Estimating Genetic Variability in Horticultural Crop Species at Different Stages of Domestication

Helena Persson
Department of Crop Science
Alnarp

Doctoral thesis
Swedish University of Agricultural Sciences
Alnarp 2001

Acta Universitatis Agriculturae Sueciae
Agraria 289

ISSN 1401-6249
ISBN 91-576-5838-2

Abstract

Persson, H. 2001. *Estimating genetic variability in horticultural crop species at different stages of domestication.* Doctoral thesis
ISSN 1401-6249, ISBN 91-576-5838-2.

Domestication may be viewed as an evolutionary process, involving mechanisms like mutation, selection, genetic drift, hybridization and polyploidization, and in the end resulting in individuals with traits profitable for man. The advent of modern plant breeding has accelerated the domestication of plants considerably. Plant breeding is essentially a selection of plant material based on the existence of genetic variation. Genetic variation within species has been assessed by many methods and from several perspectives. In the present thesis, I study some genetical aspects in five crop species at different stages of domestication, using RAPD and morphological characters. The very first step in the domestication process involves selection of plant material in nature. Often, only a small amount of the variation present in the source material is represented in the samples taken. Domesticated populations of Turk's-cap lily proved, however, to contain as high levels of genetic variation as native populations. A totally different pattern was found in black chokeberry, as no variation at all could be found in cultivated material. In such a case, it is of vital importance to broaden the genetic basis within the crop, and this may be accomplished by incorporation of new selections from nature. In order to optimize collection strategies, information must be aquired about genetic structure in these populations. Native plant material of black chokeberry turned out to contain substantial amounts of variation, however tetraploid and presumably apomictic plants produced progeny groups with much less variability than progeny groups derived from diploid plants. In native populations of lingonberry, individual clones extended at least 30m, which gives an indication of how to collect plants for maximizing the genetic variation within the material. In crops where valuable cultivars already have been developed, tools for simple and fast identification of these cultivars are needed. In the present thesis, both RAPD markers and morphology were successfully used to identify twelve cultivars of rhubarb. Moreover, to facilitate the breeding work, molecular markers linked to traits of interest are highly desirable. In this thesis, RAPD was used to search for a marker linked to sex determination in sea buckthorn.

Key words: *Aronia melanocarpa*, *Hippophae rhamnoides*, *Lilium martagon*, *Rheum*, *Vaccinium vitis-idaea*, molecular markers, genetic diversity

Author's address: Helena Persson, Department of Crop Science, The Swedish University of Agricultural Sciences, Fjälkestadsvägen 123-1, SE-291 94 Kristianstad, Sweden.

Contents

Introduction 9

Plant material - biology and utilisation 10
Turk's-cap lily, *Lilium martagon* L. 10
Black chokeberry, *Aronia melanocarpa* (Michx.) Ell. 10
Lingonberry, *Vaccinium vitis-idaea* L. 11
Rhubarb, *Rheum* spp. 12
Sea buckthorn, *Hippophae rhamnoides* L. 13

Methods and statistical measures for assessing genetic variation 14
Morphology 14
Molecular markers 15
Correlation between morphological and molecular methods 16
Statistics 16
Genetic relationships 16
Genetic and genotypic diversity 17
Partitioning of variation 17
Segregation and trait specific markers 18

Plant introduction 19
Turk's-cap lily, an ornamental containing substantial genetic variation 19
Pronounced founder effect in black chokeberry 20

Genetic enhancement 21
Continuation of the black chokeberry case 21
Genetic structure in vegetatively propagating lingonberry populations 24
Leaf shape analysis in RAPD-defined clones of lingonberry 26

Identification of genotypes 28
Discrimination among culinary rhubarb cultivars 28

Trait specific markers 30
The search for a sex-specific marker in sea buckthorn 30

Concluding remarks 32

Acknowledgements 32

References 33

Appendix

Papers I – V

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

I. Persson HA, Lundquist K and Nybom H (1998) RAPD analysis of genetic variation within and among populations of Turk's-cap lily *(Lilium martagon* L.). Hereditas 128: 213–220.

II. Persson HA and Nybom H (1998) Genetic sex determination and RAPD marker segregation in the dioecious species sea buckthorn *(Hippophae rhamnoides* L.). Hereditas 129: 45–51.

III. Persson HA, Rumpunen K and Möllerstedt LK (2000) Identification of culinary rhubarb (*Rheum* spp.) cultivars using morphological characterization and RAPD markers. Journal of Horticultural Science & Biotechnology 75: 684–689.

IV. Persson HA and Gustavsson BA (2001) The extent of clonality and genetic diversity in lingonberry (*Vaccinium vitis-idaea* L.) revealed by RAPDs and leaf-shape analysis. Molecular Ecology 10: 1385–1397.

V. Persson HA, Jeppsson N, Bartish IV and Nybom H. RAPD-based estimates of genetic variation reveal different reproductive strategies in black chokeberry [*Aronia melanocarpa* (Michx.) Ell.]. (Manuscript)

Introduction

Domestication and natural evolution show similarities as well as differences (van Raamsdonk 1993). The most important similarities are mechanisms such as mutation, selection, genetic drift, hybridization and polyploidization. The major difference is the objective of the process. Domestication results in individuals that have characters profitable for man, but that generally have a reduced fitness in nature. These individuals are consequently dependent on man for survival. By contrast, evolution is a natural change through time to reach better adaptation or fitness for a group of individuals under local circumstances (van Raamsdonk 1993).

The domestication process may thus be viewed as an evolutionary process, however more rapid than the slow evolution in plants under natural selection. The advent of modern plant breeding, with its greater understanding of the genetic systems that govern genetic variability, has accelerated this process even more. Plant breeding is essentially a selection of plant material based on the existence of genetic variability. Traditionally, plant breeding aims at improving morphological and physiological traits of the crop, but lately, traits like quality and nutritional content, and amenability for mechanical handling have become increasingly important (Sánchez-Monge 1993).

Since the domestication process shows many of the same mechanisms as natural evolution, it is possible to use methods developed for evaluation of genetic variation in natural populations also for domesticated plant material and vice versa. A combination of approaches and methods, originally developed for different research areas, may therefore lead to a considerable improvement of the results obtained.

The genetic structure of crop gene pools has been assessed by many methods and from several perspectives. Studies have been conducted on all types of plant material, from native populations to genotypes within collections frequently used in plant breeding. The aim of my thesis is to demonstrate how morphological and molecular measures of genetic variability and relationships may be applied in a selection of horticultural crop species, which are at different stages of domestication.

Plant material - biology and utilisation

Turk's-cap lily, *Lilium martagon* L.

Turk's-cap lily *(Lilium martagon* L.) is the most widely distributed of all lilies and the most common lily species in Europe (Woodcock & Stearn 1950). It is a self-incompatible diploid, 2n=24 (Lundquist 1991), which propagates predominantly by seed-set and only very rarely by bulbils (Tillge 1967). Turk's-cap lily is insect-pollinated (Lundquist 1991) and at the end of the summer, many light-weight seeds are produced and dispersed by wind (Tillge 1967). The species grows naturally up to about 2300m altitude, in beech forests and wood edges, and on the borders of pasture-lands. It can be very variable in morphological characters like colour of flowers, the number of dots on the tepals and the pubescence of flower buds, also within a single population (Feldmaier & McRae 1982).

Turk's-cap lily comprises both native and domesticated populations. The native populations are distributed in Portugal, Spain, Central and Eastern Europe, European Russia and the Caucasus. The species then extends into northern Asia Minor and across Asiatic Russia to Siberia. Domesticated and naturalized populations in ornamental parks and forests can be found in e.g. Great Britain, Belgium and Scandinavia (Woodcock & Stearn 1950). The early history of the species in Sweden and the other Scandinavian countries is not well known. It has been suggested that Turk's-cap lily was introduced already in the Middle Ages (Karling 1931, Eriksson 1969). The geographic origin of the Scandinavian populations has never been documented. Many of the Scandinavian populations are restricted in size and geographically isolated from one another. The largest population in Sweden (at the estate of Ulfåsa), and perhaps in Scandinavia, has however more than 25,000 individuals.

Black chokeberry, *Aronia melanocarpa* (Michx.) Ell.

The genus *Aronia* belongs to Rosaceae and comprises two morphologically and ecologically very similar species: *A. arbutifolia* (L.) Ell. (red chokeberry, 2n=34, 68; Darlington & Janaki 1945) and *A. melanocarpa* (Michx.) Ell. (black chokeberry, 2n=34; Darlington & Janaki 1945). The two species, which are discriminated mainly by fruit colour (red vs. black berries; Hardin 1973), occur partly sympatrically in eastern North America, and are believed to hybridise and give rise to intermediate forms. There have been reports also of a third species, *A. prunifolia* (Marsh.) Rehder. According to Hardin (1973), this species is probably a hybrid between the other two species and he suggested *A. prunifolia* (with purple black berries) to be included into black chokeberry since they are extremely difficult to tell apart.

Black chokeberry grows in wet habitats like bogs as well as in dry areas like dunes and rocky slopes (Hardin 1973). Individual plants are capable of vegetative spreading through root suckers. The species flowers in late spring or early summer when grown in southern Sweden (Jeppsson 2000, Jeppsson & Johansson 2000), and pollination is carried out by insects, mainly small bees (Hardin 1973). Experiments have shown that *Aronia* species are self-compatible (Hardin 1973), thus probably having a mixed breeding system where both outbreeding and selfing occurs. Moreover, the presumed hybrid *A. prunifolia* has shown tendencies to be apomictic (Hardin 1973). Apomictic seed set has been reported also in Russian cultivated black chokeberry material (Poplavskaya 1995).

