expert	Group	Domain	Total added value		P
			Expert-based	Random	
A. All new varieties					
1	8	414	10.0	9.7	0.445
2	15	414	20.6	18.4	0.258
3	37	414	43.0	44.6	0.630
4	24	414	28.4	29.4	0.580
5	77	414	88.2	91.8	0.693
6	95	414	130.8	112.4	0.010
CGN standard	32	414	34.2	39.0	0.827
B. New varieties known to experts					
1	8	34	10.0	9.0	0.335
2	15	56	20.6	19.7	0.407
3	37	109	43.0	35.8	0.050
4	24	212	28.4	26.7	0.337
5	77	233	88.2	73.6	0.012
6	95	324	130.8	109.9	0.002
CGN standard	32	193*	34.2	32.5	0.348

"Group" denotes the number of recommended varieties and "Domain" the number of varieties from which the selection was made. *P* represents the probability that a random selection resulted in an equal or larger total added value than the selection of the crop expert. Results according to CGN's currently used protocol to combine expert information are denoted by "CGN standard". Separate analyses were performed for all new varieties (A) and for varieties known to experts (B)

* Cultivars known to at least three crop experts

Table 5 Total added value of 40 new varieties selected based on CGN's standard protocol and on various alternative methods (see material and methods for details)

Protocol	Total added value	P
CGN standard	49.2	0.417
Expert 1	51.1	0.284
Expert 2	48.5	0.463
Expert 3	46.1	0.634
Expert 4	48.3	0.479
Expert 5	43.8	0.789
Expert 6	66.1	0.000
High 1	55.8	0.084
High 2	57.0	0.057
High 3	51.0	0.293
Average	59.0	0.026
Unknown + average	59.1	0.024
Standardized high	48.7	0.453
Standardized average	48.0	0.504
Random	48.1	

The probability that a random selection resulted in an equal or larger total added value than the investigated protocol is denoted by P in the last column

genetic diversity of a crop gene pool with a minimum level of redundancy. Genebanks always need to consider improvement of the collection composition in case new material is available for inclusion. This includes novel varieties released by the plant breeding industry. In the case of lettuce, new varieties are
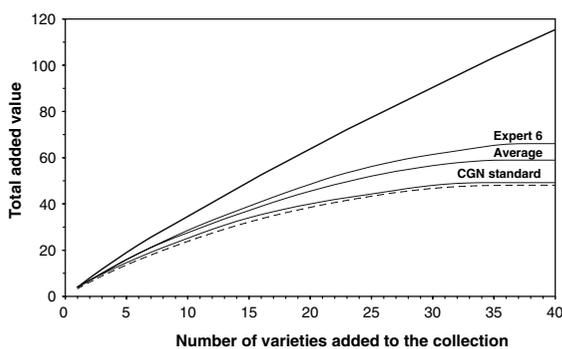
Fig. 2 Total added value to CGN's lettuce collection for different selections of 40 new varieties (see text for details). Results are presented for selections based on CGN's standard protocol, on the scores of expert 6 and on the average scores of the experts (solid lines). The dashed line represents the results for 1,000 randomly chosen selections out of the total group of 414, while the bold solid line denotes the results when selections were optimized based on microsatellite data

marketed regularly. Between 1997 and 2001, approximately 700 varieties were registered in the common European variety list. Because collection size is often problematic for many genebanks, introducing all new varieties in a collection is not a feasible option. Moreover, such practice would most likely increase the level of redundancy within the collection because often a more or less fixed gene pool is used in plant breeding. In close cooperation with crop experts, CGN only accesses a selection of available new lettuce varieties in order to maintain a manageable collection size and to avoid redundancy. The present study focused on the protocol for using information from crop experts by relating that information to microsatellite fingerprinting results.

In previous research, CGN's entire lettuce collection was screened with AFLPs and microsatellites (van Hintum 2003). Both marker systems are generally highly informative and usually yield highly reproducible results. Microsatellites are particularly useful when data generated in different experiments or labs have to be compared (Bredemeijer et al. 2002; Röder et al. 2002). The microsatellite markers that were used in the present study were developed from a library of total genomic DNA (van de Wiel et al. 1999). It has been questioned to what extent molecular marker variation is associated with variation in quantitative genetic characters and life-history traits (Reed and Frankham 2001). In the absence of such an association, only insignificant differences between the added value of expert-based selections and that of randomly chosen groups of new varieties could have been expected in the present study, but this was clearly not the case. In addition, microsatellite variation was reasonably well associated with variation in lettuce crop types. These two findings are difficult to explain if the observed marker variation is not to some extent associated with phenotypic variation.

