Genetic Diversity in Fruit and Berry Crops Estimated with Molecular Markers

Larisa Gustavsson
Faculty of Landscape Planning, Horticulture and Agricultural Science
Department of Plant Breeding and Biotechnology
Balsgård

Doctoral thesis
Swedish University of Agricultural Sciences
Alnarp 2008

Acta Universitatis Agriculturae Sueciae
2008: 4

ISSN 1652-6880
ISBN 978-91-85913-37-4

Tryck: SLU Service/Repro, Alnarp 2008

Abstract

Gustavsson, L. 2008. Genetic diversity in fruit and berry crops estimated with molecular markers. Doctoral dissertation.
ISSN 1652-6880, ISBN 978-91-85913-37-4

The amount and structure of genetic diversity has been investigated in three fruit and berry crops which differ in their biological properties and history of domestication: *Chaenomeles*, lingonberry and apple. Genetic variability has been assessed by a variety of molecular markers: isozymes, RAPD, SSR, SCAR and CAPS with the aim to develop sampling strategies, select material for *ex situ* conservation and provide lasting identification of cultivars.

Breeding of *Chaenomeles*, a newly domesticated fruit crop, is dependent on well-characterized plant material containing considerable amount of genetic variation. Phenetic relationships within the genus *Chaenomeles*, estimated on both cultivated plant material and on wild populations, were congruent and indicated that *C. japonica* is differentiated from all the other species, i.e., *C. speciosa*, *C. cathayensis* and *C. thibetica*, and the hybrid taxon, *C.* x *superba*. Furthermore, *C. japonica* and *C. speciosa* were found to be considerably more variable than *C. cathayensis* and *C. thibetica*, and thus different sampling strategies are called for.

Lingonberry breeding is still to a large extent dependent on plant material derived from wildgrowing populations. A large-scale pattern of isolation-by-distance among European populations was encountered. Within-population variability was medium high, and thus in agreement with the mixed mating system and life history traits of this species. RAPD markers in combination with the Maximum diversity computer program allowed for selection of a diverse and representative plant material for future *ex situ* conservation.

Commercially available apple cultivars must be provided with quick and lasting identification, which can guarantee that they are true-to-type. Reliable identification of apple cultivars in genetic resource collections is also very important for, e.g., scientists and breeders. In one study, 151 cultivars were characterised with RAPD markers. No pronounced groupings associated with geographic origin were found, and there were also no major changes in levels of diversity that could be associated with geographic or temporal origination of the different cultivars. In a second study, 68 mandate apple cultivars, aimed for preservation in Sweden, were provided with allelic profiles based on several SSR-loci and the self-incompatibility locus.

Key words: apple, CAPS, *Chaenomeles*, genetic diversity, isozymes, lingonberry, RAPD, SCAR, SSR.

Author's address: Larisa Gustavsson, Balsgård–Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, Fjälkestadsvägen 459, SE-291 94 Kristianstad, Sweden

Contents

List of Appendices

Papers I-V

The present thesis is based on the following papers, which will be referred to by their Roman numerals:

I. Garkava, L.P., Rumpunen, K. & Nybom, H. 2000. Genetic relationships in *Chaenomeles* Lindl. (Rosaceae) revealed by isozyme analysis. *Scientia Horticulturae* 85: 21–35.

II. Bartish, I.V., Garkava, L.P., Rumpunen, K. & Nybom, H. 2000. Phylogenetic relationships and differentiation among and within populations of *Chaenomeles* Lindl. (Rosaceae) estimated with RAPD and isozymes. *Theoretical and Applied Genetics* 101: 554–563.

III. Garkava-Gustavsson, L., Persson, H.A., Nybom, H., Rumpunen, K., Gustavsson, B.A. & Bartish, I.V. 2005. RAPD-based analysis of genetic diversity and selection of lingonberry (*Vaccinium vitis-idaea* L.) material for *ex situ* conservation. *Genetic Resources and Crop Evolution* 52: 723–735.

IV. Garkava-Gustavsson, L. & Nybom, H. 2007. Genetic diversity in a collection of apple (*Malus* x *domestica* Borkh.) cultivars as revealed by RAPD markers. *International Journal of Horticultural Science* 13: 1–11.

V. Garkava-Gustavsson, L., Kolodinska Brantestam, A., Sehic, J. & Nybom, H. Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis. *Hereditas*. In press.

Introduction

Germplasm, both wild and cultivated when available, is needed for immediate research and breeding, but should also be preserved for future use in *in situ* and/or *ex situ* conservation units. Maintenance of germplasm collections with a considerable amount of genetic diversity can therefore be regarded as a necessary prerequisite for present and future crop improvement.

In order to obtain and preserve useful germplasm, optimal sampling strategies must be developed based on overall diversity as well as presence of valuable genes. The amount, structure and distribution of genetic variability has traditionally been evaluated by the use of morphological characters, but molecular markers have become an important instrument for characterisation of wild and cultivated germplasm during the last decades.

The main objectives of my thesis were to demonstrate how genetic diversity estimators, obtained with different kinds of molecular markers, can be used in (1) development of sampling strategies, (2) creation of a gene bank for *ex situ* conservation and (3) genetic identification of cultivars in existing germplasm collections.

Crops studied: biology and utilization

Chaenomeles species

The genus *Chaenomeles* Lindl. (Rosaceae) comprises four diploid (2n = 34) species: *C. japonica* (Thunb.) Lindl. ex Spach (Japanese quince), *C. cathayensis* Hemsl. (Chinese quince), *C. speciosa* (Sweet) Nakai (Flowering quince) and *C. thibetica* Yu (Tibetan quince) (Phipps *et al.*, 1990). Plants within the genus *Chaenomeles* are moderate-sized shrubs, with abundant thorns terminating shoots. The flowers are nectar-rich but scentless, held in clusters of 1 to 6, and can vary in colour from white to darkest red through pink, orange and scarlet. All species are presumed to be outcrossing, and are often pollinated by honeybees and bumblebees. The fruits are pomes, very diverse in shape, colour and size, with numerous seeds (Rumpunen, 2001).

Chaenomeles japonica is a dwarf shrub, normally below 1.2 m, endemic to Japan. A strong self-incompatibility system prevails in this species, but artificial selfing is possible in some genotypes (Kaufmane & Rumpunen, 2002). The other three species, *C. cathayensis*, *C. speciosa* and *C. thibetica*, originate in Central Asia. *Chaenomeles speciosa* is a large shrub, 2 to 5 m, growing in Central and southern China, Tibet and Burma. There is also some evidence of self-

incompatibility in *C. speciosa* (Weber, 1964). *Chaenomeles cathayensis* is the largest shrub in the genus and it can reach up to 6 m in height. *Chaenomeles thibetica* is a smaller shrub, 1.5–3 m, growing in Tibet and western Sichuan. A mixed breeding system has been proposed for *C. cathayensis* and *C. thibetica* (paper II).

Interspecific hybridization produces fertile offspring and four hybridogenous taxa, the result of ornamental plant breeding, have been described: *C.* x *superba* (Frahm) Rehder (*C. japonica* x *C. speciosa*), *C. clarkiana* Weber (*C. cathayensis* x *C. japonica*), *C. vilmoriniana* Weber (*C. cathayensis* x *C. speciosa*) and *C. x californica* Clarke ex Weber (*C. cathayensis* x *C.* x *superba*) (Weber, 1964).

Chaenomeles species can easily be propagated by seeds, but cultivars must be propagated vegetatively, i.e. by grafting, or from root pieces, or hardwood and softwood cuttings, in order to maintain their characteristics (Rumpunen, 2001).

Chaenomeles has for a long time been appreciated both as an ornamental crop (Weber, 1964) and for medicinal purposes (Anonymous, 1989). Recently, members of this genus have also become domesticated for commercial fruit production in Europe and Asia. Several different species and species hybrids are thus grown for production of fruits with valuable aroma compounds and acidic juice in Latvia and Lithuania (Rumpunen *et al.*, 1998; Rumpunen, 2001) and China (Wang *et al.*, 1997, 1998). Fruits have been used for manufacturing of syrup, liqueur, carbonated soft drinks, marmalades and candies (Ruisa, 1996). Furthermore, a sugar-juice aroma extract has proven to be an excellent flavour in ice-cream (Rumpunen, 2001).

Many ornamental *Chaenomeles* cultivars have been successfully marketed. By contrast, only a few cultivars suitable for fruit production, are as yet commercially available. However, some superior genotypes, which may become registered as cultivars, have recently been selected in a North-European *Chaenomeles* breeding program based on heterogenous plants propagated by seed in orchards in the Baltic countries and on plants in breeding populations in Sweden (Rumpunen, 2001).

Lingonberry

Lingonberry, *Vaccinium vitis-idaea* (2n = 24), is a diploid, perennial, evergreen dwarf shrub, which belongs to the Ericaceae family. This species is represented by two subspecies: ssp. *vitis-idaea* L. and ssp. *minus* (Lodd.) Hultén (Hultén, 1949). The main difference between the two subspecies is plant size, with ssp. *minus* being considerably smaller in stature and leaf size. Lingonberry reproduces vegetatively through rhizomes (subterranean runners) and sexually by seeds. The species is partially self-fertile (Jacquemart & Thompson, 1996), but functions mainly as an outcrosser. It is pollinated by honeybees and bumblebees, and seed set is reported to be higher after cross-pollination compared to selfing (Gustavsson, 1999).

Lingonberry has ovate, dark green, thick and leathery leaves, and campanulate, white and nectariferous flowers, held in drooping racemes. Berries are bright red, 6

to 10 mm in diameter, and have a high antioxidant activity, which could potentially make this berry crop a candidate for cancer chemoprevention and treatment (Wang *et al.*, 2005). The berries are quite tart, and are therefore not well suited for fresh consumption. They do, however, taste very good when cooked, and are widely used for processing in jam, juice, liqueur and yoghurt. Lingonberry products are especially popular in the Nordic countries, but are becoming more accessible also in other parts of the world (Gustavsson, 1999).

Lingonberry is also popular for the decorative leaves, and branches are therefore harvested in the wild for making wreaths etc. In addition, plants are marketed for use in home gardens or for landscaping purposes.

Lingonberries destined for processing industries are usually harvested from plants growing wild in natural populations in coniferous forests and bogs, but there is an increasing interest in commercial plantations with registered cultivars. Still, development of commercial cultivars has been quite restricted and all the known cultivars (about 20) have been developed by selection of promising individuals either directly from wild stands or from cultivated populations obtained by open pollination. Four high-yielding and winter-hardy cultivars have been selected in Sweden: 'Sussi', 'Sanna', 'Ida' and 'Linnea'. For commercial purposes, lingonberry cultivars are propagated vegetatively mainly by stem cuttings (Gustavsson, 1999).

Apple

Apple, *Malus* x *domestica* Borkh., is the economically most important fruit crop in the temperate zone. Apple is thought to be an interspecific hybrid with *Malus sieversii* being the major maternal contributor (Harris *et al.*, 2002). Other species, which may have contributed to the lineage of cultivated apple are: *Malus orientalis* Uglitzk ex Juz., *Malus baccata* L., *Malus prunifolia* (Willd.) Borkh., and the European *Malus sylvestris* (L.) Mill. (Janick *et al.*, 1996). Apple has a gametophytic self-incompatibility mechanism, which enforces outbreeding and results in high levels of heterozygosity and genetic diversity (Kitahara *et al.*, 2005; Vanwynsberghe *et al.*, 2005). Numerous cultivars are known all over the world and the number is rapidly increasing (Sansavini *et al.*, 2004). Most of the apple cultivars are functional diploids (2n = 34), but spontaneously occurring triploid and even tetraploid cultivars are found.

Many apple varieties have arisen as chance seedlings, e.g., obtained via open pollination in fruit orchards, and therefore have an unknown origin. However, in the last century, an increasing number of cultivars have been derived from controlled crosses performed by plant breeders (Janick *et al.*, 1996). Naturally occurring mutations (bud-sports) in adapted cultivars are broadly exploited as well, and play a considerable role in the increasing number of new apple cultivars (Sansavini *et al.*, 2004). Apple cultivars are maintained by vegetative propagation and all trees belonging to the same cultivar should be genetically identical wherever they are grown.

In Sweden, modern apple breeding, based on controlled crosses, was first undertaken at Alnarp (1920-1960) and since around 1950 at Balsgård (both are nowadays parts of the Swedish University of Agricultural Sciences). Breeding and research at Balsgård rely on the largest apple germplasm collection in the country with about 1000 indigenous and foreign cultivars. This collection contains about 100 of the 220 mandate cultivars presently appointed for conservation by the 'National Program of Cultivated Plants'. The mandate status is granted to indigenous cultivars named, bred, propagated and marketed in Sweden as well as to some cultivars with a long history of being grown in the country (Hjalmarsson & Wallace, 2004).

Molecular markers used for detection of genetic diversity

Numerous molecular markers have been developed and applied for analyses of genetic diversity and relatedness, and all marker systems have their strengths and weaknesses (Schlötterer, 2004; Weising *et al.*, 2005). Choice of molecular marker system should be determined by the objectives of the study as well as mating system of the studied species, and available financial support.

Protein-based markers

Protein (enzyme- and non-enzyme) molecular markers provide indirect information about plant genome structure. Only one class of protein-based molecular markers, isozymes, were used in the diversity analyses within this thesis.

Isozymes

The term 'isozymes' was introduced by Markert & Moller (1959) and refers to protein forms of an enzyme with the same catalytic activity, and converting the same substrate, but differing in molecular weight or in electric charge. The differences in size and charge can be caused by amino-acid substitutions or post-translational modifications. Protein variants in isozyme analysis are distinguished by gel electrophoresis and visualized by an enzyme-specific staining mixture, where substrate, co-factor and an oxidized salt are included.

Isozymes are not necessarily products of the same gene, and they can be active at different life stages or in different cell compartments. Isozymes encoded by the same locus but by different alleles are usually referred to as allozymes (Weising *et al.*, 2005).

The main advantage of isozyme analysis, especially when allozymes are used, is their co-dominant inheritance, which allows the discrimination between homozygous and heterozygous genotypes. This is necessary for precise

estimations of population genetics parameters, especially in cross-pollinated species.

The main disadvantage is that isozyme analysis detects variation only in protein-coding loci and therefore provide fewer markers compared to DNA-based methods. Furthermore, isozyme systems often show low variability and in some cases no variability at all due to a low rate of mutational events.

Isozyme (allozyme) analysis has been widely used for analysing genetic diversity and population structure in a large range of plant species (Hamrick & Godt, 1989; Fady-Welterlen, 2005). The analyses are cheap and relatively easy to perform, which makes it possible to study an excessive amount of individuals. Although only one minor isozyme study had previously been published on *Chaenomeles* (Ponomarenko, 1990), we found this cost-effective method to be appropriate for diversity studies on *Chaenomeles* (papers I & II).

DNA-based markers

Of the great diversity of available DNA-based molecular markers, the following were used within this thesis: RAPD, SSR, SCAR and CAPS.

RAPD (Random Amplified Polymorphic DNA)

RAPD was first described by Williams *et al.* (1990). The method uses short arbitrary primers, usually 10 bp, to amplify anonymous PCR fragments in non-coding as well as coding regions of template DNA. The PCR products are separated by electrophoresis in an agarose gel. The gel is stained with ethidium bromide and the amplification products are visualized under UV light. A PCR product is generated when primers anneal in an inverted way and at an amplifiable distance from one another (150–3000 bp). The number of amplification products is related to the number and direction of the genome sequences complementary to the primer. RAPD polymorphism is caused by a nucleotide substitution within the target site, or an insertion or deletion of a DNA-fragment within the amplified region. These mutations result in an absence of a fragment or changes in fragment size (Weising *et al.*, 2005).

RAPD is a relatively cheap and technically simple method. It is especially suitable for analysing large numbers of samples belonging to species for which there is no genome sequence information. The main drawbacks of RAPD markers are possible problems with reproducibility (Jones *et al.*, 1997), co-migration of non-homologous fragments (Rieseberg, 1996), competitive priming (Halldén et al., 1996) and their dominant inheritance, i.e., heterozygotes cannot be distinguished from dominant homozygotes (Williams *et al.*, 1990).

Reproducibility problems within the same laboratory can be avoided by careful optimisation and adherence to a well-defined protocol, by screening large numbers of primers for reproducibility and by scoring only strong, clear and consistently appearing bands (Koller *et al.*, 1993; Mulcahy *et al.*, 1993; Belaj *et al.*, 2003). Co-migration of non-homologous fragments can be a problem at higher taxonomic

levels, but is probably negligible at the intraspecific level or when the studied species are closely related (Rieseberg, 1986). RAPD markers have, despite the above-mentioned limitations, been widely used in plant genetics and breeding, e.g., to estimate genetic diversity, detect phylogenetic relationships, and create genetic maps (Weising *et al.*, 1995, 2005).

Within this thesis, RAPD markers were used in papers II–IV to analyse genetic diversity and genetic relationships within rather large plant sets due to their technical simplicity and cost-effectiveness. They have previously proven to be useful in estimating genetic diversity in *Chaenomeles* (Bartish *et al.*, 1999), in estimating extent of clonality and genetic diversity in lingonberry populations (Persson & Gustavsson, 2001) and for cultivar identification, evaluation of genetic diversity and paternity analysis in apple (Harada *et al.*, 1993; Koller *et al.*, 1993; Zhou & Li, 2000).

SSR (Simple Sequence Repeats or microsatellites)

Microsatellites are short (1–6 bp) tandemly repeated sequences of DNA (Litt & Luty, 1989; Tautz, 1989). In microsatellite analysis, two primers are designed complementary to the sequences flanking a specific microsatellite DNA sequence. The sequence is amplified by PCR, fragments are separated by electrophoresis and allele sizes are determined. The allelelic polymorphism of microsatellite loci is usually caused by a variable number of integer repeat units.

Microsatellites are abundant in plant genomes, co-dominantly inherited, multi-allelic and highly polymorphic. Microsatellites have therefore become a marker system of choice for identification of genotypes, population genetics, paternity testing and genome mapping (Weising *et al.*, 2005).

Although SSR markers generally show highly reproducible results within and between laboratories, some discrepancies in allele sizing between laboratories have been reported, possibly caused by the many factors that can influence the migration of PCR fragments through the gel (Galli *et al.*, 2005; Amos *et al.*, 2007). The main drawback of microsatellite analysis is the high expenses at the initial stages of primer development, when sequences for primer design must be isolated and cloned. However, some cross-species transferability of primer pairs (Yamamoto *et al.*, 2001, 2004) or software-designed primers based on publicly accessible sequence information from gene-bank accessions (Tonguc & Griffiths, 2004) can decrease the initial costs considerably.

SSR analysis, used for analysis of genetic diversity, genetic relatedness and genotype identification in paper V, has previously been successfully applied in apple for i.e., cultivar identification (Galli *et al.*, 2005), evaluation of genetic diversity in germplasm collections (Guarino *et al.*, 2006; Pereira-Lorenzo *et al.*, 2007) and for parent identification (Kitahara *et al.*, 2005). SSR markers are superior to RAPD concerning the reproducibility between laboratories, which is extremely important in multilateral international projects. Still, the expences and technical facilities required for SSR analysis, restricted the investigations in Paper V to a smaller set of highly prioritized cultivars.

SCARs (Sequence-Characterized Amplified Regions)

SCARs were originally developed from RAPD-derived markers (Paran & Michelmore, 1993), but have later also been developed from AFLP-derived markers (Mienie *et al*., 2002). A DNA fragment in an agarose gel, tightly linked to the character of interest, is cloned and sequenced, and approx. 24 bp long primers, complementary to the ends of the original fragment, are designed. These primers can then be used in amplification of a single marker for the character of interest. SCAR markers demand stringent PCR conditions and thus have no problems with reproducibility. The initial stages of development of SCAR markers are laborious, but once developed they are user-friendly since the PCR-fragments can be detected in an agarose gel with ethidium bromide. Apple cultivars have been characterized with SCAR markers to detect alleles that control apple skin colour (Cheng *et al.*, 1996), incompatibility (S-alleles) (Broothaerts, 2003) and resistance against apple scab caused by the fungus *Venturia inaequalis* (Hemmat *et al.*, 2002, Mattisson & Nybom, 2005; Melounova *et al*., 2005).

SCAR markers were used in paper V for analysis of S-locus alleles. Apple cultivars which share one S-allele, have a reduced yield capacity if cross-pollinated, while those with both alleles in common produce almost no fruit at all.

PCR-RFLP or CAPS (Cleaved Amplified Polymorphic Sequences)

CAPS markers (Konieczny & Ausubel, 1993) are generated in two steps: PCR fragments are first amplified by a sequence-specific primer pair, and subsequently digested with a restriction enzyme. The resulting fragments are separated by gel electrophoresis. Mutations in the restriction sites can prevent the restriction and thus fragments of differing size are obtained. CAPS markers are codominantly inherited and have been used for detection of scab resistance controlled by the Vf gene (Mattisson & Nybom, 2005) and for characterisation of the S-locus in apple (Matsumoto *et al.*, 1999; paper V)

Correlation between methods

Different molecular markers, used for the same plant material, may reveal either similar or different patterns of diversity and genetic relationships. Thus, good correlation and congruence between isoenzymes and RAPD has been observed in apple (Royo & Itoiz, 2004) and in black spruce (*Picea mariana*) (Isabel *et al.*, 1995) but moderately low in oats (*Avena sterilis*) (Heun *et al.*, 1994). Significant correlations were obtained in soybean (*Glycine*) when RFLP, RAPD, SSR and AFLP data were compared, but correlation coefficients involving the RAPD data were lower than those obtained with the other polymorphic assays (Powell *et al.*, 1996). A similarly low correspondence between values was obtained for RAPD and SSR data in cultivated potato (*Solanum tuberosum*) (Milbourne *et al*., 1997). According to Powell *et al.* (1996), several factors may affect the estimation of genetic relationships between individuals, namely number of markers used, distribution of markers in the genome and the nature of evolutionary mechanisms underlying the variation measured. Weising *et al.* (2005) stress the fact, that four

to 10 times more markers are needed when using dominant instead of co-dominant markers in order to achieve the same accuracy in the estimation of genetic distances, and the estimation of within- and among-population variation.

Materials and Methods

Plant material

Highly diverse germplasm collections have been gathered at SLU, Balsgård, to provide a basis for breeding and research in *Chaenomeles*, lingonberry and apple. The plant material investigated within this thesis, derives mainly from these collections.

In paper I, we analysed 42 *Chaenomeles* plants, from 14 accessions representing three species, *C. japonica*, *C. speciosa*, *C. cathayensis* and a hybridogenous taxon, *C. x superba*. The plants were raised from seeds obtained from botanical gardens in Europe and Japan and from domesticated populations in the Baltic countries.

In paper II, we studied 185 *Chaenomeles* genotypes from nine wild Chinese and Japanese populations representing all four known spontaneous species: *C. japonica* (3 populations), *C. speciosa* (3 populations), *C. cathayensis* (2 populations), and *C. thibetica* (1 population). All but three populations (two of *C. speciosa* and one of *C. cathayensis*) were collected directly in the wild. The remaining three were derived from fruits purchased at the local markets, but originally collected in near-by mountains. Only one fruit was picked from each shrub and the seeds from each fruit constituted a family. The analysed families were taken randomly from the different populations.

The lingonberry plant material, analysed in paper III consisted of 211 plants derived from 15 geographically dispersed locations. At each location, berries were collected within an area of 1000 square meters. Plants raised from seeds from the same location were considered as a population. Ten to 15 plants from each population were taken randomly for the analyses. The analysed populations represent two subspecies of *Vaccinium vitis-idaea*: ssp. *vitis-idaea* and ssp. *minus.* In addition, 8 plants representing 8 commercial cultivars were analysed, increasing the total material to 219 plants.

In paper IV, 151 apple cultivars were analysed. These cultivars were divided into six groups based on their historical age and their geographic origination derived from pomological literature (Dahl, 1929; Nilsson, 1986; Svensson & Kastman, 2005): 1) ancient (originated before 1800) Swedish (17), 2) ancient foreign (21), 3) old (1800–1900) Swedish (34), 4) old foreign (29), 5) new (after 1900) Swedish (31) and 6) new foreign (19). Of these, 94 were mandate cultivars, appointed for conservation in Sweden, with 68 having a Swedish origin.

The 68 indigenous Swedish apple mandate cultivars were further analysed in paper V. In addition, a few non-mandate parents and some reference cultivars were

analysed in order to verify the pedigree information and to compare our allele-sizing data with previously published results.

Methods and data scoring

Isozyme analysis

Protein extraction

Enzymes were extracted with extraction buffer from fresh phloem tissue or leaves, that were ground in pre-cooled mortars with liquid nitrogen. A detailed extraction protocol is described in paper I. The same extraction protocol was used also in paper II.

Electrophoresis and staining

In both studies, papers I & II, isozymes were separated by electrophoresis in a gradient (5.8–11.8%) (Boscovic et al., 1994) or in 7.5% (Davis, 1964) polyacrylamide gels (described in detail in paper I).

A broad range of isozyme systems (in total 16) were originally assayed in paper I. Six of these, namely acid phosphatase (ACP), esterases (EST), glutamate oxaloacetate transaminase (GOT), malate dehydrogenase (MDH), phosphoglucomutase (PGM) and peroxidase (PRX) showed sufficient resolution and consistent banding patterns and were therefore used in diversity analyses. The same enzyme systems were applied in paper II, except for MDH, which was substituted by shikimate dehydrogenase (SKDH).

Bands for each isozyme system were scored as unordered multistate traits (Beer *et al.*, 1993; Strefeler *et al.*, 1996a, b) in both studies, and binary matrices with "1" for presence and "0" for absence were produced.

DNA-marker analysis

DNA-extraction

In all four studies with DNA-based markers (papers II–V), total cellular DNA was extracted from leaves. In paper II, DNA was extracted following a previously described protocol (Bartish *et al.*, 1999) with some modifications. In paper III, the extraction protocol of Nybom & Shaal (1990) was used. In papers IV and V, DNA was extracted using the Qiagen Dneasy™ Plant Mini Kit, following the manufacturer's protocol.

RAPD analysis

Oligonucleotide primers used in papers II–IV, were selected from the Operon Technologies sets based on polymorphism, reproducibility and clarity of the obtained patterns. PCR amplifications in paper II (*Chaenomeles*) were performed following Bartish *et al.* (1999) and in paper III (lingonberry) Persson & Gustavsson (2001). PCR protocols for apple are presented in detail in paper IV. In all studies, RAPD bands were separated in 1.8% agarose gels, stained by ethidium

bromide and scored manually. Binary data matrices were produced, based on presence “1” or absence “0” of a band.

SSR analysis

Ten SSR primer pairs from a list of 140 (Liebhard *et al.*, 2002) were used. Nine of these, CH01d03, CH01h02, CH02c06, CH02c09, CH02c11, CH02d08, CH04c06, CH04e05 and COL belong to the “standard primer set” defined by the European working group on apple genetic resources (Laurens *et al.*, 2004; Guarino *et al.*, 2006). The tenth primer pair (CH02b10) had been used previously at our laboratory (Mattisson & Nybom, 2005).

All PCR amplifications were performed according to Gianfranceschi *et al.* (1998) and the originally published PCR program was used for all but one primer pair, Ch02c06, for which a detailed cycling profile is given in paper V.

The PCR products were separated and analysed by capillary electrophoresis. Allele sizes were assigned to their repeat units equivalent. The amplified SSR fragments were scored in terms of loci and alleles. In addition, the fragments were scored phenotypically, with “1” for presence of fragment and “0” for absence.

SCAR and CAPS analyses.

In paper V, SCAR and CAPS markers were used to fingerprint apple cultivars using alleles at the S-locus. Allele-specific primers published by Matsumoto *et al.* (1999), Kitahara & Matsumoto (2002) and Broothaerts (2003) were used for the detection of a total of 14 different S-alleles. A detailed description of PCR conditions, amplification protocols and enzyme restrictions is given in paper V. The amplified and digested fragments were separated on a 2.5% agarose gel, stained with ethidium bromide and scored manually.

Flow cytometry

Ploidy level of apple cultivars (paper V), that amplified more than two alleles in an SSR and/or S-locus, was estimated by flow cytometry analysis, performed by Plant Cytometry Services (JG Schijndel, The Netherlands). Leaf samples were chopped in icecold buffer with DAPI and than analysed on a CyFlow ML using *Lactuca sativa* as an internal standard.

Statistics

Genetic relatedness (Cluster analysis, MDS, PCO)

Levels of genetic relatedness among the analysed individuals (papers I, III, IV & V) were estimated by comparing molecular marker-derived band patterns using Jaccard’s coefficient of similarity (Jaccard, 1908). Jaccard’s coefficient excludes the negative band matches, which is especially important for RAPD markers, since the absence of a band can be caused by several different events (Weising *et al.*, 2005). Dissimilarities among the analysed plants (paper V) were measured by a commonly used measure of dissimilarity, Roger’s coefficient (Reif *et al.*, 2005).

Similarities among populations (paper III) were estimated by a COSINE pattern similarity measure (Norusis, 1990). Phenetic distances among populations and among families were estimated with the Euclidean distance, while genetic distances among populations were calculated according to Nei (1972).

Matrices based on similarities or dissimilarities were subjected to, e.g., cluster analysis (papers I–V), multidimensional scaling (MDS) (papers I, III) or principal co-ordinate analysis (PCO) (papers II, IV).

Cluster analysis produces a hierarchical representation of the relationships among samples or groups of samples. The most commonly applied method for DNA marker-based cluster analysis is UPGMA (Sneath & Sokal, 1973), which was used also in my thesis. The levels of relatedness were visualised in a dendrogram. Cluster analysis generally provides a good resolution among closely related individuals and groups, but clusters are produced whether or not natural groupings exist. Compared to cluster analysis, the ordination methods provide a better overview of the relationships between groups. The two most commonly used methods with DNA-based data, MDS and PCO, were used in this thesis. MDS and PCO represent the relationships among individuals or groups of individuals in a non-hierarchical way. These relationships were illustrated in two- (papers I, III and IV) or three-dimensional plots (paper II). MDS generally reveals a larger proportion of the variability already in the first two dimensions (Weising *et al.*, 2005).

Genetic diversity estimators

Genetic diversity within a population or a group of, e.g., cultivars was evaluated by several different parameters in my thesis: alleles per locus (A), effective number of alleles per locus (A_e), expected heterozygosity (H_e), i.e., the probability that two alleles from the same locus would be different when chosen at random (Nei, 1972), and observed or direct count heterozygosity (H_o). For dominant markers, i.e., RAPD, the formulae for calculating Nei's unbiased gene diversity (Nei, 1978) have been modified by Lynch & Milligan (1994). Assumption of a Hardy-Weinberg equilibrium at each locus is required when calculating Nei's gene diversity or the Lynch & Milligan index for dominant, biallelic data in outcrossing species, where heterozygotes are common but cannot be detected (Lynch & Milligan, 1994). Another commonly used measure of within-population variability, applied in the thesis, is Shannon's index (Bussell, 1999), which does not require an assumption about Hardy-Weinberg equilibrium.

To verify the potential of chosen markers to discriminate among analysed cultivars, two measures were used in paper V: power of discrimination (PD), which estimates the probability that two randomly sampled individuals are discriminated, and a confusion probability (C), which is the probability that any two cultivars are identical in their molecular genotypes at all loci by chance alone (Kloosterman *et al.*, 1993; Aranzana *et al.*, 2003).

Partitioning of diversity

Population subdivision causes an inbreeding-like effect, which can be measured in terms of a decrease in the proportion of heterozygous genotypes. The effect of population subdivision is quantified by the fixation index, Wright's F_{ST} (Hartl, 1988). The fixation index can be calculated for different hierarchical levels, between subpopulations within populations, between populations within regions etc., using a formula:

$$F_{ST} = (H_T - H_S)/H_T,$$

where H_T is the expected heterozygosity in the total population (species) and H_S is the mean expected heterozygosity in the subpopulations (populations). Nei's coefficient of gene differentiation (G_{ST}) is equivalent to F_{ST}, but analysis is extended to multiple loci and any number of alleles per locus (Weising *et al.*, 2005).

Population differentiation is often estimated with the Analysis of Molecular Variance (AMOVA) approach (Excoffier *et al.*, 1992). AMOVA was applied in my thesis to calculate the proportion of variability among populations within a species, among populations within a genus and among species within a genus (paper II), and to estimate the partitioning of diversity within and between populations for three different data sets (paper III).

The AMOVA-derived Φ_{ST} estimators have previously been found to provide values which are very similar to those obtained with G_{ST} estimators when calculated on the same raw data (Nybom & Bartish, 2000; Nybom, 2004).

Components of diversity within and between species, populations and families in Paper II, within and between populations in paper III, and within and between predefined groups of cultivars in paper IV, were also calculated based on the Shannon's diversity index according to Bussell (1999).

Summary of Results and Discussion

The amount and structure of genetic variability in germplasm collections is highly dependent on the biological properties of the crop in question, as well as on its history of domestication. The studies presented in this thesis focus on exploration of genetic diversity in three horticultural crops: *Chaenomeles*, lingonberry and apple.

All the investigated crops have different histories of domestication. Thus, species within the genus *Chaenomeles* have been known for at least 500 years, but they were first domesticated for their ornamental value, and have been cultivated for fruit production only during the last 50–60 years (Rumpunen, 2001). Lingonberry has a very brief history of domestication and has been cultivated only for a few decades (Gustavsson, 1999). Apple has been used as a fruit crop by man for at least 2000 years (Harris *et al.*, 2002) and is today the economically most important fruit crop in Sweden.

Genetic diversity in *Chaenomeles* (papers I & II)

Genetic diversity within cultivated plant material

The specific objectives of paper I were to investigate genetic diversity among taxa, accessions, and individual *Chaenomeles* plants using isozymes.

Results obtained for 42 cultivated *Chaenomeles* plants, representing three species, *C. cathayensis*, *C. japonica* and *C. speciosa*, and a hybridogenous taxon, *C.* x *superba*, showed that *C. x superba* and one of its ancestor species, *C. speciosa*, were more similar to each other than to any of the other species. Individuals of these taxa jointly formed one group in an MDS plot (Fig. 1) and clustered together in a UPGMA dendrogram (Fig. 2). The other ancestor, *C. japonica*, was clearly differentiated from *C. x superba*. Furthermore, *C. japonica* and *C. cathayensis* formed well-separated groups in both a dendrogram and an MDS plot, and were therefore considered to be the most distantly related species. Inside the major clusters, genotypes grouped mainly according to the geographic origin of plant material. The individuals of *C. cathayensis*, obtained from three different locations, formed the most tight cluster in both dendrogram and MDS representations, indicating a comparatively low level of diversity within this species.

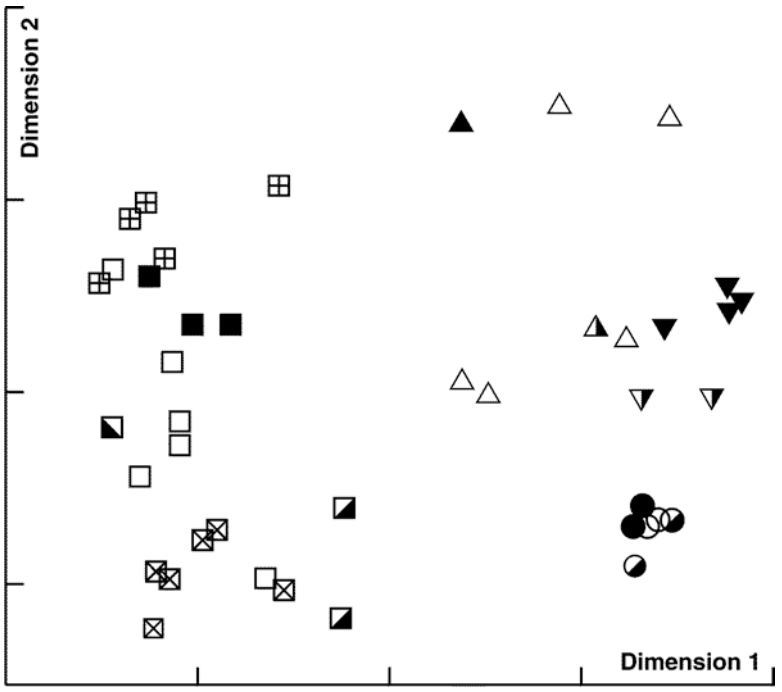


Figure 1. Plot of MDS analysis of genetic similarity estimates (Jaccard's coefficient) for isozyme data, showing groups among individual plants of *C. cathayensis* (O), *C. speciosa* (∇), *C. x superba* (Δ) and *C. japonica* (□) (Figure 2, paper I).

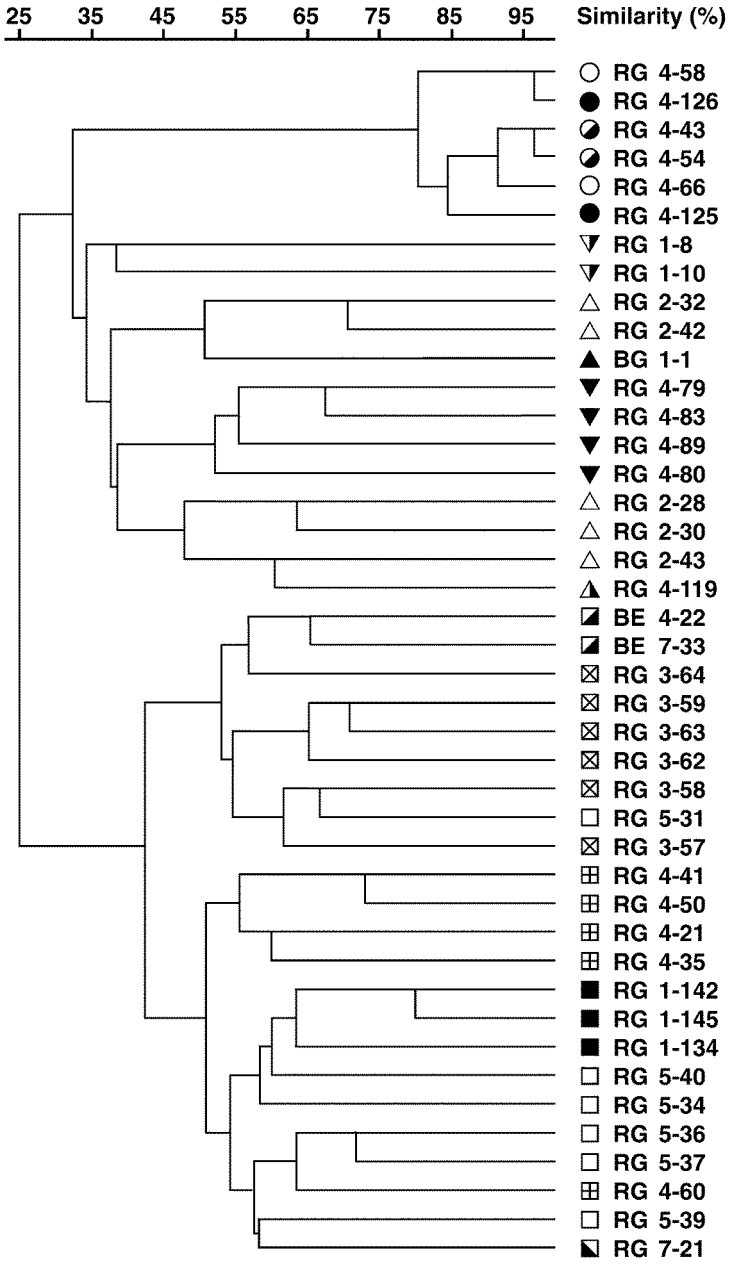


Figure 2. Dendrogram based on UPGMA analysis of genetic similarity estimates (Jaccard's coefficient) for isozyme data, showing relationships among individual plants of *C. cathayensis* (O), *C. speciosa* (∇), *C.* x *superba* (Δ) and *C. japonica* (□) (Figure 1, paper I).

Relationships revealed by isozyme analysis in this study, were in general agreement with morphological data for the same species (Weber, 1964) and RAPD data for the same set of 42 individuals (Bartish *et al.*, 1999). Correlation (Mantel test) between similarity matrices based on isozymes (paper I) and RAPD (Bartish *et al.*, 1999), was high (r = 0.74).

Genetic diversity within wild plant material

Unfortunately, there are some problems with using material obtained from botanical gardens, like uncertain origins and the possibility that spontaneous intra- and interspecific hybridization may have taken place, potentially having an impact on diversity estimators.

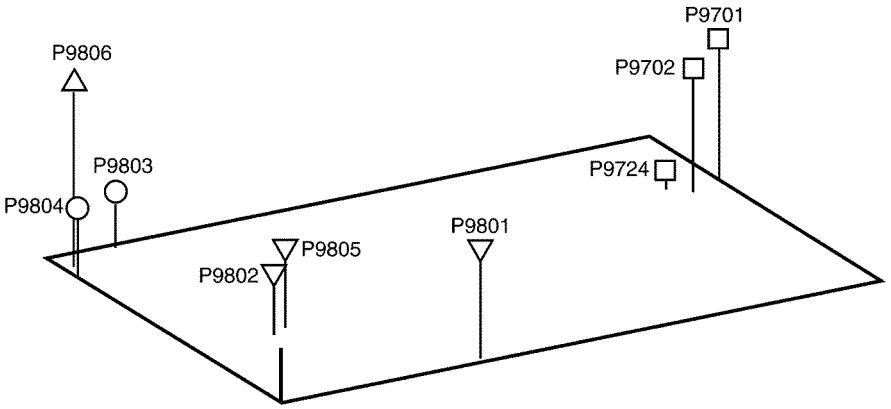
Further investigations were therefore undertaken using wild plant material. In paper II, RAPD and isozymes were used to determine genetic diversity and population structure for species and populations within the genus *Chaenomeles* and to estimate level of correspondence between the two marker systems.

Phenetic relationships within *Chaenomeles* assessed in paper II for 9 native populations representing four known species, were congruent with the results obtained in previous investigations on cultivated plant material (Bartish *et al.*, 1999; paper I). Thus, *C. japonica* was clearly differentiated from both *C. speciosa* and *C. cathayensis* (Fig. 3). The recently recognized species *C. thibetica* appeared to be rather closely related to *C. cathayensis*.

Genetic diversity, estimated by Shannon's index and the Lynch and Milligan index, indicated that populations of *C. japonica* and *C. speciosa* were considerably more diverse than populations of *C. cathayensis* and *C. thibetica* (Table 1). Correspondingly, the main part of the total variability resided within populations in the case of *C. japonica* and *C. speciosa* and between populations in the case of *C. cathayensis*. Differences in mating system among the species could be a possible explanation to these results, and a mixed mating system was thus suggested for the two most uniform species, *C. cathayensis* and *C. thibetica*.

Highly significant correlations between isozyme and RAPD data were observed for phenetic distances (r = 0.889, P < 0.001) and gene diversity estimates, although RAPD-based estimates were somewhat higher. Isozyme markers did, however, show a markedly higher rate of differentiation between populations of *C. japonica* compared to the RAPD-derived data. This discrepancy may be explained by a higher proportion of non-neutral markers for isozymes than for RAPD, thus some isozyme markers may have been subjected to diversifying selection.

a) RAPDs



b) isozymes

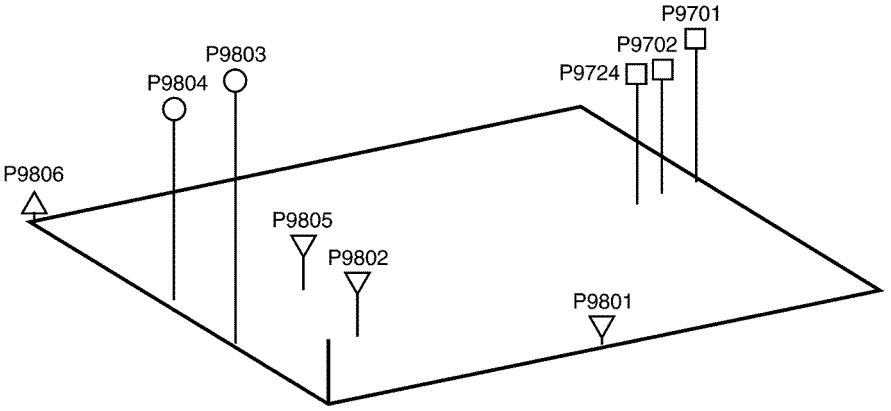


Figure 3a, b. Three-dimensional representation of principal co-ordinate analysis of phenetic relationships between populations of *Chaenomeles*. (Figure 3a, b, paper II).

