Theor Appl Genet (2007) 114:609–618
DOI 10.1007/s00122-006-0461-5

ORIGINAL PAPER

# Simple sequence repeats reveal uneven distribution of genetic diversity in chloroplast genomes of *Brassica oleracea* L. and (*n* = 9) wild relatives

**C. J. Allender · J. Allainguillaume · J. Lynn · G. J. King**

Received: 31 October 2006 / Accepted: 6 November 2006 / Published online: 2 December 2006


**Abstract** Diversity in the chloroplast genome of 171 accessions representing the *Brassica* 'C' (*n* = 9) genome, including domesticated and wild *B. oleracea* and nine inter-fertile related wild species, was investigated using six chloroplast SSR (microsatellite) markers. The lack of diversity detected among 105 cultivated and wild accessions of *B. oleracea* contrasted starkly with that found within its wild relatives. The vast majority of *B. oleracea* accessions shared a single haplotype, whereas as many as six haplotypes were detected in two wild species, *B. villosa* Biv. and *B. cretica* Lam.. The SSRs proved to be highly polymorphic across haplotypes, with calculated genetic diversity values (*H*) of 0.23–0.87. In total, 23 different haplotypes were detected in C genome species, with an additional five haplotypes detected in *B. rapa* L. (A genome *n* = 10) and another in *B. nigra* L. (B genome, *n* = 8). The low chloroplast diversity of *B. oleracea* is not suggestive of multiple domestication events. The predominant *B. oleracea* haplotype was also common in *B. incana* Ten. and present in low frequencies in *B. villosa*, *B. macrocarpa* Guss, *B. rupestris* Raf. and *B. cretica*. The chloroplast SSRs reveal a wealth of diversity within wild *Brassica* species that will facilitate further evolutionary and phylogeographic studies of this important crop genus.

Communicated by H. C. Becker.

**Electronic supplementary material** Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0461-2 and is accessible for authorized users.

C. J. Allender (✉) · J. Lynn
Warwick HRI, Wellesbourne, Warwick CV35 9EF, UK
e-mail: charlotte.allender@warwick.ac.uk

J. Allainguillaume
School of Biological Sciences, Plant Science Laboratories,
University of Reading, Whiteknights,
Reading RG6 6AS, UK

G. J. King
Rothamsted Research, Harpenden,
Hertfordshire AL5 2JQ, UK

## Introduction

The genus *Brassica* contains a number of wild and cultivated species, including six crop species that yield various types of vegetables, oilseeds and mustards of global economic significance. The genus is characterised by diploid genomes with type species *B. rapa* (A genome, *n* = 10), *B. nigra* (B genome, *n* = 8) and *B. oleracea* (C genome, *n* = 9) that have hybridised to form the amphidiploids *B. juncea* L. (AB), *B. napus* L. (AC) and *B. carinata* Braun (BC) (UN 1935). The diploid species are themselves the result of a polyploidisation event that took place 7.9–14.6 Mya (Lysak et al. 2005). The A and C genomes diverged less than 4 Mya (Inaba and Nishio 2002). Wild populations exist for most species, and the C genome is also represented by a number of related species with distinct eco-geographic centres of diversity, mostly of Mediterranean origin (Snogerup et al. 1990). Crosses between and among *B. oleracea* and C genome relatives are known to produce fertile or semi-fertile offspring (Kianian and Quiros 1992; Gómez-Campo 1999), making taxonomic assignment challenging. Indeed, it has been suggested that most of the C genome (*n* = 9) species would be better described as a single species, with sub-

specific status awarded to the populations currently classified as different species (Gladis and Hammer 2001). Another taxonomic approach was described by Harberd (1976) who used the term 'cytodeme' to describe a monophyletic group of taxa that share the same chromosome complement and karyotype. Given the morphological variation exhibited by *B. oleracea* as a crop (including the crops cabbage, kale, cauliflower, broccoli, Brussels sprouts and kohl rabi) and the fact that crops usually represent only a fraction of their ancestral genepool, it seems reasonable to expect that the inter-fertile but geographically isolated wild populations related C genome species represent a large resource of genetic diversity that could be utilised in crop improvement programmes.

The genetic diversity and evolutionary relationships of among C genome species have been well studied, based on eco-geographic, phenotypic and genotypic information (see Gómez-Campo 1999). However, it is difficult to make comparisons between molecular studies, as different genetic marker systems have been used on different populations and accessions of variable or loosely defined provenance. Both nuclear and organelle based molecular markers have been used to generate genotypic datasets. Song et al. (1990) used restriction fragment length polymorphism (RFLP) analysis to compare *B. oleracea* and nine C genome wild species with *B. rapa* and found that *B. oleracea* formed a paralogous clade with its wild relatives. However, studies using other marker systems suggest different evolutionary relationships. Dendrograms based on randomly amplified polymorphic DNA (RAPD) markers (Lázaro and Aguinagalde 1998a) and isozyme data (Lázaro and Aguinagalde 1998b) indicated that *B. oleracea* clustered with species such as *B. montana* and *B. incana*. In contrast, Tatout et al. (1999) used SINE transposons as markers and found that *B. oleracea* and *B. incana* were more similar to species such as *B. hilarionis*. An early mutational analysis of the entire *Brassica* chloroplast genome using restriction enzymes focused on the six major crop species (Palmer et al. 1983). A more detailed investigation of the relationships within the *B. oleracea* cytodeme was carried out by Lannér (1998) who used non-coding sequence from the chloroplast genome to examine diversity and relationships between *B. oleracea* and nine C genome species.

