Combined analyses of RAPDs, cpDNA and morphology demonstrate spontaneous hybridization in the plant genus *Chaenomeles*

I. V. BARTISH, K. RUMPUNEN* & H. NYBOM

Balsgård–Department of Horticultural Plant Breeding, Swedish University of Agricultural Sciences, Fjälkestadsvägen 123–1, S-291 94 Kristianstad, Sweden

Evidence of spontaneous hybridization between two partially sympatric species of *Chaenomeles*, *C. cathayensis* and *C. speciosa*, has been obtained through analysis of offspring families from these two species, as well as from two presumed interspecific hybrid populations. A combination of different methods was applied. Analysis of diagnostic RAPD markers and of chloroplast DNA haplotypes supported the notion of spontaneous hybridization, and suggested that there has been symmetrical, rather than unidirectional, introgression between *C. cathayensis* and *C. speciosa*. RAPDs and morphological characters revealed concordant patterns of genetic relatedness among the studied offspring families. Some putative hybrid families had mainly intermediate characters, whereas others appeared to be later generation hybrids as they were genetically and phenotypically rather similar to families that appeared to represent pure species. The RAPD-based proportion of between-family variability was considerably higher in the putatively hybridogenous populations than in populations of the pure species. Within-family gene diversity estimates ranged from *C. speciosa* (max. $H_j = 0.235$) to *C. cathayensis* (min. $H_j = 0.094$) with the presumed hybrid families taking intermediate values.

Keywords: *Chaenomeles*, cpDNA, interspecific hybridization, Maloideae, morphology, RAPD.

Introduction

The fact that interspecific hybridization and introgression are common phenomena in many plant groups has long been recognized, as well as the need to study these phenomena in theoretical and applied plant research, e.g. to clarify evolutionary trends (Anderson, 1949) and to determine sources of valuable diversity (Allard, 1960). However, thorough molecular documentation is available only for a restricted number of genera, reviewed for animals by Avise (1994) and for plants by Rieseberg *et al.* (1996b). A detailed study of speciation via introgression indicated mechanisms such as selection for coadapted gene complexes and selection against chromosomal rearrangements (Rieseberg *et al.*, 1995).

The subfamily Maloideae (Rosaceae), comprising approximately 940 species in 28 genera (Robertson *et al.*, 1991) and a further 33 species representing 13 hybrid genera (Phipps *et al.*, 1990), is notorious for its variability in reproductive and breeding systems. Extensive hybridization at generic and specific levels, as well as polyploidy, is remarkably frequent (Campbell *et al.*, 1991). Hybrid complexes in Maloideae, including artificially produced hybrids, have been studied mainly by traditional morphological and cytological methods (Dickinson & Campbell, 1991; Phipps *et al.*, 1991; Robertson *et al.*, 1991). Molecular evidence of natural hybridization has been reported for *Malus* (Dickson *et al.*, 1991) and for *Amelanchier* (Campbell *et al.*, 1997). In an allozyme-based study of interspecific hybridization within groups of several species of *Malus*, native to North America, and the domesticated apple (*M.* × *domestica*), Dickson *et al.* (1991) observed little gene flow from the domesticated apple into native *Malus*. However, in support of earlier conclusions based on morphology, two species, *M. platycarpa* and *M. dawsoniana*, appeared to be the result of interspecific hybridization involving the introduced *M.* × *domestica* and some native species. In *Amelanchier*, phylogenetic analysis of nuclear ribosomal DNA sequence

polymorphisms demonstrated that five eastern North American taxa in a group of hybrid microspecies differed at many nucleotide sites (Campbell *et al.*, 1997). *Amelanchier humilis* or *A. sanguinea* were proposed to represent one of the parents of this group, and various members from a sister phylogenetic clade to represent the other parent. Polyploidy and/or agamospermy, which are prevalent in the genus, were hypothesized to be responsible for stabilization of the detected sequence polymorphisms.

Although intergeneric hybrids are of common occurrence in the Maloideae, none is described with *Chaenomeles* as one of the parents, in spite of the shared chromosome number ($2n = 34$) and similar chromosome morphology within the subfamily (Weber, 1964). Moreover, spontaneous intrageneric hybridization has also not been reported in this genus, which is native to China (*C. cathayensis*, *C. speciosa* and *C. thibetica*) and Japan (*C. japonica*). By contrast, interspecific hybrids have been produced artificially in all possible directions between three of the species (*C. cathayensis*, *C. japonica*, and *C. speciosa*) which have been introduced to Europe and North America, with the main objective to develop new ornamental cultivars (Weber, 1963). Meiosis is regular in the species as well as in the hybrids (Weber, 1964).

In nature, only *C. cathayensis* and *C. speciosa* are known to grow sympatrically but detailed information regarding the distribution and ecology is still lacking. The species grow on open hillsides, rocky slopes and ravines in thickets and at the edge of forests. Outcrossing has been suggested for both species, although without documented proof (Weber, 1964). Honeybees and bumblebees pollinate the flowers. Seed dispersal is supposed to take place through birds and rodents, which ignore the very acid fruit flesh and instead eat the numerous seeds. As a consequence of the medicinal use of dried fruits of *C. speciosa* and the use of *C. cathayensis* for making vinegar, man is also a potential vector for seed dispersal, collecting fruits in the wild and cultivating small-scale orchards. Obviously there is a contradiction between the virtual absence of reproductive isolation barriers in artificial crossings between *C. cathayensis* and *C. speciosa* and the absence of described natural interspecific hybrids. Several explanations for this may be offered: (i) insufficient sampling in putative hybrid zones; (ii) strong selection against F1 hybrids, which eliminates individuals with intermediate phenotypes; and (iii) inability of morphological characters to reveal later-generation hybrids (Rieseberg & Ellstrand, 1993).

