Molecular Phylogenetics and Evolution 67 (2013) 547–559

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

AFLP-based population structure analysis as a means to validate the complex taxonomy of dogroses (*Rosa* section *Caninae*)

Jan De Riek [a,*], Katrien De Cock [a], Marinus J.M. Smulders [b], Hilde Nybom [c]

[a] Institute for Agricultural and Fisheries Research, Plant Sciences Unit – Applied Genetics and Breeding, Caritasstraat 21, 9090 Melle, Belgium
[b] Wageningen UR Plant Breeding, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands
[c] Department of Plant Breeding and Biotechnology–Balsgård, Swedish University of Agricultural Sciences, Fjälkestadsvägen 459, 29194 Kristianstad, Sweden

ARTICLE INFO

Article history:
Received 11 June 2012
Revised 7 February 2013
Accepted 25 February 2013
Available online 7 March 2013

Keywords:
Dogroses
Taxonomy
DNA markers
Population genetics
Speciation
Hybridisation

ABSTRACT

Within the genus *Rosa* numerous species have been described. Circumscription of the dogrose section *Caninae* is straightforward, but the delineation of species and subsections within this section is less clear, partly due to hybridisation between species. We have investigated the extent to which DNA marker-based information of wild populations corroborates present-day dogrose taxonomy and hypotheses about the origination of taxa. Sampling was conducted in a transect across Europe, collecting over 900 specimens of all encountered dogrose taxa. For comparison, we also included more than 200 samples of species belonging to other sections. Two lines of statistical analyses were used to investigate the genetic structure based on AFLP data: (1) an unstructured model with principal coordinate analysis and hierarchical clustering, and (2) a model with a superimposed taxonomic structure based on analysis of genetic diversity using a novel approach combining assignment tests with canonical discriminant analysis. Support was found for five of the seven subsections, whereas *R. balsamica* apparently belongs to subsection *Caninae* thus omitting the need for recognising subsection *Tomentellae*. For *R. stylosa*, a hybridogenic origin with a non-dogrose section member has been suggested, and it can be treated either as a separate subsection or within subsection *Caninae*. Within the subsection *Rubigineae*, a species cluster with low support for the taxa *R. micrantha*, *R. rubiginosa* and the putatively hybridogenous *R. gremlii* was identified. Similarly, several species in the subsection *Caninae* overlapped considerably, and are best regarded as one common species complex. This population genetic approach provides a general method to validate the taxonomic system in complex and polyploid taxa.

1. Introduction

The genus *Rosa* has attracted considerable attention from taxonomists and numerous species have been described. Presently, about 200 species are recognised (Wissemann, 2003). Many of these are thought to have arisen by species hybridization, often accompanied by polyploidization. Although criticised in numerous recent DNA-based studies, the classification system of Rehder (1940) or variations thereof (e.g., Henker, 2000; Wissemann, 2003) still constitute the standard taxonomic treatment. This system comprises four subgenera with 10 sections in the largest, i.e., subgenus *Rosa*. Based on DNA marker analysis, two major clades have been identified within this subgenus, with sections *Carolinae*, *Rosa* (formerly *Cinnamomeae*) and parts of *Pimpinellifoliae* in one clade and most of the other sections in the other clade (Jan et al., 1999; Bruneau et al., 2005). This second clade comprises most of the progenitors of our cultivated ornamental roses, i.e., sections *Synstylae*, *Gallicanae* and *Indicae*. Its largest member is, however, the section *Caninae*, also known as dogroses and only utilised to a minor degree as rootstocks, for planting in public areas and for commercial rosehip production.

Early in the last century, several hundreds of dogrose taxa were described. More critical evaluations, both in the field and in herbaria, have later prompted a reduction to approx. 50 dogrose species (Wissemann, 2003). Most DNA analyses suggest that this section constitutes a well-circumscribed monophyletic group (Matsumoto et al., 2000; Scariot et al., 2006; Koopman et al., 2008). Although sharing Internal Transcribed Spacer (ITS) sequence types with species in other sections, thereby confirming their hybridogenous origin, the *Caninae* species also have one unique ITS sequence type which is further evidence of their monophyly (Ritz et al., 2005; Kovařík et al., 2008).

All dogrose species are characterised by the peculiar *canina* meiosis (Lim et al., 2005). Regardless of ploidy level (usually 5×, but some 4× and 6× taxa also occur; Täckholm, 1922; Wissemann, 2003), only seven bivalents are formed in the first meiotic division. The remaining chromosomes form univalents and are not included

* Corresponding author.
E-mail address: jan.deriek@ilvo.vlaanderen.be (J. De Riek).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.02.024

in viable pollen grains. SSR-based analyses of different species and offspring from controlled crosses suggest that bivalent formation involves one biparentally inherited, highly homozygous diploid genome, whereas the remaining 2, 3 or 4 haploid and often highly differentiated genomes are transmitted only by the seed parent (Nybom et al., 2004, 2006). Interfertility is very high among dog-rose taxa, and they can also hybridise with species of other ploidy levels, behaving as a polyploid when a seed parent, and as a diploid when a pollen parent.

While cytological, morphological and DNA-based circumscription of section *Caninae* generally is straightforward, the delineation of species and subsections within this section is less clear. Three larger subsections, with several species each, are usually acknowledged on the basis of leaf and fruit morphology: subsections *Caninae*, *Rubigineae* and *Vestitae*. Of these, only subsection *Rubigineae* is well-defined according to AFLP data (Koopman et al., 2008; De Cock et al., 2008). In addition three or sometimes four subsections, with one or two species each, are sometimes recognised and treated within section *Caninae*: subsections *Tomentellae*, *Rubrifoliae*, *Stylosae* and *Trachyphyllae*. On the basis of many sequence-based phylogenetic studies in the genus *Rosa*, Wissemann and Ritz (2005) pointed out two major problems; (1) European rhodologists have 'over-systematised' especially the section *Caninae* since Linnaean times, and (2) nr-ITS and cp-DNA data are not sufficient for reconstruction of phylogenetic relationships at species level. Although having provided insight into the delimitation of sections, nr-ITS and cp-DNA data produce gene trees, not species trees, and therefore cannot resolve deep species relationships (Wissemann and Ritz, 2005) in allopolyploids that contain different combinations of the genomes of the ancestral diploid species (Zhang et al., 2013).

Within the GeneRose project (Van Huylenbroeck et al., 2005), estimates of genetic diversity among and within natural populations of European dogrose taxa were used to validate taxonomic subgroups and current hypotheses about the (hybridogenic) origination of some of the taxa (Henker, 2000; Wissemann, 2003). Certainly, population genetics, not taxonomy, was the primary aim of the GeneRose study. However, the taxonomic uncertainty of our collected section *Canina* accessions, created both by vague species delineations and by inevitable misclassifications in the field, would, if not tackled, constitute a major hurdle for straightforward population genetic analysis. In a first attempt, in part reported in Koopman et al. (2008), we evaluated many species but only a limited number of accessions per species (on average 2 botanically "typical" specimens) with an extensive number of polymorphic markers (520 AFLP bands). Still, except for *Rubigineae*, no support was found for *Canina* subsections since the high within-species diversity produced a polyphyletic clade.

