ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization **8**(2); 171–181
doi:10.1017/S1479262110000134

Distribution of genetic diversity in wild European populations of prickly lettuce (*Lactuca serriola*): implications for plant genetic resources management

C. C. M. van de Wiel[1]*, T. Sretenović Rajičić[2†], R. van Treuren[3], K. J. Dehmer[2], C. G. van der Linden[1] and Th. J. L. van Hintum[3]

[1]*Wageningen University and Research Centre Plant Breeding, PO Box 16, 6700 AA Wageningen, The Netherlands,* [2]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and* [3]*Centre for Genetic Resources, The Netherlands (CGN), Wageningen University and Research Centre, Wageningen, The Netherlands*

Received 15 February 2010; Accepted 21 April 2010 – First published online 25 May 2010

Abstract

Genetic variation in *Lactuca serriola*, the closest wild relative of cultivated lettuce, was studied across Europe from the Czech Republic to the United Kingdom, using three molecular marker systems, simple sequence repeat (SSR, microsatellites), AFLP and nucleotide-binding site (NBS) profiling. The 'functional' marker system NBS profiling, targeting disease resistance genes of the NBS/LRR family, did not show marked differences in genetic diversity parameters to the other systems. The autogamy of the species resulted in low observed heterozygosity and high population differentiation. Intra-population variation ranged from complete homogeneity to nearly complete heterogeneity. The highest genetic diversity was found in central Europe. The SSR results were compared to SSR variation screened earlier in the lettuce collection of the Centre for Genetic Resources, the Netherlands (CGN). In the UK, practically only a single SSR genotype was found. This genotype together with a few other common SSR genotypes comprised a large part of the plants sampled on the continent. Among the ten most frequent SSR genotypes observed, eight were already present in the CGN collection. Overall, the CGN collection appears to already have a fair representation of genetic variation from NW Europe. The results are discussed in relation to sampling strategies for improving genebank collections of crop wild relatives.

Keywords: collection; genetic diversity; *Lactuca*; molecular markers; plant genetic resources; wild relatives

Introduction

Prickly lettuce, *Lactuca serriola*, is the most closely related wild relative of cultivated lettuce (*Lactuca sativa*). Thus, *L. serriola* is of high interest for introgressing traits into cultivated lettuce, in particular, resistance against the downy mildew *Bremia lactucae*, which poses serious problems to lettuce cultivation worldwide (Crute, 1992). *L. serriola* is an example of a well-represented wild species in worldwide genebank collections. Nevertheless, based on data from the 'International *Lactuca* database' (ILDB, www.documents.plant.wur.nl/cgn/pgr/ildb/), Lebeda *et al.* (2004) showed that imbalances exist in these

* Corresponding author. E-mail: clemens.vandewiel@wur.nl
† Present address: De Ruiter Seeds, Bergschenhoek, The Netherlands

collections. For example, there is far more material of *L. serriola* from Europe than from other regions, such as Asia. Moreover, even within Europe, accessions' origins are not evenly distributed geographically.

The most important reason for maintaining wild accessions in genebanks is making them accessible to users, such as researchers and plant breeders. Particularly in *L. serriola*, conservation *per se* appears not to be an important issue, since the species is common across its distribution area and even has shown a strong increase in occurrence in recent times (D'Andrea *et al.*, 2009). To optimize users' access to *L. serriola* diversity, a genebank collection should represent the variation in nature as widely as possible within the limits of its resources. For several reasons, this is not a simple task. Samples from some regions may only have been available from other collections, which might not have optimally documented their origin. Also, regions may be underrepresented or lacking entirely. Furthermore, seeds may have been collected using different sampling strategies (Guarino *et al.*, 1995). Most genebanks sample many plants per population, mixing the sampled seeds in a single batch per population before uptake into the collection. Others believe that the seeds of individual plants sampled from a population should be kept separate to allow for proper assessment of the population structure. Sometimes, collecting a single plant per population suffices, resulting in a 'pure' line, which is preferred by breeders as starting material. Given the variation in sampling strategies, the genetic variation within field populations may be missed, or partially lost during seed multiplication, in the eventual genebank accessions. Moreover, sampling based solely on geographic representation may overlook important genetic variation that is not immediately apparent in the field from e.g. morphological assessment. In this regard, little is known about the distribution of genetic variation both within and between wild populations of *L. serriola* across its distribution area. Lebeda *et al.* (2009) assessed genetic variation across Europe using AFLP on single plants from 50 sampling places, and found that groups could be distinguished that generally correlated with eco-geographic regions. Kuang *et al.* (2008) screened about 700 plants from 41 populations from Turkey, Armenia, Israel, Europe and California using both AFLP and a microsatellite located inside an intron of the RGC2 cluster of nucleotide-binding site/leucine-rich repeat (NBS/LRR) disease resistance gene homologues. Shen *et al.* (2002) showed that genes encoding resistance against *B. lactucae* belong to this cluster. Variation in the RGC2 cluster was correlated with genome-wide variation assessed by AFLP, and the highest diversity was found in Turkey and Armenia.