Table 1. *Gene diversity estimated by Shannon's index (H') and by the Lynch and Milligan index (H) with standard error, as well as the percentage polymorphic markers (P) within populations and species of* Chaenomeles, *based on RAPD and isozyme data (Table 2, paper II).*

Group	H'_{pop} RAPD	H'_{pop}isozyme	H_{pop} RAPD	*P*(%) RAPD	*P*(%) isozyme
C.cathayensis					
P9803	0.140(0.024)	0.069(0.025)	0.066(0.012)	15.8	7.1
P9804	0.282(0.032)	0.081(0.026)	0.139(0.017)	30.3	8.9
C. japonica					
P9701	0.487(0.032)	0.422(0.047)	0.229(0.017)	54.6	50.9
P9702	0.467(0.032)	0.375(0.050)	0.212(0.016)	50.9	44.6
P9724	0.424(0.031)	0.413(0.050)	0.196(0.016)	49.1	46.4
C. speciosa					
P9801	0.547(0.030)	0.503(0.049)	0.253(0.015)	63.5	54.5
P9802	0.520(0.030)	0.419(0.044)	0.239(0.014)	62.4	53.6
P9805	0.589(0.030)	0.460(0.044)	0.258(0.014)	66.1	60.7
C. thibetica					
P9806	0.286(0.029)	0.189(0.052)	0.141(0.015)	34.3	14.3
Group	H'_{sp}RAPD	H'_{sp}isozyme	H_{sp}RAPD	*P*(%) RAPD	*P* (%) isozyme
C.cathayensis	0.337(0.032)	0.284(0.056)	0.160(0.015)	39.9	22.3
C. speciosa	0.620(0.032)	0.612(0.050)	0.278(0.016)	69.0	67.9
C. japonica	0.706(0.029)	0.655(0.043)	0.306(0.014)	78.2	75.9
Group	H'_{ge} RAPD	H'_{ge} isozyme	H_{ge} RAPD	*P*(%) RAPD	*P* (%) isozyme
Chaenomeles	0.746(0.013)	0.665(0.024)	0.282(0.008)	93.7	92.0

Genetic diversity in lingonberry (paper III)

Lingonberry breeding programs generally rely mainly on germplasm derived from natural populations. Therefore, knowledge about amount and structure of genetic variation within and between natural populations is very important for collecting and preserving genetic resources of this crop.

In paper III, the main purpose was to assess relatedness and genetic diversity for 15 recently collected lingonberry populations (from Sweden, Finland, Norway, Estonia, Russia, Japan and Canada). All but two populations belonged to ssp. *vitis-idaea* and the remaining two, from Canada and Japan, represented ssp. *minus*. Since the lingonberry breeding program at Balsgård had been temporarily halted, we also aimed to select a representative set of individuals from the analysed populations for *ex situ* conservation and future use.

Genetic diversity and relatedness

RAPD-based cluster analysis and MDS produced similar phenetic patterns, and a pronounced geographic grouping among populations was observed in most cases (Fig. 4; Fig. 5). Genetic and geographic distances were significantly correlated (Mantel test) for the entire plant material ($r = 0.532$, $P < 0.001$) and for the 13 populations of ssp. *vitis-idaea* ($r = 0.475$, $P < 0.001$) indicating isolation-by-distance. When the analysis was restricted to only Swedish populations, the correlation was non-significant ($r = 0.182$, $P = 0.320$).

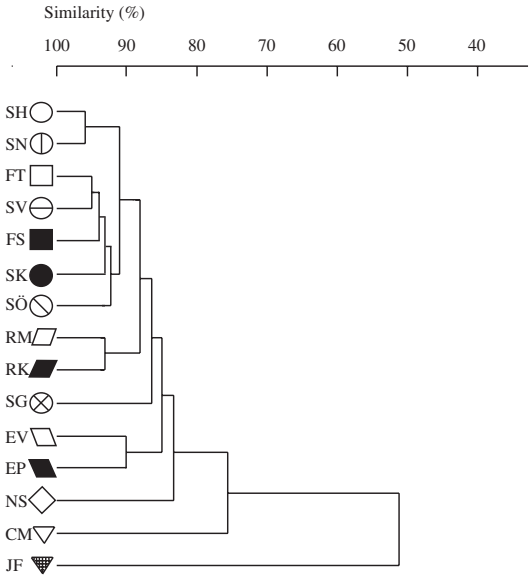


Figure 4. Dendrogram based on UPGMA analysis of genetic similarity estimated among 15 populations of lingonberry (Figure 1, paper III).

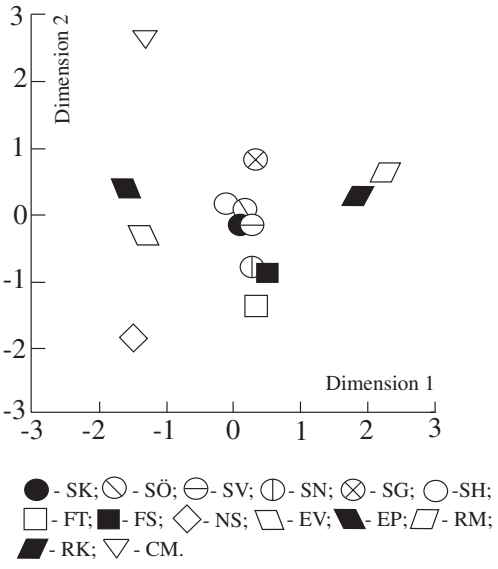


Figure 5. Two-dimensional plot of MDS analysis based on a genetic distance matrix, illustrating genetic relationships among 14 populations of lingonberry. The Japanese population of *Vaccinium vitis-idaea* ssp. *minus*, an outlier, is not included (Figure 2, paper III).

Levels of within-population genetic diversity are presented in Table 2. The mean value for within-population diversity (H_{pop} = 0.206) is in general agreement with the mixed mating system, reported for lingonberry (Jacquemart & Thompson, 1996) and with life history traits (long-lived perennials, dispersed by animal-ingested seeds) (Nybom & Bartish, 2000). The major part of the diversity was found within populations, which is in agreement with previous RAPD-based data in a study of four lingonberry populations (Persson & Gustavsson, 2001).

Table 2. *Genetic diversity (including standard error) estimated by Lynch and Milligan's index (Hpop) and Shannon's index (H'pop) within populations of lingonberry (modified Table 1, paper III).*

Sampling site no.	Sampling site code	No. of plants	Country	Location	H*pop*	H'*pop*
1	SK	15	Sweden	Kristianstad	0.227 (0.026)	0.483 (0.049)
2	SÖ	15	Sweden	Örebro	0.214 (0.025)	0.517 (0.051)
3	SV	15	Sweden	Västerbotten	0.245 (0.027)	0.513 (0.050)
4	SN	10	Sweden	Norrbotten	0.197 (0.026)	0.400 (0.051)
5	SG	13	Sweden	Gävleborg	0.248 (0.028)	0.523 (0.048)
6	SH	11	Sweden	Halland	0.219 (0.026)	0.500 (0.051)
7	FT	15	Finland	Toijala	0.187 (0.026)	0.375 (0.049)
8	FS	15	Finland	Simo	0.178 (0.026)	0.374 (0.050)
9	NS	15	Norway	Sognedal	0.225 (0.029)	0.434 (0.055)
10	EV	15	Estonia	Vöru	0.241 (0.029)	0.456 (0.055)
11	EP	10	Estonia	Pärnu	0.209 (0.027)	0.434 (0.055)
12	RM	15	Russia	Murmansk	0.110 (0.024)	0.190 (0.043)
13	RK	14	Russia	Kirov	0.135 (0.024)	0.265 (0.047)
14	JF	14	Japan	Fuji San	0.180 (0.028)	0.349 (0.056)
15	CM	15	Canada	Montreal	0.274 (0.025)	0.654 (0.048)
					x = 0.206	x = 0.431

Selection of individuals for ex situ conservation

To create a new lingonberry genebank at Balsgård, a set of 61 individuals was selected out of a total of 211 plants derived from wild populations. One plant of each of 8 commercially available cultivars were added to the collection. Two different approaches, previously developed for the establishment of core collections, were compared: a hierarchical sampling strategy (Parsons *et al.*, 1999) and the Maximum genetic diversity program (Marita *et al.*, 2000). In addition, randomly sampled datasets were created for comparisons with the two systematically selected data sets.

Both of the systematically selected datasets were significantly more diverse than randomly chosen sets (Fridman's test, $p < 0.001$), but the most diverse set of lingonberry specimens ($p < 0.001$) was obtained when using the Maximum diversity program (Table 3). Furthermore, only the subset selected with the Maximum diversity program retained all RAPD bands found in the entire set.

Table 3. *Genetic diversity in the entire plant material (219 accessions) and subsamples (69 accessions) obtained by different strategies (modified Table 4, paper III).*

	Nei's gene diversity	Shannon's index
Entire material	0.300	0.670
Randomly chosen collection (3 data sets)	0.301–0.304	0.663–0.669
Hierarchical sampling strategy-based collection	0.322***	0.698***
Maximum diversity-based collection	0.356***	0.759***

Moreover, most of the bands with frequencies above 0.5 in the entire collection showed decreased frequencies in the subset, while bands with frequencies below 0.5 increased instead (Fig. 6). Plants with rare RAPD bands should rather be over-represented than under-represented in a gene bank since they may carry rare genes, which can be valuable in plant breeding.

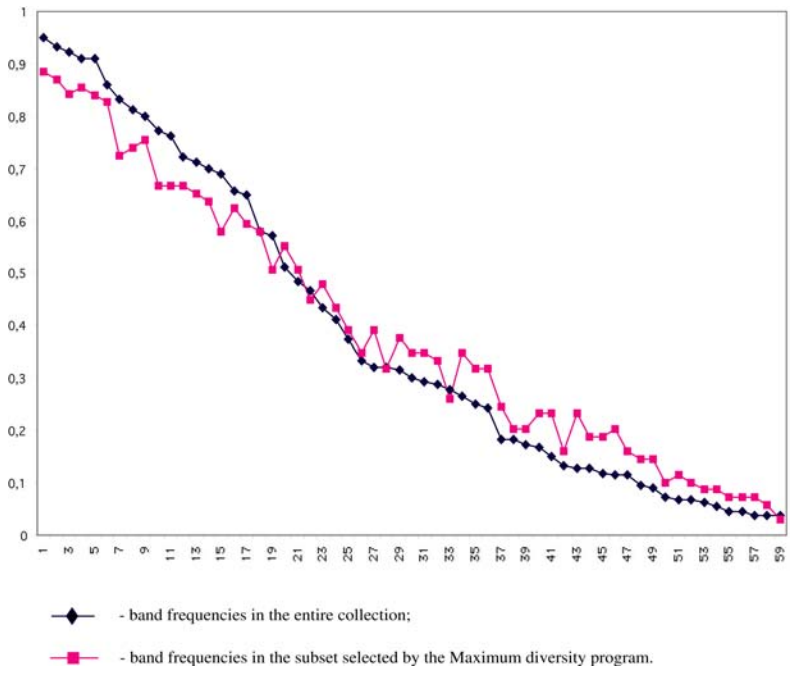


Figure 6. RAPD band frequencies in the entire collection and in the subset obtained by using the Maximum genetic diversity program (Figure 3, paper III).

Genetic diversity and molecular characterisation of apple cultivars

The apple breeding and research program at Balsgård is dependent on large germplasm collections with properly identified accessions. We therefore used RAPD, SSR and the S-locus in order to provide some cultivars with DNA-based fingerprints. In Paper IV, RAPD markers were used to discriminate among cultivars, detect duplicates and mis-labellings and identify genetic relationships among cultivars. We also analysed possible differences in inter-cultivar diversity caused by varying selection pressures over time and/or in different geographic regions. In addition, we compared the genetic diversity within the subset of Swedish mandate cultivars with the diversity in the entire set.

In paper V, the specific objectives were to fingerprint 68 mandate cultivars, originating from Sweden, using the allelic compositions at some SSR-loci and the S-locus, as well as to evaluate the diversity and relationships within this set.

DNA-based cultivar identification

RAPD and SSR markers allowed to determine if cultivars had arisen through genetic recombination or as sports. In both studies, papers IV and V, all cultivars presumed to have derived via sexual recombination were distinguished, while the known sports had identical molecular profiles with their progenitors in all cases in paper IV, and in all but one case in paper V.

Based on molecular marker profiles, we could clarify the case with ‘Spässerud’ (first described in 1903) and ‘Särsö’ (first described in 1917), which have long been known to be very similar (Nilsson, 1986). These two cultivars showed identical molecular profiles (RAPD, SSR and S-locus) and since the fruit flesh colour is reported to be somewhat different, we concluded that one of the cultivars has been derived as a sport of the other. Furthermore, one obvious case of mis-labelling was revealed in paper IV: the tree, labelled ‘Grågylling from Skokloster’ was instead ‘Alexander’. A similar case had previously been revealed in a small RAPD-based study: a tree, labelled as ‘Astrakan, White’ was instead ‘Astrakan, Stor Klar’ (Garkava-Gustavsson & Nybom, 2003). The nature of band pattern identity was checked for seven pairs of apple cultivars by Hokansson et al. (1998) using SSR markers; five pairs contained sport mutations and/or their progenitors, in one pair one tree was mis-labelled, and in the last pair, one tree probably had a synonymous name or was a sport mutation.

In paper V, 68 indigenous Swedish cultivars were provided with DNA fingerprints using the allelic compositions in 11 polymorphic SSR loci and the S-locus. Seven cultivars: ‘Frösåker’, ‘Holländaräpple’, ‘Kalmar Glasäpple’, ‘Kinnekulle Kantäpple’, ‘Norrstack’, ‘Villands Glasäpple’ and ‘Vrams Järnäpple’ were found to be triploid according to the results of SSR and S-locus analysis, and flow cytometry. The previously known ancestry was confirmed for eight cultivars: ‘Alfa 68’, ‘Alice’, ‘Aroma’, ‘Eva-Lotta’, ‘Katja’, ‘Kim’, ‘Mio’ and ‘Sylvia’.

SSR allele compositions for reference cultivars (paper V) were compared with previously published studies and a few discrepancies in allelic composition, and

several differences in allele sizes were revealed. These findings can be explained either by appearance of minor somatic mutations, since all studies have been conducted on different trees, or by experimental artefacts. Experimental artefacts can be minimized by registering relative size of alleles instead of absolute (Galli *et al.*, 2005; Amos *et al.*, 2007), by use of standard genotypes selected to display a diverse allelic range (Cryer *et al.*, 2006), and by use of cloned sequence-specific internal DNA standards. The latter is widely used in human clinic and forensic studies (Bruland *et al.*, 1999; Leclair *et al.*, 2004) and thus can potentially eliminate the problems with allele sizing in international plant research projects.

Pattern of relatedness

A RAPD-based cluster analysis and a PCO (Fig. 7) did not reveal any clear geographic patterns, even though we observed a slight tendency for Swedish and foreign cultivars to differentiate (paper IV). Similarly, an SSR-based dendrogram failed to produce well-defined clusters in paper V. A corresponding lack of association between geographical origin and clustering in a dendrogram has also been reported in other crops (Monte-Corvo *et al.*, 2000; Vijayan *et al.*, 2005). Our findings indicate that Swedish apple cultivars represent a more or less random set of molecular phenotypes and form a broad gene pool possibly due mainly to a high level of heterozygosity, numerous crosspollinations between genotypes, extensive exchange of plant material between countries and plant breeding programs, and a slow turnover of popular genotypes.

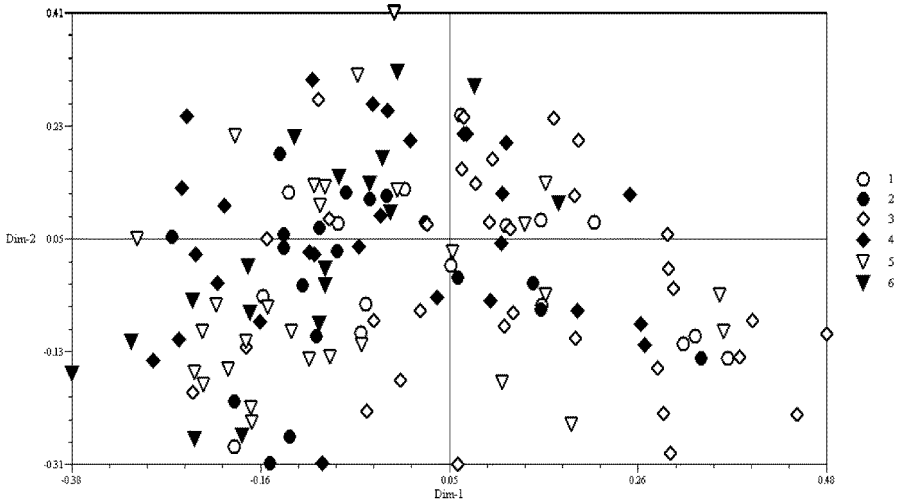


Figure 7. A two-dimensional plot of PCO analysis of 143 individual apple genotypes belonging to different historical and geographical groups (all duplicates are removed) (Figure 2, paper IV).

On the other hand, many cultivars clustered together with one of their known parents or siblings (paper IV), thus suggesting that closely clustering cultivars can be regarded as genetically more similar than average cultivars. Thus, in paper V,

comparison of allele profiles of two closely clustering cultivars, 'Hanaskog' and 'Oranie' suggested that 'Hanaskog' could be an offspring of the presumably much older 'Oranie'. The information on genetic relatedness can be useful in character screenings: only the most promising genotypes can be targeted instead of screening all available accessions.

Genetic diversity

Overall genetic diversity (paper IV), estimated with Nei's diversity index, H = 0.269 and with Shannon's index H' = 0.594 based on RAPD data, indicate that apple holds average levels of genetic diversity compared to other outcrossing species (Nybom & Bartish, 2000; paper II). No major changes in diversity were observed over time and space (paper IV), although the group with recent foreign cultivars had a significantly lower diversity value when calculated by Shannon's index (H') (Friedman's test, $P < 0.05$) (Table 4). The group with modern Swedish cultivars did however not show the corresponding tendency, which is probably due to the rather broad range of cultivars used in breeding programs.

Table 4. *Number of investigated apple cultivars (within parentheses the original number before duplicates and mislabelled samples had been deleted) used to calculate within-group genetic diversity, measured by Nei's diversity index (H) and Shannon's index (H') (including standard error) in historically and geographically different groups (modified Table 2, paper IV)*

	Group	No. of cultivars	Nei's diversity index, H	Shannon's index, H'
I	Ancient Swedish	16 (17)	0.260 (0.022)	0.539 (0.042)
II	Ancient foreign	19 (21)	0.276 (0.023)	0.568 (0.044)
III	Old Swedish	33 (34)	0.248 (0.023)	0.523 (0.045)
IV	Old foreign	28 (29)	0.281 (0.022)	0.585 (0.042)
V	New Swedish	29 (31)	0.270 (0.021)	0.567 (0.038)
VI	New foreign	18 (19)	0.235 (0.023)	0.493* (0.044)
			x=0.262	x=0.546

SSR-based estimates of expected heterozygosity (H_e) within the set of mandate cultivars (paper V), ($H_e = 0.74$, when all polymorphic loci were included and $H_e = 0.80$, when the less-polymorphic locus in a bilocus system was omitted) were in good agreement with previously published values (Liebhard *et al.*, 2002; Galli *et al.*, 2005; Guarino *et al.*, 2006). In the S-allele analysis, which is a good complement to the other two methods, we identified 14 S-alleles, and allele frequencies ranged from 1 to 18%.

Correlation between markers

SSR markers are generally scored as co-dominant markers and thus analysed as alleles in individual loci. However, some studies, especially those involving polyploids, instead use phenotypic scoring and registering of SSR fragments (McGregor *et al.*, 2000; Ghislain *et al.*, 2006). In paper V, we evaluated the extent of congruence for these two approaches. Correlation (Mantel test) between distance/similarity coefficients for these two types of scoring, was highly significant ($r = -0.863$, $P < 0.001$).

Comparison of inter-cultivar genetic distances obtained with the two different marker systems, SSR (Roger's genetic distance matrix) and RAPD (Jaccard's similarity matrix) yielded a negative correlation as expected ($r = -0.310$, $P < 0.001$). A positive correlation ($r = 0.377$, $P < 0.001$) was obtained when Jaccard similarity matrices for the RAPD data and for the phenotypically scored SSR data were compared.

Genetic diversity within the apple collection appointed for ex situ *conservation*

In paper IV, genetic diversity in the subset with 94 mandate cultivars was only slightly lower than diversity in the entire set with 151 cultivars. Analysis of RAPD band frequencies revealed no general changes: bands that were common in the entire set, remained common also in the subset with mandate cultivars, while rare bands remained rare as well (Fig. 8). Five RAPD bands were not found in the subset of mandate cultivars demonstrating that some gene regions are lacking in this subset. The obtained results indicate that the set with mandate cultivars retains almost the same amount of variation as a randomly sampled collection would have.

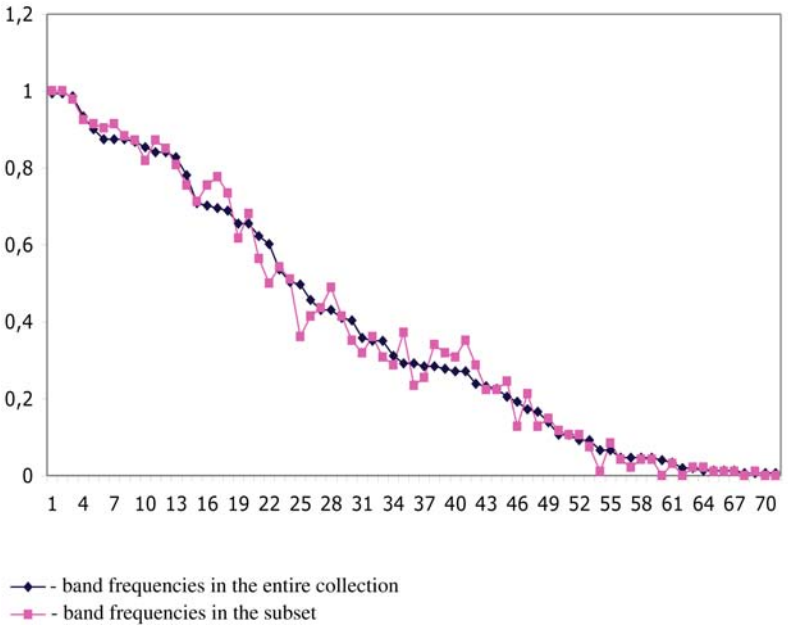


Figure 8. Comparison of RAPD band frequencies in the entire set of apple cultivars (151 cvs) and in the subset of Swedish mandate cultivars (94 cvs) (Figure 3, paper IV).

Conclusions

Studies on genetic diversity estimated by molecular markers presented within this thesis lead to the following conclusions:

- Genetic relationships revealed within genus *Chaenomeles* on cultivated and wild plant material were congruent, and in both cases *C. japonica* was clearly differentiated from other species.
- Populations of *C. japonica* and *C. speciosa* were considerably more diverse than populations of *C. cathayensis* and *C. thibetica*, probably due to different mating systems in these species. Thus, different sampling strategies are needed for *C. japonica* and *C. speciosa* on the one hand, and for *C. cathayensis* and *C. thibetica* on the other.
- The levels and structure of genetic diversity in lingonberry has been found to be in good agreement with the mixed mating system reported for this crop.
- Significant correlation between genetic and geographic distances was observed for the 15 geographically dispersed populations of two subspecies of *Vaccinium vitis-idaea* and for the 13 populations of *V. vitis-idaea* ssp. *vitis-idaea*, derived from locations in Europe and Russia, but the correlation was non-significant when 6 Swedish population were analysed.
- RAPD markers in combination with the Maximum diversity program have proven to be useful in creation of a diverse and representative lingonberry gene bank.
- The apple collection has revealed high overall diversity, which is explained by high heterozygosity due to existing self-incompatibility system, longevity and diverse origins of cultivars.
- No geographic grouping of apple cultivars was observed.
- No major differences in diversity were obtained over time and space, but the group with recent foreign cultivars had somewhat lower diversity.
- The Swedish apple mandate cultivars are apparently quite representative for overall genetic diversity in this crop, but addition of carefully selected germplasm is nevertheless needed for preservation of genes of special interest, like disease resistance etc.

References

Amos, W., Hoffman, J.I., Frodsham, A., Zhang, L., Best, S. & Hill, V.S. 2007. Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* 7, 10–14.

Anonymous, 1989. Medicinal plants in China – a selection of 150 commonly used species species. WHO Regional Publisher, Western Pacific Series No. 2, Manila, 72–73.

Aranzana, M.J., Carbó, J. & Arus, P. 2003. Microsatellite variability in peach [*Prunus persica* (L.) Batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. *Theoretical and Applied Genetics* 106, 1341–1352.

Bartish, I.V., Rumpunen, K. & Nybom, H. 1999. Genetic diversity in Chaenomeles (Rosaceae) revealed by RAPD analysis. *Plant Systematics and Evolution* 214, 131–145.

Beer, S.C., Goffreda, J., Phillips, T.D., Murphy, J.P. & Sorrells, M.E. 1993. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. *Crop Science* 33, 1386–1393.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L. & Trujillo. I. 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theoretical and Applied Genetics* 107, 736–744.

Boscovic, R., Tobutt, K.R., Arus, P. & Messeguer, R. 1994. *Isoenzymes*. In: Methods of molecular marker analysis in *Prunus*. (Ed. R. Messeguer) IRTA. Barcelona, pp. 4–25.

Broothaerts, W. 2003. New findings in apple S-genotype analysis resolve previous confusion and request the re-numbering of some S-alleles. *Theoretical and Applied Genetics* 106, 703–714.

Bruland, O., Almqwist, E.W., Goldberg, Y.P., Boman, H., Hayden, M.R. & Knappskog, P.M. 1999. Accurate determination of the number of CAG repeats in the Huntington disease gene using a sequence-specific internal DNA standard. *Clinical Genetics* 55, 198–202.

Bussel, J.D. 1999. The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobeliaceae). *Molecular Ecology* 8, 775–789.

Cheng, F.S., Weeden, N.F. & Brown, S.K. 1996. Identification of co-dominant RAPD markers tightly linked to fruit skin colour in apple. *Theoretical and Applied Genetics* 93, 222–227.

Cryer, N.C., Fenn, M.G.E., Turnbull, C.J. & Wilkinson, M.J. 2006. Allelic size standards and reference genoypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. *Genetic Resources and Crop Evolution* 53, 1643–1652.

Dahl, C.G. 1929. *Pomologi*. Albert Bonniers förlag, Stockholm. (in Swedish).

Davis, B.J. 1964. Disc electrophoresis. Annals of the New York Academy of Sciences 121, 404–427.

Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

Fady-Welterlen, B. 2005. Is there really more biodiversity in Mediterranean forest ecosystems? *Taxon* 54, 905–910.

Galli, Z., Halász, G., Kiss, E., Heszky, L. & Dobránski, J. 2005. Molecular identification of commercial apple cultivars with microsatellite markers. *HortScience* 40, 1974–1977.

Garkava-Gustavsson, L & Nybom, H. 2003. DNA-analyser avslöjar våra äpplesorter. *Fakta Trädgård-Fritid*, 94. (in Swedish)

Ghislain, M., Andrade, D., Rodriguez, F., Hijmans, R.J. & Spooner, D.M. 2006. Genetic analysis of the cultivated potato *Solanum tuberosum* L. Phureja Group using RAPDs and nuclear SSRs. *Theoretical and Applied Genetics* 113, 1515–1527.

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M. & Gessler, C. 1998. Simple sequence repeats for the genetic analysis of apple. *Theoretical and Applied Genetics* 96, 1069–1076.

Guarino, C., Santoro, S., De Simone, L., Lain, O., Cipriani, G. & Testolin, R. 2006. Genetic diversity in a collection of ancient cultivars of apple (*Malus x domestica* Borkh.) as revealed by SSR-based fingerprinting. *Journal of Horticultural Science and Biotechnology* 81, 39–44.

Gustavsson, B.A. 1999. Plant breeding and domestication of lingonberry (*Vaccinium vitis-idaea* L.). Doctoral thesis, Swedish University of Agricultural Sciences, Alnarp.

Halldén, C., Hansen, M., Nilsson, N.-O., Hjerdin, A. & Säll, T. 1996. Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics* 93, 1185–1192.

Hamrick, J.L. & Godt, M.J.W. 1989. *Allozyme diversity in plant species.* In: Plant population genetics, breeding and genetic resources. (Eds. A.H.D. Brown, M.T. Clegg, A. L. Kahler & B.S. Weir) Sinauer Associates Inc., Sunderland Massachusetts, USA, pp. 43–63.

Harada, T., Matsukawa, K., Sato, T., Ishikawa, R., Niizaki, M., & Saoto, K. 1993. DNA-RAPDs detect genetic variation and paternity in *Malus. Euphytica* 65, 87–91.

Harris, S.A., Robinson, J.P. & Juniper, B.E. 2002. Genetic clues to the origin of the apple. *Trends in Genetics* 18, 426–430.

Hartl, D.L. 1988. *A Primer of Population Genetics*. Sinauer Associates Inc, Sunderland Massachusetts, USA.

Hemmat, M., Brown, S.K. & Weeden, N.F. 2002. Tagging and mapping scab resistance genes from R12740-7A apple. *Journal of American Society for Horticultural Science* 127, 365–370.

Heun, M., Murphy, J.P. & Phillips, T.D. 1994. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theoretical and Applied Genetics* 87, 689–696.

Hjalmarsson, I. & Wallace, B. 2004. Content of the Swedish berry gene bank. *Journal of Fruit and Ornamental Plant Research* 12, 129–138.

Hoffman, J.I. & Amos, W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14, 599–612.

Hokanson, S.C., Szewc-McFadden, A.K., Lamboy, W.F. & McFerson, J.R. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. *Theoretical and Applied Genetics* 97, 671–683.

Hultén, E. 1949. On the races in the Scandinavian flora. *Svensk Botanisk Tidskrift* 43: 383–406.

Isabel, N., Beaulieu, J. & Bousquet, J. 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. *Proceedings of the National Academy of Sciences of the United States of America* 92, 6369–6373.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles* 44, 223–270.

Jacquemart, A.L. & Thompson, J.D. 1996. Floral and pollination biology of three sympatric *Vaccinium* (Ericaceae) species in the Upper Ardennes, Belgium. *Canadian Journal of Botany* 7, 210–221.

Janick, J., Cummins, J.N., Brown, S.K. & Hemmat, M. 1996. *Apples.* In: Fruit Breeding. Vol. 1. Tree and tropical fruits. (Eds. J. Janick & J. Moore). John Willey & Sons Inc., New York, 1–27.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Weil, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert R., Volcaert, G., Rueda, R., Linacero, R., Vazgues, A. & Karp, A. 1997. Reproducubility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* 3, 381–390.

Kaufmane, E. & Rumpunen, K. 2002. Pollination, pollen tube growth and fertilization in *Chaenomeles japonica* (Japanese quince). *Scientia Horticulturae* 94, 257–271.

Kitahara, K & Matsumoto, S. 2002. Cloning of S_{25} cDNA from 'McIntosh' apple and an S_{25}-allele identification method. *Journal of Horticultural Science and Biotechnology* 77, 724–728.

Kitahara, K., Matsumoto, S., Yamamoto, T., Soejima, J., Kimura, T., Komatsu, H. & Abe, K. 2005. Parent identification of eight apple cultivars by S-RNase analysis and simple sequence repeat markers. *HortScience* 40, 314–317.

Kloosterman, A.D., Budowle, B. & Daselaar, P. 1993. PCR-amplification and detection of the human D1S80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. *International Journal of Legal Medicine* 105, 257–264.

Koller, B., Lehmann, A., McDermott, J.M. & Gessler, C. 1993. Identification of apple cultivars using RAPD markers. *Theoretical and Applied Genetics* 85, 901–904.

Konieczny, A. & Ausubel, F.M. 1993. A procedure of mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. *Plant Journal*, 4, 403–410.

Laurens, F., Durel, C.E. & Lascostes, M. 2004. Molecular characterisation of French local apple cultivars using SSR. *Acta Horticulturae* 663, 639–641.

Leclair, B., Fregeau, C.J., Bowen, K.L. & Fourney, R.M. 2004 Precision and accuracy in fluorescent short tandem repeat DNA typing: assessment of benefits impared by the use of allelic ladders with the AmpF/STR®ProfilerPlus™ kit. *Electrophoresis* 25, 790–796.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C.D., Tarchini, R., Van De Weg, E. & Gessler, C. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Molecular Breeding* 10, 214–241.

Litt, M. & Luty, J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44, 397–401.

Lynch, M. & Milligan, B.G. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3, 91–99.

Marita, J.M., Rodriguez, J.M. & Nienhuis, J. 2000. Development of an algorithm identifying maximally diverse core collections. *Genetic Resources Crop Evolution* 47, 515–526.

Markert, C.L. & Moller F. 1959. Multiple forms of enzymes: Tissues, cytogenetic and species specific patterns. *Proceedings of the National Academy of Sciences of the United States of America* 45, 753–763.

Matsumoto, S., Komori, S., Kitahara, K., Imazu, S. & Soejima J. 1999. S-genotypes of 15 apple cultivars and self-compatibility of 'Megumi'. *Journal of the Japanese Society for Horticultural Science* 68: 236–241.

Mattisson, H. & Nybom, H. 2005. Application of DNA markers for detection of scab resistant cultivars ans selections. *International Journal of Horticultural Science* 11, 59–63.

McGregor, C.E., Lambert, C.A., Greyling, M.M., Louw, J.H. & Warnich, L. 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica* 113, 135–144.

Melounova, M., Vejl, P., Sedlák, P., Blazek, J., Zoufala J., Milec, Z. & Blazkova, H. 2005. Alleles controlling apple skin colour and incompatibility in new Czech apple varieties with different degrees of resistance against *Venturia inaequalis* CKE. *Plant Soil and Environment* 51, 65–73.

Mienie, C.M.S., Fourie, H., Smit, M.A., Van Staden, J., Botha, F.C. 2002. Identification of AFLP markers in soybean linked to resistance to *Meloidogyne javanica* and conversion to sequence characterized amplified regions (SCARs). *Plant Growth Regulation* 37, 157–166.

Milbourne, D., Meyer, R., Bradshaw, J.E., Baird, E., Bonar, N., Provan, J., Powell, W. & Waugh, R. 1997. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Molecular Breeding* 3, 127–136.

Monte-Corvo, L., Cabrita, L., Oliveira, C. & Leitao J. 2000. Assessment of genetic relationships among *Pyrus* species and cultivars using AFLP and RAPD markers. *Genetic Resources Crop Evolution* 47, 257–265.

Mulcahy, D.L., Cresti, M., Sansavini, S., Douglas, G.C., Linskans, H.F., Bergamini Mulcahy, G., Vignani, R. & Pancaldi, M. 1993. The use of random amplified polymorphic DNAs to fingerprint apple genotypes. *Scientia Horticulturae* 54, 89–96.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* 106, 283–292.

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

Nilsson, A. 1986. *Våra äpplesorter*. AB Allmänna Förlaget, Stockholm. (in Swedish)

Norusis, M. J. 1990. *SPSS Base System User's Guide*. SPSS Inc., Chicago.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13, 1143–1155.

Nybom, H. & Bartish, I. 2000. Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology Evolution and Systematics* 3, 93–114.

Nybom, H. & Schaal, B.A. 1990. DNA "fingerprints" reveal genotypic distributions in natural populations of blackberries and raspberries (*Rubus*, Rosaceae). *American Journal of Botany* 77, 883–888.

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

Parsons, B.J., Newbury, N.J., Jackson, M.T. & Ford-Lloyd, B.V. 1999. The genetic structure and conservation of aus, aman and boro rices from Bangladesh. *Genetic Resources Crop Evolution* 46, 587–598.

Pereira-Lorenzo, S., Ramos-Cabrer, A. M. & Diáz-Hernández, M. B. 2007. Evaluation of genetic identity and variation of local apple cultivars (*Malus* x *domestica* Borkh.) from Spain using microsatellite markers. *Genetic Resources and Crop Evolution* 54, 405–420.

Persson, H.A. & Gustavsson, B.A. 2001. The extent of clonality and genetic diversity in lingonberry (*Vaccinium vitis-idaea* L.) revealed by RAPDs and leaf-shape analysis. *Molecular Ecology* 7, 1385–1397.

Phipps, J.B., Robertson, K.R., Smith, P.G. & Rohrer, J.R. 1990. A checklist of the subfamily Maloideae (Rosaceae). *Canadian Journal of Botany* 68, 2209–2269.

Ponomarenko, N.S. 1990. Biological characteristic and intraspecific variability of Chaenomeles japonica (Thunb). Lindl ex Spach in Moldova. Synopsis of Thesis. Botanical Garden, Kishinjev. (in Russian)

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

Reif, J.C., Melchinger, A.E. & Frisch, M. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank manamegent. *Crop Science* 45, 1–7.

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

Royo, J.B. & Itoiz, R. 2004. Evaluation of the discriminance capacity of RAPD, isoenzymes and morphologic markers in apple (*Malus x domestica* Borkh.) and the congruence among classifications. *Genetic Resources and Crop Evolution* 51, 153–160.

Ruisa, S. 1996. Studies on Japanese quince (*Chaenomeles japonica*) in Latvia. Report 1992-1994, Balsgård-Department of Horticultural Plant Breeding, Swedish University of Agricultural Sciences, 204–206.

Rumpunen, K. 2001. Diversity in the plant genus Chaenomeles. Doctoral thesis, Swedish University of Agricultural Sciences, Alnarp.

Rumpunen, K., Kviklys, D., Kaufmane, E. & Garkava, L. 1998. Breeding Chaenomeles - a new aromatic fruit crop. *Acta Horticulturae* 484, 211–216.

Sansavini, S., Donati, F., Costa, F. & Tartarini, S. 2004. Advances in apple breding for enhanced fruit quality and resistance to biotic stresses: new varieties for the European market. *Journal of Fruit and Ornamental Plant Research* 12, 13–52.

Schlötterer, C. 2004. The evolution of molecular markers – just a matter of fashion? *Nature reviews/Genetics* 5, 63–69.

Sneath, P.H.& Sokal, R.M. 1973. *Numerical taxonomy. The principles and practice of numerical classification.* W.H. Freeman and Co, San Francisco.
Strefeler, M.S., Darmo, E., Becker, R.L.& Katovich, E.J. 1996a. Isozyme variation in cultivars of purple loosestrife (*Lythrum* sp.). *HortScience* 31, 279–282.
Strefeler, M.S., Darmo, E., Becker, R.L.& Katovich, E.J. 1996b. Isozyme characterization of genetic diversity in Minnesota populations of purple loosestrife, *Lythrum salicaria* (Lythraceae). *American Journal of Botany* 83, 265–273.
Svensson, H & Kastman, K. 2005. *Äpplen i Sverige.* Prisma. (in Swedish)
Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463–6471.
Tonguc, M. & Griffiths, P.D. 2004. Genetic relationships of *Brassica* vegetables determined using database derived simple sequence repeats. *Euphytica* 137, 193–201.
Vanwynsberghe, L., De Witte, K., Coart, E. & Keulemans, J. 2005. Limited application of homozygous genotypes in apple breeding. *Plant Breeding* 124, 399–403.
Vijayan, K., Nair, C. V. & Chatterjee, S. N. 2005. Molecular characterization of mulberry genetic resources indigenous to India. *Genetic Resources Crop Evolution* 52, 77–86.
Wang, J.X, Guang, Z.G. & Teng, Z.Q. 1997. The technique for high production of Yizhou quince. *China-Fruits* 3, 39–40. (in Chinese)
Wang, J.X., Wang, X.L., Guan, Z.G. & Teng, Z.Q. 1998. Preliminary investigation and classification of *Chaenomeles* cultivars. *Journal of the Bejing Forestry University* 20, 123–125. (in Chinese)
Wang, S.Y., Feng, R., Bowman, L., Penhallegon, R., Ding, M. & Lu, Y. 2005. Antioxidant activity in lingonberries (*Vaccinium vitis-idaea* L.) and its inhibitory effect on Activator protein-1, Nuclear factor-*k*B, and Mitogen-activated protein kinases activation. *Journal of Agricultural and Food Chemistry* 53, 3156–3166.
Weber, C. 1964. The genus *Chaenomeles* (Rosaceae). *Journal of the Arnold Arboretum* 45, 161-205, 302-345.
Weising, K., Nybom, H., Wolff, K. & Meyer, W. 1995. *DNA fingerprinting in plants and fungi.* CRC Press. Boca Raton.
Weising, K., Nybom, H., Wolff, K. & Kahl., G. 2005. *DNA fingerprinting in plants. Principles, Methods, and Applications.* 2nd edition. CRC Press, Tylor and Francis group. Boca Raton.
Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Research* 18, 6531–6535.
Yamamoto, T., Kimura, T., Sawamura, Y., Kotobuki, K., Ban, Y., Hayashi, T. & Matsuta, N. 2001. SSR isolated from apple can identify polymorhism and genetic diversity in pear. *Theoretical and Applied Genetics* 102, 865–870.
Yamamoto, T., Kimura, T., Soejima, J., Sanada, T., Ban, Y. & Hayashi, T. 2004. Identification of quince varieties using SSR markers developed from pear and apple. *Breeding Science* 54, 239–244.
Zhou Z.Q. & Li Y.N., 2000. The RAPD evidence for the phylogenetic relationship of the closely related species of cultivated apple. *Genetic Resources and Crop Evolution* 47, 353–357.

Acknowledgements

Now I have finished this long journey. I would like to thank everybody who helped me on my way.

I would like to express my gratitude to my supervisor, Professor Hilde Nybom. Thank you, Hilde, for giving me the possibility to work in your group, for excellent supervision, for encouraging and supporting me, for reading and correcting my manuscripts over and over again, for giving me your time.

My thanks also to my co-supervisor, Philip Forsline, for introducing me into the challenging word of genetic resources, and to my other co-supervisors, Gun Werlemark and Anders Carlsson for fruitful discussions.

My warm thanks go to Kimmo Rumpunen for inviting me to Sweden and introducing me into the world of *Chaenomeles*, and for being a cheerful colleague and a friend. Thank you, Kimmo, also for the *Chaenomeles* pictures. I thank also my co-authors, Igor Bartish, Helena Persson, Agnese Kolodinska-Branteshtam and Jasna Sehic for good cooperation and creating a fine team spririt.

My sincere thanks to Helena Mattisson and Åsa Jönsson, my PhD student mates and friends. Thanks also to Ulrika Carlson-Nilsson for interesting conversations. You made my years at Balsgård pleasant!

I thank everybody at Balsgård for friendship and support!

I thank all my friends from Alnarp, especially Shu-Chin Hysing, Salla Marttila, Rakel Berglund, Sara Bergquist and Kristina Santén. I always felt welcome when I visited Alnarp.

I am also grateful to Eva Johansson for support and valuable advices. Thanks to Jan-Erik Englund for help with statistics.

I thank Professor Arnulf Merker for giving me a possibility to work as a project coordinator, which is completely different from scientific work. It has been - and still is - a very valuable experience. Thanks also to my co-worker in this project, Carina Larsson.

When I was in the middle of the PhD studies, my little daughter was born. That was a very tough time. Therefore I would like to express my deepest gratitude to Lisbeth Olsson, Jan Bergills, Bo Selander, Beata Skanse, and the staff at the Children's hospital in Kristianstad and Lund involved in her recovery. Special thanks to Kerstin Jakobsson for the long and very valuable job she has done.