Based on the frequency of microsatellite alleles among the new varieties, the probability of finding identical genotypes by chance was estimated to be 1.04×10^{-7}. Despite this high resolving power, identical microsatellite profiles were frequently observed among varieties, which may seem somewhat surprising. However, particularly if varieties differ for only a single or a very limited number of traits, a low probability of detecting differences can be expected when studying a small subset of the

genome. Therefore, it is concluded that, given the estimated resolving power, varieties with identical microsatellite profiles can be considered closely related.

In lettuce, resistance to downy mildew is by far the most important trait in the development of new varieties (Crute 1992). Introgression of new resistance characters in existing varieties that otherwise possess desirable traits, and introduction of desirable traits in highly resistant cultivars is common practice in lettuce breeding. Therefore, new varieties may differ only for a limited number of characters from other new varieties and from existing cultivars. This was supported by the identification of 50 groups of new varieties that could not be distinguished from each other based on the investigated microsatellites. Out of these 50 groups, 24 were found to match CGN accessions. The largest group consisted of 24 varieties that matched accessions belonging to the cultivar type 'Hilde'/'Attraction'. CGN's collection includes 68 accessions of this cultivar type that is known for its good yield and absence of leaf anthocyanin. Therefore, the group of 24 similar varieties seemed to represent variations of this "popular" cultivar type.

To examine the extent to which new varieties could enrich the genetic diversity of the collection, a parameter called 'added value' was used. This parameter was defined as the mean number of differences at microsatellite loci with the five genetically most similar accessions. Focusing on five instead of a single accession enabled discrimination between common and rare genotypes occurring in the collection, resulting in lower added values for new varieties that resemble a relatively frequent genotype within the collection. Calculations were also carried out based on the three and ten genetically most similar accessions, respectively, but this showed only minor effects on the results (data not shown). Quantification of genetic differences between samples in terms of the number of different scores at microsatellite loci resembled the calculation of dissimilarity values often used to estimate genetic relatedness based on dominant marker data (e.g. González-Candelas and Palacios 1997). A drawback of this parameter is that information contained in the codominant nature of microsatellites is ignored. However, because lettuce is a self-fertilizer, heterozygosity may be expected to occur only occasionally. In the present study, heterozygosity was observed in

only 0.49% and 0.69% of the scores obtained from, respectively, the new cultivars and the CGN accessions. Therefore, calculation of added values based on the percentage allele sharing will have a negligible effect on the obtained results. Added values were calculated using an iterative procedure, virtually extending the collection step by step with the variety displaying the largest added value. The rationale behind this approach is that the similarity amongst new varieties is taken into account and that introduction of large numbers of similar material in the collection is avoided.

Added values for expert-based recommendations were analyzed for varieties that received the qualification 'very highly' or 'highly' important. Analyses were also performed for the class of very highly important varieties only, but these had little effect on the results (data not shown).

Considerations used by experts to qualify new varieties may include market impact, degree of uniqueness based on pedigree data and avoidance of introducing sister lines in the collection. The quality of the final selection depends on several factors. First, the crop experts need to have adequate knowledge about new varieties. It may be questioned whether this is a realistic demand since the number of new lettuce varieties released in a five-year period amounts several hundreds. Crop experts cannot be expected to know each of the new varieties in detail, particularly when material from other breeding companies is involved. Pedigree data for instance are often treated as confidential because of commercial interests. Because of unfamiliarity with material from other breeders, expert 1 qualified only varieties released by the company to which the expert was affiliated. Data presented in Table 4 indicated that unfamiliarity with new varieties had a significant influence on the performance of crop experts. Second, the crop experts need to have sufficient knowledge about the accessions of the collection. Also this may be questioned given the fact that the current lettuce collection of CGN comprises nearly 2,600 accessions and that detailed knowledge requires a long lasting experience with the collection. The long lasting experience in lettuce breeding may explain the consistently good performance of expert 6 (Table 4). Alternatively, one could argue that, regardless of experience, expert 6 may have tended to recommend predominantly varieties from the own company,

which appeared to release varieties with a high added value. However, this was considered unlikely based on the fact that expert 6 rejected (qualification "low" or "very low" importance) 70 out of the 131 varieties released by the own company and that 58 out of the 95 varieties recommended by expert 6 were released by other companies. Apparently, expert 6 was able to distinguish well between varieties with low and varieties with high added value, both from the own and from other companies. Third, the information of crop experts needs to be combined in a meaningful way. CGN's current protocol aims at the selection of approximately 10% of the new varieties by focusing on material that received high qualifications from a combination of three crop experts. However, this protocol did not perform any better than randomly chosen groups (Table 4). It appeared that only 47% of the new varieties were known to at least three experts, which means that the majority of the material has no probability whatsoever to be selected, regardless of a high appreciation by only a single or few experts. Therefore, the impact of a variety, as reflected by the number of experts that were familiar with it, seemed a decisive criterion in CGN's standard protocol. It was shown that the protocol could be optimized by selecting varieties based on the sole recommendations of expert 6 or on the highest average scores of all experts.

