# 8 Molecular Markers in *Allium*

**M. Klaas[1] and N. Friesen[2]**
[1]*Gotthard Müller Straβe 57, D-70794 Filderstadt-Bernhausen, Germany;* [2]*Botanical Garden of the University of Osnabrück, Albrechtstrasse 29, D-49076 Osnabrück, Germany*



## 1. Introduction: Why Molecular Markers?

Evolution is a two-phase process, in which genetic variability accumulates in a random fashion, after which morphological, biochemical or physiological changes are induced and stabilized by environmental pressure or the plant breeder's efforts (Mayr, 1969). While the evolutionist's interest lies primarily with investigating the forces directing the second set of processes, molecular markers can be used to sample the underlying genetic variability when it is not directly being subjected to the action of evolutionary pressures. This pool of information gives the opportunity to reconstruct the course of the evolutionary process while avoiding the use of phenotypic markers, such as morphological or anatomical features, which may be influenced by the very mechanisms which they are required to elucidate. Consequently, there is an increased utilization of molecular markers in evolutionary and systematic studies. However, for efficiency reasons, the use of molecular markers in these studies depends on pre-existing data, such as taxonomic classification. Increased standardization of the techniques and availability of equipment and expertise have also promoted the application of molecular technologies for other purposes, such as for quick analysis of cytoplasm types, the verification of hybrid plants

and extensive use in the construction of genetic maps. Concurrently, the recent years since Klaas (1998) reviewed the topic have seen continued and widespread application of molecular markers in *Allium* research, including the use of the most technically advanced procedures such as comparative sequencing of DNA markers and the development of amplified fragment length polymorphisms (AFLPs) and microsatellites. While the choice of a particular technique depends on both the biological question to be answered and the available laboratory equipment and expertise, the quality of data and their suitability for a particular study must be judged by the same rigorous standards. This review aims to give an overview of molecular-marker applications in *Allium*, and some judgements about the validity of the approaches taken. We hope that our reflections will be of some help in planning future studies, since there is usually no one single most useful technique. Choices have to be made based on the weighting of different markers' strengths and drawbacks and on the practical options available in a particular laboratory.

For more detailed information, several excellent monographs have been published, which include laboratory protocols of the techniques as well as broader topics, such as laboratory set-up and sampling strategies (Zimmer *et al.*, 1993) and the theory, scope and limits of applications (Hillis *et al.*, 1996). For concise descriptions of selected procedures, see Hoelzel (1992).

## 2. Markers

### 2.1 Isozymes

While isozyme analysis was historically the first application of molecular markers in *Allium*, it still holds some advantages today over the now more widely employed DNA markers in certain applications. For general introductions to the techniques, see May (1992) and Murphy *et al.* (1996). In particular, allozyme analysis, which detects polypeptide variants corresponding to different alleles at one locus, is a very cost-effective means of obtaining Mendelian molecular markers in a short time for large numbers of individuals. The isozyme investigation of large *Allium* collections (*A. sativum*: 300 accessions, Maaß and Klaas, 1995; 110 accessions, Pooler and Simon, 1993; *A. cepa* var. *ascalonicum*: 189 accessions, Arifin and Okubo, 1996; *A. cepa* and *A. fistulosum*: 188 accessions and 29 accessions, respectively, Peffley and Orozco-Castillo, 1987) and a larger study in *A. douglasii* (29 populations each with 30–60 analysed individuals, Rieseberg *et al.*, 1987), with numbers of accessions not yet paralleled in studies based on DNA markers, testify to the strength of the approach.

Briefly, plant tissue is squashed in a suitable buffer that preserves enzyme activity; this solution is applied to a starch gel and electrophoretically separated. Thereafter, protein bands with enzymatic activity are revealed by specific staining reactions. Changes of peptide amino acid sequence which result in altered electrophoretic mobility due to charge, size or conformation differences can be detected. Following separation, horizontal slicing of the gel allows for the scoring of up to three different enzyme systems, using separate staining reactions.

The genetic structures of the major enzyme systems are well characterized (Wendel and Weeden, 1989), so a thorough interpretation of the banding patterns yields Mendelian data that have been shown to correspond well with DNA-marker results (restriction fragment length polymorphism (RFLP): Chase *et al.*, 1991; randomly amplified polymorphic DNA (RAPD): Maaß and Klaas, 1995).

The major limitation of isozyme analysis is the small number (15 or less) of suitable enzyme systems, of which usually only a subset will exhibit sufficient variability. The suitability of isozyme markers for an intended plant study therefore has to be tested in advance. More isozyme alleles can be resolved by technological refinements, such as different gel-matrix pore sizes or different buffer systems, or by differential heat-stability tests (Murphy *et al.*, 1996), but then the advantage of simple and quick application is

often lost. Since fresh plant material is preferable, isozyme analysis is still a good choice if a limited number of high-quality data points are needed for a large number of individuals in a population study or for the genetic characterization of larger living collections of crops. At taxonomic levels higher than species or close species complexes, the assignment of observed bands to homologous loci based on electrophoretic mobility seems dubious.

## 2.2 DNA markers

The use of DNA-based markers avoids detection problems due to uneven expression, which has been a major problem in developing additional isozyme systems. It also allows for the development of basically unlimited numbers of markers, and it enables prolonged storage of samples for later analysis, either as frozen tissue or even as dried material kept at room temperature.

### *2.2.1 RFLP*

Apart from some earlier experiments on direct hybridization of DNAs from different taxa, yielding distance type of data (Werman *et al.*, 1996), the RFLP technique brought the first opportunity for DNA-based molecular markers. Purified DNA is cut by a restriction endonuclease at specific recognition sites, and then the digested DNA is electrophoretically separated according to size. RFLPs detect nucleotide substitution, which results in loss or gain of a recognition site, or insertions/deletions, which lengthen/shorten a specific fragment.

The direct visualization of separated restriction fragments is possible from digested purified chloroplast DNA (cpDNA) (Linne von Berg *et al.*, 1996, in a first DNA-based phylogeny of the genus *Allium*). The approach allows scoring of numerous bands from one gel, with virtually no possibility for contamination to influence the results, but it has been rarely applied, since it depends on the isolation of chloroplasts from fresh leaves prior to DNA extraction. More common is the transfer of the restricted and separated DNA fragments to a membrane, followed by hybridization with a specific labelled DNA probe. The detection of labelled bands on the membrane is much more sensitive than direct visualization, and can be extended by prolonged exposure. The hybridization with specific probes is a precondition for analysing nuclear DNA changes by RFLPs. The hybridization signal is also an indicator of overall sequence similarity, an important information aspect that is missing from polymerase chain reaction (PCR)-generated markers.

While a lot of high-quality data have been generated by RFLPs, their use was replaced to some extent by PCR after wider realization of the potential of this approach. RFLPs require larger quantities of relatively high-quality DNA, which has to be highly purified, since the restriction endonucleases are generally more sensitive to small impurities in the target DNA than the Taq DNA polymerase working at higher temperatures, and today the very same type of data can be generated faster by PCR.

### *2.2.2 PCR-based techniques*

CAPS. Cleaved amplified polymorphic sequences (CAPS) simplify the gathering of RFLP data, avoid the complicated blotting/hybridization procedures of traditional Southern blots and require only small amounts of total genomic DNA. DNA regions known or suspected to contain polymorphisms are amplified from genomic DNA by specific PCR primers, followed by restriction analysis of the purified PCR products. Although the technique is expensive in evolutionary or genetic-diversity studies, because of primer costs, large data sets can be generated in a short time (Mes *et al.*, 1998, 1999; Friesen *et al.*, 1999, evolution within genus *Allium*), and it is an economical substitute for Southern blots in repeated tests for the presence of known polymorphisms. For example, it can be used as an indicator for the presence of certain cytoplasm types (Havey, 1995, identification of *A. cepa* cytoplasms; Dubouzet *et al.*, 1998, verification of *A. giganteum* hybrids).

RAPD. RAPD analysis requires no prior knowledge of the genome investigated and can thus be readily applied to different species (Williams *et al.*, 1990). It has also been applied in numerous *Allium* investigations. A PCR reaction is carried out using a single primer, usually of ten bases, and with purified total genomic DNA as the target. Sequences between primer binding sites within a suitable distance, generally less than 2 kilobases (kb), are amplified and scored for size differences after electrophoresis. Somewhat pointedly, the technique has been likened to practising PCR without a clue (see also Wolfe and Liston, 1998, for a general discussion of the technique), referring both to the lack of any pre-experiment sequence information about the target DNA and its use by many practitioners who are oblivious to the limitations of the approach. Nevertheless, RAPD analysis offers a quick and comparatively cheap approach for the detection of small genetic differences, since a larger proportion of the genome can be sampled than with other techniques. To avoid the shortfalls of RAPD analysis, such as low reproducibility of some bands and the uncertain homology of fragments comigrating in gel electrophoresis, rigorous laboratory standards are required. All reactions should be repeated, and all reactions should be analysed on the same gel for a reliable scoring of presence and absence of bands (Friesen and Klaas, 1998; Wolfe and Liston, 1998). Impurities in the genomic DNA may prevent the reproduction of some bands, and the banding pattern is reproducible only within a specific range of DNA concentrations. This therefore requires the determination of the DNA concentration either fluorometrically or by titration in several PCR reactions. However, with new DNA-isolation kits (such as the Qiagen DNeasy kit or the Macherey-Nagel DNA Plant Nucleospin kit), these problems are easily overcome, avoiding expensive procedures, such as CsCl density-gradient purification of DNA.

For the guaranteed reproduction of specific bands – for example, if linkage to genes of interest is assumed – a RAPD band can be transformed into a sequence-characterized amplified region (SCAR) band (Paran and Michelmore, 1993). The RAPD band is cloned and adjacent bases from genomic sequence are added to the RAPD sequence in order to obtain a PCR primer that should bind only at one locus in the genome.

For an investigation of genetic diversity, in our opinion at least three scored RAPD bands per taxon are required. More than three or four usually do not add more substantial information, due to the inherent noise in the data. Preferably these bands are scored from several primers, since, if more than about ten bands are scored per reaction, less reliable bands have to be included. In this case, due to differences of base composition within a genome, the genome may not be sampled homogeneously.

Mendelian inheritance of RAPD markers chosen for analysis of genetic distances has been demanded (Bradeen and Havey, 1995; Rieseberg, 1996) but might not be feasible for a project of limited size, especially if only initial information is being gathered about a little-known group. Similarly, homology of RAPD bands has been tested by hybridization (in interspecific applications: Inai *et al.*, 1993; Yamagishi, 1995; Lannér *et al.*, 1996), but this is feasible only for a small number of bands of special interest. Other approaches have been used to increase the data value from RAPD reactions – for example, by evaluating differences in band intensity (Demeke *et al.*, 1992) and/or by using more primers of different composition. If more accessions have to be analysed than can fit on to one gel (usually not more than 40–50), other methods should be used, such as hybridization of dot blots with RAPD probes (*Allium* subgenus *Rhizirideum*: Dubouzet *et al.*, 1997), AFLPs (in *Allium*: Smilde *et al.*, 1999) or microsatellites (in *A. cepa*: Fischer and Bachmann, 1998). At a higher taxonomic level, CAPS approaches (e.g. genus *Allium*: Mes *et al.*, 1998; Friesen *et al.*, 1999) are preferable to an extension of the RAPD approach, due to its limitations. While RAPDs are, with all the necessary repetitions and optimizations, not such a cheap procedure as they have sometimes been portrayed, they can still generate, in any laboratory with standard equipment, informative and reproducible data for a medium-sized study of closely related species or popula-

tions. RAPDs have been successfully used in investigations of infraspecific variation and the differentiation of close species, but the interpretation of the data depends on the assumption that amplification products of equal size are homologues. If the relationship between the taxa within a study is not well known, this assumption of band homology is generally hazardous without further tests.