In the present study we took a different approach, using a much larger set of samples. We did not assume the existence of any taxon up front, but performed a statistical analysis to determine the level of evidence for putative taxa as present in the data. We followed two assumptions of Jacobs et al. (2011): (1) accessions that may exchange genetic material can be analysed as if they are part of one gene pool; and (2) genetic differentiation among species is expected to be higher than within species. Especially in the genetically complex dogroses, such a population genetic approach to a taxonomic problem is a methodological improvement since it starts from genetic similarities and differentiation among plants in the field and results in the construction of sets that correspond to existing natural populations, (sub)species, hybrids, etc. Nevertheless, and again due to the complex nature of dogroses, conventional population genetic measures like F_{st} values and popular tools as integrated in software suites like Structure (Pritchard et al., 2000) or Arlequin (Excoffier et al., 2005), are not applicable since underlying premises would be violated. Therefore, we combined (1) assignment tests (De Riek et al., 2001, 2007) to solve problems with "correctness of taxon names", with (2) canonical discriminant analysis to obtain unbiased structured ordinations with as few *a priori* assumptions as possible. In order to analyse the delimitation of section *Caninae* and determine the true position of some putatively hybridogenous taxa, we also included samples of species belonging to other sections; the statistical approach taken also allowed comparing these non-dogrose taxa with dogroses.

2. Materials and methods

2.1. Plant material

Material of wild rose shrubs were sampled across six North-Western European countries (Table 1; details in "Appendix") within the EU-funded GeneRose project (van Huylenbroeck et al., 2005). If available, inventories on the occurrence and distribution of indigenous rose species were used to select the sampled populations ("Appendix"). Generally, up to five plants per species were sampled in each population. In total, the data set contains 913 dogrose individuals, representing 22 species and two unnamed putative hybrids, and 255 populations. For comparison, a total of 226 samples were collected of 7 species representing 55 populations in sections *Pimpinellifoliae*, *Gallicanae*, *Rosa* (formerly *Cinnamomeae*) and *Synstylae*.

Plants were identified in the field and classified according to Henker (2000) (Table 1). Of each population, a representative herbarium specimen was collected for later verification of the initial *in situ* species determination, and deposited at the Wageningen branch (WAG) of the National Herbarium of The Netherlands. In cases of doubt, herbarium material was determined by three taxonomists independently. Young and fresh leaflets, collected on silica gel or liquid nitrogen, were lyophilised and stored at −18 °C, or frozen and kept at −80 °C until DNA extraction.

2.2. AFLP analysis

Approximately 300 ng DNA was obtained from 25 mg lyophilised leaf material (for freshly frozen material, the amount was adjusted to 100 mg) by the Qiagen DNeasy Plant Mini Kit (Westburg, The Netherlands). AFLP reactions were performed according to Vos et al. (1995); the restriction-ligation reaction was performed in one step, selective amplification included a pre-amplification with *Eco*RI-A/*Mse*I-C and final IR-labelled amplifications with three primer combinations *Eco*RI-AAG/*Mse*I-CAT, *Eco*RI-AAG/*Mse*I-CAG, and *Eco*RI-ATC/*Mse*I-CTA. Amplified fragments were separated on a Global Edition IR2 system of Li-COR (Li-COR) and visualised in automatically generated TIFF-files. Fragments were automatically scored using SAGA-MX version 3.0 (Li-COR), additional manual controls and corrections were performed. For the primer combination *Eco*RI-ATC/*Mse*I-CCG the *Eco*RI primer was ^{33}P-labelled prior to amplification, and amplified fragments were separated on a 6% polyacrylamide gel (Sequagel-6, Biozym) in 1× TBE (Tris–borate EDTA) electrophoresis buffer using a SequiGen 38 × 50 cm gel apparatus (BioRad, Hercules, California, USA) (as in Koopman et al., 2008). Gels were dried on Whatman 3MM paper, and X-ray films (Kodak X-OMAT, Rochester, New York, USA) were exposed for 1–3 wk at room temperature. The films were scanned and scored using QuantarPro (Keygene, Wageningen, The Netherlands).

In total, 137 polymorphic fragments were evaluated. The resulting data set was transformed into a binary matrix (presence 1, absence 0) as input for further statistical analyses.

2.3. Statistical analyses

In a first exploratory analysis, all samples were included, using the name given to that particular population by taxonomic experts.

Table 1
Samples of *Rosa* (taxonomic designation according to Henker (2000), collected in Belgium (B), Denmark (D), France (F), Germany (G), The Netherlands (N) and Sweden–Denmark (Sc).

Taxonomic designation	(Sub)section no.	Species no.	B	G	F	N	Sc	No. plants	No. pop.
Section *Pimpinellifoliae*	1								
R. spinosissima L.		1	8	19	11	23	2	**63**	18
Section *Gallicanae*	2								
R. gallica L.		2		10	39			**49**	10
Section *Rosa (form. Cinnamomeae)*	3								
R. majalis Herrm.		3		21			8	**29**	7
R. pendulina L.		4		10	2			**12**	4
Section *Synstylae*	4								
R. arvensis Huds.		5	15	29	6	12		**62**	12
R. multiflora Thunb.		6		3				**3**	1
R. sempervirens L.		7			8			**8**	4
Section *Caninae*									
Subsection *Trachyphyllae*	5								
R. marginata Wallr.		8		10				**10**	2
Subsection *Rubrifoliae*	6								
R. glauca Pourr.		9	1	8	3			**12**	6
Subsection *Rubigineae*	7								
R. agrestis Savi		11	10		11	10		**31**	10
R. gremlii Christ		12				25	5	**30**	8
R. elliptica Tausch		13			4			**4**	2
R. inodora Fr.		14					8	**8**	2
R. micrantha Sm.		15	6	5		14		**25**	8
R. rubiginosa L.		16	25	23	5	34	43	**130**	37
R. rubiginosa × gremlii		17				9		**9**	2
Subsection *Vestitae*	8								
R. mollis Sm.		18		15			15	**30**	14
R. pseudoscabriuscula (R. Keller) Henker & G. Schulze		19	23	5		1		**29**	7
R. sherardii Davies		20		11	1	12	6	**30**	9
R. tomentosa Sm.		21	3		4	60	3	**70**	17
R. villosa L.		22	2		2			**4**	4
Subsection *Tomentellae*	9								
R. balsamica Besser		32	12			43		**55**	14
Subsection *Caninae*	10								
R. caesia Sm.		23		1	2	3	4	**10**	5
R. canina L.		24	34	56	9	69	37	**205**	55
R. canina × stylosa		25	1					**1**	1
R. corymbifera Borkh.		26	10	38	7	61		**116**	35
R. dumalis Bechst.		27		5	1	4	52	**62**	20
R. montana Chaix		28			15			**15**	5
R. stylosa Desv.		29	4					**4**	3
R. subcanina (Christ) Vuk.		30	2	1		6		**9**	8
R. subcollina (Christ) Vuk.		31				11		**11**	5
R. × irregularis Déségl. & Guillon		33	2			1		**3**	3
TOTALS			158	270	130	398	183	**1139**	338

Principal Coordinate Analyses (PCO) were conducted in SPSS Statistics 20 (IBM Corporation) taking a Jaccard similarity matrix between all individuals as input. Partitioning of genetic variability was investigated with analyses of molecular variance (AMOVA), using the Arlequin software (Excoffier et al., 2005). F_{st} matrices generated by AMOVA, were by UPGMA clustering turned into dendrograms using NEIGHBOR and CONSENSE (Phylip; Felsenstein, 2005), albeit without bootstrap support since this is not obtained within the AMOVA procedure.

In a second approach, we combined assignment tests (De Riek et al., 2001, 2007) with canonical discriminant analysis using SPSS to obtain structured ordinations. First, an assignment table was produced, which showed for each specimen under evaluation the most related set of species. The assignment values were taken as input to a canonical discriminant analysis, targeting the classification towards membership of taxonomical sections, subsections (dogroses) or species. The independent variables were entered simultaneously. The covariance matrix within groups was used for the ordination; prior probabilities for classification were computed from the group sizes. Classifications were based both on case-wise results, and on the leave-one-out method.