Black chokeberry was introduced into Russia in the 19th century with seeds from Germany and was originally intended for berry production in home gardens. Since the 1940s, it has been grown as a commercial fruit crop in Siberia (Kask 1987), and the berries are mainly used for juice and wine production. In the 1980's, a project was initiated at SLU Balsgård concerning commercial cultivation of black chokeberry in Sweden for the production of natural food colourants. There is, however, very limited variation in seedlings derived from cultivated Russian material (Skvortsov & Maitulina 1982, Jeppsson 1999a). In Russia, black chokeberry orchards are often established from seedlings (Kask 1987), which indicates that seed propagation results in more or less homogeneous progenies.

Lingonberry, *Vaccinium vitis-idaea* L.

Lingonberry (*Vaccinium vitis-idaea* L.; 2n=24) is a perennial, evergreen dwarf shrub, which belongs to the Ericaceae family (Anderberg *et al.* 1997). There are about 450 species of *Vaccinium*, of which only a few are found in Sweden: lingonberry, bilberry (*V. myrtillus* L.), cranberry (*V. oxycoccus* L.), small cranberry [*V. microcarpum* (Turczaninow ex Ruprecht) Schmalhausen], and bog whortleberry (*V. uliginosum* L.) (Anderberg *et al.* 1997). Lingonberry prefers acid soils, and has a wide distribution in northern temperate, boreal and subarctic areas. In Sweden, the species is very common, growing mainly in bogs and coniferous forests (Anderberg *et al.* 1997). Lingonberry reproduces sexually by seeds or vegetatively through rhizomes (Teär 1972). However, Eriksson (1989) lists lingonberry among those species in which no recruitment from seeds has been observed in established populations. Lingonberry is pollinated by honeybees and bumblebees, and functions mainly as an out-crosser. The species is at least partially self-compatible, but shows poor capacity to self in the absence of pollinators (Jacquemart & Thompson 1996).

The berries are used for industrial production of e.g. jam and juice, and as aroma additives in liqueur and yoghurt, and the species is of large economical importance in Sweden. Changes in forest management, variable fruit quality from native stands, and fluctuations in annual yield have stimulated initiatives to domesticate and cultivate the species (Gustavsson 1999). At SLU Balsgård, lingonberry breeding has

taken place since the 1970's and has resulted in a number of cultivars suitable for commercial cultivation as well as for small scale cultivation in home gardens (Gustavsson 1999).

Rhubarb, *Rheum* spp.

Rheum, with the common name rhubarb, is a genus of perennials belonging to the Polygonaceae family. About 50 species have been described (Flora of China 1996) and most of the species are native to the northern and central regions of Asia (Turner 1938). Chromosome numbers differ between taxa within the genus, the most common being 2n=22 (diploids), 2n=44 (tetraploids) and 2n=66 (hexaploids). Culinary rhubarb cultivars are tetraploids, and exhibit a continuous variation in morphology between *R. rhaponticum* (2n=22) on the one hand, and *R. rhabarbarum* (2n=44, though some cases of 2n=22 have been reported) on the other hand (Englund 1983).

For thousands of years, *Rheum* has been cultivated in China for medical purposes, and the earliest record is in a herbal from about 2700 B.C. It has been suggested that this medicinal species was *R. officinale* or *R. undulatum*, but common belief holds that *R. palmatum* is the genuine medicinal rhubarb (Turner 1938). The use of rhubarb root eventually reached Europe via the Arabs and in the beginning of the Christian era, the drug was used in Greek and Roman medicine. The drug had to be imported from Asia until European rhubarb cultivation started in the 16th century. The first *Rheum* species grown in Europe was probably *R. rhaponticum* (Turner 1938). In the middle of the 17th century, rhubarb was grown for medical purposes in England and Germany and eventually, the cultivation spread to the Scandinavian countries. However, the imported roots from Asia yielded a drug with much higher quality and European rhubarb cultivation dwindled (Hintze 1951). In the 18th century, it was discovered that rhubarb had edible stalks and *R. rhaponticum, R. undulatum* (=*R. rhabarbarum*; Englund 1983) and *R. hybridum* were grown as a vegetable crop in England (Turner 1938), and soon after, also in the Scandinavian countries (Hintze 1951). At an early stage, the growers in England started to develop culinary rhubarbs and some of the cultivars are still grown, e.g. 'Victoria' and 'Prince Albert'. In the breeding of new cultivars, *R. hybridum*, which has an unknown origin, was often used as a parent. This, in combination with haphazard crossings and seldom recorded parentages, makes it almost impossible to determine the proper origin of the vegetable-type rhubarb of today (Turner 1938). In order to preserve the cultivar characteristics, rhubarb must be vegetatively propagated. This is achieved by dividing field-grown plants or by micropropagation. Cultivar identification of culinary rhubarb is based mainly on morphological traits.

In culinary rhubarb, the high oxalate content is a major drawback. The oxalate ions form stable complexes with calcium ions, which may result in pathological conditions in man and animals. There is, however, a considerable variation in oxalate content among cultivars (Libert 1987). A renewed interest in rhubarb production is

now directed towards the use of stalks from low-oxalate cultivars as a cheap filler for industrial production of marmalade, jam and syrup. Rhubarb is also well suited for organic production (Rumpunen & Pettersson 1997).

Sea buckthorn, *Hippophae rhamnoides* L.

Sea buckthorn (*Hippophae rhamnoides* L., 2n=24) is a dioecious, wind-pollinated and very polymorphic shrubby species growing mainly on sand dunes or on gravel and sand banks (Rousi 1971). It has nitrogen-fixing capacity, and can withstand drought as well as temporary flooding. The species belongs to the genus *Hippophae*, which is widely distributed on the Eurasian continent. The two main wild subspecies in Europe are subsp. *rhamnoides* (referred to as *maritima* by van Soest), which grows along the Atlantic, North Sea and Baltic coasts of Northern Europe, and subsp. *fluviatilis* van Soest, which grows mostly on river banks in the Alps and adjacent mountainous areas (Rousi 1971). The species reproduces vegetatively with root suckers and sexually with bird-dispersed seeds.

Sea buckthorn was mentioned as a medicinal plant already in the traditional Tibetan pharmacopoeia, completed 618-907 A.D. In many parts of the world, sea buckthorn has been planted to stabilize the soil and to act as a windbreak. It has also become rather popular as a garden ornamental due to its silvery green foliage and attractive, orange-coloured berries. In the former USSR, commercial cultivation has been undertaken for the production of juices, jams and medical compounds since the berries are very rich in carotenoids and vitamin C (Rousi 1971). Plant breeding programs are conducted mainly in China, Russia, Finland, Germany and Sweden (Li & Schroeder 1996). The breeding project at SLU Balsgård aims at introducing sea buckthorn as a new commercial berry crop for industrial use, and has so far resulted in the two cultivars 'Romeo' and 'Julia'. Due to the distinctive flavour, the berries may be used in products like juice and jam and as aroma additive in e.g. ice cream. Since the berries also contain high levels of bioactive compounds, they may be used in the production of functional food (Jeppsson 1999b).

Methods and statistical measures for assessing genetic variation

Morphology

In many situations, the most easily obtained assessment of genetic variation is that of measuring morphological or phenotypic variation. The sharing of phenotypic characters is interpreted as an indication of relatedness. Morphological traits are, however, often influenced by environmental conditions (e.g. Jasienski 1997, Kercher & Sytsma 2000), which in turn may influence the estimation of genetic variation and relatedness. Consequently, to be really useful, morphological measurements should be accomplished on plant material that is grown in comparative trials. This may be both expensive and time consuming, and moreover, almost impossible to accomplish for some species that are very difficult to grow. However, if morphological characters are shown to be heritable, they will nevertheless reflect the genetic structure within the plant material.

When studying a plant species, there are several sets of phenotypic characters that may be used for discrimination and relatedness. Horticulturally important traits are valuable for a species in cultivation, and form the basis for the breeders' selection of promising plant material. Other morphological traits are used mainly for identification of genotypes and cultivars, e.g. the UPOV (International Union for the Protection of New Varieties of Plants) guide-lines for evaluation of distinctness, homogeneity and stability. A third type of traits are used for evaluation of the genetic variation within a species, e.g leaf shape analysis.

The use of automated image analysis of shape has several advantages over the scoring of morphological characters by hand. It allows rapid and cost-effective scoring of variation, in that a large number of plants can be screened easily. In addition, the process is separated into two phases (image aquisition and shape description): once the outlines have been stored, different approaches to shape description can be used with the possibility to choose the descriptor system that is best suited to a particular problem (White *et al.* 1988). Different descriptor suites are moment invariants (Dudani *et al.* 1977, White *et al.* 1988) and elliptic Fourier coefficients (Kuhl & Giardina 1982, Kincaid & Schneider 1983, McLellan & Endler 1998). The moment invariants descriptor system describes leaf shape by quantities that measure the distribution of the Cartesian coordinates (x and y) of image points along the outline of the leaf. In contrast, the elliptic Fourier coefficients approach approximates the coordinates of points around the outline of the leaf to a trigonometric function. Moment invariants tend to be more efficient in describing shape differences at the between-plant level, whereas elliptic Fourier coefficients yield a better separation at the population or regional levels (Lönn & Prentice 1990, Prentice 1992).

