Genetic variation in the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers

ELS COART,* XAVIER VEKEMANS,† MARINUS J. M. SMULDERS,‡ IRIS WAGNER,§
JOHAN VAN HUYLENBROECK,* ERIK VAN BOCKSTAELE*¶ and ISABEL ROLDÁN-RUIZ*
*Department of Plant Genetics and Breeding, Agricultural Research Centre, Caritasstraat 21, 9090 Melle Belgium, †Université de Lille 1, Laboratoire de Génétique et Evolution des Populations Végétales, Bâtiment SN2, 59655 Villeneuve d'Ascq Cedex, France, ‡Plant Research International, P.O. Box 16, 6700 AA Wageningen, the Netherlands, §Pro Arbore Research Institute, Gustav-Adolf-Strasse 3, 01219 Dresden Germany, ¶University of Ghent, Faculty of Agricultural and Applied Biological Sciences, Coupure Links 653, 9000 Gent Belgium

Abstract

The genetic variation within and between wild apple samples (*Malus sylvestris*) and cultivated apple trees was investigated with amplified fragment length polymorphisms (AFLP) and microsatellite markers to develop a conservation genetics programme for the endangered wild apple in Belgium. In total, 76 putative wild apples (originating from Belgium and Germany), six presumed hybrids and 39 cultivars were typed at 12 simple sequence repeats (SSR) and 139 amplified fragment length polymorphism (AFLP) loci. Principal co-ordinate analysis and a model-based clustering method classified the apples into three major gene pools: wild *Malus sylvestris* genotypes, edible cultivars and ornamental cultivars. All presumed hybrids and two individuals (one Belgian, one German) sampled as *M. sylvestris* were assigned completely to the edible cultivar gene pool, revealing that cultivated genotypes are present in the wild. However, gene flow between wild and cultivated gene pools is shown to be almost absent, with only three genotypes that showed evidence of admixture between the wild and edible cultivar gene pools. Wild apples sampled in Belgium and Germany constitute gene pools that are clearly differentiated from cultivars and although some geographical pattern of genetic differentiation among wild apple populations exists, most variation is concentrated within samples. Concordant conclusions were obtained from AFLP and SSR markers, which showed highly significant correlations in both among-genotypes and among-samples genetic distances.

Keywords: AFLP, conservation, genetic diversity, hybridization, *Malus sylvestris*, microsatellites

Received 5 July 2002; revision received 15 November 2002; accepted 9 December 2002

Introduction

Important fruit crops cultivated in temperate regions have been domesticated within the *Rosaceae* family, e.g. apple (*Malus domestica*), pear (*Pyrus communis*), prune (*Prunus domestica*) and cherry (*Prunus avium*) cultivars.

The capacity of *Rosaceae* species for interspecific hybridization, even beyond genus borders, has been exploited in breeding programmes to incorporate desirable traits of wild populations into breeding gene pools. Hybridization

between fruit crops and their wild relatives has probably also occurred 'spontaneously' and individuals with intermediate phenotypes are known to occur throughout the European landscape. The importance of (anthropogenic-induced) hybridization processes has been underestimated by conservation biologists until recently (Allendorf *et al.* 2001). It is however, becoming more and more apparent that hybridization has led to the extinction of many populations and species and represents a severe threat especially to rare species that come into contact with other, more abundant species (Rhymer & Simberloff 1996). This may also be the case for wild European *Rosaceae* populations, which have been in contact with cultivated genotypes for

Correspondence: Els Coart. Fax: + 32 9272 29 01; E-mail: E.coart@clo.fgov.be

centuries. The extent to which hybridization processes with related crop varieties might have caused irreversible damage to the genetic identity of wild *Rosaceae* is unknown.

Other threats to the genetic diversity of natural populations of wild *Rosaceae* species, especially during the last decennia, are the elimination or reduction in size of many natural populations because of anthropogenic activities (Hokanson *et al.* 1998; Stephan *et al.*, in press). As a consequence, *in situ* and/or *ex situ* conservation programmes are being set up in different European countries to preserve the valuable genotypes and populations that might still be present in nature. For this, it is of paramount importance to discriminate between 'genuine' wild genotypes and genotypes derived from cultivars or their direct descendants.

Although the majority of studies concerning hybridization between crops and wild relatives are based on morphological evidence, (introgressive) hybridization is not necessarily indicated by the phenotypic occurrence of the characters of one taxon in another. Identical characters in cultivated varieties and their wild relatives may occur as a result of either phenotypic plasticity, convergent evolution, or simply common ancestry (Linder *et al.* 1998), the latter being the case for domesticated and wild pear for instance. Furthermore, individuals from hybrid swarms that obtained most of their genes from one of the parental taxa are often morphologically indistinguishable from that parental taxon (Allendorf *et al.* 2001). Molecular markers have become powerful tools with which to establish the extent of hybridization processes (Linder *et al.* 1998; see Jarvis & Hodgkin 1999 for a review) as well as to provide the background knowledge necessary to implement conservation genetics programmes (Haig 1998; Smulders *et al.* 2000). The advantages of using molecular markers in these fields are that most molecular markers do not exhibit plasticity, are unlikely to be similar because of convergent evolution and can generate information on many different loci. Several studies, however, have demonstrated that the results obtained using a given marker system may not fully agree with the results obtained using a different marker system, especially for loosely related genotypes (Roldán-Ruiz *et al.* 2001). This could be the result of differences among marker systems in the targeted genome (cytoplasmic vs. nuclear), in mutation rate, or in dominance characteristics. King & Burke (1999) recommended combining the information derived from different types of molecular markers to minimize the risks related to misinterpreting the data or to the inevitable drawbacks of each marker technique. Two of the currently most applied marker systems to study genetic diversity in plant species, are used here: amplified fragment length polymorphisms (AFLP) and microsatellites. However, appropriate data analysis approaches are necessary to obtain statistics that can be compared between dominant (AFLP) and co-dominant (microsatellite) markers (Lynch & Milligan 1994; Vekemans *et al.* 2002).

This paper reports on a study tackling the issues that have to be clarified prior to the development of a sound conservation programme for the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium. Recent inventories revealed that it has become one of the most endangered tree species in the region (Coart *et al.* 1998; Maes & Rövekamp 2000). The occurrence in nature of many phenotypically intermediate forms that resemble *Malus × domestica* Borkh. (cultivated apple) has led to the hypothesis that by frequent hybridization, the wild species has been replaced by a hybrid swarm, consisting of hybrid genotypes with a significant admixture of the cultivated genetic information (Kleinschmit *et al.* 1998). Since fruits on apple trees in dense forests are mostly absent, fruit characteristics cannot often be used for classification. Therefore, the hairiness of inferior leaf surfaces is considered one of the most discriminatory phenotypic characteristics between wild apples and edible cultivars [*M. × domestica*: felted leaf surface, *M. sylvestris*: sparsely hairy (on veins) in spring and hairless in autumn; Remmy & Gruber 1993, Wagner 1998; see Wagner (1996) for a review (in German) on morphological discrimination between wild and cultivated apple trees]. Individuals with different degrees of leaf hairiness are known to occur in Belgian forests and it is sometimes difficult to apply this criterion to classify a specific tree.