In a recent taxonomic study of the genus *Chaenomeles* (Bartish *et al.*, 2000), both isozyme and RAPD data suggested that two of the analysed populations had been derived through interspecific hybridization between *C. cathayensis* and *C. speciosa*. In another study,

preliminary analysis of chloroplast DNA from four species of *Chaenomeles*, with three specific primer pairs (*trn*Lc – *trn*Ff, Taberlet *et al.*, 1991; *trn*H – *trn*K and *trn*S – *trn*fM, Demesure *et al.*, 1995) revealed a mutation which could discriminate between *C. japonica* and *C. speciosa*, and between *C. cathayensis* and *C. thibetica* (Bartish, unpubl. results). The mutation appears to be a deletion of an approximately 100 bp fragment from an approximately 1600 bp amplification product, directed by primers *trn*H and *trn*K.

In the present investigation, we re-analysed the previously obtained RAPD data set to study within-family marker frequencies in five populations, including the parental species as well as the putative hybrids, to further corroborate our hypothesis. In addition, we analysed chloroplast DNA polymorphism as well as numerous morphological characters in the same five populations.

Materials and methods

Seeds labelled as either *C. cathayensis* or *C. speciosa* were obtained from four different areas of Yunnan province, China, with the co-operation of Chinese botanists. These accessions were treated as five different populations (one area with samples of both species), P1–P5 (Table 1). Populations P4 and P5 were collected in the wild (one fruit per sampled genotype), whereas populations P1, P2 and P3 were obtained from fruits purchased at the local markets but collected in nearby mountains. The seeds from one fruit were presumed to constitute a half-sib family. Seeds were stratified ($+ 4°C$) for two months and germinated in a greenhouse in February 1998. In some populations, a high number of albino seedlings was obtained (up to 44% in one family), which all died within a few weeks. The numbers of albinos and of normal seedlings, respectively, were noted separately. Three families per population, with seven normal seedlings in each family, were then sampled for molecular and morphological analyses (in population P1 two plants died after transplanting, resulting in a total of 103 genotypes). Our intention was to take these families randomly among the seedlings available (Table 1). However, putative interspecific hybrids with heterosis-induced vigour may inadvertently have been preferred as we chose families with healthy, vigorously growing seedlings that could be used for DNA isolation as soon as possible.

Analysis of morphology

Plants were kept in a greenhouse during the first growing season. Stem characters (Table 2) were measured in the autumn when plants had entered dormancy.

Table 1 Number (*n*) of available and collected families, identity of analysed families, number of analysed individuals, species determination and locality of the five *Chaenomeles* populations investigated

Families[a] *n* available (*n* collected)	Families analysed	Individuals *n* analysed	Species[b]	Locality (co-ordinates and altitude)[c]
15 (27)	P1–8, 9, 14	19	*C. speciosa*	Zhenyuan, Yunnan, China (23°50′N, 101°10′E, 1800–2300 m)
10 (12)	P2–1, 2, 11	21	Hybrid?	Hutiaoxia, Zhongdian, Yunnan, China (27°50′N, 99°30′E, 1800–2100 m)
3 (3)	P3–1, 2, 3	21	*C. cathayensis*	Hutiaoxia, Zhongdian, Yunnan, China (27°50′N, 99°30′E, 1800–2100 m)
22 (26)	P4–10, 11, 25	21	*C. cathayensis*	Caojian, Yunlong, Yunnan, China (25°35′N, 98°05′E, 2500–2600 m)
14 (18)	P5–8, 9, 14	21	Hybrid?	Dali, Yunnan, China (25°35′N, 100°10′E, 1900–2400 m)

[a]Only families with at least 7 seedlings were included in 'available'.
[b]Species identity of samples inferred from leaf morphology of seedlings (Weber, 1964), accesssions of 'Hybrid?' populations originally labelled as *C. speciosa*.
[c]Geographical coordinates broadly estimated.

Plants were kept in a frost-free greenhouse during winter. At the end of April, leaves were sampled from long shoots of the annual growth, preferably from the main stem or else from nonshaded erect shoots. Three fully expanded leaves were sampled from nodes 7 to 9, the first node counted 1 cm below the actively growing shoot apex. The leaves were pressed, digitized by scanning (HP ScanJet 6100C), and characterized (Table 2) by use of the MacFOLIA[TM] Pro software (Regent Instruments Inc, Canada). For each character, an average was calculated for the three leaves from each genotype. Finally, several combined characters were calculated (Table 2), including SIFF, which is a commonly used shape parameter measuring the deviation from a circle, and SIAL which measures the deviation from a rectangle of the same width and length as the leaf (West & Noble, 1984).

DNA extraction and RAPD analysis

DNA extraction and PCR amplification followed Bartish *et al.* (1999a). Based on reproducibility and polymorphism, 10 primers were selected out of one hundred ('Operon Technologies' sets A, B, D, E, F), which had been screened in a previous study on genetic diversity in *Chaenomeles* (Bartish *et al.*, 1999b).