Molecular markers

During the past decades, classical methods to evaluate genetic variation have been complemented by molecular techniques. Molecular markers consist of DNA sequences at specific positions on a chromosome, or their immediate products like enzyme molecules. These markers are inherited in a Mendelian manner and may therefore be used as landmarks for genome analysis. Biochemical markers like isozymes reveal polymorphisms at the protein level and have been used for studying genetic variation within a large number of species (Hamrick & Godt 1989, Hamrick & Godt 1996). Isozyme markers are generally codominant, i.e. heterozygous individuals can be distinguished from homozygotes, but since these markers only detect variation in protein coding loci, they may reveal only a small amount of the variation present in the individual or population. Moreover, isozyme markers may be dependent on the developmental stage of the plant and their expression may be influenced by the environment.

DNA markers are numerous since they potentially may cover the entire genome. They are not influenced by developmental stage or environment, and allow selection of individuals already at the seedling stage, that is as soon as the plants are large enough to yield sufficient DNA. There is a number of molecular techniques available for characterization of the variation at the DNA level, e.g. RFLP (restriction fragment length polymorphism), which is a hybridization-based methodology using locus-specific probes, and the PCR (polymerase chain reaction) based polymorphisms like e.g. RAPD (random amplified polymorphic DNA), AFLP (amplified fragment length polymorphism) and ISSR (inter simple sequence repeats). RFLP is generally considered to be a reliable method, however labour-intensive and time-consuming, and requiring large amounts of DNA. Most of the PCR-based techniques are easy to perform and requires only small amounts of DNA. Furthermore, they are able to reveal a virtually unlimited number of markers.

For genetic diversity studies, the RAPD technique (Williams *et al.* 1990) shows some important advantages. In contrast to e.g. STMS (sequence tagged microsatellite sites) analysis, prior knowledge of the DNA sequence is not needed, which makes RAPDs very suitable for investigation of species that are not well known. Unspecific primers are used to amplify non-coding as well as coding regions of the DNA. The method is fast and easy to perform but has, unfortunately, a problem with reproducibility since small changes in the PCR conditions may lead to changes in amplified fragments (Williams *et al.* 1993). This problem may, however, be minimized by careful optimization and replication of reaction conditions. Competitive priming remains a more serious problem (Halldén *et al.* 1996). The use of longer primers in AFLP is generally believed to minimize the reproducibility problems found in RAPDs. However, reproducibility problems have been reported also in AFLP (Goulão *et al.* 2001). Moreover, in studies where different types of DNA analyses have been compared, RAPD has been shown to be just as efficient at estimating genetic variation as AFLP (Virk *et al.* 2000, Goulão *et al.* 2001).

Correlation between morphological and molecular methods

The use of different methods to evaluate genetic diversity may reveal different patterns of variation. Phenotypic differences are not necessarily correlated with the number of underlying gene mutations, and differences in phenotypic characters are not necessarily reflections of different genetic events (Bachmann 1992). RAPDs can potentially cover the entire genome (coding as well as non-coding regions), and since most of the genome is composed of non-coding DNA, it is plausible that the majority of the amplified fragments are from these regions. Mutations in non-coding DNA are selectively neutral and therefore, the RAPD analysis is able to detect even small differences in DNA which are not associated with phenotypic variation.

Morphological traits are prone to selection since they often are related to fitness. However, RAPD markers have also been shown to be under selective pressure. In their analysis of *Triticum dicoccoides*, Li *et al.* (1999) found a substantial amount of plant differentiation at the DNA level, which was associated with microclimatic stress. Moreover, quantitatively inherited characters like leaf morphology are often influenced by phenotypic plasticity (Widén *et al.* 1994), but see McLellan (2000). Consequently, a combination of morphological and molecular analyses may be the most useful alternative when trying to understand all aspects of genetic variation within a species (e.g. Olsson 1999).

Statistics

The use of RAPD markers presents some practical problems since they do not allow differentiation between homozygotes and heterozygotes. Consequently, many statistical procedures normally used in population genetics are not easily applicable to RAPD data. However, there are statistical analyses developed also for dominant markers, based on the assumption that each locus can be treated as a two-allele system, i.e. presence or absence of a band.

Genetic relationships

Once the morphological traits or the generated molecular marker profiles have been evaluated, there are different strategies how to estimate the similarity or dissimilarity between the analysed individuals. Similarity indices measure the amount of closeness between two individuals, the larger the value the more similar are the two individuals. There is a variety of alternative measures for expressing similarity, like Jaccard's coefficient of similarity which can be used for binary data and often is applied in RAPD-based studies. This coefficient is based on number of positive matches between two individuals whereas joint abscences are excluded. Dissimilarity coefficients instead estimate the distance or unlikeness of two individuals, the larger the value the more different are the two individuals. The Euclidean distance and the squared Euclidean distance are two commonly used measures of dissimilarity in both morphological and molecular analyses.

Matrices of similarities or dissimilarities between pairs of individuals may then be used as a starting point for statistical procedures such as cluster analysis, multidimensional scaling (MDS), or principal coordinate analysis (PCO). In a cluster analysis relatively homogeneous groups of individuals cluster together in a hierarchical way and this clustering is visually displayed in a dendrogram. MDS and PCO summarize dissimilarity data between individuals, or groups of individuals, in a non-hierarchical manner and then display the variation in an ordination plot.

Genetic and genotypic diversity

Two major measures of variation can be applied for analyses within a population: *genetic* variation (based on number and frequency of alleles) and *genotypic* variation (based on number and frequency of individual genets). These measures can be used for morphological as well as molecular characters. However, in plant population genetics, most applications involve formulae developed for co-dominant molecular markers.

The most widely used index for calculation of genetic diversity is H_e (expected heterozygosity), which is equivalent to Nei's unbiased gene diversity (Nei 1978). When used in mainly selfing species where heterozygotes are rare, this index should yield rather accurate estimations also for dominant markers (Lynch & Milligan 1994). However, for outcrossing species Hardy-Weinberg equilibrium must be assumed for each locus. A generally more unbiased gene diversity estimate for dominant markers in such cases was proposed by Lynch & Milligan (1994). An alternative approach for calculation of within-population variation is Shannon's diversity index (e.g. Bussell 1999), which does not assume Hardy-Weinberg equilibrium.

Plant species capable of vegetative, or clonal, growth produce offspring that are genetically identical to each other and to the maternal plant. The identical offspring of a single plant are called *ramets* and a group of ramets with the same genotype is called *genet* or *clone* (Cook 1983). Apart from variation due to somatic mutations, genetic variation occurs at the genet level. The genotypic variation within a clonal species may be calculated as the proportion of distinguishable genets (*PD*), the Simpson's diversity index (*D*) and the evenness measure (*E*) (Ellstrand & Roose 1987). *D* is 0 in a population composed of a single genet and 1 in a population where every sampled plant has a unique genotype. *E* is 0 in a population where all plants represent different genotypes or where one genotype is dominating and all the other genotypes are represented by a single plant, and 1 in a population where all genotypes are represented by the same number of plants.

Partitioning of variation

When a set of populations is investigated, the amount of genetic variability can be expressed at different hierarchical levels, e.g. between regions, between populations within regions and within populations.

Partitioning of variation in morphometric characters may be performed using canonical variates analysis (CVA). CVA provides a measure of the proportion of the total variation that is due to within-group variation, referred to as Wilks' lambda (Λ). The proportion of the total variation due to variation between groups is then given by $1-\Lambda$.

For molecular data, the AMOVA (analysis of molecular variance) procedure developed by Excoffier *et al.* (1992) has been widely used (see reviews in e.g. Bussel 1999, Nybom & Bartish 2000). This method was originally developed for haplotype data, but has recently become much used also for dominant markers like RAPDs. AMOVA has been applied mostly for estimation of population structure in diploid organisms, but has also been shown to give good estimation of population variability in tetraploid individuals (Jenczewski *et al.* 1999). AMOVA is based on squared Euclidean distances among individuals, and assumes that the studied populations are in Hardy-Weinberg equilibrium. An alternative approach to AMOVA is to calculate hierarchical components of the Shannon's diversity index (Bussell 1999). This index does not require any assumption about Hardy-Weinberg equilibrium.

Segregation and trait specific markers

Linkage disequilibrium is a condition in which certain alleles at two linked loci are non-randomly associated with each other, either because of their presence close together on the same chromosome or because of inbreeding or selection (Zamir & Tadmor 1986). High levels of linkage between markers may distort the estimates of genetic relationships, and such marker data must be treated with caution. Molecular markers have often been described as rather prone to segregation distortion. Generally, intraspecific crossings tend to yield lower rates of distortion than interspecific crossings (Jenczewski *et al.* 1997).

The major application of molecular markers in plant breeding involves indirect selection of plant material through markers linked to a trait of economical importance, for example disease resistance or sex determination. This method, which is called marker assisted selection (MAS), is relatively easy to accomplish for monogenic characters with simple Mendelian inheritance. One essential requirement for MAS is that the marker(s) should be easy to score and tightly linked to the gene of interest.

RAPD bands are generally inherited as dominant markers, with no possibility to distinguish between homozygous dominant genotypes and heterozygotes. Mendelian segregation ratios, i.e. 3:1 or 1:1, are expected if crosses have taken place between two heterozygotes or between a heterozygote and a homozygous recessive, respectively. Chi-square analyses are carried out to assess the goodness of fit of the segregation in the progeny to the expected ratios.