The objectives of the study were to (i) investigate the level of genetic differentiation between apple cultivars and Belgian wild apple trees using AFLP and simple sequence repeat (SSR) markers; (ii) specifically test for the occurrence of cultivated genotypes or hybrid genotypes in the wild; (iii) investigate the population genetic structure of wild apple trees from Belgian and other origins and test whether the Belgian trees form a distinct gene pool. With these objectives in mind, apple trees were inventoried and sampled in Belgian forests. Also genotypes with felted hairy leaves were sampled in the wild. Additional apple trees from three origins were also analysed to generate reference groups: (i) edible apple cultivars; (ii) ornamental cultivars; and (iii) wild individuals of *M. sylvestris* from Germany.

Materials and methods

Studied genotypes

In total, 44 Belgian putative wild apple trees were sampled at three different locations in forests, including trees with different degrees of leaf hairiness. In addition, six presumed hybrids, all with felted hairy leaves, were collected. These individuals were sampled near forests, on roadsides and in meadows.

Table 1 Description of the sampled apple genotypes

Origin	Code	Coord-inates	n	Material	Typed with	Description of sampling site
Wild Belgian						
Meerdaal	WBM	4°41′/50°48′	28	G	AFLP + SSR	*M. sylvestris* genotypes sampled in one forest, more genotypes present (estimated at 100 individuals, no inventory available), total area of collection: 6 km²
Voeren	WBVo	5°52′/50°44′	9	G	AFLP + SSR	*M. sylvestris* genotypes sampled in forests, all known genotypes in neighbourhood sampled, total area of collection: 3 km²
Viroin	WBVi	4°28′/50°04′	7	L	SSR	*M. sylvestris* genotypes sampled in forests, presumably more genotypes present (but no inventory of area available), total area of collection: ±20 km²
Belgian hybrids						
Limburg	WBL	5°15′/51°03′	6	G	AFLP + SSR	Presumed hybrids between *M. sylvestris* and *M. × domestica*, sampled near forests, total area of collection: 6 km²
Wild German						
Gene bank NFV Escherode (G)	WGL	/	23	G	AFLP + SSR	*M. sylvestris* genotypes of German Federal District Lower Saxony and Shleswic Holstein
Gene bank Forstgenbank Arnsberg (G)	WGF	/	6	L	AFLP + SSR	*M. sylvestris* genotypes of German Federal District North Rhein West Phalia
Wild private collection						
Private collection (B, F, GR)	WPr	/	3	L	SSR	*M. sylvestris* genotypes from the south of Belgium, south of France and Greece
Edible apple cultivars						
Gene bank CRA-Gembloux	ECV	/	11	G	AFLP + SSR	Old apple cultivars: 'Gris Braibant', 'Belle Fleur De Brabant', 'Oogstappel', 'Gueule de Mouton', 'Court Pendu Rose', 'Calville Des Prairies', 'Pomme De Sucre', 'Reinette De Wattripont', 'Pomme De Douce', 'Blanc Braibant', 'Jerusalem'
Ornamental apple cultivars						
Gene bank DvP	OCV	/	28	G	AFLP + SSR	Ornamental cultivars, including many different species: *M. mandsurica*, *M. × zumi* 'Calocarpa', *M.* 'Van Eseltine', *M.* 'Golden Hornet', *M.* 'Maypole', *M.* 'Adirondack', *M. baccata* 'Street parade', *M. baccata* 'Yellow Syberian', *M.* 'Professor Sprenger', *M.* 'Red Sentinel', *M.* 'John Downie', *M.* 'Liset', *M.* 'Butterball', *M.* 'Profusion', *M.* 'D.V.P. Obel' (Red Obelisk®), *M.* 'Eleyi', *M.* 'Evereste', *M. floribunda*, *M.* 'Hartuigii', *M. pumila* 'Hopa New', *M.* 'Makamik', *M.* 'Neville Copeman', *M.* 'Red Jade', *M.* 'Royalty', *M. toringo* 'Rosea', *M. toringo* var. *sargentii*, *M. tschonoskii*

Origin: B, Belgium; G, Germany; F, France; GR, Greece.
n, number of genotypes sampled.
Material: G, grafts; L, leaves.

Additional genotypes of *Malus sylvestris* were obtained from various sources: 29 individuals from two German gene banks, and three European presumed *M. sylvestris* genotypes from a private Belgian collection. In addition, grafts from 11 old edible apple cultivars were obtained from the gene bank of CRA-Gembloux. These ancient cultivars were chosen because they were cultivated at the time of germination of the putative wild Belgian apple trees sampled (the average age of the sampled apple trees in the forest is estimated at 100 years). One additional cultivar, 'Elstar', was added as a standard genotype for the microsatellite analysis (allele sizes of 'Elstar' are available for all used microsatellite loci, K. Kenis personal communication). Finally, 28 ornamental apple cultivars were sampled from the *Malus* collection of the Department of Plant Genetics and Breeding (DvP; Melle, Belgium). These ornamental cultivars have been derived from different *Malus* species, and are used in the breeding programme of ornamental apple trees. Locations and sample sizes are described in Table 1.

Table 2 Allelic diversity of the nuclear microsatellite loci scored in 119 apple genotypes

Locus	Repeat motif	Chromosome	Original publication		This study	
			No. of alleles	Range of sizes (bp)	No. of alleles	Range of sizes (bp)
NZ02b01*	$(GA)_{14}$	15†	7	212–238	18	194–246
NZ04h11*	$(GA)_{23}$	9†	6	201–233	13	200–242
NZ05g08*	$(GA)_{18}$	4†	6	115–141	19	90–156
NZ23g04*	$(GA)_{19}$	6†	9	84–116	17	82–125
NZ28f04*	$(GA)_{18}$	12†	4	98–112	11	91–123
CH01h10‡	$(AG)_{21}$	8§	7	93–119	26	88–147
CH01e12‡	$(AG)_{32}$	8¶	8	243–248	14	223–275
CH01f02‡	$(AG)_{22}$	12§	11	168–222	23	153–227
CH01h01‡	$(AG)_{25.5}$	17§	9	107–141	21	89–145
CH02b12‡	$(GA)_{26}$	5§ + 10¶	8	124–142	18	109–159
CH02c06‡	$(GA)_{21}(GA)_{17}$	2§	10	216–254	35	206–309
CH02d12‡	$(GA)_{19}$	11§	9	175–205	20	175–219

Normalized nomenclature of the loci follows Liebhard *et al.* (2002).
*From Guilford *et al.* (1997).
†Mapped in Maliepaard *et al.* (1998).
‡From Gianfranceschi *et al.* (1998).
§Mapped in Liebhard *et al.* (2002).
¶Mapped by Eric van de Weg *et al.*, in preparation.

Most of the genotypes included in this study were grafted at the nursery of the DvP (column 4 of Table 1). Young leaves were used for DNA extraction. The leaves were immediately frozen in liquid nitrogen and stored at –80 °C before freeze-drying for 48 h. Dried material was stored at room temperature under vacuum conditions until DNA extraction. For genotypes that were not grafted in the nursery, DNA had to be extracted from often badly preserved leaves collected in the field, and this was of insufficient quality for AFLP analysis. The DNA was of appropriate quality for SSR analysis, which is less sensitive to DNA quality. As a result, for the 10 individuals of WBVi and WPr no clean AFLP fingerprints could be obtained after several trials and only SSR data are available.