MICROSATELLITES. For the development of microsatellite markers, genomic DNA is fragmented and cloned. From this material, clones can be detected by hybridization which carry highly repetitive sequences of two or three base-pair unit length. After sequencing, flanking non-repetitive sequences can be determined: these are used for the generation of PCR primers. They amplify the repetitive region. Changes in copy number of the repeat can be detected as length variations of the PCR product (Gupta and Varshney, 2000). The mutation of repeat copy numbers occurs at a rate several orders of magnitude higher than nucleotide substitution (Aquadro, 1997), and is therefore useful at the level of population studies. Even with enrichment procedures available that facilitate the detection of suitable repetitive DNA clones (Edwards *et al.*, 1996; for *A. cepa*: Fischer and Bachmann, 1998), the generation of microsatellites is still cumbersome. It is justified only for long-term projects with crops of economic importance – for example, for finding markers linked to genes of special interest – or to contribute towards the construction of a genetic map. The running costs are higher than those of other markers, since the alleles can only be separated on gels prepared from expensive high-resolution agaroses, such as Metaphor™, or, preferably, on polyacrylamide sequencing gels. A specific advantage is the detection of allelic variants at the same locus (as with isozyme analysis), but with virtually unlimited numbers of markers, limited only by material and manpower. In crops other than *Allium*, large data sets for diversity investigation have been generated by microsatellites (e.g. Chavarriaga-Aguirre *et al.*, 1999, also with comparisons with other techniques), but the usefulness of the technique is clearly limited by the availability of funds.

AFLPS. AFLPs basically transform RFLP-type data into PCR-generated markers (Vos *et al.*, 1995). Restriction-enzyme recognition sites are extended by adapter sequences, rendering a PCR reaction on genomic DNA templates less ubiquitous, so that a discrete number of amplification products are generated. These products are separated on sequencing-type gels, and the pattern is detected via labelled primers (radioactively or with a fluorescent label) or directly via silver staining of unlabelled PCR products. While the procedure is technically demanding, advance preparation of the markers is not required. As with other PCR markers, virtually unlimited numbers of markers can be generated in a short time by using different primer extensions flanking the restriction-site core. The technique has been successfully applied to the generation of molecular mapping data and for the generation of nuclear DNA markers for relationship and hybrid analysis. Once the technique is established in a laboratory, the generation of large data sets is straightforward, as the applications in *Allium* testify (Smilde *et al.*, 1999; van Raamsdonk *et al.*, 2000). Compared with microsatellites, AFLP's strength is in gathering large numbers of data points for smaller numbers of taxa – for example, in mapping experiments. Microsatellites should yield allelic markers with higher certainty across a larger number of investigated accessions, since the length and nucleotide sequence of both primer sequences of one microsatellite marker add to the specificity of the amplified locus.

DNA FINGERPRINTING. DNA fingerprinting has earlier been called the hybridization of a genomic DNA blot with labelled microsatellite sequences, which makes corresponding target sequences visible throughout the genome (see Bruford *et al.*, 1992, for an overview). The technique has been tried on *A. cepa* without success (Sharon *et al.*, 1995; M. Klaas, unpublished results), even though the presence of corresponding sequences in

*A. cepa* has been demonstrated by the successful generation of microsatellite markers (Fischer and Bachmann, 1998). Possibly the large genome size of *Allium* makes direct visualization of microsatellite sequences by hybridization under standard conditions, as tested in other plant groups, difficult. The term fingerprinting has subsequently been used in the literature for various PCR-based techniques, such as RAPD, microsatellite analysis (termed single-locus fingerprinting by Bruford *et al.*, 1992) and even isozyme application. Today, therefore, it stands more for the purpose, rather than for a specific technique, of characterizing a genome down to the level of a cultivar, since identification of individuals is usually not an issue in plant science. In *Allium*, cultivar or line identities have been checked by RAPD analysis (*A. cepa*: Campion *et al.*, 1995; Havey, 1995; *A. sativum*: Bradley *et al.*, 1996; Al-Zahim *et al.*, 1999, in tests for somaclonal variation; triploid onion: Puizina *et al.*, 1999). However, as discussed below, the genetic differences between even recognized botanical varieties might be too small to be detected by these general approaches. RFLPs with nuclear probes were also successfully used to distinguish *A. cepa* commercial inbreds (King *et al.*, 1998b). While the approach required considerable experimental effort, a high resolution was achieved by use of 69 anonymous complementary DNA (cDNA) probes and an alliinase clone.

COMPARATIVE DNA SEQUENCING. Potentially the most informative but also the most laborious marker technique is comparative DNA sequencing of specific loci, which has been greatly facilitated by use of PCR techniques. It has been applied in a number of studies on the molecular evolution of *Allium*. The technique is restricted to phylogenetic applications at the section level and above, and will be dealt with in Section 3.3. The comparison between nuclear DNA markers and chloroplast markers, in particular, allows insights into reticulate evolution and hybridogenic speciation common in *Allium*.

The third genome – mitochondrial DNA (mtDNA) – has not been used as a marker for molecular evolution. In plants, the nucleotide substitution is much slower than in animals, where mtDNA has often been applied to molecular evolution studies. In plants, mtDNA is prone to frequent rearrangements, which makes interpretation of data difficult. Since mtDNA is implicated in cytoplasmic male sterility (CMS) systems, RFLP-based detection systems have been developed to distinguish between different mtDNA types (*A. ampeloprasum*: Kik *et al.*, 1997; *A. cepa*/*A. ampeloprasum*: Buiteveld *et al.*, 1998; *A. schoenoprasum*: Engelke and Tatlioglu, 2000).

## 3. Applications in *Allium* Research

### 3.1 Phylogeny/taxonomy

During the late 1980s and the 1990s, molecular phylogenetics has dramatically reshaped our views of the relationships between organisms and of their evolution. Numerous DNA regions representing the nuclear and chloroplast genomes are now routinely used for phylogenetic inference for plants. Revised concepts of relationships based on phylogenetic analyses are resulting in revised classification in many groups of plants (Soltis and Soltis, 2000).

#### *3.1.1 The genus* Allium *and its subdivisions*

The position of the genus within the *Alliaceae* was investigated by Fay and Chase (1996), through a phylogenetic analysis of plastid DNA sequences coding for the large subunit of ribulose-1,5-biphosphate carboxylase (rbcL). This data set, comprising 52 species, also included sequences of *A. subhirsutum*, *A. altaicum* and *Nectaroscordum siculum*. According to Fay and Chase (1996), *N. siculum* should be included in the genus *Allium*, and *Milula spicata*, a rare Central Himalayan–south-eastern Tibetan endemic species, is the closest relative to the genus *Allium*. Its status has recently been revised, also by other molecular markers (Friesen *et al.*, 2000, discussed below).

A first approach to structuring the genus *Allium* itself by molecular markers was published by Linne von Berg *et al.* (1996). From

48 species representing the major subgenera, plastid DNA was isolated and digested with restriction enzymes and the fragment patterns were analysed phenetically, i.e. the presence/absence of each fragment is counted as an independent character of equal weight, contributing to an overall measure of genetic similarity based on the shared proportion of fragments. The major subgenera were identified as clusters in the UPGMA (unweighted pair-group method using arithmetic averages) dendrogram, with the notable exception that species of subgenera *Amerallium* and *Bromatorrhiza* were joined in a loosely associated cluster. RFLP experiments with heterologous plastid DNA probes were applied to investigate more closely the interrelationship of the *Amerallium–Bromatorrhiza* complex (Samoylov *et al.*, 1995, 1999). The subgenus *Bromatorrhiza*, originally circumscribed by Ekberg (1969) by the occurrence of fleshy roots as storage organs and the lack of true storage bulbs or rhizomes, again proved to be polyphyletic and is now partly integrated into the subgenus *Amerallium* (all species with x = 7) and partly included into subgenus *Rhizirideum* (species with x = 8). The distribution of *Amerallium* species between Old World and New World habitats was well reflected in the phylogenetic data, which was also supported by internal transcribed spacer (ITS) sequence analysis (Dubouzet and Shinoda, 1999).

Ohri *et al.* (1998) undertook a survey of the nuclear DNA content (2C values) in 86 species of all subgenera of the genus *Allium*. However, contrary to some earlier assumptions, little indication of phylogenetic information was found in these data; significant loss or gain of DNA amounts per genome was observed, and the 2C values seemed to be related more to ecological factors than to systematic affiliation. Some generalizations, such as a larger or smaller DNA content in certain subgenera, were possible, but there were no distinct discontinuities defining certain groups. In an earlier limited study of 25 *Allium* species, Jones and Rees (1968) had already found considerable differences between 2C values, but they did not attempt to investigate the possible correlation of DNA loss or gain with phylogeny, since at that time the taxonomy of *Allium* was little understood.

Mes *et al.* (1998) included 29 species of *Allium* and seven species of related genera in a phylogenetic study using RFLP data from PCR-amplified cpDNA. In a cladistic analysis, the large subgenera *Rhizirideum* and *Allium*, which had remained largely intact in the phenetic analysis of RFLP bands (Linne von Berg *et al.*, 1996), proved to be polyphyletic, and *N. siculum* was clearly placed in the genus *Allium*. Some deviating sections are affiliated to other groups: the subgenus *Rhizirideum* sect. *Anguinum* with *A. tricoccum* and *A. victorialis* and sect. *Butomissa* with *A. tuberosum* are now associated with subgenus *Melanocrommyum*, and the two subgenus *Allium* sections *Allium* and *Scorodon* are separated by several *Rhizirideum* sections. Earlier, unification or separation of taxa was based on morphological traits, thus leading to mistaken classifications. Hence, using molecular markers, Mes *et al.* (1999) confirmed the artificial nature of subgenera *Rhizirideum*, *Bromatorrhiza* and *Allium*. Some sections in the monophyletic subgenus *Melacrommyum* are also artificial. The subgenus *Bromatorrhiza* is subdivided between the x = 7 and the x = 8 species, in agreement with the earlier studies (Samoylov *et al.*, 1995, 1999; Linne von Berg *et al.*, 1996). In these studies the taxonomy at the level of sections remains more or less intact, but the affiliation of some deviating groups to larger-order structures is changed by the cladistic analysis of molecular markers. The phenetic analysis of RFLP data for a UPGMA clustering (Linne von Berg *et al.*, 1996) gave less reliable grouping at the level of subgenera. Their approach could also lead to the inclusion of misleading data, since bands of the same size were treated as homologues without verification by probe hybridization.

Dubouzet *et al.* (1997) proposed a first phylogeny of subgenus *Rhizirideum* based on nuclear DNA markers. Dot blots with genomic DNA of 44 species were successively hybridized with 55 RAPD fragments. These probes were isolated from separate PCR reactions from 11 *Rhizirideum* species and 11 RAPD primers. Most probes

hybridized to sets of species from several sections, resulting in continuous rather than binomial signal distribution. The intensity of the hybridization signals was determined densitometrically and transformed into a distance matrix. The resulting UPGMA dendrogram largely confirmed the taxonomy as detailed in Hanelt *et al.* (1992). Unfortunately, no species from outside subgenus *Rhizirideum* were included, which might have been needed in order to reliably structure the subgenus itself, which is not a monophyletic group (Mes *et al.*, 1998, 1999). The approach of Dubouzet *et al.* (1997) avoids problems of band homology, as in standard RAPD experiments or RFLPs without hybridization, but the analysis is restricted to distance methods.

Other recent publications on molecular taxonomy (Dubouzet and Shinoda, 1998, subgenus *Melanocrommyum*; van Raamsdonk *et al.*, 2000, subgenus *Rhizirideum*) have the same shortcomings: no species from outside the studied subgenera were included. We regard this as crucial for the adequate positioning of taxa from polyphyletic groups. Another very important aspect in a molecular taxonomy study is the origin and quality of the studied plants. Often researchers collect seeds from botanical gardens, seed companies or other sources and use it without further checks. In the experience of the Gatersleben taxonomic group, about 50% of such material is incorrectly determined or has hybridogenic origins. Our experience indicates that species from genus *Allium* in particular are frequently hybridized in collections and are often wrongly named (N. Friesen, personal observation).

In an ongoing investigation of the phylogeny of *Allium* using molecular markers, we searched for a suitable outgroup taxon as close as possible to but outside the ingroup being studied, to be a part of the cladistic analysis. The results of Fay and Chase (1996) and the general morphological similarity indicated that *Milula* should be the appropriate candidate for this purpose. Phylogenetic relationships between *Allium* and the monotypic Himalayan genus *Milula* were analysed using sequences of the nuclear ribosomal DNA (rDNA) ITS region and of the intergenic spacers from the chloroplast *trn*D(GUC)–*trn*T(GGU) region (Friesen *et al.*, 2000). The comparison of ITS data with the independent cpDNA data set unambiguously placed *M. spicata* within *Allium* subgenus *Rhizirideum*, close to *A. cyathophorum*. Two major clades were found in *Allium* based on both data sets: subgenera *Nectaroscordum* (x = 9) and *Amerallium* (x = 7) on one side, with subgenera *Caloscordum*, *Rhizirideum* and *Milula* (all x = 8) on the other side. This result supports the division of *Allium* into two large groups, as suggested by earlier cpDNA analyses (Linne von Berg *et al.*, 1996; Mes *et al.*, 1999), and the breaking up of subgenus *Bromatorrhiza*, which appears to be an artificial taxon (Samoylov *et al.*, 1995, 1999). Only two small differences in the positions of *A. kingdonii* and *A. insubricum* between the analyses based on nuclear DNA and cpDNA data were found. Hybridization events or sample errors could explain the different positioning of these taxa. To resolve these conflicts a much larger sample of species would have to be analysed to avoid errors introduced by taxon selection.