3. Results

3.1. General genetic structure and attribution of genetic variation

The sampling strategy of the GeneRose survey focussed on the genetic diversity within and between wild rose populations present in Europe. Species delineation, especially within the dogroses, appeared to be vague. Possible explanations are over-classification of the section *Caninae* creating intrinsic taxonomic noise, as well as ordinary misnaming or mislabelling during collection of samples. Especially the hunting for rare species may have yielded incorrectly determined samples. An initial analysis of the AFLP data during the GeneRose project detected obvious mistakes in

determination; erroneous samples were rechecked by three taxonomists (coded as "Field determination" versus "Final determination" in "Appendix").

PCO analysis on all 1130 rose samples explained in total 81% of the present variation; components 1 (67.1%) and 2 (7.5%) produced three major clusters (Fig. 1). Samples classified to the sections *Pimpinellifoliae* and *Rosa* (formerly *Cinnamomeae*) were mingled in one cluster, while *Synstylae* and *Gallicanae* formed a second cluster, in which each section could be regarded as a subcluster. Finally, the largest and most dense cluster consisted of all *Caninae* samples. The combination of pentaploidy (in most taxa) and dominantly inherited AFLP markers rendered dogroses very similar in these analyses, although they strongly outnumbered the other sections in plants analysed. All *Caninae* samples were intermingled along the first two components but subsection *Rubigineae* was clearly different from the other *Caninae* taxa on the third component (6.1% of the variation). A second PCO on only the dogrose samples explained 52.1%, 14.4% and 6.8% of the variation on the first three components (Fig. 2) and revealed two loosely defined clusters, with subsection *Rubigineae* samples forming one cluster and the remaining samples in the other, or unclustered.

Using AMOVA, the relevance of different taxonomic substructures for explaining the genetic variation among samples could be compared. When all the material was entered into the same analysis 35–40% of the variation was encountered at the uppermost level, i.e. among the total of five sections, or among the total of 32 species and hybrids (Table 2). In the classification based on sections, populations within sections explained 35% of the variation, while 27% occurred within populations. A slightly lower differentiation, 30%, was found among populations within species while 35% occurred within populations.

Analysing only the dogrose taxa resulted in a very different partitioning of the genetic variation. Differentiation among subsections explained only 20% of the variation while populations

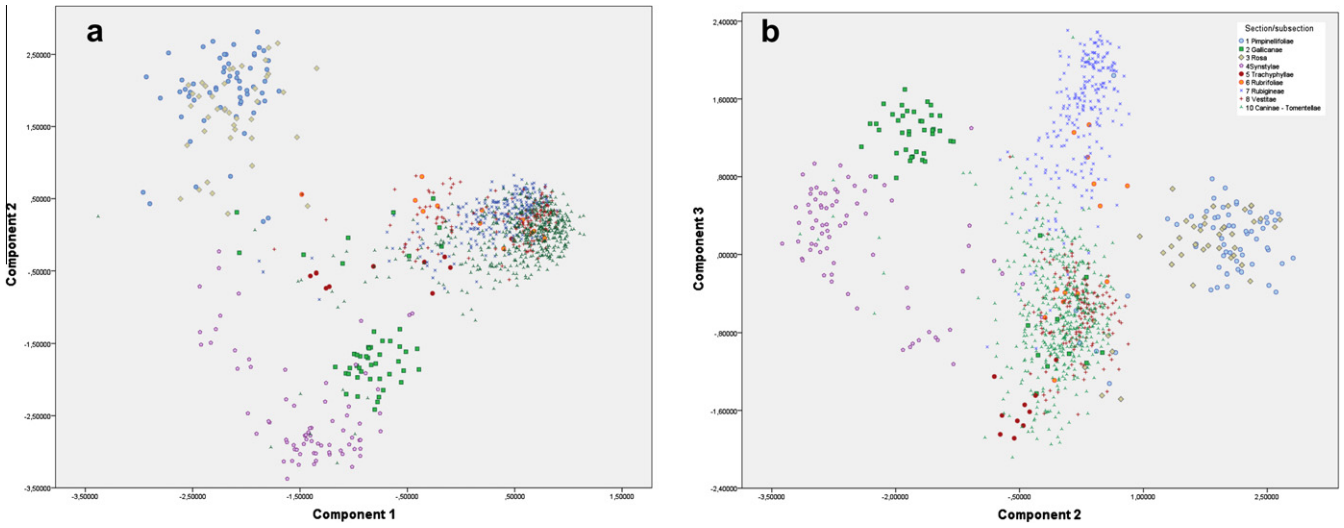

Fig. 1. PCO plots of the European subgenus *Rosa* based on 137 polymorphic AFLP markers (a) the first two components; (b) the second and third component. The first three components explained 67.1%, 7.5%, and 6.1%, respectively, of the variation. Sections *Pimpinellifoliae* (1), *Gallicanae* (2), *Rosa* (formerly *Cinnamomeae*) (3), *Synstylae* (4), and section *Caninae* with subsections *Trachyphyllae* (5), *Rubrifoliae* (6), *Rubigineae* (7), *Vestitae* (8), *Tomentellae* (included in 10) and *Caninae* (10) are given.

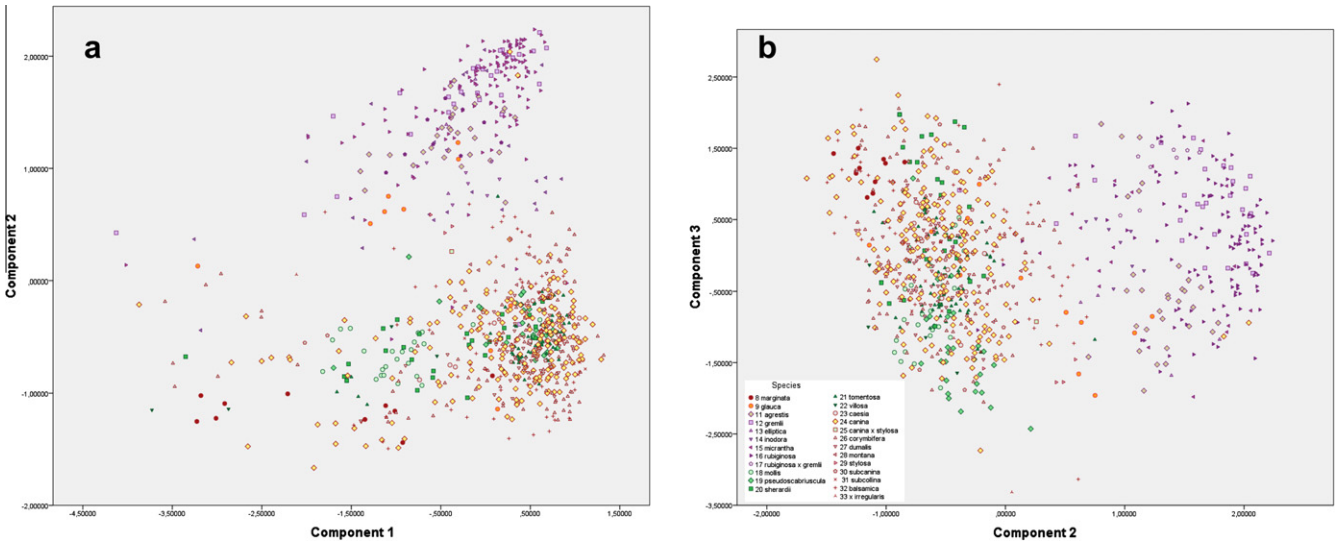

Fig. 2. PCO plots of the section *Caninae* species based on 137 polymorphic AFLP markers (a) the first two components; (b) the second and third component. The first three components explained 52.1%, 14.4%, and 6.8%, respectively, of the variation. Species indication codes according to Table 1.

Table 2
Distribution of molecular variance over different taxonomic substructures (SS sum of squares; d.f. degrees of freedom).