AFLP analysis

Total plant DNA was isolated from 30 mg of ground dried leaf material, using the CTAB extraction procedure described in Dumolin *et al.* (1995). DNA concentrations were estimated and standardized against known concentrations of λ DNA on 1.5% agarose gels. AFLP analysis was performed as described in Coart *et al.* (2002), using a Gene ReadIR 4200 analyser Sequencer (Li-Cor, Inc.). Primer combinations used for generation of fingerprints were *Eco*RI-ACA + *Mse*I-CAG, *Eco*RI-ACT + *Mse*I-CTG and *Eco*RI-AGC + *Mse*I-CAT. AFLP bands were scored as present (1) or absent (0), within the size range of 65–540 base pairs (bp). A reproducibility test was performed for five trees with all three primer combinations. For each sample, three independent DNA extractions were carried out and AFLP fingerprints were generated using all different DNA templates (a total of 45 AFLP fingerprints).

Microsatellite analysis

Twelve primer pairs of microsatellite loci (Table 2) which had previously been shown to display easy to read band patterns, which mapped to different linkage groups and which displayed a high degree of polymorphism in *Malus* species, were used for polymerase chain reactions (PCR; Guilford *et al.* 1997; Gianfranceschi *et al.* 1998; Maliepaard *et al.* 1998; Liebhard *et al.* 2002). All amplification reactions were performed as described in Gianfranceschi *et al.* (1998), using a PCR-Core-kit I (Promega). PCR reactions were carried out on a Perkin Elmer Geneamp PCR System 9600. Forward primers of each primer pair were labelled with a fluorescent near-infrared dye (IRD 700 or IRD 800). Two microsatellite PCR products, with the same IRD-label and different size ranges of amplification products, were multiplexed after PCR reaction and loaded on the gel together, enabling the visualization of four microsatellite loci in a single run (two loci by each 700 and 800 laser). Sample denaturation and gel electrophoresis were performed as described for AFLP, but the gels were run for 1.5 h at 1500 V. GENE IMAGIR software (Li-Cor, Inc.) was used to determine fragment sizes and to define allele bins. Alleles were scored according to their molecular weight and the occurrence of between-gel differences was checked using cultivar 'Elstar', which was present on each gel.

Morphological description

The hairiness of leaves was described in autumn for all the wild apple trees and the edible cultivars grafted in the nursery, according to Wagner (1998). Ornamental trees were not described as they belong to different species. The hairiness of the inferior leaf surfaces was scored from 0 to 3:

Score 0: hairless

Score 1: sparsely hairy, difficult to recognize (magnifying lens needed); only hairs on main nerves

Score 2: moderately hairy, easy to recognize by eye; hairs on nerves and entire leaf surface

Score 3: felted leaf surface

The grafted trees were too young to describe other morphological traits that are thought to discriminate between wild and cultivated trees, such as thorns on twigs and several flower and fruit characters.

AFLP data analysis

After elimination of monomorphic markers, Spearman correlations were calculated between pairs of markers. Using the software AFLP-SURV v1.0 (Vekemans *et al.* 2002; available at http://www.ulb.ac.be/sciences/lagev/), estimates of pairwise relatedness coefficients (*r*) were calculated between individuals according to Lynch & Milligan (1994). A principal coordinate (PCO) analysis was performed based on this matrix using NTSYS (Rohlf 2000) and the first two axes were plotted graphically together with 90% confidence ellipses for each group (STATISTICA 6.0, StatSoft Inc., Tulsa, OK). After exclusion of outliers (see below), pairwise genetic distances (Nei 1987) were calculated between samples. A neighbour-joining tree was computed, based on these distance measurements and a thousand bootstraps were performed over AFLP loci using AFLP-SURV and PHYLIP (Felsenstein 1993).

Furthermore, allelic variation, genetic diversity and differentiation statistics were computed using the software AFLP-SURV v1.0. Allelic frequencies at AFLP loci were calculated from the observed frequencies of fragments using the Bayesian approach proposed by Zhivotovsky (1999) for diploid species and assuming some deviation from Hardy–Weinberg genotypic proportions as estimated from SSR data. A nonuniform prior distribution of allelic frequencies was assumed with its parameters derived from the observed distribution of fragment frequencies among loci (see note 4 in Zhivotovsky 1999). These allelic frequencies were used as input for the analysis of genetic diversity within and between samples following the method described in Lynch & Milligan (1994). The significance of the genetic differentiation between groups was tested by comparison of the observed F_{ST} with a distribution of F_{ST} under the hypothesis of no genetic structure, obtained by means of 1000 random permutations of individuals among groups. In addition, average fragment size (with standard deviation) and Pearson correlation coefficients between fragment sizes and fragment frequencies (together with the *P*-value associated with the correlation) were calculated on the overall sample.

Microsatellite data analysis

Relatedness coefficients between pairs of individuals were estimated by computing the multilocus Moran's I statistic (Hardy & Vekemans 1999) using the program SPAGeDI 0.0 (available at http://www.ulb.ac.be/sciences/lagev/). A PCO analysis was performed based on this similarity matrix using NTSYS (Rohlf 2000); the first two axes were plotted graphically together with 90% confidence ellipses for each group (STATISTICA 6.0, StatSoft Inc., Tulsa, OK). After exclusion of outliers (see below), pairwise distances [Nei's (1978) standard genetic distance, and $\delta\mu^2$, Goldstein *et al.* 1995] were calculated between samples (SPAGeDI 0.0) and neighbour-joining trees with 1000 bootstraps over SSR loci were computed, based on both distance measures using PHYLIP (Felsenstein 1993) and MICROSAT (http://hpgl.stanford.edu/projects/microsat). Deviation of genotypic frequencies from Hardy–Weinberg proportions was tested with the program GENEPOP version 3.3 (Raymond & Rousset 1995).

The following statistics of genetic variation within samples were computed as average over loci using the software GEN-SURVEY (Vekemans & Lefèbvre 1997): mean number of alleles per locus, *A*; average observed heterozygosity H_O; average gene diversity H_E, computed according to Nei (1978); and Wright's inbreeding coefficient, F_{IS} corrected for small sample sizes (Kirby 1975). The overall mean value of F_{IS} was used as input for the estimation of AFLP-marker allelic frequencies (see above).

The population genetic structure of the wild genotypes was analysed using hierarchical *F*-statistics computed with the software ARLEQUIN (Schneider *et al.* 2000). The proportion of genetic variation was determined for the following components: among wild samples in the overall wild gene pool (F_{ST}); among samples with German and Belgian origins (F_{SC}), and between samples of German and Belgian origin (F_{CT}). The significance of each variance component was tested with permutation tests (Excoffier *et al.* 1992).

A model-based clustering method was applied on multilocus SSR data to infer genetic structure and define the number of clusters (gene pools) in the dataset using the software STRUCTURE (Pritchard *et al.* 2000). Individuals are assigned probabilistically to inferred gene pools, or jointly to two or more gene pools if their genotypes indicate that they are admixed. This allows the identification of hybrids between inferred gene pools and the detection of genotypes that are outliers in their sample of origin and in fact belong to another gene pool.