For the present study performed in the context of the European research project, GENE-MINE (www.genemine.org), *L. serriola* seeds were sampled separately from individual plants of a series of populations across NW Europe from the Czech Republic to the United Kingdom (Lebeda *et al.*, 2007), to assess both intra- and inter-population variation. All populations were characterized for morphological traits, resistance against pathotypes of *B. lactucae* and lettuce mosaic virus (LMV), respectively, and molecular genetic variation. For the latter, we used two systems targeting 'neutral' genomic diversity, the dominant AFLP (review: Meudt and Clarke, 2007) and the co-dominant microsatellites (simple sequence repeat, SSR, review: Schlötterer, 2004). To explore the possibilities of assessing agronomically relevant genomic variation, we also applied a relatively novel method referred to as NBS-directed profiling (NBSp). NBSp targets variation in and around members of the NBS/LRR disease resistance gene family, using PCR primers designed from conserved parts of this gene family (van der Linden *et al.*, 2004), to which the RGC2 cluster mentioned above belongs. The geographic distribution of variation within and between populations is described and compared to that represented in the CGN collection, which has a fair representation of wild lettuce material. Previously, the molecular genetic variation of this collection has been screened using AFLP and microsatellites (van Hintum, 2003), and so a comparison can be made between variation *in situ* and that in an *ex situ* collection. Implications of the results for efficient sampling strategies for *L. serriola* are discussed.

Materials and methods

Plant material

Seeds were sampled from 16 plants in each of 50 populations, 10 in the United Kingdom (UK), 8 in the Netherlands (NL), 16 in Germany (D) and 16 in the Czech Republic (CZ). The geographical distribution of the sampled populations and more detailed information about the sampling procedures, the morphological determination of the plants as either *L. serriola* f. *serriola* or *L. serriola* f. *integrifolia*, and the ecology of the populations can be found in Lebeda *et al.* (2007). Seedlings were grown, young leaves were harvested, and DNA was extracted in 96-well plates according to Doyle and Doyle (1990), with modifications for a robotic liquid handling system (Sretenović Rajičić *et al.*, 2008).

AFLP

The AFLP analysis was performed using the procedure as described in Sretenović Rajičić *et al.* (2008).

Three primer combinations were used for selective amplification, namely the EcoRI primer E00 + ACA in combination with each of the MseI primers M00 + CTAT, M00 + CTTC and M00 + CTTT, respectively. Amplification products were analysed on a MegaBACE 1000 sequencer, following the manufacturer's protocol, and peak presence/absence was scored using Fragment Profiler 1.2 software.

NBS profiling

NBSp was performed according to the procedure originally described by van der Linden *et al.* (2004), with modifications according to Syed *et al.* (2006). For amplification, the standard adapter primer was combined with NBS-specific primer NBS5a (YYTKRTHGTMITKGATGAYGTITGG) or NBS3 (GTWGTYTTICCYRAICCISSCATICC), respectively. In each case, the NBS-specific primer was labelled with γ^{33}P-ATP for the visualization by autoradiography of amplification products after electrophoresis on 6% polyacrylamide sequencing gels.

Simple sequence repeat (microsatellites)

Eleven microsatellites (van de Wiel *et al.*, 1999) were amplified by multiplex PCR of two to four SSR markers, labelled with HEX, NED or 6-FAM. PCR was performed in a 20 μl reaction volume containing 10 ng of genomic DNA, 2–4 pmol of each primer, 100 μM of each dNTP, 10 mM of Tris–HCl pH 9.0, 20 mM of $(NH_4)_2SO_4$, 0.01% of Tween 20, 1.5 mM of $MgCl_2$ and 0.4 Unit of Goldstar Taq DNA polymerase (Eurogentec, Maastricht, The Netherlands). The optimized PCR conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 120 s and a final extension at 72°C for 3 min. After amplification, the fluorescently labelled products were detected using a capillary-based ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Fragment sizes and peak areas were determined automatically using Genescan Analysis Software (Release 1.1 3700 software, Perkin Elmer Biosystems). Two loci were found to also have null alleles, LsE003 and LsD103, the latter to such a large extent that it was removed from the data analysis.

Comparison with genetic diversity present in the CGN collection

For comparing genetic marker data on these 50 populations to genetic diversity observed in the CGN collection, SSR data previously obtained from this collection were used (van Hintum, 2003). The other marker systems could not be used for such a comparison, since the CGN collection has not been characterized with NBSp and because of the difficulty in aligning the AFLP scores from the two studies due to the different visualization methods used (fluorescent labelling *vs.* autoradiography). At the time of the earlier SSR study, the CGN collection comprised 1487 accessions of *L. sativa* and 627 accessions of *L. serriola*. Because the microsatellite analyses were carried out on an ABI377 gel-based sequencer (instead of the capillary-based 3700) in the earlier study, twenty samples from the CGN collection and three standard crop lettuce samples (Balisto, Madrilene and Tianjin Big Stem) were rerun in order to facilitate the calibration of allele assignments between the two studies (cf. van de Wiel *et al.*, 1999; van Treuren *et al.*, 2008).

Data analysis

Nei's (1973) expected heterozygosity/gene diversity, *F* statistics and distance matrices, based on Nei' (1978) unbiased genetic distance measure, were calculated using Popgene 32 (www.ualberta.ca/~fyeh/). Principal coordinate (PCO) plots were constructed using Genstat version 7 (VSN International Ltd, Hemel Hempstead, UK) and NTSYSpc version 2.1 (Exeter Software, Setauket, NY, USA). Mantel tests were performed using Genstat version 7. To estimate the quality of the marker datasets with respect to the ability to describe the genetic structure within the study material, data resolution (DR) values were calculated according to the methods described by van Hintum (2007). The DR calculates the average correlation between two similarity matrices after repeatedly partitioning of the data set in two random halves. The method has been shown to estimate the quality of the marker datasets with respect to the ability to describe the genetic structure within the material studied.

Results

Overall diversity

For microsatellites, ten loci distributed across all nine linkage groups of lettuce (cf. Truco *et al.*, 2007) were scored. The number of alleles per locus ranged from 3 to 23, with a total of 128 across all nine loci. In line with lettuce's highly selfing nature, observed heterozygosity (average proportion of heterozygotes per locus) was very low: 0.0042. Overall, 3.1% of the plants were heterozygous for one or multiple loci. Expected

heterozygosity for the total sample was 0.32, and F_{st} was 0.54, indicating a high level of population differentiation. For AFLP, three primer combinations were used from which 173 polymorphic bands could be scored. Overall expected heterozygosity with AFLP was 0.05, and overall G_{st} was 0.54. For NBSp, two primer combinations were used from which 91 polymorphic bands could be scored. Overall expected heterozygosity with NBSp was 0.09. The overall G_{st} was 0.54, which was in line with the high population differentiation found with the other marker systems.