### 3.1.2 Comparison of cpDNA and nuclear DNA

To study the relationships in the entire genus *Allium*, the ITS region of nuclear rDNA was sequenced from 216 samples that represented 195 *Allium* species, two species of *Nothoscordum*, and one species of each of the genera *Milula*, *Ipheion*, *Dichelostemma* and *Tulbaghia* (N. Friesen, M. Klaas and F.R. Blattner, unpublished data). The subgenera *Rhizirideum* and *Allium*, which are not monophyletic according to the cpDNA analysis, were represented by 162 accessions, and representatives of each section of the subgenera *Amerallium*, *Caloscordum*, *Nectaroscordum* and *Melanocrommyum* were also included. In all cases where species were placed in an unexpected position in the preliminary phylogenetic tree, we analysed more accessions from the particular species to avoid errors from taxon selection. Within the 195 *Allium* species, the lengths of the ITS regions are in a range from 612 bp in *A. cyathophorum* to 661 bp in *A. triquetrum*. Pairwise genetic

distances (Kimura, 1980) are between 1% Kimura distance (between species from one section) to 53% (*A. haneltii* (subgen. *Allium*) to *A. bulgaricum* (subgen. *Nectaroscordum*)). These are unusually large intrageneric genetic distances for the ITS data within *Allium*: Kimura distances above 40% often characterize the most distant genera within subfamilies or even families. Intrageneric distances in other plant families are mostly less than 10%. These findings make *Allium* of very ancient origin, and molecular evolution has not been accompanied by the rise of comparable numbers of taxonomic categories.

A phylogenetic analysis of ITS sequences (Fig. 1.1 in Fritsch and Friesen, Chapter 1, this volume) supported a monophyletic origin of most circumscribed sections, with some exceptions (the morphologically variable sections *Reticulato-Bulbosa*, *Oreiprason* and *Scorodon* are polyphyletic) and a polyphyletic origin of subgenera *Rhizirideum* and *Allium*. Subgenera *Rhizirideum* and *Allium* are subdivided into six monophyletic groups which have different relationships: section *Anguinum* is a sister group of subgenus *Melanocrommyum*; sect. *Butomissa* (including some species from sect. *Reticulato-Bulbosa*) is a sister group to all other sections of subgenus *Rhizirideum* and *Allium*; sect. *Rhizirideum*, *Caespitosoprason*, *Tenuissima* and *A. eduardii* (sect. *Reticulato-Bulbosa*) are sister groups to all the other sections of subgenus *Rhizirideum* and *Allium*; most species from subgenus *Allium* form a monophyletic clade, excluding species from sect. *Scorodon sensu stricto* and *A. turkestanicum*.

In parallel to the beginning of our ITS project described above, a set of cpDNA sequences was gathered in order to detect possible differences between the two data sets, indicative of reticulate evolution events not detected by a single marker. The non-coding rbcL–atpB spacer from cpDNA was amplified and sequenced from 60 accessions belonging to 50 species of genus *Allium* and one species each of the genera *Nothoscordum*, *Tulbaghia* and *Bloomeria* (Fig. 8.1; see Table 8.1 for European Molecular Biology Laboratory (EMBL)* sequence accession numbers and information on the plant accessions, original data not yet published). The resolution of the *rbcL–atpB* marker overlaps to some extent with that of the ITS, but is generally more useful at a somewhat higher taxonomic level – below section level too few substitutions are found – on the other hand, inclusion into the alignment of related species outside *Allium* is still possible, due to conserved portions of the region (see Soltis and Soltis, 1998, for a discussion of different sequencing markers in phylogenetic studies). Three major clades were found in *Allium* based on these sequences: subgenera *Nectaroscordum* (x = 9) and *Amerallium* (x = 7, including species from former subgenus *Bromatorrhiza*); subgenera *Caloscordum*, *Melanocrommyum* and section *Anguinum* (all x = 8); and subgenera *Rhizirideum* (including species from former subgenus *Bromatorrhiza*) and *Allium* (all x = 8). The cpDNA data largely agree with the phylogeny based on the nuclear ITS sequences. Remarkably, in section *Cepa*, the species *oschaninii* and *pskemense* are not included and are more distant than in the ITS tree, indicating a different evolutionary origin of their nuclear and cytoplasmic DNA. The species of subgenus *Allium* are divided into three groups, compared with their monophyletic appearance in the ITS tree of these taxa (see Fig. 1.1, Fritsch and Friesen, Chapter 1, this volume). In theory, all phylogenies based on chloroplast markers should yield the same tree; however, the *rbcL/atpB* intergene sequence data provide a far better taxonomic resolution at the genus level compared with the CAPS-based analysis (Fig. 8.2; Mes *et al.*, 1999). At higher taxonomic levels of 'old genera' like *Allium*, restriction data such as those generated by CAPS apparently include increasingly homoeological characters, resulting in

*European Molecular Biology Laboratory (EMBL), an international network of research institutes funded by 15 countries, is dedicated to research in molecular biology. Apart from computational services such as analysing data sets and providing DNA-analysis software, one of its main goals is to establish a central computer database of DNA sequences as a resource for the scientific community.

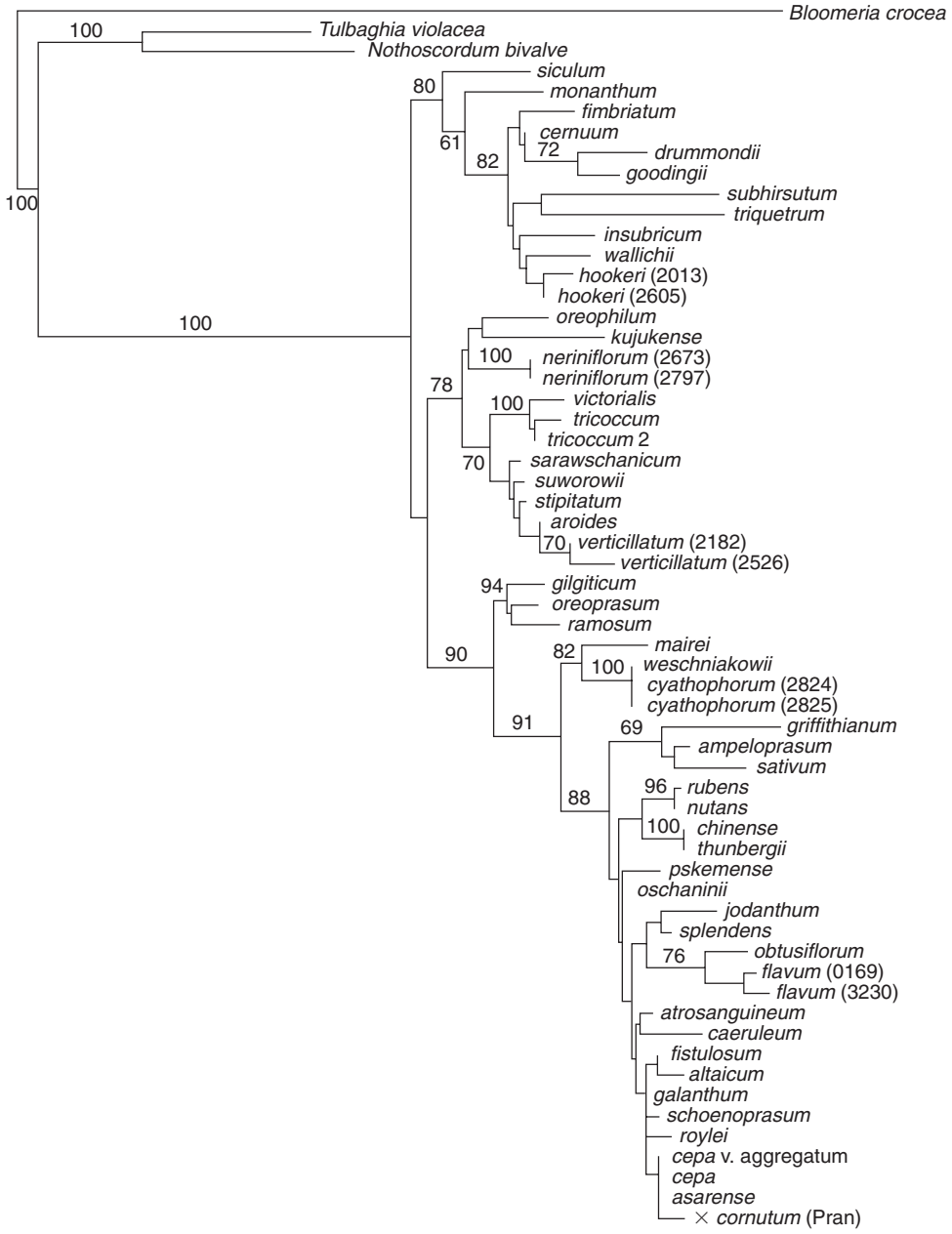


**Fig. 8.1.** Strict consensus tree of maximum parsimony analysis of the *rcbL–atpB* intergenic region, based on 388 trees. Consistency index (CI) 0.760; retention index (RI) 0.790. The non-coding region between the *rcbL* and *atpB* chloroplast genes was amplified from conserved sequences within the genes, as described by Savolainen *et al.* (1994), and manually sequenced in both directions. Intervening primers were synthesized based on *Allium*-specific sequences. From the sequence files, *c.* 880 bp per species, an alignment of *c.* 1100 bp was constructed with the CLUSTALW program (Thompson *et al.*, 1994), and this was used for a cladistic analysis of the data with PAUP 3.1 (Swofford, 1993). The figures above the branches indicate bootstrap values; only figures > 50 are given. bp, base pairs.

**Table 8.1.** The origin and taxonomy of the investigated accessions of the genus *Allium*.

| Subgenus, Section | Species | 2*n* | TAX | Origin | EMBL |
|---|---|---|---|---|---|
| *Allium* | | | | | |
| *Allium* | *ampeloprasum* L. | 48 | 1025 | Zugdidi, Caucasus, Georgia | AJ299086 |
| | *sativum* L. | 16 | 1319 | North Tajikistan | AJ299088 |
| *Avulsea* | *griffithianum* Boiss. | 16 | 3660 | Zaravshan Mts, Takhta-Karachi Pass, Uzbekistan | AJ299128 |
| *Caerulea* | *caeruleum* Pall. | 16 | 1525 | BG Moscow, Russia | AJ299141 |
| *Codonoprasum* | *flavum* L. | 16 | 3230 | BG Linz, Austria | AJ299120 |
| | *flavum* L. | 16 | 0169 | Dizderica, Croatia | AJ299091 |
| | *obtusiflorum* DC | 16 | 3101 | Piserra dello Zingaro-Scopollo, Italy | AJ299119 |
| *Amerallium* | | | | | |
| *Amerallium* | *drummondii* Regel | 14 | 0200 | BG Uppsala, Sweden | AJ299144 |
| *Briseis* | *triquetrum* L. | 16 | 0933 | BG Liege, Belgium | AJ299137 |
| *Bromatorrhiza* | *hookeri* Thwaites | 22 | 2013 | Kunming, China | AJ299095 |
| | *hookeri* Thwaites | 22 | 2506 | SW Lijiang, China | AJ299105 |
| | *wallichii* Kunth | 14 | 2441 | BG Gatersleben, Germany | AJ299104 |
| *Caulorhizideum* | *goodingii* Ownbey | 14 | 3471 | Arizona, USA | AJ299124 |
| *Lophioprason* | *cernuum* Roth | 14 | 0497 | BG Strasbourg, France | AJ299133 |
| *Microscordum* | *monanthum* Maxim. | 32 | 5457 | Vladivostok, Russia | AJ299135 |
| *Molium* | *subhirsutum* L. | 14 | 0023 | Adiacenze di Petralia, Italy | AJ299103 |
| *Narcissoprason* | *insubricum* Boiss. et Reut. | 14 | 0230 | BG Marburg, Germany | AJ299101 |
| *Rhopetoprason* | *fimbriatum* Wats. v. *purdyi* Eastw. | 14 | 3487 | Lake County, California, USA | AJ299146 |
| *Caloscordum* | | | | | |
| *Caloscordum* | *neriniflorum* Herbert | 16 | 2379 | Somon Chalchgol, Mongolia | AJ299102 |
| | *neriniflorum* Herbert | 16 | 2797 | Dauria, Russia | AJ299115 |
| *Nectaroscordum* | | | | | |
| *Nectaroscordum* | *siculum* Ucria | 18 | 0093 | Garden in Gatersleben, Germany | AJ299138 |
| *Melanocrommyum* | | | | | |
| *Acmopetala* | *suworowii* Nabelek | 16 | 3652 | Alma-Ata–Dzhambul road, Kazakhstan | AJ299127 |
| *Aroidea* | *aroides* M. Pop. et Vved. | 16 | 2517 | BG Tashkent, Uzbekistan | AJ299106 |
| *Megaloprason* | *sarawschanicum* Regel | 16 | 3673 | Zaravshan Mts, Tajikistan | AJ299129 |
| | *stipitatum* Regel | 16 | 2257 | Kholmon Valley, Tajikistan | AJ299100 |
| *Miniprason* | *karataviense* Regel | 18 | 2989 | Chilchenboa Mts, Uzbekistan | AJ299118 |
| *Porphyroprason* | *oreophilum* C.A. Mey. | 16 | 0348 | BG Graz, Austria | AJ299125 |