Plant introduction

The very first step in the domestication process involves selection of plant material in nature. A planned introduction of a species from one region to another involves many of the same genetic processes that operate during colonizing events in native populations (Barrett & Husband 1989). When a population is first established (e.g., founder events) a small and, in some cases, genetically isolated group is produced. Often, only a small amount of the genetic information present in the source is represented in the colonizing group. The domestication of plant species frequently involves genetic bottlenecks, which result in a reduction of genetic diversity in comparison with wild relatives (Barrett & Husband 1989). The genetic consequences of small population size are random fluctuations in allele frequencies and inbreeding, through selfing and biparental inbreeding. Generally speaking, genetic drift on the one hand results in decreased intrapopulational variation, and on the other hand increased interpopulational differentiation (Ellstrand & Elam 1993).

Turk's-cap lily, an ornamental containing substantial genetic variation

In **Paper I**, domesticated and native populations of Turk's-cap lily from Sweden, Denmark, Norway, Lithuania, Switzerland and Italy were analysed with RAPDs (Table 1). The domesticated and naturalized populations of today are basically descendants from plants that were collected from nature many years ago. Most domesticated populations are restricted in size and geographically isolated from each other. Consequently, Turk's-cap lily presents an opportunity to examine whether smaller populations have reduced genetic variation and whether genetic diversity in domesticated populations is different from the diversity in native populations.

The domesticated populations contained surprisingly high levels of variation in spite of at least one century spent in restricted and genetically isolated populations, and were as variable as the native populations (Table 1). Population size was significantly correlated with the amount of genetic variation; larger populations had higher levels of gene diversity than smaller populations. We also found a significant correlation between gene diversity and sample size. Due to restricted permission to collect plants, some of our samples were quite small, especially those collected in small populations. This may have lowered the gene diversity estimates for the smallest populations. However, in general there is no association between sample sizes and genetic diversity parameters (Nybom & Bartish 2000).

Table 1. Studied populations of Turk's-cap lily (D stands for domesticated and N for native), number of individual plants analysed, population size and gene diversity values (Lynch & Milligan diversity index)

Population/collection site	Individuals analysed	Pop. size	Gene diversity H
Ulfåsa, Östergötland, Sweden (D)	17	>25,000	0.24
Skokloster, Uppland, Sweden (D)	10	1500-3000	0.26
Linnés Hammarby, Uppland, Sweden (D)	7	1500-3000	0.25
Svartsjö slottspark, Uppland, Sweden (D)	10	1500-3000	0.23
Tyresö slottspark, Södermanland, Sweden (D)	6	1500-3000	0.20
Siggeberg/Gössäter, Västergötland, Sweden (D)	4	50-150	0.20
Råbäcks munkängar, Västergötland, Sweden (D)	4	200-300	0.17
Irup, Jylland, Denmark (D)	10	50-150	0.24
Levring, Jylland, Denmark (D)	9	200-300	0.19
Nes verk, Tvedestrand, Norway (D)	6	50-150	0.19
Vilnius, Lithuania (N)	6	50-150	0.20
Kaunas, Lithuania (N)	2	10-50	–
Botanical Garden, Kaunas, Lithuania (D)	2	10-50	–
Private garden, Kaunas, Lithuania (D)	2	10-50	–
Luzern, Switzerland (N)	5	200-300	0.15
Monte Baldo, Italy (N)	8	1500-3000	0.24

In Turk's-cap lily, selfing is prevented by an efficient self-incompatibility system (Lundquist 1991), which in turn contributes to the maintenance of high levels of genetic variation. Furthermore, the overlapping generations of perennials delay the loss of genetic variability that may occur in small populations (Ranker 1994). Turk's-cap lily reproduces predominantly by seed set (Tillge 1967), and has a prolonged generation time caused by a delay in reproductivity (Lundquist 1991), which could also promote the retaining of rather high gene diversity values in spite of small population sizes. In addition, the plant material was most probably highly heterogeneous already from the beginning and since then, no selections of plant material have been made. Consequently, populations of Turk's-cap lily may be quite useful for plant conservation and breeding purposes even after a long history of domestication.

Pronounced founder effect in black chokeberry

Whereas surprisingly high levels of variation were found in domesticated populations of Turk's-cap lily (**Paper I**), a totally different pattern was found when we analysed cultivated material of black chokeberry (**Paper V**). No RAPD band variation at all was encountered either among four cultivars ('Aron', 'Nero', 'Viking', and *Aronia melanocarpa* var. *elata*) or among seedlings derived from material collected in a seed-propagated plantation in Russia. Differences within the cultivated plant material have, however, been found when studying the chemical content of the berries

(Jeppsson 2000). Some of these differences may have been caused by somatic mutations, which are well known for affecting fruit colour yet leaving DNA marker profiles intact (Nybom 1990). Most likely, the absence of molecular variability in the cultivated material of black chokeberry is partly a result of founder effect. The cultivation in Russia started with the use of seeds from Germany. Large scale orchards were established in Siberia with plant material most likely derived from this initial introduction, and the cultivation of the crop eventually spread all over the former Soviet Union (Kask 1987). This severe founder event resulted in a small genetic basis already from the beginning. Our analyses also showed that the cultivated material is tetraploid and probably facultatively apomictic.

Genetic enhancement

Continuation of the black chokeberry case

As previously mentioned, no molecular variation at all was found in the cultivated material of black chokeberry. In such a case, breeding efforts based on controlled crosses will be futile and consequently, it is of vital importance to broaden the genetic basis. This may be achieved by incorporation of new selections from nature. However, quite often there is not much knowledge of the genetic structure in these native populations. An accurate and complete study of patterns of genetic variation within black chokeberry would facilitate the development of proper breeding strategies considerably.

Eight native populations from North America (three or five mother plants/population and five seedlings/mother plant, the latter equivalent to a family) were analysed with RAPDs (**Paper V**). Within this native plant material, there were two types of mother plants: those that produced very heterogeneous offspring and those that produced more or less homogeneous offspring (Table 2).

Analysis of ploidy level of one randomly chosen plant from each family, revealed that the families with many polymorphic bands consisted of diploid plants whereas the families with few or no polymorphic bands consisted of tetraploids. Some of the populations contained only diploid or tetraploid plants, whereas two populations had plants of both ploidy levels. Within-population variation differed among populations, and was associated with ploidy levels. Tetraploid populations had the lowest amount of diversity and diploid populations somewhat higher, whereas populations consisting of both tetraploids and diploids contained the highest amount of diversity (Table 2). Partitionings of variability attributed approx. 20% of the variation to the among-population level in diploids, compared to approx. 55% in the tetraploids.

Table 2. Number of polymorphic bands and ploidy level in eight populations (D=diploid, T=tetraploid and M=mixed) and 26 families of *A. melanocarpa*, and gene diversity values (including standard error) for each population, estimated by the Lynch and Milligan index (H_{pop}) and by Shannon's index (H'_{pop})

Pop.	Number of polymorphic bands Within populations		Within families	Ploidy level	H_{pop}	H'_{pop}
3D:	29	Fam a:	21	2x	0.197 (0.029)	0.377 (0.055)
		Fam b:	18	2x		
		Fam c:	16	2x		
6M:	40	Fam a:	17	2x	0.233 (0.026)	0.535 (0.052)
		Fam b:	23	2x		
		Fam c:	6	4x		
8D:	35	Fam a:	23	2x	0.223 (0.028)	0.470 (0.056)
		Fam b:	24	2x		
		Fam c:	20	2x		
16T:	26	Fam a:	5	4x	0.165 (0.024)	0.386 (0.059)
		Fam b:	1	4x		
		Fam c:	0	4x		
20T:	20	Fam a:	4	4x	0.138 (0.026)	0.305 (0.056)
		Fam b:	0	4x		
		Fam c:	0	4x		
23T:	26	Fam a:	2	4x	0.190 (0.027)	0.393 (0.059)
		Fam b:	8	4x		
		Fam c:	1	4x		
24M:	48	Fam a:	22	2x	0.281 (0.024)	0.640 (0.047)
		Fam b:	15	4x		
		Fam c:	21	2x		
32T:	21	Fam a:	3	4x	0.141 (0.026)	0.281 (0.050)
		Fam b:	1	4x		
		Fam c:	1	4x		
		Fam d:	1	4x		
		Fam e:	0	4x		

The pronounced discrepancy in levels of genetic variability in a plant material that looked very homogeneous in its native populations, prompted additional analyses of RAPD variation in open-pollinated offspring obtained from seed set in an experimental field at Balsgård. The analyses showed that tetraploid plants produced offspring that, with few exceptions, were identical whereas the offspring of diploid plants were highly heterogeneous.

The difference in genetic variation for diploid and tetraploid families points to different modes of reproduction. In the diploids, outcrossing presumably generates a high level of polymorphism, whereas the very restricted amount of polymorphism within tetraploid progeny groups indicates repeated selfing or apomixis. For differences in level of homogeneity between ploidy levels to be caused by differences in

amount of selfing, we must assume that there are concordant differences in self-compatibility between ploidy levels. However, as shown by Hardin (1973), all *Aronia* species are self-compatible.

In apomicts, the asexual reproduction of seeds results in offspring that are genetically an exact copy of the mother plant. The maternal-like progeny from tetraploid plants of black chokeberry suggests occurrence of apomixis in the analysed populations. Apomictic seed set has previously been suggested to occur in *Aronia* (Hardin 1973, Poplavskaya 1995), and is quite common in Rosaceae. However, since sexual recombination obviously occurred in the progeny (two of the progeny groups had one offspring with a deviating RAPD profile), the tetraploid form of black chokeberry appears to be a facultative apomict.

In our study of black chokeberry, we found the diploid and the tetraploid gene pools to be somewhat dissimilar yet overlapping (Fig. 1). A possible explanation is that the tetraploids contain genetic material from more than one diploid species. Black chokeberry and red chokeberry are sympatric over a rather wide area and are capable of hybridization (Hardin 1973). Our tetraploid plant material may consist of, or include, the allopolyploid hybrid *A. prunifolia*, which morphologically is very difficult to discriminate from black chokeberry.