Finally, I would like to express my warmest thanks to my husband Björn. You have always supported and loved me. You were always sure that I would manage it, even though I sometimes doubted it myself. Thanks! I thank my children, Beata and Oskar. My little girl, you have taught me to never give up. My son, thank you

for your interest in my experiments and for helping me to fill up the boxes with pipet tips. Thanks to all of you for your patience with my work.

Thanks to my mother, Lidija and my mother-in-law Marianne for love, support and for helping with the children. My thoughts go to my brother, father and my father-in-law, who unfortunately are not with us today. You would have been proud of me.

Thanks!

I

ELSEVIER

Scientia Horticulturae 85 (2000) 21–35

SCIENTIA HORTICULTURÆ

www.elsevier.com/locate/scihorti

Genetic relationships in *Chaenomeles* (*Rosaceae*) revealed by isozyme analysis

Larisa P. Garkava[a,b], Kimmo Rumpunen[b,*], Igor V. Bartish[b]

[a]*Institute of Horticulture, Kiev-27, 252027, Ukraine, Russia*
[b]*Balsgård — Department of Horticultural Plant Breeding, Swedish University of Agricultural Sciences, Fjälkestadsvägen 123-1, S-29194 Kristianstad, Sweden*

Accepted 23 August 1999

Abstract

Genetic relationships were studied among four taxa in the genus *Chaenomeles* using isozyme analysis. The band patterns obtained with six polymorphic isozyme systems provided 108 reliable markers, which were scored as unordered multistate traits. A cluster analysis as well as a multidimensional scaling analysis grouped the taxa in agreement with previously published results obtained with RAPD (random amplified polymorphic DNA) analysis; *C. japonica* and *C. cathayensis* were the most distantly related species, whereas *C. speciosa* took an intermediate position together with the hybrid taxon *C.* x *superba*. Similarity matrices obtained with isozymes and RAPDs, respectively, were closely correlated, $r = 0.74$. The previously noted low level of RAPD variability in *C. cathayensis* was indicated also with isozymes. However, the isozyme data were less efficient than the RAPD data for intraspecific grouping of the genotypes according to the origin of the plant material.

Keywords: *Rosaceae*; *Chaenomeles*; Genetic diversity; Isozymes; RAPD; Quince

1. Introduction

Domestication of the genus *Chaenomeles* (Maloideae, *Rosaceae*) as a horticultural crop has recently been initiated in Europe and in Asia. Various species and species hybrids of *Chaenomeles* are now grown for the production of

* Corresponding author. Tel.: +46-44-75533; fax: +46-44-75530.
E-mail address: kimmo.rumpunen@hvf.slu.se (K. Rumpunen)

0304-4238/00/$ – see front matter
PII: S0304-4238(99)00122-3

valuable aroma compounds and an acid juice in Latvia and Lithuania (Rumpunen et al., 1998) and China (Wang et al., 1997, 1998). The large amount of high quality polysaccharides in the fruits makes this crop a promising candidate also for the manufacture of marmalades and pectins provided that well adapted, high yielding varieties can be developed.

All *Chaenomeles* species are diploid, $2n = 34$, and presumed to be outcrossing but the mating system is not yet well studied. Interspecific hybridization produces fertile offspring in all combinations so far tried, and several hybrid taxa have consequently been described (Weber, 1964). A strong self-incompatibility system prevails in *C. japonica* (Rumpunen et al., 1998). There is also some evidence of self-incompatibility in *C. speciosa* and in various hybrids (Weber, 1964).

A large collection of *Chaenomeles* genotypes has been gathered at Balsgård — Department of Horticultural Plant Breeding, Sweden, from orchards and botanical gardens around the world to provide a basis for a joint North-European plant breeding program. Three species, *C. japonica* (Thunb.) Lindl. (Japanese quince), *C. speciosa* (Sweet) Nakai (flowering quince), and *C. cathayensis* (Hemsl.) Schneider (Chinese quince) as well as the hybrid taxon *C.* x *superba* are represented in this collection. The fourth *Chaenomeles* species mentioned in the most recent check list of the subfamily *Maloideae* (Phipps et al., 1990), *C. thibetica* Yü, has unfortunately not yet been incorporated in this plant collection.

The present study investigates the feasibility of using isozyme analysis for characterization of genetic diversity among taxa, accessions, and individual plants of *Chaenomeles*. Only one minor isozyme study has previously been reported in *Chaenomeles* (Ponomarenko, 1990). We also aimed to compare isozyme data with previously published RAPD data on the same plant material (Bartish et al., 1999), and to estimate the efficiency of these two marker systems.

2. Materials and methods

2.1. Plant material

A total of 42 genotypes were investigated representing four taxa in *Chaenomeles*: *C. cathayensis, C. japonica*, *C. speciosa* and *C.* x *superba*. The plant material used (Table 1) was the same as described in Bartish et al. (1999) except for genotypes RG 4-37 and RG 4-48 (replaced by RG 4-43 and RG 4-54), RG 4-29 and RG 4-30 (replaced by RG 4-126 and RG 4-125). The new genotypes were from the same accessions as the previously analysed plants which had unfortunately been killed by winter frost. The plant material was obtained from the *Chaenomeles* core collection maintained at Balsgård — Department of Horticultural Plant Breeding. One-year-old shoots were taken in the field in January when dormancy had been terminated. Phloem tissue was sampled from

Table 1
Plant accessions of *Chaenomeles*

Taxon	Accession	Symbol	Genotype, origin
C. cathayensis	8	●	RG 4-37/125, RG 4-48/126 Botanisk Have, Denmark
C. cathayensis	9	○	RG 4-58, RG 4-66 Botanischer Garten, Essen, Germany
C. cathayensis	7	◒	RG 4-29/43, RG 4-30/54 Botanischer Garten, Stuttgart, Germany
C. speciosa	10	▼	RG 4-79, RG 4-80, RG 4-83, RG 4-89 Hortus Zoologius Pragensis, The Czech Republic
C. speciosa	11	▽	RG 1-8, RG 1-10 Kyoto Takeda Herbal Garden, Japan
C. x *superba*	12	△	RG 2-28, RG 2-30, RG 2-32, RG 2-42, RG 2-43 Botanischer Garten, Stuttgart, Germany
C. x *superba*	13	◭	RG 4-119 Botanischer Garten, Stuttgart, Germany
C. x *superba* 'Pink Lady'	14	▲	BG 1-1 The Elite Plant Station, Sweden
C. japonica	1	⊠	RG 3-57, RG 3-58, RG 3-59, RG 3-62, RG 3-63, RG 3-64 Botanical Garden, Uppsala, Sweden
C. japonica	4	◪	BE 4-22, BE 7-33 Domesticated population, Babtai, Lithuania
C. japonica	5	□	RG 5-31, RG 5-34, RG 5-36, RG 5-37, RG 5-39, RG 5-40 Domesticated population, Smiltene, Latvia
C. japonica	6	◹	RG 7-21 Domesticated population, Smiltene, Latvia
C. japonica	2	■	RG 1-134, RG 1-142, RG 1-145 Sendai Botanical Garden, Japan
C. japonica	3	⊞	RG 4-21, RG 4-35, RG 4-41, RG 4-50, RG 4-60 Salaspils Botanical Garden, Latvia

debarked shoots whereas young leaves were obtained from shoots that had been forced in a greenhouse at 15–20°C for 3–4 weeks.

2.2. Enzyme extraction

Fresh phloem tissue (1 g) or leaves (1 g) were ground with liquid nitrogen in pre-cooled mortars and 3 ml of extraction buffer was added. Extraction from phloem tissue was carried out according to Boscovic et al. (1994). Extraction from leaves was carried out with an 0.05 M sodium phosphate buffer (adjusted to pH 7.2 with HCl), containing 0.03% DL-dithiothreitol (DTT), 0.075% β-

mercaptoethanol, 0.1% ascorbic acid, 4.5% polyvinylpyrrolidone and 8.0% sucrose. For both kinds of tissue, samples were centrifuged (7000 rpm) for 20 min at 2°C and the supernatant was collected and kept at −20°C until used.

2.3. Electrophoresis

Isozyme analysis was carried out by electrophoresis in polyacrylamide gels (16 cm × 18 cm, 15 wells, 0.75 mm thick). Density gradient gels, 5.8–11.8%, were prepared according to Boscovic et al. (1994) and 7.5% gels were prepared according to Davis (1964). Electrophoretic separation was carried out at 4°C with a running time of 3.5–4 h at 30 mA on LKB 2001 vertical electrophoresis unit with LKB 2003 power supply. All samples were electrophoresed twice.

2.4. Gel staining

The enzyme assay procedures for acid phosphatase (ACP, EC 3.1.3.2), superoxide dismutase (SOD, EC 1.15.1.1), malic enzyme (ME, EC 1.1.1.40), and esterases (EST, EC 3.1.1.-) were as described by Wendel and Weeden (1989). Alcohol dehydrogenase (ADH, EC 1.1.1.1), alkaline phosphatase (AKP, EC 3.1.3.1), glucose-6-phosphate dehydrogenase (G-6-PDH, EC 1.1.1.49), glucose-6-phosphate isomerase (GPI, EC 5.3.1.9), glutamate dehydrogenase (GDH, EC 1.4.1.2), glutamate oxaloacetate transaminase (GOT, EC 2.6.1.1), isocitrate dehydrogenase (IDH, EC 1.1.1.42), malate dehydrogenase (MDH, EC 1.1.1.37), shikimate dehydrogenase (SKDH, EC 1.1.1.25), phosphogluconate dehydrogenase (PGD, EC 1.1.1.44) and phosphoglucomutase (PGM, EC 5.4.2.2) were stained according to Vallejos (1983). Peroxidase (PRX, EC 1.11.1.7) was stained with 0.091% benzidine in 0.32 M sodium acetate buffer adjusted to pH 5.4, and bands were revealed with 0.01% hydrogen peroxide. All isozyme systems except SOD were stained in the dark at 37–38°C. SOD were stained under light after dark treatment at the same temperature as the other systems. Gels stained for PRX, ACP, EST and GOT were fixed in 95% ethanol and 30% glycerol, whereas gels stained for the other enzyme systems were fixed in 40% glycerol when a suitable band intensity had been attained.

2.5. Data analysis

Data were collected immediately when gels had been fixed. Only clear and consistent isozyme bands were scored. Band patterns from each zone of activity were numbered in ascending order starting with the fastest migrating band for each enzyme system and scored as unordered multistate traits (Beer et al., 1993; Strefeler et al., 1996a,b). Thus a matrix for genotypes and isozyme systems was constructed with ‘‘1’’ (presence) and ‘‘0’’ (absence) for each isozyme band.

SPSS data analysis package (Norusis, 1990) and NTSYS-pc (Rohlf, 1997) were used for the statistical analyses. Cronbach's α (procedure reliability, SPSS) was used to investigate the raw data and remove unreliable bands before further statistical treatment.

Jaccard's coefficient of similarity was calculated for all pairwise comparisons between individual samples to provide a distance matrix. A cluster analysis based on average linkage between groups (unweighted pair group method algorithm, UPGMA) was constructed based on this matrix, and a dendrogram was produced.

A multidimensional scaling analysis (MDS) was conducted (procedure alscal, SPSS) using a genetic distance matrix obtained from the binary data set with negated and rescaled (0–1) values for Jaccard's coefficient of similarity. A scaling solution with three dimensions was used in our study.

Correspondence between the two distance matrices obtained here with isozymes and previously with RAPDs (Bartish et al., 1999), respectively, was analysed with the MXCOMP procedure (NTSYS) which produces the Pearson product-moment correlation, *r*, and the Mantel test statistic, *Z*. If the two matrices show similar relationships, *Z* is large in comparison to chance expectation. The estimated *Z* was compared with its permutational distribution obtained from 9999 random samples of all possible permutations of the matrices.

A combined matrix was also produced, including the present isozyme data as well as the previously obtained RAPD data. The MXCOMP procedure was used on each data set separately as well as on the combined data set to obtain the cophenetic correlation between the UPGMA matrix and the distance matrix as a measure of goodness of fit for the cluster analysis.

2.6. Experimental error

The fact that four *C. cathayensis* genotypes had to be replaced in the present isozyme study due to death of plants used in the RAPD study, may have biased some of the results. However, this bias is probably insignificant since *C. cathayensis* appears to be the least variable species, and since the new genotypes were selected from the same accessions as the original plants.

3. Results

Initially, starch gel electrophoresis was used for some systems but sufficient resolution was only obtained for MDH and ACP. For other systems the resolution was poor or there was no activity. Moreover, the isozyme patterns obtained for MDH and ACP with starch gel electrophoresis were less polymorphic than the patterns obtained with polyacrylamide gels. Therefore, only data derived from polyacrylamide gels were used in this study.

3.1. Polymorphism of isozyme systems

Thirteen (ACP, ADH, AKP, EST, GDH, GOT, G-6PDH, IDH, MDH, PGM, PRX, SKDH and SOD) of the 16 isozyme systems initially tested showed enzyme activity. Based on these preliminary investigations, four systems (ACP, EST, MDH and PRX) were selected which consistently yielded good resolution, clear bands and reproducible profiles both for phloem and leaf tissue. In addition, PGM was selected for analysis of leaf extracts and GOT was selected for analysis of phloem extracts (Table 2). Other isozyme systems did not result in sufficient resolution or produced inconsistent band patterns. For PRX, a discontinuous electrophoretic system was used while for ACP, EST, GOT, MDH and PGM a gradient system resulted in the best resolution.

More than one zone of activity was observed for all the selected isozyme systems. These zones of activity were probably the result of different isozyme loci but this was not analysed further.

Acid phosphatase. Four monomeric zones and one dimeric zone of activity were observed for ACP in the phloem extracts. Twelve bands were scored in the monomeric zones but no sufficiently clear bands could be scored in the dimeric zone. All *C. cathayensis* genotypes had the same ACP profile, but this system was polymorphic for the other taxa.

Leaf extracts yielded four monomeric zones that seemed homologous to the zones produced by phloem extracts. However, with leaf extracts one additional fast migrating zone was obtained. Leaf extract-derived bands in the dimeric zone stained faintly. Altogether 14 polymorphic bands were selected for scoring from leaf extracts.

Esterases. α-esterase activity was detected in both leaf and phloem extracts whereas β-esterase activity was detected only in the phloem extracts.

Table 2
Number of total bands and number of polymorphic bands for each of the six isozyme systems studied

Isozyme system	Total bands		Polymorphic bands	
	Leaves	Phloem	Leaves	Phloem
ACP (EC 3.1.3.2)	15	12	14	12
EST (EC 3.1.1.-)	17	12	14	12
GOT (EC 2.6.1.1)	–	10	–	10
MDH (EC 1.1.1.37)	8	9	7	9
PER (EC 1.11.1.7)	17	19	15	16
PGM (EC 5.4.2.2)	6	–	6	–
Total bands	63	62	56	59

Two cathodal zones and one anodal zone of α-esterase activity were observed in the phloem extracts. In total, seven clear and consistent bands were scored, of these six were in the cathodal zone. All *C. cathayensis* genotypes had the same cathodal α-esterase profile, which differed from profiles of other species. All *C. speciosa* and *C.* x *superba* genotypes had one intensely stained α-esterase band in the anodal zone. The *C. cathayensis* genotypes also had this band, but stained less intensely. Only three out of 23 genotypes of *C. japonica* had this band.

β-esterase showed two zones of activity in the phloem extracts, one fast migrating and one slowly migrating. The slowly migrating zone had poor resolution and was not taken into account for this study. In total five polymorphic bands were scored. *C. cathayensis* had one monomorphic band in the fast migrating zone, whereas *C.* x *superba* and *C. japonica* genotypes had another band. *C. speciosa* genotypes each had two or three bands in this zone.

At least seven zones of α-esterase activity were observed in the leaf extracts. Six of these zones together yielded 17 well separated bands, but the three fastest migrating bands were monomorphic. Thus, fourteen polymorphic bands were scored.

Glutamate oxaloacetate transaminase. Ten bands representing four zones were found for GOT in the phloem extracts. Leaf extracts yielded only one active zone.

Malate dehydrogenase. Up to four zones of MDH activity were observed in the phloem extracts and nine bands were scored, all of them polymorphic. In the leaf extracts, eight bands were well separated and seven of these were polymorphic.

Peroxidase. Seven zones (PRX-1–PRX-7) and 19 bands of peroxidase activity were observed in the phloem extracts. The same seven zones were found also in the leaf extracts but here the activity in PRX-1 and PRX-2 zones was very low. PRX-5 and PRX-6 zones were monomorphic. *C. cathayensis* had one major band with very high staining intensity in the PRX-4 zone. This band, however, less intensely stained, was also found in the *C. speciosa* and *C.* x *superba* genotypes but not in any of the *C. japonica* genotypes.

Phosphoglucomutase. Two zones of activity were observed for PGM in the leaf extracts, but only one of them was well resolved. Six bands were consistent and, therefore, selected for scoring. PGM activity was found also in the phloem extracts, but resolution was not sufficient to allow reliable scoring.

Based on the calculations of Cronbach's α, seven out of 115 initially scored bands (6%) were removed from our data set. The remaining bands resulted in a matrix of 42 genotypes and 108 bands, representing six isozyme systems, which were then used for the further analyses.

3.2. Multivariate analyses

The cophenetic correlation for the cluster analysis was high, $r = 0.90$, indicating a very good fit of the data. The 42 genotypes were divided into three

major clusters (Fig. 1). *C. cathayensis* isozyme phenotypes appear to be rather similar to each other and formed the most distant cluster. Another major cluster was formed by *C. japonica*. Genotypes of *C. speciosa* and *C.* x *superba* clustered together. Inside the major clusters, genotypes generally grouped according to the origin of the plant material.

The three-dimensional MDS yielded a plot of dimension 1 vs 2 (Fig. 2) with three main groups, which were in good agreement with the cluster analysis. *C.*

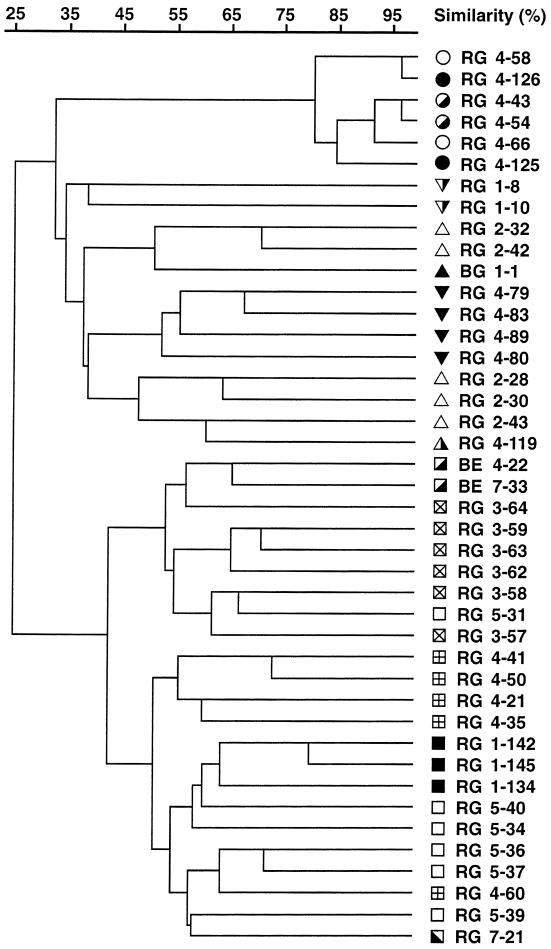


Fig. 1. Dendrogram based on UPGMA analysis of genetic similarity estimates (Jaccard's coefficient) for isozyme data, showing relationships among individual plants. The 42 genotypes are classified into three major clusters. *C. cathayensis* (○) form the most distant cluster. Another major cluster is formed by *C. japonica* (□). Genotypes of *C. speciosa* (▽) and *C.* x *superba* (△) cluster together.

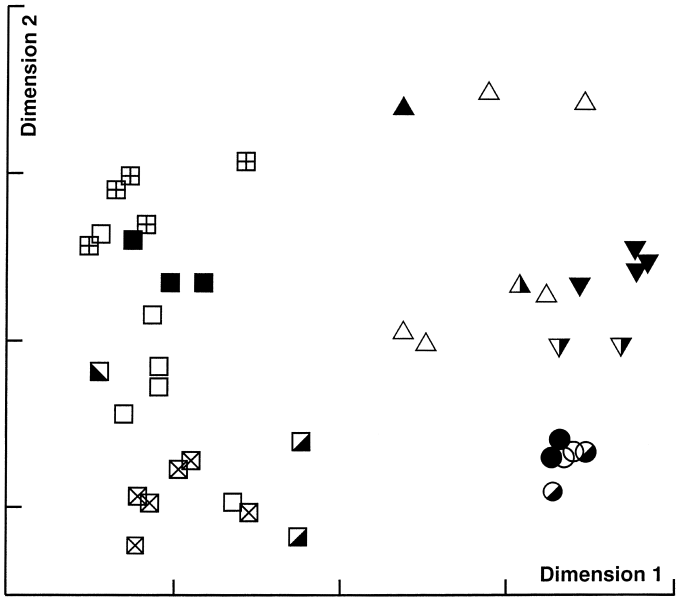


Fig. 2. Plot of MDS analysis of genetic similarity estimates (Jaccard's coefficient) for isozyme data, showing groups among individual plants. *C. cathayensis* (○) and *C. japonica* (□) form two well-separated groups, whereas *C. speciosa* (▽) and *C.* x *superba* (△) together form a third intermediate group.

cathayensis and *C. japonica* formed two well-separated groups, whereas *C. speciosa* and *C.* x *superba* together formed a third group.

3.3. Comparing isozyme and RAPD data

The Pearson product-moment correlation coefficient calculated between the elements of the distance matrices for the isozyme data here presented and the RAPD data previously reported (Bartish et al., 1999) was $r = 0.74$. In addition, no estimate of Z from 9999 random permutations of the matrices was equal to or larger than the observed Z (16.9). In conclusion, the two data sets thus provided highly corresponding estimates of genetic relationships among the 42 genotypes.

3.4. Combining isozyme and RAPD data

When the isozyme and RAPD data sets were combined in one matrix, the cophenetic correlation coefficient reached $r = 0.91$ indicating a very good fit of data for the cluster analysis. The resulting dendrogram (Fig. 3) was in very good

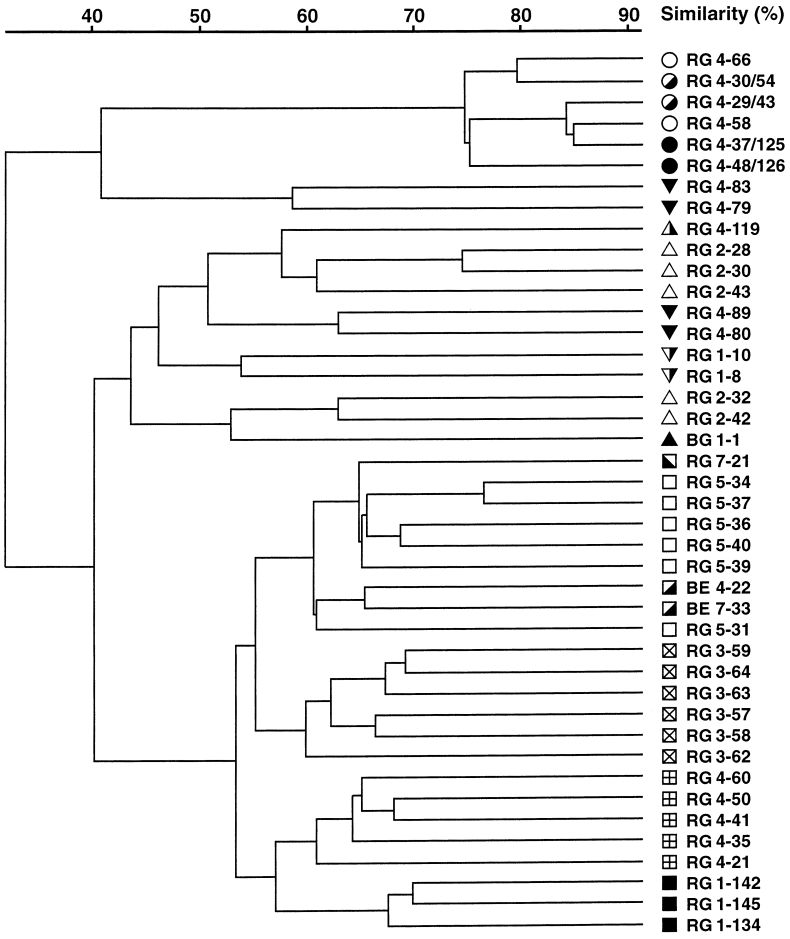


Fig. 3. Dendrogram based on UPGMA analysis of genetic similarity estimates (Jaccard's coefficient) for the combined data set (isozymes and RAPDs), showing relationships among individual plants. The dendrogram is in good agreement with the taxonomy, and the genotypes in most cases cluster according to accession or postulated relatedness. *C. cathayensis* (○) form the most distant cluster. Another major cluster is formed by *C. japonica* (□). Genotypes of *C. speciosa* (▽) and *C.* x *superba* (△) cluster together except for one accession of *C. speciosa* (▼) which is split in two groups of which genotype RG 4-83 and RG 4-79 form an intermediate group to *C. cathayensis*.

agreement with the taxonomy, and the genotypes in most cases clustered according to accession or postulated relatedness. However, one accession (number 10) of *C. speciosa* was split in two different groups in the combined data set (Fig. 3) whereas in the isozyme data set (Fig. 1) the four genotypes (RG 4-79, RG 4-80, RG 4-83 and RG 4-89) clustered together.

4. Discussion

Six isozyme systems were found that are useful for characterization of genetic resources and to investigate genetic relationships among and within groups of closely related *Chaenomeles* taxa. In addition, some of the initially tested isozyme systems could probably also yield useful data with further improvement of the methodology. Some other isozyme systems, that have been used successfully in apples (Chevreau et al., 1999), pears (Chevreau et al., 1997) and different *Prunus* species (Mowrey et al., 1990, Mowrey and Werner, 1990; Granger, 1996), could also be tried on *Chaenomeles*.

When integrating isozyme band patterns for the same isozyme system for different tissues (such as leaves and phloem) in the same data set there is a possibility that results may be biased. Some loci may be expressed in both tissues and thus could be scored twice. However, the profiles obtained for the different tissues were clearly different within the isozyme systems studied except for peroxidase. To completely avoid such bias, only the tissue yielding the most polymorphic isozyme pattern should be selected but at the same time valuable information from additional tissue specific loci may be lost. A better strategy would thus be to score the tissue specific loci from each tissue together with the common tissue unspecific loci. To perform such an investigation unambiguously it is, however, necessary to have complete knowledge on the inheritance of the loci. Such information can only be obtained from an analysis of plant material derived through controlled crosses (Manganaris and Alston, 1992c).

The bands were not scored as loci and alleles since the genetic background of the different isozyme systems employed is not yet known in *Chaenomeles*. However, scoring of band patterns as unordered multistate traits has previously proved to be useful for studies where loci cannot be unambiguously identified (Beer et al., 1993; Strefeler et al., 1996a,b). Consequently, this approach was considered appropriate also for the present investigation. In addition, highly variable multilocus systems such as esterases and acid phosphatase are difficult to score in any other way since homologous loci often cannot be identified between taxa (Weeden and Lamb, 1987). Nevertheless, these systems are quite useful for providing numerous, highly polymorphic bands.

4.1. Multivariate analysis

Cluster analysis and MDS of the isozyme data yielded rather similar results. *C. cathayensis* and *C. japonica* appear to be the most distantly related species, whereas *C. speciosa* clustered together with *C.* x *superba*. However, despite the large number of isozyme markers employed in this study, only some association was found between origin of the plant material and the intraspecific variability according to the cluster analysis and MDS, respectively.

4.2. Comparing isozyme and RAPD data

Our isozyme data partitioned genetic diversity in *Chaenomeles* in almost the same way as previously reported with RAPD markers (Bartish et al., 1999). However, some differences were observed. Isozyme data yielded a more compact clustering of *C. cathayensis* genotypes, thus suggesting even less genetic variability than indicated with RAPD data. By contrast, *C. japonica* genotypes clustered less densely when isozyme data was used compared to RAPD data. Both data sets placed *C. speciosa* and *C.* x *superba* close together, and in an intermediate position between *C. cathayensis* and *C. japonica*. However, a more compact cluster was obtained for *C. speciosa* and *C.* x *superba* with isozyme data than with RAPD data. Intraspecific variation was more closely associated with origin of the analysed plant material when based on RAPD data compared to when based on isozyme data. These findings are in agreement with studies on several other species (Heun et al., 1994; Maass and Klaas, 1995; Staub et al., 1997).

Rather few investigations have been published on the correlation between gene diversity estimates obtained with isozymes and RAPDs on woody plants. RAPDs are considered to be selectively neutral and to sample the total genome randomly (Dawson et al., 1995) but there are drawbacks due, for example, to their dominant nature. The distance matrices based on isozyme and RAPD data, respectively, were well correlated ($r = 0.74$) in our investigation. Complete congruence in diversity estimates was found in black spruce (*Picea mariana*) when enzyme loci and RAPD loci were studied in haploid plant tissue, whereas biased estimates were obtained with dominant RAPD phenotypes (Isabel et al., 1995). Estimates on intraspecific variability should, therefore, be considered as preliminary when based on unordered isozyme data or on RAPD phenotypes. Moreover, the sampling error may be rather large because of unknown genetic relationships among our genotypes and because of unknown linkage between isozyme loci.

4.3. Combining isozyme and RAPD data

Combining isozyme data and RAPD data may not always be appropriate for analysis at the interspecific level since co-migrating RAPD bands from different species could be non-homologous (Rieseberg, 1996). In our investigation, the same argument can be applied to the isozymes since we scored these as unordered multistate traits. However, goodness of fit for the cluster analyses, measured as the cophenetic correlation coefficient, was even slightly higher when based on combined data ($r = 0.91$) as compared to only isozyme data ($r = 0.90$) or only RAPD data ($r = 0.85$). Moreover, the genotypes of the most studied taxon, *C. japonica*, clustered more closely according to accession when the combined data set was used. Thus genotypes from Sendai, Japan (RG 1-142, RG1-145 and RG 1-

134) formed one cluster, genotypes from Uppsala Botanical Garden, Sweden (RG4-21, RG 4-35, RG 4-41, RG 4-50 and RG 4-60) formed a second cluster, genotypes from Salaspils Botanical Garden, Latvia (RG 3-57, RG 3-58, RG 3-59, RG 3-62, RG 3-63 and RG 3-64) formed a third cluster, and partly domesticated plants originating from three fields in Latvia and Lithuania formed a fourth cluster. Obviously the mass selection applied in open pollinated populations in farmers' fields has had an impact on the plant material which, it was claimed, had originated from European botanical gardens about three to four generations ago. Morphologically the domesticated plant material differed from wild plants by having almost thornless twigs.

Based on the combined data set two genotypes of *C. speciosa* (accession number 10) were clustered in an intermediate position between the *C. cathayensis* cluster and the *C. speciosa* / *C.* x *superba* main cluster whereas based on isozyme data all four genotypes in the accession were clustered into one group within the *C. speciosa* / *C.* x *superba* main cluster. This result demonstrates the higher discriminating power of RAPDs compared to isozymes since the same separation was obtained also when only RAPD data were used (Bartish et al., 1999). The result indicates a hybrid origin of the accession which is a possible consequence of spontaneous cross pollination taking place in the botanical garden from which the seeds were obtained.

5. Conclusion

The isozyme analysis resulted in almost the same estimates of genetic diversity as previously reported for RAPD markers. Taking into account the comparatively low cost of chemicals and the high efficiency and reliability, isozyme analysis appears to be useful for genetic and taxonomic investigations of the genus *Chaenomeles*. A drawback is the relatively low number of useful systems, which limit the number of polymorphic loci to be studied.

Further investigations are needed to clarify the genetic control of the isozyme systems in *Chaenomeles*, as previously conducted in several closely related genera like *Pyrus* (Chevreau et al., 1997,1999) and *Malus* (Weeden and Lamb, 1987; Manganaris, 1989; Manganaris and Alston, 1992a,b,c; Manganaris and Alston, 1997). This would allow the interpretation of zymograms as alleles and loci, thus raising the information content considerably.

Acknowledgements

This study was the carried out with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD

programme, CT 97-3894, Japanese quince (*C. japonica*) — a new European fruit crop for production of juice, flavour and fibre. It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. Financial support was also received from The Swedish Institute. Valuable comments on the manuscript were received from H. Nybom.

References

Bartish, I.V., Rumpunen, K., Nybom, H., 1999. Genetic diversity in *Chaenomeles* (*Rosaceae*) revealed by RAPD analysis. Pl. Syst. Evol. 214, 131–145.

Beer, S.C., Goffreda, J., Phillips, T.D., Murphy, J.P., Sorrells, M.E., 1993. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. Crop. Sci. 33, 1386–1393.

Boscovic, R., Tobutt, K.R., Arus, P., Messeguer, R., 1994. Isoenzymes. In: Messeguer, R. (Ed.), Methods of Molecular Marker Analysis in *Prunus*. IRTA, Barcelona, pp. 4–25.

Chevreau, E., Leuliette, S., Gallet, M., 1997. Inheritance and linkage of isozyme loci in pear (*Pyrus communis* L.). Theor. Appl. Genet. 94, 498–506.

Chevreau, E., Manganaris, A.G., Gallet, M., 1999. Isozyme segregation in five apple progenies and potential use for map construction. Theor. Appl. Genet. 98, 329–336.

Davis, B.J., 1964. Disc electrophoresis. Ann. NY Acad. Sci. 121, 404–427.

Dawson, I.K., Simons, A.J., Waugh, R., Powell, W., 1995. Diversity and genetic differentiation among subpopulations of *Gliricidia sepium* revealed by PCR-based assays. Heredity 74, 10–18.

Granger, A.R., 1996. Inheritance and linkage of isozymes in sweet cherry (*Prunus avium* L.). Theor. Appl. Genet. 93, 426–430.

Heun, M., Murphy, J.P., Phillips, T.D., 1994. A comparison of RAPD and isoenzyme analysis for determining the genetic relationships among *Avena sterilis* L. accessions. Theor. Appl. Genet. 87, 689–696.

Isabel, N., Beaulieu, J., Bousquet, J., 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc. Natl. Acad. Sci. 92, 6369–6373.

Maass, H.I., Klaas, M., 1995. Infraspecific differentiation of garlic (*Allium sativum* L.) by isozyme and RAPD markers. Theor. Appl. Genet. 91, 89–97.

Manganaris, A.G., 1989. Isoenzymes as genetic markers in apple breeding. Ph.D. Thesis. London University.

Manganaris, A.G., Alston, F.H., 1992a. Genetics of leucine aminopeptidase in apple. Theor. Appl. Genet. 83, 345–352.

Manganaris, A.G., Alston, F.H., 1992b. Genetics of esterase isoenzymes in *Malus*. Theor. Appl. Genet. 83, 467–475.

Manganaris, A.G., Alston, F.H., 1992c. Inheritance and linkage relationships of peroxidase isoenzymes in apple. Theor. Appl. Genet. 83, 392–399.

Manganaris, A.G., Alston, F.H., 1997. Genetics of superoxide dismutase in apple. Theor. Appl. Genet. 95, 484–489.

Mowrey, B.D., Werner, D.J., 1990. Phylogenetic relationships among species of *Prunus* as inferred by isozyme markers. Theor. Appl. Genet. 80, 129–133.

Mowrey, B.D., Werner, D.J., Byrne, D.H., 1990. Isozyme survey of various species of *Prunus* in the subgenus *Amygdalus*. Sci. Hort. 44, 251–260.

Norusis, M.J., 1990. SPSS Base System User's Guide. SPSS Inc., Chicago.

Phipps, J.B., Robertson, K.R., Rohrer, J.R., 1990. A check list of the subfamily *Maloidea* (*Rosaceae*). Can. J. Bot. 68, 2209–2269.

Ponomarenko, N.S., 1990. Biological characteristic and intraspecific variability of *Chaenomeles japonica* (Thunb). Lindl ex Spach in Moldova. Synopsis of Thesis. Botanical Garden, Kishinjev, 22 pp. (in Russian).

Rieseberg, L.H., 1996. The homology among RAPD fragments in interspecific comparisons. Mol. Ecol. 5, 99–105.

Rohlf, F., 1997. NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System. Ver. 1.80. Exeter Software.

Rumpunen, K., Kviklys, D., Kaufmane, E., Garkava, L., Breeding *Chaenomeles* — a new aromatic fruit crop. In: Tobutt, K.R., Alston, F.H. (Eds.), Proceedings of the Eucarpia Symposium on Fruit Breeding and Genetics, vol. 484. Acta Hort., 1998, pp. 211–216.

Staub, J.E., Box, J., Meglic, V., Horejsi, T.F., McGreight, J.D., 1997. Comparison of isozyme and random amplified polymorphic DNA for determining intraspecific variation in *Cucumis*. Genet. Res. Crop Evol. 44, 257–269.

Strefeler, M.S., Darmo, E., Becker, R.L., Katovich, E.J., 1996a. Isozyme variation in cultivars of purple loosestrife (*Lythrum* sp.). Hort. Sci. 31, 279–282.

Strefeler, M.S., Darmo, E., Becker, R.L., Katovich, E.J., 1996b. Isozyme characterization of genetic diversity in Minnesota populations of purple loosestrife, *Lythrum salicaria* (*Lythraceae*). Am. J. Bot. 83, 265–273.

Vallejos, C.E., 1983. Enzyme activity staining. In: Tanksley, S.D., Orton, T.J. (Eds.), Isoenzymes in Plant Genetics and Breeding, Vol. A. Elsevier, Amsterdam, pp. 469–516.

Wang, J.X., Guan, Z.G., Teng, Z.Q., 1997. The techniques for high production of Yizhou quince. China-Fruits 3, 39–40 (in Chinese).

Wang, J.X., Wang, X.L., Guan, Z.G., Teng, Z.Q., 1998. Preliminary investigation and classification of *Chaenomeles* cultivars. J. Beijing Forest. Univ. 20, 123–125 (in Chinese).

Weber, C., 1964. The genus *Chaenomeles* (*Rosaceae*). J. Arnold Arbor. 45 161–205, 302–339.

Weeden, N.F., Lamb, R.C., 1987. Genetics and linkage analysis of 19 isozyme loci in apple. J. Am. Soc. Hort. Sci. 112, 865–872.

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

II

Theor Appl Genet (2000) 101:554–563

I.V. Bartish · L.P. Garkava · K. Rumpunen
H. Nybom

Phylogenetic relationships and differentiation among and within populations of *Chaenomeles* Lindl. (Rosaceae) estimated with RAPDs and isozymes

Received: 23 December 1999 / Accepted: 17 January 2000

Abstract RAPD and isozyme analyses based on numerous markers have been used for the first time to investigate patterns of phenetic and genetic differentiation among and within nine wild populations of the genus *Chaenomeles* represented by the species *C. japonica*, *C. speciosa*, *C. cathayensis* and *C. thibetica*. Highly significant correlations were found between the two different marker systems for both phenetic distances and gene diversity estimates. In agreement with previous studies on cultivated *Chaenomeles* material, *C. japonica* was clearly differentiated from *C. speciosa* and *C. cathayensis*. The recently recognised species *C. thibetica* appeared to be rather closely related to *C. cathayensis*. Populations of *C. japonica* and *C. speciosa* were considerably more diverse than populations of *C. cathayensis* and *C. thibetica*. Correspondingly, most of the total variability could be attributed to the within-population differentiation in the case of *C. japonica* and *C. speciosa*, and to the between-population differentiation in the case of *C. cathayensis*. Differences in mating systems among the species can be suggested as a possible explanation of the results. A discordant pattern was found between RAPDs and isozymes in the analyses of population structure within *C. japonica*. This may be explained by a higher proportion of non-neutral markers for isozymes than for RAPDs. This finding also shows the importance of using multiple molecular marker systems in studies of population structure.

Key words Gene diversity · Isozyme · Non-neutrality · RAPD · Rosaceae

Communicated by P.M.A. Tigerstedt

I.V. Bartish · L.P. Garkava · K. Rumpunen (✉) · H. Nybom
Balsgård–Department of Horticultural Plant Breeding,
Swedish University of Agricultural Sciences,
Fjälkestadsvägen 123-1, S-291 94 Kristianstad, Sweden
e-mail: kimmo.rumpunen@hvf.slu.se
Fax: +46 44 755 30

Introduction

Chaenomeles is a genus within the subfamily Maloideae comprising four diploid (2n=34) species (Phipps et al. 1990). One species is endemic to Japan, and three originate in central Asia. Presently they are being investigated for possible domestication and improvement through breeding in northern Europe. The plants within the genus *Chaenomeles* are all moderate-sized shrubs adaptable to cultivation in temperate areas. *C. japonica* (Thunb.) Lindl. ex Spach (Japanese quince) is outcrossing and strongly self-incompatible (Rumpunen et al. 1999). The mating system has not yet been properly investigated for the other species in the genus. *C. speciosa* (Sweet) Nakai (flowering quince) is a traditional medicinal plant (Anonymous 1989) and grows in Central and southern China, Tibet and Burma. Recently it has also become cultivated in eastern China (Wang et al. 1998). In the province of Yunnan, *C. speciosa* grows sympatrically with *C. cathayensis* (Hemsl.) Schneider (Chinese quince), the largest shrub in the genus, which can also be found in Bhutan and Burma. The fourth species, *C. thibetica* Yü (Tibetan quince), first described in 1963 (Yü and Kuan) and included in the most recent check list of the subfamily Maloideae (Phipps et al. 1990), grows in Tibet and western Sichuan. *C. japonica*, *C. cathayensis* and *C. speciosa* have been used to create several interspecific hybrids for 400 years, resulting in more than 500 cultivars (Weber 1963), which today are widely appreciated as ornamentals.

Efficient methods to clarify the taxonomic status of both the wild and the cultivated material are much needed. Insights into the relative gene diversity among and within wild populations of *Chaenomeles* would be useful in plant breeding and also for the development of strategies for *ex situ* conservation of plant genetic resources. Molecular markers [random amplified polymorphic DNAs (RAPDs) and isozymes] have recently been used to evaluate gene diversity in plant samples of three species of *Chaenomeles*: *C. cathayensis*, *C. japonica* and *C. speciosa*. The accessions were obtained from several botanical gardens, and included samples of domesticated

Table 1 Accession identity (ID), number of individuals and origin of the nine *Chaenomeles* populations investigated

ID	Individuals	Species	Origin (co-ordinates[a,b] and altitude)
P9701	21	*C. japonica*	Shionomuro, Imaichi, Tochigi, Japan (36°45'N, 139°48'E[a]; 275 m)
P9702	19	*C. japonica*	Noguchi, Nikko, Tochigi, Japan (36°45'N, 139°43'E[a]; 450 m)
P9724	21	*C. japonica*	Ibuki, Shiga, Japan (35°15'N, 136°20'E[b]; 900 m)
P9801	19	*C. speciosa*	Zhenyuan, Yunnan, China (23°50'N, 101°10'E[b]; 1800–2300 m)
P9802	21	*C. speciosa*	Hutiaoxia, Zhongdian, Yunnan, China (27°50'N, 99°30'E[b]; 1800–2100 m)
P9803	21	*C. cathayensis*	Hutiaoxia, Zhongdian, Yunnan, China (27°50'N, 99°30'E[b]; 1800–2100 m)
P9804	21	*C. cathayensis*	Caojian, Yunlong, Yunnan, China (25°35'N, 98°05'E[b] , 2500–2600 m)
P9805	21	*C. speciosa*	Dali, Yunnan, China (25°35'N, 100°10'E[b]; 1900–2400 m)
P9806	21	*C. thibetica*	Yi'ong, Bomi, Tibet (30°00'N, 95°00'E[b]; 2500 m)

[a] Precise
[b] Estimated

populations (Bartish et al. 1999b, Garkava et al. 2000). Both marker systems appeared to be useful in the discrimination of accessions. Cluster and multidimensional scaling analyses yielded similar results for the two molecular marker methods (Garkava et al. 2000). The two methods also showed a concordant pattern of relative diversity within species. *C. japonica* and *C. speciosa* were found to be considerably more diverse than *C. cathayensis*. However, because of possible interspecific hybridisation among accessions from botanical gardens, it was not clear to what extent the obtained results would be representative for indigenous populations of *Chaenomeles*.