*Continued.*

**Table 8.1.** *Continued.*

| Subgenus, Section | Species | 2*n* | TAX | Origin | EMBL |
|---|---|---|---|---|---|
| *Melanocrommyum* (continued) | | | | | |
| *Verticillata* | *verticillatum* Regel | 16 | 2182 | Gazimajlik Mts, Tajikistan | AJ299099 |
| | *verticillatum* Regel | 16 | 2526 | Dushanbe, Tajikistan | AJ299107 |
| *Vvedenskya* | *kujukense* Vved. | 20 | 3625 | Karatau Mts, Kuyuk Pass, Kazakhstan | AJ299147 |
| *Rhizirideum* | | | | | |
| *Anguinum* | *tricoccum* Sol. | 16 | 2582 | BG Glencoe, Minnesota, USA | AJ299109 |
| | *victorialis* L. | 16 | 2673 | Caucasus, Georgia | AJ299112 |
| *Annuloprason* | *atrosanguineum* Kar. et Kir. | 16 | 2560 | Kusawlisai Valley, Tajikistan | AJ299108 |
| *Butomissa* | *ramosum* L. | 32 | 2735 | BG Alma-Ata, Kazakhstan | AJ299114 |
| | *gilgiticum* Wang et Tang | | | Karakorum, Pakistan (Herbarium, Gatersleben) | AJ299140 |
| *Campanulata* | *jodanthum* Vved. | 16 | 1330 | Kondara Valley, Tajikistan | AJ299089 |
| *Cepa* | *altaicum* Pall. | 16 | 0339 | BG Kaunas, Lithuania | AJ299121 |
| | *cepa* L. | 16 | A878 | cv. Stuttgarter Riesen | AJ299139 |
| | *cepa* Aggregatum group | 16 | 1810 | 1986, No.4–1 | AJ299093 |
| | *fistulosum* L. | 16 | 0266 | Wisley Gardens, UK | AJ299111 |
| | *galanthum* Kar. et Kir. | 16 | 1729 | BG Alma-Ata, Kazakhstan | AJ299092 |
| | *oschaninii* B. Fedtsch. | 16 | 2177 | Varsob Valley, Tajikistan | AJ299098 |
| | *pskemense* B. Fedtsch. | 16 | 1994 | BG Copenhagen, Denmark | AJ299094 |
| | *roylei* Stearn | 16 | 5152 | Olomouc, Czech Republic | AJ299142 |
| | *asarense* R.M.Fritsch et Matin | 16 | 3900 | Central Kopetdag, Turkmenistan | AJ299131 |
| | × *cornutum* G.C. Clementi ex Vis. | 24 | 5193 | 'Pran', Kashmir, India | AJ299134 |
| *Reticulato-Bulbosa* | *oreoprasum* Schrenk | 16 | 3643 | Transili Mts, Turgen Valley, Kazakhstan | AJ299126 |
| | *splendens* Schult. et Schult. f. | 48 | 1288 | BG Kyoto, Japan | AJ299087 |
| *Rhizirideum* | *nutans* L. | 32 | 2080 | Gorno-Altaisk, Altai, Russia | AJ299096 |
| | *rubens* Schrad. | 16 | 1609 | Temirtau, Kazakhstan | AJ299145 |
| *Sacculiferum* | *chinense* G. Don | 32 | 3407 | Fukui, Japan | AJ299122 |
| | *thunbergii* G. Don | 16 | 3408 | Kumamoto, Japan | AJ299123 |
| *Schoenoprasum* | *schoenoprasum* L. | 16 | 4214 | Garden in Gatersleben, Germany | AJ299132 |
| *Coleoblastus* | *mairei* Levl. | 16 | 2104 | BG Zurich, Switzerland | AJ299097 |
| *Cyathophora* | *cyathophorum* Bur. et Franch. | 16 | 2824 | BG Oslo, Norway | AJ299116 |
| | *cyathophorum* Bur. et Franch | 16 | 2825 | BG Jena, Germany | AJ299117 |
| | *weschniakowii* | | | BG Moscow University; Tienshan (Lake Issuk-Kul, Kyrgyzstan) | AJ299143 |

| | | | | | |
|---|---|---|---|---|---|
| *Bloomeria* | *crocea* (Torrey) Coville | | 2697 | BG Santa Barbara, California, USA | AJ299113 |
| *Nothoscordum* | *bivalve* (L.) Britton | 18 | 594 | BG Palermo, Italy | AJ299136 |
| *Tulbaghia* | *violaceae* Harv. | | 1467 | Chelsea Physic Garden, London, UK | AJ299090 |

TAX, accession numbers of the Department of Taxonomy of the Institute for Plant Genetic and Crop Plant Research, Gatersleben; EMBL, sequence accession numbers at the European Molecular Biology Laboratory; BG, botanical garden.

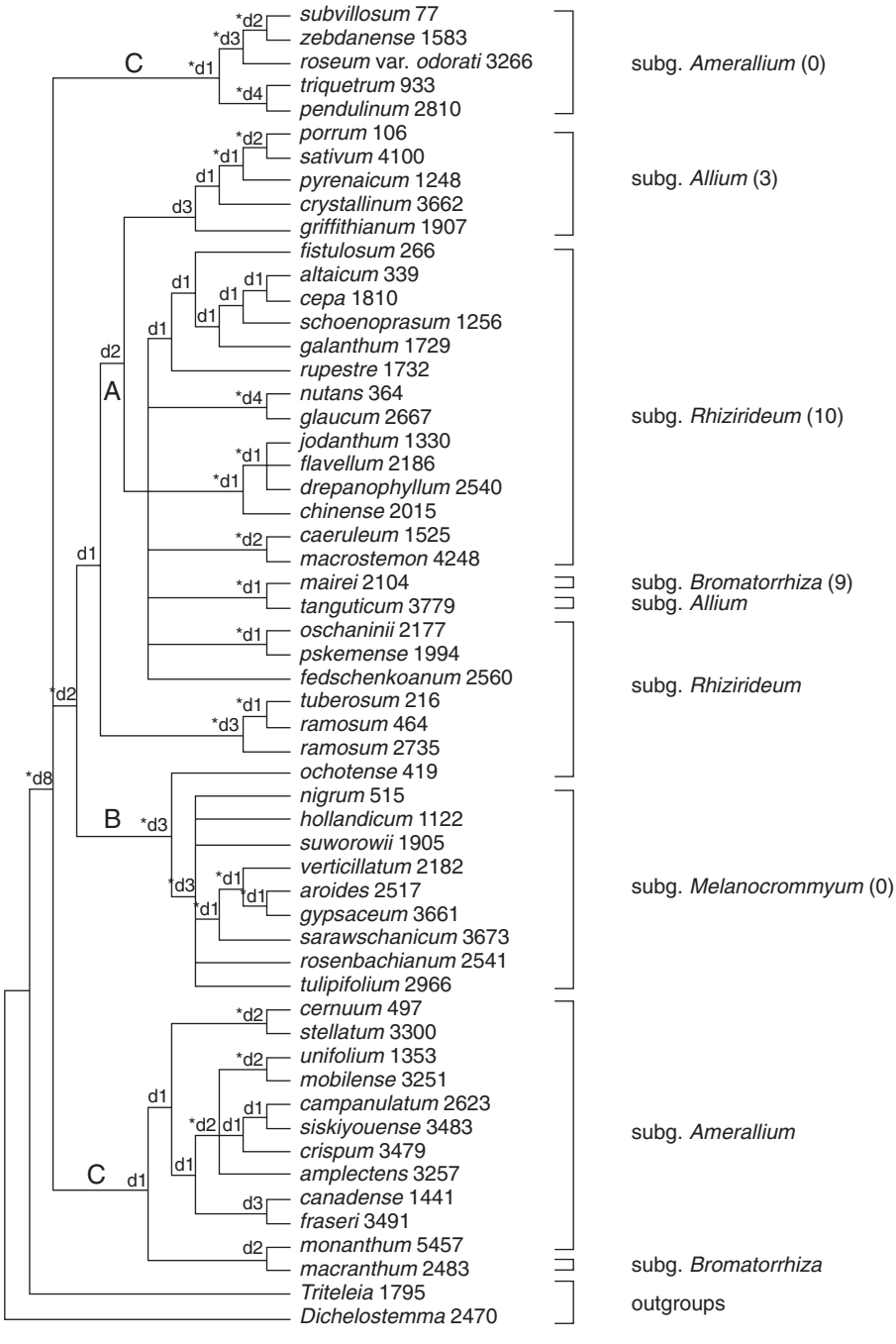


**Fig. 8.2.** Consensus cladogram based on restriction sites and length variants in the chloroplast DNA from *Allium* species (Mes *et al.*, 1999, with the permission of the publishers of *Genome*).

higher noise in the data and unresolved trees. However, CAPS data enable the reconstruction of phylogenies with excellent resolution at lower taxonomic levels (at least within subgenera), and large data sets can be generated in far less time compared with a sequence project with a comparable number of taxa. Comparative sequencing necessarily involves the generation of data from conserved (i.e. uninformative) sequences, which is required to ensure a correct alignment of sequences as the most important step in phylogenetic analysis.

#### 3.1.3 GISH

Detailed information on the chromosomal composition of hybrid plants is possible by genomic *in situ* hybridization (GISH) analysis, a powerful method for the analysis of differentiation between genomes (Schwarzacher and Heslop-Harrison, 2000). Chromosome spreads from metaphase plates are hybridized with total labelled genomic DNA from one of the suspected parent species. By addition of different ratios of unlabelled blocking DNA from the other parent, even closely related genomes can be distinguished, so that single chromosomes or even parts thereof can be attributed to one or the other parent species.

### 3.2 Infraspecific applications

The molecular approach is often the only means of obtaining a sufficient number of unbiased markers for infraspecific investigations. In populations of wild species, an assessment of genetic diversity (usually only in part reflected in morphological differentiation) is essential for the investigation of the status of subspecies groups and problems of recent or ongoing speciation events. In an extensive isozyme study, Rieseberg *et al.* (1987) sampled populations from four varieties of *A. douglasii* (subgen. *Amerallium*) from a limited region in the north-western USA. With 12 enzyme systems, 22 loci were scored and allelic frequencies for a total of 26 populations determined. These molecular data clearly separated the two varieties *douglasii* and *nevii*, while vars *columbianum* and *constrictum* populations formed a third group in the clustering dendrogram. From two isozyme autapomorphies (derived character states that define a new evolutionary line) found in *constrictum*, this variety was concluded to be a recent derivative of *columbianum*. In a later report, Smith and Vuong Pham (1996) applied RAPD data in a similar investigation of the rare *A. aaseae*, endemic to Idaho, and its more common sister species *A. simillimum*. From 12 selected primers, 65 variable markers were scored in 14 populations from both species, but in this case the RAPD dendrogram did not confirm the species status of the populations as determined by morphology. Since the species are defined by ecological and morphological data, an explanation of the RAPD results could be a recent speciation event or might indicate multiple origins of *A. aaseae* from *A. simillimum*. Hybridization and introgression occur in common habitats but do not explain the lack of genetic differentiation in geographically distant populations of the two species.