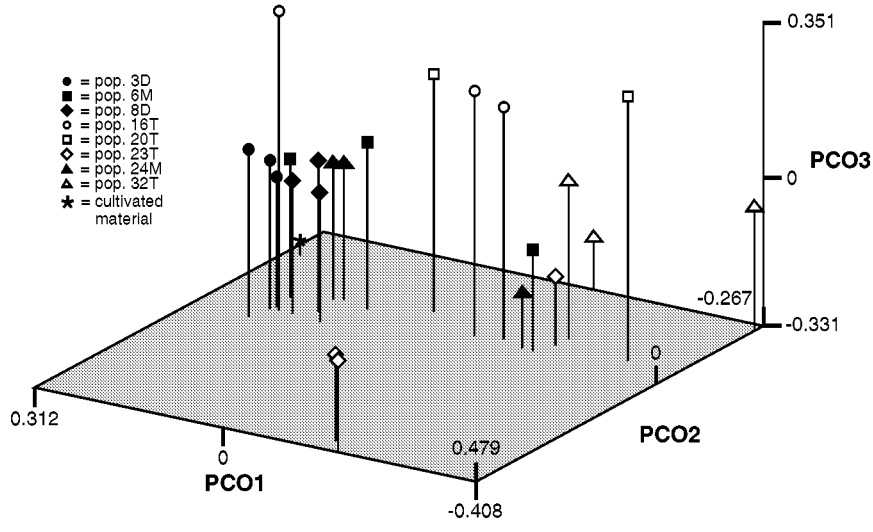

Fig. 1. Three-dimensional plot from a PCO showing the variation among cultivated plant material and 26 natural offspring families of black chokeberry.

In conclusion, for plant breeding purposes, it is necessary to broaden the genetic basis of black chokeberry. Since the native populations contained substantial amounts of variation, this may be achieved by the introduction of germplasm from natural stands. Great care should, however, be taken to make sure that diploid sexual genotypes and polyploid apomictic genotypes are properly identified prior to being used in breeding programs.

Genetic structure in vegetatively propagating lingonberry populations

For breeding purposes, new plant material is continuously collected from native populations of lingonberry. However, to be able to collect material in an optimal way, more information is needed about the extent of clonality and genetic variation in these populations. It is generally assumed that recruitment from seeds is rare and infrequent in clonal plant species (Eriksson 1989), which implies that they should have a lower level of genetic variation than non-clonal plants. Several studies have, however, found clonal plants to possess the same amount of genetic variation as that reported in non-clonal plants (e.g. Ellstrand & Roose 1987, Hamrick & Godt 1989).

For investigations of population biology in clonal plants, discrimination of genets is an important first step. Identification of different genets can in some cases be achieved by evaluation of morphological traits (e.g. Kemperman & Barnes 1976). A more efficient tool for distinguishing genets is the use of molecular markers, such as RAPDs (Hsiao & Rieseberg 1994, Gabrielsen & Brochmann 1998).

In **Paper IV**, RAPDs were used to evaluate the extent of clonality and genetic diversity in two native Swedish populations of lingonberry. We were able to identify 29 different genets among 129 analysed plants (ramets). The larger clones extended at least 30m, whereas the smallest clones were represented by a single ramet (Fig 2). Some of the smaller clones may be the result of recent seedling establishments or somatic mutations, whereas others may have had their main occurrence outside the analysed plots. Due to the linear sampling procedure, we were able to determine the clonal size in one direction only. Sampling was however, undertaken in very homogeneous areas, where the clones should have had the opportunity to spread rather evenly in all directions.

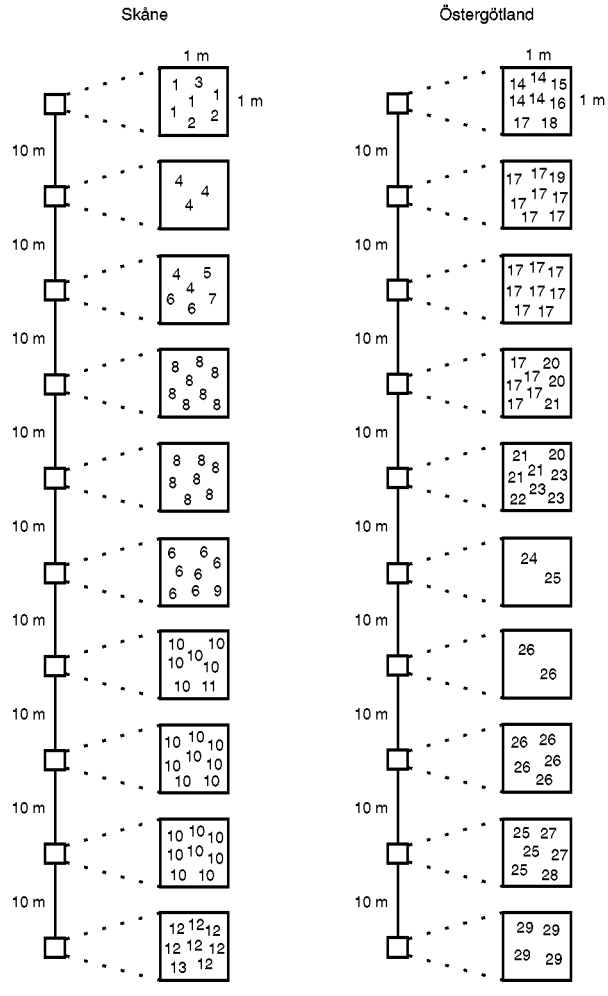

Fig. 2. Spatial distribution of the 29 putative clones of lingonberry. Since we do not know the exact position of the plants within each plot, the ramets representing the different clones are placed arbitrarily.

Measures of genotypic diversity (i.e. *PD*, *D* and *E*) were all within the range of estimates reported from other predominantly clonal plants (Ellstrand & Roose 1987). However, these estimates of genotypic diversity are not entirely comparable among studies since they may be influenced by the scale of sampling (Gabrielsen & Brochmann 1998). Furthermore, several of the studies included in the calculation of clonal plant averages are based on a very small set of allozyme loci, a factor that may influence the number of genotypes detected and hence also the *PD*, *D* and *E* estimates (Ellstrand & Roose 1987).

Within-population genetic diversity values, calculated on genets only, appeared to be similar to values obtained for other non-clonal outcrossing plants, indicating that sexual reproduction has played an important role at some time during the history of the populations. This was also indicated by AMOVA, which attributed a substantial amount of the variability (89.2%) to the within-population component.

Previous work indicates that seedlings of lingonberry are able to establish in sites where there is little competition from already established clones (Eriksson & Fröborg 1996). These optimal growing conditions are probably rare and unpredictable, since few observations on seedling recruitment in wild stands have been made. Nevertheless, the high levels of variation found in the present study indicate that the rate of seedling recruitment is sufficiently high to maintain genetic diversity.

Our data on the extent of clonality and genetic diversity in wild plant populations offers some indications for how to collect plant material for breeding purposes. To maximize variation in plant material from lingonberry, collections should be made with a minimum inter-plant distance of 20-30m. Collecting material from geographically widely separated populations will not necessarily increase total variability, since our study shows that the major part of the variation in lingonberry lies within populations. Still, the search for climatic adaptation and other, more specific characters may necessitate collection over wider areas.

Leaf shape analysis in RAPD-defined clones of lingonberry

As previously mentioned, the use of automated image analysis has several advantages over the scoring of morphological characters by hand. It allows a rapid and cost-effective scoring of a large number of plants, and it is also possible to choose the descriptor system that is best suited to a particular problem.

In **Paper IV**, plants investigated with both morphometric and molecular methods were plotted on the first two canonical variates obtained with elliptic Fourier coefficients, with RAPD-defined clones as groups and plants as replicates (Fig. 3a and b). The plots revealed a tendency for grouping of plants which belonged to the same clone. Moreover, morphometric analyses based on moment invariants and molecular analyses revealed a rather good concordance between partitions of diversity.

Classification tests revealed, however, that elliptic Fourier coefficients were more efficient than moment invariants in assigning the plant material to the correct clone.

The lingonberry plants in our study were grown under controlled conditions in the greenhouse, which should minimize the effect of environmental factors. Moreover, if plasticity is a major source of variation, one would expect a much larger within-plant component of variation than that observed in our study ($\Lambda<0.1$).

In conclusion, there was a rather good concordance between morphometric and molecular analyses in this study and consequently, a combination of these methods may be useful when trying to understand different aspects of genetic variation within lingonberry.