Intra-population diversity

A broad range of genetic variability within populations was observed, with gene diversities varying from 0.00 to 0.63 with SSR, from 0.00 to 0.15 with AFLP and from 0.00 to 0.21 with NBSp. The amount of within-population variation differed markedly between countries (Table 1). The UK showed the lowest gene diversities with all the marker systems, whereas Germany and the Czech Republic had the highest values. Gene diversity values for the Netherlands were intermediate between those of Germany and the Czech Republic on the one hand, and the UK on the other hand.

Inter-population diversity

Genetic relationships among all populations are presented by PCO plots in Fig. 1. The low variability and close relationships of the UK populations are similar for all the three marker systems. Furthermore, a geographical trend was observed with all the marker systems. The group of UK populations and the group of German and Czech populations are located on opposite sides of the plots, while the Dutch populations occupy an intermediate position. Accordingly, Mantel tests indicated a siginificant correlation between geographic and genetic distances for all the marker systems, with the correlation coefficient for SSR (0.323) being smaller than that for AFLP and NBSp (0.637 and 0.651, respectively). The German and Czech populations show more overlap in the plots than the other countries.

Table 1. Average gene diversities (H_e) within *Lactuca serriola* populations sampled within the four investigated countries

	Within-population He		
Country	SSR	AFLP	NBSp
United Kingdom	0.05 (0.05)	0.01 (0.01)	0.01 (0.01)
The Netherlands	0.25 (0.19)	0.05 (0.02)	0.08 (0.05)
Germany	0.46 (0.14)	0.08 (0.02)	0.14 (0.05)
Czech Republic	0.37 (0.17)	0.08 (0.03)	0.11 (0.05)
Overall	0.32 (0.21)	0.05 (0.04)	0.09 (0.07)

SSR, simple sequence repeat; NBSp, nucleotide-binding site-directed profiling.
Standard deviations of H_e-values are given between brackets.

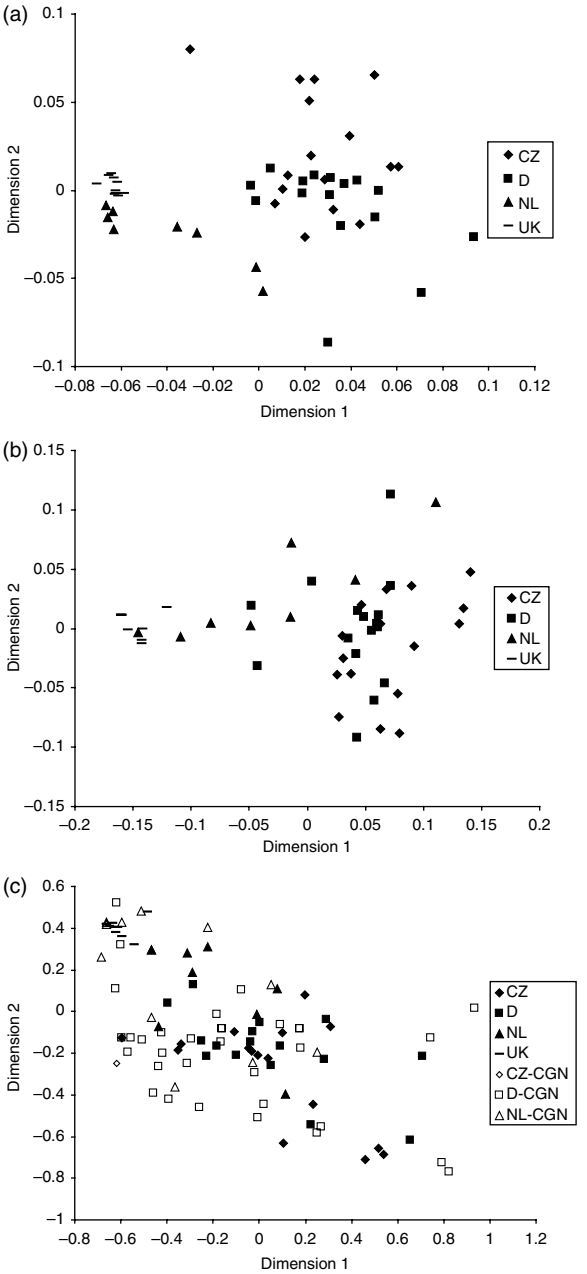

Fig. 1. PCO plots of *Lactuca serriola* populations from the NW European study area based on AFLP (a), NBSp (b) and SSRs (c). For AFLP and NBSp, data are presented for the newly sampled field populations, while for SSR, the CGN accessions from the same area are also included (indicated by CGN in the key). CZ, Czech Republic; D, Germany; NL, the Netherlands and UK, the United Kingdom.

Comparison between marker systems

DR values for the three marker systems were 0.738 for the 173 AFLP markers, 0.762 for the 91 NBSp markers and 0.646 for the 10 SSR markers, respectively. It concerned the same highly structured material, and thus the influence of population structure can be considered to be similar for each of the DR estimates. This would mean that in order to reach a DR value of 0.762 as observed for the 91 NBSp markers, 17.6 SSR markers or 196.4 AFLP markers would be needed. This implied that the informativeness per AFLP data point is lower compared with the other dominant multilocus marker system, NBSp. Due to the higher degree of allelic diversity for co-dominant SSR loci, the level of informativeness per locus was highest for SSR. However, because of the low number of markers studied, the overall DR value was lower than the levels observed for both AFLP and NBSp, respectively.