Most studies of infraspecific differentiation in *Allium* have been aimed at crop plants of economic importance. Questions of crop evolution and the interrelationships of cultivars and varieties are addressed, and the relation of crops to close or ancestral wild species can be clarified. The determination of the genetic diversity of crop accessions is of direct use in a gene bank, both to assess the value of a collection and to direct future collecting missions.

#### 3.2.1 Chives

*Allium schoenoprasum* is extremely widespread in Eurasia and North America. Different morphological types have been described, and section *Schoenoprasum* contains several closely related species, in part of polyploid nature, with partly unclear species status. In such a situation, molecular markers could bring some clarification difficult to obtain by other means. Using 233 RAPD markers derived from 11 primers, Friesen and Blattner (2000) investigated 38 accessions from section

*Schoenoprasum*, including 29 accessions from *A. schoenoprasum* (covering its geographical range in Eurasia), and representatives of seven other species. The molecular markers indicated genetic differentiation within *A. schoenoprasum* according to geographical distribution; however, the morphological types of chives described earlier (Stearn, 1978; Friesen, 1996) were not reflected in the dendrograms.

The raw data were analysed in several ways and their relative merits discussed. Pairwise genetic distances were calculated for the construction of UPGMA and neighbour-joining trees. In principal coordinate data analysis (PCA), new independent coordinates' axes were calculated in a process analogous to the construction of a regression line from a cloud of data points, so as to explain a maximum of the diversity of the underlying data. The taxa are graphically represented as points in a three-dimensional (3-D) space defined by the first three coordinates (see Fig. 8.3). This representation was well suited to demonstrating the reflection of geographical origin in the genetic grouping. Cladistic analysis is based on the reconstruction of a series of phylogenetic splitting events, each defined by gain or loss of characters common to at least two offspring taxa. The temporal order of these events is deduced from comparison with an outgroup species as close as possible to but outside the investigated group. The procedure is more commonly applied in analysing DNA sequence data, where it is generally

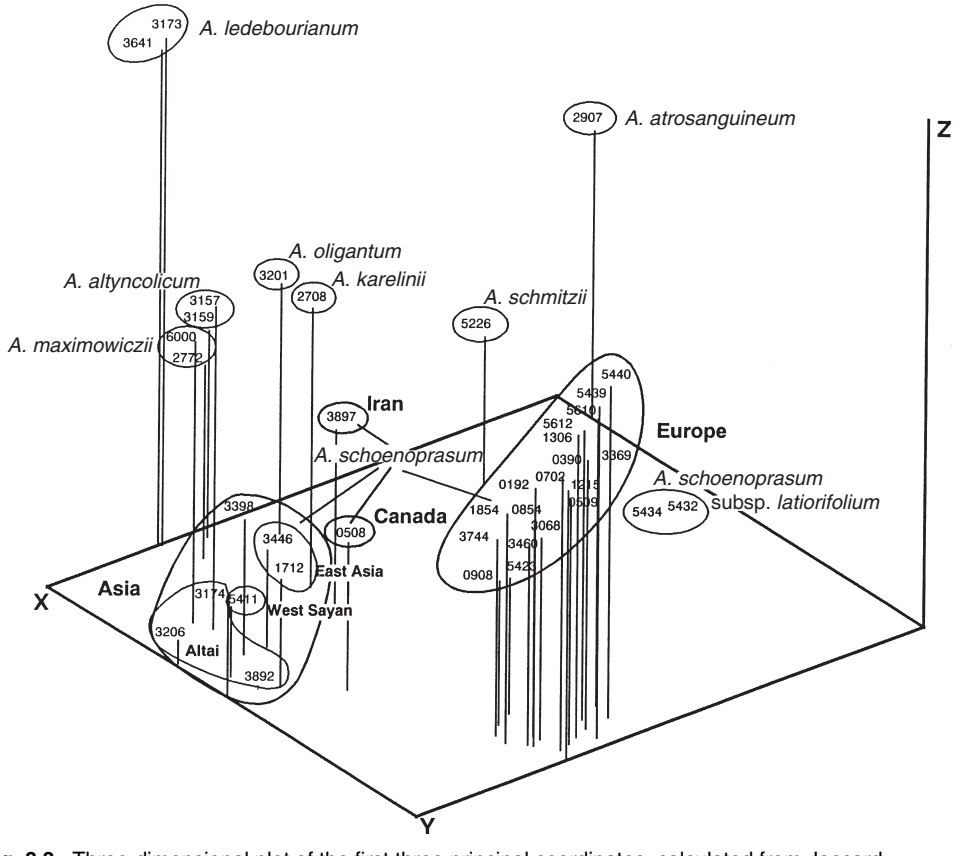


**Fig. 8.3.** Three-dimensional plot of the first three principal coordinates, calculated from Jaccard distances of 38 accessions of seven species of *Allium* sect. *Schoenoprasum* and *A. atrosanguineum*, based on 233 RAPD bands. The four-digit numbers here and in Fig. 8.4 are accession numbers of the living collection of the Department of Taxonomy, IPK, Gatersleben.

clearer what constitutes a change, and it is compared with the gain or loss of a RAPD band. Nevertheless, a cladistic tree of the *Schoenoprasum* RAPD data was constructed. The consensus tree yielded a grouping very similar to the neighbour-joining tree based on pairwise distances. The removal of known polyploid species based on theoretical objections in a phylogenetic analysis did not change the grouping of the remaining species.

### 3.2.2 Garlic

*Allium sativum* is a predominantly sterile species known only in cultivation (see Etoh and Simon, Chapter 5, this volume). Nevertheless, there is great variability in morphological and physiological features and varying degrees of bolting and flower formation, which led to the proposition of three botanical varieties. Presumably *A. longicuspis* is the wild progenitor, but whether its status should be as a separate species or just as feral plants derived from crops has been disputed. Pooler and Simon (1993) investigated a collection of 110 garlic clones with morphological and isozyme methods for an infraspecific classification. Thirteen isozyme systems were tested, although, because of inconsistent staining or lack of variability, only four were useful, and 17 different enzyme groups were detected. While flower characteristics correlated well with isozyme data, bulb-related traits or geographical origin had little predictive value for the genetic relationship of accessions. Maaß and Klaas (1995) tested 300 clones with isozymes, and 48 of these were tested with RAPDs as well, to compare the two marker systems. Their gene pool contained many accessions from areas close to the centre of origin in Central Asia and was suitable for investigating the genetic relationship between cultivated clones with primitive features, derived strains and a feral accession of *A. longicuspis*. Twelve isozyme systems were tested which identified 22 loci, ten of which were polymorphic and defined 16 isozyme groups. Predictably, the 125 RAPD markers allowed a more detailed distinction, but generally both markers gave a good delimitation of varieties *sativum* (bolting and non-bolting types could be separated) and *ophioscorodon*. The third variety, *pekinense* was not distinguishable by either marker from *longicuspis*-type plants; nor was an accession determined as *A. longicuspis* separated from more primitive (i.e. partially fertile) garlics, based on molecular markers.

A similar range of accessions was investigated by Al-Zahim *et al.* (1997). Their results differed in some important aspects. Twenty-seven named garlic cultivars were structured with 63 polymorphic RAPD bands generated from 26 primers. Eleven accessions were assigned to variety *ophioscorodon*, 11 to variety *sativum* and five to *A. longicuspis*. In agreement with Maaß and Klaas (1995), the accessions of var. *sativum* (only non-bolting accessions were included) grouped together: however, these workers found genetic differentiation within var. *ophioscorodon* and interspersal with *A. longicuspis* accessions. These findings were in contrast to the genetic homogeneity of the *ophioscorodon* group (80 accessions were investigated by isozymes, seven of these by RAPDs), being genetically clearly distinct from *longicuspis*-type accessions, as reported by Maaß and Klaas (1995). The different results can probably be explained by the different morphological classification of the material prior to the molecular study, rather than by a misapplication of the RAPD markers in either case, since a comparable number of primers and markers per taxon was used in both laboratories. In the well-characterized collection in Gatersleben, *ophioscorodon* was morphologically clearly distinguishable from *A. longicuspis* (Helm, 1956; Maaß and Klaas, 1995; Maaß, 1996b), while Al-Zahim *et al.* (1997) reported difficulties in distinguishing *ophioscorodon* from *A. longicuspis* based solely on exserted anthers. An interspersal of var. *ophioscorodon* accessions with plants from the *longicuspis* group would explain these data.

Bradley *et al.* (1996) investigated a collection of 20 Australian garlic accessions with five RAPD primers, resulting in 65 marker bands. The approach was well suited to grouping the major Australian cultivars according to bolting behaviour, early and late types, and places of origin.

### 3.2.3 Common onion and related crops

*Allium cepa* is the most important *Allium* crop in terms of economics and areas of production worldwide. Apart from the common bulb onion, shallots and some hybrid crops are derived from this species. Wilkie *et al.* (1993) demonstrated the applicability of RAPD markers in *Allium* in investigations of seven cultivars of *A. cepa* and one accession of each of four other species. Between all the species, 91 polymorphic band positions were scored in reactions with 20 random primers, but within *cepa* only seven bands were polymorphic, resulting in limited resolution at the infraspecific level. Roxas and Peffley (1992) also reported the successful application of RAPDs to onion-cultivar identification using six random primers, but no details were given.

RAPDs and one isozyme locus have been used to test the genetic integrity of a doubled-haploid (DH) line derived from the open-pollinated 'Dorata di Parma' onion cultivar (Campion *et al.*, 1995). While a high degree of RAPD polymorphism was observed among individuals of the parent cultivar population, no differences were found among individuals of the DH line. In a second gynogenic line, a haploid line derived from the Japanese cultivar 'Senshyu Yellow', no RAPD-detectable incidence of genetic instability was found during micropropagation (Campion *et al.*, 1995).

Hybrid-onion seeds are produced from inbred lines: these necessarily retain a relatively high heterozygosity level, since inbreeding depression leads to rapid loss of vigour in bulb onions. In a detailed study, Bradeen and Havey (1995) investigated the use of RAPDs for the testing of the integrity of inbred lines, which is essential to hybrid performance. From a cross between two distant cultivars differing in pungency, soluble solids and storage properties, 59 $F_3$ families were analysed for the segregation of RAPD markers. Of 580 tested random primers, only 53 detected polymorphisms, and 12 of these gave bands in the 3 : 1 segregation ratio expected for genetic markers inherited in a Mendelian way. In a test of four inbred lines, they were not clearly separated in a UPGMA clustering dendrogram based on data from these genetically characterized markers, and some incidence of contamination was found. Data generated in parallel from markers which were not genetically characterized (i.e. not segregating in a Mendelian fashion) agreed only poorly with these results and were discarded from the final analysis. The use of these and other markers in onion for the construction of a low-density genomic map was summarized (Havey *et al.*, 1996; King *et al.*, 1998a).

Considerably more polymorphisms were detected in genomic RFLP blots probed with random nuclear cDNAs, even though this approach met with some technical difficulties due to the high 2C value of onion (Bark and Havey, 1995). These workers investigated the genetic diversity in 17 open-pollinated populations of onions that bulbed under short-day (SD) and long-day (LD) conditions, and two inbred lines (SD and LD). Of 104 cDNA clones, 60 detected at least one polymorphism. In total, 146 fragments were scored for presence or absence. The raw data were analysed cladistically (parsimonious evolution) and by methods based on genetic distances (UPGMA, PCA). The populations were not clearly separated according to their day-length response, a trait conventionally used to classify onion groups, and yet populations known to be closely related were recognized from the DNA data. Generally, the SD populations were the more diverse, and it appeared that LD onions are a derived group.

A shallot was included in the analysis as an outgroup, but was possibly too close to the SD onions. One accession of *A. fistulosum* was also included, but only 14% of the detected fragments were identical to those of *A. cepa*. King *et al.* (1998b) applied the same technique to the investigation of 14 commercial *A. cepa* inbreds in an RFLP study with 69 anonymous cDNA clones. As few as ten polymorphic restriction enzyme/probe combinations were able to distinguish all the investigated inbreds, indicating a high resolving power of suitably chosen nuclear RFLP probes for the characterization of onion lines.