Chloroplast DNA RFLP analysis

Chloroplast DNA primers *trn*H and *trn*K (Demesure *et al.*, 1995) were used to amplify DNA samples, extracted as above, from one seedling per family, i.e. 15 seedlings in total. The reaction mixtures were prepared as in our RAPD analysis, except that: the

concentration of primers was 0.2 μM; and the reaction was carried out in a 15 μL reaction mixture. The reaction mixture was subjected to amplification at 94°C for 3 min, 35 cycles at 94°C for 45 s, 58°C for 1 min and 72°C for 1 min 30 s followed by an 8-min extension time at 72°C. The amplified fragments were restricted by *Hin*d III endonuclease (1 u μL^{-1}) and separated by electrophoresis (90 min) on a 1.8% agarose gel containing ethidium bromide and using TPE buffer. The gel was photographed under UV light.

Data analysis

Morphological characters were analysed using the SPSS v.4.0 software package (SPSS Inc., Chicago, Illinois, USA) using the MANOVA and DISCRIMINANT procedures to investigate the differentiation among and within the studied populations and families.

Binary data matrices were produced from the RAPD marker data for individual plants. Within-family variation was calculated as the unbiased expected heterozygosity (H_j) according to Lynch & Milligan (1994), as in Bartish *et al.* (2000), and as Shannon's Index (H'_j) according to Bussell (1999), but with family as group variable. Variation among families was calculated as G'-statistics (Bussell, 1999) and as Φ-statistics using AMOVA (Huff *et al.*, 1993).

Null allele frequencies were calculated on the RAPD data set for each family separately in accordance with the formulae of Lynch & Milligan (1994). The resulting data set was used to compute Nei's genetic distances (Nei, 1972) between families. Standardized mean per family estimates of morphological

Table 2 Definition of characters used to estimate morphology in *Chaenomeles*

Characters	Definitions
A20[a]	leaf base angle at 20% of LENGTH (°)
A30[a]	leaf base angle at 30% of LENGTH (°)
APEX[a]	W90/W50 (calculated ratio)
AREA[a]	leaf area (cm)
ATH[a]	average tooth height (mm)
ATW[a]	average tooth width (mm)
BASE[a]	A20/A30 (calculated ratio)
BLEN	average branch length calculated as BTOT/BNUM (cm/branch)
BNUM	total number of first order branches (count)
BSPI	number of spines on first-order branches (count)
BTOT	total length of first-order branches (cm)
LENGTH[a]	maximum leaf length, including the petiole (cm)
MAXP[a]	PMBW/LENGTH (calculated ratio)
MBW[a]	maximum leaf blade width (cm)
MWL[a]	MBW/LENGTH (calculated ratio)
NOFT[a]	number of leaf teeth (count)
PERM[a]	leaf perimeter (cm)
PHGT	maximum vertical plant height (cm)
PMWB[a]	position of maximum leaf blade width (cm)
SDIA	stem diameter (mm)
SIAL[a]	leaf shape index calculated as AREA/LENGTH MBW
SIFF[a]	leaf shape index calculated as 4π AREA/PERM^2
SLDM	SLEN/SDIA (calculated ratio)
SLEN	total length of main stem (cm)
SPI	spininess calculated as (SSPI + BSPI)/ (SLEN + BTOT) (spines/cm)
SSPI	number of spines from nodes on the main stem (count)
W50[a]	leaf blade width at 50% of LENGTH (cm)
W90[a]	leaf blade width at 90% of LENGTH (cm)

[a]characters measured on digitized leaves.

characters were used to calculate Euclidean distances between families. The SAHN program from NTSYS-pc v.1.8 software (Rohlf, 1997) was then used to produce UPGMA dendrograms from each of the distance matrices, i.e. (i) Nei's genetic distances based on RAPD, and (ii) Euclidean distances based on morphology. A principal co-ordinate analysis (using DCENTER and EIGENVECTOR programs) was run for each of the distance matrices, and the results were viewed in 3D plots.

Pearson product–moment correlation analyses were performed between estimates of gene diversity obtained through different methods of data analysis (H_j and H'_j), and between gene diversity and percentage of germination within families.
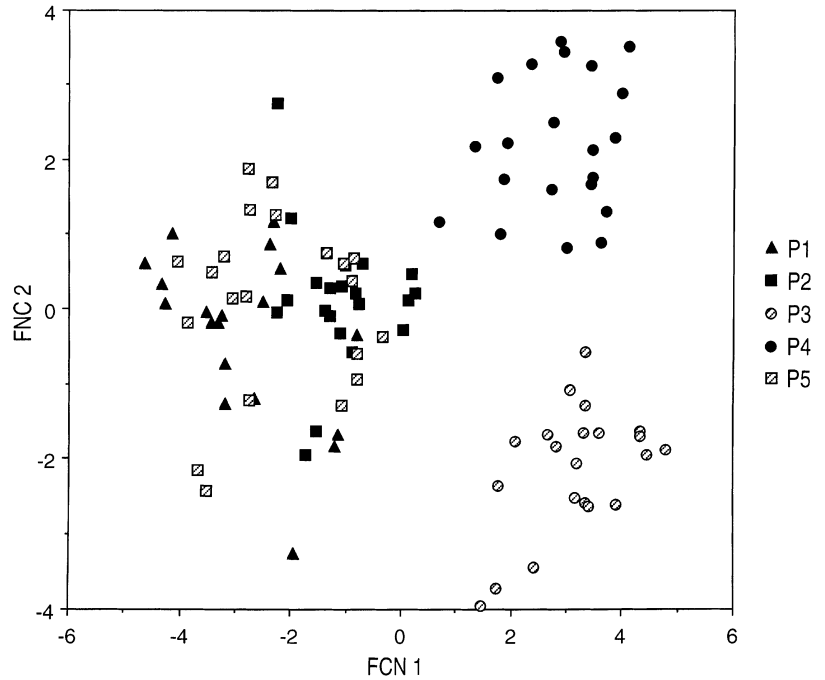
Results

Discriminant analysis based on morphology

A multivariate analysis of variance (MANOVA) of the morphological characters resulted in highly significant values, both at the population (Wilks' lambda = 0.02218, $P < 0.0001$) and at the family (Wilks' lambda = 0.00021, $P < 0.0001$) level. The efficiency of each character in separating the group means for populations and families, respectively, was estimated by a univariate F-test based on Wilks' lambda (Table 3). Leaf serration (NOFT) and plant height (PHGT) were highly efficient both at population and family levels. Spininess (SPI) and shape of leaf apex (APEX) were the least efficient characters for separating populations, as were APEX and average branch length (BLEN) for separating families.