Table 3 Description of hairiness of inferior leaf surfaces according to Wagner (1998)

Sample	Score 0	Score 1	Score 2	Score 3
Wild Belgian				
WBM	20	8	1	0
WBVo	5	3	1	0
Wild German				
WGL	13	1	1	0
Belgian hybrids				
WBL	0	0	1	5
Edible cultivars				
ECV	0	0	0	11

Score 0, hairless; score 1, sparsely hairy; score 2, moderately hairy; score 3, felted hairy leaf surface.
Figures are number of trees with a given score for hairiness.

Furthermore, Mantel tests with 999 999 permutations were performed to check for correlation between AFLP and SSR data sets, both on the genotype and the sample level (NTSYS, Rohlf 2000). These computations are, respectively, based on 103 genotypes and seven samples, typed for both marker systems.

Results

Description of hairiness of inferior leaf surfaces

Leaf hairiness was described for the 70 apples trees (wild trees and edible cultivars) that were grafted in the nursery at DvP. The results are summarized in Table 3. The majority of putative wild trees had hairless (score 0) or sparsely hairy (score 1) leaves; only three putative wild individuals were given a score of 2 (moderately hairy). All edible cultivars and all but one hybrid had felted inferior leaf surfaces (score 3), one hybrid was given score 2.

Allelic variation at AFLP and microsatellite loci

The use of three AFLP primer combinations on 110 *Malus* genotypes resulted in 139 markers that could be scored, of which 126 (91%) were polymorphic (least common state at least 5%). All trees were characterized by a unique banding pattern. The overall reproducibility value was high: 97.51% of the markers were scored identically in the three repeats of each of five individuals tested. A negative correlation between fragment sizes and frequencies (-0.29, $P < 0.01$) was detected, which indicates that some degree of homoplasy might be present in the data set (Vekemans *et al.* 2002). Therefore, data analysis was repeated with only AFLP fragments larger than 150 bp. For this reduced data set (90 markers), correlation between fragment sizes and

frequencies was also negative (-0.17, $P = 0.10$) and differentiation values for all groups considered were similar to those obtained with the complete data set (results not shown), suggesting that the potential presence of size homoplasy of AFLP fragments does not result in underestimating genetic divergence between samples. The low average pairwise correlation value between markers of 0.014 ± 0.174 (SD) indicates that only a limited amount of information in the data set is redundant.

All 12 SSR loci analysed in this study were revealed as highly polymorphic, displaying many alleles (minimum 11 and maximum 35 alleles per locus) and a wide size range of PCR products (Table 2). More alleles and wider size ranges were recorded than in the original publications, where only a limited set of cultivated varieties and species was used.

Relationships among genotypes

Both PCO plots (based on, respectively, AFLP and SSR data, Fig. 1) show congruent groupings. Individuals from ornamental cultivars and those from edible apple cultivars clearly formed two separate groups. Most individuals from the German gene banks, from the Belgian wild samples and the three European wild individuals from the private collection (collected at different geographical locations and typed only for SSR) are positioned within a third group, suggesting that this group represents the wild genotypes of *Malus sylvestris*. Within this group, the individuals were not distributed according to their geographical origin and not even German and Belgian origins could be differentiated with either SSR or AFLP markers.

The two PCO plots also clearly indicate that the presumed hybrid trees from Limburg (WBL) are indistinguishable from edible apple cultivars. Moreover, one individual from the German gene bank (from WGL) and one individual from the Belgian forests (from WBM) seemed to be more related to the edible apple cultivars than to the *M. sylvestris* group in both PCO plots.

The main difference between the AFLP-derived and SSR-derived PCO plots is the larger heterogeneity of the groups, particularly the wild group, in the plot based on SSR data. This may be the result of the high number of SSR alleles present within this group: the number of alleles present at each locus is higher than for the edible cultivar group and the average number of alleles per locus is the highest within the wild gene pool (the average number of alleles per locus is 13, 12.5 and 6.6 within the wild, ornamental cultivar and edible cultivar group, respectively).

Inference of genetic structure and assignment of genotypes to inferred gene pools

Using the model-based clustering method of Pritchard *et al.* (2000), the highest estimate of the likelihood of the data,
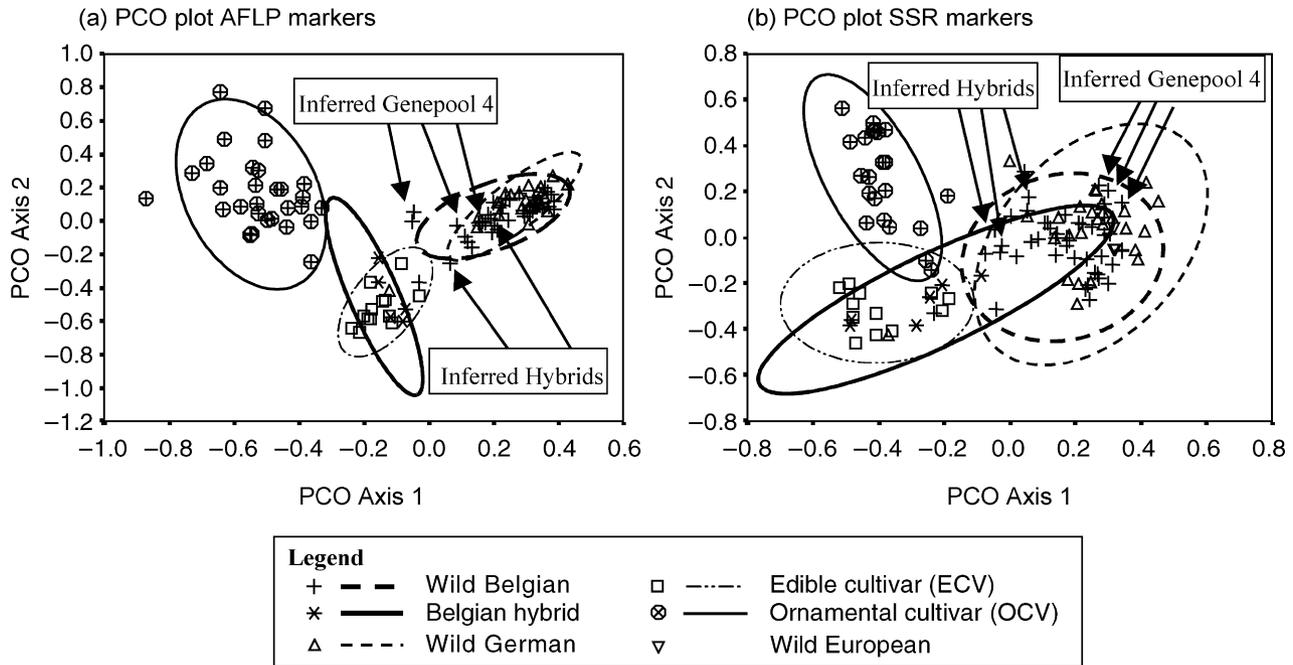
Fig. 1 Principle coordinate analysis plot of first two principal coordinates with 90% confidence ellipses for each group. (a) Calculated based on 126 polymorphic AFLP markers and using the relatedness coefficient *r* between genotypes (Lynch & Milligan 1994). Explained variation: 6% (axis 1), 4% (axis 2). (b) Calculated from 12 polymorphic SSR loci and the Moran's I relatedness coefficient between genotypes (Hardy & Vekemans 1999). Axes explain 9% and 5% of the variation. No AFLP data were available for the WBVi and WPr genotypes and hence these genotypes are not displayed in (a). Coding of samples according to Table 1. 'Gene pool 4' and 'hybrids' (gene pool 2) individuals diverging from the wild gene pool are as revealed with the clustering method (Table 4).