Shallots, while formerly considered a separate species (*A. ascalonicum*), are now con-

sidered part of *A. cepa* as the Aggregatum group. Maaß (1996a) compared the isozyme patterns of 30 individuals of a distinct type, the French grey shallot, with those of 466 bulb onions and other shallots, 15 *A. oschaninii* and 22 *A. vavilovii*. The allele distribution at four isozyme loci suggests that the grey shallot is more closely related to either *A. vavilovii* or *A. oschaninii* than to *A. cepa* and the other shallots, which appear as a closely related assemblage (Fig. 8.4; Messiaen *et al.*, 1993). Within *cepa*, only two loci were polymorphic, while all four loci were polymorphic within the wild species.

The relation of common onions to different types of shallots was investigated by RAPD markers from four primers and morphological traits (Le Thierry D'Ennequin *et al.*, 1997). The seed-propagated types of shallot proved to be closely related to the common onions, while the vegetative shallots grouped separately. In agreement with the isozyme data from Maaß (1996a), the grey shallot was clearly distinct from both types of shallots as well as from common onions.

Arifin and Okubo (1996) structured a large collection of 189 tropical shallots and the sterile wakegi accessions with five isozyme systems. They identified 25 enzyme patterns of wakegi and 18 patterns of shallots; the two groups were clearly distinct, even though the two crops are grown interchangeably in many areas where plants were collected.

While the number of easily scorable isozyme loci is clearly not sufficient for a detailed infraspecific analysis of *A. cepa* compared with the DNA-based approaches described earlier, isozyme analysis provides a powerful technique for investigating the relationships between close species, when large numbers of accessions have to be analysed in order to take into account the infraspecific variation within each species.

#### 3.2.4 Japanese bunching onion

Three varietal groups of *A. fistulosum* were differentiated with seven loci from five enzyme systems (Haishima and Ikehashi, 1992; Haishima *et al.*, 1993) when 23 Japanese, one Chinese and one Korean accessions were analysed. Many loci were uniform among the Japanese accessions, probably due to a loss of diversity over the introduction and selection of this crop. The fixation of several otherwise rare alleles indicated random genetic drift due to 'founder' effects resulting from a population bottleneck.

### 3.3 Hybrids

Interspecific hybridization is known in many *Allium* groups as a mode of speciation through the evolution of the genus. Usually, the interspecific hybrid and offsprings were recognized from intermediate morphological features and increased chromosome numbers. In crops, deliberate hybridizations of close species have been carried out for the construction of maps (van Heusden *et al.*, 2000), to explore the possibility of introgression of desirable foreign traits such as disease resistance not available in the crop's germplasm (Khrustaleva and Kik, 2000), or to probe the relationships between species (van Raamsdonk *et al.*, 1992). In experimental crossings with known parents, it is usually sufficient to validate the hybrid plants by the detection of a few markers characteristic of each parent. These include morphological (e.g. leaf shape, pigmentation), physiological (resistance to pests, growth habit) and/or molecular markers, such as RAPD bands, ITS restriction-enzyme sites, or isozyme markers (van der Valk *et al.*, 1991; Buiteveld *et al.*, 1998; Dubouzet *et al.*, 1998). Similarly, the cytoplasm of the hybrid can be characterized by one or a few parent-specific PCR or RFLP markers from mtDNA or cpDNA (Holford *et al.*, 1991; Satoh *et al.*, 1993; Havey, 1995). More detailed information on the chromosomal composition of hybrid plants is possible by GISH analysis, a powerful method for the analysis of differentiation between genomes. Khrustaleva and Kik (2000) were able to identify by GISH the parental species to chromosomal regions in experimental hybrids with genomic contributions from three parents – *cepa*, *fistulosum* and *roylei*. (For a detailed discussion, see Kik, Chapter 4, this volume.)

For the reconstruction of species hybridization in wild populations or crops in the more distant past, this approach gives increasingly unreliable results. Processes of sequence elimination, intergenomic exchanges and independent molecular evolution in the lineages of putative parent species and the suspected hybrid tend in time to blur the clear distinction between all involved genomes, thus making clear identification by GISH more difficult (Friesen and Klaas, 1998). In these cases, discrete markers, such as RAPDs or AFLPs, may be the most useful, since these data are amenable to the detection of underlying information. For example, PCA can be used to identify segments of the genome that are the most closely related to those of another species (Fig. 8.4; see Fig. 8.3 for an intraspecific application; Friesen and Hermann, 1998; Friesen and Klaas, 1998; Friesen and Blattner, 2000).

### *3.3.1 Genetic structure of species complexes*

INVESTIGATION OF WILD HYBRIDOGENIC SPECIES. Several cytological and molecular techniques have been tested to investigate the hybrid nature of *A. altyncolicum*, suspected to be a (4*n*) allopolyploid species derived from a spontaneous cross between the diploid species *schoenoprasum* × *ledebourianum* (Friesen *et al.*, 1997a). C-banding, ITS sequencing, PCR-RFLP of plastid DNA, GISH, RAPD analysis and rDNA RFLP were applied. GISH revealed the segmental allopolyploid nature of *A. altyncolicum* by specific hybridization of one parent's labelled genomic DNA only to the corresponding chromosomes in the hybrid. Due to the closeness of the parental genomes and their mutual adaptation (since this species originated many generations ago), the ratios

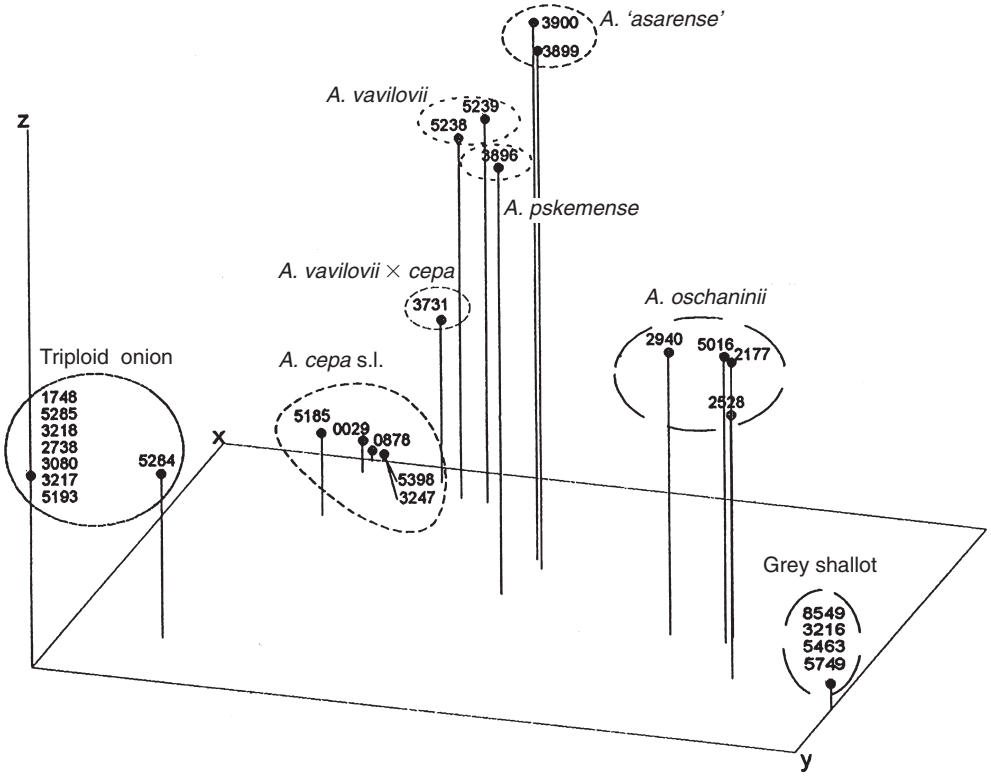


**Fig. 8.4.** Three-dimensional plot of the first three principal coordinates, calculated from a distance matrix based on RAPD data (Friesen and Klaas, 1998).

of labelled DNA and blocking DNA from the other parent required careful calibration to distinguish the parental chromosomes. The chromosomes were also identified by C-banding, the other approaches gave only little or no species-specific distinction. In a related study, ornamentals from subgenus *Melanocrommyum* which had been generated by uncontrolled pollination in breeder's fields were investigated (Friesen *et al.*, 1997b). Initial RAPD screens of the suspected hybrid plants and suspected parent species identified (or excluded) putative genomic contributors. In subsequent GISH experiments, the presence (or absence) of the parental genomes was unequivocally demonstrated (Friesen *et al.*, 1997b).

ANALYSIS OF HYBRID CROPS. Several vegetatively propagated crop species in *Allium* are of hybrid origin. They often arose spontaneously, to be subsequently selected and maintained by gardeners for their unusual properties. *Allium wakegi* is a sexually sterile ancient garden crop in Japan and China. Its hybrid nature (*A. fistulosum* × *A. cepa*) was long suspected because of the intermediate morphology of leaves, bulbs and flowers. The hybrid nature of this species has been proven by GISH (Hizume, 1994). Additional evidence for the hybrid character of *A. wakegi* was gathered by localization of 5S-RNA loci at chromosomal positions corresponding to *A. cepa* and *A. fistulosum* (Hizume, 1994). *A. fistulosum* was identified as the maternal parent of *A. wakegi* by RFLP experiments on purified plastid DNA that was hybridized to an *A. fistulosum* cpDNA probe (Tashiro *et al.*, 1995). Tested by this limited approach, all investigated *A. wakegi* accessions had an identical cytoplasm with a *fistulosum*-like RFLP pattern.

Top onions, or topsetting onions, and viviparous onions are other hybrid species of suspected *A. fistulosum* × *A. cepa* origin, long known from European botanical gardens and gardeners' books as locally cultivated minor garden crops. Havey (1991a) analysed two (2*n*) accessions with RFLP probes for the plastid and the nuclear genome. From the six restriction-enzyme sites in the cpDNA, *A. fistulosum* was determined as the seed parent, while seven out of 11 restriction-enzyme nuclear rDNA fragments were found both from *A. fistulosum* and from *A. cepa*. Maaß (1997a) used six isozyme assays to analyse a large collection of 164 top-onion accessions, six accessions of *A. wakegi*, the parental species *A. cepa* (59 accessions), *A. fistulosum* (27 accessions, including one population of *A. altaicum*) and some artificial hybrids. All allele combinations from the top onions were also found in the hybrids, in addition to some others. The recombination of the hybrids' genome (in crops as well as in artificial hybrids) from the parent species could also be verified by this approach. This required the prior analysis of the allelic diversity of representative collections of the parent species. Finally, both parental genomes of the topsetting onion were clearly identified in a GISH experiment (Friesen and Klaas, 1998).

Less clear is the parentage situation in the (3*n*) viviparous onion *Allium* × *cornutum* (known as 'Pran'), which has long been cultivated in Kashmir and proved to be a rather widespread garden crop, even cultivated sporadically in Europe. Isozyme analysis (Maaß, 1997b) and RAPD analysis (Friesen and Klaas, 1998; Puizina *et al.*, 1999) found no differences between 'Pran' and the Croatian cultivar 'Ljutica'. This crop had been suspected to be either an allotriploid (AAB) (Singh *et al.*, 1967) or a segmental allotriploid (AA′A″) (Koul and Gohil, 1971). The cpDNA pattern could not be attributed to any of the species analysed by Havey (1991b), but is identical with the pattern found in the S cytoplasm of CMS bulb onions (Havey, 1993, 1995), and the nuclear rDNA fragments indicate *A. cepa* as one parent. These data suggested that 'Pran' originated from a cross between a so far unknown seed parent and *A. cepa*. Isozyme analysis failed to identify the second parent of 'Pran' in a comparative study including accessions of sect. *Cepa* species *cepa*, *fistulosum*, *galanthum*, *pskemense*, *vavilovii* and *oschaninii* (Maaß, 1997b). However, *A. fistulosum* was excluded as the second parent (Havey, 1991b; Maaß, 1997b), as also were with some probability *A. schoenoprasum* and *A. roylei*. Friesen and Klaas (1998) investi-

gated the relation of the (3*n*) onion to several sect. *Cepa* species with RAPD and GISH. They found GISH hybridization signals mainly from *A. cepa* (or the close *A. vavilovii*) probes in experiments with different blocking DNAs, rendering the crop a segmental allopolyploid; no GISH signal was obtained from *A. roylei*, which had given no common RAPD bands. The unknown parent contributing the non-*cepa* fraction of the (3*n*) onion remained unknown. Using GISH, Puizina *et al.* (1999) confirmed that a third of the 'Ljutica' genome belongs to *cepa*, another third to *roylei* and the remainder to an as yet unknown third parent. These workers observed signal (in eight chromosomes) from a *roylei* probe with *cepa* blocking DNA. These data are in complete disagreement with the above results of Friesen and Klaas (1998). The latter did not get a clear signal, even without blocking DNA, and found no common RAPD bands between *roylei* and eight studied triploid accessions. While the lack of any common RAPD bands argues against a close relation of *roylei* and triploid onion, the two groups' contradictory GISH results should be resolved by reanalysis of the enigmatic plant.