A discriminant analysis based on morphological characters, with population as group variable, clearly separated the *C. cathayensis* populations P3 and P4 from one another and from a combined cluster of P1, P2 and P5 (Fig. 1). Two extracted functions together explained 92.0% of the variation, whereas three functions explained 97.1%. A reclassification test, comparing the actual population membership with the predicted population membership based on the discriminant analysis, resulted in overall 82.5% correctly classified

Table 3 Test of univariate equality of group means to estimate the efficiency of different characters in separating the groups. Low values for Wilks' lambda indicate that the within-group variability is small compared to the total variability, thus most of the total variability is attributable to differences between the means of the populations and families, respectively. All values are highly significant ($P < 0.0001$). Values of Wilks' lambda corresponding to the most efficient characters are printed in bold

|	Wilks' lambda	
Character	Populations	Families
APEX	0.772	0.566
ATH	0.709	0.437
ATW	0.562	0.360
BASE	0.526	0.481
BLEN	0.702	0.551
MAXP	**0.461**	0.368
MWL	0.734	0.304
NOFT	**0.418**	**0.135**
PHGT	**0.446**	**0.254**
SIAL	0.699	0.523
SIFF	0.607	**0.287**
SLDM	0.551	0.407
SPI	0.786	0.444

Fig. 1 Two-dimensional plot of the result of a discriminant analysis based on the morphological data set with populations as groups and genotypes as cases. The different populations are represented by triangles (*C. speciosa*), circles (*C. cathayensis*) and squares (hybrids). Percentage explained variation for the extracted functions: FCN1 70.8% and FCN2 21.2%.

cases. From this analysis, the two putative hybrid populations (P2 and P5) were the least well-defined, with only 76.2 and 57.1% correctly classified cases. The corresponding values for populations P1, P3 and P4 were 84.2, 100.0 and 95.2%.

A discriminant analysis based on morphological characters, with family as group variable, resulted in a separation of the families in the putative hybrid popu-lations P2 and P5 into three different groups, whereas the families within the *C. cathayensis* populations P3 and P4 grouped together (Fig. 2). Family P5–14 was placed rather far away from the two major groups of *C. speciosa* (P1) and *C. cathayensis* (P3 and P4), whereas families P5–9, P2–2 and P2–11 fell into a separate group closer to the *C. cathayensis* cluster. Two extracted functions together explained 77.9% of the variation,
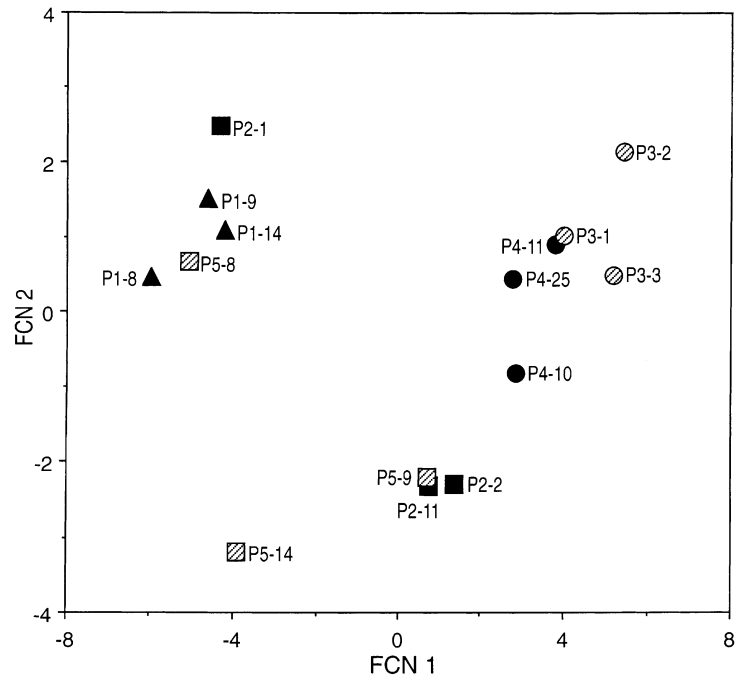
Fig. 2 Two-dimensional centroid plot of the result of a discriminant analysis based on the morphological data set with families as groups. The different popu-lations are represented by triangles (*C. speciosa*), circles (*C. cathayensis*) and squares (hybrids). Percentage explained variation for the extracted functions: FCN1 65.9% and FCN2 12.0%.

whereas three functions explained 86.2%. A reclassification test of the families resulted in 83.5% correctly classified cases. Misclassification for populations P3 and P4 (*C. cathayensis*) occurred only among families within the correct population. For population P1 (*C. speciosa*) only one genotype was misclassified to another population, whereas in the same way three genotypes were misclassified for populations P2 and P5, respectively. Discriminant analysis with families as group variable thus yielded at least as good or even better resolution than when population was used as group variable.