AMOVA analysis on samples from five sections within the genus Rosa				
Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among sections	4	4230.872	10.40891	38.32
Among pop. w. section	321	12756.183	9.44473	34.77
Within populations	801	5855.307	7.31000	26.91
Among species	31	8461.872	7.23568	34.63
Among pop. w. sp.	295	8543.588	6.34866	30.38
Within populations	801	5855.307	7.31000	34.99
AMOVA analysis on samples from subsection Caninae				
Among subsections	5	2395.759	3.70067	20.41
Among pop. w. subsect.	265	8423.304	7.48603	41.29
Within populations	635	4409.200	6.94362	38.30
Among species	24	3932.495	3.98657	23.02
Among pop. w. species	249	6951.837	6.38663	36.88
Within populations	635	4409.200	6.94362	40.10

among subsections explained 41% with the remaining 38% occurring within populations. Structuring the data according to species produced only a marginally higher differentiation, 23%, at the uppermost level followed by 37% among populations within species and 40% within populations.

F_{st} matrices derived from the AMOVA analyses were used for calculating a set of dendrograms. When the five sections were used as OTUs (operational taxonomic units), sections *Pimpinellifoliae* and *Rosa* (formerly *Cinnamomeae*) were the closest. Sections *Gallicanae* and *Synstylae* also formed a cluster, with *Caninae* more loosely attached (Fig. 3a). Restricting the data to only section *Caninae* and using six subsections as OTUs produced a dendrogram in which *Tomentellae* and *Caninae* formed a tight cluster, with the other four subsections added to it in the order: *Vestitae*, *Rubigineae*, *Rubrifoliae* and *Trachyphyllae* (Fig. 3b). Finally, one dendrogram was produced with the 32 species and putative hybrids as OTUs (Fig. 3c). All dogrose species clustered together, with two notable exceptions; *R. glauca* in subsection *Rubrifoliae* clustered with the three species in sections *Rosa* (formerly *Cinnamomeae*) and *Pimpinellifoliae*, while *R. marginata* Wallr. (synonym *R. jundzillii* Besser) in subsection *Trachyphyllae* split off from the remainder just after the first split-off, i.e. *R. multiflora* Thunb. ex Murr. (section *Synstylae*). Among the dogroses, all taxa in subsection *Caninae* except *R. stylosa* Desvaux clustered together, as did all taxa in subsection *Rubigineae* except *R. inodora* Fr.

3.2. Data management and fine-check of the classification

In contrast to PCO or hierarchical clustering, a structured analysis imposes *a priori* a certain (taxonomic) grouping and subsequently evaluates the validity of the imposed framework by applying statistics. We have combined assignment tests and canonical discriminant analysis in a novel approach. For the assignment test, each specimen was targeted individually: the initially given species name was recorded for each of the individual genotypes that showed a Jaccard similarity $\geqslant 0.60$ to the specimen under evaluation. The distribution (percentage different species names in the group of similar plants) was used to assign each plant to the most likely species. This assignment was carried out with a canonical discriminant analysis using the columns of the assignment test (the "Species scores" in "Appendix") as independent variables. Each specimen was thus classified into species and section, or subsection (for dogrose species) ("Appendix"). Several specimens had a closer affiliation with another species than with the species to which it belonged according to taxonomic expertise. A total of 102 samples were identified as possibly misclassified ("Appendix"). Sixty-one of these samples had originally been determined as belonging to a taxon within section *Caninae* (i.e. species 8–33 in Table 1) but nevertheless possessed AFLP profiles that complied better with either a dogrose taxon from another subsection (in this context, the very similar subsections *Tomentellae* and *Caninae* were regarded as one subsection only), or with a taxon from one of the other four *Rosa* sections. These possibly misclassified specimens were treated as 'unclassified' in the subsequent discriminant analyses described below for analyses of subsets of the material. These specimens were entered as unmarked symbols in the biplots, so as not to detract from the attempts to delineate the analysed taxa in more detail.

3.3. Phenetic relationships in subsets of the material

Twelve taxonomic issues regarding placement in (sub)sections, distinction between species, or putative hybrid origin of taxa were studied by analysing the genetic differentiation in various combinations of taxonomic units with biplots showing the first two axes of the canonical discriminant ordination; ordinations were set, – i.e. structured –, according to the taxonomic level of (sub)sections or species (Table 3). See the legends of Fig. 4a–l for the issues tackled. Eigenvalues for the discriminant axes together with an estimate of the variance explained by the model and related statistics were used to judge the relevance of the observed clustering in the biplots obtained (in "Appendix").

Within the six dogrose subsections, three major groups were found (Fig. 4a), with subsection *Rubigineae* being the most well-defined. The other major groups comprised all species in subsection *Vestitae* in one group, and all species in subsections *Caninae*, *Tomentellae* and *Trachyphyllae* in the other. A fourth, smaller group containing samples of subsection *Rubrifoliae* was placed in between of the *Caninae*/*Tomentellae*/*Trachyphyllae* group on the one hand, and the *Rubigineae* group on the other hand. Between these two major groups, another small cluster was found with *R. stylosa* and samples of a putative hybrid between *R. canina* L. and *R. stylosa*. As for the variation among subsections, only three of these (*Rubigineae*, *Vestitae* and *Rubrifoliae*) appear to be well differentiated while the remaining three subsections formed one composite cluster.

In more recent classifications, subsection *Rubrifoliae*, here only represented by *R. glauca* Pourr., is placed within section *Caninae* since its members have the *canina* meiosis although morphologically *R. glauca* belongs in section *Rosa* (formerly *Cinnamomeae*; Wissemann, 2003). The *R. glauca* samples formed a small subgroup close to the major groups of dogroses (Fig. 4b). However, the *R. glauca* subgroup was closer to *R. majalis* Herrm. and *R. pendulina* L. than any of the other dogrose samples was, suggesting a true affiliation and possibly a hybridogenous origin involving both dogroses and a species of section *Rosa* (formerly *Cinnamomeae*).