In principle, NBSp and AFLP target different types of variation, i.e. in and around disease resistance gene (analogues) of the NBS/LRR type for NBSp and in principle random molecular genetic variation for AFLP. Therefore, it was examined whether marked differences could be found between AFLP and NBSp in their estimates of genetic diversity. For SSR, the different, co-dominant scoring system and the limited number of loci used in the analysis (nine) did not allow a comparison with NBSp. Using the same correlation method as the DR between AFLP and NBSp, a value of 0.514 was found, which was lower than the values found above within AFLP and NBSp (0.738 and 0.762, respectively). For a more detailed comparison of AFLP and NBSp, the frequency distributions of the level of population differentiation (G_{st}) and gene diversity (H_e) values across loci for all populations taken together were compared. A Kolmogorov–Smirnov test showed the maximum difference between the cumulative distributions, D, to be 0.299 for G_{st} and 0.266 for H_e, which pointed to a significant difference between AFLP and NBSp results in both cases ($P = 0.000$). A graphical overview of these distributions (Supplementary Fig. S1, available online only at www.journals.cambridge.org) showed the following differences between AFLP and NBSp: (1) with regard to G_{st} values, AFLP displayed a relatively high percentage of loci showing low differentiation between populations compared with NBSp (Fig. S1(a), available online only at www.journals.cambridge.org); (2) with regard to gene diversity, NBSp showed a relatively high number of loci with values between 0.45 and 0.5, representing the highest class of potential H_e values for biallelic loci; both multilocus marker systems furthermore had a preponderance of loci with low H_e (Fig. S1(b), available online only at www.journals.cambridge.org). However, the difference between AFLP and NBSp in the distribution of gene diversities H_e among loci was considerably smaller when gene diversities of individual loci were plotted for each population separately (data not shown): in the latter case, 6.5% of the AFLP and 9.2% of NBSp gene diversities fell in the 0.45–0.5 category. These values were 7.5% for AFLP loci and 19.8% for NBSp loci in the case of all populations taken together (cf. Fig. S1(b), available online only at www.journals.cambridge.org).

Comparison with the CGN collection at the population level

Not all accessions scored in the CGN dataset could be used for a comparison with the newly collected populations. For instance, for accessions obtained from botanical gardens in the study area, information on original sampling site was often lacking and closer examination often showed highly different SSR scores compared with the rest of the populations from the region (data not shown). Therefore, these plants may have originated from outside our present study area. This was corroborated by the finding based on SSR comparisons that several of these accessions appeared to be duplicates of accessions from botanical gardens outside of our study area. This implies that botanical gardens could have obtained their plants by exchange with other botanical gardens instead of directly from field sampling in the close surroundings. After removal of accessions without adequate origin data, 46 accessions (listed in Supplementary Table S1, available online only at www.journals.cambridge.org) from the study area remained for comparison of the SSR variation with the newly sampled populations (Fig. 1(c)). The CGN accessions covered about the same area of the plot as the newly sampled populations, and therefore, the new field populations appeared of little added value in terms of the SSR variation already present in the CGN collection. The CGN accessions also showed a similar geographic trend as observed for the newly collected materials, i.e. with the Dutch populations occupying an intermediate position between the UK populations on the one hand, and the group of German and Czech populations on the other hand. The UK populations did not appear to cover unique variation, as they clustered with some Dutch and a few German accessions.

Comparison with the CGN collection at the level of individual genotypes

All previous analyses were performed at the population level, i.e. a collection of plants sampled at the same location. This may overlook the fact that such populations may consist of different genotypes with limited

genetic interaction because of the autogamous breeding system of *L. serriola*. Indeed, intra-population genotypic variation in the newly sampled material ranged from complete homogeneity to 15 different genotypes among the total of 16 plants sampled per population.

Overall, the CGN collection contained 954 different SSR genotypes of *L. serriola*, of which, according to the passport data, 83 genotypes originated from the same area as the newly collected populations. The newly sampled populations consisted of 184 different genotypes, of which 164 did not match any of those in the CGN collection. Nevertheless, the newly sampled populations contained only three alleles that were not observed in the entire CGN lettuce collection, all at the highly variable locus LsD108. These three alleles represented large PCR products that likely exhibited a higher chance of missing values in the earlier screening of the CGN accessions, and so it could still be held possible that such alleles do occur in the CGN collection. Thus, the new genotypes found in this study did not so much represent new allelic variation, but rather represent new combinations of alleles already present in the CGN collection. Such new combinations could nevertheless be relevant for genebank collections when representing new characteristics relevant for agronomic performance.

The majority of the 184 SSR-based genotypes were represented by only one plant (103). On the other hand, the majority of plants (644) shared their genotype with other plants in the new collection. This varied from two plants sharing a genotype (35 cases) till up to 108 (one case). The distribution of the individual plants across genotypes found is graphically represented in Fig. 2(a). The typical genotype distribution was also reflected in the number of newly sampled populations in which a specific genotype is found (Fig. 2(b)). The great majority of the genotypes (160) were found in a single population only, but also quite a few in several populations, up till 13 and 16 populations, respectively, for the two most common genotypes (represented by 108 and 83 plants, respectively). In the CGN collection, a comparable distribution of the genotypes across individual plants and populations was observed for the same study area (Fig. 2). In this case, the highest number of plants sharing a single SSR genotype was 14. Since a number of the newly sampled populations showed a high level of genetic homogeneity, a preponderance of the genotype category represented by 15 or 16 plants (i.e. the number of plants sampled per population) was observed. This effect was even more pronounced with the CGN accessions for the genotype category represented by five plants per genotype (i.e. the number of plants per accessions screened in the previous study). In Fig. 2(a), this stronger bias explains the linear course of the curve between