Diagnostic RAPDs

One marker (F04.1250) occurred only in population P3 (*C. cathayensis*) and was therefore diagnostic for this population. Seventeen markers out of the total set of 222 were diagnostic between population P1 (*C. speciosa*), and populations P3 and P4 (*C. cathayensis*) (Table 4). Seven of these markers were present in all three families of population P1 (19 individual genotypes) and absent in all six families of populations P3 and P4 (42 individual genotypes). The reverse distribution pattern was observed for 10 other markers.

Families in populations P2 and P5 (putative hybrid populations originally labelled as *C. speciosa*, Table 1) showed rather complicated patterns for the above-mentioned diagnostic markers (Table 4). Deviation by one or a few markers from the otherwise dominating type (either *C. cathayensis* or *C. speciosa*, as hypothesized from cpDNA data, Table 5) can be considered as an introgression of alien loci into the genetic background of one of the species. We observed several such cases in family P2-1 (which is of *C. speciosa* type but with five introgressed markers of *C. cathayensis* type, of which two were fixed, in family P5-08 (*C. speciosa* type but with seven introgressed markers of *C. cathayensis* type, of which two were fixed), in families P2-2 (*C. cathayensis* type but with five introgressed markers of *C. speciosa* type, of which four were fixed) and P5-9 (*C. cathayensis* type but with five introgressed markers of *C. speciosa* type, of which four were fixed). Even more numerous cases of putative marker introgression were noticed for the two remaining families, which may, in fact, represent F1 hybridizations; for family P2-11 (*C. cathayensis* type but with nine introgressed markers of *C. speciosa* type, of which six were fixed) and family P5-14 (originally considered as a *C. speciosa* type based on morphology and cpDNA with 13 introgressed

Table 4 Frequencies of diagnostic RAPD markers within families. Families are ordered from predominantly *C. speciosa* types to predominantly *C. cathayensis* types (with direction of hybridization hypothesized from cpDNA data). Apparently introgressed markers are printed in bold

RAPD markers	C. speciosa			Putative hybrids						C. cathayensis					
	P1-8	P1-9	P1-14	P2-1	P5-8	P5-14	P2-11	P2-2	P5-9	P4-10	P4-11	P4-25	P3-1	P3-2	P3-3
C. speciosa type															
D08.1440	1	1	1	1	1	1	**1**	0	0	0	0	0	0	0	0
E07.700	1	1	1	1	1	**0**	0	0	0	0	0	0	0	0	0
E07.770	1	1	1	1	**0.86**	**0**	0	0	0	0	0	0	0	0	0
B07.570	1	1	1	**0.86**	**0.71**	**0.71**	**1**	**1**	**1**	0	0	0	0	0	0
E07.270	1	1	1	**0.71**	**0.43**	**0**	0	0	0	0	0	0	0	0	0
E07.760	1	1	1	**0**	**0**	**0**	0	0	0	0	0	0	0	0	0
E07.800	1	1	1	**0**	**0**	**0**	0	0	0	0	0	0	0	0	0
C. cathayensis type															
A18.430	0	0	0	**0.29**	**0.29**	**0.29**	**0.86**	1	1	1	1	1	1	1	1
A18.830	0	0	0	0	**0.14**	**0.57**	**0**	**0**	**0**	1	1	1	1	1	1
E07.690	0	0	0	0	0	**1**	1	1	1	1	1	1	1	1	1
E07.320	0	0	0	0	0	**0.71**	1	1	1	1	1	1	1	1	1
E07.1000	0	0	0	0	0	**0.57**	**0.86**	1	1	1	1	1	1	1	1
D08.410	0	0	0	0	0	**0.57**	**0**	**0**	**0**	1	1	1	1	1	1
E07.900	0	0	0	0	0	**0.43**	**0.29**	**0**	**0**	1	1	1	1	1	1
E07.780	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
D08.1480	0	0	0	0	0	0	**0**	1	1	1	1	1	1	1	1
D08.1700	0	0	0	0	0	0	**0**	**0.14**	**0.29**	1	1	1	1	1	1

Table 5 Gene diversity estimated as expected heterozygosity (H_j), and Shannon's Index of diversity (H'_j) within families in rank order as well as percentage germination and percentage of albino seedlings (SE = standard error)

Family	cpDNA (haplotype)	H_j	SE	Rank	H'_j	SE	Rank	Seeds (n)	Germination (%)	Albinos (%)
P1-14	II	0.235	0.016	1	0.479	0.027	1	98	27	15
P1-8	II	0.222	0.015	2	0.450	0.027	2	105	28	10
P1-9	II	0.218	0.015	3	0.387	0.026	5	65	45	21
P5-8	II	0.200	0.015	4	0.395	0.026	4	96	38	44
P5-14	II	0.192	0.015	5	0.402	0.026	3	97	36	11
P2-1	II	0.180	0.015	6	0.355	0.025	6	50	68	18
P2-11	I	0.162	0.014	7	0.295	0.024	7	125	19	0
P4-11	I	0.137	0.013	8	0.247	0.022	8	83	74	0
P4-25	I	0.126	0.013	9	0.196	0.021	9	89	64	0
P2-2	I	0.124	0.013	10	0.187	0.021	10	63	82	0
P5-9	I	0.109	0.012	11	0.143	0.018	12	189	—	—
P4-10	I	0.108	0.012	12	0.169	0.020	11	150	73	0
P3-3	I	0.096	0.012	13	0.120	0.015	15	162	85	0
P3-1	I	0.095	0.012	14	0.135	0.017	13	163	97	0
P3-2	I	0.094	0.012	15	0.132	0.016	14	131	95	0

markers of *C. cathayensis* type, of which six were fixed). Coherent introgression of the same diagnostic markers was evident, even among families sampled from different geographical areas. It is remarkable that two random decamer primers (E07 and D08), out of the total set of eight primers used in the present study, were responsible for 82% of the diagnostic markers (Table 4).