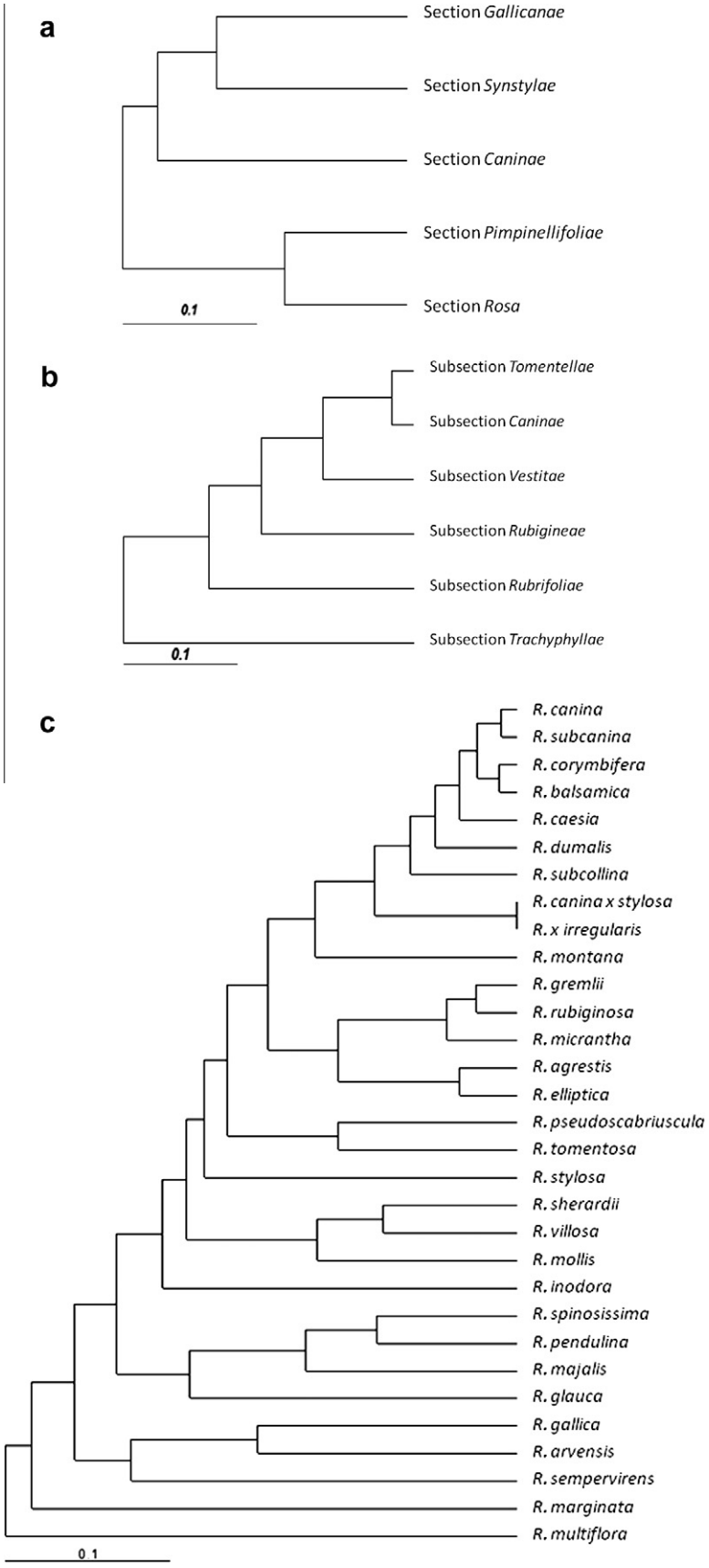

Fig. 3. Phenograms based on the pairwise Fst matrix between *Rosa* sections (a), Section *Caninae* subsections (b) or *Rosa* species and interspecific hybrids (c). The Fst-matrices were derived from the AMOVA analysis (UPGMA clustering).

Table 3
Species with a supposed hybridogenic origin.

Taxon	Proposed position	References	Compared to	See
R. glauca	*Sections Rosa* (morphology) and *Caninae* (meiosis)	Wissemann (2003)	All dogroses, *R. majalis* and *R. pendulina*	Fig. 4b
R. stylosa	Hybrid between *Synstylae* and *Caninae*	De Cock et al. (2007) and vander Mijnsbrugge et al. (2010)	*Caninae* and *Synstylae*	Fig. 4l
	Hybrid between *Rubigineae* and *Caninae*	Wissemann (2000)	*Caninae* and *Rubigineae*	Fig. 4d
R. balsamica	Hybrid between *Rubigineae* and *Caninae*	Wissemann (2000)	*Caninae* and *Rubigineae*	Fig. 4a and d
R. marginata	Hybrid between *Caninae* and *R. gallica*	Wissemann (1999, 2003)	Major *Caninae* subsections and *R. gallica*	Fig. 4e
R. tomentosa	Intermediate between *R. sherardii* and *R. canina*	Nilsson (1967)	*R. pseudoscabriuscula*, *R. sherardii*, *R. canina*	Fig. 4h
R. pseudoscabriuscula	Hybrid between *R. canina* and *R. tomentosa*	Kurtto et al. (2004)	*R. tomentosa*, *R. sherardii*, *R. canina*	Fig. 4h
R. gremlii	Hybrid between *R. rubiginosa* and *R. micrantha*		*R. rubiginosa*, *R. micrantha*, *R. rubiginosa* × *R. gremlii*	Fig. 4i and j
R. × *irregularis*	Hybrid between *Synstylae* and *Caninae*	De Cock et al. (2007) and vander Mijnsbrugge et al. (2010)	*R. arvensis*, *R. canina*, *R. tomentosa*	Fig. 4l
R. dumalis	(*R. canina* or *R. corymbifera*) × *R. rubiginosa*	Ritz and Wissemann (2003) and Wissemann and Ritz (2007)	*R. rubiginosa*, *R. micrantha*, *R. corymbifera* and *R. canina*	Fig. 4k
R. micrantha	*R. rubiginosa* × (*R. canina* or *R. corymbifera*)	Ritz and Wissemann (2003) and Wissemann and Ritz (2007)	*R. rubiginosa*, *R. corymbifera*, *R. canina* and *R. dumalis*	Fig. 4k

In Fig. 4c, again, *R. stylosa* and its putative hybrid with *R. canina* formed a group of their own; five of the 7 samples in subsection *Trachyphyllae* also grouped together, separate from the large subsection *Caninae* cluster. Subsection *Trachyphyllae* contained only one species, *R. marginata*. The third subsection, *Tomentellae*, was represented by only one species, *R. balsamica* Besser (synonym *R. tomentella* Léman ex Cassini) but samples had been collected from 14 populations in Belgium and Netherlands. These samples occurred in the outskirts of the large subsection *Caninae* cluster but overlapped considerably with the latter and thus showed poor differentiation. One of the subsection *Caninae* species, *R. montana* Chaix ex Vill., actually showed a stronger separation from the main subsection *Caninae* cluster. Although a hybridogenous origin of *R. stylosa* is supported by a somewhat intermediate position in the plot (Fig. 4d), the samples showed more affinity with subsection *Caninae* than with subsection *Rubigineae*.

The putatively hybridogenous origin of *R. marginata* (Wissemann, 1999, 2003) was investigated in an analysis together with samples of subsection *Caninae* and *R. gallica* (Fig. 4e). While most samples of *R. marginata* clearly differed from subsection *Caninae*, there was no evidence of an origin involving *R. gallica*.

The most close-knit taxa in subsection *Caninae* were analysed together, i.e. after exclusion of the somewhat deviant *R. stylosa* and *R. montana* (Fig. 4f). The three largest taxa in terms of number of samples, *R. dumalis*, *R. canina* and *R. corymbifera*, overlapped strongly with an intermediate (central) position for *R. canina*. Close to these three species, also *R. subcanina* (H. Christ.) R. Keller was found. Somewhat further away but still overlapping considerably was a subgroup comprising *R. caesia* Sm., *R. subcollina* (H. Christ.) R. Keller and *R.* × *irregularis* Déségl. & Guillon.

Similarly, all species in subsection *Vestitae* were analysed together (Fig. 4g). *Rosa villosa* L., represented here by only four samples, was the best-defined species, followed by the tetraploid *R. mollis* Sm. which, however, overlapped somewhat with *R. sherardii* Davies. Rather close to the latter, *R. tomentosa* Sm. and *R. pseudoscabriuscula* (R. Keller) Henker & G. Schulze occurred at a short distance from one another. The five investigated species showed some overlap, except for *R. villosa*, but a structure with five major groups is still evident. Moreover, each species overlapped with only one or two other species, not with three or more as the species in subsection *Caninae* did.

No clear-cut evidence of hybridization was found for *R. tomentosa* or *R. pseudoscabriuscula* (Fig. 4h); *R. tomentosa* occurred between *R. pseudoscabriuscula* and *R. sherardii*, and all three of these were well separated from *R. canina*. However, a number of ungrouped cases, originally determined mainly as *R. canina*, bridged the gap between *R. canina* and *R. tomentosa*, and may indicate a relationship. By contrast, the gap between *R. canina* and *R. pseudoscabriuscula* appeared to be larger.

Both *R. stylosa* and *R.* × *irregularis* have been described as hybrids between section *Synstylae*, especially *R. arvensis*, and subsection *Caninae*, especially *R. canina* or *R. corymbifera* (de Cock et al., 2007; vander Mijnsbrugge et al., 2010). Analysis of these five taxa and a putative *R. canina* × *R. stylosa* sample showed that *R. stylosa* and the putative hybrid sample take an intermediate position between the two *Caninae* taxa and the putative *R. arvensis* parent along the first function (Fig. 4l). The two correctly assigned *R.* × *irregularis* samples clustered with *R. canina* but a third sample, that had been deleted due to a higher AFLP similarity with a different taxon, actually clustered with *R. arvensis*.