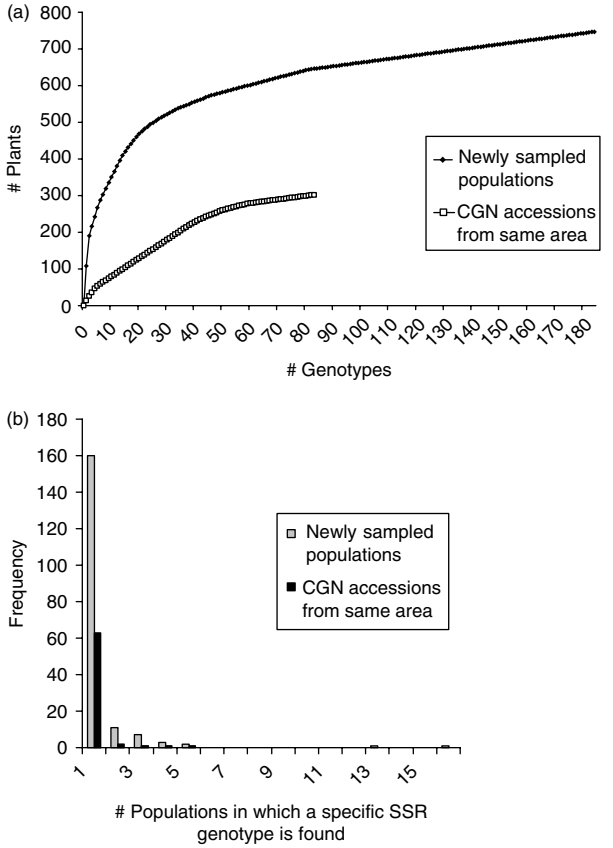

Fig. 2. Distribution of SSR genotypes across plants and populations: (a) cumulative number of plants are plotted against the cumulative number of observed SSR genotypes. On the *X*-axis, the genotypes are ordered according to the number of plants by which they are represented, starting with the category with the highest number of plants per genotype. (b) Frequency distribution of the number of newly sampled populations and the CGN accessions in which a specific SSR genotype was observed.

genotypes 6 and 38 of the CGN material, with each genotype increasing the cumulative number of plants with five.

The geographic distribution of the two most frequent genotypes is shown in Fig. 3(a), in each case combined with very closely related genotypes, that is, genotypes varying by only one repeat unit at a single SSR locus. The most frequently observed one of 108 plants (193 in combination with the related genotypes) was mostly found in the UK, but also on the continent, in the Netherlands and in the CGN accessions originating from the Netherlands, Germany, Belgium and France (additional CGN accessions of Fig. 3 outside the study area are listed in Supplementary Table S1, available online only at www.journals.cambridge.org). Only five UK plants showed more different genotypes. The SSR genotype coming second in frequency (83 plants, together with closely related genotypes, 124, cont83) was widespread

on the continent, mainly in the Central European region, including the CGN collection accessions from the Netherlands and Germany. Figure 3(b) shows the distribution of other frequent (i.e. 16–26 plants) SSR genotypes, mostly also having counterparts in the CGN collection, except for two Czech genotypes (CZ16 and CZ26). However, the number of accessions with a confirmed origin in the Czech Republic was relatively low in the CGN collection.