Data from the chloroplast genome differ from those provided by the nuclear genome in several respects. The chloroplast genome is non-recombining and inherited from only one parent (normally maternally in angiosperms). A comparison of nuclear and chloroplast data for a given population can therefore allow the influence of seed and pollen dispersal on observed population structure to be inferred (McCauley 1995). However, the overall nucleotide mutation rate of the chloroplast genome is relatively low, with silent substitution rates being half those found in nuclear DNA (Wolfe et al. 1987), and so gene sequence or RFLP markers do not generally detect high levels of intraspecific polymorphism. Chloroplast simple sequence repeats (SSRs), consisting of stretches of mononucleotide repeats have proved to be useful tools for studies of population structure and introgression due to their highly mutable and hence polymorphic nature (Provan et al. 2001). They have been used to investigate the maternal origin of major crop species such as soybean (Xu et al. 2002) and the phylogeography of Norway spruce (Vendramin et al. 2000). In contrast to chloroplasts, plant mitochondrial DNA is characterised by low levels of nucleotide substitution but high levels of changes in gene content and order, together with frequent recombination of mitochondrial genomes within individuals (Avise 1994). Wolfe et al. (1987) calculated that the silent mutation rate in chloroplast DNA is three times that of plant mitochondrial DNA. These characteristics generally render plant mitochondrial DNA unsuitable for ecological genetic and phylogeographic studies.

The aim of this study was to assess inter- and intraspecific diversity within the chloroplast genome of C genome ($n = 9$) *Brassica* crop species and wild relatives and to explore the utility of chloroplast SSRs as markers for diversity and phylogeographic studies. The germplasm accessions used in this investigation represent the founding parents (a mixture of fixed and non-fixed single lines from selected accessions) of two reference *Brassica* diversity core collections [diversity foundation sets (DFSs), see King et al. http://www.*Brassica*.info/diversity/diversity_sets.htm] under development at Warwick HRI. These collections are designed to represent allelic diversity among crop types (in the case of *B. oleracea*) and wild taxa (the $n = 9$ C genome species).

## Methods

### Plant material and DNA extraction

Eighty accessions were selected to provide a representative sampling of the diversity within the crop species *B. oleracea* (Table S1, online supplementary) and 91 representatives of ten wild C genome species were selected. Three *B. nigra*, and six accessions of *B. rapa* which represent the range of haplotype diversity

within 78 accessions (data not shown) of this species were included for comparison. Figure 1 indicates the original collecting sites for the 63 wild C genome accessions for which data were available. Species identity was confirmed for mature plants according to the descriptions given by Snogerup et al. (1990) or by flow cytometry to determine genome size (carried out by Plant Cytometry Services Ltd, The Netherlands). Where the original species description was found to be incorrect, accessions have been analysed on the basis of their confirmed morphological identity (Table S1, online supplementary).

Seeds for the accessions tested were obtained from genetic resource collections both in the public domain and from within Warwick HRI. DNA was extracted from the first true leaf of a single plant from each accession using a Qiagen DNEasy 96 kit. In some cases the seed was allowed to germinate on filter paper and DNA was extracted from the whole seedling. The seed obtained for a small number of accessions was not viable. In these cases DNA was extracted from a whole single seed after imbibing the seed on damp filter paper for 24 h. A single plant from each accession was sampled in order to maximise the number of accessions surveyed whilst enabling the exact allelic composition of each haplotype to be identified. A lack of within accession diversity was detected when three plants from each of five additional *B. montana* accessions were tested using the chloroplast SSRs, with only a single haplotype detected in each accession (data not shown).

## Chloroplast SSR analysis

Primers were designed either directly from *Arabidopsis thaliana* chloroplast genome sequence (Genbank accession AP000423) or from *B. napus* chloroplast sequence obtained using primers designed from