Differentiation among species in subsection *Rubigineae* was also investigated (Fig. 4i). *Rosa inodora* was well separated from the remainder, while *R. agrestis* Savi and *R. elliptica* Tausch formed one close-knit group and *R. gremlii* (synonym *R. columnifera* (Schwertschlager) Henker & G. Schulze non-Fries), *R. micrantha* Borrer ex Sm., *R. rubiginosa* L. and some samples referred to the

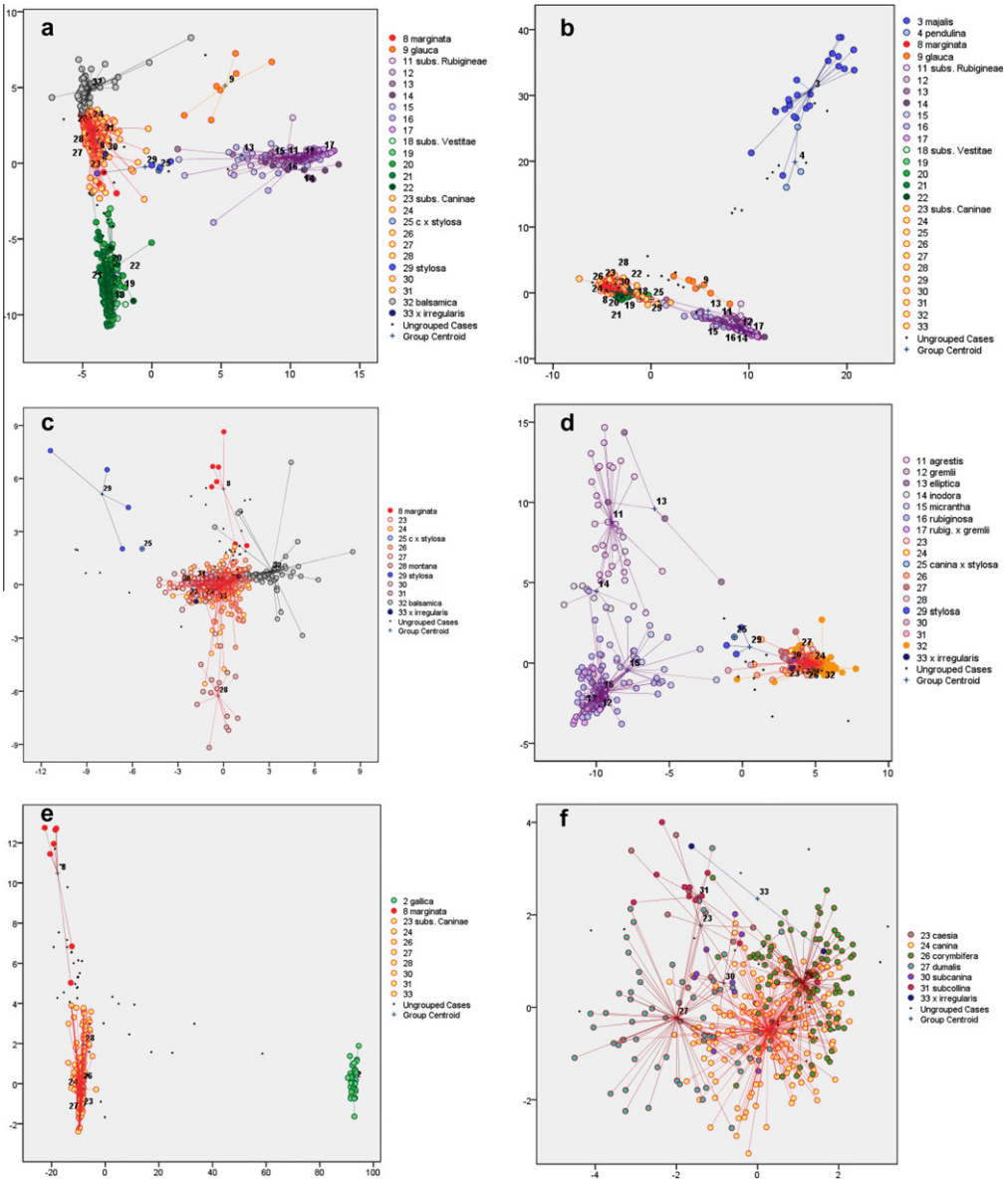

Fig. 4. Various discriminant analysis-derived plots (“ungrouped cases” are samples with non-matching (sub)sections). (a) Differentiation of all dogrose samples: positioning of subsections *Rubigineae*, *Rubrifoliae* (*R. glauca*), *Trachyphyllae* (*R. marginata*) and *Tomentellae* (*R. balsamica*) towards subsection *Caninae* (*Axis 1*: Eigenvalue = 42.738; % of variance = 50.3; *Axis 2*: Eigenvalue = 15.165; % of variance = 17.8). (b) Dogrose subsections compared to section *Rosa* (formerly *Cinnamomeae*; *R. majalis* and *R. pendulina*): positioning of subsection *Rubrifoliae* (*R. glauca*) and subsection *Rubigineae* (*Axis 1*: Eigenvalue = 41.165; % of variance = 36.8; *Axis 2*: Eigenvalue = 34.962; % of variance = 31.3). (c) Positioning of the subsections *Trachyphyllae*, *Tomentellae* (*R. balsamica*) and *R. stylosa* versus subsection *Caninae* (*Axis 1*: Eigenvalue = 2.351; % of variance = 24.0; *Axis 2*: Eigenvalue = 2.082; % of variance = 21.2). (d) Positioning of *R. stylosa* versus subsections *Caninae* and *Rubigineae* (*Axis 1*: Eigenvalue = 46.157; % of variance = 75.2; *Axis 2*: Eigenvalue = 5.173; % of variance = 8.4). (e) Positioning of *R. marginata* versus *R. gallica* and subsection *Caninae* (*Axis 1*: Eigenvalue = 871.252; % of variance = 99.3; *Axis 2*: Eigenvalue = 2.044; % of variance = 0.2). (f) Inner structure of subsection *Caninae*: *R. dumalis*, *R. subcollina*, *R. subcanina*, *R. canina*, *R. caesia*, *R. corymbifera*, *R.* × *irregularis* (*Axis 1*: Eigenvalue = 1.145; % of variance = 54.2; *Axis 2*: Eigenvalue = 0.486; % of variance = 23.0). (g) Species differentiation within subsection *Vestitae* (*Axis 1*: Eigenvalue = 6.100; % of variance = 58.4; *Axis 2*: Eigenvalue = 2.393; % of variance = 22.9). (h) Positioning of *R. sherardii*, *R. tomentosa* and *R. pseudoscabriuscula* versus *R. canina* (*Axis 1*: Eigenvalue = 19.288; % of variance = 85.4; *Axis 2*: Eigenvalue = 2.256; % of variance = 10.0). (i) Species differentiation within subsection *Rubigineae* (*Axis 1*: Eigenvalue = 7.495; % of variance = 54.2; *Axis 2*: Eigenvalue = 4.525; % of variance = 32.7). (j) Positioning of *R. rubiginosa*, *R. gremlii* and their putative hybrid versus *R. micrantha* (*Axis 1*: Eigenvalue = 1.184; % of variance = 57.0; *Axis 2*: Eigenvalue = 0.641; % of variance = 30.9). (k) Positioning of *R. micrantha*, *R. rubiginosa*, *R. canina*, *R. corymbifera* versus *R. dumalis* (*Axis 1*: Eigenvalue = 58.705; % of variance = 96.7; *Axis 2*: Eigenvalue = 1.063; % of variance = 1.8). (l) Positioning of *R. arvensis*, *R. tomentosa*, *R. canina*, *R. stylosa*, *R. canina* × *stylosa* and *R.* × *irregularis* (*Axis 1*: Eigenvalue = 123.202; % of variance = 85.5; *Axis 2*: Eigenvalue = 20.762; % of variance = 14.4).

putative hybrid *R. rubiginosa* × *R. gremlii* formed another close group.