Genetic relatedness among families

Whereas the described diagnostic RAPD markers are useful for differentiating accessions of *C. cathayensis* and *C. speciosa*, genetic relatedness between families in the entire set of studied populations (P1–P5) cannot be inferred from these markers alone. Therefore, the complete data set of 222 RAPD markers was used for this analysis, as well as the morphological characters. Relatedness between all 15 families from the five populations was represented in UPGMA dendrograms (Figs 3 and 4). These dendrograms are consistent in: (i) clustering of all families from population P1 (*C. speciosa*) and families P2-1 and P5-8 in one group with family P5-14 more loosely associated, and (ii) clustering of families from populations P3, P4 (both *C. cathayensis*) and families P2-2, P2-11 and P5-9, respectively, into three different but closely positioned groups. The RAPD-based dendrogram is however, more efficient in: (i) separation of families from population P1 and families P2-1 and P5-8 into different groups, and (ii) revealing a close relationship between populations P3 and P4.
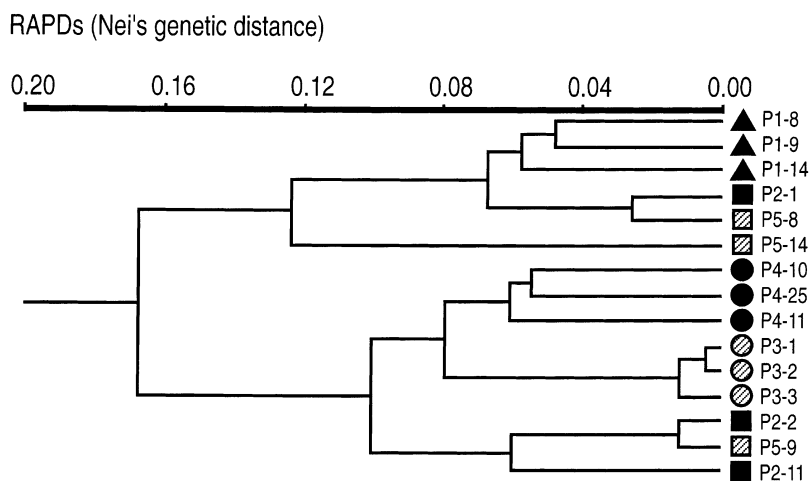
Fig. 3 UPGMA dendrogram of a RAPD-based data set (Nei's genetic distances) for families of *Chaenomeles*. The different populations are represented by triangles (*C. speciosa*), circles (*C. cathayensis*) and squares (hybrids).

Morphology (Euclidean distance)
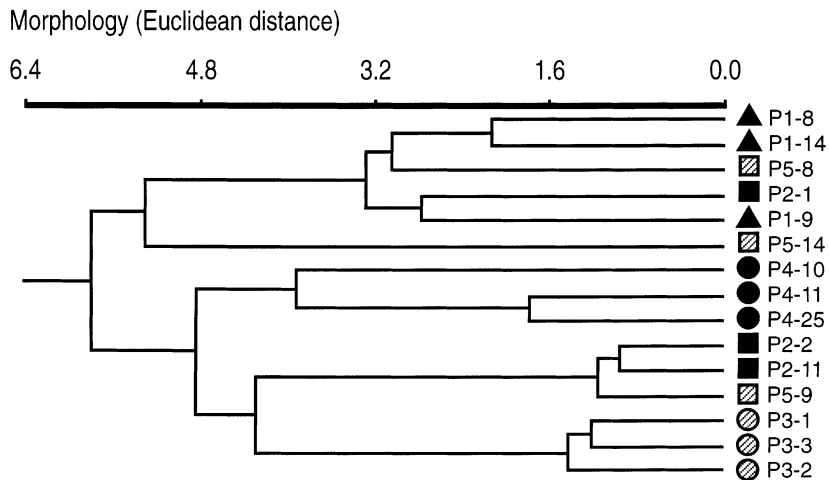
Fig. 4 UPGMA dendrogram of a morphological data set (Euclidean distances) for families of *Chaenomeles*. The different populations are represented by triangles (*C. speciosa*), circles (*C. cathayensis*) and squares (hybrids).

The principal coordinate analyses yielded results rather similar to the UPGMA dendrograms (results not shown). There were two main groups: population P1 together with families P5-8 and P2-1, and populations P3 and P4 together with families P2-2, P2-11 and P5-9. Family P5-14 again occupied an intermediate position.

The correlation between matrixes of pairwise distances based on RAPD-data (Nei's genetic distance) and on morphology-data (Euclidean distance) was $r = 0.79$ ($P < 0.0001$).

Analysis of cpDNA polymorphism

Analysis of chloroplast DNA resulted in two haplotypes, I and II, which revealed a pattern consistent with the genetic relatedness analyses (Table 5). Families P5-8, P5-14 and P2-1 had the same haplotype (II) as families from population P1 (*C. speciosa*) whereas families P2-2, P2-11 and P5-9 had the same haplotype (I) as families from populations P3 and P4 (*C. cathayensis*).