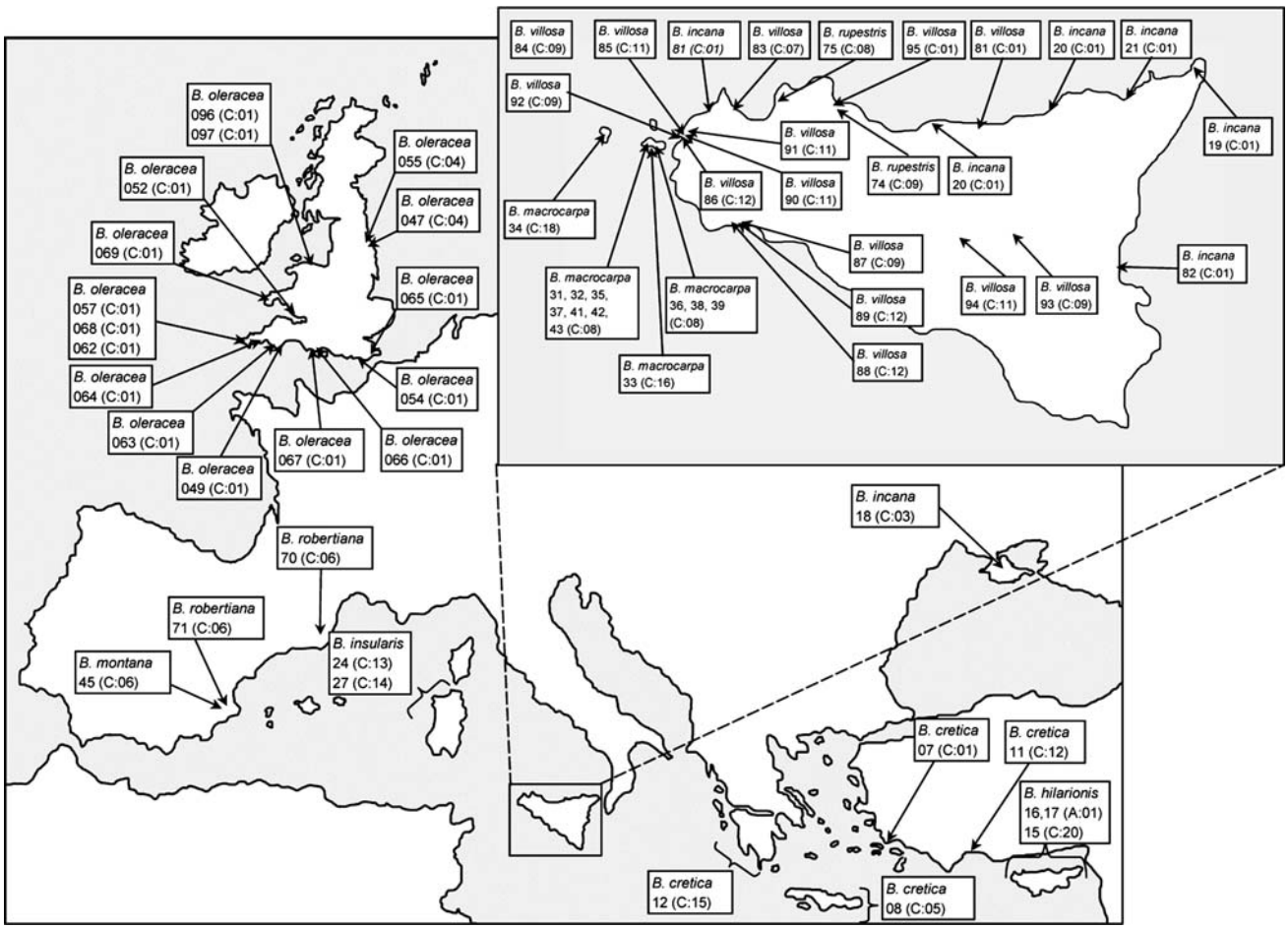


**Fig. 1** Sketch map of Mediterranean region indicating the original collecting location of *Brassica* C genome accessions for which such data were available. These data were obtained from the Genetic Resource Unit at Warwick HRI or from records maintained with research seed stocks. *Numbers* refer to the accession number assigned to each of the BCgDFS accessions. The chloroplast SSR haplotype is indicated in *brackets*

*A. thaliana*. Primer sequences are given in Table 1. Six mononucleotide chloroplast SSRs were amplified from a one tenth dilution of the genomic DNA extract. A Qiagen Multiplex PCR kit was used for amplification and reactions were carried out in 5 µl volumes containing 0.1 µM each primer, 2.5 µl Qiagen master mix and 1 to 10 ng DNA. PCR products were diluted to a ratio of 1 in 30 before loading on an ABI 3100 DNA sequencer (Applied Biosystems) as per manufacturer's instructions. Genescan ROX 500 internal size standard (Applied Biosytstems) was also added to each sample. Trace files from the sequencer were then scored using GeneMapper v3.5 software (Applied Biosystems). Comparing the traces yielded by the same 95 samples on two different occasions showed that the primers produced consistent, robust results (data not shown).

### Diversity analysis

A comparison of both locus and species diversity was made by calculating the diversity index *H* (Nei 1973) for each SSR (across all samples) and within each species (for those that were represented by five or more accessions) according to the following formula:

$$H = 1 - \Sigma p_i^2$$

where $p_i$ is the frequency of the *i*th allele. *H* for each SSR was calculated both from all C genome individuals and from the 23 unique haplotypes detected in the C genome species.

The relatedness of the haplotypes was investigated by considering the number of mutational steps between them to produce a network diagram (Fig. 3). For each pair of haplotypes, the SSR loci at which the haplotypes differed were counted. A network diagram was constructed, in which pairs of haplotypes differing at only one SSR were connected with a solid line. Where pairs of haplotypes which differed at two SSRs were not already implied by this diagram, they were added to it using dashed lines. Relationships between more distant pairs of haplotypes were not shown in order to minimise visual complexity.

## Results

The chloroplast SSRs revealed a wide range of diversity within the *Brassica* C genome genepool. The number of different alleles present at each of the chloroplast SSR loci ranged from two to ten. In total, 23 different haplotypes (Table 2) were detected among the 171 C genome accessions tested. *H* values for each SSR calculated across all C genome individuals were markedly lower (0.11–0.50) than those calculated across the set of unique C genome haplotypes (0.23–0.87), mainly due to the large number of accessions with the C:01 haplotype (Fig. 2). Some haplotypes included null alleles at one or two of the SSR loci—these nulls were counted as an allelic state rather than missing information.

The level of diversity varied amongst the different *Brassica* C genome species. *H* values were calculated based on the number of different haplotypes detected within a species (Table 3). It is apparent that the range of *H* values is not solely related to the number of accessions sampled per species, since the most extensively sampled species (*B. oleracea*, 105 accessions) was the least diverse. The most diverse species (with six haplotypes) represented within the accessions sampled were *B. cretica* and *B. villosa*.

The distribution of chloroplast haplotypes across the *Brassica* C genome genepool (Fig. 2) reveals a considerable degree of overlap between species. This was particularly evident with the most common *B. oleracea* haplotype (C:01). The three *B. nigra* (B genome) accessions sampled shared a common distinct haplotype (B:01). Five of the *B. rapa* haplotypes (A:02, A:03, A:04, A:05 and A:06) were unique and not found in any C genome taxa. They were, however, similar enough to other haplotypes to be included within the network diagram (Fig. 3). The haplotype detected in three *B. hilarionis* accessions is identical to the *B. rapa* haplotype A:01. The geographic distribution of the haplotypes (for wild C genome accessions where the original collection location is known) is indicated on Fig. 1. It is apparent that different haplotypes are found in different zones within the Mediterranean region, reflecting the individual geographical ranges for

**Table 1** Origin and primer sequences of the chloroplast SSRs used in this study

| Name | Forward primer | Reverse primer | Species of origin |
|---|---|---|---|
| Chla16 | aagtatgggaaacttaatgg | gaccattccaatgctccttt | *A. thaliana* |
| Chloro35 | tggaaaaggggagttgtccg | gaatattactctcgaacagac | *A. thaliana* |
| Chloro39 | catgaattagtaactgcatcc | tcctattcatggggattccg | *A. thaliana* |
| ChloroP | ccccaaggttggttgattag | cgaatcgcacgtagagatat | *B. napus* |
| ChloroO | aactttcttcacttcatatg | gtactggaagacaatcacta | *B. napus* |
| ChloroQ | ctgcccaagcctggctgagg | ggccactccactctatttcg | *B. napus* |

the species tested. However, this may represent stochastic evolutionary processes in isolated populations rather than selection for particular haplotypes in different eco-geographic regions.