The diploid grey shallot is a distinct form of shallot long cultivated in France and Italy. Isozyme studies (Maaß, 1996a) proved insufficient to identify the genomic composition, but suggested a closer relationship to *A. vavilovii* and *A. oschaninii* than to *A. cepa*. While a first RAPD study with 24 markers (Le Thierry D'Ennequin *et al.*, 1996) indicated affiliation of grey shallots with other normal shallots belonging to *A. cepa*, both GISH and RAPD data (Friesen and Klaas, 1998) show that most of the chromosomes of grey shallot belong to *A. oschaninii*, with only one and a half chromosome arms derived from either *cepa* or *vavilovii* (see Fig. 8.4 for a 3-D representation of genetic distances between these species; see also Rabinowitch and Kamenetsky, Chapter 17, this volume).

## 4. Conclusions

During recent years, the application of molecular markers has become routine in *Allium* research, due to the increased ease of use and the standardization of the biochemical techniques and of the procedures for the evaluation of results. The power of established techniques, such as RFLP and repetitive DNA analysis, has been enhanced by combination with PCR approaches, enabling increased resolution in less experimental time. While undeniably substantial progress has been made, the extent of a diversity survey is still limited by the necessary effort involved in the generation of molecular markers and analysis for each sample. A breakthrough in this field, e.g. if a survey on large germplasm collections is attempted, will only be achieved by complete automatization of marker generation and analysis. The adaptation of microarray techniques as used at present for expression profiling might be suitable, or further development of genetic bit analysis as presented in *Allium* (Alcala *et al.*, 1997), which is able to detect single-site allelic polymorphisms colorimetrically.

The framework of the genus's phylogeny can be considered as validated, especially if the same groupings are resolved by nuclear as well as chloroplast markers. This also applies to the relationships of the subgenera within the genus and their circumscription. However, with the finer detail now available (see Figs 8.1 and 8.2 and the ITS-based tree in Fritsch and Friesen, Chapter 1, this volume), some arbitrariness has become apparent in the decisions made as to which groups are elevated to subgenus level. Final classification will depend not only on phylogenetic conclusions but on practical considerations.

More troubling are the contradictory groupings obtained in spite of thorough analysis with different markers: plastid DNA analysis should yield approximations of the very same tree, regardless of its origin from RFLP, CAPS or DNA sequences. Excluding experimental errors, these differences are most probably founded on use of the marker at levels of taxonomic resolution not suitable for its resolving power, i.e. when the phylogenetic signal from observed mutations is hidden by multiple changes affecting the same restriction site. The only alternative explanation would be the existence of irregular recombination of plastid DNA not yet documented.

Differences between nuclear DNA and cpDNA phylogenies are explained by reticulate evolution, leading to new recombinant types in the nuclear DNA but not the cpDNA. However, a slight unease remains at the interpretation of ITS sequences, which is at present the only nuclear marker established for investigations at the genus level on a larger scale. Its specific type of sequence evolution – concerted evolution in the repetitive rDNA cluster leading to homogenization if different types of ITS sequences are present – is not invariably representative of molecular evolution throughout the genome, even though ITS analysis yields reasonable groupings in agreement with other types of data, such as morphological and anatomical studies.

## References


Alcala, J., Giovannoni, J.J., Pike, L.M. and Reddy, A.S. (1997) Application of Genetic Bit Analysis (GBA™) for allelic selection in plant breeding. *Molecular Breeding* 3, 495–502.

Al-Zahim, M.A., Newbury, H.J and Ford-Lloyd, B.V. (1997) Classification of genetic variation in garlic (*Allium sativum* L.) revealed by RAPD. *HortScience* 32, 1102–1104.

Al-Zahim, M.A., Ford-Lloyd, B.V. and Newbury, H.J. (1999) Detection of somaclonal variation in garlic (*Allium sativum* L.) using RAPD and cytological analysis. *Plant Cell Reports* 18, 473–477.

Aquadro, C.F. (1997) Insight into the evolutionary process from patterns of DNA sequence variability. *Current Opinions in Genetics and Development* 7, 835–840.

Arifin, N.S. and Okubo, H. (1996) Geographical distribution of allozyme patterns in shallot (*Allium cepa* var. *ascalonicum* Backer) and wakegi onion (*A.* × *wakegi* Araki). *Euphytica* 91, 305–313.

Bark, O.H. and Havey, M.J. (1995) Similarities and relationships among populations of the bulb onion as estimated by nuclear RFLPs. *Theoretical and Applied Genetics* 90, 407–414.

Bradeen, J.M. and Havey, M.J. (1995) Randomly amplified polymorphic DNA in bulb onion and its use to assess inbred integrity. *Journal of the American Society for Horticultural Science* 120, 752–758.

Bradley, K.F., Rieger, M.A. and Collins, G.G. (1996) Classification of Australian garlic cultivars by DNA fingerprinting. *Australian Journal of Experimental Agriculture* 36, 613–618.

Bruford, W.M., Hanotte, O., Brookfield, J.F.Y. and Burke, T. (1992) Single-locus and multilocus DNA fingerprinting. In: Hoelzel, A.R. (ed.) *Molecular Genetic Analysis of Populations. A Practical Approach*. IRL Press, Oxford, pp. 225–269.

Buiteveld, J., Kassies, W., Geels, R., van Lookeren Campagne, M.M., Jacobsen, E. and Creemers-Molenaar, J. (1998) Biased chloroplast and mitochondrial transmission in somatic hybrids of *Allium ampeloprasum* L. and *Allium cepa* L. *Plant Science* 131, 219–228.

Campion, B., Bohanec, B. and Javornik, B. (1995) Gynogenic lines of onion (*Allium cepa* L.): evidence of their homozygosity. *Theoretical and Applied Genetics* 91, 598–602.

Chase, C.D., Ortega, V.M. and Vallejos, C.E. (1991) DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. *Theoretical and Applied Genetics* 81, 806–811.

Chavarriaga-Aguirre, P., Maya, M.M., Tohme, J., Duque, M.C., Iglesias, C., Bonierbale, M.W., Kresovich, S. and Kochert, G. (1999) Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA based markers to maintain germplasm collections. *Molecular Breeding* 5, 263–273.

Demeke, T., Adams, R.P. and Chibbar, R. (1992) Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case study in *Brassica*. *Theoretical and Applied Genetics* 84, 990–994.

Dubouzet, J.G. and Shinoda, K. (1998) Phylogeny of *Allium* L. subg. *Melanocrommyum* (Webb et Berth.) Rouy based on DNA sequence analysis of the internal transcribed spacer region of nrDNA. *Theoretical and Applied Genetics* 97, 541–549.

Dubouzet, J.G. and Shinoda, K. (1999) Relationships among Old and New World *Allium*s according to ITS DNA sequence analysis. *Theoretical and Applied Genetics* 98, 422–433.

Dubouzet, J.G., Shinoda, K. and Murata, N. (1997) Phylogeny of *Allium* L. subgenus *Rhizirideum* (G. Don ex Koch) Wendelbo according to dot blot hybridization with randomly amplified DNA probes. *Theoretical and Applied Genetics* 95, 1223–1228.

Dubouzet, J.G., Shinoda, K. and Murata, N. (1998) Interspecific hybridization of *Allium giganteum* Regel: production and early verification of putative hybrids. *Theoretical and Applied Genetics* 96, 385–388.

Edwards, K.J., Barker, J.H.A., Daly, A., Jones, C. and Karp, A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* 20, 758–760.

Ekberg, L. (1969) Studies in the genus *Allium* II. A new subgenus and new sections from Asia. *Botaniska Notiser (Lund)* 122, 57–68.

Engelke, T. and Tatlioglu, T. (2000) Mitochondrial genome diversity in connection with male sterility in *Allium schoenoprasum* L. *Theoretical and Applied Genetics* 100, 942–948.

Fay, M.F. and Chase, M.W. (1996) Resurrection of *Themidaceae* for the *Brodiaceae* alliance, and circumscription of *Alliaceae*, *Amaryllidaceae* and *Agapanthoideae*. *Taxon* 45, 441–451.

Fischer, D. and Bachmann, K. (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *BioTechniques* 24, 796–802.

Friesen, N. (1996) A taxonomic and chorological revision of the genus *Allium* L. sect. *Schoenoprasum* Dumort. *Candollea* 51, 461–473.

Friesen, N. and Blattner, F.R. (2000) Geographical isolation predominates over ecological differentiation in the phylogeny of *Allium schoenoprasum*. *Plant Biology* 2, 297–305.

Friesen, N. and Hermann, N. (1998) Taxonomy, chorology and evolution of *Allium lusitanicum* – the European '*A. senescens*'. *Linzer Biologische Beiträge* 30, 815–830.

Friesen, N. and Klaas, M. (1998) Origin of some minor vegetatively propagated *Allium* crops studied with RAPD and GISH. *Genetic Resources and Crop Evolution* 45, 511–523.

Friesen, N., Borisjuk, N., Mes, T.H.M., Klaas, M. and Hanelt, P. (1997a) Allotetraploid origin of *Allium altyncolium* (sect. *Schoenoprasum*) as investigated by karyological and molecular markers. *Plant Systematics and Evolution* 206, 317–335.

Friesen, N., Fritsch, R. and Bachmann, K. (1997b) Hybrid origin of some ornamentals of *Allium* subgenus *Melanocrommyum* verified with GISH and RAPD. *Theoretical and Applied Genetics* 95, 1229–1238.

Friesen, N., Pollner, S., Bachmann, K. and Blattner, F.R. (1999) RAPDs and non-coding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (*Alliaceae*). *American Journal of Botany* 86, 554–562.

Friesen, N., Fritsch, R.M., Pollner, S. and Blattner, F.R. (2000) Molecular and morphological evidence for an origin of the aberrant genus *Milula* within Himalayan species of *Allium* (*Alliaceae*). *Molecular Phylogenetics and Evolution* 17, 209–218.

Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185.

Haishima, M. and Ikehashi, H. (1992) Identification of isozyme genes in native varieties of Japanese bunching onion (*Allium fistulosum* L.). *Japanese Journal of Breeding* 42, 497–505.

Haishima, M., Kato, J. and Ikehashi, H. (1993) Isozyme polymorphism in native varieties of Japanese bunching onion (*Allium fistulosum* L.). *Japanese Journal of Breeding* 43, 537–547.

Hanelt, P., Schultze-Motel, J., Fritsch, R., Kruse, J., Maaß, H.I., Ohle, H. and Pistrick, K. (1992). Infrageneric grouping of *Allium* – the Gatersleben approach. In: Hanelt, P., Hammer, K. and Knüpffer, H. (eds) *The Genus Allium – Taxonomic Problems and Genetic Resources*. P*roceedings of an International Symposium, Gatersleben, 11–13 June 1991*. IPK, Gatersleben, Germany, pp. 107–123.

Havey, M.J. (1991a) Molecular characterization of the interspecific origin of viviparous onion. *Journal of Heredity* 82, 501–508.

Havey, M.J. (1991b) Phylogenetic relationships among cultivated *Allium* species from restriction enzyme analysis of the chloroplast genome. *Theoretical and Applied Genetics* 81, 752–757.

Havey, M.J. (1993) A putative donor of S-cytoplasm and its distribution among open-pollinated populations of onion. *Theoretical and Applied Genetics* 86, 128–134.

Havey, M.J. (1995) Identification of cytoplasms using the polymerase chain reaction to aid in the extraction of maintainer lines from open-pollinated populations of onion. *Theoretical and Applied Genetics* 90, 263–268.

Havey, M.J., King, J.J., Bradeen, J.M. and Bark, O. (1996) Molecular markers and mapping in bulb onion, a forgotten monocot. *HortScience* 31, 1116–1118.

Helm, J. (1956) Die zu Würz- und Speisezwecken kultivierten Arten der Gattung *Allium* L. *Kulturpflanze* 4, 130–180.

Hillis, D.M., Moritz, C. and Mable, B.K. (eds) (1996) *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts, 655 pp.