Among- and within-family variability

Although G'-statistics and Φ-statistics (based on Shannon's Index and on AMOVA, respectively) differed considerably in absolute values, they nevertheless revealed a relatively consistent pattern (Table 6). Thus both statistical analyses found the among-families component of variance to be highest in the two putative hybrid populations P2 and P5, whereas it was lowest in population P3 (*C. cathayensis*). However, the ranking of populations P1 and P4 differed between the two methods.

The Lynch and Milligan expected heterozygosity (H_j) and Shannon's Index (H'_j) values for within-family gene diversity were closely associated (Pearson coefficient of correlation: $r = 0.99$, $P < 0.0001$). Ranking of families

Table 6 Partitioning of genetic variability between families within populations, estimated for the RAPD data set as G'-statistics (derived from Shannon's Index) with standard error (SE) and Φ-statistics (AMOVA) with P-value. Values in bold refer to the largest estimates of variability

Population	G'	(SE)	Φ	P-value
P1	0.520	(0.024)	0.193	(0.0001)
P2	**0.740**	(0.032)	**0.530**	(0.0001)
P3	0.182	(0.020)	0.025	(0.1965)
P4	0.289	(0.022)	0.371	(0.0001)
P5	**0.630**	(0.027)	**0.530**	(0.0001)

by their relative diversity showed that families from population P1 (*C. speciosa*) were the most diverse, and families from population P3 (*C. cathayensis*) the least diverse (Table 5). Families with chloroplast DNA haplotype II were consistently more diverse than families with haplotype I. Families P2-11 and P4-11 had haplotype I, but their diversity values were still considerably higher compared to families from population P3 (Table 5).

Rate of germination and albino seedlings

Within-family rates of seed germination (except family P5-09 for which data are missing) varied from 19 to 97% (Table 5). These rates were negatively correlated with the two estimates of within-family gene diversity ($r = -0.86$, $P < 0.0001$ for both H_j and H'_j), suggesting that more homogeneous families have higher germination rates. It should also be noticed that all but one of the lower germination rates was found in families with chloroplast DNA haplotype II. Furthermore, albino seedlings were very common (10–44%) in families with this haplotype, whereas they were completely absent in families with chloroplast DNA haplotype I.

Discussion

The high discriminatory power of leaf morphology characters, revealed in our study, had previously been noted in *Chaenomeles* by Weber (1964), who also recognized their potential in determining species hybrids. However, had we relied only on morphological characters at the population level, we would not have been able to find clear evidence for hybridization between *C. cathayensis* and *C. speciosa*. Instead, a combination of morphological studies of offspring families and molecular data, have shown that two of our populations are, in all likelihood, derived from interspecific hybridization. A reason that natural hybrids have not previously been reported in *Chaenomeles* may be connected with the well-known mechanism of selection against first-generation hybrids (Rieseberg *et al.*, 1995), and the difficulty in detecting the subtle signs of later-generation introgression using morphological characters alone.

Putatively neutral RAPDs and morphological characters, which may be subjected to natural selection, revealed surprisingly consistent patterns of relatedness between families. However, it is premature to infer the possible influence of natural selection on the process of hybridization between *C. cathayensis* and *C. speciosa* because the plant sampling was not designed specifically for this purpose.

Some introgressed diagnostic markers revealed a highly correlated pattern of fixation and polymorphism, irrespective of the putative direction of gene flow or the geographical origin of the analysed hybrid families. This may indicate either tight genetic linkage between different sets of the diagnostic RAPDs, or selection for coadapted gene complexes in genetic backgrounds of both species. The latter mechanism has been suggested by Rieseberg *et al.* (1996a) in their study of interspecific hybrids in *Helianthus*. Breeding system can also be an important factor in determining whether markers remain correlated in hybrid zones (Rieseberg & Ellstrand, 1993). However, at the moment, we do not have any experimental data to determine breeding system of our plant material.

Unilateral (directional) introgression of both nuclear and chloroplast DNA is a quite frequent and well-known phenomenon (reviewed by Rieseberg *et al.*, 1996b). This kind of introgression does not, however, seem to prevail in the present study. Both chloroplast DNA haplotypes, otherwise diagnostic for the two studied species, were found within the presumed hybrid populations P2 and P5. We also found that the RAPD markers were associated with the correspondent chloroplast DNA haplotype, so that there were no cases where the chloroplast genome of one species was combined with nuclear DNA markers of mainly the other species. Only two cases (P5-14 and P2-11) represented intermediate nuclear genotypes.

Our results show a strong negative correlation between percentage of seed germination and gene diversities within families: the more diverse the family, the lower the percentage seed germination. It is remarkable that family P2-11, which is the most heterogeneous of the families with a *C. cathayensis* cpDNA haplotype, at the same time had the lowest seed germination among all families studied. Moreover, we found an absolute concordance between the percentage of (inviable) albinos and the cpDNA haplotype. Albino plants were thus always present in families with the *C. speciosa* haplotype and never present in families with the *C. cathayensis* haplotype. The chloroplast genome in *Chaenomeles* is probably uniparental in origin and maternally inherited like as in most other angiosperms. Consequently, an active selection/linkage mechanism of postzygotic isolation between *C. speciosa* and *C. cathayensis* is indicated.

The distribution of molecular markers and the patterns of genetic relatedness between families indicate that at least four of the six hybrid families included in our study represent later-generation backcrosses. Similarly, Rieseberg & Ellstrand (1993) emphasize in their review on interspecific hybridization in plants, that hybrid zones are seldom limited solely to F1s. Furthermore, substantial evidence that chromosomal and genic sterility barriers are responsible for strong selection against F1 hybrids has been presented (reviewed by Rieseberg & Carney, 1998).