The network diagram based on identity at each of the six chloroplast SSR loci revealed a complex pattern of relationships between the haplotypes (Fig. 3). Some conspecific clustering of haplotypes is present, for example those found in *B. oleracea* and *B. villosa*. In contrast, the haplotypes detected in the *B. cretica* accessions tested did not cluster at all and were distributed widely across the network. Four of the haplotypes differed from each other at more than two of the SSR loci and could therefore not be included within the network diagram. The haplotypes from *B. rapa* are integral to the network, linking the haplotypes found in *B. oleracea* to those found in *B. villosa*.

## Discussion

Chloroplast SSR markers have revealed a contrasting range of diversity between *B. oleracea* and other C genome species. One striking feature of the distribution of variation is the relative lack of chloroplast diversity found within *B. oleracea*, even though this was the most intensively sampled species in the study. The low calculated *H* value of 0.07 reflects the fact that out of the 105 accessions sampled, only four different haplotypes were detected, and moreover three of these were found in just four accessions. All but 3 of the 25 wild UK *B. oleracea* accessions exhibited the same haplotype (C:01) found in the cultivated samples. Panda et al. (2003) also found no variation in chloroplast PCR-RFLP profiles between a cultivated accession and nine wild populations from Spain, France and the UK.

**Table 2** Haplotypes and allele sizes detected in C genome species. *H* was calculated based on the alleles and haplotypes present in the C genome. Haplotypes detected in *B. rapa* and *B. nigra* are included at the bottom of the table for comparison

| Haplotype | Allele size | | | | | |
|---|---|---|---|---|---|---|
| | Chla16 | ChloroP | Chloro35 | Chloro39 | ChloroO | ChloroQ |
| A:01 | 113 | 265 | 84 | 84 | 89 | 150 |
| C:01 | 114 | 271 | 83 | 84 | 89 | 148 |
| C:02 | 114 | 270 | 83 | 84 | 89 | 148 |
| C:03 | 114 | 272 | 83 | 84 | 89 | 148 |
| C:04 | 114 | 266 | 84 | 84 | 89 | 148 |
| C:05 | 113 | 266 | 84 | 84 | 89 | 148 |
| C:06 | 113 | 264 | 84 | 84 | 91 | 148 |
| C:07 | 111 | 264 | 84 | 84 | 91 | 148 |
| C:08 | 111 | 264 | 84 | 84 | 91 | 154 |
| C:09 | 111 | 264 | 84 | 84 | 91 | 155 |
| C:10 | 110 | 264 | 84 | 84 | 91 | 155 |
| C:11 | 111 | 263 | 84 | 84 | 91 | 155 |
| C:12 | 111 | 264 | 84 | 85 | 91 | 155 |
| C:13 | 113 | Null | 84 | 84 | 91 | 147 |
| C:14 | 113 | Null | 84 | 84 | 90 | 147 |
| C:15 | 114 | 268 | 84 | 84 | 89 | 144 |
| C:16 | 114 | 268 | 84 | 83 | 89 | 144 |
| C:17 | 114 | 268 | 84 | 83 | Null | 144 |
| C:18 | 115 | 262 | 84 | 83 | 89 | 144 |
| C:19 | 112 | 265 | 84 | 83 | 89 | 138 |
| C:20 | 114 | 269 | 84 | 85 | 89 | 150 |
| C:21 | 115 | 271 | 84 | 84 | 90 | 144 |
| C:22 | 114 | Null | 84 | 84 | 100 | 140 |
| Total no. of alleles | 6 | 11 | 2 | 3 | 5 | 7 |
| *H* (all individuals) | 0.41 | 0.50 | 0.42 | 0.11 | 0.35 | 0.41 |
| *H* (haplotypes) | 0.74 | 0.87 | 0.23 | 0.42 | 0.64 | 0.81 |
| A:02 (A genome) | 113 | 264 | 84 | 84 | 89 | 150 |
| A:03 (A genome) | 113 | 264 | 84 | 84 | 89 | 148 |
| A:04 (A genome) | 113 | 265 | 84 | 84 | 89 | 148 |
| A:05 (A genome) | 113 | 265 | 84 | 83 | 89 | 148 |
| A:06 (A genome) | 113 | 272 | 84 | 83 | 91 | 148 |
| B:01 (B genome) | 116 | 263 | 83 | 83 | Null | 123 |

| Haplotype | Species | n |
|---|---|---|
| A:01 | *B. hilarionis* | 3 |
| | (*B. rapa*) | (1) |
| C:01 | *B. oleracea* | 101 |
| | *B. villosa* | 1 |
| | *B. rupestris* | 1 |
| | *B. macrocarpa* | 1 |
| | *B. incana* | 10 |
| | *B. cretica* | 1 |
| C:11 | *B. oleracea* | 1 |
| | *B. villosa* | 4 |
| C:08 | *B. rupestris* | 1 |
| | *B. macrocarpa* | 12 |
| C:09 | *B. villosa* | 3 |
| | *B. rupestris* | 2 |