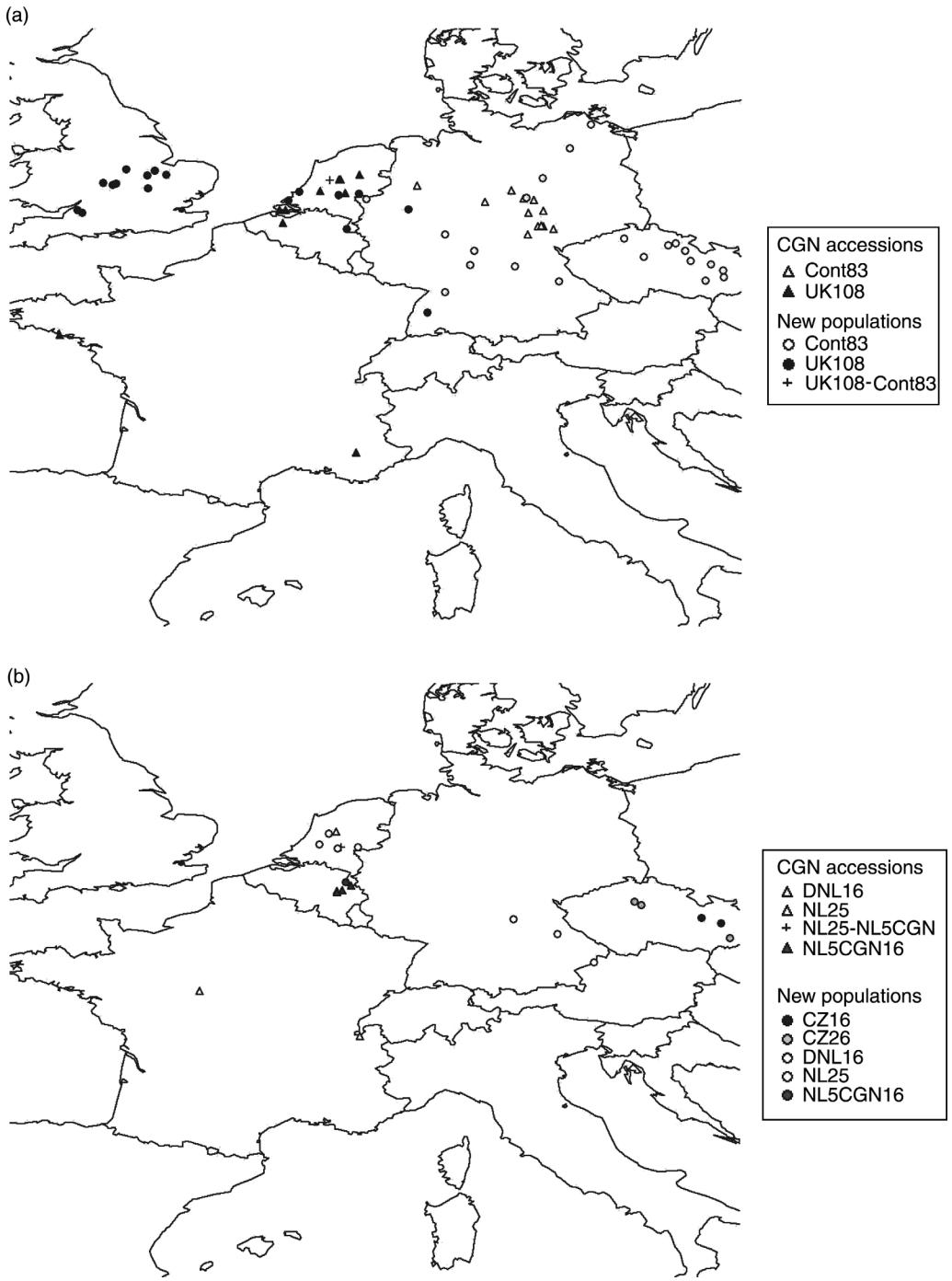

Fig. 3. Geographic locations in Europe of common SSR genotypes in the newly sampled populations and the CGN accessions: (a) the two most frequently observed genotypes, (b) the remaining frequently observed genotypes. Genotypes are denoted by a combination of the region or the CGN collection where they most commonly occur and their respective number of plants in the new collection. CZ, Czech Republic; D, Germany; NL, the Netherlands; UK, the United Kingdom and cont, continental; populations containing more than one genotype are indicated by a cross on the map and a combination of (abbreviated) genotype names.

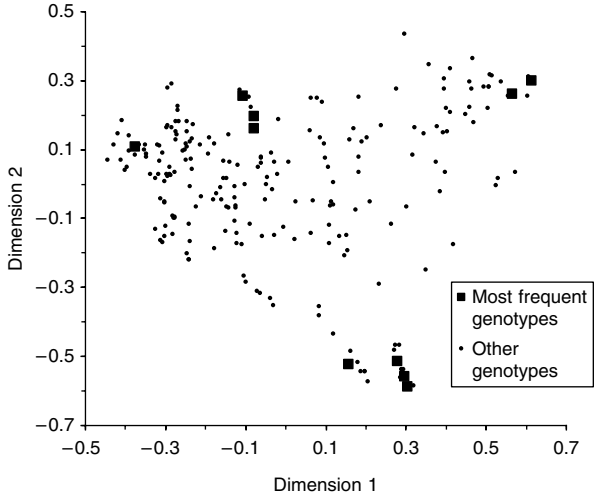

Fig. 4. PCO plot of the SSR genotypes from the newly sampled populations and the CGN collection from the NW European study area.

Some of the other frequent genotypes were widespread, and some were localized to specific regions, for example, the distribution in eastern Belgium of the one genotype found more commonly in the CGN collection, but also in two newly sampled populations from the Netherlands (NL5CGN16). The genetic relationships among the ten most common genotypes (including three very closely related to UK108 and one to cont83), comprising 335 of the total 747 plants in the new collection, and the CGN accessions from the study area are shown in Fig. 4. Interestingly, the most frequent SSR genotypes are positioned at the borders of the PCO plot.

Discussion

Genetic variation was studied in prickly lettuce (*L. serriola*), the closest wild relative of cultivated lettuce, across Europe from the Czech Republic to the UK, using three different molecular marker systems, SSR (microsatellites), AFLP and NBSp. Lettuce is known as a predominantly autogamous species, and this was also clearly observed for *L. serriola* in the low observed heterozygosity and high population differentiation. The most striking observations included the finding of practically a single SSR genotype in the UK, which was also present on the continent, and the finding of a few frequently occurring continental genotypes that comprised a large part of the sampled plants. These frequently occurring genotypes occupied extreme sides of the genetic variation assessed within the study area. Possibly, the remaining samples, occupying intermediate positions, represent inbred descendants from infrequent crosses between the most common genotypes. A similar finding was reported by a fine-scale population genetic analysis of the highly selfing plant species *Medicago truncatula* in a more limited area (Siol *et al.*, 2008). Concerning UK populations, strong founder or bottleneck effects may have occurred during the overseas colonization. It remains to be determined whether such effects have also played a role in the observed preponderance of a few genotypes in the study area, where a large expansion of *L. serriola* has occurred in recent decades (Hooftman *et al.*, 2006; D'Andrea *et al.*, 2009). Furthermore, it cannot be excluded that selection for favourable genotypes has contributed to this phenomenon. In this regard, it is worth noting that several of the more frequent genotypes occurred only in a relatively small area. New angles to the study of wild germplasm genetic diversity, i.e. the comparison of a 'functional' molecular marker system, NBSp, to the classical 'neutral' system of AFLP, and the assessment of the representation in the CGN lettuce collection of *L. serriola* genetic diversity in the field, are discussed in the following sections.