At this stage we cannot distinguish between genetic and ecological factors, which may have created the extant pattern of introgression in the sampled populations. Because *C. speciosa* is a traditional medicinal plant, and both species are cultivated to a minor extent, there is a possibility that man has influenced the natural pattern of gene flow. However, the selection/linkage hypothesis (Rieseberg *et al.*, 1996b) seems to be the most plausible explanation to our findings. Experimental crosses using plants from the families P2-11 and P5-14, which appear to be the most intermediate between *C. cathayensis* and *C. speciosa*, as one parent, and plants from the other families as the other parent, could provide estimates of genetic linkages between scored RAPD markers. Comparison of linkage maps of different species and interspecific hybrids could be useful in studying genetic mechanisms involved in the putative selection against interspecific hybrids (Rieseberg *et al.*, 1995). Such analyses could also provide additional information about mechanisms of speciation via interspecific hybridization and introgression.

Acknowledgements

References

ALLARD, R. W. 1960. *Principles of Plant Breeding*. Wiley & Sons, New York.

ANDERSON, E. 1949. *Introgressive Hybridisation*. Wiley & Sons, New York.

AVISE, J. C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

BARTISH, I. V., JEPPSSON, N. AND NYBOM, H. 1999a. Population genetic structure in the dioecious pioneer plant species *Hippophae rhamnoides* investigated by random amplified polymorphic DNA (RAPD) markers. *Mol. Ecol.*, **8**, 791–802.

BARTISH, I. V., RUMPUNEN, K. AND NYBOM, H. 1999b. Genetic diversity in *Chaenomeles* (Rosaceae) revealed by RAPD analysis. *Plant Syst. Evol.*, **214**, 131–145.

BARTISH, I. V., GARKAVA, L. P., RUMPUNEN, K. AND NYBOM, H. 2000. Phylogenetic relationships and differentiation among and within populations of *Chaenomeles* Lindl. (Rosaceae), estimated with RAPD and allozyme markers. *Theor. Appl. Genet*, **101**, 554–563.

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

CAMPBELL, C. S., GREENE, C. W. AND DICKINSON, T. A. 1991. Reproductive biology in subfam. Maloideae (Rosaceae). *Syst. Bot.*, **16**, 333–349.

CAMPBELL, C. S., WOJCIECHOWSKI, M. F., BALDWIN, B. G., ALICE, L. A. AND DONOGHUE, M. J. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol. Biol. Evol.*, **14**, 81–90.

DEMESURE, B., SODZI, N. AND PETIT, R. J. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.*, **4**, 129–131.

DICKINSON, T. A. AND CAMPBELL, C. S. 1991. Population structure and reproductive ecology in the Maloideae (Rosaceae). *Syst. Bot.*, **16**, 350–362.

DICKSON, E. E., KRESOVICH, S. AND WEEDEN, N. F. 1991. Isozymes in North American *Malus* (Rosaceae): hybridisation and species differentiation. *Syst. Bot.*, **16**, 363–375.

HUFF, D. R., PEAKALL, R. AND SMOUSE, P. E. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.]. *Theor. Appl. Genet.*, **86**, 927–934.

LYNCH, M. AND MILLIGAN, B. G. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, **3**, 91–99.

NEI, M. 1972. Genetic distance between populations. *Am. Nat.*, **106**, 283–292.

PHIPPS, J. B., ROBERTSON, K. R., ROHRER, J. R. AND SMITH, P. G. 1991. Origins and evolution of subfam. Maloideae (Rosaceae). *Syst. Bot.*, **16**, 303–332.

PHIPPS, J. B., ROBERTSON, K. R., SMITH, P. G. AND ROHRER, J. R. 1990. A checklist of the subfamily Maloideae (Rosaceae). *Can. J. Bot.*, **68**, 2209–2269.

RIESEBERG, L. H. AND ELLSTRAND, N. C. 1993. What can molecular and morphological markers tell us about plant hybridisation? *Crit. Rev. Plant. Sci.*, **12**, 213–241.

RIESEBERG, L. H., VAN FOSSEN, C. AND DESROCHERS, A. M. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, **375**, 313–316.

RIESEBERG, L. H., SINERVO, B., LINDER, C. R., UNGERER, M. C. AND ARIAS, D. M. 1996a. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, **272**, 741–745.

RIESEBERG, L. H., WHITTON, J. AND LINDER, C. R. 1996b. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Bot. Neerl.*, **45**, 243–262.

RIESEBERG, L. H. AND CARNEY, S. E. 1998. Tansley review no. 102. Plant hybridisation. *New Phytol.*, **140**, 599–624.

ROBERTSON, K. R., PHIPPS, J. B., ROHRER, J. R. AND SMITH, P. G. 1991. A synopsis of genera in Maloideae (Rosaceae). *Syst. Bot.*, **16**, 376–394.

ROHLF, F. J. 1997. NTSYS/pc. *Numerical taxonomy and multivariate analysis system*, Version 1.80. Exeter Software, Setauket, NY.

TABERLET, P., GIELLY, L., PAUTOU, G. AND BOUVET, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.*, **17**, 1105–1109.

WEBER, C. 1963. Cultivars in the genus *Chaenomeles*. *Arnoldia*, **23**, 17–75.

WEBER, C. 1964. The genus *Chaenomeles* Rosaceae. *J. Arnold. Arbor.*, **45**, 161–205, 302–345.

WEST, J. G. AND NOBLE, I. R. 1984. Analyses of digitised leaf images of the *Dodonaea viscosa* complex in Australia. *Taxon*, **33**, 595–613.