Similarity among the four taxa in the above-mentioned close-knit group within subsection *Rubigineae* was also examined (Fig. 4j). While *R. gremlii* and *R. rubiginosa* remained very close together, some differentiation was shown by *R. micrantha* and the putative hybrid *R. rubiginosa* × *R. gremlii*, which thus appeared to have a different origin.

Finally, interspecific hybrids between *R. canina* or *R. corymbifera* as seed parent and *R. rubiginosa* as pollen parent have been claimed to be morphologically indistinguishable from *R. dumalis*, while the reciprocal hybrids *R. rubiginosa* × *R. canina* (or *R. corymbifera*) would be morphologically identical to *R. micrantha* (Ritz and Wissemann, 2003; Wissemann and Ritz, 2007). An analysis of our samples from these five taxa resulted in two well-defined clusters; one with *R. rubiginosa* and *R. micranta*, and another with *R. corymbifera*,

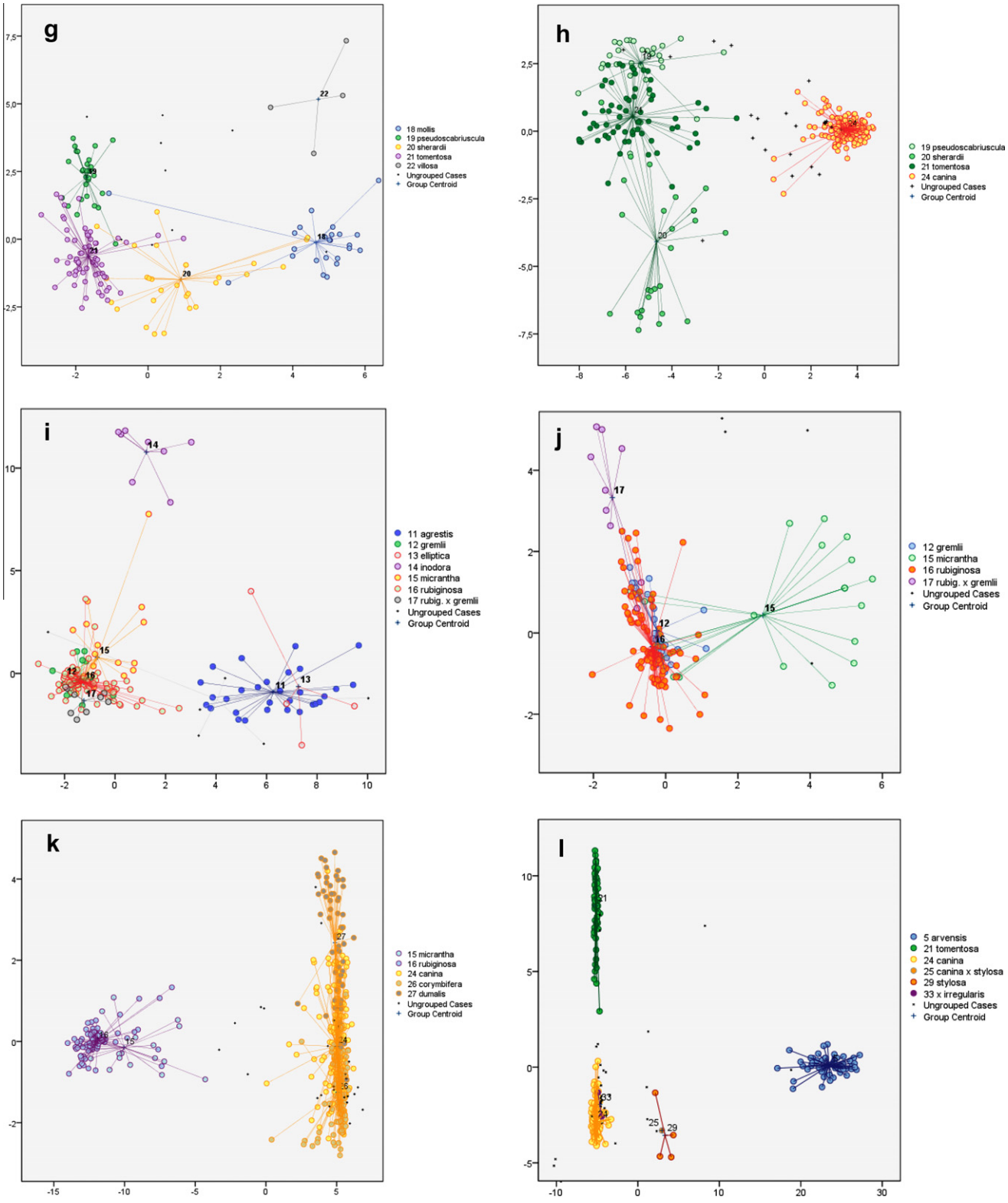

Fig. 4. (*continued*)

R. canina and *R. dumalis* (Fig. 4k). The few intermediate, ungrouped samples in between of *R. canina* and *R. micrantha* may indicate some gene flow. By contrast, *R. dumalis* did not seem to have any relationship with the *Rubigineae*.

4. Discussion

Since our aim was to elucidate genetic differentiation within a close-knit group of taxa, we analysed fewer AFLP bands compared

to a taxonomic study (e.g., Koopman et al., 2008), but on many plants per taxon, rather than on a single representative of each taxon. We consider this a first attempt, not only to validate the species delineation but also to determine the relevance of the different taxonomic structures within the genus *Rosa*, i.e., sections and subsections. We agree with Jacobs et al. (2011) in the concept that, as gene flow occurs mainly within a species, accessions of one species are expected to share more alleles with each other than with accessions from other species. As a result, genetic differentiation among species is expected to be higher than within species, and genetic variation will be optimally distributed with exactly the right number of species. Both overclassification and underclassification will decrease genetic differentiation among groups (commonly represented as F_{st}). Jacobs et al. (2011) operationalised this by maximising F_{st} while merging putatively overclassified taxa in *Solanum* section *Petota*.

Dogroses are polyploid, thus conventional genetic distances do not perform as expected. In addition, we used dominantly inherited AFLP markers. Two lines of analyses were conducted; unstructured data were analysed with PCO and hierarchical clustering (dendrograms), while data with a superimposed taxonomic structure were analysed by AMOVA, assignment tests and discriminant analysis. Unlike true assignment tests (Paetkau et al., 1997; De Riek et al., 2007), the commonly applied STRUCTURE (Pritchard et al., 2000) method only differs from (non-structured) clustering and PCO analysis in that *a priori* the number of groups K is set (which is again a variable over the different simulations); individuals are attributed to the different clusters according to minimised linkage disequilibrium (LD). LD between (dominant) markers does not have predictable properties in polyploids. We decided to apply assignment and discriminant analysis for the analysis of taxonomic structure rather than STRUCTURE for two additional reasons: (1) a similarity measure $Sa_{x,y}$, directly derived from assignment values, has been documented (De Riek et al., 2007) as a robust method for assessment of genetic conformity, and has been shown to be insensitive to the molecular marker technique applied (AFLP versus SSR or CAPS) or the similarity measure used (Jaccard versus Simple Matching) in a complex group of diploid and triploid sugar beet accessions; in this set it was superior to the well-documented assignment method of Paetkau et al. (1997); (2) in contrast to STRUCTURE, no prior assumptions are needed about the number of groups or whether the samples are from populations in Hardy–Weinberg equilibrium, since the assignment values or their derived $Sa_{x,y}$-measures are merely database tools revealing relationships between the analysed accessions from the existing marker data as such.