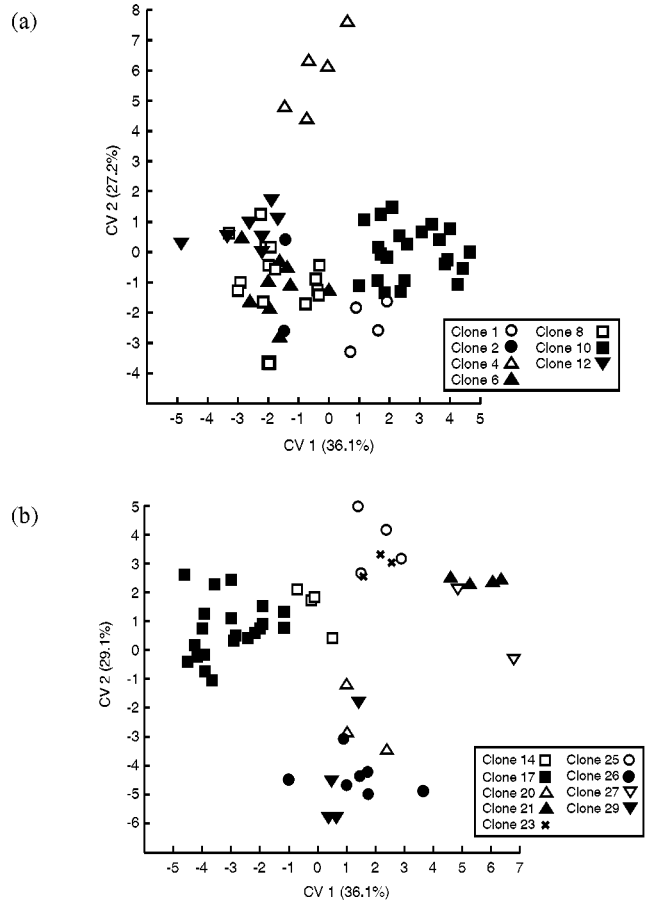

Fig. 3a and b. Between-genet variation in leaf shape in lingonberry based on elliptic Fourier coefficients. (a) The population from Skåne (b) The population from Östergötland.

Identification of genotypes

Unambiguous identification of plant cultivars and other genotypes of interest is important for practical plant breeding purposes. Horticulturally important traits or other morphological traits, e.g. the UPOV guide-lines, are often used for identification. However, also molecular markers have been used for identification in a large number of crops (e.g. Ling *et al.* 1997, Obara-Okeyo & Kako 1998, Sedra *et al.* 1998).

Discrimination among culinary rhubarb cultivars

Cultivar identification of culinary rhubarb is based mainly on morphological traits. This approach is, however, limited in its usefulness, since the variation in morphological traits often is influenced by environmental factors. Therefore, morphological characterization may need to be supplemented by more stable yet polymorphic markers for e.g. identification of cultivars and for estimation of genetic relationships.

In **Paper III** morphological characteristics and RAPD data were compared with respect to their ability to discriminate and evaluate genetic relationships among twelve different culinary rhubarb cultivars. The cultivars analysed were 'Bond', 'Canada Red', 'Early Sunrise', 'Elmsblitz, 'Elmsfeuer', 'German Wine', 'Marshall's Early Red', 'McDonald', 'Rosenhagen', 'Spangsbjerg', 'Tilden' and 'Victoria'. The cultivars were morphologically characterized using the UPOV guide-lines for evaluation of distinctness, homogeneity and stability (Anonymous 1978). Twelve reliable and easily scorable characters were selected for evaluation: leaf blade (colour; blistering; undulation of margin; shape of apex; anthocyanin coloration of main veins), petiole (attitude; length; width; ratio width/thickness; type of cross-section; ribs at back; flesh colour).

A significant association was found between the morphological and molecular distance matrices (r=0.41, P=0.001). However, when visually displayed in dendrograms, the groupings based on RAPD data were only to a limited extent supported by the morphological data (Figs. 4a and b). In other studies where data on genetic similarity obtained by RAPD analysis have been compared with classifications based on morphological traits, agreements as well as discrepancies have been found (e.g. Lerceteau *et al.* 1997, Sedra *et al.* 1998, Harrison *et al.* 1997, Papa *et al.* 1998, Martinello *et al.* 2001). The use of morphological traits is often not the most informative method when evaluating genetic relationships. However, in our study of rhubarb, the correlation between the two types of methods was at least moderate, suggesting that the selected morphological characters yield some useful information about genetic relationships.

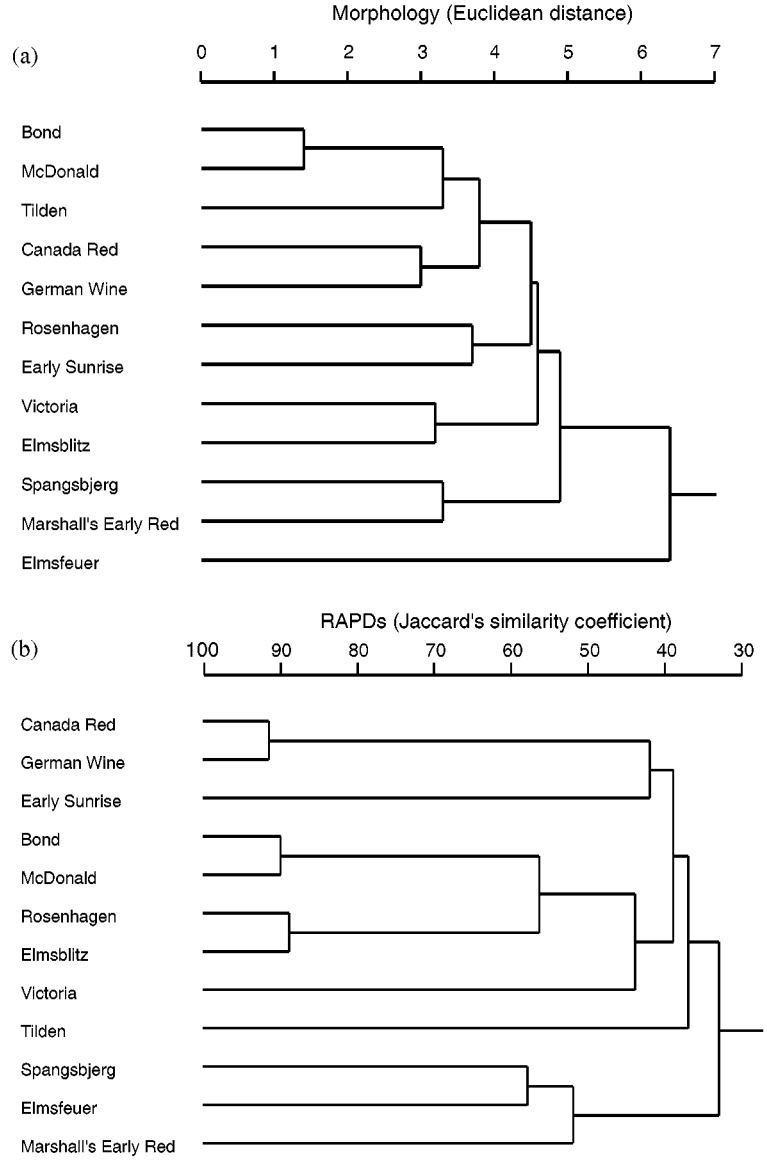

Fig. 4. Dendrograms of 12 rhubarb cultivars using (a) morphological characters and (b) RAPD markers.

Trait specific markers

The genetic improvement of a crop species depends on the ability to select promising plant material. To facilitate the selection process, molecular markers that are associated with important traits can be used as selection tools. The markers can then be used to establish genetic maps, which in turn are important tools for more refined marker-assisted selection in breeding programs as well as for in-depth genetic and systematic analyses.

The search for a sex-specific marker in sea buckthorn

In the commercial production of sea buckthorn berries, quality of the female plants is critical whereas the ca 10% male plants needed in the field are required only to produce large amounts of fertile pollen. Consequently, breeding efforts are directed mainly towards the development of superior female genotypes. Unfortunately, gender of sea buckthorn seedlings cannot be determined until flowering, which usually takes place after 3–4 years in the field. This represents a serious problem for plant breeders who have to retain large numbers of superfluous males for several years. An easily scored genetic marker, which could be used at an early stage for screening sea buckthorn seedlings and discarding the males, would be very useful in plant breeding programs.

In general, gender is genetically determined in dioecious plants, either by the occurrence of distinguishable sex chromosomes or, more commonly, by the expression of alleles at one or several autosomal loci (Irish & Nelson 1989, Durand & Durand 1990), but environmentally induced sex determination has also been demonstrated (Irish & Nelson 1989). The mechanism governing sex determination in sea buckthorn has not been determined. However, the occurrence of distinguishable sex chromosomes has been reported with the males being heteromorphic, suggesting that gender is determined by an X/Y system (Shchapov 1979).

In **Paper II**, two different F1 progenies derived from the crosses 'Leikora' (female) x 'Pollmix I' (male), and BHi 10224 (female) x 2-24 (male) were analysed with RAPDs. The analysis of marker segregation revealed markers with genotypic frequencies that corresponded to the expected ratios, i.e 1:1 or 3:1, as well as markers that showed skewed segregation ratios (Table 3). Deviating segregation ratios may be the product of linkages between molecular markers and distorting factors that operate in pre- and postzygotic phases of reproduction (Zamir & Tadmor 1986). In our study of *Hippophae*, postzygotic selection is unlikely, as seedling survival was > 90%. However, markers with skewed ratios may still be used in linkage analyses. Conner *et al.* (1997) found that markers with distorted segregation ratios mapped with about the same efficiency as markers which segregated in a Mendelian manner. Consequently, they decided to include both types of markers in their linkage map of apple.