Comparison of marker systems

In addition to the well-established SSR and AFLP, variation was also assessed by NBSp, a novel marker system that targets disease resistance genes of the NBS/LRR gene family, in order to test whether a 'functional' marker system would score another type of genetic variation and even perhaps would give cues to the presence of variation important to disease resistance occurrence. However, even markers generated within or around disease resistance genes may still measure neutral DNA variation to the extent that the sequence variation detected does not influence the phenotype. In the present study, we found some statistical support for NBSp and AFLP representing different types of variation, but overall, the results from the NBSp-based population genetic analysis were not markedly different from those obtained by AFLP. In a comparison of several neutral and targeted marker systems on a set of 80 CGN *Lactuca* accessions, van Treuren and van Hintum (2009) also found no marked differences in diversity between AFLP or NBSp. Likewise, Mantovani *et al.* (2006) obtained similar results in a genetic diversity study comparing SSR, AFLP and NBSp on a broad set of accessions of durum wheat. Also in reconstructing phylogenies of tuber-bearing *Solanum* species, AFLP and NBSp were shown to produce similar results by Wang *et al.* (2008). Nevertheless, NBSp is expected to offer a higher chance of finding individual markers tightly linked to resistance traits, which could be reflected in patterns of variation of such specific markers within populations or across the area sampled. However, the combination of the dominant scoring of

AFLP and NBSp and the selfing habit of *L. serriola*, by which significant influences from founder effects and drift may be expected, makes it difficult to assess indications for selection, or even deviation from neutrality, from population genetic analysis of individual molecular marker scores. The only indications for differences in behaviour between AFLP and NBSp were the following observations: (1) NBSp showed a higher level of intra-population as well as overall variation than AFLP; (2) AFLP produced more bands showing a low population differentiation; (3) NBSp produced more individual bands showing a higher diversity than AFLP. Previous studies on patterns of variation in NBS/LRR disease resistance genes in *Arabidopsis* by Stahl *et al.* (1999) and in lettuce by Kuang *et al.* (2004, 2008) have ascribed their results to the occurrence of balancing (diversifying) selection. It remains to be established whether the higher levels of NBSp diversity observed in the present study are due to balancing (diversifying) selection operating on at least part of the NBSp-generated markers. For instance, when individual marker diversities in our study were compared between AFLP and NBSp per individual population, differences between the two marker systems were much smaller, whereas balancing selection would be expected to operate at the level of individual populations as well. In addition, as indicated by the DR values of the marker systems, the AFLP data overall showed a somewhat lower quality than the NBSp data, and therefore, a direct comparison of gene diversities may not be warranted in this case. In order to investigate more directly whether certain NBSp markers are subject to some form of selection, future analyses need to relate the molecular marker data to the disease resistances scored for *Bremia* and LMV in related studies of the GENE-MINE project.

Implications for plant genetic resources management

The CGN lettuce collection presently (March 2010) comprises 78% of the internationally available *ex situ* accessions of *L. serriola* in the ILDB, which probably is even an underestimation as this percentage did not take into account the accessions duplicated between the collections. Across the total range of *L. serriola*, Europe is relatively overrepresented in the ILDB with 458 accessions, compared with 344 from Asia, 14 from Africa and 16 from America. The latter two continents are also relatively less important for *L. serriola*, although particularly in North Africa (Mediterranean part), additional sampling would be warranted (see also Lebeda *et al.*, 2004, 2007). In general, the genetic variation observed in the newly sampled populations from the present study area in NW Europe appeared to be already well represented in the CGN collection. Nevertheless, several recommendations for the future sampling of *L. serriola* can be derived from the present study. In the first place, sampling should take into account the potential presence of relevant genetic variation within populations, i.e. a number of plants growing in each others' vicinity in the same habitat. However, the amount of intra-population diversity in *L. serriola* appeared to vary widely among populations, which cannot be easily assessed in the field. In the present study, the level of variation ranged from completely homogeneous population samples, where sampling only a single plant per population would have been sufficient, to nearly completely heterogeneous population samples. Whether high levels of population variability can be maintained during maintenance of the samples in genebanks depends on the multiplication procedures. Several genebanks, including CGN, use 16 plants per accession for seed multiplication in autogamous crops, such as *L. serriola*, making loss of variation from highly heterogeneous populations unavoidable (Boukema *et al.*, 1990).

The high frequency of populations harbouring only a limited number of genotypes may imply that sampling from multiple populations, rather than from multiple plants per population, is a more efficient strategy to increase the diversity obtained. The average chance of obtaining a unique genotype within a certain field population was about one third. However, this will depend on the location, as intra-population variation was found to increase towards central Europe. This area also contained the major part of the variation observed in the European outskirts, the UK and the Netherlands. However, despite the observed more or less clinal distribution of diversity across the study area, identical genotypes were observed across large geographic distances. Thus, sampling from multiple remote populations may still result in the introduction of duplicates within collections. This was illustrated by the observation that 335 out of the 747 newly sampled plants (45%), represented just ten frequently occurring SSR genotypes, four of which moreover being very closely related to either of the two most frequently occurring genotypes, and that most of these ten frequently occurring genotypes were also for the most part already represented in the CGN collection. Furthermore, genotypic diversity was not evenly distributed across our study area. For the distribution area in the UK, a few samples could be considered sufficient to cover the total variation. However, this variation was most probably derived from the continent, since identical genotypes were observed in samples and/or accessions from the Netherlands, Belgium and France. On the other hand, the most frequent central European genotype was observed in an area ranging from the Netherlands to

the Czech Republic, and this genotype was also represented by several accessions in the CGN collection. The two examples of frequently occurring genotypes that were not observed in the CGN collection originated from the Czech Republic. Although the Czech Republic is poorly represented with accessions in the CGN collection on the basis of passport data, this was in line with the observed higher level of variation in central Europe, where variation unexplored by genebank collections may still exist. This suggests to direct sampling efforts to Central Europe, when *L. serriola* collections are to be extended with material from the studied area.