Table 4 Summary of the results obtained using the assignment procedure based on SSR data

| | Inferred gene pools | | | | |
	1	2	3	4	5
(a) Category of origin					
Wild Belgian apples	0.863	0.048	0.006	0.077	0.005
Belgian hybrids (WBL)	0.009	0.967	0.005	0.013	0.006
Wild German apples	0.942	0.041	0.004	0.008	0.005
Edible cultivars (ECV)	0.013	0.974	0.006	0.004	0.004
Ornamental cultivars (OCV)	0.007	0.168	0.748	0.011	0.067
(b) Hairiness					
Score 0	36*	0	0	2	0
Score 1	11	0	0	1	0
Score 2	1	3	0	0	0
Score 3	0	16	0	0	0

(a) Figures are the proportion of estimated membership to each of five inferred gene pools for genotypes of a given category of origin. Category Wild German, samples WGL and WGF; Wild Belgian, samples WBM, WBVo and WBVi.
(b) Figures are the number of trees with a given score for hairiness that were assigned to each inferred gene pool. Only 70 trees (see Table 3 for details on their category of origin) could be scored for this trait. *Two trees show admixture and are partially assigned to gene pool 1 (74 and 71%) and partially to gene pool 2 (24 and 23%).

conditional on a given number of clusters, was obtained when clustering all genotypes into five gene pools. The assignment of individuals from the different samples to these five gene pools is given in Table 4 and described below.

Gene pool 1 comprises the majority of wild individuals, both from Belgian and German origin. It is worth noting that it includes 12 wild individuals that have hairy inferior leaf surfaces (11 individuals with score 1, one individual with score 2). Although admixture of genotypes between gene pools is allowed by the method, most individuals were fully assigned to this gene pool. Strikingly three individuals (from samples WBVi, WBVo and WBM) were only partially assigned to this wild gene pool (respectively 50%, 74% and 71%) and have a significant proportion of their genetic information assigned to the gene pool of edible cultivars (respectively 38%, 24% and 23%). For two of these individuals (from WBVo and WBM) we also scored leaf hairiness and both displayed hairless leaves (score 0).

Gene pool 2 comprises all edible cultivars (ECV), all presumed Belgian hybrids (WBL), three ornamental cultivars (OCV), one Belgian individual sampled in the wild (from sample WBM), and one German individual (from sample WGL, a *M. sylvestris* gene bank). The latter two individuals were those identified as 'hybrids' in the PCO plots (Fig. 1) and are characterized by moderate hairiness scores
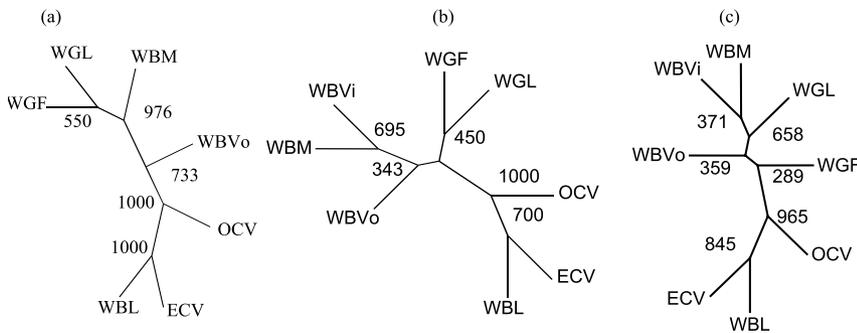
Fig. 2 Neighbour-joining trees with bootstrap support values (based on 1000 bootstraps) at forks. Calculated from (a) AFLP data and Nei's genetic distance; (b) SSR data and Nei's standard genetic distance; (c) SSR data and $\delta\mu^2$ distance. Sample codes are described in Table 1.

(score 2). All these individuals [even the presumed hybrids (WBL) and the two individuals sampled as wild] were found to be assigned completely to this gene pool, showing no evidence of genetic admixture.

Gene pool 3 comprises most of the ornamental cultivars (OCV), with the exception of three individuals assigned to gene pool 2. Four genotypes indicate admixture between the edible and ornamental cultivars with, respectively, 18, 24, 37 and 38% of their genes assigned to the edible cultivar gene pool.

Gene pool 4 consists of only three closely related wild Belgian trees (from sample WBVo), the three genotypes are assigned completely to this cluster. One individual has sparsely hairy leaves (score 1), two trees have hairless leaves (score 0).

Gene pool 5 consists of only one ornamental cultivar, *M. tschonoskii*, assigned completely to this cluster. This Japanese apple is the only individual of the studied trees that does not belong to the section *Malus* (but to the section *Docyniopsis*, classification according to Phipps *et al.* 1990).

The same clustering method was applied on AFLP data by treating each class of genotypes as being a haploid allele, but no admixture can be taken into account for dominant data. This analysis resulted in very similar clustering of genotypes (results not shown).

The correlation between the pairwise relatedness values calculated on the basis of AFLP data and SSR data was tested using a Mantel test. A highly significant positive correlation was found between both datasets ($r = 0.214$, $P = 0.0001$).

Relationship among samples

Based on the results presented in the previous section, reduced samples were defined. The two 'wild' individuals that were assigned completely to the edible cultivar gene pool were excluded from the data set. The sample of presumed hybrids (WBL) that was assigned completely to the edible cultivar gene pool was included for studying relationships between samples but was not included in the analysis of genetic diversity within and among wild samples.

Results of neighbour-joining trees based on AFLP and SSR data are shown in Fig. 2. As with the PCO plots, highly

concordant results were found based on both marker techniques. Very similar trees were also obtained with both distance measures used to analyse the SSR data (Nei standard genetic distance and $\delta\mu^2$). In all trees, samples from the wild clustered separate from cultivars (both ornamental and edible cultivars), with bootstrap support values ranging from 77 to 100%. As expected from the results presented above, only the formerly presumed hybrid trees from Limburg (sample WBL) clustered together with the edible cultivars (bootstrap support values from 70 to 100%). Both German samples clustered together in the trees based on AFLP data and SSR data (although with low bootstrap values, 55 and 45%, respectively) but only when distances were computed according to Nei. The sample of ornamental cultivars (OCV) was always positioned in between the wild samples and the edible cultivars (ECV).

Divergence between wild samples and edible cultivars was further investigated (Table 5). The differentiation among all samples was significant for both marker systems applied ($F_{ST} = 0.14$, $P < 0.0001$ for AFLP data, $F_{ST} = 0.097$, $P < 0.0001$ for SSR data). Also differentiation among wild samples only was small but significant (AFLP: $F_{ST} = 0.046$, $P < 0.0001$, SSR: $F_{ST} = 0.060$, $P < 0.0001$). Of the total differentiation among wild samples, AMOVA-analysis based on SSR data attributed 43% ($P = 0.016$) to the divergence between German and Belgian origins, while 57% ($P < 0.0001$) was explained by differentiation of samples within both origins. Divergence between the wild gene pool and edible cultivars was large based on both marker systems (AFLP: $F_{ST} = 0.19$, $P < 0.0001$, SSR: $F_{ST} = 0.11$, $P < 0.0001$).