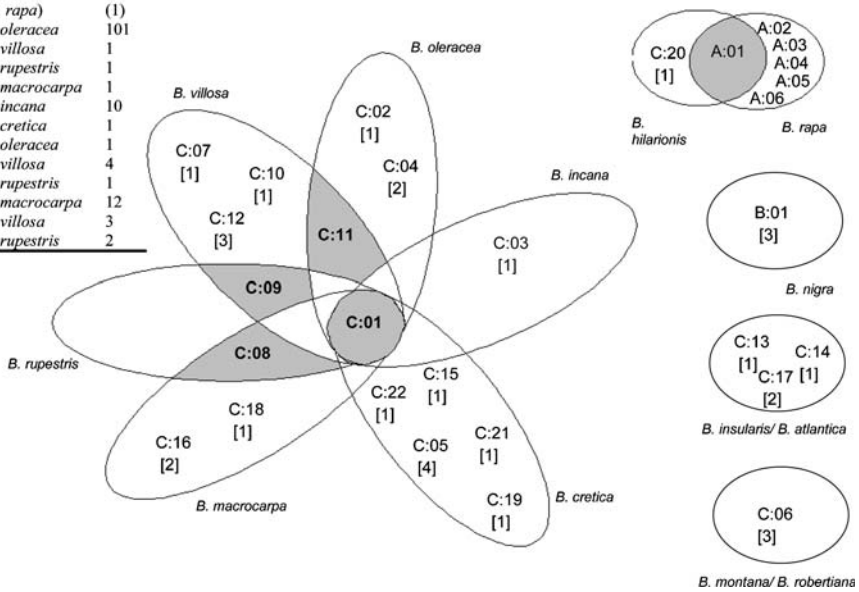


**Fig. 2** Distribution of haplotypes among the *Brassica* species tested. *Shaded areas* indicate where a haplotype is shared between two or more species. The frequency of haplotype occurrence per species is indicated in *square brackets* and that for the shared haplotypes is shown in the table

The contrast in chloroplast diversity between *B. oleracea* and the other C genome species may be at least partially explained by the process of domestication. Gómez-Campo and Prakash (1999) postulate that *B. oleracea* was first cultivated as kale originating from wild populations on European Atlantic coasts. These early kale cultivars were then likely to have been introduced to the Eastern Mediterranean region around 3,000 to 4,000 years ago, where a massive diversification in crop type subsequently occurred. However, there is scant reliable historical or biological evidence of the nature and location of initial *B. oleracea* domestication. Thompson (1979) suggested a similar series of events, and that a shift towards cold tolerance and a biennial growth habit occurred as the cultivation of *B. oleracea* proceeded northwards. The status of the wild populations around the British coast was investigated by Mitchell (1976) who concluded that they mostly originated as escapes from cultivation, based on historical evidence from regional floras and geographic locations. This would explain the lack of chloroplast diversity in the UK wild accessions tested in the current study. Gómez-Campo and Prakash (1999) also noted that the populations of wild kales found in the Mediterranean region may be escapes from cultivation. It therefore appears unclear whether any truly wild population of *B. oleracea* remains, as the phylogeography of the species has been blurred by introgression and escape of domesticated forms. Two of the wild accessions which possess a different haplotype originate from the north eastern coast of England. This haplotype (C:04) appears to be intermediate between the other *B. oleracea* haplotypes and those found in other species (Fig. 3), and thus may represent a phylogenetic link between *B. oleracea* and the other C genome species studied here. It may be possible to substantiate this through further investigations based on a combination of nuclear and plastid genome markers, together with sequence comparisons.

The degree of intra-specific similarity between haplotypes varied widely. All of the six *B. villosa* haplotypes differed from each other by only a small number of mutational steps (Fig. 3). In contrast, *B. cretica* has a diverse aggregation of haplotypes which show very little similarity to each other. Snogerup et al. (1990) described several sub-species of *B. cretica* which may explain the pattern of diversity seen here, although the accessions used were not classified in this way. Other authors have also noted that *B. cretica* is an extremely diverse species, with genetic distances between populations of sub-species, calculated from nuclear molecular markers, often exceeding that found between species in the *B. villosa*/*B. rupestris*/*B. macrocarpa* group (Lázaro and Aguinagalde 1998b). Figure 1 reveals the contrasting geographic origins for *B. cretica* (dispersed Aegean islands and Turkey) and members of the *B. rupestris*/*B. villosa* group (restricted to Sicily). *B. cretica* populations may have been heavily influenced by 'isolation by distance' processes leading to a high level of diversity within the species as a whole. The specific/sub-specific status of the various *B. villosa* populations has also been in dispute. No subspecies were recognised by Snogerup et al. (1990), although such classifications were applied to the material

**Table 3** Details of the accessions used in this study

| Species | Sub-classification | No. of accessions tested | *H* |
|---|---|---|---|
| *B. atlantica*[a](Coss) Schultz | – | 2 | – |
| *B. bourgaei* (Webb) Kuntze | – | 1 | – |
| *B. cretica* Lam. | – | 9 | 0.74 |
| *B. hilarionis* Post | – | 4 | – |
| *B. incana* Ten. | – | 11 | 0.17 |
| *B. insularis* Morris | – | 2 | – |
| *B. macrocarpa* Guss. | – | 16 | 0.41 |
| *B. montana* Pourret | – | 1 | |
| *B. oleracea* | var. *acephala* DC. | 11 | – |
| *B. oleracea* | ssp. *alboglabra* Bailey | 4 | – |
| *B. oleracea* | var. *botrytis* L. | 21 | – |
| *B. oleracea* | var. *capitata* L. | 15 | – |
| *B. oleracea* | var. *gemmifera* DC. | 8 | – |
| *B. oleracea* | var. *gongylodes* L. | 3 | – |
| *B. oleracea* | var. *italica* Plenck | 14 | – |
| *B. oleracea* | var. *tronchuda* Bailey | 4 | – |
| *B. oleracea* | Wild | 24 | – |
| *B. oleracea* total | | 105 | 0.07 |
| *B. robertiana*[b] Gay | – | 2 | – |
| *B. rupestris* Raf. | – | 4 | – |
| *B. villosa* Biv. | – | 2 | – |
| *B. villosa* | ssp. *bivoniana* | 6 | – |
| *B. villosa* | ssp. *drepanensis* | 3 | – |
| *B. villosa* | ssp. *tinei* | 2 | – |
| *B. villosa* | ssp. *villosa* | 1 | – |
| *B. villosa* total | | 13 | 0.78 |
| *B. rapa* (A genome) | Various | 6 | – |
| *B. nigra* (B genome) | – | 3 | – |