Isozymes and RAPDs are among the most frequently used molecular markers for taxonomic and systematic analyses of plants. Correspondence between results with nuclear DNA and isozyme markers using cluster and ordination analyses of the same plant sample is generally moderate (Heun et al. 1994), but highly correlated estimates of genetic distances have been reported between RAPD and isozyme datasets (Lifante and Aguinagalde 1996). Some authors have considered RAPD-based analysis to be more definitive in its separation of clusters than isozyme-based analysis (Heun et al. 1994), mostly because RAPDs yield lower coefficients of variation than isozymes and thus allow for a higher level of discrimination (Staub et al. 1997). A different view was expressed by Chan and Sun (1997), who suggested that the reliability of RAPD and isozyme data is comparable at the intraspecific level but that at the interspecific level RAPD markers are less suitable for studying phylogenetic relationships than both isozyme and restriction fragment length polymorphism (RFLP) markers.

When population structure is being investigated, different molecular marker systems and the interpretation of the scored fingerprints as phenetic or genetic data may produce different results. Gene diversity estimates obtained from RAPDs may be higher than those obtained from isozymes (Aagaard et al. 1998) because "an inherently higher rate of detectable mutations and weaker degree of selective constraint at RAPD compared to allozyme loci". Furthermore, estimates of among-population differentiation could be inflated when based on RAPD phenotypes compared to RAPD and isozyme genotypes (Isabel et al. 1999). Several studies have also questioned the neutrality of isozyme loci, suggesting that both balancing (Karl and Avise 1992; Raybould et al. 1996) and diversifying (Berry and Kreitman 1993; Lönn 1993; Le Corre et al. 1997) selection may operate on some of the enzymes. The simultaneous application of both techniques thus seems to be a more reliable approach when estimating population parameters and phylogenetic relationships, as indeed has been advocated by, for example, Chan and Sun (1997) and Ayres and Ryan (1999).

The purpose of our investigation was: (1) to determine population structure and gene diversity for species and populations within *Chaenomeles*, and (2) to investigate the correspondence between RAPD- and isozyme-based datasets in analyses of phylogenetic relationships, population structure and relative gene diversity estimates.

Materials and methods

Plant material

All of the accessions were collected in the wild in China and Japan during the autumn of 1997, except for P9801, P9802 and P9803, which were purchased from local markets (the fruits had been collected in the nearby mountains). Only one fruit, with numerous seeds, was picked from each shrub. The seeds from one fruit were presumed to constitute a family. Following cold stratification, the seeds were germinated in a greenhouse. When the seedlings were 3 months old, 7 plants from each of three randomly selected families per population were sampled for molecular analyses. However, one of the populations (P9724), was represented by only one family (of 21 seedlings to achieve a more balanced plant sampling). Altogether, 189 plants representing nine populations of *C. japonica*, *C. speciosa*, *C. cathayensis* and *C. thibetica* were initially sampled (Table 1). One-year old plants were used for isozyme extractions. Since 4 plants died before being sampled for isozyme analysis (1 plant per family in two families of population P9702 and in two families of population P9801), the total number of plants used in both analyses was 185.

Isozyme analysis

Enzymes were extracted from leaflets and phloem tissue, and polyacrylamide gel electrophoresis was carried out as previously described (Garkava et al. 2000).

The enzyme assay procedures for acid phosphatase (ACP, EC 3.1.3.2), and esterases (EST, EC 3.1.1.-) were as described by Wendel and Weeden (1989). Glutamate oxaloacetate transaminase (GOT, EC 2.6.1.1), shikimate dehydrogenase (SKDH, EC 1.1.1.25), and phosphoglucomutase (PGM, EC 5.4.2.2) were stained according to Vallejos (1983). Peroxidase (PRX, EC 1.11.1.7) was stained with 0.09% benzidine in 0.32 *M* sodium acetate buffer adjusted to pH 5.4, and bands were revealed with 0.01% hydrogen peroxide. All isozyme systems were stained in the dark at 37°–38°C. Bands were scored visually as presence and absence of a marker.

DNA extraction and polymerase chain reaction (PCR) amplification

Young unexpanded leaves (5–15 mg per individual) were collected, placed into an Eppendorf tube (1.5 ml) and stored at –80°C. Leaves were ground with a minipestle to a homogenate in 200–250 µl of 2×CTAB extraction buffer and incubated for 45 min at 65°C. From this point on all procedures for DNA extraction and PCR amplification followed the previously described method of Bartish et al. (1999b). The only exception was the use of a different thermocycler, PTC–100 TM (MJ Research).

Ten primers were selected out of 100 (Operon Technologies sets A, B, D, E, F) that had been tested previously (Bartish et al. 1999b). The main criterion for selection was reproducibility of amplification. However, some bias in the direction of higher polymorphism among RAPD markers generated by the primers used in the present study (compared to random sampling of primers) cannot be excluded.

Products of amplification from DNA samples of families from different populations were run on the same gel. In all cases, except population P9803, separate families from the same population were analysed on separate gels. For each primer, three samples of each population were replicated in different PCR experiments and analysed on separate gels. Products of PCR amplification were scored visually as presence or absence of a marker. Bands of identical size, amplified with the same primer, were considered to be homologous. Bands, which were very faint and rare (frequency less than 3%), were not included into the analysis.

Data analysis

Binary data matrices were produced from the scored isozyme and RAPD markers for individual plants. Phenotype frequencies of molecular markers within families of *C. japonica* and within the total sample of populations and species were calculated and used to create datasets of marker frequencies. The datasets were used to compute phenetic (Euclidean) distances among populations (or families in the case of *C. japonica*) according to the method of Harrison et al. (1997) and Hancock and Bringhurst (1979) for RAPD-based and isozyme-based data, respectively. This approach allows evaluation of phylogenetic relationships between populations when precise genetic data for molecular markers are absent.

In addition, a RAPD-based dataset of null allele frequencies within populations was produced for the whole material. The markers were considered to represent separate loci. Null allele frequencies were calculated for these markers in accordance with Lynch and Milligan (1994) and Bartish et al. (1999a). The dataset of RAPD null allele frequencies for populations was then used to compute Nei's genetic distances (Nei 1972) among populations.

All datasets were analysed with NTSYS-pc software (version 1.8, Rohlf 1997). The SAHN programme from this package was used to produce UPGMA dendrograms from each of two phenetic distance matrices and one genetic distance matrix to show the amount of relatedness among populations and species. Principal co-ordinate analysis (DCENTER and EIGENVECTOR programmes) was run for each of the two matrices of phenetic distances among populations, and for two more matrices of phenetic distances among families within *C. japonica*. We used this analysis to obtain additional information about interpopulation relations, which was revealed by different methods. For graphical representation of principal co-ordinate analyses, two-dimensional plots for families and three-dimensional plots for populations were produced by NTSYS.

We used the matrix of null allele frequencies within populations from the RAPD dataset for cluster analysis in PHYLIP 3.573c software (Felsenstein 1993). The matrix was bootstrapped to produce 100 random matrices, and Nei's genetic distances (1972) were calculated by GENDIST from PHYLIP for each of the 100 bootstraps. These matrices were subjected to cluster analysis in the NEIGHBOR programme from PHYLIP with the neighbour-joining algorithm. A strict consensus tree, summarising the relationships described by all 100 resampled datasets (with bootstrap support for each node), was finally produced.

Two of the most frequently used approaches in RAPD analysis were employed to evaluate gene diversities on both RAPD and isozyme datasets: estimates of unbiased values of expected heterozygosity, as in Lynch and Milligan (1994), and Shannon's index, as in Bussell (1999). Monomorphic markers were included into the Shannon's index calculations as suggested by Bussell (1999), in order to standardise values and enable comparisons between studies in which a different number of markers were recorded. We used the modified standardisation procedure suggested by Liu and Furnier (1993) because the diversity of the total plant sample may have a substantial influence on within-population diversity estimations (Bartish et al. 1999c). Since portions of the genome that are monomorphic for the recessive allele cannot be observed, doubling the number of markers monomorphic for the dominant allele (presence of a band) should yield a good estimate of the number of monomorphic markers and hopefully improve the total estimate of within-species polymorphism. We calculated the total number of monomorphic and polymorphic molecular markers separately for each species; thus, the combined number of markers included into all gene diversity estimates was different for each of the species.

The Lynch and Milligan index was calculated for each RAPD marker, i, separately for each population, $H_{pop}(i)$, species, $H_{sp}(i)$, and the genus, $H_{ge}(i)$. The mean Lynch and Milligan index for a population, H_{pop}, was then calculated by averaging $H_{pop}(i)$ over all markers. H_{sp} and H_{ge} were calculated similarly.

Shannon's index was calculated for each RAPD or isozyme marker, i, separately for each population, $H'_{pop}(i)$, species, $H'_{sp}(i)$, and the genus, $H'_{ge}(i)$. The average Shannon's index for a population H'_{pop} was then calculated by averaging $H'_{pop}(i)$ over all markers as in Monaghan and Halloran (1996). H'_{sp} and H'_{ge} were calculated similarly. The mean Shannon's index was then calculated for each marker within a species, $H'_{sp}(i)$, by averaging $H'_{pop}(i)$ over all populations within each species (3 populations for *C. japonica* and *C. speciosa*, 2 populations for *C. cathayensis*) . Similarly $\underline{H}'_{ge/sp}(i)$ was calculated by averaging $H'_{sp}(i)$ over all species, and $H'_{ge/pop}(i)$ by averaging $H'_{pop}(i)$ over all populations within the genus.

We calculated the components of diversity between groups for different levels of taxonomic hierarchy analogous to Bussel (1999). For each locus the component of diversity between populations within species, $G'_{pop/sp}(i)$, was $[H'_{sp}(i)–H'_{sp}(i)]/H'_{sp}(i)$, and the component between populations within the genus, $G'_{pop/ge}(i)$, was $[H'_{ge}(i)–H'_{ge/pop}(i)]/H'_{ge}(i)$. The component of diversity between species within genus, $G'_{sp/ge}(i)$, was correspondingly $[H'_{ge}(i)–H'_{ge/sp}(i)]/H'_{ge}(i)$. Mean estimates of $G'_{pop/sp}$, $G'_{pop/ge}$ and $G'_{sp/ge}$ were calculated by averaging $G'_{pop/sp}(i)$, $G'_{pop/ge}(i)$ and $G'_{sp/ge}(i)$ over all markers. We then calculated 99% confidence intervals for $G'_{pop/sp}$ of each species to reveal markers that deviated significantly from the parametric mean of all markers. The same procedure was also applied to only two populations (P9701 and P9702) of *C. japonica*.

Two matrices of phenetic distances among all individual plants (RAPD and isozyme datasets of polymorphic markers) were used as input distance matrices in AMOVA (Huff et al. 1993), and Φ-statistics was calculated from the variance components. In this paper we only report the proportions of variance differentiating among populations within species ($\Phi_{pop/sp}$), among populations

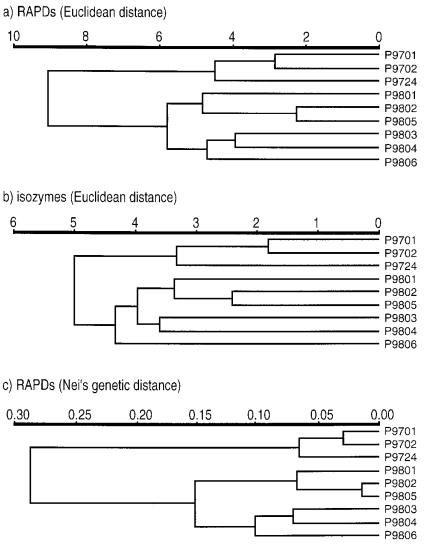


Fig. 1a–c UPGMA dendrograms of RAPD-based (Euclidean distances) (**a**), isozyme-based (Euclidean distances) (**b**) and RAPD-based (**c**) Nei's genetic distances between populations of *Chaenomeles*

within the genus ($\Phi_{pop/ge}$) and among species within genus ($\Phi_{sp/ge}$), which correspond to the *G'*-statistics.

We used Pearson product-moment correlation analysis and the standard *t*-test (paired comparisons), respectively, to evaluate associations between estimates of within-population gene diversity obtained with different molecular datasets. This analysis was also used to estimate associations between some of the molecular markers.

The product-moment correlation and the Mantel test statistic were computed with NTSYS (MXCOMP programme) to measure the degree of relationship between the distance matrices.

Results

Phenetic relationships among populations

For the total plant sample, 271 (including 17 monomorphic) RAPD and 103 (9 monomorphic) isozyme markers were scored. Phenetic (Euclidean) distances between populations varied in the range 2.87–10.07 for RAPD and 1.84–5.77 for isozyme datasets, respectively. Genetic (Nei's) distances based on RAPDs varied from 0.024 to 0.332. The complete matrices of phenetic and genetic distances for all datasets are available on request. All matrices were significantly ($P<0.001$) correlated with each other according to Mantel's tests. Coefficients of correlation varied from $r=0.889$ (between matrices of RAPD and isozyme phenetic distances) to $r=0.921$ (between the matrix of RAPD genetic distances and the matrix of RAPD phenetic distances).

In accordance with the high values of correlation between the matrices of phenetic and genetic distances,

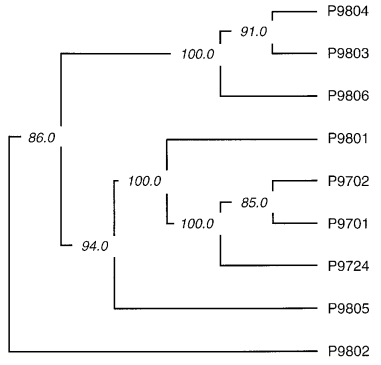


Fig. 2 Dendrogram of phylogenetic relationships between populations of *Chaenomeles* calculated on RAPD null allele frequencies by the neighbour-joining method with bootstrap support (%) for each node of the tree

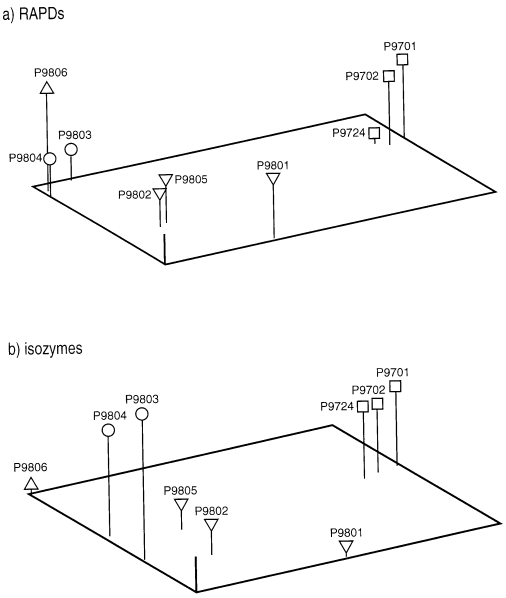


Fig. 3a, b Three-dimensional representation of principal co-ordinate analysis of phenetic relationships between populations of *Chaenomeles*. Percentage explained variability was for **a** the phenetic RAPD-based dataset: PC1 62.4%, PC2 16.0% and PC3 5.9%; **b** the isozyme-based dataset: PC1 44.2%, PC2 17.2% and PC3 12.9%

UPGMA analysis of phylogenetic relationships between populations of *Chaenomeles* produced congruent dendrograms (Fig. 1). Analysis of molecular markers grouped populations from the same species (as inferred from their morphological characters) in the same cluster. This result confirmed a monophyletic origin for all

Table 2 Gene diversity estimated by Shannon's index (*H*') and by the Lynch and Milligan index (*H*) with standard error, as well as the percentage polymorphic markers (*P*) within populations and species of *Chaenomeles*, based on RAPD and isozyme data

Group	H'_{pop} RAPD	H'_{pop} isozyme	H_{pop} RAPD	*P*(%) RAPD	*P*(%) isozyme
C. cathayensis					
P9803	0.140 (0.024)[a]	0.069 (0.025)	0.066 (0.012)	15.8	7.1
P9804	0.282 (0.032)	0.081 (0.026)	0.139 (0.017)	30.3	8.9
C. japonica					
P9701	0.487 (0.032)	0.422 (0.047)	0.229 (0.017)	54.6	50.9
P9702	0.467 (0.032)	0.375 (0.050)	0.212 (0.016)	50.9	44.6
P9724	0.424 (0.031)	0.413 (0.050)	0.196 (0.016)	49.1	46.4
C. speciosa					
P9801	0.547 (0.030)	0.503 (0.049)	0.253 (0.015)	63.5	54.5
P9802	0.520 (0.030)	0.419 (0.044)	0.239 (0.014)	62.4	53.6
P9805	0.589 (0.030)	0.460 (0.044)	0.258 (0.014)	66.1	60.7
C. thibetica					
P9806	0.286 (0.029)	0.189 (0.052)	0.141 (0.015)	34.3	14.3
Group	H'_{sp} RAPD	H'_{sp} isozyme	H_{sp} RAPD	*P*(%) RAPD	*P*(%) isozyme
C. cathayensis	0.337 (0.032)	0.284 (0.056)	0.160 (0.015)	39.9	22.3
C. japonica	0.620 (0.032)	0.612 (0.050)	0.278 (0.016)	69.0	67.9
C. speciosa	0.706 (0.029)	0.655 (0.043)	0.306 (0.014)	78.2	75.9
Group	H'_{ge} RAPD	H'_{ge} isozyme	H_{ge} RAPD	*P*(%) RAPD	*P*(%) isozyme
Chaenomeles	0.746 (0.013)	0.665 (0.024)	0.282 (0.008)	93.7	92.0

[a] Standard error in parenthesis

Table 3 Correlation (Pearson's) between within-population gene diversity values estimated by Shannon's index (H'_{pop}) and by the Lynch and Milligan index (H_{pop}) and percentage polymorphic markers (*P*), based on RAPD and isozyme datasets

	H'_{pop} RAPD	H'_{pop} isozymeH_{pop}	RAPD	*P* RAPD
H'_{pop} isozyme	0.946***			
H_{pop} RAPD	0.996***	0.942**		
P RAPD	0.995***	0.952***	0.993***	
P isozyme	0.963***	0.979***	0.950***	0.963*

* $P<0.05$, ** $P<0.01$, *** $P<0.001$

conspecific populations. Populations of *C. japonica* always clustered separately from the other populations. Results of cluster analysis based on RAPD genetic distances between populations (Fig. 1c) were in good agreement with cluster analyses based on phenetic distances obtained from RAPD and isozyme datasets, respectively (Fig. 1a, b).

High values of bootstrap support were obtained for each node of the phylogenetic tree based on the dataset of genetic distances and the neighbour-joining algorithm (Fig 2). However, populations of *C. speciosa* were widely scattered in this dendrogram.

The pattern of phenetic relationships, revealed by UPGMA cluster analyses, was further supported by principal co-ordinate analysis (Fig. 3). RAPD- and isozyme-based phenetic distances produced results which were concordant in: (1) grouping populations P9701, P9702 and P9724 (*C. japonica)* separately from all other populations; (2) joining populations P9802 and P9805 between population P9801 (*C. speciosa*) and populations P9803 and P9804 (*C. cathayensis*), respectively; (3) showing the relative closeness of population P9806 (*C. thibetica*) to populations P9803 and P9804.

Gene diversity

Analysis of gene diversity by the Lynch and Milligan index (*H*) and Shannon's index (*H*') within populations and species of *Chaenomeles* showed that *C. japonica* and *C. speciosa* were substantially more diverse than *C. cathayensis* and *C. thibetica*, independent of the method used for evaluation (Table 2). Percentages of polymorphic markers (*P*) and values of Shannon's index were similar when RAPD- and isozyme-based datasets were compared, although RAPD-based estimates were generally somewhat higher. The only clear difference among estimates of within-population gene diversity was found in population P9804. According to RAPD-based estimates, P9804 was approximately twice as diverse as P9803,

Table 4 Components of diversity among populations within species, among populations within genus and among species within genus, estimated from phenetic RAPD and isozyme datasets by AMOVA (Φ-statistics with *P*-value) or derived from Shannon's index (*G'*-statistics with standard error)

Group variable	$G'_{pop/sp}$ RAPD	$G'_{pop/sp}$ isozyme	$\Phi_{pop/sp}$ RAPD	$\Phi_{pop/sp}$ isozyme
Populations within				
C. cathayensis	0.606 (0.036)	0.830 (0.026)	0.229 (0.001)	0.615 (0.001)
C. japonica	0.278 (0.031)	0.323 (0.036)	0.126 (0.367)	0.178 (0.001)
C. speciosa	0.176 (0.027)	0.268 (0.032)	0.100 (0.257)	0.181 (0.001)
Group variable	$G'_{pop/ge}$ RAPD	$G'_{pop/ge}$ isozyme	$\Phi_{pop/ge}$ RAPD	$\Phi_{pop/ge}$ isozyme
Populations	0.535 (0.013)	0.491 (0.019)	0.325 (0.001)	0.404 (0.001)
Group variable	$G'_{sp/ge}$ RAPD	$G'_{sp/ge}$ isozyme	$\Phi_{sp/ge}$ RAPD	$\Phi_{sp/ge}$ isozyme
Species	0.446 (0.015)	0.435 (0.018)	0.303 (0.001)	0.312 (0.001)

[a] Standard error in parenthesis

whereas values of gene diversity for these two populations were similar in the isozyme dataset (Table 2).

In the absence of precise genetic information on individual loci and the lack of random sampling of molecular markers since selected primers were used for RAPD analysis, we refrained from trying to compare absolute values of gene diversity estimates between the two molecular methods. Instead, the Shannon's index values for population diversity estimated by RAPDs and isozymes, respectively, were tested by a *t*-test (paired comparisons). Our results clearly showed (Table 3) that the null hypothesis of the two sets of molecular markers being independent of each other could be rejected with a high level of significance ($P<0.001$). This was also true for the comparison of the RAPD-based Lynch and Milligan (H_{pop}) and Shannon's index (H'_{pop}) diversity estimates ($P<0.001$, Table 3).

Partitioning of components of molecular variance

G'-statistics or Φ-statistics were used to evaluate the variability within and between taxa for RAPDs and isozymes, respectively. The analysis was carried out for three species separately (*C. thibetica* was not included since it was represented by only one population) and for the whole genus. Between-population variability was always lower when estimated by Φ-statistics than when estimated with *G'*-statistics (Table 4). *G'*-statistics obtained with the RAPD data resulted in higher estimates of differentiation among species and among populations across the entire genus, whereas isozyme-based values resulted in higher differentiation within species. Estimates from Φ-statistics for isozymes provided a better discrimination between groups of plants at all levels of taxonomic hierarchy than RAPDs.

The major part of the molecular variability could be attributed to the between-population component for *C. cathayensis*, as estimated from *G'*-statistics and from Φ-statistics for isozyme markers (Table 4). In contrast, much less of the total molecular variability was attributed to this component for *C. japonica* and *C. speciosa*. Within- and between-species components of total vari-

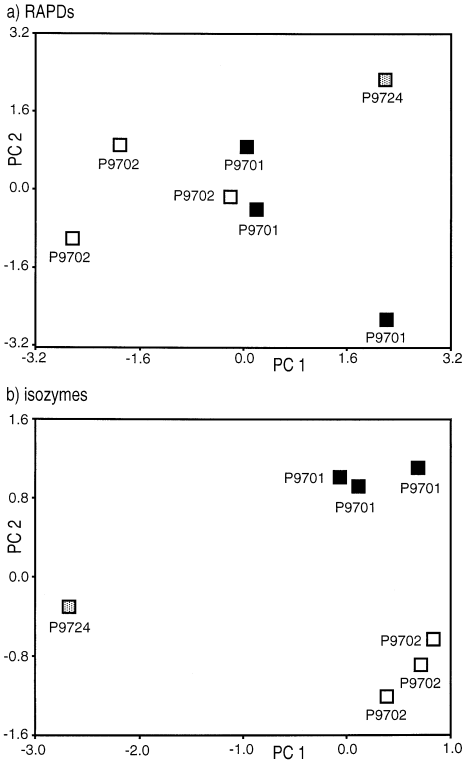


Fig. 4a, b Plot of principal co-ordinates of phenetic distances between families of *C. japonica.* **a** RAPD-based dataset: PC1 27.3%, PC2 20.1%; **b** isozyme-based dataset: PC1 39.3%, PC2 25.9%

Table 5 Differentiation between conspecific populations derived from Shannon's index [$G'_{pop/sp}(i)$] for individual markers, which deviate significantly ($P<0.01$) from the parametric means for the total set of markers. Markers beginning with OP are RAPD markers, the reminder are isozyme markers

Species	Marker	$G'_{pop/sp}(i)$
C. japonica	PEst–2	0.726
	PPrx–7	0.842
	PAcp–8	0.741
	OPA18.220	0.682
	OPF08.800	0.837
C. speciosa	PAat–9	0.689
	PAat–10	0.689
	OPE07.380	0.726
C. japonica (P9701 and P9702)	LEst–2	0.753
	LEst–3	0.676
	OPE07.1050	0.473
	OPD08.1070	0.508

ability in *Chaenomeles* were partitioned more evenly when derived from Shannon's index compared to AMOVA, with slightly more of the total variance distributed within species. The within-species component was considerably higher in the *Φ*-statistics (Table 4).

Correlations between estimates of the among-population (species) components were relatively high and significant when different methods of calculations were compared (RAPDs: *G'* vs. *Φ*, $r=0.805$, $P<0.05$; isozymes : *G'* vs. *Φ*, $r=0.983$, $P<0.01$).

Differentiation between families of *C. japonica*

RAPD- and isozyme-based matrices of phenetic distances between families of *C. japonica* were not significantly correlated ($r=0.479$, $P=0.061$). We also found substantial discrepancies between results of principal co-ordinate analyses applied to different matrices of phenetic distances. A two-dimensional plot of principal co-ordinates for families of *C. japonica* revealed only a weak population structure when RAPDs were analysed (Fig. 4a). In contrast, a clear subdivision of families into separate groups was observed for isozymes, with families from the same population grouping together (Fig. 4b).

Deviating markers

G'-statistics for individual molecular markers were used to reveal those markers, which deviated significantly ($P<0.01$) from the parametric mean for the total set of markers (Table 5). Three such isozyme markers were found for *C. japonica* (PEst–2, PPrx–7, and PAcp–8; 3.7% of the total set of markers) and 2 for *C. speciosa* (PAat–9, PAat–10; 2.3% of the total set of markers). Only 2 such RAPD markers were found for *C. japonica* (OPF08.800, OPA18.220) and 1 for *C. speciosa* (OPE07.380), comprising about 1% and 0.5% of the total set of markers, respectively. No isozyme or RAPD marker deviated significantly from the parametric mean of the *G'*-statistics for *C. cathayensis*.

When only populations P9701 and P9702 were analysed, 2 esterase markers (LEst–2 and LEst–3) and 2 RAPD markers (OPE07.1050 and OPD08.1070) deviated significantly from the parametric means for the total set of markers (Table 5).

Markers that deviated significantly from the parametric mean were found to co-occur in the same plants. Within *C. japonica* 3 isozyme (PEst–2, PPrx–7, and PAcp–8) and 1 RAPD marker (OPA18.220) were strongly associated (correlation among markers ranged from 0.74 to 0.83). Associations of deviating markers within *C. speciosa* were even stronger. Both isozyme markers (PAat–9, PAat–10) were absolutely negatively correlated ($r=-1.0$) and the correlation between OPE07.380 and the isozyme markers was high ($r=0.81$).

Discussion

Phylogenetic relationships among populations

From previous publications it is still not clear to what extent results from the different molecular marker systems are congruent. In direct comparative analyses, correlations between estimates of genetic distances, obtained from isozyme- and RAPD-based data, have proven to be significant, with estimates of correlation coefficients from moderate (Heun et al. 1994, $r=0.34$) to high (Lifante and Aguinagalde 1996, $r=0.83$). In general, isozyme- and RAPD-based estimates of gene diversity have also proven to be in good correspondence, but some discrepancies have been found in virtually all studies.

In the present study we found that correlation between distance estimates, obtained from isozyme and RAPD datasets, varied according to the geographic distances between the plant samples analysed. It could be very high, as in the case of the total sample of nine populations from four species ($r=0.891$, $P<0.001$), and it could be lower and insignificant ($r=0.479$, $P=0.061$), as in the case of families from the three populations of *C. japonica*.

The phenetic estimates of phylogenetic relationships are independent of Hardy-Weinberg equilibrium (Bussel 1999). Conversely, RAPD-based genetic estimates are dependent on this assumption. Therefore, the high correspondence between phenetic and genetic estimates obtained for all populations in the RAPD-based dataset may indicate that (1) the underlying assumption of Hardy-Weinberg equilibrium for calculation of genetic estimates do not influence the phylogenetic relationships revealed in this study or (2) the studied populations are at an equilibrium. We believe that the first alternative is the most likely since *C. cathayensis* and *C. thibetica* are supposed to have a mixed mating system and, furthermore, we have found some indications of interspecific hybridisation between *C. speciosa* and *C. cathayensis*. Both of these phenomena are known to cause deviations from the Hardy-Weinberg equilibrium. Violation of the equilibrium can, however, not be confirmed directly in this study

since we did not have access to precise genetic information about the isozyme markers that would enable the calculation of fixation indices.

The most important discrepancy between the dendrograms of phylogenetic relationships between populations was observed for *C. speciosa*. Two populations of this species (P9802 and P9805) clustered together with a third population (P9801) in the UPGMA dendrograms, but in the neighbour-joining based dendrogram they were intermediate between the *C. cathayensis* (P9803, P9804) – *C. thibetica* (P9806) cluster on one hand and the *C. japonica* (P9701, P9702, P9724) – *C. speciosa* (P9801) cluster on the other hand (Fig. 2). However, based on a more detailed analysis of genetic relationships between families in these two populations and in populations P9801, P9803 and P9804, spontaneous interspecific hybridisation between *C. cathayensis* and *C. speciosa* can be suspected (manuscript in preparation). This is probably the reason for the discrepancies observed between the two different representations.

The phylogenetic relationships between species of *Chaenomeles*, inferred from previous studies based on plant accessions from botanical gardens (Bartish et al. 1999b, Garkava et al. 2000), were completely congruent with the present results based on samples of naturally growing plants. This finding makes it possible to infer that the possible spontaneous hybridisations among plants cultivated in botanical gardens have not changed the taxonomic status of the species and that the taxonomic identities of the previously analysed accessions from several botanical gardens (Bartish et al. 1999b, Garkava et al. 2000) were correct.

Gene diversity

In accordance with previous reports (Liu and Furnier 1993; Peakall et al. 1995; Aagaard et al. 1998), our results showed that, in general, RAPDs and isozymes reveal similar patterns of gene diversity and that these markers probably experience similar evolutionary forces in *Chaenomeles*. Estimates of polymorphism of both RAPD and isozymes markers were very high and similar for the whole genus (93.7% and 92.0%, respectively, Table 2). Thus, relatively few molecular markers were fixed in the sampled populations.

Contrasting gene diversity estimates were obtained for *C. japonica* and *C. speciosa* on one hand and for *C. cathayensis* and *C. thibetica* on the other hand (Table 2). This is rather common in plants when congeneric species have contrasting mating systems (Barrett 1989; Gottlieb 1973). Our findings thus imply that outcrossing may be the prevailing mating system for *C. japonica* and *C. speciosa*, whereas *C. cathayensis* and *C. thibetica* may have a mixed mating system. Both selfing and biparental inbreeding may be suggested for the latter species to explain the observed distribution of variance, but self-incompatibility has been suspected for *C. cathayensis* (Weber 1964). Therefore, biparental inbreeding seems to be a more probable mating system for this species. Direct investigation of the mating systems within *Chaenomeles* would enable us to draw more convincing conclusions. Since our *C. cathayensis* and *C. thibetica* plant material is still juvenile, we have not yet been able to perform the necessary study. Using the more conventional method to calculate H_{pop}-values (i.e. including only markers which are polymorphic within species) yielded less differentiation between *C. japonica* (0.254) and *C. speciosa* (0.275) on the one hand, and *C. cathayensis* (0.214) and *C. thibetica* (0.238) on the other hand (results not shown in further detail). Obviously, the inclusion of monomorphic markers into our main set of calculations, according to Liu and Furnier (1993), instead emphasizes the homogeneity of the latter two species.

Shannon's index gene diversity estimates obtained with isozymes and RAPDs separately, for populations of *C. japonica* and *C. speciosa* were relatively similar within each species (Table 2). By contrast, diversity estimates based on RAPD markers showed considerably more differentiation between populations of *C. cathayensis* than did isozymes. Furthermore, a relative difference in gene diversity estimates between populations of *C. japonica* and *C. speciosa* on one hand and *C. cathayensis* and *C. thibetica* on the other was much more pronounced for isozyme markers, with two- to threefold higher relative differences (Table 2). This may indicate balancing selection on some of the isozyme markers in the presumably partly inbreeding species *C. cathayensis* and *C. thibetica*. A similar phenomenon has been suggested for oysters by Karl and Avise (1992) based on gene diversity estimates with isozymes and nuclear RFLP.

Relative gene diversity estimates derived by the Lynch and Milligan formula and by Shannon's index, respectively, were similar. The assumption of Hardy-Weinberg equilibrium therefore did not seem to influence the relative estimates when RAPD null-allele frequencies were used in the diversity analyses for the total sample of populations. This is in accordance with our findings in the phylogenetic analysis.

Partitioning of components of molecular variance

G'-statistics and *Φ*-statistics are analogous to *F*-statistics and can be used to analyse haplotypic diversity of molecular markers at different levels of hierarchial subdivisions. We found that the between-population estimates from *Φ*-statistics were always lower than corresponding estimates from *G'*-statistics (Table 4). In one case we obtained a particularly low estimate of the between-population component of total variance (*C. cathayensis*, RAPD dataset, Table 4). However the *G'*-statistics and the *Φ*-statistics were highly correlated (RAPDs: r=0.805, P<0.05; isozymes r=0.983, P<0.01).

Isozyme analysis (reviewed by Hamrick and Godt 1989) and RAPD analysis (Bartish et al. 1999a) have shown that outcrossing species, in general, retain most of their genetic variability within populations. We found that within-population components of total variability for *C.*

japonica and *C. speciosa* were always higher than between-population components, independent of kind of molecular markers or method for statistical evaluation. This is in agreement with the general result for outcrossing species. In contrast, *C. cathayensis* almost always yielded comparatively higher values for between-population components. This may indicate a mixed mating system and partial inbreeding in this species, as has also been suspected from the overall estimates of gene diversity.

Deviation of some isozyme markers from the neutral model

Contrasting patterns of genetic differentiation between conspecific populations have been previously revealed in plants as a result of comparison between isozyme and DNA markers (Raybould et al. 1996; Latta and Mitton 1997). In both these studies, DNA markers (RFLP or RAPD) suggested a higher rate of differentiation between populations than did isozymes. Balancing selection on some of the isozyme markers was brought forward as a possible explanation.

In our study, a comparative analysis of population structure within *C. japonica* by RAPD and isozyme markers also revealed substantial differences between the methods. However, isozyme markers were, in general, more differentiated than RAPDs. If we assume neutrality of RAPDs, then diversifying selection acting on some of the isozyme markers could be suspected. A similar explanation has been suggested by Le Corre et al. (1997) who compared RAPD and isozyme diversity estimates among populations of sessile oak. This is also in accordance with the general prediction made by Lewontin and Krakauer (1973) and the diversifying selection found to be acting on some plant isozymes (Lönn 1993).

The component of diversity between populations derived from Shannon's index for individual isozyme and RAPD markers was useful to find markers that deviated significantly from the parametric mean for the total set of markers (Table 5). Such markers, especially isozymes, may be involved in the adaptation to environmental stress factors (Lönn 1993). Alternatively, none of these markers represent genes directly involved in adaptation to local environment, but they are "hitchhiking" with such genes through selection due to linkage disequilibrium (Aquadro et al. 1994). The association between most of the deviating markers to such a high extent that 4 of 5 markers within *C. japonica* and all 3 markers within *C. speciosa* may be associated with the same linkage groups supports such a scenario. A note of precaution should be made however. Since our confidence interval for distribution of $G'_{pop/sp}(i)$ values accounts for sampling variation only and not for stochastic variation (Slatkin and Arter 1991), it may be incorrect to assume that selection is the only possible source of significant deviations. Difference in mutation rates between markers should also be taken into account.

Acknowledgements We are most grateful to Dr. Munetaka Sugiyama, Botanical Gardens, University of Tokyo, Japan, and Prof. Roland von Bothmer, Department of Plant Breeding, Swedish University of Agricultural Sciences, Svalöv, Sweden, who collected seeds of *C. japonica*. We are likewise very grateful to Prof. Hong De Yuan and Dr. Hong Yu, Institute of Botany, Chinese Academy of Sciences, Beijing, China, for providing seeds of the Chinese species. IVB and LPG obtained grants from the Swedish Institute and the Swedish Royal Academy of Sciences. This study was also carried out with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD programme, CT 97–3894, 'Japanese quince (*Chaenomeles japonica*) – a new European fruit crop for production of juice, flavour and fibre'. It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. The experiments performed comply with current Swedish laws.

References

Aagaard JE, Krutovskii KV, Strauss SH (1998) RAPDs and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. Heredity 81: 69–78

Anonymous (1989) Medicinal plants in China – a selection of 150 commonly used species, WHO Regional Publ, Western Pacific Ser No. 2, Manila, pp 72–73

Aquadro CF, Begun DJ, Kindahl EC (1994) Selection, recombination, and DNA polymorphism in *Drosophila*. In: Golding B (ed) Non-neutral evolution: theories and molecular data. Chapman and Hall, New York, pp 46–56

Ayres DR, Ryan FJ (1999) Genetic diversity and structure of the narrow endemic *Wyethia reticulata* and its congener *W. bolanderi* (Asteraceae) using RAPD and allozyme techniques. Am J Bot 86:344–353

Barrett SCH (1989) Mating system evolution and speciation in heterostylous plants. In: Otte D, Endler JA (eds) Speciation and its consequences. Sinauer, Sunderland, Mass., pp 257–283

Bartish IV, Jeppsson N, Nybom H (1999a) Population genetic structure in the dioecious pioneer plant species *Hippophae rhamnoides* investigated by random amplified polymorphic DNA (RAPD) markers. Mol Ecol 8:791–802

Bartish IV, Rumpunen K, Nybom H (1999b) Genetic diversity in *Chaenomeles* (Rosaceae) revealed by RAPD analysis. Plant Syst Evol 214:131–145

Bartish IV, Jeppsson N, Bartish GI, Lu R, Nybom H (1999c) Inter- and intraspecific genetic variation in *Hippophae* (Elaeagnaceae) investigated by RAPD markers In: Jeppson N (ed) Genetic variation and fruit quality in sea buckthorn and black chokeberry. PhD thesis, Swedish University of Agricultural Sciences, Agraria 199, Alnarp

Berry A, Kreitman M (1993) Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. Genetics 134:869–893

Bussell JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobeliaceae). Mol Ecol 8:775–789

Chan KF, Sun M (1997) Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of *Amaranthus*. Theor Appl Genet 95:865–873

Felsenstein J (1993) PHYLIP (Phylogeny interference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, Wash.

Garkava LP, Rumpunen K, Bartish IV (2000) Genetic relationships in *Chaenomeles* (Rosaceae) revealed by isozyme analysis. Sci Hortic 85:21–35

Gottlieb LD (1973) Genetic differentiation, sympatric speciation and the origin of a diploid species of *Stephanomeria*. Am J Bot 60:545–553

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, Mass., pp 43–63

Hancock JF Jr, Bringhurst RS (1979) Ecological differentiation in perennial octoploid species of *Fragaria*. Am J Bot 66: 367–375

Harrison RE, Luby JJ, Furnier GR, Hancock JF (1997) Morphological and molecular variation among populations of octoploid *Fragaria virginiana* and *F. chiloensis* (Rosaceae) from North America. Am J Bot 84:612–620

Heun M, Murphy JP, Phillips TD (1994) A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. Theor Appl Genet 87: 689–696

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.) Engelm.]. Theor Appl Genet 86: 927–934

Isabel N, Beaulieu J, Theriault P, Bousquet J (1999) Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. Mol Ecol 8:477–483

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256:100–102

Latta RC, Mitton JB (1997) A comparison of population differentiation across four classes of gene marker in Limber pine (*Pinus flexilis* James). Genetics 146:1153–1163

Le Corre V, Dumolin-Lapegue S, Kremer A (1997) Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. Mol Ecol 6:519–529

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195

Lifante ZD, Aguinagalde I (1996) The use of random amplified polymorphic DNA (RAPD) markers from the study of taxonomical relationships among species of *Asphodelus* sect. *Verinea* (Asphodelaceae). Am J Bot 83:949–953

Liu Z, Furnier GR (1993) Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor Appl Genet 87:97–105

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Mol Ecol 3:91–99

Lönn M (1993) Genetic structure and allozyme-microhabitat associations in *Bromus hordeaceus*. Oikos 68:99–106

Monaghan BG, Halloran GM (1996) RAPD variation within and between natural populations of morama [*Tylosema esculentum* (Burchell) Schrieber] in Southern Africa. S Afr J Bot 62: 287–291

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Mol Ecol 4: 135–147

Phipps JB, Smith PG, Rohrer JR (1990) A checklist of the subfamily Maloideae (Rosaceae). Can J Bot 68:2209–2269

Raybould AF, Mogg RJ, Clarke RT (1996) The genetic structure of *Beta vulgaris* ssp. *maritima* (sea beet) populations: RFLPs and isozymes show different patterns of gene flow. Heredity 77:245–250

Rohlf FJ (1997) NTSYS/pc. Numerical taxonomy and multivariate analysis system. Version 1.80. Exeter Software, Setauket, N.Y.

Rumpunen K, Kviklys D, Kaufmane E, Garkava L (1999) Breeding *Chaenomeles* – a new aromatic fruit crop. In: Tobutt KR, Alston FH (eds) Proc Eucarpia Symp Fruit Breed Genet. Acta Hortic 484:211–216

Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population genetics. Am Nat 138:499–517

Staub JE, Box J, Meglic V, Horejsi T, McCreight JD (1997) Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in *Cucumis*. Genet Resour Crop Evol 44:257–269

Vallejos CE (1983) Enzyme activity staining. In: Tanksley SD, Orton TJ (eds) Isoenzymes in plant genetics and breeding, Part A. Elsevier Science Publ, Amsterdam, pp 469–516

Wang, JX, Wang XL, Guan ZG, Teng ZQ (1998) Preliminary investigation and classification of *Chaenomeles* cultivars (in Chinese). J Beijing Forestry Univ 20:123–125

Weber C (1963) Cultivars in the genus *Chaenomeles*. Arnoldia 23:17–75

Weber C (1964) The genus *Chaenomeles* (Rosaceae). J Arnold Arbor 45:161–205, 302–345

Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, London, UK, pp 5–45

Yü TT, Kuan KC (1963) Taxa nova Rosacearum sinicarum I. Act Phytotaxon Sin 8:214–220

III

Genetic Resources and Crop Evolution (2005) 52: 723–735
DOI 10.1007/s10722-003-6123-4

RAPD-based analysis of genetic diversity and selection of lingonberry (*Vaccinium vitis-idaea* L.) material for *ex situ* conservation

L. Garkava-Gustavsson[1,*], H.A. Persson[1], H. Nybom[1], K. Rumpunen[1], B.A. Gustavsson[1] and I.V. Bartish[2]

[1]*Balsgård–Department of Crop Science, Swedish University of Agricultural Sciences, Fjälkestadsvägen 459, 29194 Kristianstad, Sweden;* [2]*Department of Phanerogamic Botany, Swedish Museum of Natural History, Box 50007, 10405 Stockholm, Sweden; *Author for correspondence (e-mail: Larisa.Gustavsson @vv.slu.se; fax: +46 44265830)*

Received 29 July 2003; accepted in revised form 28 November 2003

Key words: Genetic relatedness, Germplasm collection, Molecular markers, Population genetics

Abstract

Random amplified polymorphic DNA markers were used to assess relatedness and genetic diversity for 15 lingonberry (*Vaccinium vitis-idaea*) populations. Seven primers yielding 59 polymorphic bands were used to analyse 13 populations, representing ssp. *vitis-idaea* from Sweden, Finland, Norway, Estonia and Russia, and two populations, representing ssp. *minus* from Japan and Canada. A cluster analysis and a multi-dimensional scaling analysis (MDS) showed similar phenetic patterns among populations, with a pronounced geographic grouping in most cases. Significant correlations were obtained between geographic and genetic distances for the entire set of populations as well as for the 13 ssp. *vitis-idaea* populations. Mean within-population diversity was 0.206 when estimated with Lynch and Milligan's index, and 0.431 when estimated with Shannon's index, which is in agreement with the mixed mating system reported for lingonberry. Within-population variability accounted for 68.6% of the total variance when all populations were included, and for 78.8% when only populations of ssp. *vitis-idaea* were analysed. Two different approaches were applied to the selection of plant material for a potential gene bank: (1) a hierarchical sampling strategy based on a cluster analysis and (2) the Maximum genetic diversity program, developed for the establishment of core collections. Random sampling was undertaken for comparisons with the selected data sets. The most diverse and representative set of lingonberry specimens was obtained when samples were selected with the Maximum diversity program.