A Mantel test on distance matrices between samples (Nei's distance) based on AFLP and SSR data was performed, resulting in a highly significant positive correlation with a much higher r-value than that based on relatedness between individual genotypes ($r = 0.872$, $P = 0.0035$).

Analysis of genetic variation within wild samples and comparison with cultivated samples

Diversity statistics are summarized in Table 6. High genetic variation at microsatellite loci was observed, with a mean number of alleles per locus equal to 7.2 (range:

Table 5 Analysis of genetic differentiation based on 139 AFLP markers and 12 SSR loci

Comparison	AFLP						SSR		
	n	H_T	H_W	H_B	F_{ST}	P-value	n	F_{ST}	P-value
(a) Among all samples	6	0.2871	0.247	0.0401	0.140	< 0.0001	7	0.0970	< 0.0001
(b) Among all wild samples	4	0.2303	0.2197	0.0106	0.0464	< 0.0001	5	0.0602	< 0.0001
(c) Between wild and edible cultivated origin	2	0.2870	0.2325	0.0547	0.1880	< 0.0001	2	0.1085	< 0.0001

(a) Among all samples except for the ornamental sample WBM, WBVo, WBL, WGL, WGF, Wvi (for SSR only) (ECV).
(b) Among wild samples only WBM, WBVo, WVi (for SSR only), WGL (WGF).
(c) Between wild WBM, WBVo, Wvi (for SSR only), WGF (WGL) and cultivated (ECV).
n, number of samples; H_T, total diversity; H_W, average diversity within populations; H_B, average diversity between populations; F_{ST}, differentiation between defined groups.

Table 6 Statistics of genetic diversity within samples of *Malus sylvestris* for AFLP and microsatellite loci

Samples	Microsatellites					AFLP			
	n	A	H_O	H_E	F_{IS}†	n	NPL	PLP	H_J
Wild samples									
WBVo	9	6.7	0.691	0.742	0.065*	9	112	80.6	0.233
WBM	28	9.9	0.706	0.810	0.127***	27	108	77.7	0.204
WBVi	7	6.2	0.708	0.784	0.102*	/	/	/	/
WGL	21	8.3	0.652	0.759	0.140***	21	98	70.5	0.208
WGF	6	4.8	0.636	0.703	0.092**	6	108	77.7	0.253
Mean		7.17	0.679	0.721	0.105***		106.5	76.63	0.225
SD		1.96	0.033	0.056	0.030		3.0	4.31	0.015
CV		0.27	0.049	0.077	0.29		0.03	0.06	0.065
WBL	6	5.4	0.659	0.709	0.134*	5	107	77.0	0.298
Cultivated samples									
ECV	12	6.4	0.729	0.775	0.058**	12	122	87.8	0.265
OCV	21	11.7	0.749	0.841	0.126***	29	116	83.5	0.287

n, number of individuals typed; A, mean number of alleles per locus; H_O, average proportion of heterozygotes; H_E, average gene diversity; F_{IS}, average inbreeding coefficient; NPL, number of polymorphic loci at the 5% level; PLP, proportion of polymorphic loci; H_J, same as H_E for AFLP markers.
†Exact test of departure from Hardy–Weinberg genotypic proportions: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, other values not significant.

4.8–9.9) and a mean gene diversity of 0.721 (range: 0.703–0.810). Very similar levels of genetic variation were found in all wild samples [coefficient of variance (CV) of gene diversity is only 7.7%]. The edible cultivar sample (ECV) and the presumed hybrid sample (WBL) have similar gene diversities (respectively 0.78 and 0.71), whereas higher diversity was detected within the ornamental cultivars (0.84). High variation at AFLP loci was also recorded, with on average 76.6% of polymorphic loci within wild samples. It should be noted however, that only loci polymorphic in the overall data set were included in the analysis. Very similar values of gene diversities (mean 0.225, range: 0.204–0.253) were found in all wild samples (CV is only 6.5%). The hybrid sample, edible and ornamental cultivar samples had higher gene diversities (respectively 0.30, 0.27 and 0.29).

The average multilocus inbreeding coefficient (F_{IS}) for all wild samples was 0.105, showing a significant ($P < 0.001$) overall departure from Hardy–Weinberg proportions with an excess of homozygotes. A significant excess of homozygote genotypes was also detected within each wild sample (Table 6).

Discussion

Comparison of results from AFLP and microsatellite marker systems

A strong congruence between the results from both marker systems was revealed. Mantel tests on relatedness matrices between genotypes or samples showed highly significant

positive correlations. This is in agreement with former observations, where similar results were obtained with AFLP and SSR at the highest taxonomic levels (Powell *et al.* 1996; Maguire *et al.* 2002). As in the study of Maguire *et al.* (2002), the correlation between pairwise genetic distance estimates from AFLP and SSR markers was lower at the intragroup level as compared to the intergroup level, but was still significant. As a result of the nature of markers derived from both systems (binary markers for AFLP, multiallelic markers for SSR), different values for genetic diversity statistics were obtained (higher for SSR than for AFLP). However, similar estimates of the relative differentiation among samples (F_{ST}) were observed (Table 5).

Despite the drawbacks of the AFLP technique, data from AFLP markers generated with only three primer combinations led to the same conclusions as data from 12 SSR loci. However, if more apple samples were to be typed and results from different laboratories were to be compared, microsatellites might be the marker system of choice because of their easy exchangeability between different laboratories and amenability for creation and management of databases. Nevertheless, the availability and congruence of the two data sets gives us more confidence in the interpretation of our results on differentiation between wild and cultivated gene pools and their level of hybridization.

Delineation of the Malus sylvestris *gene pool*

Both marker systems resulted in concordant groupings of genotypes and a gene pool of *Malus sylvestris*, clearly divergent from cultivated material, could be delineated. The high degrees of genetic similarity between *M. sylvestris* individuals from Germany, Belgium and three European individuals from a private collection further support this observation. Furthermore, the SSR-based assignment methodology applied, as well as the AFLP- and SSR-PCO plots, allowed us to identify two putative wild apple trees (one tree collected in the wild from the WBM sample and one German tree from an established gene bank, WGL sample) that most likely represent 'escaped' edible cultivars. Moreover, the model-based clustering method identified three putative hybrids between the wild and cultivated gene pools, as well as three peculiar wild genotypes that make up a different gene pool and which could not be detected on either of the PCO plots (these genotypes are marked on plots in Fig. 1). These results clearly demonstrate that a molecular approach is necessary for a conservation programme in *M. sylvestris*, to confirm the genetic identity of putative wild apple trees and to discriminate 'genuine' wild genotypes from cultivated genotypes occurring in the wild.