*H* was calculated only for those species where five or more accessions were sampled. *H* was not calculated for *B. rapa* as the accessions were selected to represent maximum diversity

[a] *B. atlantica* is a non-accepted synonym for *B. insularis* (Warwick et al. 2000)

[b] *B. robertiana* is a non-accepted synonym for *B. montana* (Warwick et al. 2000)

used in the current study by the original collectors. It is not within the scope of this study to support or reject taxonomic revisions, as morphological and nuclear marker data would be required.

The C:01 haplotype is found in four other species (Fig. 2). Ten of the eleven *B. incana* accessions possess the C:01 haplotype, in addition to one each of *B. villosa*, *B. macrocarpa* and *B. cretica*. There are three possible explanations for this. The accessions concerned could be misclassified or contaminated, the C:01 haplotype was present as a common ancestor of the five species concerned, or post-speciation introgression has occurred. For certain species, one or the other of these scenarios seems more likely to be correct. For example, of the 11 *B. incana* accessions tested, 10 possessed the C:01 haplotype. It is unlikely that all ten accessions could be misclassified, since the species identity was verified for this study, and so an ancestral origin or introgression are plausible explanations. Indeed, several previous studies (Lázaro and Aguinagalde 1998a, b; Tatout et al. 1999) have shown *B. oleracea* and *B. incana* to be closely related. Introgression from *B. oleracea* is certainly a possibility across most of the range of *B. incana* (Snogerup et al. 1990). However, if the two species share the same chloroplast lineage this would be difficult to detect without further data from nuclear markers. Interestingly, the one accession of *B. incana* from outside the Mediterranean region has a divergent haplotype (C:03), and this population was thought to be the result of introgression (Snogergup et al. 1990), possibly from cultivated forms. In the case of *B. cretica*, only a single accession from the nine tested possessed the C:01 haplotype, suggesting introgression might be a more likely explanation. *B. oleracea* is a very widespread crop species and crop to wild species gene flow may account for the widespread distribution of this haplotype.

In general, the overlapping haplotype distribution seen in Fig. 2 supports the patterns of clustering previously reported based on molecular and morphological characteristics. *B. macrocarpa*, *B. villosa* and *B. rupestris* cluster together (Snogerup et al. 1990; Tatout et al. 1999; Lázaro and Aguinagalde 1998a, b; Warwick and Black 1991; Geraci et al. 2004). These species are all native to Sicily (*B. villosa* and *B. macrocarpa* are endemic to the island), and so the possibility of co-ancestry and/or recent gene flow should not be surprising. Snogerup et al. (1990) included *B. incana* as part of the *B. macrocarpa*/*B. rupestris* group, although the chloroplast SSRs used in this study revealed no haplotype in common between these species other than C:01. Two or three unique haplotypes were detected in *B. insularis* depending on whether the two *B. atlantica* accessions actually represent *B. insularis*. Lannér (1998) identified three different chloroplast types in four accessions of *B. insularis*. One was identical to a *B. macrocarpa* chloroplast type, whilst the other two were more closely related to *B. oleracea*. The chloroplast SSR haplotypes identified here do not follow the pattern observed by Lannér (1998), although haplotype C:17 (*B. atlantica*) appears to be related to haplotype C:16 (*B. macrocarpa*—see Fig. 3). Lázaro and Aguinagalde (1998b) found that *B. insularis* fell outside of both the *B. oleracea* and *B. macrocarpa* groups based on isozyme data. *B. insularis*, like *B. cretica*, also has a wide distribution including Corsica, Sardinia and coastal areas of Tunisia and Algeria.

The haplotype detected in the four *B. hilarionis* accessions is intriguing as it is identical to a *B. rapa* haplotype. This raises questions about the relationship of the C genome species to A genome *B. rapa*, and the patterns of gene flow between them. In general, the six haplotypes detected in 78 wild and cultivated accessions of *B. rapa* from across its global range are not found in any of the C genome species with the exception of *B. hilarionis*. Natural hybridisation between *B. oleracea* and *B. rapa* that results in a stable allotetraploid appears to be very rare, although it can be achieved in the laboratory through embryo rescue and ovule culture (Song et al. 1993; Abel et al. 2005). Song et al. made synthetic hybrids and found maternal transmission of chloroplast in all but one hybrid plants resulting from 14 crosses. Where biparental transmission of organelles was observed in an $F_1$, the paternal complement was not present in the $F_2$. The conserved segmental nature of A and C genome chromosomes (Parkin et al. 1995), appears to lead to exchange of chromosome segments between the A and C nuclear genomes in natural and synthetic hybrids, as a result of non-reciprocal homoeologous translocations (Parkin et al. 1995; Sharpe et al. 1995). In natural populations this process could lead to progressive loss of one or other diploid genome and effective gene flow of cytoplasm between species. A strong sporophytic self incompatibility (SI) system is in operation in many *Brassica* species, including *B. oleracea* and related C genome species. Such a system would increase the likelihood of new genotypes and chloroplast haplotypes (from dispersal of pollen and seed, respectively) becoming successfully established in populations and act to increase genetic diversity.