Introduction

Lingonberry, *Vaccinium vitis-idaea* ($2n = 24$), is a perennial, evergreen dwarf shrub, which belongs to the Ericaceae family. It grows preferably on acid soils in coniferous forests and bogs. Lingonberry is widely distributed in boreal and subarctic regions of the northern hemisphere. The species is represented by two subspecies: ssp. *vitis-idaea* L. and ssp. *minus* (Lodd.) Hultén (1949). Plant size constitutes the main difference between the two subspecies. Plants of *V. vitis-idaea* ssp. *vitis-idaea* are 25–30 cm in height, whereas plants of ssp. *minus* are approximately 20 cm high (Fernald 1950). The evergreen leaves of lingonberry are ovate, dark green, thick and leathery. The flowers are campanulate, white and nectariferous, held in drooping racemes. Lingonberry reproduces vegetatively through rhizomes (subterranean runners) and sexually by seeds. The species is partially self-fertile (Jacquemart and Thompson 1996), but seed-set is reported to be higher after cross-pollination

(B.A. Gustavsson, pers. observ.). Lingonberry attracts primarily honey-bees and bumble bees as pollinators (Jacquemart 1993). The berries are bright red, 6–10 mm in diameter. They are quite tart, and therefore not suitable for the fresh fruit market. By contrast, they taste very good when cooked, and are widely used for processing in jam, juice, liqueur and yoghurt. Lingonberry products are especially popular in the Nordic countries, but are becoming more accessible in several other countries (Gustavsson 1999). Lingonberries for processing are usually picked from plants growing wild in natural populations, but there is increased interest in commercial plantations of registered cultivars (Gustavsson 1999).

A breeding program for improvement of lingonberry cultivars was initiated at Balsgård, The Swedish University of Agricultural Sciences, in the 1970s. This program was based mainly on evaluation and selection of superior plants raised from seeds from wild-picked berries. Extensive collections of lingonberry germplasm were gathered from natural populations in Sweden, Finland, the Baltic countries, Russia, North America and Japan (Gustavsson 1997). These plants were grown in comparative field trials at Balsgård. Growth habit, fruit yield, fruit size, fruit quality, winter hardiness, regeneration ability and disease resistance are important characters that were considered in the screening. Selected plants were used to raise a second generation through open-pollination to provide even more superior plants (Gustavsson 1999). Unfortunately, limited resources resulted in a temporary interruption of the lingonberry program at Balsgård. At present, the most important task is to select and preserve a representative germplasm collection of field-grown and genetically identifiable genotypes for future use in, for example, chemical screenings and applied plant breeding. This collection should have the same properties as a 'core collection', a limited set of accessions with maximal genetic diversity and a minimum of repetitions (Frankel 1984).

Morphological characters have traditionally been used to obtain information on variation within plant species. These characters are usually controlled by many loci and may be affected by environment, which can complicate the evaluation of genetic diversity. By contrast, molecular markers are not generally influenced by environment; they are often but not always selectively neutral and, if chosen carefully throughout the genome, supposedly unbiased. Several classes of molecular markers have been used to evaluate the genetic diversity in collections of genetic resources in horticultural crops, such as isozymes, amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), sequence tagged microsatellite sites (STMS) and random amplified polymorphic DNA (RAPD). We chose RAPD, since this method does not require previous knowledge of DNA sequences, is easy to perform and is one of the most cost-effective methods for obtaining polymorphic markers in many plant genera (Lu et al. 1996; Virk et al. 2000; Goulao et al. 2001; Nybom 2003). Studies, in which RAPD-based results have been compared to results obtained with other classes of DNA-based markers in the same plant material, have demonstrated high levels of correspondence, both for genetic distances and for relative estimates of between- and within population diversity (Nybom in press). Despite their dominant nature, RAPDs have already proven to be suitable in studies on the extent of clonality and the amount and partitioning of genetic diversity in lingonberry (Persson and Gustavsson 2001).

In the present study, RAPD markers were used (1) to assess the relationships among 15 lingonberry populations assembled from Sweden, Finland, Norway, Estonia, Russia, Japan and Canada; (2) to estimate levels of genetic diversity and its partitioning into within- and between-population components; (3) to select a representative set of plant specimens for *ex situ* conservation from these 15 populations.

Materials and methods

Plant material

A total of 211 plants, representing 15 populations, was analysed (Table 1). Populations 1 to 13 represent *V. vitis-idaea* ssp. *vitis-idaea*; populations 14 and 15 represent *V. vitis-idaea* ssp. *minus*. In addition, one plant was analysed from each of eight vegetatively propagated cultivars, chosen so as to represent a large geographic range and possibly also large genetic diversity: 'Sanna', 'Sussi', 'Linnea',

Table 1. Analysed populations of lingonberry (*Vaccinium vitis-idaea* L.), number of plants sampled, location of population, within-population gene diversity (including standard error) estimated by Lynch and Milligan's index (H_{pop}) and Shannon's index (H'_{pop}).

Sampling site no.	Subspecies	Sampling site code	No. of plants	Country	Location	Latitude (N)	Longitude (E/W)	H_{pop}	H'_{pop}
1	*vitis-idaea*	SK	15	Sweden	Kristianstad	56°13′	14°12′	0.227 (0.026)	0.483 (0.049)
2	*vitis-idaea*	SÖ	15	Sweden	Örebro	59°24′	14°39′	0.214 (0.025)	0.517 (0.051)
3	*vitis-idaea*	SV	15	Sweden	Västerbotten	63°37′	19°51′	0.245 (0.027)	0.513 (0.050)
4	*vitis-idaea*	SN	10	Sweden	Norrbotten	66°42′	19°33′	0.197 (0.026)	0.400 (0.051)
5	*vitis-idaea*	SG	13	Sweden	Gävleborg	60°18′	16°46′	0.248 (0.028)	0.523 (0.048)
6	*vitis-idaea*	SH	11	Sweden	Halland	57°05′	13°20′	0.219 (0.026)	0.500 (0.051)
7	*vitis-idaea*	FT	15	Finland	Toijala	60°13′	24°10′	0.187 (0.026)	0.375 (0.049)
8	*vitis-idaea*	FS	15	Finland	Simo	65°41′	25°01′	0.178 (0.026)	0.374 (0.050)
9	*vitis-idaea*	NS	15	Norway	Sognedal	61°12′	7°05′	0.225 (0.029)	0.434 (0.055)
10	*vitis-idaea*	EV	15	Estonia	Võru	57°55′	27°03′	0.241 (0.029)	0.456 (0.055)
11	*vitis-idaea*	EP	10	Estonia	Pärnu	58°25′	24°40′	0.209 (0.027)	0.434 (0.055)
12	*vitis-idaea*	RM	15	Russia	Murmansk	68°55′	33°05′	0.110 (0.024)	0.190 (0.043)
13	*vitis-idaea*	RK	14	Russia	Kirov	58°53′	49°30′	0.135 (0.024)	0.265 (0.047)
14	*minus*	JF	14	Japan	Fuji San	35°20′	138°45′	0.180 (0.028)	0.349 (0.056)
15	*minus*	CM	15	Canada	Montreal	45°50′	73°50′	0.274 (0.025)	0.654 (0.048)
								$\bar{x} = 0.206$;	$\bar{x} = 0.431$.

'Ida', 'Koralle', 'Splendor', 'Erntesegen' and 'Masovia'. Development of commercial lingonberry cultivars has been quite restricted in comparison with other small fruit crops and only 17 named varieties were known in 1999, all of which had been derived by open pollination and usually by selection directly in the wild (Gustavsson 1999). 'Sanna', 'Sussi', 'Linnea' and 'Ida' are Swedish cultivars, selected for superior horticultural characters at Balsgård from the wild stands in the province of Småland, Sweden. 'Koralle', originally an ornamental cultivar, was selected in Netherlands from wild stands. 'Splendor' was selected in Wisconsin, USA, from plant material collected in Finland. 'Erntesegen' was selected from wild stands in Germany, and 'Masovia' was selected in Poland.

RAPD analysis

Young leaves were collected and stored at −80 °C. The extraction procedure followed the protocol of Nybom and Schaal (1990). PCR reactions were performed as described by Persson and Gustavsson (2001). The DNA fragments were separated by electrophoresis in 1.8% agarose gels with a Tris–phosphate–EDTA (TPE) buffer system. The gels were stained with ethidium bromide, and the amplification products were visualised under UV light and documented with Polaroid photography for further analyses. Molecular Weight Marker VI (Roche Diagnostics Corp., Indianapolis, IN) was used to determine the size of the DNA fragments.

Eighty decamer primers (Operon Technologies, Alameda, CA) had previously been checked for their suitability in terms of polymorphism, reproducibility and clarity of the obtained patterns (Axelsson 2000). Based on these results, seven primers (OPA-10, OPA-11, OPB-19, OPC-15, OPD-05, OPE-10 and OPF-07) were chosen for further analysis of all 219 plants. To monitor the reproducibility between runs, DNA of the same three plants was included in every run. In addition, amplification of two different samples was carried out twice in each PCR run.

Statistical analyses

Amplification products were scored manually. Each RAPD band was treated as an independent locus with two alleles, presence (1) or absence (0) of a band. A binary matrix based on 59 polymorphic RAPD bands was thus generated.

To evaluate the informativeness of each RAPD primer, a RAPD primer index was produced as in Ghislain et al. (1999): polymorphic index content (PIC) was calculated for each RAPD band as $PIC = 1 - p^2 - q^2$ where p is band frequency and q is no-band frequency. The RAPD primer index is

the sum of PIC values for all bands generated by the same primer.

Two approaches were used to analyse the relationships among populations: cluster analysis and multidimensional scaling analysis (MDS) (SPSS Data Analysis Package 4.0, Norusis 1990). Cluster analysis produces a hierarchical representation of the relationships among groups of organisms, whereas MDS produces a non-hierarchical view of these relationships. Frequencies for each RAPD band were calculated for each of the 15 populations separately. A COSINE pattern similarity measure (Norusis 1990) was used to produce a similarity matrix for comparisons between populations. The matrix was used to perform a cluster analysis (UPGMA, Norusis 1990) and produce a dendrogram, which illustrates the similarity among populations. For the MDS analysis, a Euclidean dissimilarity coefficient matrix was generated from the band frequencies. A scaling solution with three dimensions was used (Norusis 1990).

To analyse possible relationships between genetic and geographic distances, a matrix of geographic distances between population was generated as in Bartish et al. (2000b). Pairwise Φ_{ST} coefficients, which are interpreted as being analogous to F_{ST} values, were computed and a population-based matrix was produced (AMOVA, WINAMOVA version 1.55; Excoffier et al. 1992). Correlation analyses between the geographic distance matrix and the matrix of pairwise Φ_{ST} coefficients were carried out with Mantel tests (MXCOMP in NTSYS-pc, 9999 permutations were used to compute the significance of a given correlation). These analyses were conducted: (1) on the entire plant material, (2) on 13 populations of *V. vitis-idaea* ssp. *vitis-idaea* and (3) on six Swedish populations.

Two approaches were used to evaluate amount of diversity within populations: Shannon's diversity index (Bussell 1999) and Lynch and Milligan's index (Lynch and Milligan 1994). Lynch and Milligan's index requires the assumption of a Hardy–Weinberg equilibrium for each locus, whereas Shannon's index does not require this assumption. Based on the frequency of a given RAPD band, p_i, Shannon's index was calculated for each RAPD marker, i, separately for each population, $H'_{pop(i)}$. The average Shannon's index for a population, H'_{pop}, was then calculated by averaging $H'_{pop(i)}$ over all markers.

Unbiased null-allele frequencies were calculated and then used to estimate the Lynch and Milligan gene diversity within populations. This index was calculated for each RAPD marker, i, separately for each population, $H_{pop(i)}$, over all markers. The mean Lynch and Milligan gene diversity index for a population, H_{pop}, was then calculated by averaging $H_{pop(i)}$ over all markers. Pearson product-moment correlation analyses were performed between estimates of gene diversity obtained through the two different methods of data analysis (Shannon's and Lynch and Milligan's indexes).

Two different methods, Shannon's index and AMOVA (WINAMOVA version 1.55; Excoffier et al. 1992) were used to partition the variability within our samples into different hierarchical levels, that is, within- and between-populations. The components of variability were calculated for three data sets, comprising (1) all populations, (2) populations of *V. vitis-idaea* ssp. *vitis-idaea* and (3) the Swedish populations only.

Shannon's index was calculated for each marker, i, separately for each population, $H'_{pop(i)}$, and for the species, $H'_{sp(i)}$. The average Shannon's diversity index per locus for populations within the species, $H'_{sp/pop}$, was calculated by averaging $H'_{pop(i)}$ over all populations. Diversity among populations, $G'_{pop/sp(i)}$, was calculated for each locus, $G'_{pop/sp(i)} = [H'_{sp(i)} - H'_{sp/pop(i)}]/H'_{sp(i)}$. Mean $G'_{pop/sp}$ (G'–statistics) was then calculated by averaging $G'_{pop/sp(i)}$ over all markers.

AMOVA was used to compute Φ-statistics (Excoffier et al. 1992). A matrix of squared Euclidean distances was produced and used as input matrix in AMOVA.

Selection of samples for ex situ *conservation*

With the goal of establishing a lingonberry gene bank with high genetic diversity, two different strategies were compared for selecting plants for preservation: (1) a hierarchical sampling strategy (Parsons et al. 1999) and (2) the Maximum genetic diversity program (Marita et al. 2000). In addition, random sampling was used for comparisons with the selected data sets. Each time, 61 plants were chosen among the 211 accessions. In addition, eight plants representing different cultivars were included in each data set.

Table 2. Selected primers used for RAPD analysis.

Primer	Sequence (5′ to 3′)	Number of scored bands	Number of polymorphic bands	Polymorphic bands (%)	RAPD primer index
OPA-10	GTGATCGCAG	12	11	91.7	3.64
OPA-11	CAATCGCCGT	12	12	100.0	3.30
OPB-19	ACCCCCGAAG	11	9	81.8	2.33
OPC-15	GACGGATCAG	9	9	100.0	2.97
OPD-05	TGAGCGCACA	9	8	88.9	2.41
OPE-10	CACCAGGTGA	4	4	100.0	1.18
OPF-07	CCGATATCCC	9	6	66.7	1.62

For the hierarchical sampling strategy, Jaccard's coefficient of similarity was calculated for all pairwise comparisons between individual samples (Maguire and Sedgley 1997). A UPGMA cluster analysis was performed, and a dendrogram was produced for the 211 plants. By truncating the hierarchy at a similarity level of 0.60, we obtained 61 groups. One randomly chosen individual from each group was included in our genebank collection.

For the second strategy, genetic distances among individual samples were calculated using the complement to the simple matching coefficient (Marita 1998; Marita et al. 2000). The distances were used as an input matrix for the Maximum genetic diversity program (Marita et al. 2000).

To compare the different methods, mean Nei's gene diversity and Shannon's diversity index were calculated for the initial plant set (219 plants = 211 + 8) and for the data sets obtained with the two different strategies as well as with random sampling (61 + 8 plants). In addition, the frequency of each RAPD band was calculated for each data set to examine whether the various strategies had resulted in any major changes of band frequencies as evidenced from ocular inspection of band frequency histograms.

A cluster analysis (UPGMA) was performed on the most diverse data set, and a dendrogram was produced, which illustrated the relationships among chosen plants.

Results

RAPD amplification

Seven selected primers produced 66 clear and reproducible bands; 59 (89.4%) of these were polymorphic. Individual primers produced from 4 to 12 bands, whereas the proportion of polymorphic bands varied from 66.7% to 100%. RAPD primer index ranged from 1.18 (OPE-10) to 3.64 (OPA-10) (Table 2). Only polymorphic bands were included in the statistical analyses.

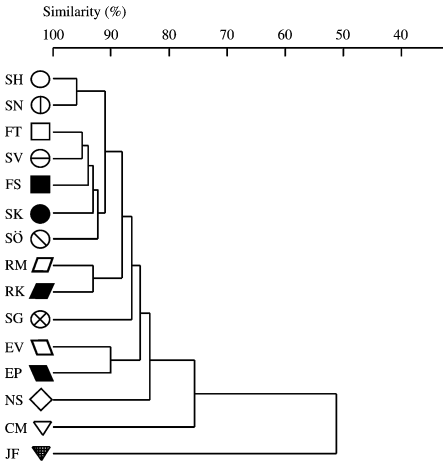


Figure 1. Dendrogram based on UPGMA analysis of genetic similarity estimates among 15 populations of lingonberry.

Genetic relatedness among populations

A cluster analysis was performed to examine the relatedness among all 15 populations, and the resulting dendrogram clearly demonstrated an association between genetic relatedness and geographic distances (Figure 1). The Japanese population was the most divergent, followed by the Canadian and

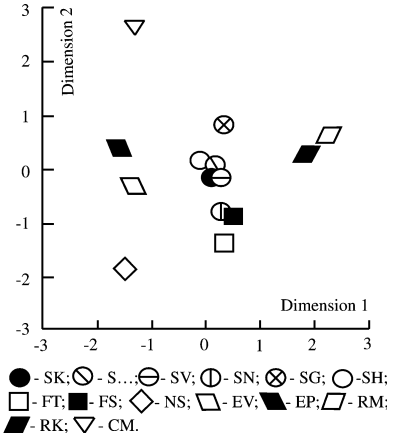


Figure 2. Two-dimensional plot of MDS analysis, based on a genetic distance matrix, illustrating genetic relationships among 14 populations of lingonberry. The Japanese population of *Vaccinium vitis-idaea* ssp. *minus*, which is an outlier, is not included.

Norwegian populations. Among the remaining populations, the two Estonian populations formed one cluster, and the Swedish, Finnish and Russian formed another. Within the last cluster, the two Russian populations clustered together, whereas the Swedish and Finnish populations were completely intermingled.

An MDS analysis with three dimensions was applied for all 15 population (results not shown). The Japanese population was a distant outlier and therefore another MDS analysis was conducted for the remaining 14 populations. The proportion of variance explained by the first three axes (RSQ) was 0.952 with Kruskal's stress = 0.096, indicating a good fit between our data and the three-dimensional representation of variation among populations (Bartish et al. 1999b). The MDS analysis supported the genetic relationships revealed by the cluster analysis. A plot of dimension 1 vs. dimension 2 yielded a main group consisting of six Swedish and two Finnish populations (Figure 2). These were placed in the centre with almost identical values on the first dimension. The two Russian populations formed one pair and the two Estonian another. The Norwegian population was relatively isolated from the other populations.

Table 3. Diversity among populations estimated as Φ-statistics by analysis of molecular variance (AMOVA) and G'-statistics derived from Shannon's index, both with standard error.

Source of diversity	Φ-statistics	G'- statistics
All populations	0.314 (0.002)	0.398 (0.024)
Populations of *V. vitis-idaea* spp. *vitis-idaea*	0.213 (0.002)	0.298 (0.017)
Swedish populations	0.094 (0.002)	0.188 (0.017)

The most distant population was the one from Canada.

We found a significant correlation between geographic and genetic distances when the whole plant material was analysed ($r = 0.532$, $P < 0.001$). The Japanese and Canadian populations belong to ssp. *minus*, and therefore another correlation analysis was conducted on the 13 populations of ssp. *vitis-idaea*. As in the previous case, a significant correlation between genetic and geographic distances was observed ($r = 0.475$, $P < 0.001$). By contrast, there was no significant correlation in the analysis of the six Swedish populations ($r = 0.182$, $P = 0.320$).

Gene diversity

Evaluation of gene diversity using Lynch and Milligan's index (H_{pop}) and Shannon's index (H'_{pop}) revealed the Canadian population to be the most diverse ($H_{pop} = 0.274$; $H'_{pop} = 0.654$). The Russian populations RM and RK were comparatively homogeneous and showed the lowest diversity values with both evaluation methods (Table 1). When calculated over all populations, a high correlation was found between Lynch and Milligan's index and Shannon's index ($r = 0.956$, $P < 0.001$).

AMOVA analysis attributed 31.4% of the variability to the among-population component when all 15 populations were included (Table 3). When an analysis was performed on only the 13 populations of ssp. *vitis-idaea*, 21.3% of the variability was allocated among populations. When calculated for the Swedish populations, only 9.4% of the molecular variation was attributed to the among-population component. Diversity between populations was also estimated with Shannon's index,

which resulted in higher values but showed the same overall pattern (Table 3).

Selection of samples for ex situ *conservation*

A potential genebank collection of 61 samples, that is, 31.5% of the 211 plants analysed, was selected using two different strategies. For comparison with these strategies, random sampling was performed three times to create three different data sets.

For the first selection strategy, hierarchical sampling, a cluster analysis was performed on 211 plants (data not shown). Plants from all six Swedish and the two Finnish populations were distributed rather evenly over the entire dendrogram. Individuals from the other populations formed clusters mainly in accordance with their geographic provenance. The Japanese population formed a close-knit cluster, which was an outlier in comparison with the remaining samples. Hierarchy was truncated at a similarity level of 0.60, selected to result in 61 groups. A single randomly chosen accession from each group was then included in our potential genebank collection.

Application of the second selection strategy, the Maximum diversity program, requires the user to select the first sample and to determine the total number of samples to be included in the core collection. We chose to keep the number constant at 62 accessions, but we varied the first sample. The program was run 11 times. Samples that were selected at least seven times (59 in total) were included in our genebank collection. No plants from populations RM and SN were included in any of the program-generated collections. However, we chose plants RM7 and SN198 as initial samples in two of the runs and therefore decided to include these plants in our final data set. Eight plants representing cultivars were added to each of the data sets, thus increasing the number of individuals to 69 for our potential genebank collection.

Nei's gene diversity and Shannon's index were used to estimate the extent of diversity within the entire plant set and within the different subsets chosen for *ex situ* conservation (Table 4). The subset obtained with the Maximum diversity program had the highest diversity values (Nei's gene diversity = 0.356; Shannon's index = 0.759). Nei's gene diversity value for the hierarchical sampling strategy-based collection (0.322) was higher than the values for the entire material (0.300) and for the randomly chosen collections (0.301–0.304) whereas Shannon's index values were proportionally somewhat more similar but showed the same overall pattern: 0.670 for the entire material, 0.663–0.669 for the randomly chosen collections and 0.698 for the hierarchical sampling strategy-based collection.

Table 4. Genetic diversity in the entire plant material (219 accessions) and subsamples (69 accessions) obtained by different strategies.

	Nei's gene diversity	Shannon's index
Entire material	0.300	0.670
Randomly chosen collection (3 data sets)	0.301–0.304	0.663–0.669
Hierarchical sampling strategy-based collection	0.322	0.698
Maximum diversity-based collection	0.356	0.759

To investigate how well we preserved the RAPD bands revealed in the entire collection, we compared the frequencies of bands in the entire collection and in each of the different subsets. One band was lost in one of the three randomly chosen subsets and in the hierarchical strategy-based subset. In addition, there was no clear pattern in band frequency changes when the hierarchical strategy-based subset and the randomly chosen subsets were compared with the entire material. In the Maximum diversity program-based subset, no bands were lost. Moreover, those bands which had frequencies above 0.5 in the entire collection, showed decreased frequencies in the subset, while bands with frequencies below 0.5 in the entire material increased their frequencies in the subset (Figure 3).

A cluster analysis was performed and a dendrogram was produced illustrating the relationships among the 69 plants included in the most diverse subset, which was obtained using the Maximum diversity program (Figure 4). The selected individuals clustered in a way that closely resembled clustering among samples in the entire collection. Again, plants from the Japanese, Canadian, Estonian and Norwegian populations clustered mainly in accordance with their provenance, whereas plants from Swedish and Finnish populations were scattered over the dendrogram. Pairwise similarity values in the subset ranged from 0.23 to

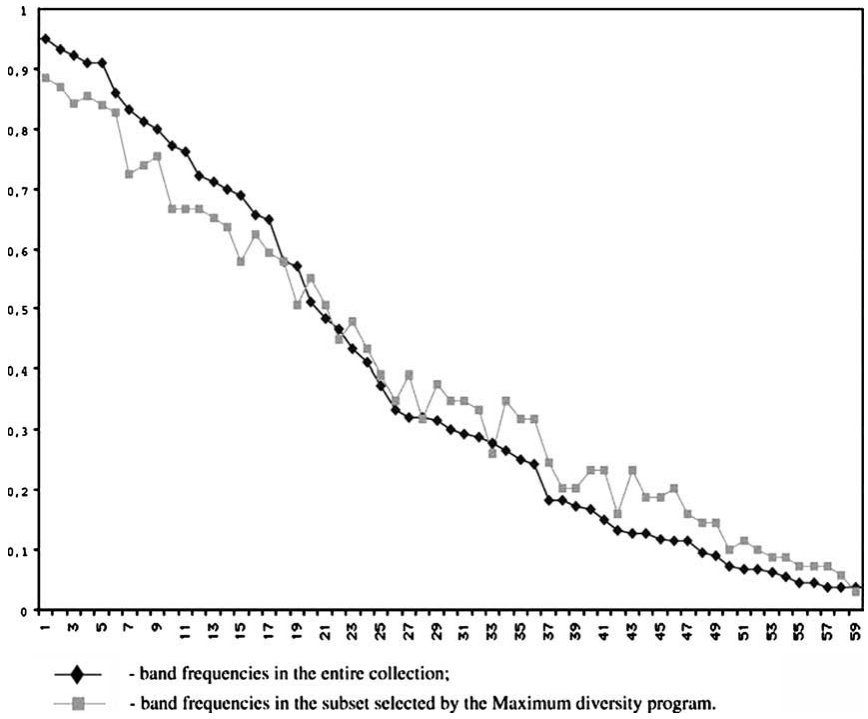


Figure 3. RAPD band frequencies in the entire collection and in the subset obtained by using the Maximum genetic diversity program.

0.81, as compared to 0.22 to 1.00 in the entire collection, where two plants from the population RK were identical.

Discussion

Lingonberry has a relatively brief history of domestication and cultivation, but for millennia people have picked berries in the forests. In present day in Scandinavia, large quantities of lingonberries are still gathered annually from native stands in coniferous forests. Most of these berries are used for processing, and the species is of considerable economical importance, especially in Finland, Sweden and Russia. However, recent changes in forest management, variable fruit quality from native stands, and fluctuations in annual yield have stimulated initiatives to cultivate the species in several North European countries (Gustavsson 1999). Lingonberry orchards have been established recently also in the United States of America. So far, plant breeding efforts have been modest, and most of the 17 hitherto recognized cultivars were derived from selections in the wild (Gustavsson 1999).

Proper evaluation of genetic resources in native material is especially important at the initial stages of domestication. Apart from the identification of material with desirable traits, there is also a need to estimate the amount of genetic diversity and its partitioning within and among populations. Recent results obtained using DNA markers have been reviewed for temperate fruit and berry crops in the Northern hemisphere (Antonius-Klemola 1999; Hokanson 2001; Nybom et al. 2003). In lingonberry, RAPD analysis was applied to analyse

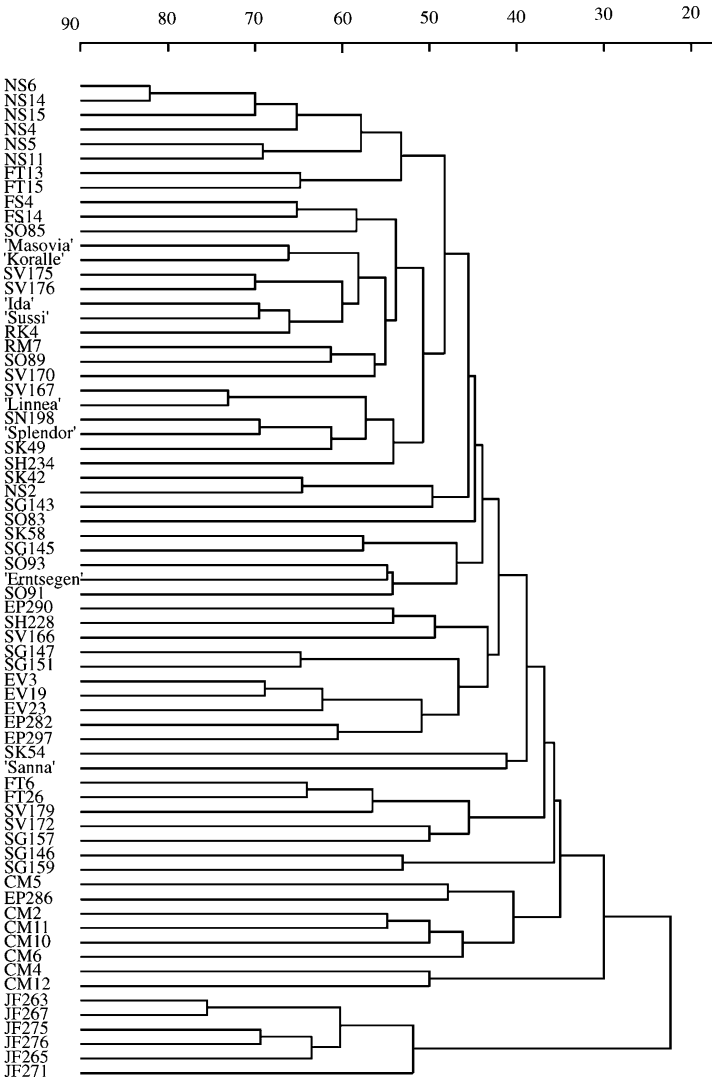


Figure 4. UPGMA based dendrogram showing the relationships among 69 individuals in the subset obtained by using the Maximum genetic diversity program.

genetic and genotypic diversity in material collected from native Swedish lingonberry populations (Persson and Gustavsson 2001). The use of 43 polymorphic RAPD markers was sufficient for the identification of 29 different genotypes (= putative clones) among the 129 plants analysed. These

clones stretched linearly up to 30 m. In studies of clonal species, wild populations are often described in terms of their frequency of unique genotypes (D) and the evenness (E) among these genotypes (Ellstrand and Roose 1987). In lingonberry, both D (mean 0.84) and E (mean 0.81) were higher than the average for clonal plant species (Persson and Gustavsson 2001). In addition to RAPD analysis, an automated image analysis of leaf shape was used to assess the amount of phenotypic variation and partition it among and within populations. Both approaches revealed that most of the variation could be attributed to within-populational variation (Persson and Gustavsson 2001).

Genetic relatedness among populations

In our study, both the dendrogram and the MDS indicated close relationships between the Swedish and Finnish populations. By contrast, the Norwegian population, which also grows in the Scandinavian peninsula, appeared to be relatively genetically distant from the Finnish and the Swedish populations. This population was collected in the western part of Norway, where it was separated from the Swedish populations by high mountains, which may have acted as a barrier for gene migration. The Japanese and Canadian populations, which belong to ssp. *minus* were, as expected, genetically distant from the other populations, which all represent ssp. *vitis-idaea*. The same Japanese population was also included in a study on genetic variation in horticulturally important traits where it differed in many traits from all the other populations (Gustavsson 2001).

Genetic structuring of plant populations is strongly influenced by both common ancestry and current patterns of inter- and intrapopulational genetic exchange (Schaal et al. 1998). We found a significant correlation between RAPD-based estimates of genetic distances and geographical distances except when the analysis was limited to only the six Swedish populations. Similar evidence of isolation-by-distance has been reported for many other outcrossing plant species (Gabrielsen et al. 1997; Graham et al. 1997; Le Corre et al. 1997; Tollefsrud et al. 1998) but is usually lacking in inbreeding species (Fahima et al. 1999) or when the analysed samples derive from closely growing populations (Fischer and Matthies 1998; Gustafson et al. 1999; Stevens et al. 1999).

Gene diversity

RAPD-based estimations of diversity within and among wild-growing populations are generally associated closely with life-history traits according to a compilation of 108 studies (Nybom and Bartish 2000). Nei's gene diversity (in some cases modified according to Lynch and Milligan 1994) denoted as H_{pop}, was used for comparisons of levels of within-population diversity in that compilation. Breeding system had a profound effect, with mean $H_{pop} = 0.12$ for selfing species, 0.22 for species with a mixed breeding system (selfing as well as outcrossing), and 0.26 for outcrossing species. The mean value for within-population diversity (H_{pop}) was 0.206 in our study, which is in agreement with a mixed breeding system, as previously reported for lingonberry (Jacquemart and Thompson 1996). In the compilation (Nybom and Bartish 2000), mean values for H_{pop} were 0.24 for long-lived perennials, 0.23 for species dispersed by animal-ingested seeds, and 0.20 for species characterized by a mid-successional status in vegetation development, all of which are traits that seem to be valid for lingonberry (Gustavsson 1999). As suggested also in Persson and Gustavsson (2001), sexual reproduction appears to have played a significant role in the establishment and growth of lingonberry populations despite low levels of seedling recruitment being observed in present-day populations (Eriksson 1989).

Shannon's index produced values that were about twice as high as the values obtained with Lynch and Milligan's index for within-population diversity. The mean Shannon's index was 0.431 in our study. For comparison, Shannon's index was 0.568 when calculated for four Swedish populations (Persson and Gustavsson 2001). In both cases, values are quite similar to those reported in non-clonal, prevailingly outcrossing plant species, such as *Chaenomeles japonica* (0.459) and *C. speciosa* (0.552) (Bartish et al. 2000a).

Population differentiation is usually estimated as the AMOVA-derived Φ_{st} or as G_{st}, which yield very similar values when calculated on the same material (Nybom and Bartish 2000). In general, population differentiation is negatively correlated with

within-population diversity (Nybom and Bartish 2000). The partitioning of molecular variance within and among populations is influenced directly by the overall diversity of the plant sample. Between-population variability accounted for only 15% of the total variance within *Hippophae rhamnoides* ssp. *rhamnoides* (Bartish et al. 1999a) but for 59.6% when the analysis was enlarged to the entire genus (Bartish et al. 2000b). In our study, between-population variability for the Swedish populations was 9.4%, which is very similar to the 10.8% obtained in a previous study (Persson and Gustavsson 2001). When our analysis was performed on the entire set of populations, between-population variability accounted for 31.4%. Our Φ-statistic values resemble the means reported by Nybom and Bartish (2000) for outcrossing species (0.28), for long-lived perennials (0.25), for species with ingested fruits (0.29), and for mid-successional species (0.40).

Selection of samples for ex situ *conservation*

Insufficient funding is often a problem in plant breeding programs, and preservation of the most valuable accessions becomes a priority when resources are limited. These accessions can then be used to help conserve the diversity of a crop for future use by plant breeders and researchers. To choose optimal material for a new gene bank, criteria similar to those for the selection of 'core collections' of already established gene banks, can be used. Many studies have developed and compared methods and strategies of assembling core collections in different species (Schoen and Brown 1993; Diwan et al. 1995; Lerceteau et al. 1997; Parsons et al. 1999; Marita et al. 2000; Chandra et al. 2002; Dhanaraj et al. 2002). In our study, we compared potential genebank collections chosen by the use of (1) a hierarchical sampling strategy (Parsons et al. 1999) and (2) the Maximum genetic diversity program (Marita et al. 2000).

Various criteria can be used to determine the extent to which degree of diversity has been maintained or even increased in a chosen subset out of a larger, pre-existing plant collection. One of these criteria is marker diversity estimated as Nei's gene diversity or Shannon's index. Both of the systematically selected subsets in our study were considerably more diverse than randomly generated subsets, with the highest values obtained for both Nei's gene diversity and Shannon's index when using the Maximum genetic diversity program.

A second criterion is the number of bands retained in the chosen subset. All revealed RAPD bands were preserved in the subset based on the Maximum genetic diversity program in our study. Similarly, only 1.4 AFLP bands were lost in a Maximum diversity program-generated subset representing 31.9% of the initial plant collection in sweetpotato *Ipomoea batatas* (Fajardo et al. 2002).

A third, and possibly the most important criterion, is the extent to which the relative frequencies of rare bands have been retained or even increased in the chosen subset. Although rare genes are sometimes considered to be peripheral in conserving endangered species, they can be valuable in plant breeding (Li et al. 2002). When acquiring crop accessions for *ex situ* conservation, it is important to preserve genes that govern traits of potential value in commercial production. If the plant material has not been sufficiently well characterized for horticulturally important traits, as is the case with lingonberry, we may avoid losing such genes by preserving at least one specimen from each population and by increasing the number of rare molecular markers in the hope that these will occur in individuals that are also more likely to carry rare genes. By using the Maximum diversity program, we were able to select a subset that preserved all rare RAPD bands, and their frequencies were increased over those of the initial plants sampled.

Cultivated and wild accessions clustered separately in a study by Bradeen et al. (2002), where AFLP and ISSR markers and cluster analysis were used to study the genetic diversity in cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations. As expected in our more recently domesticated species, cultivated accessions were instead intermixed with wild accessions (Figure 4). Two cultivars, 'Ida' and 'Sussi', which derive from wild stands in the province of Småland in Sweden, clustered together.

In conclusion, the use of RAPD and the Maximum diversity program was successful in choosing a diverse and representative subset of plants that can be used for establishing a new genebank for lingonberry.

Acknowledgements

We thank Dr Jane M. Marita, US Dairy Forage Research Center, USDA–ARS, USA, who kindly supplied the Maximum genetic diversity program. We also thank Prof. Susan Brown, Department of Horticultural Sciences, Cornell University, USA and Dr Philip L. Forsline, Plant Genetic Resources Unit, Cornell University, USA, for critical reading and valuable comments on the manuscript. Financial support was received from FORMAS (The Swedish Council for Environment, Agricultural Sciences and Spatial Planning).

References

Antonius-Klemola K. 1999. Molecular markers in *Rubus* (Rosaceae) research and breeding. J. Hort. Sci. Biotechnol. 74: 149–160.

Axelsson A. 2000. Random amplified polymorphic DNA base-rad variation inom och mellan populationer av lingon (*Vaccinium vitis-idaea* L.). Undergraduate thesis, Department of Biomedical Laboratory Technology, Malmö University, Malmö, Sweden.

Bartish I.V., Jeppsson N. and Nybom H. 1999a. Population genetic structure in the dioecious pioneer plant species *Hippophae rhamnoides* investigated by RAPD markers. Mol. Ecol. 8: 791–802.

Bartish I.V., Rumpunen K. and Nybom H. 1999b. Genetic diversity in *Chaenomeles* (Rosaceae) revealed by RAPD analysis. Plant Syst. Evol. 214: 131–145.

Bartish I.V., Garkava L.P., Rumpunen K. and Nybom H. 2000a. Phylogenetic relationships and differentiation among and within populations of *Chaenomeles* Lindl. (Rosaceae) estimated with RAPDs and isozymes. Theor. Appl. Genet. 101: 554–563.

Bartish I.V., Jeppsson N., Bartish G.I., Lu R. and Nybom H. 2000b. Inter- and intraspecific genetic variation in *Hippophae* (Elaeagnaceae) investigated by RAPD markers. Plant Syst. Evol. 225: 85–101.

Bradeen J.M., Bach I.C., Briard M., Le Clerc V., Grzebelus D., Senalik D.A. and Simon P.W. 2002. Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. J. Amer. Soc. Hort. Sci. 127: 383–391.

Bussell J.D. 1999. The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobeliaceae). Mol. Ecol. 8: 775–789.

Chandra S., Huaman Z., Hari Krishna S. and Ortiz R. 2002. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data – a simulation study. Theor. Appl. Genet. 104: 1325–1334.

Dhanaraj A.L., Rao E.V.V.B., Swamy K.R.M., Bhat M.G., Prasad D.T. and Sondur S.N. 2002. Using RAPDs to assess the diversity in Indian cashew (*Anacardium occidentale* L.) germplasm. J. Hort. Sci. Biotechnol. 77: 41–47.

Diwan N., McIntosh M.S. and Bauchan G.R. 1995. Methods of developing a core collection of annual *Medicago* species. Theor. Appl. Genet. 90: 755–761.

Ellstrand N.C. and Roose M.L. 1987. Patterns of genotypic diversity in clonal plant species. Am. J. Bot. 74: 123–131.

Eriksson O. 1989. Seedling dynamics and life histories in clonal plants. Oikos 55: 231–238.

Excoffier L., Smouse P.E. and Quattro J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.

Fahima T., Sun G.L., Beharav A., Krugman T., Beiles A. and Nevo E. 1999. RAPD polymorphism in wild emmer wheat populations, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 98: 434–447.

Fajardo D.S., La Bonte D.R. and Jarret R.L. 2002. Identifying and selecting for genetic diversity in Papua New Guinea sweetpotato *Ipomoea batatas* (L.) Lam. germplasm collected as botanical seed. Genet. Resour. Crop Evol. 49: 463–470.

Fernald M.L. 1950. Gray's Manual of Botany. 8th edn. Dioscorides Press, Portland, OR, 1632 pp.

Fischer M. and Matthies D. 1998. RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae). Amer. J. Bot. 85: 811–819.

Frankel O.H. 1984. Genetic perspectives of germplasm conservation. In: Arber W., Llimensee K., Pecock W.J. and Starlinger P. (eds), Genetic Manipulation: Impact on Man and Society, Cambridge University Press, Cambridge, UK, pp. 161–170.

Gabrielsen T.M., Bachmann K., Jakobsen K.S. and Brochmann C. 1997. Glacial survival does not matter: RAPD phylogeography of Nordic *Saxifraga oppositifolia*. Mol. Ecol. 6: 831–842.

Ghislain M., Zhang D., Fajardo D., Huamán Z. and Hijmans R.J. 1999. Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genet. Resour. Crop Evol. 46: 547–555.

Goulao L., Cabrita L., Oliveira C.M. and Leitao J.M. 2001. Comparing RAPD and AFLP (TM) analysis in discrimination and estimation of genetic similarities among apple (*Malus domestica* Borkh.) cultivars. Euphytica 119: 259–270.

Graham J., Squire G.R., Marshall B. and Harrison R.E. 1997. Spatially dependent genetic diversity within and between colonies of wild raspberry *Rubus idaeus* detected using RAPD markers. Mol. Ecol. 6: 1001–1008.

Gustafson D.J., Gibson D.J. and Nickrent D.L. 1999. Random amplified polymorphic DNA variation among remnant big bluestem (*Andropogon gerardii* Vitman) populations from Arkansas' Grand Prairie. Mol. Ecol. 8: 1693–1701.

Gustavsson B.A. 1997. Breeding strategies in lingonberry culture (*Vaccinium vitis-idaea*). Acta Horticult. 446: 129–137.

Gustavsson B.A. 1999. Plant breeding and domestication of lingonberry (*Vaccinium vitis-idaea* L.). Doctoral thesis, Swedish University of Agricultural Sciences, Alnarp.