The results based on molecular markers are to some extent concordant with the morphological trait of leaf hairiness. Trees with a felted hairy inferior leaf surface (score 3) were assigned to the same gene pool, and represented edible cultivars and the hybrids from sample WBL. Three trees displaying moderate hairiness (score 2) were identified as derived from edible cultivars and the majority of trees with hairless leaves (score 0) were identified as *M. sylvestris* genotypes. However, one tree with a hairiness score of 2, and many trees with sparsely hairy leaves (score 1, 11 trees), were assigned completely to the wild gene pool and two genotypes showing admixture between wild and edible cultivar gene pools displayed hairless leaves. These observations suggest that some genetic variation in the degree of hairiness exists within the wild populations. Thus the degree of hairiness cannot be interpreted as a clear-cut 'degree of wildness' as speculated by Remmy & Gruber (1993) and often used by botanists as a rule of thumb in the field. Our results indicate that the degree of hairiness can be used as a first indication of the origin of an apple tree but the resolution of assignment of individuals to the wild and/or cultivated gene pool reached by molecular markers is much higher. In practice, a tree with felted hairy leaves (score 3) can be regarded as cultivated but the origin of trees with hairless or intermediate hairy leaves (score 0, 1 and 2) cannot be derived from this character only. More *M. sylvestris* genotypes will have to be studied to establish the genetic variation of this trait. In addition, more phenotypic characters should be evaluated for their discriminative value between wild and cultivated apple trees. Fruit characteristics especially, which have been the most important selection criteria throughout the apple selection process by man, may prove to be useful for identification purposes.

Occurrence of hybrids and cultivated genotypes in the wild

All presumed hybrids sampled in the wild and the two aberrant genotypes from the wild samples (classified in the edible cultivar gene pool) were not found to form an intermediate (hybrid) group, but instead completely merged into the edible cultivar gene pool. The whole study therefore detected only three genotypes displaying admixture between wild and edible cultivar gene pools. No genotypes related to ornamental cultivars or hybrids between wild and ornamental gene pools have been detected in the wild. Three putative wild individuals from the same locality (WBVo) show very different fingerprints and make up a distinct gene pool in the clustering method. No explanation can be suggested for this observation. Therefore, it can be concluded that gene flow between wild and cultivated gene pools is very rare, and that wild genotypes have not lost their genetic identity through hybridization with cultivated genotypes, despite the presence of cultivated varieties in the landscape (in plantations, in gardens, etc.) and the occurrence of escaped cultivars in the wild. The conclusion that *M. sylvestris* and *M. domestica* represent

different gene pools is further supported by isozyme studies (Wagner & Weeden 2000). Differences in flowering time between wild and cultivated genotypes might form an efficient barrier to massive hybridization, as the cultivars tend to flower later than wild apple trees at this latitude. In this respect, similar studies of spontaneous hybridization between wild and cultivated apples at lower latitudes would be worth investigating. Alternatively, the reproductive fitness of interspecific hybrids might be lower. This could be tested by comparison of the interspecific hybrids identified in this study with that of typical wild individuals under a range of conditions.

However, in this study we have only investigated the genetic composition of apple trees in the forest and no results are available on the genetic make-up of the fruits originating from these wild trees. Most of the wild apple trees in Belgian forests, however, occur as solitary trees or small groups of old trees and more fragmentation and decline of populations might have occurred since they germinated. This means that the nearest sexually compatible tree may currently be a cultivated tree and hence, the next generation may show a higher level of hybridization with edible cultivars than the level reported here. This might pose an effective threat because a preliminary investigation (unpublished results I. Wagner, based on isozyme data) indicates that the embryos of one *M. sylvestris* tree growing in nature might be of hybrid origin, but more research is needed to establish the possibilities of hybridization under natural conditions. However, no spontaneous rejuvenation of wild apple trees was detected during collection of the material in Belgian forests and fruits were only found on a few trees. Before applying forestry measures to increase fruit set and natural rejuvenation (e.g. opening the dense canopy cover surrounding the apple trees), restocking of the relict locations with wild genotypes from the region would reduce the risk that seeds from hybrid origin are formed.

Similar results of low hybridization rates between a native and introduced cultivated species were revealed for the highly endangered tree species *Populus nigra* L. (black poplar). Although no apparent fertilization barriers between the massively introduced *P. × euramericana* clones and *P. nigra* exist (Heinze 1998), very low levels of introgression in *P. nigra* are reported (Heinze 1997; Arens *et al.* 1998; Winfield *et al.* 1998; Benetka *et al.* 1999). Moreover, none of the studies found evidence for the occurrence of (introgressed) hybrids in open pollinated progenies of female trees of *P. nigra*. The currently detected low levels of gene flow from cultivated to wild gene pools are encouraging to increased conservation efforts. But a better understanding of the possibilities and barriers for spontaneous hybridization between native tree species and their cultivated relatives is necessary to develop an effective conservation programme.

Population genetic structure in Malus sylvestris

As expected for an outcrossing tree species (Hamrick & Godt 1989), a low overall differentiation among wild samples was observed (F_{ST} = 0.046 (AFLP) and 0.06 (SSR), both $P < 0.0001$). High diversity values were obtained with both marker systems for all samples. The values of overall genetic diversity obtained agreed with diversity values in other studies of outcrossing tree species from similar geographical regions (H_E = 0.72 in this study, H_E = 0.73 for *Fraxinus excelsior*, Heuertz *et al.* 2001 using SSR data; H_J = 0.225 in this study, H_J = 0.29 for *Quercus petraea* and *Quercus robur*, Coart *et al.* 2002; using AFLP data). For all wild samples a positive inbreeding coefficient was observed (mean F_{IS} = 0.11). An often reported cause of positive values of the inbreeding coefficient is the presence of null alleles (Bruford *et al.* 1998). Null alleles can be detected by studying the progeny of one parent or from a controlled cross, or by studying pedigrees. The microsatellite markers used in this study have been applied in only a few controlled crosses as far as they were polymorphic in the parents and could be mapped. CH01e12 had a null allele in cultivar Fiesta (E. van de Weg, personal communication). In a study of family trees of several apple cultivars, three out of six microsatellite markers tested (all of them different from the ones used here) had a null allele (E. van de Weg *et al.*, in preparation). So null alleles do occur and this may partly explain the shortage of heterozygotes. However, in the (multipopulation) exact test for departure from Hardy–Weinberg expectations, nine out of 12 microsatellite loci showed a significant deficiency of heterozygotes. It is not so likely that each locus has null alleles at high frequencies. Therefore, a more plausible cause for the shortage of heterozygotes detected is that the remaining individuals of wild apples in the Belgian forests studied here were collected from scattered locations that never formed one random mating population, and that this led to a Wahlund effect.

The clustering method grouped Belgian and German samples in one gene pool and the majority of differentiation between wild samples (57%) was attributed to divergence among samples from the same country. However, some level of neutral genetic differentiation was observed within and between Belgian and German origins and thus potentially differentiation for adaptive traits may have occurred.

Only a large-scale comparison of European wild genotypes, local cultivars and old landraces can help us understand better the genetic variation within *M. sylvestris* and make a finer delineation of the wild gene pool. Co-operation between European institutes is already occurring within the framework of EUFORGEN, a conservation programme for forest genetic resources (http://www.ipgri.cigiar.org/networks/euforgen). This

co-operation should be intensified to study and protect endangered species such as *Malus sylvestris*.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Arens P, Coops H, Jansen J, Vosman B (1998) Molecular genetic analysis of black poplar (*Populus nigra* L.) along Dutch rivers. *Molecular Ecology*, **7**, 11–18.