Table 3. Segregation and Chi-square values of RAPD markers in two different crosses of sea buckthorn. *, $p < 0.05$; **, $p < 0.01$ and ***, $p < 0.001$ indicate significant departure from expected Mendelian segregation ratios

Cross	Marker	Present	Absent	Expected ratio	$\chi 2$
'Leikora' x 'Pollmix I'	OPA11-900	27	7	3:1	0.35
	OPA11-600	31	3	1:1	23.06 ***
	OPA11-400	10	24	1:1	5.76 *
	OPB10-1030	22	12	1:1	2.94
	OPB18-1500	17	17	1:1	0.00
	OPB18-1030	18	16	1:1	0.12
	OPB18-250	32	2	3:1	6.63 *
	OPD15-1100	13	21	1:1	1.88
	OPD15-600	17	17	1:1	0.00 male specific
	OPD15-400	17	17	1:1	0.00
BHi 10224 x 2-24	OPA11-1500	6	16	1:1	4.54 *
	OPA11-1250	16	6	3:1	0.06
	OPA11-600	19	3	3:1	1.52
	OPA11-400	2	20	1:1	14.73 ***
	OPB10-1030	5	17	1:1	6.54 *
	OPB18-1100	16	6	3:1	0.06
	OPB18-1030	7	15	1:1	2.91
	OPB18-1000	19	3	3:1	1.52
	OPB18-520	19	3	3:1	1.52
	OPB18-250	18	4	3:1	0.54
	OPD15-1100	1	21	1:1	18.18 ***
	OPD15-600	1	21	1:1	18.18 ***
	OPD15-400	19	3	3:1	1.52

In one of the crosses, ('Leikora' x 'Pollmix I'), we found one marker that was present in all males and in the male parent, but not in any of the females or in the female parent. In the other cross, (BHi 10224 x 2-24), this marker was present in only one of the males and not in any of the females. The male specific band appears to be linked to a sex determining region in 'Pollmix I' and therefore this band is useful as a genetical marker for gender in progenies derived from 'Pollmix I' or, possibly, from other related genotypes. Gender in sea buckthorn is most likely genetically determined, although the presence of sex chromosomes could not be verified in the present study. Previous attempts to identify sex-linked RAPD-markers in plant species have met with variable success (e.g. Mulcahy *et al.* 1992, Hormaza *et al.* 1994, McLetchie & Tuskan 1994, Cipriani *et al.* 1996). In general, to find a good marker, it seems crucial to screen a large amount of primers to obtain sufficiently many bands. Moreover, even if a sex-specific marker is identified in a particular cross, it is not necessarily informative in another cross.

Concluding remarks

Identification and utilization of diverse germplasm is the central issue in plant breeding. Basic information about the genetic variation in domesticated and native populations of a crop species will enhance our ability to utilise crop gene pools in an effective way in different breeding programs. Collections of plant material from natural populations will efficiently capture a major part of the genetic variation present in nature, provided that the genetic structure in these populations is well known.

The investigations included in my thesis show that genetic variation may be very different in different crops, e.g. domesticated populations of the obligate outbreeder Turk's-cap lily contained surprisingly high levels of variation, whereas domesticated and presumably apomictic populations of black chokeberry showed no variation at all. In general, native populations appear to contain substantial amounts of genetic variation and thus form a valuable contribution to plant breeding programs. Moreover, the combination of morphological and molecular analyses can be useful when trying to understand different aspects of variation within a species, as shown in the rhubarb and the lingonberry studies. In crops where valuable cultivars already have been developed, e.g. culinary rhubarb, molecular methods may be used as a complement to morphological traits, for simple and fast identification of cultivars. Finally, when analysing segregating RAPD markers in sea buckthorn we found a sex-specific marker, which, however, only worked in one cross. Still, the detection of such a marker indicates that gender is genetically determined within this species, and that a more general marker could be used for efficient screening of gender in future breeding projects.

Acknowledgements

Yes, I am finally there... after four years of struggling I have reached the finishing tape with a thesis in my pocket - and it feels good! Sometimes I have been running against the wind, but most often I have had the wind behind me. However, most important of all is that I have not been running alone. Quite a number of people have helped me and backed me up, and all of you deserve a heartfelt thank you.

First of all, I would like to thank *Hilde Nybom* for being such an enthusiastic and competent supervisor. Thank you *Hilde* for your continuous encouragement and interest in my work.

Thanks also to my co-supervisor *Stefan Andersson* for reading the manuscripts and giving valuable comments and statistical advice.

Furthermore, I would like to thank:

Gun Werlemark, for being such a cheerful companion in the lab. It has been fun, whether we have been talking about genetics or more worldly matters.

Björn Gustavsson, Niklas Jeppsson, Kjell Lundquist, Kimmo Rumpunen and *Katarina Wahlberg* for kindly giving me access to the different crop species, and for helping me with various analyses.

Igor Bartish for helping me with different statistical calculations and interpretations.

Åsa Olsson for helping with leaf shape analysis in the lingonberry study.

Thanks also to the rest of the staff at Balsgård and to *Bengt-Olof Bjurman* for his angelic patience with all my stupid questions about computers.

Finally, I would like to thank my friends and relatives for support and encouragement during these four years, and especially I would like to thank my mother and my sister *Christina*. Ett stort tack till min mamma och till min syster *Christina* för att ni finns där och stöttar mig.

Jag tänker också med värme och tacksamhet på min pappa som fanns i mitt liv då jag började den långa och mödosamma färden mot en doktorstitel, men som inte fick vara med då jag äntligen nådde fram. Jag vet att du hade varit stolt över mig idag.

References

Anderberg AA, Anderberg A-L, Ehn Salter M, Jacobson A and Larsdotter K (1997) Den virtuella floran. http://linnaeus.nrm.se/flora/ © Swedish Museum of Natural History, Sweden (in Swedish, information taken 2000-12-29).

Anonymous (1978) Guidelines for the conduct of tests for distinctness, homogeneity and stability, Rhubarb (*Rheum rhabarbarum* L.). UPOV.

Bachmann K (1992) Phenotypic similarity and genetic relationship among populations of *Microseris bigelovii* (Asteraceae: Lactuceae). Botanica Acta 105: 337–342.

Barrett SCH and Husband BC (1989) The genetics of plant migration and colonization. In: Plant population genetics, breeding, and genetic resources (eds. Brown AHD, Clegg MT, Kahler AL and Weir BS). Sinauer, Sunderland, Massachusetts, pp. 254–277.

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

Cipriani G, Di Bella R and Testolin R (1996) Screening RAPD primers for molecular taxonomy and cultivar fingerprinting in the genus *Actinidia*. Euphytica 90: 169–174.

Conner PJ, Brown SK and Weeden NF (1997) Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. Journal of the American Society of Horticultural Science 122: 350–359.
Cook RE (1983) Clonal plant populations. American Scientist 71: 244–253.
Darlington CD and Janaki EK (1945) Chromosome atlas of cultivated plants. George Allen & Unwin Ltd., London.
Dudani SA, Breeding KJ and McGhee RB (1977) Aircraft identification by moment invariants. IEEE Transactions on Computers C-26: 39–45.
Durand R and Durand B (1990) Sexual determination and sexual differentiation. Critical Reviews in Plant Science 9: 295–316.
Ellstrand NC and Roose ML (1987) Patterns of genotypic diversity in clonal plant species. American Journal of Botany 74: 123–131.
Ellstrand NC and Elam DR (1993) Population genetic consequences of small population size: Implications for plant conservation. Annual Reviews in Ecology and Systematics 24: 217–242.
Englund R (1983) Odlad rabarber och taxonomiska problem inom släktet *Rheum*, speciellt sektionen *Rhapontica*. Report, Dept. of Systematic Botany, Uppsala University, Sweden (in Swedish).
Eriksson G (1969) Botanikens historia i Sverige intill år 1800. Almqvist & Wiksell, Stockholm (in Swedish).
Eriksson O (1989) Seedling dynamics and life histories in clonal plants. Oikos 55: 231–238.
Eriksson O and Fröborg H (1996) "Windows of opportunity" for recruitment in long-lived clonal plants: experimental studies of seedling establishment in *Vaccinium* shrubs. Canadian Journal of Botany 74: 1369–1374.
Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.
Feldmaier C and McRae J (1982) Lilien. Verlag Eugen Ulmer, Stuttgart.
Flora of China (1996) http://flora.huh.harvard.edu/china/ © Flora of China Project (information taken 2001-08-07).
Gabrielsen TM and Brochmann C (1998) Sex after all: high levels of diversity detected in the arctic clonal plant *Saxifraga cernua* using RAPD markers. Molecular Ecology 7: 1701–1708.
Goulão L, Cabrita L, Oliveira CM and Leitão JM (2001) Comparing RAPD and AFLP™ analysis in discrimination and estimation of genetic similarities among apple (*Malus domestica* Borkh.) cultivars. Euphytica 119: 259–270.
Gustavsson B (1999) Lingonodling – använd mikroförökade plantor och täck med torv. Fakta Trädgård 2. Report, Swedish University of Agricultural Sciences, Sweden (in Swedish).
Halldén C, Hansen M, Nilsson N-O, Hjerdin A and Säll T (1996) Competition as a source of errors in RAPD analysis. Theoretical and Applied Genetics 93: 1185–1192.
Hamrick JL and Godt MJW (1989) Allozyme diversity in plant species. In: Plant population genetics, breeding, and genetic resources (eds. Brown AHD, Clegg MT, Kahler AL and Weir BS). Sinauer, Sunderland, Massachusetts, pp. 43–63.
Hamrick JL and Godt MJW (1996) Effects of life history traits on genetic diveristy in plant species. Philosophical Transactions of the Royal Society London, Series B 351: 1291–1298.
Hardin JW (1973) The enigmatic chokeberries (*Aronia*, Rosaceae). Bulletin of the Torrey Botanical Club 100: 178–184.
Harrison RE, Luby JJ, Furnier GR and Hancock JF (1997) Morphological and molecular variation among populations of octoploid *Fragaria virginiana* and *F. chiloensis* (Rosaceae) from North America. American Journal of Botany 84: 612–620.