Services of genebanks include providing access to materials that meet the short-term needs of the plant breeding industry as well as the long term conservation of genetic resources for the future. Plant breeders are generally more interested in variation in specific characters that may result in varieties with a high market impact, while curators are concerned also about the broad genetic diversity. Therefore, the ideal composition of a collection will depend on the perspective. To meet the demands from both groups, it is suggested to improve the acquisition protocol by taking both genetic diversity and impact in plant breeding into account. This can be achieved by selecting a subset of varieties based on the average qualifications of all experts, supplemented with a selection based on the qualifications from a combination of three crop experts.

References

Bonnier FJM, Reinink K, Groenwold R (1992) New sources of major gene resistance in *Lactuca* to *Bremia lactucae*. Euphytica 61:203–211

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bredemeijer GMM, Cooke RJ, Ganal MW, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Genet 105:1019–1026

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resource management. Plant Breed Rev 13:11–86

Brown SM, Kresovich S (1996) Molecular characterization for plant genetic resources conservation. In: Paterson H (ed) Genome mapping of plants. Academic Press, San Diego, pp 85–93

Crute IR (1992) The role of resistance breeding in the integrated control of downy mildew (*Bremia lactucae*) in protected lettuce. Euphytica 63:95–102

Engels JMM, Visser L (2003) A guide to effective management of germplasm collections. IPGRI handbook for Genebanks No. 6. International Plant Genetic Resources Institute, Rome, Italy

FAO (1996) FAO state of the world's plant genetic resources for food and agriculture. Food and Agriculture Organisation of the United Nations, Rome

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209

González-Candelas F, Palacios C (1997) Analyzing molecular data for studies of genetic diversity. In: Ayad WG, Hodgkin T, Jaradat A. Rao VR (eds) Molecular genetic techniques for plant genetic resources. Report of an IPGRI Workshop, 9–11 October 1995, Rome, Italy. International Plant Genetics Resources Institute, Rome, Italy, pp 55–81

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T (1997) Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy

Lund B, Ortiz R, Skovgaard IM, Waugh R, Anderson SB (2003) Analysis of potential duplicates in barley genebank

collections using re-sampling of microsatellite data. Theor Appl Genet 106:1129–1138

Milbourne D, Meyer R, Bradshaw JE, Baird R, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3:127–136

Phippen WB, Kresovich S, Candelas FG, McFerson JR (1997) Molecular characterization can quantify and partition variation among genebank holdings: a case study with phenotypically similar accessions of *Brassica oleracea* var. *capitata* L. (cabbage) 'Golden Acre'. Theor Appl Genet 94:227–234

Plucknett DL, Smith NJH, Williams JT, Anishetty NM (1987) Gene banks and the world's food. Princeton University Press, Princeton, New Jersey, USA

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Sicard D, Woo SS, Arroyo-Garcia R, Ochoa O, Nguyen D, Korol A, Nevo E, Michelmore R (1999) Molecular diversity at the major cluster of disease resistance genes in cultivated and wild *Lactuca* spp. Theor Appl Genet 99:405–418

Sokal RR, Rohlf FJ (1981) Biometry, 2nd edn. WH Freeman and Company, New York

Spooner D, van Treuren R, de Vicente MC (2005) Molecular markers for genebank management. IPGRI technical bulletin No. 10. International Plant Genetic Resources Institute, Rome, Italy

Van de Wiel CCM, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.) Genome 42:139–149

Van Hintum ThJL (2003) Molecular characterization of a lettuce germplasm collection. In: van Hintum ThJL, Lebeda A, Pink D, Schut JW (eds) Eucarpia leafy vegetables 2003, Proceedings of the Eucarpia meeting on leafy vegetables, Noordwijkerhout, The Netherlands, 19–21 March 2003. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands, pp 99–104

Van Treuren R, van Hintum ThJL (2003) Marker-assisted reduction of redundancy in germplasm collections: genetic and economic aspects. Acta Hort 623:139–149

Van Treuren R, van Soest LJM, van Hintum ThJL (2001) Marker-assisted rationalisation of genetic resources collections: a case study in flax using AFLPs. Theor Appl Genet 103:144–152

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (1995) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90:1049–1055

Waycott W, Fort SB (1994) Differentiation of nearly identical germplasm accessions by a combination of molecular and morphologic analyses. Genome 37:577–583