Acknowledgements

This work was a part of the EU-funded project GENE-MINE (QLK5-CT-2000-00722). The authors wish to thank Yolanda Noordijk for assistance in running SSR samples, Natalia Rhyzova for assistance in running NBSp samples and Roel Hoekstra for assistance in making the geographic Fig. 3.

References

Boukema IW, Hazekamp Th and van Hintum ThJL (1990) *CGN Collection Reviews: The CGN Lettuce Collection*. Wageningen: Centre for Genetic Resources, The Netherlands (CGN).

Crute IR (1992) The role of resistance breeding in the integrated control of downy mildew (*Bremia lactucae*) in protected lettuce. *Euphytica* 63: 1–2.

D'Andrea L, Broennimann O, Kozlowski G, Guisan A, Morin X, Keller-Senften J and Felber F (2009) Climate change, anthropogenic disturbance and the northward range expansion of *Lactuca serriola* (Asteraceae). *Journal of Biogeography* 36: 1573–1587.

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

Guarino L, Ramanatha Rao V and Reid R (eds) (1995) *Collecting Plant Genetic Diversity: Technical Guidelines*. Wallingford: CABI.

Hooftman DAP, Oostermeijer JGB and Den Nijs JCM (2006) Invasive behaviour of *Lactuca serriola* (Asteraceae) in the Netherlands: spatial distribution and ecological amplitude. *Basic and Applied Ecology* 7: 507–519.

Kuang HH, Woo SS, Meyers BC, Nevo E and Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16: 2870–2894.

Kuang HH, van Eck HJ, Sicard D, Michelmore R and Nevo E (2008) Evolution and genetic population structure of prickly lettuce (*Lactuca serriola*) and its RGC2 resistance gene cluster. *Genetics* 178: 1547–1558.

Lebeda A, Doležalová I and Astley D (2004) Representation of wild *Lactuca* spp. (Asteraceae, Lactuceae) in world genebank collections. *Genetic Resources and Crop Evolution* 51: 167–174.

Lebeda A, Doležalová I, Křístková E, Dehmer KJ, Astley D, van de Wiel CCM and van Treuren R (2007) Acquisition and ecological characterization of *Lactuca serriola* L. germplasm collected in the Czech Republic, Germany, the Netherlands and United Kingdom. *Genetic Resources and Crop Evolution* 54: 555–562.

Lebeda A, Kitner M, Dziechciarková M, Doležalová I, Křístková E and Lindhout P (2009) An insight into the genetic polymorphism among European populations of *Lactuca serriola* assessed by AFLP. *Biochemical Systematics and Ecology* 37: 597–608.

Mantovani P, van der Linden G, Maccaferri M, Sanguineti MC and Tuberosa R (2006) Nucleotide-binding site (NBS) profiling of genetic diversity in durum wheat. *Genome* 49: 1473–1480.

Meudt HM and Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* 12: 106–117.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70: 3321–3323.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics* 5: 63–69.

Shen KA, Chin DB, Arroyo Garcia R, Ochoa OE, Lavelle DO, Wroblewski T, Meyers BC and Michelmore RW (2002) Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. *Molecular Plant Microbe Interactions* 15: 251–261.

Siol M, Prosperi J, Bonnin I and Ronfort J (2008) How multilocus genotypic pattern helps to understand the history of selfing populations: a case study in *Medicago truncatula*. *Heredity* 100: 517–525.

Sretenović Rajičić T, van Hintum T, Lebeda A and Dehmer KJ (2008) Analysis of wild *Lactuca* accessions: conservation and identification of redundancy. *Plant Genetic Resources: Characterization and Utilization* 6: 153–163.

Stahl EA, Dwyer G, Mauricio R, Kreitman M and Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. *Nature* 400: 667–671.

Syed NH, Sorensen AP, Antonise R, van de Wiel CCM, van der Linden CG, van't Westende W, Hooftman DA, den Nijs HC and Flavell AJ (2006) A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *Theoretical and Applied Genetics* 112: 517–527.

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJW, Kesseli RV, Lindhout P, Michelmore RW and Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theoretical and Applied Genetics* 115: 735–746.

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

van der Linden CG, Wouters DCAE, Mihalka V, Kochieva EZ, Smulders MJM and Vosman B (2004) Efficient targeting of plant disease resistance loci using NBS profiling. *Theoretical and Applied Genetics* 109: 384–393.

van Hintum TJL (2003) Molecular characterisation of a lettuce germplasm collection. In: van Hintum TJL, Lebeda A, Pink D and Schut J (eds) *EUCARPIA Leafy Vegetables 2003 Proceedings of the EUCARPIA Meeting on Leafy Vegetables Genetics and Breeding, Noordwijkerhout, Netherlands,*

19–21 March 2003. Wageningen: Centre for Genetic Resources the Netherlands (CGN), pp. 99–104.

van Hintum TJL (2007) Data resolution: a jackknife procedure for determining the consistency of molecular marker datasets. *Theoretical and Applied Genetics* 115: 343–349.

van Treuren R and van Hintum TJL (2009) Comparison of anonymous and targeted molecular markers for the estimation of genetic diversity in *ex situ* conserved *Lactuca*. *Theoretical and Applied Genetics* 119: 1265–1279.

van Treuren R, van Hintum TJL and van de Wiel CCM (2008) Marker-assisted optimization of an expert-based strategy for the acquisition of modern lettuce varieties to improve a genebank collection. *Genetic Resources and Crop Evolution* 55: 319–330.

Wang MQ, van den Berg R, van der Linden CG and Vosman B (2008) The utility of NBS profiling for plant systematics: a first study in tuber-bearing *Solanum* species. *Plant Systematics and Evolution* 276: 137–148.