Gustavsson B.A. 2001. Genetic variation in horticulturally important traits of fifteen wild lingonberry *Vaccinium vitis-idaea* L. populations. Euphytica 120: 173–182.

Hokanson S.C. 2001. SNiPs, chips, BACs, and YACs: Are small fruits part of the party mix? HortSci. 36: 859–871.

Hultén E. 1949. On the races in the Scandinavian flora. Svensk Botanisk Tidskrift 43: 383–406.

Jacquemart A.L. 1993. Floral visitors of *Vaccinium* species in the High Ardennes, Belgium. Flora 188: 263–273.

Jacquemart A.L. and Thompson J.D. 1996. Floral and pollination biology of three sympatric *Vaccinium* (Ericaceae) species in the Upper Ardennes, Belgium. Can. J. Bot. 74: 210–221.

Le Corre V., Dumolin-Lapègue S. and Kremer A. 1997. Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl: the role of history and geography. Mol. Ecol. 6: 519–529.

Lerceteau E., Robert T., Pétiard V. and Crouzillat D. 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. Theor. Appl. Genet. 95: 10–19.

Li Q., Xu Z. and He T. 2002. *Ex situ* genetic conservation of endangered *Vatica guangxiensis* (Dipterocarpaceae) in China. Biol. Cons. 106: 151–156.

Lu J., Knox M.R., Ambrose M.J., Brown J.K.M. and Ellis T.H.N. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Theor. Appl. Genet. 93: 1103–1111.

Lynch M. and Milligan B.G. 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3: 91–99.

Maguire T.L. and Sedgley M. 1997. Genetic diversity in *Banksia* and *Dryandra* (Proteaceae) with emphasis on *Banksia cuneata*, a rare and endangered species. Heredity 79: 394–401.

Marita J.M. 1998. Characterization of *Theobroma cacao* using RAPD-marker based estimates of genetic distance and recommendations for a core collection to maximize genetic diversity. M.S. thesis, University of Wisconsin, Madison, WI, USA.

Marita J.M., Rodriguez J.M. and Nienhuis J. 2000. Development of an algorithm identifying maximally diverse core collections. Genet. Resour. Crop Evol. 47: 515–526.

Norusis M.J. 1990. SPSS Base System User's Guide. SPSS Inc., Chicago.

Nybom H. 2003. Navigating the 'Smörgåsbord' of DNA marker methods for improvement of horticultural crops. Acta Horticult. 625: 15–28.

Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol. Ecol (in-press).

Nybom H. and Bartish I.V. 2000. Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Persp. Plant Ecol. Evol. Syst. 3(2): 93–114.

Nybom H., Bartish I., Garkava-Gustavsson L., Persson H. and Werlemark G. 2003. Evaluating genetic resources in minor fruits. Acta Horticult. 622: 81–94.

Nybom H. and Schaal B.A. 1990. DNA "fingerprints" reveal genotypic distributions in natural populations of blackberries and raspberries (*Rubus*, Rosaceae). Amer. J. Bot. 77: 883–888.

Parsons B.J., Newbury N.J., Jackson M.T. and Ford-Lloyd B.V. 1999. The genetic structure and conservation of aus, aman and boro rices from Bangladesh. Genet. Resour. Crop Evol. 46: 587–598.

Persson H.A. and Gustavsson B.A. 2001. The extent of clonality and genetic diversity in lingonberry (*Vaccinium vitis-idaea* L.) revealed by RAPDs and leaf-shape analysis. Mol. Ecol. 10: 1385–1397.

Schaal B.A., Hayworth D.A., Olsen K.M., Rauscher J.T. and Smith W.A. 1998. Phylogeographic studies in plants: problems and prospects. Mol. Ecol. 7: 465–474.

Schoen D.J. and Brown A.H.D. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. USA 90: 10623–10627.

Stevens M.T., Turner M.G., Tuskan G.A., Romme W.H., Gunter L.E. and Waller D.M. 1999. Genetic variation in postfire aspen seedlings in Yellowstone National Park. Mol. Ecol. 8: 1769–1780.

Tollefsrud M.M., Bachmann K., Jakobsen K.S. and Brochmann C. 1998. Glacial survival does not matter – II: RAPD phylogeography of Nordic *Saxifraga cespitosa*. Mol. Ecol. 7: 1219–1232.

Virk P.S., Zhu J., Newbury H.J., Bryan G.J., Jackson M.T. and Ford-Lloyd B.V. 2000. Effectiveness of different classes of molecular marker for classifying and revealing variation in rice (*Oryza sativa*) germplasm. Euphytica 112: 275–284.

IV

International Journal of Horticultural Science 2007, 13 (3): 1–11
Agroinform Publishing House, Budapest, Printed in Hungary
ISSN 1585-0404

Genetic diversity in a collection of apple (*Malus* × *domestica* Borkh.) cultivars as revealed by RAPD markers

Garkava-Gustavsson, L. & Nybom, H.

Balsgård–Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, Fjälkestadsvägen 459, SE-29194 Kristianstad, Sweden E-mail: larisa.gustavsson@ltj.slu.se

Summary: A collection of 151 apple cultivars was investigated with 7 RAPD primers generating 71 informative bands, to evaluate genetic variability and relatedness. All cultivars presumably derived through genetic recombination were distinguished whereas identical DNA profiles indicated that some cultivars had arisen as sports. A cluster analysis and a PCO did not reveal any distinct geographic patterns, but there was a weak tendency for Swedish and foreign cultivars to differentiate. Many cultivars however clustered together with either one of their parents or with siblings. Overall genetic diversity among the 151 cultivars was estimated with Nei's diversity index (H), 0.269, and with Shannon's index (H'), 0.594. The cultivars were also analysed in six groups, according to time of origination and country of origin, with an average H = 0.262 and H' = 0.546. No major differences in genetic diversity were observed over time or space, although the group with recent, foreign cultivars had the lowest diversity (H = 0.235, H' = 0.493). Comparison between the entire material and a subset with 94 mandate cultivars chosen for preservation in Sweden, showed similar genetic diversity: H_{ENTIRE} = 0.268, H'_{ENTIRE} = 0.593 and $H_{MANDATE}$ = 0.263, $H'_{MANDATE}$ = 0.575. No major differences in band frequencies were observed between these two sets, but 5 RAPD bands were missing in the set with mandate cultivars..

Key words: apple, genetic diversity, germplasm, *Malus domestica*, RAPD

Introduction

Apple, *Malus* x *domestica* Borkh., is one of the economically most important fruit crops in temperate zones. It is also a very diverse fruit crop, with numerous different cultivars all over the world, and the number is increasing rapidly (*Sansavini* et al., 2004). Many apple cultivars have arisen as open-pollinated seedlings, often of unknown origin. The exploitation of naturally occurring mutations (bud-sports) in adapted cultivars has also been important. In the last century, an increasing number of cultivars have, however, been created by crosses performed by plant breeders (*Janick* et al., 1996).

Apple cultivars are maintained by vegetative propagation and are monoclonal, which means that all individuals, belonging to the same cultivar, are genetically identical. By contrast, genetic diversity among cultivars, obtained by sexual recombination, is expected to be rather high because of self-incompatibility which enforces outbreeding and results in heterozygosity (*Kitahara* et al., 2005). Many recently developed apple cultivars have been designed to incorporate genetically determined resistance towards apple scab and some other fungal diseases and insect pests (*Crosby* et al., 1992; *Holb*, 2000). Genes for disease resistance have been obtained from wild relatives to our cultivated apple, thus contributing to the genetic diversity in apple.

From its centre of origin in Central Asia, the apple was introduced into Europe by Romans and for the last 2000 years, the domesticated apple has diversified and flourished worldwide (*Harris* et al., 2002). Apples were brought to Sweden from Central and Southern Europe, and the first apple trees were planted around the 12th century, mainly in monastery orchards. Apple growing with the aim to produce fruits for commercial sale started in the 16th century. In addition to foreign cultivars brought in from e.g. Germany and England, new local cultivars originating as chance seedlings were also grown when found to possess desirable characters like large and tasty fruits. Such seedlings were selected, propagated by grafting and distributed to other growers.

Modern plant breeding, based on controlled crosses, has been undertaken in Sweden first at Alnarp (1920–1960) and then at Balsgård (from around 1950), both nowadays part of the Swedish University for Agricultural Sciences. Both foreign and indigenous cultivars have been used in these plant breeding programs. At Balsgård, there is presently about 1000 different apple cultivars in a germplasm collection, which includes old and new Swedish varieties as well as foreign varieties which are adapted to the Swedish climate and/or contain genes of special interest for plant breeding.

At present, publicly funded conservation of clonally propagated plant genetic resources in Sweden is managed by a governmentally appointed unit, the 'National Program for Diversity of Cultivated Plants', which has defined a set of mandate cultivars. Mandate cultivars are indigenous varieties which have been named, bred, propagated and marketed in Sweden. Some foreign cultivars with a long history of being grown in Sweden are also included (*Hjalmarsson* & *Wallace*, 2004). In apple, 220 mandate cultivars have been appointed. These cultivars are presently conserved mainly in smaller

clone archives all over the country, usually at outdoor museums or other public places. About 100 of these cultivars are also present in the Balsgård collection.

Extensive collections of clonally propagated crops are difficult and expensive to maintain. Accurate and permanent genetic identification of individual genotypes is therefore of outmost importance. All unnecessary duplicates, synonyms and mis-labelled genotypes can then be identified and removed. Proper characterisation also ensures that genotypes are true-to-type, and enables users to refer character screenings to unambiguously identifiable genotypes. Furthermore, estimates of genetic relatedness among genotypes may be useful for character screening: instead of screening all available accessions, only those genotypes which appear to be the most promising according to relatedness information, can be targeted.

Historically, so called pomological (morphological) characters have been used for identification of apple cultivars (*Nilsson*, 1986), but most of these characters are heavily influenced by the environment. During the past few decades, molecular markers have therefore became increasingly popular in the characterization of apple collections, e.g., isozymes (*Weeden & Lamb*, 1985), RFLP (*Nybom & Shaal*, 1990), RAPD (*Koller* et al., 1993), AFLP (*Xu & Korban*, 2000), ISSR (*Goulao* & Oliveira, 2001) and SSR (*Gianfranceschi* et al., 1998, *Liebhard* et al., 2002). The most user-friendly of these methods in terms of need for technical equipment, skills and funding is RAPD, which has been used for identification of apple cultivars (*Koller* et al., 1993; *Mulcahy* et al., 1993) and rootstocks (*Autio* et al., 1998), to study genetic diversity in the genus *Malus* (*Dunemann* et al., 1994; *Zhou & Li*, 2000), and for paternity analysis (*Harada* et al., 1993). RAPD has also been used in the early stages of genomic mapping projects (*Conner* et al., 1997). Specific RAPD bands have been used as markers of horticulturally important traits and have sometimes been converted into co-dominant SCAR markers (*Cheng* et al., 1996; *Yang* et al., 1997; *Kim* et al., 2003). Possible problems with reproducibility within the same laboratory can be avoided if the same protocol is applied and followed carefully (*Mulcahy* et al., 1993) and only strong, clear and consistently amplified bands are scored (*Koller* et al., 1993).

In this study, RAPD-markers were applied to some of the cultivars in the Balsgård apple collection to: i) discriminate among cultivars; ii) detect duplicates and mis-labellings; iii) identify genetic relationships among cultivars; iv) evaluate the genetic diversity and possible effects over time (ancient, old and more recent cultivars) and space (Swedish and foreign cultivars).

Materials and methods

Plant material

In total 151 apple cultivars were analysed (*Table 1*). Of these, 94 are mandate cultivars with 68 originating in Sweden. These 151 cultivars were divided into groups based on their historical age: ancient (originated before 1800), old

Table 1 Apple cultivars analysed and their origination, marked with '?' if only putative. Cultivars were divided into groups (I–VI), which are defined in Table 3. Cultivars regarded as 'mandate cultivars' in Sweden are marked with 'M' after the group number

	Cultivar	Origin	Group	Descendence
1	'Alexander'	Russia	II, M	
2	'Alfa 68'	Sweden	V, M	'Boskoop' x 'Filippa'
3	'Algott' (B:0654)	Sweden	V	'Astrakan, Gyllenkrok's' x 'Worcester Pearmain'
4	'Alice'	Sweden	V, M	Seedling of 'Ingrid Marie'
5	'Annero'	Sweden	III, M	
6	'Antonovka Kamenichka'	Ukraina	IV	
7	'Antonovka Pamtorutka'	Russia	IV, M	
8	'Aroma'	Sweden	V, M	'Ingrid Marie' x 'Filippa'
9	'Arvidsäpple'	Sweden	III, M	
10	'Aspa'	Sweden	III, M	
11	'Astrakan, Gyllenkrok's'	Sweden	I, M	
12	'Astrakan, White'	Russia	II, M	
13	'Astrakan, Red'	Sweden	I, M	
14	'Astrakan, Stor Klar'	Sweden	III, M	Progeny of 'Astrakan, White' ?
15	'Birgit Bonnier'	Sweden	V, M	'Cortland' x 'Lord Lambourne'
16	'Blenheim Orange'	England	II	
17	'Boiken'	Germany	II	
18	'Borgherre'	Netherlands? Germany?	II, M	
19	'Borsdorfer'	Germany	II, M	
20	'Boskoop'	Netherlands	IV	
21	'Brunnsäpple, Halland'	Sweden	I, M	
22	'Cellini'	England	IV, M	Seedling of 'Langton's Nonesuch'
23	'Charlamovsky'	Russia	II, M	
24	'Classic Red Delicious'	USA	IV	
25	'Close'	USA	VI	
26	'Cortland'	USA	VI	'Ben Davis' x 'McIntosh'
27	'Cox's Orange Pippin'	England	IV, M	Seedling of 'Ribston'?
28	'Cox's Pomona'	England	IV, M	Seedling of 'Ribston'?
29	'Discovery'	England	VI	'Worcester Pearmain' x 'Beauty of Bath'
30	'Domö Favorit'	Sweden	III, M	
31	'Drakenberg'	Sweden	III, M	
32	'Dronning Louise'	Denmark	IV	
33	'Edsele'	Sweden	V, M	
34	'Elise' (syn. 'Roblos')	Netherlands	VI	'Septer' x 'Cox's Orange Pippin'
35	'Elstar'	Netherlands	VI	'Golden Delicious' x 'Ingrid Marie'
36	'Eva-Lotta'	Sweden	V, M	'Cortland' x 'James Grieve'
37	'Fagerö'	Sweden	III, M	
38	'Farmors Juläpple'	Sweden	V, M	
39	'Fiholms Ribston'	Sweden	V	
40	'Filippa'	Denmark	IV, M	
41	'Flädie'	Sweden	III, M	Seedling of 'Gravensteiner' ?
42	'Fredrik'	Sweden	V	'Aroma' x selection from USA
43	'Frida'	Sweden	V	'Aroma' x selection from USA
44	'Frösåker'	Sweden	III, M	
45	'Fullerö'	Sweden	III, M	
46	'Förlovningsäpple'	Sweden	III, M	
47	'Gelber Richard'	Germany	II	
48	'Golden Delicious'	USA	IV	

	Cultivar	Origin	Group	Descendence
49	'Goldparmain'	England	IV, M	
50	'Granatäpple, Kungsbacka'	Sweden	I, M	
51	'Gravensteiner'	Italy?		
		Denmark?	II, M	
52	'Gravensteiner, Red'	Italy?		
		Denmark?	II, M	
53	'Gravensteiner of Fusa'	Norway	IV	
54	'Grågylling'	Sweden	I, M	
55	'Grågylling from Skokloster'	Sweden	I	
56	'Guldborg'	Denmark	IV	
57	'Göteborgs Flickäpple'	Sweden	III, M	
58	'Hanaskog'	Sweden	III, M	Seedling of 'Oranie' ?
59	'Hannaäpple'	Sweden	V	
60	'Hedenlunda'	Sweden	III, M	
61	'Himmelstalund'	Sweden	III, M	
62	'Holsteiner Cox'	Germany	VI	Seedling of 'Cox's Orange Pippin'
63	'Holländaräpple'	Sweden	III, M	
64	'Hugoäpple'	Sweden	V	
65	'Hausmütterchen'	Germany	II, M	
66	'Höstkalvill, Gul'	Germany	IV, M	
67	'Ingrid Marie'	Denmark	VI	Possibly 'Cox's Orange Pippin' x unknown
68	'Ivö'	USA/Sweden	VI, M	Synonym: 'Monroe seedling', came to Sweden as budwood
69	'James Grieve'	Scotland	IV	Seedling of 'Pott's Seedling'
70	'Jonathan'	USA	IV	Seedling of 'Esop Spizenburg'
71	'Josefiner'	Sweden	III, M	
72	'John-Georg'	Sweden	V	'Golden Delicious' x 'James Grieve'
73	'Julyred'	USA	VI	
74	K:1016	Sweden	V	'Aroma' x selection from USA
75	K:1016, Red	Sweden	V	Possible sport of K:1016
76	K:1160	Sweden	V	'Katja' x 'Priscilla'
77	K:1343	Sweden	V	Seedling of Coop14 (USA)
78	'Kalmar Glasäpple'	Sweden	I, M	
79	'Katja'	Sweden	V, M	'James Grieve' x 'Worcester Pearmain'
80	'Kavlås'	Sweden	III, M	
81	'Kesäter'	Germany	II, M	
82	'Kramforsäpple'	Sweden	V	
83	'Kim'	Sweden	V, M	'Cortland' x 'Ingrid Marie'
84	'Kingston Black'	England	II	
85	'Kinnekulle Kantäpple'	Sweden	V, M	
86	'Landskronaäpple'	Sweden	V, M	
87	'Langton's Nonesuch'	England	IV	
88	'Larsmässeäpple'	Sweden	I	
89	'Laxton's Superb'	England	IV	'Cox's Orange Pippin' x 'Wyken Pipping'
90	'Linda'	Canada	VI	Seedling of 'Langford Beauty'
91	'Linnaeus' Apple'	Sweden	I, M	
92	'Lobo'	Canada	IV	Seedling of 'McIntosh'
93	'Maglemer'	Denmark	II, M	
94	'Mank's Codlin'	England	IV	
95	'McIntosh, Rogers'	Canada	II	Seedling of 'Fames'
96	'Melon'	Germany	II, M	
97	'Melon, Red'	Germany	IV, M	
98	'Melonkalvill'	Sweden	III, M	
99	'Menigasker'	Sweden	III, M	
100	'Mio'	Sweden	V, M	'Worcester Pearmain' x 'Oranie'
101	'Mutsu'	Japan	VI	'Golden Delicious' x 'Indo'
102	'Mälsåker'	Sweden	III, M	
103	'Nanna'	Norway	VI	'Katja' x 'Buckley Giant'
104	'Norrstack'	Sweden	III, M	
105	'Norrviken'	Sweden	III, M	
106	'Oranie'	Sweden	I,M	
107	'Oretorp'	Sweden	V, M	
108	'Pigeon'	Denmark	IV	
109	'Prima'	USA	VI	PRI 14-510 x NJ 123249
110	'Prinsessäpple'	Netherlands	II, M	
111	'Queen Cox'	England	VI	Sport of 'Cox's Orange Pippin'
112	'Reinette de Blenheim'	England	II	
113	'Rescue'	Canada	VI	Seedling of 'Blushed Calville'
114	'Ribston'	England	II	
115	'Ringstad'	Sweden	III, M	
116	'Risäter'	Sweden	III, M	
117	'Rosen Crab'	Russia	IV	
118	'Rödluvan'	Sweden	V, M	'Lobo' x 'Barhatnoe'
119	'Sandbergs Röda'	Sweden	V, M	
120	'Signe Tillisch'	Denmark	IV, M	
121	'Silva'	Sweden	V, M	'Melba' x 'Stenbock'
122	'Siv'	Norway	VI	'Katja' x 'Buckley Giant'
123	'Snövit'	Sweden	V, M	'Stenbock' x 'Pfirsichroter Sommerapfel'
124	'Sparreholm'	Sweden	III, M	
125	'Spässerud'	Sweden	I, M	
126	'Stenkyrke'	Sweden	I, M	
127	'Stäringe Karin'	Sweden	III, M	
128	'Svanetorp'	Germany	IV, M	
129	'Suislepper'	Estonia	IV, M	
130	'Summerred'	Canada	VI	Seedling of 'Summerland'
131	'Sylvia'	Sweden	V, M	'Astrakan, Gyllenkrok's' x 'Worcester Pearmain'
132	'Särsö'	Sweden	V, M	
133	'Sävstaholm'	Sweden	III, M	
134	'Sörmlandsäpple'	Sweden	I	
135	'Titovka'	Russia	IV	
136	'Transparente Blanche'	Russia?	II, M	
137	'Trogsta'	Sweden	III, M	
138	'Vallda'	Sweden	III, M	
139	'Veseäpple'	Sweden	I, M	
140	'Villands Glasäpple'	Sweden	III, M	
141	'Vista Bella'	USA	VI	NJ 77359 x 'Julyred'
142	'Vitgylling'	Netherlands	II, M	
143	'Vittsjö'	Sweden	III, M	
144	'Vrams Järnäpple'	Sweden	I, M	
145	'Värmlands Sötäpple'	Sweden	III, M	
146	'Wealthy, Red'	USA	IV	
147	'Worcester Pearmain'	England	IV	Seedling of 'Devonshire Quarrenden'
148	'Åkerö'	Sweden	I, M	
149	'Åkerö, Gripsholm'	Sweden	V, M	Sport of 'Åkerö'
150	'Ökna Lökäpple'	Sweden	I, M	
151	'Ökna Vita Vintergylling'	Sweden	III, M	

(1800–1900) and new (after 1900) and their geographic origination (Swedish or foreign) according to pomological literature (*Dahl*, 1929; *Nilsson*, 1986; *Svensson* & *Kastman*, 2005). In total, six groups were thus defined: ancient Swedish (17), ancient foreign (21), old Swedish (34), old foreign (29), new Swedish (31) and new foreign (19) (*Table 2*).

Table 2 Number of investigated apple cultivars (within parentheses the original number before duplicates and mislabelled samples had been deleted) used to calculate within-group genetic diversity, measured by Nei's diversity index (H) and Shannon's index (H') (including standard error) in historically and geographically different groups

Group	No. of cultivars	Nei's diversity index, H	Shannon's index, H'
I Ancient Swedish	16 (17)	0.260 (0.022)	0.539 (0.042)
II Ancient foreign	19 (21)	0.276 (0.023)	0.568 (0.044)
III Old Swedish	33 (34)	0.248 (0.023)	0.523 (0.045)
IV Old foreign	28 (29)	0.281 (0.022)	0.585 (0.042)
V New Swedish	29 (31)	0.270 (0.021)	0.567 (0.038)
VI New foreign	18 (19)	0.235 (0.023)	0.493 (0.044)
		x = 0.262	x = 0.546

RAPD analysis

Young leaves were collected in April-May and stored at -80 °C until use. Leaves were ground to a powder with liquid nitrogen in pre-cooled mortars. Approximately 100 mg of the powder was used for isolation of total plant DNA using the Qiagen Dneasy™ Plant Mini Kit and following the Qiagen protocol.

PCR reactions were performed in volumes of 25 µL, containing 20 ng of DNA, 1 x reaction buffer IV (Advanced Biotechnologies), 2.5 mM $MgCL_2$ (Advanced Biotechnologies), 0.5 µM primer (Operon Technologies), 0.2 µM PCR Nucleotide Mix (Roche Diagnostics Corp.) and 1.0 unit *Taq* DNA Polymerase (Advanced Biotechnologies). The steps for PCR amplification were: one cycle of 5 min at 94 °C followed by 40 cycles of 1 min at 94 °C, 1 min at 36 °C and 2 min at 72 °C and finally by one cycle of 7 min at 72 °C. DNA fragments were separated by electrophoresis in 1.8% agarose gel with a Tris-EDTA-Acetic acid buffer (TEA). The gels were stained with ethidium bromide and the amplification products were visualised under UV light and documented with Polaroid photography for further analyses. Molecular Weight Marker VI (Roche Diagnostics Corp.) was used to determine the size of the DNA fragments.

A total of 186 decamer primers (Operon Technologies) were checked for polymorphism, reproducibility and clarity of the obtained patterns on a subset of 4 cultivars. Seven primers were subsequently chosen for further analysis of all the 151 cultivars (*Table 3*). To check the reproducibility between runs, DNA of the same three plants was included in every run. In addition, amplification of two different samples was carried out twice in each PCR run.

Table 3 Selected primers used for RAPD analysis

Primer	Sequence (5′ to 3′)	Number of scored bands	Number of polymorphic bands	Polymorphic bands (%)	Size range	RAPD primer index
OPA-08	GTGACGTAGG	15	15	100.0	180–1300	3.69
OPA-19	CAAACGTCGG	12	12	100.0	240–1200	2.66
OPF-11	TTGGTACCCC	7	4	57.1	240–1500	1.75
OPF-19	CCTCTAGACC	13	13	100.0	453–2170	4.20
OPK-16	GAGCGTCGAA	14	12	85.7	270–1230	2.90
OPM-09	GTCTTGCGGA	8	8	100.0	430–1176	1.97
OPY-17	GACGTGGTGA	9	7	77.8	250-850	1.79

Statistical analyses

Amplification products were scored manually. Each RAPD band was treated as an independent locus with two alleles, presence (1) or absence (0) of a band. Seventyone polymorphic RAPD bands were entered into a binary matrix.

The informativeness of each RAPD primer was evaluated using the Polymorphic Index Content (PIC) (*Ghislain* et al., 1999) calculated as in *Garkava-Gustavsson* et al. (2005).

To assess levels of molecular relatedness, Jaccard's coefficient of similarity was calculated for all pairwise comparisons between cultivars. A distance matrix was then used to perform a cluster analysis based on average linkage between groups (unweighted pair group method algorithm, UPGMA) (SPSS Data Analysis Package 11.0 for Macintosh). A large dendrogram, representing the relatedness among all 151 analysed cultivars, was produced. Another Jaccard similarity matrix was obtained for 143 individual genotypes (duplicates were not included) and used to perform a Principal coordinate analysis (PCO) (NTSYS–pc statistical package, *Rolf*, 1998). A two-dimensional plot was produced.

To evaluate the amount of genetic diversity within groups of cultivars, in the entire plant set and in the set of mandate cultivars, two diversity indices were used: Nei's gene diversity index, H (*Nei*, 1987) and Shannon's diversity index, H' (*Bussell*, 1999). The Nei's gene diversity was calculated as in *Marita* et al. (2000) and Shannon's diversity index as in *Garkava-Gustavsson* et al. (2005). Shannon's index was also used for partitioning of diversity in its within- and betweengroup components. The index was calculated for each locus $G'_{GROUP/ENTIRE(i)} = (H'_{ENTIRE(i)} - H'_{GROUP(i)})/H'_{ENTIRE(i)}$, where $H'_{GROUP(i)}$ is the average Shannon's index per locus, calculated by averaging $H'_{GROUP(i)}$ over all groups. Mean value of $G'_{GROUP/ENTIRE}$ (basically the same as G'-statistics) was then calculated by averaging $G'_{GROUP/ENTIRE(i)}$ over all markers. In addition, frequencies of individual RAPD bands were calculated in both the entire plant material (151 cvs) and in the subset of Swedish (but not necessarily indigenous) mandate cultivars (94 cvs) in order to reveal any overall changes, as in *Garkava-Gustavsson* et al. (2005).

Results

RAPD polymorphism

Out of 186 oligonucleotide primers initially screened with four apple cultivars, 9 primers showed high levels of polymorphism and good

reproducibility. Two of them, OPF-13 and OPG-06, were difficult to score unambiguously because of differences in band intensity and were therefore excluded. Thus, seven primers which detected distinct, clearly resolved and consistently reproducible amplification products were selected for further analyses. These seven primers generated a total of 77 reliable fragments. The band size ranged from 180 bp to 2170 bp. Number of polymorphic bands ranged from 4 to 15, while the proportion of polymorphic bands varied from 57.1% to 100% (*Table 3*). In total, 71 bands were polymorphic in the entire set of cultivars. Five of these were unique, i.e. present in one cultivar but not in any other. In addition, nine bands were rare, here defined as present in less than 5% of all cultivars.

Based on pairwise comparisons with Jaccard's coefficient of similarity, we checked band-by-band DNA-profiles for all cultivars for which the Jaccard value was equal to 1. As expected, the sports 'Melon, red', 'Gravensteiner, red', 'K:1016, red' and 'Åkerö from Gripsholm' had profiles identical with those obtained for their progenitors ('Melon', 'Gravensteiner', K:1016 and 'Åkerö'). We also compared RAPD banding patterns for two trees labelled with the synonymous names 'Blenheim Orange' and 'Reinette de Blenheim' respectively, and these were also identical as expected. The cultivar 'Fagerö' was identical to 'Grågylling' as suspected since these have been described as very similar, with 'Fagerö' possibly being a red sport of the latter (*Nilsson*, 1986). It was somewhat more surprising to find that 'Spässerud', an old Swedish cultivar from the province of Värmland, was identical to the slightly younger cultivar 'Särsö' originating from the province of Östergötland but described as being rather similar to 'Spässerud' (*Nilsson*, 1986). In this case we performed some additional analyses, in which we compared RAPD profiles of 'Spässerud' and 'Särsö' from our collections with profiles of 'Spässerud' from three different locations in Sweden (Mårbacka, Årnäs and Gränna), and with 'Särsö' from Finland. All these samples showed identical RAPD profiles. The cultivars 'Grågylling from Skokloster' and 'Alexander' also had identical DNA profiles, but these cultivars are quite distinct according to pomological literature. Observation of the two trees in the Balsgård collection indicated that both represent true 'Alexander', thus suggesting that the tree previously regarded as'Grågylling from Skokloster' was mislabelled. Based on our results, the now documented sports or duplicates 'Gravensteiner, red', 'Melon, red', 'K:1016, red', 'Åkerö from Gripsholm', 'Reinette de Blenheim', 'Fagerö', 'Särsö' and 'Grågylling from Skokloster' were deleted from the statistical analyses of genetic diversity.

Another case, where we expected to find identical profiles, was the comparison of 'Cox's Orange Pippin' with its sport 'Queen Cox'. One of the bands amplified by primer OPM-09 was, however, found only in 'Cox's Orange Pippin'. Whether this band was amplified from a region that truly differs between the ancestral cultivar and its derivative, or whether the band difference is artefactual is not yet known. We decided, however, to retain both of these cultivars in the further analyses.

Cluster analysis and Principal coordinate analysis (PCO)

A UPGMA dendrogram, illustrating the molecular relatedness in the entire set of 151 cultivars (*Figure 1*) was constructed. No major clusters were observed in the dendrogram, and there was little grouping that could be associated with geographic origination or historical age: Swedish and foreign, old and new cultivars were completely intermingled.

Many cultivars grouped in accordance to their known descendance: either together with one of their parents or together with the other cultivars with common ancestors. 'Cox's Orange Pippin' of course clusters closely with its sport 'Queen Cox' from which it differed by only one DNA band, and also with another offspring, namely 'Holsteiner Cox'. By contrast, its alleged mother 'Ribston', its alleged sibling 'Cox's Pomona' and three other offspring, namely 'Elise', 'Ingrid Marie' and 'Laxton's Superb' occur further apart in the dendrogram.

The analysed material contained three cases of sibling cultivars. First, K:1016 and 'Fredrik' belong to the same cluster as their mother 'Aroma', while a third sibling, 'Frida', occurs somewhat further apart, as also 'Filippa' which is the mother of 'Aroma'. Second, B:0654 and 'Sylvia', which derive from a cross between 'Astrakan, Gyllenkrok's' and 'Worcester Pearmain', cluster together with one another but not with either of the parents. Interestingly, 'Worcester Pearmain' does not cluster closely with any of its other offspring ('Discovery' and 'Katja') either. Finally, the siblings 'Siv' and 'Nanna' occur quite far apart from one another, and from their mother 'Katja', which instead clusters with 'Mio' with which it shares one parent, namely the above-mentioned 'Worcester Pearmain'.

Other cases of clustering between parents and offspring involve 'Boskoop' and its offspring 'Alfa 68', 'Cortland' and its offspring 'Birgit Bonnier' and 'Eva-Lotta', 'Golden Delicious' and its offspring 'Elstar', 'John-Georg' and 'Mutsu', and 'Ingrid Marie' and its offspring 'Alice' and 'Kim'. By contrast, cases where a parent does not cluster closely with its offspring include 'Cortland' and its offspring 'Kim', 'Filippa' and its offspring 'Alfa', 'Ingrid Marie' and its offspring 'Aroma' and 'Elstar', 'James Grieve' and its offspring 'Eva-Lotta', 'John-Georg' and 'Katja', 'Julyred' and its offspring 'Vista Bella', 'Katja' and its offspring K:1160, 'Lobo' and its offspring 'Rödluvan', and 'McIntosh' and its offspring 'Cortland' and 'Lobo'. Interestingly, of all 12 cultivars having both parents included in the analyses, all but the three Balsgård varieties B:0654, 'Katja' and 'Sylvia' (all of them offspring of 'Worcester Pearmain') clustered with one of the parents.

Most of the older, indigenous mandate cultivars have an unknown origin. According to the dendrograms, a closer relationship might be suspected between the following pairs: 'Annero' and 'Sparreholm', 'Astrakan, Stor Klar' and 'Arvidsäpple', 'Aspa' and 'Kalmar Glasäpple', 'Farmors Juläpple' and 'Astrakan, red', 'Åkerö' and 'Vitgylling', 'Hedenlunda' and 'Frösåker', 'Sävstaholm' and 'Kramforsäpple'.

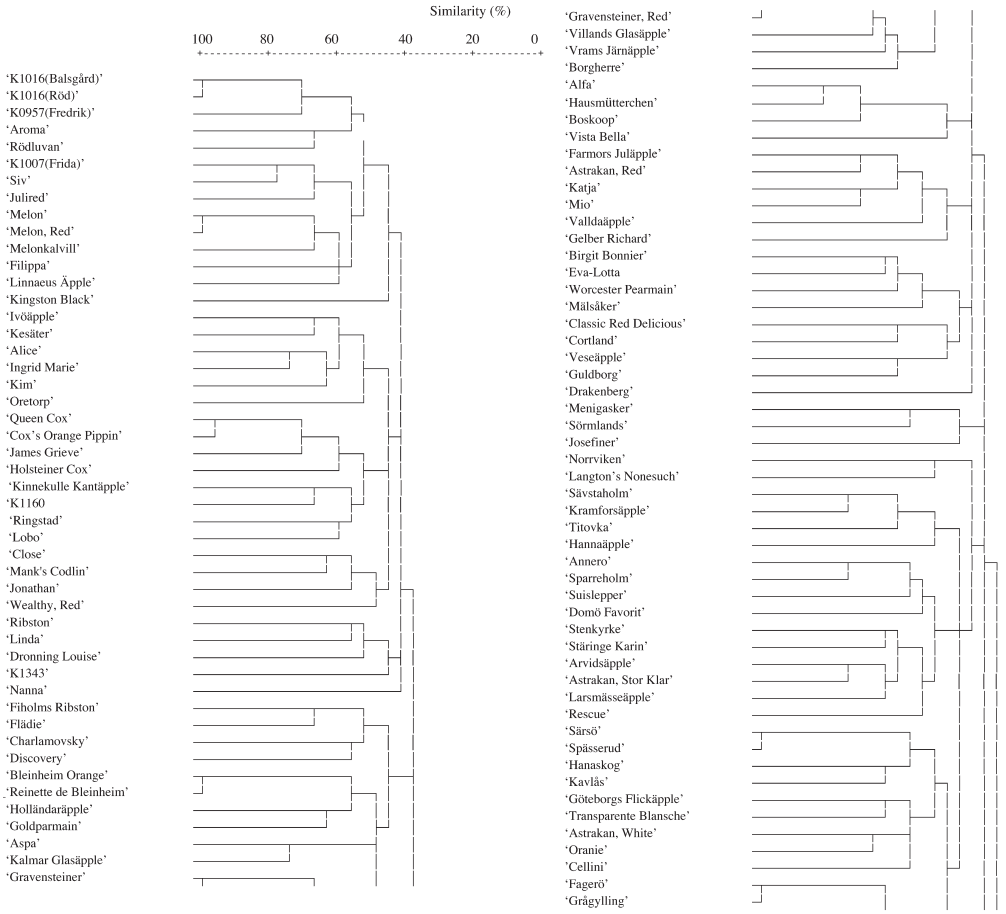


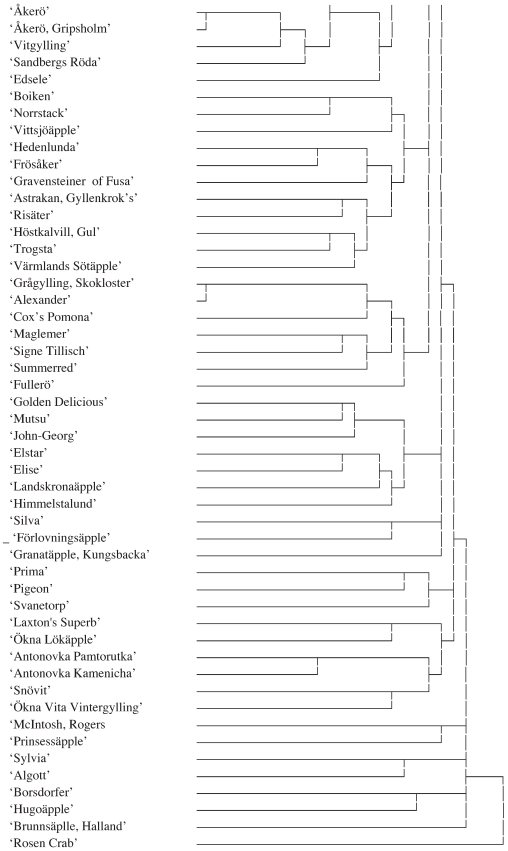
Figure 1. A dendrogram, illustrating molecular relationships among 151 analysed apple cultivars

The PCO analysis, applied to 143 individual genotypes, explained only 11% of the diversity on the first two principal components and confirmed the general pattern of intermingling among cultivars (*Figure* 2). There was, however, a clear tendency for the foreign cultivars (especially the most recent ones) to group in the leftmost half of the plot whereas the Swedish cultivars were more evenly distributed.

Genetic diversity

Genetic diversity value in the entire set of 151 cultivars measured with Nei's diversity index, H_{ENTIRE}, was 0.269. Corresponding value for Shannon's index was $H'_{ENTIRE} = 0.594$. The diversity estimators yielded only slightly lower values when calculated for the subset of 94 mandate cultivars: $H_{MANDATE} = 0.263$ and $H'_{MANDATE} = 0.575$. These diversity estimators were also calculated for each of the 6 groups of cultivars, yielding the mean values of H = 0.262, and H' = 0.548 (Table 3). The highest level of genetic diversity was observed for group 4, old foreign cultivars (H = 0.281; H' = 0.585) and the lowest for group 6, new foreign cultivars (H = 0.235; H' = 0.493). Only 14.6% of the total diversity resided between groups, which indicates a high diversity within groups compared to a rather low degree of differentiation between groups.

Another way to compare diversity in the entire set of cultivars with the subset of mandate cultivars is to analyse RAPD-band frequencies. Plotting these band frequencies, from the most common band to the least common band in both data sets, shows that there were very few discrepancies (*Figure 3*). Bands, that were common in the entire set, remained common also in the subset of mandate cultivars, and the rare bands did in most cases neither increase nor



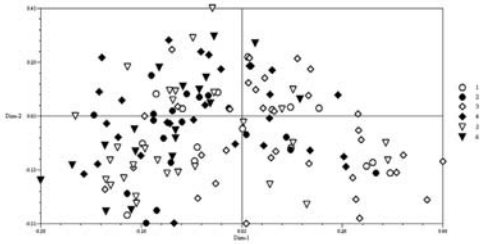


Figure 2. A two-dimensional plot of PCO analysis of 143 individual apple genotypes belonging to different historical and geographical groups (all duplicates are removed).

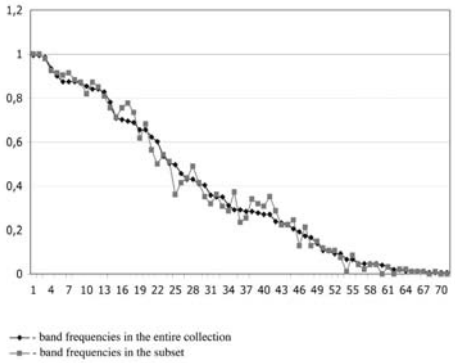


Figure 3 Comparison of RAPD band frequencies in the entire set of apple cultivars (151 cvs) and in the subset of Swedish mandate cultivars (94 cvs).

decrease in frequency. Five RAPD bands were, however, absent in the subset of mandate cultivars, showing that some gene regions are lacking in this subset.

Discussion

The purpose of our study was to obtain a quick preliminary evaluation of genetic diversity in the Balsgård germplasm collection and to find possible duplicates, synonyms and mislabellings, and therefore we chose to use RAPD. To ensure good reproducibility, great care was taken in primer selection: a large number of primers were screened and only those providing consistently amplified bands were chosen. In a different study, the same set of apple cultivars have been analysed with SSR markers (*Garkava-Gustavsson* et al., unpublished). Compared to RAPD, SSR-based analysis lends itself better to the setting up of shared marker score sheets but the high mutation rates can make these markers less useful for relatedness studies (*Weising* et al., 2005).

Identification and relatedness among cultivars

Unambiguous identification of research material becomes especially important when costly screenings are made of e.g. content of phenolic compounds and allergenic proteins (*Nybom* et al., in press) and disease resistance (*Mattisson* & *Nybom*, 2005). For plant breeders and other users, it is similarly important to have access to correctly identified plant material, which does, in fact, contain the genes and traits expected from previous investigations and analyses. RAPD markers have previously proven to be useful for identification of putative duplicates and misclassifications in the collection of e.g. yam cultivars (*Dansi* et al., 2000).

As expected, we could easily distinguish all the cultivars except for most of the somatic mutations (sports) like 'Gravensteiner, red, 'Melon, red', 'K:1016, red, 'Åkerö from Gripsholm' and 'Fagerö'. These sports were omitted from further diversity analyses. We also obtained identical RAPD

profiles for the Swedish cultivars 'Spässerud' (first described in 1903) and 'Särsö' (first described in 1917). Apparently, these cultivars are therefore either completely identical, or – perhaps more likely – one of them has been derived as a sport of the other since fruit flesh colour is reported to be somewhat different (*Nilsson*, 1986). We found one obvious case of mislabelling; the tree labelled 'Grågylling from Skokloster' was instead 'Alexander'. A similar case was revealed in a small pilot study (*Garkava-Gustavsson & Nybom*, 2003), in which 'Astrakan, white' appeared to be identical to 'Astrakan, stor klar'. The RAPD-profile of our 'Astrakan, white' was compared to a profile derived from analysis of a tree in the collection at Pikkiö in Finland, and appeared to be different. We concluded that our 'Astrakan, white' was mislabelled, and in this study it was therefore substituted with 'Astrakan, white' from Finland. Interestingly we did, however, find a one-band difference between our samples of 'Cox's Orange Pippin' and its alleged sport 'Queen Cox'. Since sexual recombination usually results in much larger band profile differences, it appears that our material of 'Queen Cox' does, indeed, represent a sport. Whether the band difference is artefactual or actually reflects a genetic difference in a mutated region must be further investigated.

RAPD markers have usually failed to differentiate among sports also in previous studies of apple identification (*Harada* et al., 1993; *Goulao* et al., 2001). Nevertheless, RAPD markers were efficient in differentiating between the original cultivar and its new radiomutants in chrysantemum (*Lema-Ruminska* et al., 2004). Experimentally applied mutagenics are, however, likely to cause considerably more changes in the genome than in the case of spontaneous sport mutations.