Hizume, M. (1994) Allodiploid nature of *Allium wakegi* Araki revealed by genomic *in situ* hybridization and localization of 5S and 18S rDNAs. *Japanese Journal of Genetics* 69, 407–415.

Hoelzel, A.R. (1992) *Molecular Genetic Analysis of Populations, A Practical Approach*, 1st edn. IRL Press, Oxford, 315 pp.

Holford, P., Croft, J.H. and Newbury, H.J. (1991) Differences between, and possible origins of, the cytoplasms found in fertile and male-sterile onions (*Allium cepa* L.). *Theoretical and Applied Genetics* 82, 737–744.

Inai, S., Ishikawa, K., Nunomura, O. and Ikehashi, H. (1993) Genetic analysis of stunted growth by nuclear–cytoplasmic interaction in interspecific hybrids of *Capsicum* by using RAPD markers. *Theoretical and Applied Genetics* 87, 416–422.

Jones, R.N. and Rees, H. (1968) Nuclear DNA variation in *Allium*. *Heredity* 23, 591–605.

Khrustaleva, L.I. and Kik, C. (2000) Introgression of *Allium fistulosum* into *A. cepa* mediated by *A. roylei*. *Theoretical and Applied Genetics* 100, 17–26.

Kik, C., Samoylov, A.M., Verbeek, W.H.J. and van Raamsdonk, L.W.D. (1997) Mitochondrial DNA variation and crossability of leek (*Allium porrum*) and its wild relatives from the *Allium ampeloprasum* complex. *Theoretical and Applied Genetics* 94, 465–471.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16, 111–120.

King, J.J., Bradeen, J.M., Bark, O., McCallum, J.A. and Havey, M.J. (1998a) A low-density map of onion reveals a role for tandem duplication in the evolution of an extremely large diploid genome. *Theoretical and Applied Genetics* 96, 52–62.

King, J.J., Bradeen, J.M. and Havey, M.J. (1998b) Variability for restriction fragment-length polymorphisms (RFLPs) and relationships among elite commercial inbred and virtual hybrid onion populations. *Journal of the American Society for Horticultural Science* 123, 1034–1037.

Klaas, M. (1998) Applications and impact of molecular markers on evolutionary and diversity studies in the genus *Allium*. *Plant Breeding* 117, 297–308.

Koul, A.K. and Gohil, R.N. (1971) Further studies on natural triploidy in viviparous onion. *Cytologia* 36, 253–261.

Lannér, C., Bryngelsson, T. and Gustafsson, M. (1996) Genetic validity of RAPD markers at the intra- and inter-specific level in wild *Brassica* species with $n = 9$. *Theoretical and Applied Genetics* 93, 9–14.

Le Thierry D'Ennequin, M., Panaud, O., Robert, T. and Ricroch, A. (1997) Assessment of genetic relationships among sexual and asexual forms of *Allium cepa* using morphological traits and RAPD markers. *Heredity* 78, 403–409.

Linne von Berg, G., Samoylov, A., Klaas, M. and Hanelt, P. (1996) Chloroplast DNA restriction analysis and the infrageneric grouping of *Allium* (*Alliaceae*). *Plant Systematics and Evolution* 200, 253–261.

Maaß, H.I. (1996a) About the origin of the French grey shallot. *Genetic Resources and Crop Evolution* 43, 291–292.

Maaß, H.I. (1996b) Morphologische Beobachtungen an Knoblauch. *Der Palmengarten* 60, 65–69.

Maaß, H.I. (1997a) Genetic diversity in the top onion, *Allium* × *proliferum*, analysed by isozymes. *Plant Systematics and Evolution* 208, 35–44.

Maaß, H.I. (1997b) Studies on triploid viviparous onions and their origin. *Genetic Resources and Crop Evolution* 44, 95–99.

Maaß, H.I. and Klaas, M. (1995) Infraspecific differentiation of garlic (*Allium sativum* L.) by isozyme and RAPD markers. *Theoretical and Applied Genetics* 91, 89–97.

May, B. (1992) Starch gel electrophoresis of allozymes. In: Hoelzel, A.R. (ed.) *Molecular Genetic Analysis of Populations. A Practical Approach*. IRL Press, Oxford, pp. 1–27.

Mayr, E. (1969) Grundgedanken der Evolutionsbiologie. *Naturwissenschaften* 56, 14–25. Reprinted in: Mayr, E. (1976) *Evolution and the Diversity of Life*, 2nd printing. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, pp. 9–29.

Mes, T.H.M., Friesen, N., Fritsch, R.M., Klaas, M. and Bachmann, K. (1998) Criteria for sampling in *Allium* (*Alliaceae*) based on chloroplast DNA PCR-RFLPs. *Systematic Botany* 22, 701–712.

Mes, T.H.M., Fritsch, R.M., Pollner, S. and Bachmann, K. (1999) Evolution of the chloroplast genome and polymorphic ITS regions in *Allium* subg. *Melanocrommyum*. *Genome* 42, 237–247.

Messiaen, C.M., Cohat, J., Leroux, J.P., Pichon, M. and Beyries, A. (1993) *Les Allium Alimentaires Reproduits par Voie Végétative*. INRA, Paris, 230 pp.

Murphy, R.W., Sites, J.W., Buth, D.G. and Haufler, C.H. (1996) Proteins: isozyme electrophoresis. In: Hillis, D.M., Moritz, C. and Mable, B.K. (eds) *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts, pp. 51–120.

Ohri, D., Fritsch, R. and Hanelt, P. (1998) Evolution of genome size in *Allium* L. (*Alliaceae*). *Plant Systematics and Evolution* 210, 57–86.

Paran, I. and Michelmore, W. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* 85, 985–993.

Peffley, E.B. and Orozco-Castillo, C. (1987) Polymorphism of isozymes within plant introductions of *Allium cepa* L. and *A. fistulosum* L. *HortScience* 22, 956–957.

Pooler, M.R. and Simon, P.W. (1993) Characterization and classification of isozyme and morphological variation in a diverse collection of garlic clones. *Euphytica* 68, 121–130.

Puizina, J., Javornik, B., Bohanec, B., Schweizer, D.M., Maluszynska, J. and Papeš, D. (1999) Random amplified polymorphic DNA analysis, genome size, and genomic *in situ* hybridization of triploid viviparous onions. *Genome* 42, 1208–1216.

Rieseberg, L.H. (1996) Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5, 99–105.

Rieseberg, L.H., Peterson, P.M., Soltis, D.E. and Annable, C.R. (1987) Genetic divergence and isozyme number variation among four varieties of *Allium douglasii* (*Alliaceae*). *American Journal of Botany* 74, 1614–1624.

Roxas, V.P. and Peffley, E.B. (1992) Short-day onion varietal identification using molecular (RAPD) markers. *Allium Improvement Newsletter* 2, 15–17.

Samoylov, A., Klaas, M. and Hanelt, P. (1995) Use of chloroplast DNA polymorphisms for the phylogenetic study of the subgenera *Amerallium* and *Bromatorrhiza* (genus *Allium*). *Feddes Repertorium* 106, 161–167.

Samoylov, A., Friesen, N., Pollner, S. and Hanelt, P. (1999) Use of chloroplast DNA polymorphisms for the phylogenetic study of *Allium* subgenus *Amerallium* and subgenus *Bromatorrhiza* (*Alliaceae*) II. *Feddes Repertorium* 110, 103–109.

Satoh, Y., Nagai, M., Mikami, T. and Kinoshita, T. (1993) The use of mitochondrial DNA polymorphism in the classification of individual onion plants by cytoplasmic genotypes. *Theoretical and Applied Genetics* 86, 345–348.

Savolainen, V., Manen, J.F., Douzery, E. and Spichiger, R. (1994) Molecular phylogeny of families relating to *Celastrales* based on rbcL 5′ flanking sequences. *Molecular Phylogeny and Evolution* 3, 27–37.

Schwarzacher, T. and Heslop-Harrison, P. (2000) *Practical* in situ *Hybridization*. BIOS Scientific Publishers, Oxford, 203 pp.

Sharon, D., Adato, A., Mhameed, S., Lavi, U., Hillel, J., Gomolka, M., Epplen, C. and Epplen, J.T. (1995) DNA fingerprinting in plants using simple-sequence repeat and minisatellite probes. *HortScience* 30, 109–112.

Singh, F., Ved Brat, S. and Khoshoo, T.N. (1967) Natural triploidy in viviparous onions. *Cytologia* 32, 403–407.

Smilde, W.D., van Heusden, A.W. and Kik, C. (1999) AFLPs in leek (*Allium porrum*) are not inherited in large linkage blocks. *Euphytica* 110, 127–132.

Smith, J.M. and Vuong Pham, T. (1996) Genetic diversity of the narrow endemic *Allium aaseae* (*Alliaceae*). *American Journal of Botany* 83, 717–726.

Soltis, D.E. and Soltis, P.S. (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis, D.E., Soltis, P.S. and Doyle, J.J. (eds) *Molecular Systematics of Plants. II. DNA Sequencing*. Kluwer Academic Publishing, Boston, Massachusetts, pp. 1–42.

Soltis, D.E. and Soltis, P.S. (2000) Contributions of plant molecular systematics to studies of molecular evolution. *Plant Molecular Biology* 42, 45–75.

Stearn, W.T. (1978) European species of *Allium* and allied genera of *Alliaceae*: a synonymic enumeration. *Annales Musei Goulandris* 4, 83–198.

Swofford, D.L. (1993) *PAUP: Phylogenetic Analysis using Parsimony, Version 3.1.1*. Illinois Natural History Survey, Champaign, Illinois.

Tashiro, Y., Oyama, T., Iwamoto, Y., Noda, R. and Miyazaki, S. (1995) Identification of maternal and paternal plants of *Allium wakegi* Araki by RFLP analysis of chloroplast DNA. *Journal of the Japanese Society for Horticultural Science* 63, 819–824.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choices. *Nucleic Acids Research* 22, 4673–4680.

van der Valk, P., Kik, C., Verstappen, F., Everink, J.T. and de Vries, J.N. (1991) Independent segregation of two isozyme markers and inter-plant differences in nuclear DNA content in the interspecific backcross (*Allium fistulosum* L. × *A. cepa* L.) × *A. cepa* L. *Euphytica* 55, 151–156.

van Heusden, A.W., van Ooijen, J.W., Vrielink-van Ginkel, R., Verbeek, W.H.J., Wietsma, W.A. and Kik, C. (2000) A genetic map of an interspecific cross in *Allium* based on amplified fragment length polymorphism (AFLP™) markers. *Theoretical and Applied Genetics* 100, 118–126.

van Raamsdonk, L.W.D., Wietsma, W.A. and de Vries, J.N. (1992) Crossing experiments in *Allium* L. section *Cepa*. *Botanical Journal of the Linnean Society* 109, 293–303.

van Raamsdonk, L.W.D., Vrielink-van Ginkel, M. and Kik, C. (2000) Phylogeny reconstruction and hybrid analysis in *Allium* subgenus *Rhizirideum*. *Theoretical and Applied Genetics* 100, 1000–1009.

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

Wendel, J.F. and Weeden, N.F. (1989) Visualization and interpretation of plant isozymes. In: Soltis, D.E. and Soltis, P.S. (eds) *Isozymes in Plant Biology*. Dioscorides Press, Portland, Oregon, pp. 5–45.

Werman, S.D., Springer, M.S. and Britten, R.J. (1996) Nucleic acids I: DNA–DNA hybridization. In: Hillis, D.M., Moritz, C. and Mable, B.K. (eds) *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts, pp. 169–203.

Wilkie, S.E., Isaac, P.G. and Slater, R.J. (1993) Random amplified polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. *Theoretical and Applied Genetics* 86, 497–504.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18, 6531–6535.

Wolfe, A.D. and Liston, A. (1998) Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Soltis, D.E., Soltis, P.S. and Doyle, J.J. (eds) *Molecular Systematics of Plants, II. DNA Sequencing*. Kluwer Academic Publishers, Boston, Massachusetts, pp. 43–86.

Yamagishi, M. (1995) Detection of section-specific random amplified polymorphic DNA (RAPD) markers in *Lilium*. *Theoretical and Applied Genetics* 91, 830–835.

Zimmer, E.A., White, T.J., Cann, R.L. and Wilson, A.C. (eds) (1993) *Molecular Evolution: Producing the Biochemical Data*. *Methods in Enzymology* 224, Academic Press, San Diego, California.