Benetka V, Mottl J, Vacková K, Pospíšková M, Dubský M (1999) Estimation of the introgression level in *Populus nigra* L. populations by means of isozyme gene markers. *Silvae Genetica*, **48**, 218–223.

Bruford MW, Ciofi C, Funk SM (1998) Characteristics of microsatellites. In: *Molecular Tools for Screening Biodiversity* (eds Karp A, Isaac PG, Ingram DS), pp. 202–205. Chapman & Hall, London.

Coart E, Vanden Broeck A, Van Slycken J (1998) *Behoud van autochtone bomen en struiken: een verkenning*. Mededelingen van het Instituut voor Bosbouw en Wildbeheer 1998–1, Institute of forestry and gence Management, Geraardsbergen, Belgium.

Coart E, De Lamote V, Loose M, Van Bockstaele E, Lootens P, Roldán-Ruiz I (2002) AFLP markers demonstrate local genetic differentiation between two indigenous oak species (*Quercus robur* L. & *Quercus petraea* (Matt.) Liebl.) in Flemish populations. *Theoretical and Applied Genetics*, **105**, 431–439.

Dumolin S, Demesure B, Petit J (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, **91**, 1253–1256.

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

Felsenstein J (1993) PHYLIP, Phylogenetic inference package, Version 3.5.7. A computer program distributed by the author, http://evolution.genetics.washington.edu/phylip.html. Departmentofgenetics, University of Washington, Seattle.

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.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Guilford P, Prakash S, Zhu JM *et al.* (1997) Microsatellites in *Malus × domestica*: abundance, polymorphism and cultivar identification. *Theoretical and Applied Genetics*, **94**, 249–254.

Haig SM (1998) Molecular contributions to conservation. *Ecology*, **79**, 413–425.

Hamrick JL, Godt MJW (1996) Effect of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London series B-Biological Sciences*, **351**, 1291–1298.

Hardy O, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Heinze B (1997) A PCR Marker for a *Populus deltoides* allele and its use in studying introgression with native European *Populus nigra*. *Belgian Journal of Botany*, **129**, 123–130.

Heinze B (1998) PCR-Based chloroplast DNA assays for the identification of native *Populus nigra* and introduced poplar hybrids in Europe. *Forest Genetics*, **5**, 31–38.

Heuertz M, Hausman J-F, Tsvetkov I, Frascaria-La Coste N, Vekemans X (2001) Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). *Molecular Ecology*, **10**, 1615–1623.

Hokanson SC, Swezc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus × domestica* borkh. core subset collection. *Theoretical and Applied Genetics*, **97**, 671–683.

Jarvis DI, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. *Molecular Ecology*, **8**, S159–S173.

King TL, Burke T (1999) Special issue on gene conservation: identification and management of genetic diversity. *Molecular Ecology*, **8**, S1–S3.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, **8**, 31–48.

Kleinschmit J, Stephan R, Wagner I (1998) Wild fruit trees (*Prunus avium, Malus sylvestris* and *Pyrus pyraster*). In: *Noble Hardwoods Network; IPGRI/EUFORGEN Report of the Second Meeting, 22–25 March 1997, Lourizan, Spain*. (eds Turok J, Collin E, Demesure B, *et al.*), pp. 51–60. (http://www.ipgri.cgiar.org/networks/euforgen/ Networks/Noble_hardwoods/Strategies/nhwildstrat.htm).

Liebhard R, Gianfrancschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus × domestica* Borkh.). *Molecular Breeding*, **10**, 217–241.

Linder CR, Taha I, Seiler GS, Snow AA, Rieseberg LH (1998) Long-term introgression of crop genes into wild sunflower populations. *Theoretical and Applied Genetics*, **96**, 339–347.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Maes N, Rövekamp C (2000) *Onderzoek naar autochtone genenbronnen in de Regionale Landschappen Vlaamse Ardennen en West-Vlaamse heuvels. Vertrouwelijke rapporten*. AMINAL, afdeling Bos en Groen, Brussel.

Maguire TL, Peakall R, Saenger P (2002) Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (*Avicenniaceae*) detected by AFLPs and SSRs. *Theoretical and Applied Genetics*, **104**, 388–398.

Maliepaard C, Alston FH, Van Arkel G, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics*, **97**, 60–73.

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

Nei M (1987) *Molecular Evolutionary Genetics*. Colombia University Press, New York.

Phipps JB, Robertson KR, Smith PG, Rohrer JR (1990) A checklist of the subfamily *Maloideae* (*Rosaceae*). *Canadian Journal of Botany*, **68**, 2209–2269.

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.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Raymond M, Rousset F (1995) GENEPOP (version12) Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Remmy K, Gruber F (1993) Untersuchungen zur Verbreitung und Morphologie des Wild-Apfels (*Malus sylvestris* (L.) Mill.). *Mitteilungen der Deutschen Dendrologischen Gesellschaft*, **81**, 71–94.

Rhymer JM, Simberloff D (1996) Extinction by hybridisation and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rohlf FJ (2000) *NTSys: numerical taxonomy and multivariate analysis system*, Version 2.1. State University of New York, New York.

Roldán-Ruiz I, van Eeuwijk FA, Gilliland TJ *et al.* (2001) A comparative study of molecular and morphological methods describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. *Theoretical and Applied Genetics*, **103**, 1138–1150.

Schneider S, Roessli D, Excoffier L (2000) Arlequin, Version 2000: A software program for population genetics data analysis. Distributed by the authors, http://anthro.unige.ch/arlequin.Geneticsandbiometrylaboratory, University of Geneva, Geneva.

Smulders MJM, Van der Schoot J, Geerts RHEM, Antonisse-de Jong AG, Korevaar H, van der Werf A, Vosman B (2000) Genetic diversity and the reintroduction of meadow species. *Plant Biology*, **2**, 447–454.

Stephan BR, Wagner I, Kleinschmit J (in press) Wild apple (*Malus sylvestris* (L.) Mill.) and pear (*Pyrus pyraster* (L.) Burgsd.), www. Ipgri.Cgiar.Org/Networks/Euforgen/Euf_Home.Htm, in press.

Vekemans X, Lefèbvre C (1997) On the evolution of the heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology*, **10**, 175–191.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Wagner I (1996) Zusammenstellung morphologischer Merkmale und ihrer Ausprägungen zur Unterscheidung von Wild- und Kulturformen des Apfel-(*Malus*) und des Birnbaumes (*Pyrus*). *Mitteilungen der Deutschen Dendrologischen Gesellschaft*, **82**, 87–108.

Wagner I (1998) Artenschutz bei Wildapfel — Die Blattbehaarung von 116 Apfelklonen auf zwei Samenplantagen. *Forst und Holz*, **53**, 40–43.

Wagner I, Weeden NF (2000) Isozymes in *Malus sylvestris*, *Malus domestica* and in related *Malus* species. *Acta Horticulturae.*, **538**, 51–56.

Winfield MO, Arnold GM, Cooper F, Le Ray M, White J, Karp A, Edwards KJ (1998) A study of genetic diversity in *Populus nigra subsp betulifolia* in the Upper Severn area of the UK using AFLP markers. *Molecular Ecology*, **7**, 3–10.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.