By contrast, RAPD as well as many other DNA-based markers have proven very useful in determining whether different cultivars have arisen by recombination or as sports. Using SSR, the nature of genetic identity was determined for seven pairs of apple cultivars in the core subset collection at USDA-ARS (*Hokansson* et al., 1998). The authors concluded that five pairs contained sport mutations and/or their progenitors, one accession was mislabelled and another one had probably a synonymous name or was a sport mutation.

The clustering of cultivars in our dendrogram, and in our two-dimensional PCO plot, showed very weak associations with geographic origin or the historical age of the investigated cultivars. Similarly, an SSR-based dendrogram, produced for the subset of 68 indigenous Swedish mandate cultivars, failed to produce well defined clusters (*Garkava-Gustavsson* et al., unpublished). A corresponding lack of association between geographical origin and clustering in a dendrogram has also been reported for e.g., mulberry (*Vijayan* et al., 2005), pear (*Monte-Corvo* et al., 2000) and olive (*Grati-Kamoun* et al., 2006). The fact that the Swedish apple cultivars were scattered all over the dendrogram, as well as the PCO plot, indicates that they represent a more or less random set of molecular phenotypes and derive from a broad gene pool, obtained through hybridisation between highly heterozygous apple genotypes during several centuries of apple cultivation. Some of the most popular cultivars have a slow turnover and are more than a hundred years old. The pattern of diversity observed in our study is typical of longlived crops, for which germplasm has been shared extensively around the world.

By contrast, *Pereira-Lorenzo* et al. (2007), using SSR markers, revealed regional differentiation among local Spanish cultivars based on PCA (Principal Component Analysis) and cluster analysis. In Spain, apple growing and cultivar development has, however, a much longer history than in Sweden. Still the differentiation values (Fst) were low, suggesting the origination of local cultivars from a common gene pool, high gene flow between regions and minimal genetic isolation among populations.

We found a strong tendency for cultivars to cluster with one of their parental cultivars. Many other studies have similarly reported that apple genotypes cluster on UPGMA dendrograms in accordance with pedigree information (*Goulao* et al., 2001; *Gardner & Hokanson*, 2005). These results suggest that closely clustering cultivars can be regarded as genetically more similar than average cultivars. Consequently, information on clustering cultivars can become useful in character screenings; only those genotypes which appear to be most promising can be targeted instead of screening all available accessions.

Genetic diversity

Diversity values obtained with Shannon's index in our study were about twice as high as the values obtained with Nei's diversity index. The mean value of Shannon's index for 6 groups of cultivars was 0.546, and slightly higher, 0.593, when calculated across all cultivars. These values are similar to the mean value for wildgrowing populations of the outcrossing flowering quince *Chaenomeles speciosa* (0.552) (*Bartish* et al., 2000*)*, somewhat higher than the mean value for collection sites of cultivated clones of the outcrossing Ethiopian crop plant *Ensete ventricosum* (0.498) (*Birmeta* et al., 2002) and somewhat lower than the mean value for wildgrowing populations of *E. ventricosum* (0.630) (*Birmeta* et al., 2004).

Mean value for Nei's genetic diversity was 0.262 in our study of six groups of apple cultivars and 0.268 when calculated for the entire set. These values are very similar to those obtained by *Marita* et al. (2000) for randomly sampled clones of cacao (0.305) and *Capsicum* (0.269), and approximately twice as high as for a collection of sour orange accessions (0.122) (*Siragusa* et al., 2006). All in all, apple appears to hold average values of genetic diversity when compared to other outcrossing crop species, and variously defined subsets hold almost the same amount of diversity as a larger set with cultivars originating at different points of time from a large array of different countries.

During the last twenty years, the number of apple breeding programs and released cultivars has increased remarkably. However, of the 1000 new cultivars released in these two decades, a majority are sports (mutant clones) (*Sansavini* et al., 2004) and therefore do not contribute to a

widening of the gene pool. Despite the recent activity in apple breeding, it has instead been suggested that the actually utilized gene pool is becoming dangerously narrow (*Noiton & Alspach*, 1996).

In our study, a small tendency towards a narrowing of the gene pool can be seen in the lower diversity for recent foreign cultivars compared to the other investigated groups of cultivars. The 18 cultivars belonging to the group 'recent foreign', represent 8 countries (USA, Canada, England, Netherlands, Norway, Denmark and Japan). Both parents are known for 8 of these cultivars while one parent is known for 6 cultivars, three have completely unknown pedigrees and one ('Queen Cox') is a sport. Pedigree analysis shows that 'Cox's Orange Pippin', 'Golden Delicious', 'McIntosh' and 'Worcester Pearmain' are the most commonly occurring cultivars in the pedigrees for the recent foreign group.

Interestingly, a corresponding decrease of genetic diversity could not be seen in the group with recent Swedish cultivars, which is probably due to the rather broad range of cultivars used in breeding programs. Ancient indigenous cultivars like 'Astrakan, Gyllenkrok's' have been used along with old foreign cultivars and novel selections.

Choice of mandate cultivars

Many studies have been devoted to the development and comparison of strategies and methods for assembling gene-banks and core collections in different plant species (*Dwivedi* et al., 2005; *Garkava-Gustavsson* et al., 2005; *van Raamsdonk & Wijnker*, 2000). At Balsgård, an active germplasm collection with approximately 1000 cultivars is being used in plant breeding and research. In addition, a set of cultivars, mostly of Swedish origination, have been granted 'mandate cultivar' status in Sweden and are now conserved mainly at outdoor museums and other public places, although many of them are also present in the Balsgård collection. The mandate cultivar status was accorded on historical merits and no priority has been given to the preservation of genetic diversity or to the availability of particular genes of interest for research and breeding. We therefore compared a set of cultivars from the active collection (151 cvs) with a subset containing only mandate cultivars (94 cvs). Both diversity values (Nei's genetic diversity and Shannon's index) were only marginally lower in the collection with mandate cultivars, suggesting that it has the same amount of variation as a randomly sampled collection would have. By contrast, collections created to preserve maximum genetic variation should have more diversity than randomly assembled collections. Thus, the mean genetic diversity in a collection of *Theobroma cacao*, created by the Maximum genetic diversity program was 0.377 compared to 0.305 in the case of random sampling (*Marita* et al., 2000). For *Capsicum* corresponding values were 0.361 and 0.269 (*Marita* et al., 2000) and for *Vaccinium vitis-idaea* 0.356 and 0.303 (*Garkava-Gustavsson* et al., 2005).

We also compared band frequencies in our two sets of cultivars, and there were no major changes although 5 bands were missing in the set of mandate cultivars. However, bands that were common in the entire collection remained common in the subset, and rare bands remained rare. By contrast, when a subset of lingonberry (*Vaccinium vitis-idaea*) genotypes had instead been choosen by a Maximum diversity algorithm procedure, all bands were preserved and moreover, the rare bands increased in frequency while common bands decreased (*Garkava-Gustavsson* et al., 2005).

Usually, mandate cultivars, aimed at preservation of mainly indigenous genetic resources, contain genotypes chosen for their cultural and historical values. This has clearly been the case for the mandate apple cultivars in Sweden. Although one could expect some overall similarities due e.g., to their adaptation to a cold climate, these cultivars appear to genetically constitute a random sampling and contain the same amount of variation as the larger set of Swedish and foreign cultivars analysed in our study.

A rather different approach for preservation of genetic resources has been taken in several other countries. Although conservation of historic cultivars is one of the purposes for gene banks in e.g., the National Plant Germplasm System (NPGS, www.ars-grin.gov/npgs) in USA, one of the main goals is to provide plant material for basic plant genetic research and breeding (*Postman* et al., 2006). Consequently, this gene bank as well as others, e.g., the National Fruit Collections at Brogdale in England (www.brogdale.org) and The Vavilov Institute in Russia (www.vir.nw.ru), contain carefully chosen cultivars from all around the world, including genotypes with especially important genes governing e.g., resistance against various diseases. The core collections developed at e.g., PGRU, which is a part of the NPGS, can therefore be regarded as sources of genes, rather than sources of genotypes and clones (*Volk* et al., 2005). Obviously in countries like Sweden, preservation of mandate cultivars must be complemented with preservation of germplasm that has been selected to increase genetic variation and to provide important genes in order to better fulfill the needs of researchers and plant breeders.

Acknowledgements

We thank Dr. Görel-Kristina Näslund and Gullmar Henäng for fruitful discussions about the origination of apple cultivars, Dr. Inger Hjalmarsson for information about the Swedish mandate cultivars, and Prof. T. Kondratenko for providing additional information about Russian and Ukrainian cultivars. Assistance with the PCO analysis was given by Dr. A. Kolodinska-Brantestam. Financial support was received from FORMAS (The Swedish Council for Environment, Agricultural Sciences and Spatial Planning).

References

Autio, W.R., Schupp, J.R., Ferree, D.C., Glavin, R. & Mulcahy, D.L. (1998): Application of RAPDs to DNA extracted from apple rootstocks. HortScience 33: 333–335.

Bartish, I.V., Garkava, L.P., Rumpunen, K. & Nybom, H. (2000): Phylogenetic relationships and differentiation among and within populations of *Chaenomeles* Lindl. (Rosaceae) estimated with RAPDs and isozymes. Theor. Appl. Genet. 101: 554–563.

Birmeta, G., Nybom, H. & Bekele, E. (2002): RAPD analysis of genetic diversity among clones of the Ethiopian crop plant *Ensete ventricosum*. Euphytica 124: 315–325.

Birmeta G., Nybom H. & Bekele, E. (2004): Distinction between wild and cultivated enset (*Ensete ventricosum)* gene pools in Ethiopia using RAPD markers. Hereditas 140: 139–148.

Bussel, J.D. (1999): The distribution of random amplified polymorphic DNA (RAPD) amongst populations of *Isotoma petraea* (Lobeliaceae). Mol. Ecol. 8: 775–789.

Cheng, F.S., Weeden, N.F. & Brown, S.K. (1996): Identification of co-dominant RAPD markers tightly linked to fruit skin colour in apple. Theor. Appl.Genet. 93: 222–227.

Conner, P.J., Brown, S.K. & Weeden, N.F. (1997): Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. J. Am. Soc. Hort. Sci: 122, 350–359.

Crosby, J.A., Janick, J., Pecknold, P.C., Korban, S.S., O'Connor, P.A., Ries, S.M., Goffreda, J. & Voordeckers, A. (1992): Breeding apple for scab resistance: 1945-1990. Fruit Var. J. 46: 145–166.

Dahl, C.G. (1929): Pomologi. Albert Bonniers förlag, Stockholm.

Dansi, A., Mignouna, H.D., Zoundjihékpon, J., Sangare, A., Ahoussou, N. & Asiedu, R. (2000): Identification of some Benin Republic's Guinea yam (*Dioscorea cayenensis/Dioscorea rotundata* complex) cultivars using Randomly Amplified Polymorphic DNA. Genet. Resour. Crop Evol. 47: 619–625.

Dunemann, F., Kahnau, R. & Schmidt, H. (1994): Genetic relationships in *Malus* evaluated by RAPD 'fingerprinting' of cultivars and wild species. Plant Breed. 113: 150–159.

Dwivedi S.L., Upadhyaya H.D. & Hegde D.M. (2005): Development of core collection using geographic information and morphological descriptors in safflower (*Carthamus tinctorius* L.) germplasm. Genet. Resour. Crop Ev. 52: 821–830.

Gardner, N., Hokanson, C. (2005): Intersimple sequence repeat fingerprinting and genetic variation in a collection of *Clematis* cultivars and commercial germplasm. HortSci. 40: 1982–1987.

Garkava-Gustavsson, L. & Nybom, H. (2003): DNA-analyser avslöjar våra äppelsorter. Fakta Trädgård-Fritid (SLU) 94: 1–4.

Garkava-Gustavsson, L., Persson, H.A., Nybom, H., Rumpunen, K., Gustavsson, B.A. & Bartish, I.V. (2005): RAPD-based analysis of genetic diversity and selection of lingonberry (*Vaccinium vitis-idaea* L.) material for *ex situ* conservation. Genet. Resour. Crop Evol. 52: 723–735.

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M. & Gessler, C. (1998): Simple sequence repeats for the genetic analysis of apple. Theor. Appl. Genet. 96: 1069–1076.

Ghislain, M., Zhang, D., Fajardo, D., Huamán, Z. & Hijmans, R.J. (1999): Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genet. Resour. Crop Evol. 46: 547–555.

Goulao, L. & Oliveira, C.M. (2001): Molecular characterisation of cultivars of apple (*Malus* x *domestica* Borkh.) using microsatellite (SSR and ISSR) markers. Euphytica 122: 81–89.

Goulao, L., Cabrita, L., Oliveira, C., & Leitao, J.M. (2001): Comparing RAPD and AFLP™ analysis in discrimination and estimation of genetic similarities among apple (*Malus domestica* Borkh.) cultivars. Euphytica 119: 259–270.

Grati-Kamoun, N., Lamy Mahmoud, F., Reba, A., Gargouri, A., Panaud, O. & Saar, A. (2006): Genetic diversity in Tunisian olive tree (*Olea europaea* L.) cultivars assessed by AFLP markers. Genet. Resour. Crop Evol. 53: 265–275.

Harada, T., Matsukawa, K., Sato, T., Ishikawa, R., Niizaki, M. & Saoto, K. (1993): DNA-RAPDs detect genetic variation and paternity in *Malus*. Euphytica 65: 87–91.

Harris, S.A., Robinson, J.P., & Juniper, B.E. (2002): Genetic clues to the origin of the apple. Trends in Genet. 18: 426–430.

Holb I.J. (2000): Disease progress of apple scab caused by *Venturia inaequalis* in environmentally friendly growing systems. International Journal of Horticultural Science 6. (4): 56–62.

Hjalmarsson, I. & Wallace, B. (2004): Content of the Swedish berry gene bank. J. Fruit Ornamental Plant Res. 12: 129–138.

Hokanson, S.C., Szewc-McFadden, A.K., Lamboy, W.F. & McFerson, J.R. (1998): Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. Theor. Appl. Genet. 97: 671–683.

Janick, J., Cummins, N., Brown, S.K. & Hemmat, M. (1996): Apples. In: Janick J. and Moore J. (eds.). Fruit Breeding – Vol.1. Tree and tropical fruits. John Wiley & Sons Inc., New York, 1–27.

Kim, M.Y., Song, K.J., Hwang, J.H., Shin, Y.U. & Lee, H.J. (2003): Development of RAPD and SCAR markers linked to the *Co* gene conferring columnar growth habit in apple (*Malus pumila* Mill.). J. Hort. Sci. Biotech. 78: 512–517.

Kitahara, K., Matsumoto, S., Yamamoto, T., Soejima, J., Kimura, T., Komatsu, H. & Abe, K. (2005): Molecular characterization of apple cultivars in Japan by S-RNase analysis and SSR markers. J. Am. Soc. Hort. Sci.130: 885–892.

Koller B., Lehmann, A., McDermott, J.M. & Gessler, C. (1993): Identification of apple cultivars using RAPD markers. Theor. Appl. Genet. 85: 901–904.

Lawson, D.M., Hemmat, M. & Weeden, N.F. (1995): The use of molecular markers to analyze the inheritance of morphological and developmental traits in apple. J. Am. Soc. Hort. Sci. 120: 532–537.

Lema-Ruminska, J., Zalewska, M. & Sadoch, Z. (2004): Radiomutants of chrysantemum (*Dendranthema grandiflora* Tzvelev) of the Lady group: RAPD analysis of the genetic diversity. Plant Breed. 123: 290–293.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C. D., Tarchini, R., van de Weg, E. & Gessler C. (2002): Development and characterisation of 140 new microsatellites in apple (*Malus* x *domestica* Borkh.). Mol. Breed. 10: 217–241.

Marita, J.M., Rodriguez, J.M. & Nienhuis, J. (2000): Development of an algorithm identifying maximally diverse core collections. Genet. Resour. Crop Ev. 47: 515–526.

Mattisson, H. & Nybom, H. (2005): Application of DNA markers for detection of scab resistant apple cultivars and selections. International J. Horticult. Sci. 11: 59–63.

Monte-Corvo, L., Cabrita, L., Oliveira, C. & Leitao, J. (2000): Assessment of genetic relationships among *Pyrus* species and cultivars using AFLP and RAPD markers. Genet. Resour. Crop Evol. 47: 257–265.

Mulcahy, D.L., Cresti, M., Sansavini, S., Douglas, G.C., Linskens, H.F., Bergamini Mulcahy, G., Vignani, R. & Pancaldi,

M. (1993): The use of random amplified polymorphic DNAs to fingerprint apple genotypes. Sci. Hortic. 54: 89–96.

Nei, M. (1987): Molecular evolutionary genetics. Columbia University Press, New York, NY.

Nilsson, A. (1986): Våra äpplesorter. AB Allmänna Förlaget, Stockholm.

Noiton, D.A.M. & Alspach, P.A. (1996): Founding clones, inbreeding, coancestry and status number of modern apple cultivars. J. Am. Soc. Hort. Sci. 121: 773–782.

Nybom, H. & Schaal, B.A. (1990): DNA „fingerprints“ applied to paternity analysis in apples *(Malus* x *domestica)*. Theor. Appl. Genet. 79: 763–768.

Nybom, H., Rumpunen, K., Persson Hovmalm, H., Marttila, S., Rur, M., Garkava-Gustavsson, L. & Olsson, M.E. Towards a healthier apple – chemical characterisation of an apple gene bank. Acta Hort., in press.

Pereira-Lorenzo, S., Ramos-Cabrer, A.M. & Díaz-Hernández, M.B. (2007): Evaluation of genetic identity and variation of local apple cultivars (*Malus x domestica* Borkh.) from Spain using microsatellite markers. Genet. Resour. Crop Evol. 54: 405–420.

Postman, J., Hummer, K., Stover, E., Krueger, R., Forsline, P., Grauke, L.J., Zee, F., Ayala-Silva, T. & Irish, B. (2006): Fruit and nut genebanks in the U.S. National Plant Germplasm System. HortScience 41: 1188–1194.

Raamsdonk, L.W.D.van & Wijnker, J. (2000): The development of a new approach for establishing a core collection using multivariate analyses with tulip as a case. Genet. Resour. Crop Evol. 47: 403–416.

Rohlf, M. (1998): NTSYS-pc: numerical taxonomy and multivariate analysis system. Version 2.2. Dept. of Ecology and Evolution. State Univ. of New York.

Sansavini, S., Donati, F., Costa, F. & Tartarini S. (2004): Advances in apple breeding for enhanced fruit quality and resistance to biotic stresses: new varieties for the European market. J. Fruit Ornam. Plant Res. 12: 13–52.

Siragusa, M., De Pasquale, F., Abbate, L. & Tusa N. (2006): Identification of sour orange accessions and evaluation of their genetic variability by molecular marker analyses. HortScience 41: 84–89.

Svensson H. & Kastman K. (2005): Äpplen i Sverige. Bokförlaget Prisma, Stockholm.

SPSS 11.0 for Macintosh Brief Guide, Copyright © 2002 by SPSS Inc.

Vijayan, K., Nair, C. V. & Chatterjee, S. N. (2005): Molecular characterization of mulberry genetic resources indigenous to India. Genet. Resour. Crop Ev. 52: 77–86.

Volk, G.M., Richards, C.M., Reilley, A.A., Henk, A.D., Forsline, P.L. & Aldwinckle, H.S. (2005): *Ex situ* conservation of vegetatively propagated species: Development of a seed-based core collection for *Malus sieversii.* J. Am. Soc. Hort. Sci. 130: 203–210.

Weeden, N.F. & Lamb, R.C. (1985): Identification of apple cultivars by isozyme phenotypes. J. Am. Soc. Hort. Sci. 110: 509–515.

Weising, K., Nybom H., Wolff K. & Kahl G. (2005): DNA fingerprinting in plants. Principles, methods and applications. Second edition. Taylor and Francis Group/CRC Press, Boca Raton

Yang, H.Y., Korban, S.S., Kruger, J. & Schmidt H. (1997): A randomly amplified polymorphic DNA (RAPD) marker tightly linked to the scab-resistance gene *Vf* in apple. J. Am. Soc. Hort. Sci. 122: 47–52.

Xu, M.L. & Korban, S.S. (2000): Saturation mapping of the apple scab resistance gene *Vf* using AFLP markers. Theor. Appl. Genet. 101: 844–851.

Zhou, Z.Q. & Li, Y.N. (2000): The RAPD evidence for the phylogenetic relationship of the closely related species of cultivated apple. Genet. Resour. Crop Evol. 47: 353–357.

V

Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis

L. Garkava-Gustavsson[1], A. Kolodinska Brantestam[2], J. Sehic[1] and H. Nybom[1]

[1]Balsgård–Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, Kristianstad, Sweden
[2]Nordic Gene Bank, Alnarp, Sweden

Abstract

Trees of 68 apple cultivars, aimed for preservation by the 'National Program for diversity of cultivated plants' as mandate cultivars, were analysed using a set of 10 SSR (Simple Sequence Repeat) primer pairs and the self-incompatibility (S-)locus to evaluate genetic diversity and reveal inter-cultivar relationships. The 12 polymorphic SSR loci exhibited 2 to 15 alleles, with expected heterozygozity (H_e) ranging from 0.36 to 0.88 and a mean of 0.74. Numerous alleles were classified as rare or unique (35% and 18% respectively). For the S-locus, a total of 14 alleles were identified in this study. Five alleles, S_1–S_3, S_5 and S_7 had frequencies ranging from 11 to 18%, whereas the remaining 9 alleles were below 6%. All sexually obtained cultivars could be distinguished with the set of SSR loci. Sports were identical with their progenitors in two cases, but differed in one SSR allele in a third case. An SSR-based dendrogram, based on Roger's genetic distances, did not reveal any clear pattern of clustering. The genetic distances were, however, correlated with a corresponding matrix obtained in a previously conducted RAPD-based study of the same cultivars. Non-mandate parents of Swedish mandate cultivars together with some other reference cultivars were included in this study to check the accuracy of allele scoring, verify parentage and compare the results of this study with those presented in previously published studies. Some discrepancies in allele sizing were revealed and the possibilities of avoiding this problem are discussed.

Introduction

Accurate and permanent identification of plant material within a germplasm collection is of utmost importance, especially for vegetatively propagated cultivars which are expensive to maintain, and should consist of a single genotype in the whole distribution area. In Sweden, the preservation of clonally propagated plants is presently managed by a governmentally appointed unit, the 'National Program for Diversity of Cultivated Plants'. Mandate status has been granted to indigenous cultivars which have been named, bred, propagated and marketed in Sweden. Some foreign cultivars with a long history of being grown in Sweden are also included (Hjalmarsson & Wallace, 2004). In apple (*Malus* x *domestica* Borkh.),

about 220 mandate cultivars have been appointed and are now conserved mainly in 11 clone archives (each with 10–50 mandate cultivars) located at outdoor museums or other public places. The largest apple collection in Sweden, with about 1000 cultivars in total, is however found at Balsgård in the southernmost province Skåne, where both research and applied breeding of new cultivars is undertaken. Of the 100 mandate cultivars in the germplasm collection at Balsgård, 68 have apparently been developed in Sweden, and are regarded as having a Swedish origin although one or both parents may have been developed elsewhere.

Unfortunately, many of the older (local) cultivars have not been properly identified, and there are problems with e.g., synonyms and mis-labellings in these plant collections. In order to ensure that our research data, obtained from rather costly analyses of e.g., chemical contents in apple cultivars (Nybom *et al.*, in press), can be unambiguously connected with well-identified genotypes, we need to provide these cultivars with proper and lasting identification. Morphological characters have historically been, and still are, used for identification of apple cultivars (Nilsson, 1986; Svensson & Kastman, 2005), but environmentally induced variation makes correct identification difficult. Therefore, molecular protein-based markers like isozymes (Weeden & Lamb, 1985), and DNA-based markers have become increasingly popular in the characterisation of apple genetic resources.

Several molecular DNA techniques have been employed for characterisation of apple germplasm collections, as well as in apple genetics and breeding, i.e., RFLP (Nybom & Shaal, 1990), RAPD (Koller *et al.*, 1993), ISSR (Goulao & Oliveira, 2001), AFLP (Xu &Korban, 2000) and SSR (Liebhard *et al.*, 2002). Each of these methods has its strengths and weaknesses. During the past 10–15 years, microsatellites or Simple Sequence Repeats have become the markers of choice for verification of cultivar identity and for diversity studies due to their abundance in plant genomes, large number of alleles per locus and high informativeness, codominant inheritance, and suitability for automatization (Weising *et al.*, 2005). SSR loci are relatively easy to score, and alleles are inherited in a Mendelian manner, which allows the verification and reconstruction of cultivar pedigrees. In addition, triploid and tetraploid individuals are revealed by the appearance of more than two alleles per locus. In apple, SSR markers have been used for assessment of genetic diversity in germplasm collections (Guarino *et al.*, 2006; Pereira-Lorenzo *et al.*, 2007), cultivar identification (Guilford *et al.*, 1997; Galli *et al.*, 2005), fingerprinting of apple rootstocks (Oragusie *et al.*, 2005), construction of genetic linkage maps (Kennis & Keulemans, 2005; Silfverberg-Dilworth *et al.*, 2006) and for parent identification (Kitahara *et al.*, 2005).

Another, highly polymorphic locus is the S-locus, which encodes for the different S-RNases that determine the S-alleles of different genotypes and causes self-incompatibility. Apple cultivars that share one S-allele have reduced compatibility and do not achieve their potential yield capacity if planted together with no other cultivars close-by (Schneider *et al.*, 2005). If both S-alleles are shared, the cultivars are usually incompatible and do not yield at all except for occasional fruits obtained by bypassing the self-incompatibility system through pollination between genetically incompatible genotypes or through selfing. In

apple, allele-specific primer pairs have been developed and used in a number of cultivar screenings (Janssens *et al.*, 1995; van Nerum *et al.*, 2001; Broothaerts 2003; Broothaerts *et al.*, 2004; Melounova *et al.*, 2005; Matsumoto *et al.*, 2007). Information about S-locus allele composition can be used for identification of apple cultivars and determination of parentage (Kitahara *et al.*, 2005) if complemented with other marker systems.

The purpose of this study was i) to fingerprint 68 mandate apple cultivars with SSR and S-locus markers and ii) to determine genetic diversity and examine genetic relationships within the set of mandate cultivars.

Material and Methods

Plant material

One tree of each of 68 apple cultivars, designated as 'Swedish mandate cultivar' and originating from Sweden, were analysed in this study (Table 1). These trees had previously been analysed by RAPD markers as part of another study (Garkava-Gustavsson & Nybom, 2007). The material included some cultivars with known ancestry: 'Alfa 68' ('Boskoop' x 'Filippa'), 'Alice' ('Ingrid Marie' open pollination), 'Aroma' ('Ingrid Marie' x 'Filippa'), 'Eva-Lotta' ('Cortland' x 'James Grieve'), 'Katja' ('James Grieve' x 'Worcester Pearmain'), 'Kim' ('Cortland' x 'Ingrid Marie'), 'Mio' ('Worcester Pearmain' x 'Oranie'), and 'Sylvia ' ('Astrakan, Gyllenkrok's' x 'Worcester Pearmain'). The non-mandate parents were added to the study to check the pedigree information. In addition, several reference cultivars were included to facilitate comparisons with previously published allele-size data. However, only the basic set of 68 cultivars was used in the statistical analyses.

DNA-isolation

Newly expanded leaves were collected in April–May and stored at -80°C until use. Leaves were powdered with liquid nitrogen in pre-cooled mortars. Genomic DNA was isolated from approximately 100 mg of leaf powder using the Qiagen Dneasy^{TM} Plant Mini Kit and following the manufacturer's protocol.

SSR analysis

Ten SSR primer pairs were used in this study (Table 2). All primer pairs were obtained from the list of 140 (Liebhard *et al.*, 2002): CH01d03, CH01h02, CH02b10, CH02c06, CH02c09, CH02c11, Ch02d08, CH04c06, CH04e05, COL. All but one, CH02b10, belong to the "standard set" defined by the European working group on apple genetic resources (Laurens *et al.*, 2004; Guarino *et al.*, 2006). The primer pair CH02b10 had previously been used successfully at our laboratory (Mattisson & Nybom, 2005) and was therefore included in this study.

All PCR amplifications were performed in a 15 µl volume according to Gianfranceschi *et al.* (1998). A Px2 Thermal Cycler (Thermo Hybaid) was programmed according to the original protocol (Gianfrancheschi *et al.*, 1998) for all primer pairs, except CH02c06. The cycling profile for Ch02c06 was modified according to K. Antonius, MTT, Finland (personal communication) and consisted of an initial denaturation at 94°C for 2 min 30 s followed by 34 cycles of 94°C for 30 s, 58°C for 1 min, 72°C for 1 min. A final extention at 72°C for 5 min was included. PCR products were preliminary checked by running on 2% agarose gel in a 1x TBE buffert, stained with ethidium bromide and visualised by UV light.

The forward primers were fluorescently labelled at the 5'-end with either FAM (Ch01d03, CH04c06, Ch04e05) or HEX (CH01H02, CH02b10, CH02c06, Ch02c09, CH02c11, CH02d08, COL). The PCR products were separated and analysed on a 3730 DNA Analyzer (Applied Biosystems). The size of the amplified products was calculated based on an internal standard (500ROXTM Size Standard (Applied Biosystems)) with GeneMapper® Software v 3.0 (Applied Biosystems). A manual binning step was included to assign all detected alleles to their repeat units equivalent.

To monitor the reproducibility between runs, we included the same three cultivars, namely 'Discovery', 'Golden Delicious' and 'Guldborg' in every run. In addition, two cultivars (different in each run) were amplified twice.

S-allele analysis

All mandate cultivars except 'Åkerö, Gripsholm' (sport of 'Åkerö') were analysed, together with four parental non-mandate cultivars (Table 1) and 11 additional reference cultivars: 'Boskoop' ($S_2S_3S_5$), 'Cox's Orange' (S_5S_9), 'Delicious' (S_9S_{28}), 'Gravensteiner' ($S_4S_{13}S_{20}$), 'James Grieve' ($S_5S_?$), 'Jonathan' (S_7S_9), 'Lobo' ($S_{10}S_{22}$), 'Maypole' ($S_{10}S_{16}$), 'Mutsu' ($S_2S_3S_{20}$), 'Ribston' ($S_1S_9S_{21}$) and 'Wealthy' (S_3S_9).

Allele-specific primers (MWG) were employed for detection of alleles S1–S5, S7, S9, S10, S16, S20, S22, S24, S25 and S28 using the nucleotide sequences published by Broothaerts (2003), except for S5 and S25 which were detected with primers published by Matsumoto *et al.* (1999) and by Kitahara & Matsumoto (2002), respectively.

PCR conditions used for all primers except for S5 and S25, included 1xPCR buffer (ABgene), 1.75 mM of MgCl2 (1.25 mM for S20), 200 µM dNTPs, 1 µM of each primer and 0.6 U of Taq DNA polymerase in 20 µl total volume. Approximately 40 ng of genomic DNA template were used per reaction. For primers S5 and S25, a 35 µl total volume was used including 1x buffer, 1.07 mM MgCl2,, 200 µM dNTPs, 1 µM of each primer and 2.5 U of Taq DNA polymerase. Five different PCR amplification protocols (Table 3) were run on a Px2 Thermal Cycler.

A protocol published by Broothaerts (2003) for the discrimination of five different alleles, S6, S14, S17, S20 and S21 using the single primer pair

S6/14/17/20/21, did not work properly in our laboratory. By lowering the amount of MgCl2 to 1.25 mM, a clear band was however produced in controls reported to have S20 and not in any other. This procedure was therefore used to detect S20 in the investigated cultivars.

Detection of the three alleles S_4, S_{16} and S_{22} was carried out using a single primer pair, $S_{4/16/22}$, and subsequently digesting 20 µl of the amplification product with 10 U restriction endonuclease *Taq*1 (Sigma) in 3 µl 10x restriction endonuclease buffer and a total volume of 25 µl for 2 h at 65 ºC.

The amplified and digested fragments were separated on a 2.5% agarose gel stained with ethidium bromide. The gels were run for approximately 2–3 hours, for the digested fragments 4 hours, at 120–150 V, together with a DNA molecular weight marker, and then analyzed on a UV-light table.

Flow cytometry

Cultivars that amplified more than two alleles in an SSR locus and/or the S-locus were analysed for ploidy level. Leaf samples were subjected to flow cytometry analysis by Plant Cytometry Services (JG Schijndel, The Netherlands). The leaves were chopped in icecold buffer with DAPI. Flow cytometry was performed on a CyFlow ML (Partec GmbH, Münster, Germany) using *Lactuca sativa* as internal standard.

Data analysis

The SSR fragments were scored in terms of loci and alleles and thus allelic composition of each cultivar was determined. To evaluate the genetic diversity within our set of cultivars, we used the following parameters: alleles per locus (A), the effective number of alleles per locus (A_e), expected heterozygosity (H_e) and observed heterozygosity (H_o).

Effective number of alleles per locus (A_e) was calculated as in Aranzana *et al.* (2003):

$$A_e = 1/\Sigma(p_i)^2$$

where p is the frequency of the i^{th} allele at a locus.

Expected heterozygosity (H_e), i.e., the probability that two alleles from the same locus would be different when chosen at random, was calculated for each SSR locus according to Nei (1973):

$$H_e = 1 - \Sigma(p_i)^2.$$

Observed heterozygosity (H_o) was calculated by dividing the number of heterozygous individuals by the number of individuals scored.

Power of discrimination (PD), which estimates the probability that two randomly sampled individuals are discriminated, was calculated for each locus according to Kloosterman *et al.* (1993):

$$PD = 1 - \Sigma(P_i)^2,$$

where P_i is the frequency of the i^{th} genotype.

Confusion probability, which is the probability that any two cultivars are identical in their SSR genotypes at all loci by chance alone, was calculated based on PD values as in Aranzana *et al.* (2003):

$$C = \Pi(1 - PD_i),$$

where PD_i is the PD value at the i^{th} locus.

Genetic distances using Rogers' dissimilarity coefficient (Rogers 1972) were calculated and a matrix was produced. Rogers' coefficient is linearly related to the coefficient of coancestry, and is regarded as appropriate for the uncovering of pedigree relationships among operational taxonomic units such as the detection of essentially derived varieties in plant breeding or the identification of duplicates and collection gaps in seed banks (Reif *et al.*, 2005). The matrix was used in an UPGMA analysis performed using the NTSYS-pc statistical package v. 2.2 (Rohlf, 2005) and a dendrogram, illustrating the genetic relationships among cultivars, was produced.

In addition, the SSR fragments were also scored phenotypically, with 1 for presence of a fragment and 0 for absence. Jaccard's coefficient of similarity was used to produce a similarity matrix.

Correlation between the two matrices obtained with Rogers' genetic distances and with Jaccard's similarity coefficients, respectively, was investigated with a Mantel test (MXCOMP in NTSYS-pc, using 9999 permutations to compute the significance of a given correlation). In addition, Mantel tests were also carried out to compare each of the two SSR-based matrices with a RAPD-based Jaccard's similarity matrix produced from data on the same trees in a previous study (Garkava-Gustavsson & Nybom, 2007).

Results

SSR polymorphism

All but one of the primer pairs amplified clear and easily scored fragments. The accuracy of scoring was checked by comparing SSR profiles of cultivars with the profiles of their known or putative parents when available, and with the reference cultivars. We did, however, experience difficulties when scoring fragments amplified by primer pair CH01d03. According to Liebhard *et al.* (2002), this primer pair amplifies more than one locus and it was difficult to assign some of the alleles to the correct locus. Therefore, fragments amplified by this primer pair were not considered in the statistical analyses, although they can still be used for identification purposes.

Primer pairs CH01h02, CH02c11 and CH04c06 amplified two loci each. The main locus, i.e. the locus described in Liebhard *et al.* (2002), is denoted as locus 2 (L2) in the present study. The other locus, denoted as locus 1 (L1), had shorter fragments and was less polymorphic than L2 for CH01h02 and CH04c06, and was monomorphic for CH02c11. The monomorphic locus CH02c11-L1 was not included in the statistical analyses.

Thus, 11 polymorphic loci amplified by 9 primer pairs were analysed statistically, yielding a total of 113 polymorphic alleles in the set of 68 cultivars (Table 2). Number of alleles per polymorphic locus ranged from 2 (Ch04c06-L1) to 15 (CH02c06 and CH02b10) with a mean of 10.4. Number of effective alleles (Ae) ranged from 1.56 (CH04c06-L1) to 8.50, with a mean of 4.92. Power of discrimination (PD) varied from 0.60 (Ch04c06-L1) to 0.97 (CH02b10), with a mean of 0.89. Probability of confusion, based on PD values, was estimated as 9.941 x 10-13. Expected heterozygozity (He) varied from 0.36 (CH04c06-L1) to 0.88 (CH02b10) with a mean of 0.74, while observed heterozygosity (Ho) varied from 0.42 (CH01h02-L1) to 0.96 (CH02b10), also with a mean of 0.74.

In general, allele frequencies were distributed unevenly within the investigated loci (Fig. 1). Each locus had 1 to 4 more common alleles, which in some cases (CH01h02-L1, CH01h02-L2, CH02d08 and CH04e05) had considerably higher frequencies than the other alleles in these loci. All but one (CH01h02-L1) had rare alleles, here defined as having a frequency below 0.05. All but two loci, L1 and L2, amplified by primer CH01h02, had unique alleles, i.e. present in one cultivar but not in any other. Loci CH02c09, CH02c11 and CH04c06-L2 had 1 unique allele each, and 2, 4 and 5 rare alleles respectively. Loci CH04e05 and CH02b10 had 3 unique alleles each, and 7 and 5 rare alleles respectively. Finally, loci CH02c06 and CH02d08 had 5 unique alleles each, and 5 and 4 rare alleles respectively. Thus, in the whole set of cultivars, only 15 alleles (approx. 13%) were found at frequencies above 0.20, 38 alleles (approx. 34%) were found at frequencies between 0.05 and 0.20, 39 alleles (35%) were rare and, finally, 21 alleles (18%) were unique.

SSR-based cultivar identification

Allelic compositions for each analysed apple cultivar are presented in Table 1. Diploid cultivars amplified one or two fragments in each locus. When only one fragment was detectable in a diploid cultivar, we considered the cultivar to be homozygous. Still, presence of null alleles cannot be excluded, and therefore our heterozygosity values can be underestimated. A putative homozygous null allele was detected in cultivar ‘Granatäpple, Kungsbacka’ at L1 amplified by primer pair CH01h02. The amplification reaction was repeated four times using two different 96-well plates but no PCR-product was obtained.

Three distinct alleles were revealed at 3 to 8 loci, indicating triploidy, in one of the reference cultivars (‘Boskoop’) and in 7 of the Swedish cultivars (‘Frösåker’, ‘Holländaräpple’, ‘Kalmar Glasäpple’, ‘Kinnekulle Kantäpple’, ‘Norrstack’, ‘Villands Glasäpple’ and ‘Vrams Järnäpple’) (Table 1). The triploid status of these cultivars was confirmed by flow cytometry analyses. We also checked one cultivar that was already known to be tetraploid, ‘Alfa 68’. By comparing the SSR profiles of this cultivar with the profiles of its two parents, ‘Boskoop’ and ‘Filippa’, we concluded that the maternal parent, ‘Boskoop’, had contributed an unreduced triploid egg cell while 'Filippa' had contributed a normal, haploid pollen. In total, ‘Alfa 68’ amplified 4 different alleles at 6 loci, 3 alleles at 4 loci and 2 alleles in 2 loci. In those cases where only two alleles were amplified in a heterozygous

triploid cultivar, we usually managed to determine which allele occurred in two copies by comparing allele peak areas. However, the two alleles COL-230 or COL-232 had similar peak areas in our reference triploid cultivar ‘Boskoop’ and in its offspring, ‘Alfa 68’.

Analysis of parents for some of the cultivars in our material contributed to the accuracy of allele scoring. All alleles found in ‘Alfa 68’, ‘Eva-Lotta’, ‘Katja’, ‘Kim’, ‘Mio’ and ‘Sylvia’ were thus found also in their parents except for two loci (CH01h02-L1 and COL). At locus CH01h02-L1, one of the parents of ‘Eva-Lotta’, 'Kim' and ‘Sylvia’ showed only one allele and this was not found in the offspring, which also had only one allele. Similar cases were revealed for ‘Mio’ and ‘Sylvia’ at locus COL. Although presence of null alleles can be suspected, segregating populations should be analysed before drawing any definite conclusions.

Comparison of SSR data for some of the presently analysed cultivars with data from previously published studies revealed a few discrepancies in allelic composition, and several differences in allele sizes (Table 4). Since these studies have been based on different trees, minor somatic mutations cannot be ruled out, but experimental artefacts are also quite likely, especially concerning allele sizes.

SSR-based cluster analysis

A dendrogram based on Rogers’ genetic distances did not produce any major clustering (Fig. 2). Clear grouping was also missing in the plot produced by a Principal Coordinate Analysis (PCO) (data not shown). The majority of bifurcations in the dendrogram occurred at genetic distances above 0.30, indicating that the indigenous Swedish mandate cultivars are generally not closely related to one another. A similar lack of well-defined clusters was observed in a previous study based on RAPD (Garkava-Gustavsson & Nybom, 2007).

Two pairs of cultivars, which could not be distinguished in the previous RAPD-based study, namely ‘Spässerud’ and ‘Särsö’, and ‘Åkerö’ and ‘Åkerö, Gripsholm’, remained identical in the present study as well. However, ‘Grågylling’ and its red-coloured mutant ‘Fagerö’ were identical according to their RAPD band patterns but differed in one SSR allele; ‘Grågylling’ had allele composition 220:241 in locus COL, whereas ‘Fagerö’ did not have allele 241 (the most common allele in this locus) but instead the unique allele 228. The PCR reactions were repeated four times using two different 96-well plates to confirm this fact.

Cultivars ‘Hanaskog’ and ‘Oranie’ clustered at a distance of 0.17 in the dendrogram. Thus a close relationship between these two cultivars can be suspected. We compared the SSR allelic profiles of ‘Hanaskog’ and ‘Oranie’ and found that these cultivars shared 2 alleles in 7 loci, and 1 allele in 4 loci. Thus, ‘Hanaskog’ may well be an offspring of the presumably much older ‘Oranie’.
We found a highly significant negative correlation between the two SSR-based matrices, calculated with Rogers’ genetic distance and with Jaccard’s similarity coefficient, respectively ($r = -0.863$, $P < 0.001$). We also found a significant negative correlation ($r = -0.310$, $P < 0.001$) between Rogers’ genetic distances, obtained in this study and Jaccard’s coefficient of similarity, obtained in the

RAPD-based study (Garkava-Gustavsson & Nybom 2007). Finally, we found a positive correlation between the two Jaccard's coefficient of similarity matrices, for RAPD and for phenotypically scored SSR fragments, respectively, $r = 0.377$, $P < 0.001$.

S-allele configuration

A total of 14 different S-alleles were analysed in the mandate cultivars and non-mandate parents (Table 1). The S-alleles found in 11 reference cultivars were identical with those previously reported (Broothaerts *et al.*, 2004; Matsumoto *et al.*, 2007) except that we did not screen for S_{13} (present in 'Gravensteiner') or S_{21} (present in 'Ribston'). One of the mandate cultivars ('Katja') and three of the non-mandate parents ('Ingrid Marie', 'Cortland' and 'Worcester Pearmain' had also been investigated before (Broothaerts *et al.*, 2004; Matsumoto *et al.*, 2007) and found to have the same S-allele configurations as in the present study.

In general, two (or fewer) S-alleles were found in diploids and three (or fewer) in triploids, with one exception; the putatively diploid 'Ökna Vita Vintergylling' had three alleles possibly resulting from the ancient amphidiploidy of domestic apple.
Complete S-allele constitution was determined for 42 of the 67 analysed mandate cultivars. Another 23 cultivars lacked one allele, and 2 cultivars (both triploid), lacked two alleles. In total, this means that there were 19% missing alleles provided that all the cultivars are heterozygous in the S-locus.