The accessions used in this study form part of two diversity core collections that are under construction at Warwick HRI, (see King et al. http://www.*Brassica*.info/diversity/diversity_sets.htm). These DFSs are designed to represent ''an informative set of genetically fixed lines representing a structured sampling of diversity across a genepool'', and are based on founder accessions sourced from ex situ genetic resource collections. They are being developed by genetic 'fixing' in order to provide immortal public domain reference sets that can form the basis of collated genotype and phenotype data. The BolDFS is designed to capture the diversity present within the *B. oleracea* crop genepool, whilst the BCgDFS represents diversity among wild populations of *B. oleracea* and other C genome species. Our finding that the set of accessions derived from *B. oleracea* crop types is almost monomorphic in terms of chloroplast type is consistent with the results of Panda et al. (2003). However, the level and distribution of chloroplast diversity in the wild species is previously unreported. One other study has indicated the presence of multiple chloroplast haplotypes among species; Lannér (1998) found two haplotypes among six accessions of *B. cretica*. In the current study six haplotypes were detected in nine accessions.

Other authors have also found chloroplast SSRs to be highly polymorphic in comparison with other chloroplast markers. Provan et al. (1999a) found only three haplotypes in 245 accessions of *Hordeum vulgare* ssp. *spontaneum* using chloroplast RFLPs, whilst chloro-

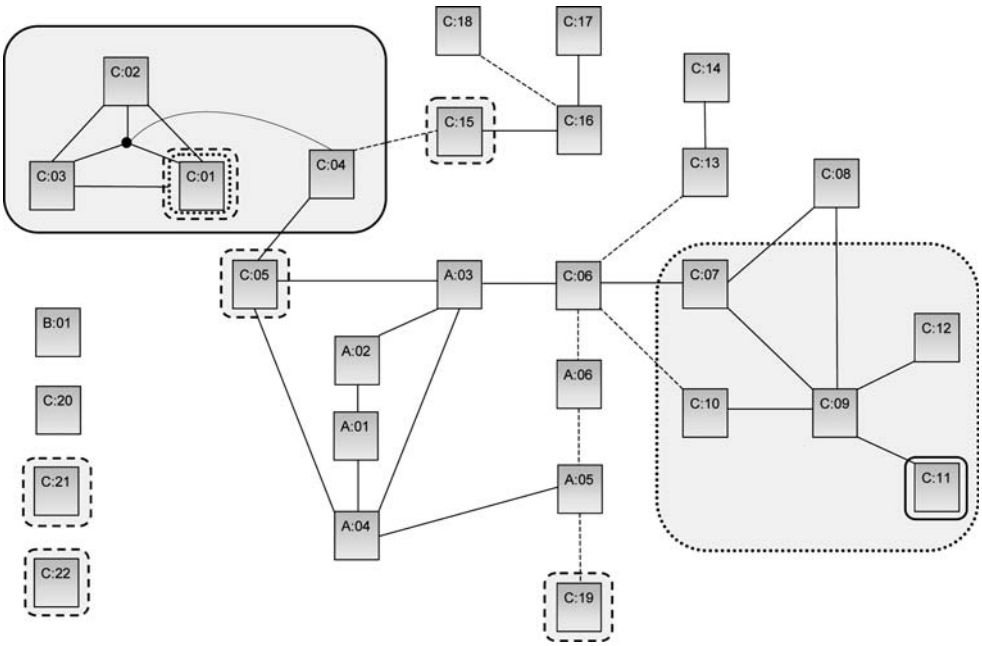


**Fig. 3** Network diagram indicating the relationships among similar haplotypes. Only those haplotypes that differ by allelic state at one (*solid line*) or two (*dashed line*) SSR loci are included in the network. More divergent haplotypes are located in a *column* to the *left* of the network. Haplotypes C:02, C:01 and C:03 differ from each other by one SSR and they are all two SSRs different from haplotype C:04. Haplotypes present in three species are contained within the *shaded areas* as follows: *B. oleracea* (*solid border*), *B. villosa* (*dotted border*), *B. cretica* (*dashed border*)

plast SSRs revealed six distinct haplotypes in just twelve accessions. *B. oleracea* as a crop appears to be somewhat impoverished in terms of chloroplast diversity compared to other major crops, although the number of seed-producing individual plants compared to cereals is very low. Chloroplast SSRs have been recently used to determine a polyphyletic origin for cultivated barley (Molina-Cano et al. 2005), rice (Garris et al. 2005) and the common bean (Chaçon et al. 2005). All of these crops appear to result from multiple domestication events in different geographic regions. In contrast, *B. oleracea* seems to have only a single centre of domestication in the eastern Mediterranean region (Thompson 1979; Gómez-Campo and Prakash 1999), and this appears to be reflected in the low observed chloroplast haplotype diversity. Just one broccoli accession has a haplotype (C:02) that differs from C:01. However, the two haplotypes are very similar and could plausibly have arisen by a single mutation. Whilst in theory this additional haplotype could represent the signature of a second domestication event, given the number of *B. oleracea* accessions sampled one would expect to encounter it at much higher frequencies if different chloroplast lineages were indeed involved in domestication. Provan et al. (1999b) report a maximum mutation rate of chloroplast SSRs in torrey pine (*Pinus torreyana*) of $3.2 \times 10^{-5}$ to $7.9 \times 10^{-5}$ mutations at each locus per generation per individual. If accurate, then this would suggest that the 101 accessions with haplotype C:01 descended from only a small number of common ancestors a few tens of generations ago—an unlikely timescale considering the breadth of accessions sampled. The mutation rate reported by Provan et al. (1999b) represents an upper threshold rate consistent with the monomorphic nature of their samples; therefore, in reality the mutation rate of chloroplast SSRs could be much lower. A lower rate would be more consistent with the lack of diversity detected in cultivated and wild *B. oleracea* accessions in this study.