Hintze S (1951) Rabarber. In: Svensk växtförädling Del II Trädgårdsväxterna Skogsväxterna. Natur och Kultur, Stockholm, pp. 389–391 (in Swedish).
Hormaza JI, Dollo L and Polito VS (1994) Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis. Theoretical and Applied Genetics 89: 9–13.
Hsiao J-Y and Rieseberg LH (1994) Population genetic structure of *Yushania niitakayamensis* (Bambusoideae, Poaceae) in Taiwan. Molecular Ecology 3: 201–208.
Irish EE and Nelson T (1989) Sex determination in monoecious and dioecious plants. Plant Cell 1: 737–744.
Jacquemart A-L and Thompson JD (1996) Floral and pollination biology of three sympatric *Vaccinium* (Ericaceae) species in the Upper Ardennes, Belgium. Canadian Journal of Botany 74: 210–221.
Jasienski M, Ayala FJ and Bazzaz FA (1997) Phenotypic plasticity and similarity of DNA among genotypes of an annual plant. Heredity 78: 176–181.
Jenczewski E, Gherardi M, Bonnin I, Prosperi JM, Olivieri I and Huguet T (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). Theoretical and Applied Genetics 94: 682–691.
Jenczewski E, Prosperi JM and Ronfort J (1999) Differentiation between natural and cultivated populations of *Medicago sativa* (Leguminosae) from Spain: analysis with random amplified polymorphic DNA (RAPD) markers and comparison to allozymes. Molecular Ecology 8: 1317–1330.
Jeppsson N (1999a) Evaluation of black chokeberry, *Aronia melanocarpa*, germ plasm for production of natural food colourants. Acta Horticulturae 484: 193–198.
Jeppsson N (1999b) Genetic variation and fruit quality in sea buckthorn and black chokeberry. PhD dissertation, Swedish University of Agricultural Sciences, Alnarp, Sweden.
Jeppsson N (2000) The effect of cultivar and cracking on the fruit quality in black chokeberry (*Aronia melanocarpa*) and the hybrids between chokeberry and rowan (*Sorbus*). Gartenbauwissenschaft 65: 93–98.
Jeppsson N and Johansson R (2000) Changes in fruit quality in black chokeberry (*Aronia melanocarpa*) during maturation. Journal of Horticultural Science & Biotechnology 75: 340–345.
Karling S (1931) Trädgårdskonstens historia i Sverige intill Le Nôtrestilens genombrott. Albert Bonniers Förlag, Stockholm (in Swedish).
Kask K (1987) Large-fruited black chokeberry (*Aronia melanocarpa*). Fruit Varieties Journal 41: 47.
Kemperman JA and Barnes BV (1976) Clone size in American aspens. Canadian Journal of Botany 54: 2603–2607.
Kercher SM and Sytsma KJ (2000) Genetic and morphological variation in populations of the rare prairie annual *Agalinis skinneriana* (Wood) Britton (Scrophulariaceae). Natural Areas Journal 20: 166–175.
Kincaid DT and Schneider RB (1983) Quantification of leaf shape with a microcomputer and Fourier transform. Canadian Journal of Botany 61: 2333–2342.
Kuhl FP and Giardina CR (1982) Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18: 236–258.
Lerceteau E, Quiroz J, Soria J, Flipo S, Pétiard V and Crouzilat D (1997) Genetic differentiation among Ecuadorian *Theobroma cacao* L. accessions using DNA and morphological analyses. Euphytica 95: 77–87.
Li TSC and Schroeder WR (1996) Sea buckthorn (*Hippophae rhamnoides* L.): a multipurpose plant. HortTechnology 6: 370–380.
Li YC, Fahima T, Beiles A, Korol AB and Nevo E (1999) Microclimatic stress and adaptive DNA differentiation in wild emmer wheat, *Triticum dicoccoides*. Theoretical and Applied Genetics 98: 873–883.

Libert B (1987) Breeding a low-oxalate rhubarb (*Rheum* sp. L.). Journal of Horticultural Science 62: 523–529.
Ling J-T, Sauve R and Gawel N (1997) Identification of poinsettia cultivars using RAPD markers. HortScience 32: 122–124.
Lönn M and Prentice HC (1990) Mosaic variation in Swedish *Petrorhagia prolifera* (Caryophyllaceae): the partitioning of morphometric and electrophoretic diversity. Biological Journal of the Linnean Society 41: 353–373.
Lundqvist A (1991) Four-locus S-gene control of self-incompatibility made probable in *Lilium martagon* (Liliaceae). Hereditas 114: 57–63.
Lynch M and Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Molecular Ecology 3: 91–99.
Martinello GE, Leal NR, Amaral Júnior AT, Pereira MG and Daher RF (2001) Comparison of morphological characteristics and RAPD for estimating genetic diveristy in *Abelmoschus* spp. Acta Horticulturae 546: 101–104.
McLellan T (2000) Geographic variation and plasticity of leaf shape and size in *Begonia dregei* and *B. homonyma* (Begoniaceae). Botanical Journal of the Linnean Society 132: 79–95.
McLellan T and Endler JA (1998) The relative success of some methods for measuring and describing the shape of complex objects. Systematic Biology 47: 264–281.
McLetchie DN and Tuskan GA (1994) Gender determination in *Populus*. Norwegian Journal of Agricultural Sciences Suppl. No. 18: 57–66.
Mulcahy DL, Weeden NF, Kesseli R and Carroll SB (1992) DNA probes for the Y-chromosome of *Silene latifolia*, a dioecious angiosperm. Sexual Plant Reproduction 5: 86–88.
Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590.
Nybom H (1990) DNA fingerprints in sports of 'Red Delicious' apples. HortScience 25: 1641–1642.
Nybom H and Bartish IV (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 3/2: 93–114.
Obara-Okeyo P and Kako S (1998) Genetic diversity and identification of *Cymbidium* cultivars as measured by random amplified DNA (RAPD) markers. Euphytica 99: 95–101.
Olsson Å (1999) Morphometric and molecular variation in the Nordic dogroses (*Rosa* Sect. *Caninae*, Rosaceae). PhD dissertation, Lund University, Sweden.
Papa R, Attene G, Barcaccia G, Ohgata A and Konishi T (1998) Genetic diversity in landrace populations of *Hordeum vulgare* L. from Sardinia, Italy, as revealed by RAPDs, isozymes and morphophenological traits. Plant Breeding 117: 523–530.
Poplavskaya TK (1995) *Aronia*, its economic significance and current status. In: A program and methods of breeding fruit, small-fruit and nut bearing crops. Oryol, Russia, pp. 457–459 (in Russian).
Prentice HC (1992) The structure of morphometric and allozyme variation in relict populations of *Gypsophila fastigiata* (Caryophyllaceae) in Sweden. Biological Journal of the Linnean Society 47: 197–216.
Raamsdonk LWD van (1993) Wild and cultivated plants: the parallelism between evolution and domestication. Evolutionary Trends in Plants 7: 73–84.
Ranker TA (1994) Evolution of high genetic variability in the rare Hawaiian fern *Adenophorus periens* and implications for conservation management. Biological Conservation 70: 19–24
Rousi A (1971) The genus *Hippophaë* L. A taxonomic study. Annales Botanicae Fennici 8: 177–227.
Rumpunen K and Pettersson M-L (1997) Renässans för rabarber - sortbedömningar och odlingsråd. Fakta Trädgård 57. Report, Swedish University of Agricultural Sciences, Sweden (in Swedish).

Sánchez-Monge E (1993) Introduction. In: Plant breeding. Principles & prospects (eds. Hayward MD, Bosemark NO and Romagosa I). Chapman & Hall, London, pp. 3–5.
Sedra My H, Lashermes P, Trouslot P, Combes M-C and Hamon S (1998) Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. Euphytica 103: 75–82.
Shchapov NS (1979) On the Karyology of *Hippophaë rhamnoides* L. Tsitologiya i Genetika 13: 45–47 (in Russian, with English abstract).
Skvortsov AK and Maitulina YK (1982) On the differences between the cultivated chokeberry and its wild progenitors. Bulletin of the Central Botanical Gardens, Moscow. Nauke Publishers 126: 35–40 (in Russian).
Teär J (1972) Vegetativ och fruktifikativ utveckling hos vildväxande och odlade lingon. PhD dissertation, Lantbrukshögskolan, Alnarp, Sweden (in Swedish).
Tillge L (1967) Primitivformer af udvalgte arter inden for Liliales og deres forhold til de respektive kulturformer, historisk, botanisk og cytologisk belyst. Licentiatopgave i faget Havebrugsbotanik ved Den kgl. Veterinær og Landbohøjskole, København (in Danish).
Turner DM (1938) The economic rhubarbs: a historical survey of their cultivation in Britain. Journal of the Royal Horticultural Society 63: 355–370.
Virk PS, Zhu J, Newbury HJ, Bryan GJ, Jackson MT and Ford-Lloyd BV (2000) Effectiveness of different classes of molecular marker for classifying and revealing variation in rice (*Oryza sativa*) germplasm. Euphytica 112: 275–284.
White RJ, Prentice HC and Verwijst T (1988) Automated image acquisition and morphometric description. Canadian Journal of Botany 66: 450–459.
Widén B, Cronberg N and Widén M (1994) Genotypic diversity, molecular markers and spatial distribution of genets in clonal plants, a literature survey. Folia Geobotanica et Phytotaxonomica, Praha 29: 245–263.
Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531–6535.
Williams JGK, Hanafey MK, Rafalski JA and Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology 218: 704–740.
Woodcock HBD and Stearn WT (1950) Lilies of the world. Their cultivation & classification. Country Life Limited, London.
Zamir D and Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Botanical Gazette 147: 355–358.