The most commonly occurring S-allele of those that could be identified was S_3 (18%), followed by S_5 and S_7 (16%), S_2 (14%) and S_1 (11%) while the remaining 9 alleles had frequencies 6% (S_{20}), 4% (S_{22}), 3% (S_4, S_{10}, S_{28}), 2% (S_{16}, S_{24}, S_{25}) and 1% (S_9).

Discussion

Identification of SSR alleles

Proper choice of primer pairs is a very important component in SSR-based analyses, especially when results from several laboratories are pooled together in a larger database. For *Theobroma cacao,* 15 SSR primers were chosen as an international molecular standard because they had the highest reproducibility and consistency within a common genotype, while still allowing good differentiation of separate genotypes (Saunders *et al.*, 2004). When 46 grape cultivars were analysed by ten partners at six SSR loci without any efforts to standardize equipment or protocols, results obtained by some partners were very similar but different allele sizes were obtained in other cases (This *et al.*, 2004). A strategy for data comparison by means of reference to the alleles detected in well-known cultivars was therefore proposed.

In apple, different sets of primers have been used, mainly from the list of Liebhard *et al.* (2002). The primers used in this study have previously been applied by Gianfrancheschi *et al.* (1998), Liebhard *et al.* (2002), Guarino *et al.* (2006) and Ramos-Cabrer *et al.* (2007). Silfverberg *et al.* (2006) proposed to include two or three reference cultivars, namely 'Fiesta', 'Discovery' and 'Prima', which have been tested with most apple SSR markers and have been involved in many genetic studies, being parental cultivars of various mapping populations in Europe. We included 'Discovery' and 'Prima' and several other reference cultivars in the present study (Table 1), and could therefore compare our results directly with previously published data concerning genetic diversity estimators, allele sizes and allele compositions for some internationally well-known cultivars. Allelic composition was usually similar, but allele size differences of 1 to 4 bp were found in all but 3 locus/study combinations (Table 4). Moreover, pairwise comparisons of data from different studies revealed a serious lack of consistency between loci and sometimes even within the same locus. The differences are probably due mainly to the use of different internal size standards and to differences in binning procedures.

Since we can expect true alleles to differ by only 1 bp due to mutations in repeat flanking regions (Guarino *et al.* 2006), questionable allele sizing must be avoided. Most apple SSR primers contain dinucleotide repeats, which are characterized by intensive stuttering, sometimes leading to misscoring of homozygous versus heterozygous alleles. By contrast, for SSR loci containing trinucleotide repeats, is not always possible to distinquish SSR amplicons from other PCR products due to lack of stutter bands. In such cases, all amplicons are usually reported, which may lead to an overestimation of the level of polymorphism of these SSRs (Silfverberg-Dilworth *et al.*, 2006).

Several studies have advocated the merits of registering relative instead of absolute size of alleles since many factors can influence the migration of PCR fragments through the gel (Galli *et al.*, 2005; Amos *et al.*, 2007). Two methods of standardisation have been applied to SSR fingerprinting in *Theobroma cacao*: i) the use of a partial allelic ladder through the production of cloned and sequenced allelic standards and ii) the use of standard genotypes selected to display a diverse allelic range (Cryer *et al.*, 2006). The use of cloned sequence-specific internal DNA standards to calibrate the size of microsatellite alleles could eliminate the problem with allele sizing in plants in international projects and is already widely used in human clinic and forensic studies (Bruland *et al.*, 1999; Leclair *et al.*, 2004).

Identification of S-alleles

Fourteen S-alleles were identified in our study, with allele frequencies ranging from 1 to 18%. Possibly, there are many, albeit rather rare, S-alleles in apple that have not yet been detected. Broothaerts *et al.* (2004) report allelic configurations that include 3% unidentified alleles and Matsumoto *et al.* (2007) have 4% unidentified alleles. In our material, number of unidentified alleles is much higher, 19%. One reason is probably that two very popular cultivars in Scandinavia,

'Ingrid Marie' and 'James Grieve', each have one allele for which there is no DNA marker yet (Broothaerts *et al.*, 2004). In addition, S_{13} and S_{21} which are present in another two important cultivars in this geographic region, 'Gravensteiner' and 'Ribston', respectively, were not investigated in our analyses.

In comparison with previously published data for DNA-derived identification of S-allele composition in apple (Broothaerts *et al.*, 2004; Melounova *et al.*, 2005; Matsumoto *et al.*, 2007) some geographically based discrepancies in allele frequencies can be noted. In our study, S_5 and S_7 are more common and S_9 more rare than in the previous studies. One major reason for this difference is probably the occurrence of S_5 and S_7 in several cultivars that are well-known in the Scandinavian countries (e.g. 'Aroma', 'Astrakan red', 'Ingrid Marie', 'James Grieve' and 'Katja') and are likely to have some relatives in the analysed material. By contrast, S_9 occurs in e.g. 'Cox's Orange Pippin' and 'Red Delicious' which have had a large influence on the development of cultivars in North America, Central and South Europe and Japan, but probably have much fewer relatives in the Scandinavian countries with their harsher climate.

Genetic diversity among cultivars

A set of apple mandate cultivars had been appointed and approved for preservation already before the initiation of this study. We decided to analyse all those mandate cultivars that had a Swedish origin and were present in the Balsgård apple collections, including also those that were suspected to be sports only. The SSR-based values of expected heterozygosity (H_e), obtained in this study, are generally in good agreement with those published by Liebhard *et al.* (2002), Guarino *et al.* (2006), and Galli *et al.* (2005). When the less polymorphic loci in multilocus SSR markers were omitted (as in Guarino *et al.* 2006), the average H_e values were almost identical in all three studies based on the same markers: 0.80 (in this study), 0.81 in Guarino *et al.* (2006) and 0.80 in Liebhard *et al.* (2002), in spite of the different numbers of cultivars analysed. The average PIC value, which is basically the same as expected heterozygosity, was 0.72 in Galli *et al.* (2005), and 0.74 in the present study when all loci (also the less polymorphic loci in the bilocus systems) were included.

Number of SSR alleles per locus in our study was generally in agreement with those reported by Guarino *et al.* (2006) but in some cases we found considerably more allele variants. Thus we found 15 alleles for primer CH02c06, while Guarino *et al.* (2006) report only 10 alleles, and for primer CH04e05 we found 14 alleles, as compared to 8 in the Guarino *et al.* (2006) study. These primers amplified mainly unique or rare (frequency at or below 0.05) alleles, namely 10 out of 15 (67%) in the case of CH02c06 and 10 out of 14 (71%) in the case of CH04e05. Consequently, the higher number of analysed individual genotypes, 66 in this study compared to 27 in the study by Guarino *et al.* (2006), may explain the differences in allele number.

The expected heterozygosity (H_e) is influenced by the number of alleles and by the distribution of allele frequencies. Thus, the most polymorphic SSR locus in this study, CH02b10, (H_e = 0.88) revealed 15 rather evenly distributed alleles.

Interestingly, locus CH02c09 with only 10 evenly distributed alleles had the same heterozygosity as locus CH02c06 (H_e = 0.85), which amplified 15 alleles, 10 of which were found at a frequency below 0.05. Even though loci CH02d08 and CH04e05 amplified 13 and 14 alleles, respectively, they had one very common allele each (frequency approx. 0.50) and, consequently, lower expected heterozygosity values, 0.70 and 0.71 respectively. The pattern of allele distribution seems to have a stronger impact on expected heterozygosity values than the number of alleles does. Galli *et al.* (2005) pointed out the importance of an even distribution of alleles, and our study confirms this fact.

The large number of relatively abundant alleles in the S-locus makes this an important tool in identification of cultivars. In addition, valuable information about compatibility relationships is obtained. S-alleles have also been applied for paternity assignment in studies on fruit tree pollination and fruit set (Schneider *et al.*, 2001a, 2001b; 2005; Zisovich *et al.*, 2005). Still, the discriminatory power of S-allele analysis is much smaller than for e.g. SSR analysis (where more loci are available) and this analysis can therefore be regarded only as a complement to other methods.

Genetic relatedness among cultivars

For analysis of closer relationships, both SSR alleles and S-locus alleles are very useful. Sports have almost identical allele composition and can easily be distinguished from closely related (parent-offspring) cultivars derived by sexual recombination. For assessment of phenetic similarities among cultivars, high multiplex ratio markers, like AFLP or ISSR would probably produce better results than SSR (Goulao & Oliveira, 2001). In the case of highly heterozygous and diverse apple cultivars, it is, however, doubtful whether any other type of molecular marker would produce well-defined clusters.

The indigenous Swedish mandate cultivars probably represent a broad gene pool due to e.g. numerous cross-pollinations between highly heterozygous genotypes during the long period of apple cultivar development, frequent exchange of plant material between countries and plant breeding programs, and a slow turnover of popular cultivars, many of which are more than a hundred years old. Indigenous Swedish mandate cultivars may therefore originate from foreign progenitors, and represent several centuries of apple cultivation even in the most recent generations. Not surprisingly, the SSR-based UPGMA dendrogram failed to produce a clear group structure in our study, thus confirming the previous results obtained with RAPD markers (Garkava-Gustavsson & Nybom, 2007).

Still, the discrepancies in S-allele frequency distribution when Swedish cultivars are compared to mainly North American, Central and South European and Japanese cultivars (an over-representation of S_5 and S_7 in Sweden, and an under-representation of S_9) suggest that allele frequencies can change considerably due to genetic drift.

Other SSR-based studies in apple have also failed to produce a clear grouping structure except in a Spanish study of locally derived cider apple cultivars

(Pereira-Lorenzo *et al.*, 2007). In the analysis of 66 apple rootstock clones, two broad groups but no distinct clusters were thus observed (Oraguzie *et al.*, 2005). A genetic distance-based dendrogram, obtained in Guarino *et al.* (2006) in a study of local Italian apple cultivars, showed no relationships with the morphological traits. Furthermore, eight reference cultivars from other countries were spread across the whole dendrogram. This distribution was explained by the fact that all the plant material probably came from the same genetic pool of apple germplasm, collected, bred and exchanged across Europe and Central Asia over many centuries.

Acknowledgements – This work was supported financially by FORMAS (The Swedish Council for Environment, Agricultural Sciences and Spatial Planning).

References

Amos, W., Hoffman, J. I., Frodsham, A., Zhang, L., Best, S & Hill, V.S. 2007. Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* 7, 10–14.

Aranzana, M. J., Carbó, J. & Arus, P. 2003. Microsatellite variability in peach [*Prunus persica* (L.) Batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. *Theoretical and Applied Genetics* 106, 1341–1352.

Broothaerts, W. 2003. New findings in apple S-genotype analysis resolve previous confusion and request the re-numbering of some S-alleles. *Theoretical and Applied Genetics* 106, 703–714.

Broothaerts, W., Van Nerum, I. & Keulemans, J. 2004. Update on and review of the incompatibility (S-) genotypes of apple cultivars. *HortScience* 39, 943–947.

Bruland, O., Almqvist, E. W., Goldberg, Y. P., Boman, H., Hayden, M.R., & Knappskog, P.M. 1999. Accurate determination of the number of CAG repeats in the Huntington disease gene using a sequence-specific internal DNA standard. *Clinical Genetics* 55, 198–202.

Cryer, N. C., Fenn, M. G. E., Turnbull, C. J. & Wilkinson, M.J. 2006. Allelic size standards and reference genoypes to unify international cocoa (*Theobroma cacao* L.) microsatellite data. *Genetic Resources and Crop Evolution* 53, 1643–1652.

Galli, Z., Halász, G., Kiss, E., Heszky, L., & Dobránski, J. 2005. Molecular identification of commercial apple cultivars with microsatellite markers. *HortScience* 40, 1974–1977.

Garkava-Gustavsson, L. & Nybom, H. 2007. Genetic diversity in a collection of apple (*Malus* x *domestica* Borkh.) cultivars as revealed by RAPD markers. *International Journal of Horticultural Science* 13, 1–11.

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M & Gessler, C. 1998. Simple sequence repeats for the genetic analysis of apple. *Theoretical and Applied Genetics* 96, 1069–1076.

Goulao, L. & Oliveira, C. M. 2001. Molecular characterisation of cultivars of apple (*Malus* x *domestica* Borkh.) using microsatellite (SSR and ISSR) markers. *Euphytica* 122, 81–89.

Guarino, C., Santoro, S., De Simone, L., Lain, O., Cipriani, G. & Testolin, R. 2006. Genetic diversity in a collection of ancient cultivars of apple (*Malus x domestica* Borkh.) as revealed by SSR-based fingerprinting. *Journal of Horticultural Science and Biotechnology* 81, 39–44.

Guilford, P., Prakash, S., Zhu, J. M., Rikkerink, E, Gardiner, S., Bassett, H.& Forster, R. 1997. Microsatellites in *Malus* x *domestica* (apple): abundance, polymorphism and cultivar identification. *Theoretical and Applied Genetics* 94, 249–254.

Hjalmarsson, I. & Wallace, B. 2004. Content of the Swedish berry gene bank. *Journal of Fruit and Ornamental Plant Research* 12, 129–138.

Janssens, G. A., Goderis, I. J., Broekaert, W. F. & Broothaerts, W. 1995. A molecular method for S-allele identification in apple based on allele-specific PCR. *Theoretical and Applied Genetics* 91, 691–698.

Kenis, K. & Keulemans, J. 2005. Genetic linkage maps of two apple cultivars (*Malus* x *domestica* Borkh.) based on AFLP and microsatellite markers. *Molecular Breeding* 15, 205–219.

Kitahara, K. & Matsumoto, S. 2002. Cloning of S_{25} cDNA from 'McIntosh' apple and an S_{25}-allele identification method. *Journal of Horticultural Science and Biotechnology* 77, 724–728.

Kitahara, K., Matsumoto, S., Yamamoto, T., Soejima, J., Kimura, T., Komatsu, H. & Abe, K. 2005. Parent identification of eight apple cultivars by S-RNase analysis and simple sequence repeat markers. *HortScience* 40, 314-317.

Kloosterman, A. D., Budowle, B. & Daselaar, P. 1993. PCR-amplification and detection of the human D1S80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. *International Journal of Legal Medicine* 105, 257–264.

Koller, B., Lehmann, A., McDermott J. M. & Gessler, C. 1993. Identification of apple cultivars using RAPD markers. *Theoretical and Applied Genetics* 85, 901–904.

Laurens, F., Durel, C. E. & Lascostes, M. 2004. Molecular characterisation of French local apple cultivars using SSR. *Acta Horticulture* 663, 639–641.

Leclair, B., Fregeau, C. J., Bowen, K. L. & Fourney, R.M. 2004. Precision and accuracy in fluorescent short tandem repeat DNA typing: assessment of benefits impared by the use of allelic ladders with the AmpF/STR®ProfilerPlus™ kit. *Electrophoresis* 25, 790–796.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C.D., Tarchini, R., Van de Weg, E. & Gessler, C. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Molecular Breeding* 10, 214–241.

Matsumoto, S., Komori, S., Kitahara, K., Imazu, S. & Soejima, J. 1999. S-genotypes of 15 apple cultivars and self-compatibility of 'Megumi'. *Journal of the Japanese Society for Horticultural Science* 68, 236–241.

Matsumoto, S., Eguchi, T., Bessho, H. & Abe, K. 2007. Determination and confirmation of the S-RNase genotypes of apple pollinators and cultivars. *Journal of Horticultural Science and Biotechnology* 82, 323–329.

Mattisson, H. & Nybom, H. 2005. Application of DNA markers for detection of scab resistant apple cultivars and selections. *International Journal of Horticultural Science* 11, 59–63.

Melounova, M., Vejl, P., Sedlák, P., Blazek, J., Zoufala, J., Milec, Z. & Blazkova, H. 2005. Alleles controlling apple skin colour and incompatibility in new Czech apple varieties with different degrees of resistance against *Venturia inaequalis* CKE. *Plant Soil and Environment* 51, 65–73.

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

Nilsson, A. 1986. *Våra äpplesorter*. AB Allmänna Förlaget, Stockholm. (in Swedish).

Nybom, H. & Schaal, B. A. 1990. DNA "fingerprints" applied to paternity analysis in apples (*Malus x domestica*). *Theoretical and Applied Genetics* 79, 763–768.

Nybom, H., Rumpunen, K., Persson Hovmalm, H., Garkava-Gustavsson, L. & Olsson,

M. E. Towards a healthier apple - chemical characterisation of an apple gene bank. *Acta Horticulturae*, in press.

Oraquzie, N. C., Yamamoto, T., Soejima, J. *et al.* 2005. DNA fingerprinting of apple (*Malus* spp.) rootstocks using Simple Sequence Repeats. *Plant Breeding* 124, 197–202.

Pereira-Lorenzo, S., Ramos-Cabrer, A. M. & Diáz-Hernández, M. B. 2007. Evaluation of genetic identity and variation of local apple cultivars (*Malus* x *domestica* Borkh.) from Spain using microsatellite markers. *Genetic Resources and Crop Evolution* 54, 405–420.

Ramos-Cabrer, A. M., Diaz-Hernandez, M. B. & Pereira-Lorenzo, S. 2007. Morphology and microsatellites in Spanish apple collections. *Journal of Horticultural Science and Biotechnology* 82, 257–265.

Reif, J. C., Melchinger, A. E. & Frisch, M. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Science* 45, 1–7.

Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. – In: *Studies in genetics* VII. Publ. 7213. University of Texas, Austin, p. 145–153.

Rohlf, M. 2005. *NTSYSpc: numerical taxonomy and multivariate analysis system*. Version 2.2. State Univ. of New York.

Saunders, J. A., Mischke, S., Leamy, E. A. & Hemeida A. A. 2004. Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theoretical and Applied Genetics* 110, 41–47.

Schneider, D., Stern, R. A., Eisikowitch, D. & Goldway, M. 2001a. Analysis of S-alleles by PCR for determination of compatibility in the 'Red Delicious' apple orchard. *Journal of Horticultural Science and Biotechnology* 76, 596–600.

Schneider, D., Stern, R. A., Eisikowitch, D. & Goldway, M. 2001b. Determination of the self-fertilization potency of 'Golden Delicious' apple. *Journal of Horticultural Science and Biotechnology* 76, 259–263.

Schneider, D., Stern, R. A. & Goldway, M. 2005. A comparison between semi- and fully compatible apple pollinators grown under suboptimal pollination conditions. *HortScience* 40, 1280–1282.

Silfverberg-Dilworth, E., Matasci, C. L., Van de Weg, W. E., Van Kaauwen, M. P. W., Walser, M., Kodde, L. P., Soglio, V., Gianfranceschi, L., Durel, C. E., Costa, F., Yamamoto, T., Koller, B., Gessler, C. & Patocchi, A. 2006. Microsatellite markers spanning the apple (*Malus* x *domestica* Borkh.) genome. *Tree Genetics and Genomes* 2, 202–224.

Svensson, H. and Kastman, K. 2005. *Äpplen i Sverige*. Prisma (in Swedish).

This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., Crespan., M., Dangl, G. S., Eisenheld, C., Ferreira-Monteiro, F., Grando, S., Ibanez, J., Lacombe, T., Laucou, V., Magalhaes, R., Meredith, C. P., Milani, N., Peterlunger, E., Regner, F., Zulini, L. Maul., E. 2004. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theoretical and Applied Genetics* 109, 1448–1458.

Van Nerum, I., Geerts, M., Van Haute, A., Keulemans, J., Broothaerts, W. 2001. Re-examination of the self-incompatibility genotype of apple cultivars containing putative 'new' S-alleles. *Theoretical and Applied Genetics* 103, 584–591.

Weeden, N.F. & Lamb, R. C. 1985. Identification of apple cultivars by isozyme phenotypes. *Journal of American Society for Horticultural Science* 110, 509–515.

Weising, K., Nybom, H., Wolff, K. & Kahl, G. 2005. *DNA fingerprinting in plants: principles, methods and applications*. Taylor & Francis/CRC press.

Xu, M. L. & Korban, S. S. 2000. Saturation mapping of the apple scab resistance gene V_f using AFLP markers. *Theoretical and Applied Genetics* 101, 844–851.

Zisovich, A. H., Stern, R. A., Shafir, S. & Goldway, M. 2005. Fertilization efficiency of semi- and fully-compatible European pear (*Pyrus communis*) cultivars. *Journal of Horticultural Science and Biotechnology* 80, 143–146.

Table 1. *SSR and S-allele profiles of 68 native Swedish mandate cultivars, SSR alleles in bold apparently occur in duplicate while SSR alleles in bold and italics apparently occur in triplicate*
** Triploid cultivars, ** Tetraploid cultivar*

	Analysed cultivars	CH01h02 L1	CH01h02 L2	CH02c06	CH02c09	CH02c11 L2	CH02d08	CH04c06 L1	CH04c06 L2	CH04e05	COL	CH02b10	S-Locus
1	'Alfa 68'**	201:202:**203**	235:243:245:247	228:**238**:248	231:247:253:255	218:**230**:232	213:225:229:254	157:**170**	174:178:186:193	**174**:**209**	220:230:232?	117:119:121:129	S_2 S_3 S_5
2	'Alice'	**204**	235:247	228:252	231:255	206:222	**254**	**170**	174:178	174:222	**220**	129:142	S_3 S_5
3	'Annero'	**203**	**235**	216:252	243:255	216:234	**211**	**170**	174.191	**174**	**241**	119:129	S_5 S_7
4	'Aroma'	202:204	235:247	216:252	231:255	216:230	211:254	**170**	178:193	198:209	220:230	136:142	S_5 S_7
5	'Arvidsäpple'	201:203	**235**	228:252	241:243	206:216	211:254	**170**	174:191	**174**	**241**	119:129	S_2 S_7
6	'Aspa'	**203**	**235**	216:230	241:243	214:222	215:225	157:170	**174**	174:209	220:232	121:157	S_3 S_5
7	'Astrakan, Gyllenkrok's'	**201**	235:247	228:250	231:243	206:232	225:254	**170**	174:191	174:224	220.241	123:129	S_3 S_{22}
8	'Astrakan, Red'	**203**	**235**	216:252	241:255	216:236	211:225	**170**	180:191	**174**	234:241	142:157	S_3 S_7
9	'Astrakan, Stor Klar'	201:203	235:239	**252**	241:247	**216**	215:254	**170**	**191**	**174**	232:241	119:142	S_7 S_{22}
10	'Birgit Bonnier'	**203**	245:247	228:252	**231**	232:236	211:250	157:170	**178**	**201**	**230**	127:145	S_2 S_5
11	'Brunnsäpple, Halland'	**203**	243:247	248:250	231:251	**214**	211:217	157:170	174:183	205:209	220:230	136:157	S_3 $S_?$
12	'Domö Favorit'	**203**	235:239	**252**	241:247	214:216	**211**	**170**	174:191	174:203	**232**	**142**	S_1 S_5
13	'Drakenberg'	**203**	**247**	216:252	239:253	214:232	211:246	**170**	**186**	174:211	220:232	129:142	S_2 $S_?$
14	'Edsele'	**203**	239:247	216:252	243:247	214:216	211:215	157:170	**174**	**174**	232:241	129:142	S_1 S_2
15	'Eva-Lotta'	**201**	**247**	238:252	231:247	232:236	229:254	157:170	178:187	200:201	230:239	129:145	S_{25} $S_?$
16	'Fagerö'	201:204	**247**	216:242	239:243	**214**	**211**	157:170	178:186	**209**	220:228	**129**	S_2 $S_?$
17	'Farmors Juläpple'	**203**	**235**	252:263	239:255	230:236	225:246	**170**	176:180	**174**	239:241	127:142	S_3 S_7
18	'Flädie'	**203**	235:255	238:242	253:255	214:232	217:229	157:170	174:186	**174**	222:241	117:129	S_1 $S_?$
19	'Frösåker'*	201:**203**	**247**	**216**:252	231:**243**	214:222:232	**211**	**157**:170	174:176:178	**174:**201	218:232:241	**142**:157	S_1 S_5 $S_?$
20	'Fullerö'	**203**	235:239	238:252	241:243	**236**	**211**	**170**	174:183	**174**	230:232	129:142	S_5 $S_?$
21	'Förlovningsäpple'	**203**	235:255	206:252	247:253	216:230	211:217	157:170	167:191	203:222	232:241	111:142	S_{20} $S_?$
22	'Granatäpple, Kungsbacka'	N/n	**235**	250:252	237:253	226:228	217:254	**170**	174:183	198:201	220:232	131:142	S_3 S_5
23	'Grågylling'	201:204	**247**	216:242	239:243	**214**	**211**	157:170	178:186	**209**	220:241	**129**	S_2 S_{20}
24	'Göteborgs Flickäpple'	201:203	235:247	244:250	241:247	216:236	211:225	157:170	174:176	174:201	**232**	125:131	S_7 $S_?$

25	‘Hanaskog’	**203**	235:248	244:250	241:253	206:216	211:254	**170**	174:186	174:211	**241**	121:131	S_3 S_7
26	‘Hedenlunda’	201:203	235:247	**252**	231:241	214:216	244:254	**170**	174:186	174:201	232:241	119:123	S_{16} $S_?$
27	‘Himmelstalund’	**201**	**247**	228:242	231:255	206:232	246:254	**170**	**186**	174:201	232.239	121:131	S_3 S_4
28	‘Holländaräpple’*	202:203:204	**235**:247	**216**:250	231:239:253	214:222:236	**211**:254	157:**170**	178:186:191	**174**:222	**232**	129:131:157	S_1 S_2 S_7
29	‘Josefiner’	201:203	**247**	252:263	237:243	214:230	**211**	157:170	167:186	**174**	220:230	119:125	S_{28} $S_?$
30	‘Kalmar Glasäpple’*	201:202:204	235:243:247	**216**	241:243:255	230:234:236	**211**	**170**	176:178:191	174:209:222	**232**	142:155:157	S_2 S_3 S_{10}
31	‘Katja’	203:204	**247**	228:250	**231**	216:226	211:229	**170**	186:187	200:201	**230**	127:131	S_5 S_{24}
32	‘Kavlås’	201:203	239:247	216:228	239:241	214:216	211:225	157:170	178:191	205:209	232:241	127:129	S_{28} $S_?$
33	‘Kim’	**203**	**247**	238:252	231:255	216:232	229:254	157:170	174:187	174:201	230:239	117:145	S_5 S_{25}
34	‘Kinnekulle Kantäpple’*	**203**	**235**:247	228:248:250	**255**	214:216:236	211:217:254	**170**	167:176:178	174:207:209	**232**	127:129:131	S_1 S_4 S_{10}
35	‘Landskronaäpple’	**202**	235:243	228:252	231:243	206:226	211:217	**170**	174:178	**174**	204:232	121:142	S_5 S_7
36	‘Linnaeus’ Apple’	201:203	235:247	216:250	231:249	216:232	**211**	157:170	**186**	174:201	218:241	131:136	S_3 $S_?$
37	‘Melonkalvill’	202:204	235:247	242:252	253:255	214:236	**211**	**170**	178:187	**174**	232:241	131:142	S_2 S_7
38	‘Menigasker’	**201**	245:247	216:232	231:239	214:216	211:225	157:170	167:186	200:209	220:230	129:157	S_5 S_9
39	‘Mio’	**203**	235:247	228:252	231:241	218:226	211:250	**170**	174:178	174:201	**222**	119:127	S_2 S_7
40	‘Mälsåker’	201:203	235:247	250:252	231:253	228:236	225:254	157:170	174:178	174:201	**220**	127:131	S_3 S_5
41	‘Norrstack’*	201:**203**	235:**247**	**216**:228	**243**:253	214:216:222	205:211:217	**170**	174:178:186	**209**	232:**241**	127:129:157	S_4 $S_?$ $S_?$
42	‘Norrviken’	**203**	235:247	244:252	**253**	222:236	211:254	**170**	**174**	**174**	220:230	125:142	S_5 $S_?$
43	‘Oranie’	**203**	235:248	244:252	241:253	206:218	211:254	**170**	174:186	174:211	222:241	119:121	S_3 S_7
44	‘Oretorp’	**203**	235:247	228:248	231:243	**206**	211:248	**170**	176:178	174:209	**239**	129:131	S_1 S_3
45	‘Ringstad’	201:203	235:247	**216**	**243**	214:236	211:225	**170**	167:178	174:209	232:241	129:157	S_1 S_{20}
46	‘Risäter’	**203**	235:247	230:250	**253**	216:234	215:225	157:170	174:186	174:201	204:241	129:131	S_3 S_7
47	‘Rödluvan’	201:203	235:243	**252**	237:255	216:226	211:254	**170**	174:183	**222**	230:239	119:142	S_7 $S_?$
48	‘Sandbergs Röda’	**203**	239:247	242:250	241:243	214:216	**211**	157:170	186:191	203:209	232:241	129:131	S_1 S_2

49	'Silva'	**203**	**235**	216:250	**243**	214:216	215:217	157:170	**174**	182:203	230:241	131:157	S_1 $S_?$
50	'Snövit'	**203**	235:247	216:222	241:255	212:214	211:225	157:170	167:191	174:222	204:232	127:136	S_3 $S_?$
51	'Sparreholm'	**201**	235:247	252:263	241:243	206:214	211:254	**170**	174:191	**174**	232:241	119:125	S_{22} $S_?$
52	'Spässerud'	**203**	235:239	242.250	**247**	214:216	**211**	157:170	**174**	174:209	220:241	127:131	S_1 S_7
53	'Stenkyrke'	201:204	239:247	230:252	243:255	**216**	211:225	**170**	178:183	174:209	230:232	117:142	S_2 S_5
54	'Stäringe Karin'	201:204	235:247	216:252	239:243	206:214	**211**	157:170	174:186	174:209	**241**	119:129	S_2 S_{22}
55	'Sylvia'	**203**	**247**	228:230	**231**	222:232	250:254	**170**	178:191	**174**	**220**	129:136	S_3 S_{24}
56	'Särsö'	**203**	235:239	242:250	**247**	214:216	**211**	157:170	**174**	174:209	220:241	127:131	S_1 S_7
57	'Sävstaholm'	**203**	235:239	197:228	231:241	216:222	211:254	**170**	174:191	174:201	222:241	121:129	S_1 $S_?$
58	'Trogsta'	201:204	235:247	250:252	241:255	206:232	211:225	**170**	183:191	**174**	239:241	119:157	S_{22} S_{28}
59	'Vallda'	**203**	247:255	244:252	241:253	214:216	225:246	157:170	167:174	**174**	220:232	119:125	S_7 $S_?$
60	'Veseäpple'	**203**	243:247	230:250	231:239	206:226	211:254	**170**	178:195	209:224	220:230	123:157	S_5 S_{20}
61	'Villands Glasäpple'*	201:202:203	**235**:247	216:242:252	**247**:251	216:**236**	**211**	157:**170**	176:178:183	**174**:201	**232**	127:136:142	S_5 $S_?$
62	'Vittsjö'	**201**	235:247	228:256	243:253	206.222	211:246	**170**	186:191	174:207	239:241	121.129	S_3 $S_?$
63	'Vrams Järnäpple'*	201:**203**	235:243:248	238:242:252	**231**	214:218:236	**211**:225	**170**	**174**:178	174:209:219	**222**:239	**121**:129	S_{20} $S_?$ $S_?$
64	'Värmlands Sötäpple'	201:203	**235**	228:230	241:243	206:230	211:252	**170**	167:191	174:198	220:241	123:129	S_4 S_{20}
65	'Åkerö'	201:203	239:247	242:250	241:243	214:216	**211**	157:170	178:191	174:209	**241**	129:131	S_1 $S_?$
66	'Åkerö, Gripsholm'	201:203	239:247	242:250	241:243	214:216	**211**	157:170	178:191	174:209	**241**	129:131	-
67	'Ökna Lökäpple'	**203**	235:245	234:250	247:255	228:236	211:246	**170**	178:186	174:217	218:239	121:133	S_3 S_{16}
68	'Ökna Vita Vintergylling'	201:203	**235**	250:252	241:255	228:232	211:246	**170**	174:186	174:222	232:239	121:142	S_2 S_3 S_{10}
	Reference cultivars:												
	'Boskoop '	201:**203**	243:245:247	228:**238**	231:247:253	218:230:232	213:225:229	157:**170**	174:178:193	**174**:209	230/232?	117:119:121	
	'Cox's Orange'	203:204	243:247	**238**	231:255	**216**	**254**	**170**	**186**	174:200	**230**	117:129	
	'Discovery'	**203**	245:247	228:244	**255**	222:226	229:250	**170**	167:178	198:201	**239**	**127**	

'Golden Delicious'	**204**	247:249	234:238	241:255	218:232	223:225	**170**	174:178	**174**	220:232	119:123	
'James Grieve'	201:204	**247**	238:250	231:247	216:236	229:254	**170**	176:187	174:200	**230**	129:131	
'Prima'	201:204	235:245	234:238	231:241	228:232	**254**	157:170	174:178	174:209	220:234	119:123	
'Summerred'	**203**	235:247	228/234	**255**	214:218	211:229	**170**	174:178	174:222	**232**	123.127	
Known parents:												
'Filippa'	**202**	**235**	216:248	231:255	218:230	211:254	**170**	186:193	**209**	220:241	129:136	S_7 S_{24}
'Ingrid Marie'	**204**	**247**	238:252	**255**	216:222	**254**	**170**	178:187	174:198	**230**	117:142	S_5 $S_?$
'Cortland'	**203**	245:247	216:252	231:255	226:232	211:229	157:170	174:178	**201**	230:239	145:157	S_5 S_{25}
'Worcester Pearmain'	**203**	**247**	228:230	**231**	222:226	211:250	**170**	178:186	174:201	**230**	127:136	S_2 S_{24}

Table 2. *Genetic diversity estimators of 11 polymorphic SSR loci.*

SSR locus	Size range	Number of alleles (A)	Effective number of alleles (A_e)	Power of discrimi-nation (PD)	Number of geno-types	Expected heterozygo-sity (H_e)	Observed heterozy-gosity (H_o)
CH01h02, L1	201–204	4	2.38	0.76	14	0.58	0.42
CH01h02, L2	235–255	7	3.02	0.87	18	0.67	0.72
CH02c06	197–263	15	6.59	0.96	36	0.85	0.91
CH02c09	231–255	10	6.83	0.96	37	0.85	0.85
CH02c11, L2	206–236	12	7.03	0.95	38	0.86	0.90
CH02d08	205–254	13	3.34	0.91	28	0.70	0.74
CH04c06, L1	157, 170	2	1.56	0.60	7	0.36	0.44
CH04c06, L2	167–195	11	5.73	0.95	34	0.83	0.84
CH04e05	174–224	14	3.47	0.90	29	0.71	0.69
COL	204–241	10	5.73	0.94	27	0.83	0.71
CH02b10	111–157	15	8.50	0.97	46	0.88	0.96
Mean value		10.27	4.93	0.89	28.50	0.74	0.74

Table 3. *The five different PCR amplification programmes used for identification of S-alleles. The denaturation-annealing-elongation parts in all three programmes were run for 30 cycles.*

		denaturation	annealing	elongation		
PCR standard	3 min, 94º	15 s, 94º	15 s, 60º	30 s, 72º	2 min, 72º	Cooling 4º
PCR S_3-specific	3 min, 94º	15 s, 94º	15 s, 57º	30 s, 72º	2 min, 72º	Cooling 4º
PCR S_{20}-specific	3 min, 94º	15 s, 94º	15 s, 60º	45 s, 72º	2 min, 72º	Cooling 4º
PCR S_5-specific	3 min, 94º	1 min, 94º	1 min, 58º	1 min, 72º	10 min, 72º	Cooling 4º
PCR S_{25}-specific	3 min, 94º	1 min, 94º	1 min, 60º	1 min, 72º	10 min, 72º	Cooling 4º

Table 4. *Comparison of allele data for reference cultivars obtained in the present study and in four previously published studies. Major discrepancies in allele compositions are underlined. 'B.p. dif.' indicate differences in size between fragments amplified in this study and in previous studies*

		Prima		Discovery			Golden Delicious		Boskoop		
		This study	Liebhard *et al*., 2002	This sudy	Liebhard *et al*. 2002	b.p. dif	This study	Guarino *et al*., 2006	This study	Guarino *et al*., 2006	
2	Ch01H02,1	201/204	N/d	203	N/d		204	206	201/203	203/205	-2
2	CH01H02,2	235/245	236/246	245/247	246/248	-1	247/249	250/252	243/245/247	245/250	-2/-3
3	Ch02C06	234/238	236/240	228/244	230/246	-2	234/238	238/242	228/238	232/242/252	-4
4	Ch02C09	231/241	233/243	255	245/257	-2	241/255	245/259	231/247/253	234/251/257	-3/-4
5	Ch02C11,1	194	N/d	194	N/d		194	197	194	197	-3
5	Ch02C11,2	228/232	231/235	222/226	225/229	-3	218/232	221/236	218/230/232	221/234/236	-3/-4
6	Ch02D08	254/254	254/254	229/250	228/250	0/+1	223/225	225/227	213/225/229	216/227/231	-2/-3
7	Ch04C06,1	157/170	155/170	170/170	170/170	0/+2	170/170	171/171	157/170	157/171	-1/0
7	Ch04C06,2	174/178	174/178	167/178	166/178	0/+1	174/178	175/179	174/178/193	175/179/193	-1/0
8	Ch04E05	174/209	174/181/209	198/201	197/201	+1/0	174/174	175/175	174/209	175/210	-1
9	COL	220/234	222/232	239	230/240	-1/+2	220/232	220/232	230/232	230/232	0
10	Ch02b10	119/123	123/127	127/127	131/131	-4	119/123	N/d	1197123	N/d	
		James Grieve		Cox's Orange			Golden Delicious		Boskoop		
		This study	Gianfran-ceschi *et al*., 1998	This study	Gianfran-cheschi *et al*., 1998		This study	Gianfran-cheschi *et al*., 1998	This study	Gianfrancheschi *et al*., 1998	
2	CH01H02,2	247/247	248/248	243/247	244/248	-1	247/249	248/250	243/245/247	243/247	-1/0
3	Ch02C06	238/250	240/252	238/238	240/240	-2	234/238	236/240	228/238	230/240/250	-2
9	COL	230/230	229/229	230/230	217/231	0/+1	220/232	217/231	230/232	229/231	+1/+3
		Prima		Summer Red			Golden Delicious		Boskoop (Belle de Boskoop)		
		This study	Ramos-Cabrer *et al*., 2007	This study	Ramos-Cabrer *et al*., 2007		This study	Ramos-Cabrer *et al*., 2007	This study	Ramos-Cabrer *et al*., 2007	
6	Ch02D08	254/254	254/254	211/229	212/228	0/+1	223/225	222/224	213/225/229	214/224/228	+1
8	Ch04E05	174/209	174/209	174/222	174/223	0/-1	174/174	174/174	174/209	174/209	0
9	COL	220/234	222/232	232/232	232/232	-2/0	220/232	222/232	230/232	230/232	-2/0

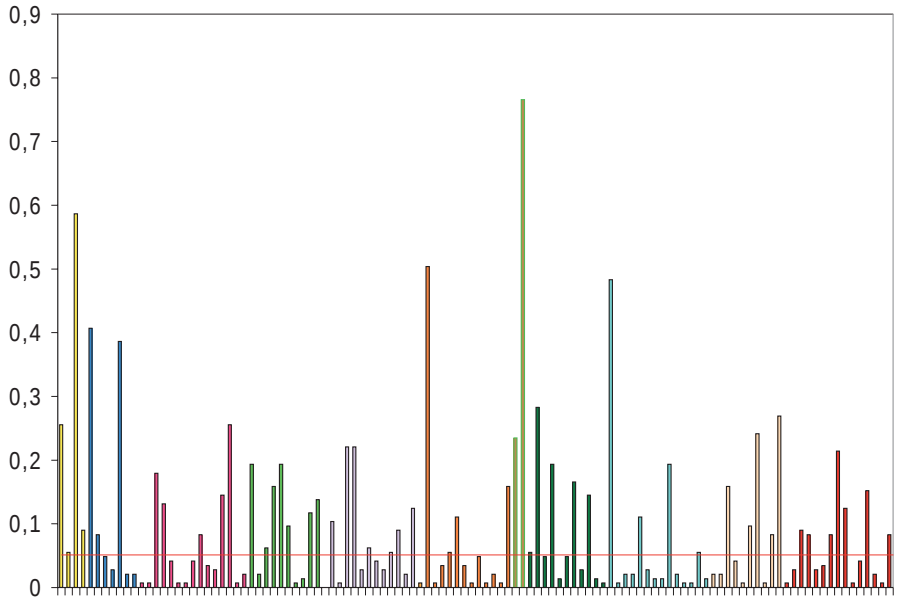


Figure 1. Allele frequencies for 11 analysed SSR loci, each represented by a different colour. Loci follow the same order as in Table 1.

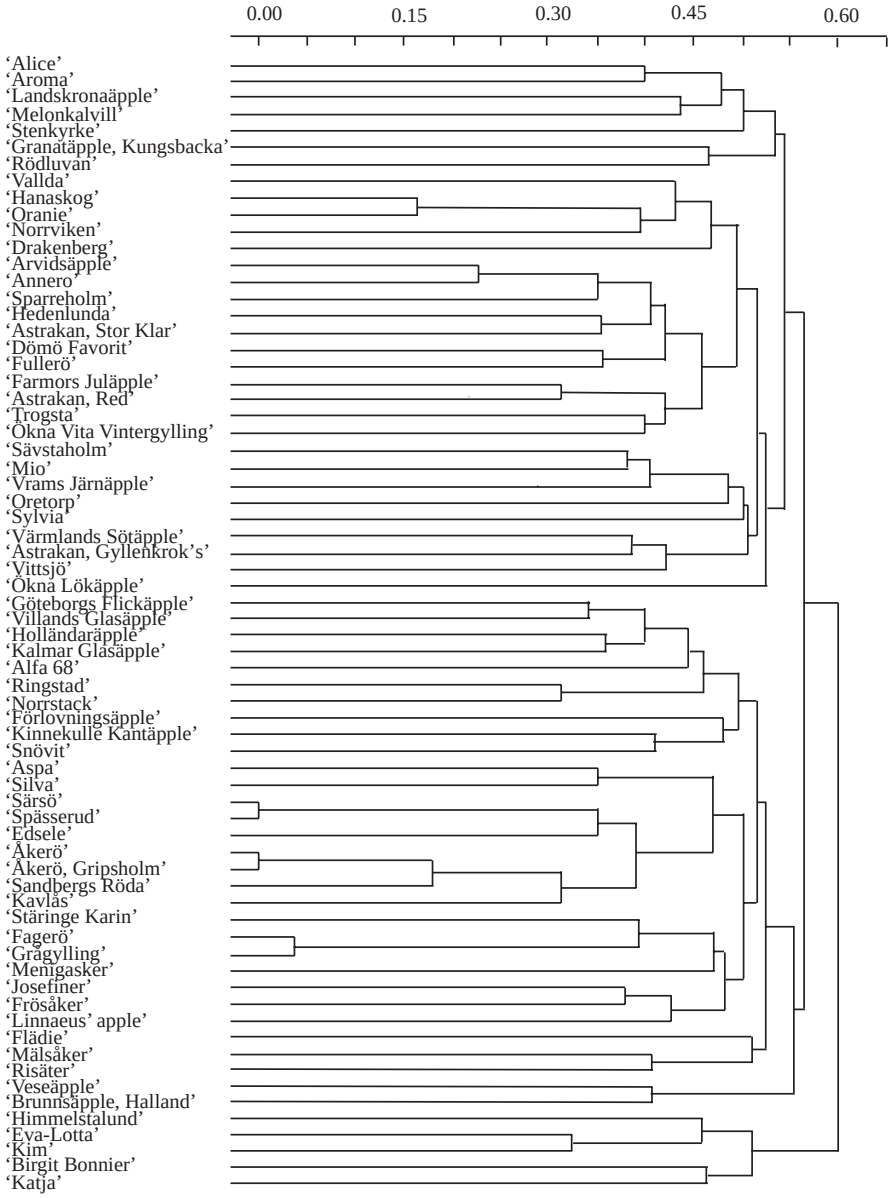


Figure 2. A UPGMA dendrogram, representing genetic relationships among 68 native Swedish mandate cultivars.