The six chloroplast SSRs used for this study have proven to be a rapid and highly sensitive method of assaying chloroplast diversity in *Brassica* species. However, the data obtained do not lend themselves to a phylogenetic analysis of the species tested, for two reasons. Firstly, the sampling design was limited to one plant per accession, and so intra-population diversity has not been tested; however, a lack of intra-accession diversity in *Brassica* C genome species was noted by Lannér (1998). If undetected within-population variation exists, this may cause evolutionary relationships to be incorrectly inferred (Tatout et al. 1999). Secondly, whilst the polymorphic nature of chloroplast SSRs makes them particularly useful for investigating intra-species variation and phylogeography, they are more susceptible to convergent mutations and homoplasy than other marker systems such as RFLPs (Provan et al. 2001). Thus, the mutation rate of chloroplast SSRs may be too high to maintain a phylogenetic signal over longer stretches of evolutionary time. Indeed the patterns of alleles in the haplotypes detected in this study (Table 2, haplotypes C:11, C:12, C:20 and C:22) provide some evidence for convergent mutations.

The diversity revealed by chloroplast SSRs among the C genome *Brassica* species is skewed towards the Mediterranean wild species and is apparently almost absent from contemporary UK natural populations of *B. oleracea* itself. This has implications both for the conservation of natural genetic diversity and for the search for novel sources of alleles for crop improvement programmes.

**Acknowledgments** This work was funded by the UK Biotechnology & Biological Sciences Research Council and the Natural Environment Research Council. The authors would like to thank Dave Pink, Dave Astley and Graham Teakle for advice and provision of accessions and Neale Grant, Sandy McClement and Jason Pole for DNA extraction and plant husbandry.

## References

Abel S, Mollers C, Becker HC (2005) Development of synthetic *Brassica napus* lines for the analysis of ''fixed heterosis'' in allopolyploid plants. Euphytica 146:157–163

Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York

Chacon MI, Pickersgill B, Debouck DG (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. Theor Appl Genet 110:432–444

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Geraci A, Chèvre A-M, Divaret I, Eber F, Raimondo FM (2004) Isozyme analysis of genetic diversity in wild Sicilian populations of *Brassica* sect *Brassica* in view of genetic resources management. Genet Resour Crop Evol l51:137–146

Gladis T, Hammer K (2001) Nomenclatural notes on the *Brassica oleracea*-group. Genet Resour Crop Evol 48:7–11

Gómez-Campo C (1999) Taxonomy. In: Biology of *Brassica* Coenospecies. Elsevier, Amsterdam

Gómez-Campo C, Prakash S (1999) Origin and domestication. In Biology of *Brassica* Coenospecies. Elsevier, Amsterdam

Harberd DJ (1976) Cytotaxonomic studies of *Brassica* and related genera. In: Vaughan JG, Macleod AJ, Jones BMG (eds) The biology and chemistry of the cruciferae. Academic, London

Inaba R, Nishio T (2002) Phylogenetic analysis of the *Brassiceae* based on the nucleotide sequences of the *S*-locus related gene *SLR1*. Theor Appl Genet 105:1159–1165

Kianian SF, Quiros CF (1992) Trait inheritance, fertility and genomic relationships of some $n = 9$ *Brassica* species. Genet Resour Crop Evol 39:165–175

Lannér C (1998) Relationships of wild *Brassica* species with chromosome number $2n = 18$, based on the comparison of the DNA sequence of the chloroplast intergenic region between *trnL* (UAA) and *trnF* (GAA). Can J Bot 76:228–237

Lázaro A, Aguinagalde I (1998a) Genetic diversity in *Brassica oleracea* L. (Cruciferae) and wild relatives ($2n = 18$) using RAPD markers. Ann Bot 82:829–833

Lázaro A, Aguinagalde I (1998b) Genetic diversity in *Brassica oleracea* L. (Cruciferae) and wild relatives ($2n = 18$) using isozymes. Ann Bot 82:821–828

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe *Brassiceae*. Gen Res 15:516–525

McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol Evol 10:198–202

Mitchell ND (1976) The status of *Brassica* oleracea L. subsp. *oleracea* (wild cabbage) in the British Isles. Watsonia 11:97–103

Molina-Cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005) Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor Appl Genet 110:613–619

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65:181–189

Panda S, Martín JP, Aguinagalde I (2003) Chloroplast and nuclear DNA studies in a few members of the *Brassica oleracea* L. group using PCR-RFLP and ISSR-PCR markers; a population genetic analysis. Theor Appl Genet 106:1122–1128

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* oilseed rape. Genome 38:1122–1131

Provan J, Russell JR, Booth A, Powell W (1999a) Polymorphic simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. Mol Ecol 8:505–511

Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999b) A low mutation rate for chloroplast microsatellites. Genetics 153:943–947

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147

Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38:1112–1121

Snogerup S, Gustafsson M, von Bothmer R (1990) *Brassica* sect. *Brassica* (*Brassica*ceae) I. Taxonomy and variation. Willdenowia 19:271–365

Song K, Osborn TC, Williams PH (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor Appl Genet 79:497–506

Song K, Tang K, Osborn T C (1993) Development of synthetic *Brassica* amphidiploids by reciprocal hybridization and comparison to natural amphidiploids . Theor Appl Genet 86:811–821

Tatout C, Warwick S, Lenoir A, Deragon J-M (1999) SINE insertions as clade markers for wild crucifer species. Mol Biol Evol 16:1614–1621

Thompson KF (1979) Cabbages, kales etc. Brassica oleracea (Cruciferae). In: Simmonds NW (ed) Evolution of crop plants, Longman Group, London

UN (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jap J Bot 7:389–452

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). Genome 43:68–78

Warwick SI, Black LD (1991) Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassiceae)—chloroplast genome and cytodeme congruence. Theor Appl Genet 82:81–92

Warwick SI, Francis A, La Fleche J (2000) Guide to the wild germplasm of *brassica* and allied crops (tribe Brassiceae, *Brassica*ceae), 2nd edn. http://www.*Brassica*.info/resources/crucifer_genetics/guidewild.htm

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Xu DH, Abe J, Gai JY, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans; evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–653