4.1. Taxonomic differentiation of European wild roses

Our study supports the overall genetic structure of subgenus *Rosa* with five sections in Europe (Table 2 and Fig. 1). The previously reported (Koopman et al., 2008) higher similarity between sections *Gallicanae* and *Synstylae* on the one hand and sections *Pimpinellifoliae* and *Rosa* (formerly *Cinnamomeae*) on the other hand was confirmed in both the PCO (Fig. 1) and the dendrogram (Fig. 3a). In accordance with Jan et al. (1999) and Bruneau et al. (2005), section *Caninae* showed somewhat higher affinity to sections *Gallicanae* and *Synstylae* than to *Pimpinellifoliae* and *Rosa* (formerly *Cinnamomeae*). Both Henker (2000) and Graham and Primavesi (1993) described the occurrence of interspecific hybridisation between section *Caninae* and diploid section *Synstylae* species, and sequencing of the nrDNA internal transcribed spacer (nrITS-1) region has also indicated that the section *Synstylae* forms a direct sister group to the section *Caninae* (Wissemann and Ritz, 2005).

Although the whole section *Caninae* may have been derived by interspecific hybridization, a hierarchical subdivision was observed in our data with AMOVA (Table 2). Excluding all species that do not belong to section *Caninae* decreased the differentiation among species (only 23% of the variation compared to 35% when the entire material was analysed) and increased the differentiation among populations (37% compared to 30%) and within (40% compared to 35%), which suggests that species delimitation is more difficult to achieve in section *Caninae*. By contrast, hierarchical canonical variates analysis of morphological data from a comparative garden trial placed most of the variation between the six studied dogrose taxa (80%) while the remaining variability occurred mainly between populations within taxa; 17% of the total diversity in manually measured reproductive characters and 20% of the total diversity in image analysis-derived leaflet shape (Olsson and Prentice, 2001). In general, morphological data appear to be more consistent with current taxonomy than are the AFLP data.

In the present study, subsection *Rubigineae* appeared to be the most distinct group, as suggested also by Koopman et al. (2008). Although the other subsections overlapped largely, and lacked clear and well-defined boundaries on the initial PCO plots (Fig. 1), a structured genetic analysis was able to confirm the status of 6 subsections (Fig. 4a and c). It should, however, be noted that 61 samples had been deleted from these analyses after being assigned to different subsections compared to the original determinations. This procedure is likely to have 'sharpened' the boundaries between the different subsections. In addition, for subsections *Stylosae*, *Trachyphyllae* and *Rubrifoliae* only a small number of individual plants were analysed. Presently, there is, however, no evidence against their present taxonomic status. By contrast, subsection *Tomentellae* was not supported in spite of being represented by material from 14 populations in two different countries.

Differentiation among the *Caninae* subsections might be explained by the very strict conditions under which hybrids are fertile. According to Nybom et al. (2006) interspecific hybrids are fertile and able to contribute their genetic material to the next generation only if the bivalent-forming chromosomes are sufficiently homologous to recombine and thus maintain the *canina* meiosis. Whereas dogrose species usually have a viable pollen stainability of 20–30%, hybridisation between distantly related dogrose taxa (i.e., taxa in different subsections) produces offspring with only approx. 5% pollen stainability although seed viability remains at the same level as in the parental species (Werlemark, 2000; Werlemark and Nybom, 2001).

Even when most of the *Caninae* subsections can be considered as genetically founded (at least based on the set of AFLP markers analysed here), the existence of multiple species within the various subsections is more cumbersome to analyse, since it concerns genetic differentiation between accessions that in principle could produce fully fertile offspring. The analysis is even more complicated by the possible occurrence of hybrids within and between *Caninae* subsections and other sections. If our sampling had been wider, including also Central and Eastern Europe and adjoining areas in Asia where subsection *Caninae* has its centre of diversity and is likely to have originated, species delimitation would probably have been even more difficult. Below we discuss the results subsection by subsection.

4.2. Support for taxa in subsections Caninae and Tomentellae

Samples of nine species and a possible hybrid in subsection *Caninae* were analysed, with *R. canina*, *R. corymbifera*, *and R. dumalis* having the largest number of samples. Of the *Tomentellae*, only *R. balsamica* was included. Although both Henker (2000) and Wissemann (2003) made a distinction between the subsections *Tomentellae* and *Caninae*, the systematic position of *R. balsamica* and the

second species in this subsection, *R. abietina* Gren ex. Christ, is known to be uncertain (Wissemann, 2000). The nomenclature of *R. balsamica*, *R. obtusifolia*, or *R. tomentella* has also been a subject of discussion. As proposed by Kurtto et al. (2004), *R. balsamica* is the correct name and *R. tomentella* is suggested to be a synonym; *R. obtusifolia*, which both Graham and Primavesi (1993) and Nilsson (1967) used as a synonym of *R. tomentella*, is mentioned as a synonym of *R. corymbifera*. Other evidence also indicates a high similarity between subsections *Caninae* and *Tomentellae*. Firstly, the epicuticular wax type found in the majority of the section *Caninae* taxa is observed also in *R. balsamica* and *R. abietina* (Wissemann, 2000). Secondly, cpDNA sequence analysis shows that *R. abietina* (*R. balsamica* not included in this study) clusters within the subsection *Caninae* clade (Wissemann and Ritz, 2005).

Based on our genetic structure analysis, only two taxa in subsections *Tomentellae* and *Caninae*, namely *R. montana* and *R. stylosa*, were discerned as separate entities (Fig. 4c). The other taxa grouped closely together (Fig. 4c and f). These also display few clear morphological subsection- and species-related characters (e.g., presence and frequency of pubescence and glands on leaflets, hips, and pedicels). An AFLP-based study on diversity within and among taxa and localities showed that, e.g., samples of *R. canina* and *R. corymbifera* collected at the same locality were more similar to one another than were con-specific samples collected at different localities (De Cock, 2008).

In conclusion, we propose that *R. balsamica* (and possibly also *R. abietina*) are included in the subsection *Caninae*, since the morphological and genetic similarity of taxa in these subsections appears to be very high, and that the taxa in subsection *Caninae* are best regarded as one common species complex.

4.3. Support for taxa in subsection Vestitae

The genetic structure analysis was able to distinguish all five taxa in subsection *Vestitae* albeit with some overlap (Fig. 4g), while the dendrogram divided them into two distinct clusters (Fig. 3c) that are supported by species-related morphological characters. The first cluster consists of *R. pseudoscabriuscula* and *R. tomentosa*, both of which are characterised by uni- to multiserrated leaflets, and a narrow orifice (diameter smaller than, or equalling, 1 mm); the second cluster contains *R. sherardii*, *R. villosa*, and *R. mollis*, all characterised by a broader orifice (larger than 1 mm), erect and persistent sepals, and (irregular) multiserrated leaflet margins.

According to Henker (2000), only a few well-defined traits distinguish *R. pseudoscabriuscula* and *R. tomentosa*, and these were also the two most similar taxa in our analysis (Fig. 4g and h). The kinship between the morphologically very similar taxa *R. mollis* and *R. villosa* is stressed in the taxonomy of Nilsson (1967), who classified these taxa as subspecies: *R. villosa* ssp. *mollis* and *R. villosa* ssp. *villosa*. Nevertheless, these taxa are clearly distinguishable on the basis of the genetic structure analysis (Fig. 4g). RAPD- and SSR-based analyses have previously suggested that the pentaploid *R. sherardii* is derived by hybridization involving the tetraploid *R. mollis* as maternal parent (Olsson et al., 2000; Nybom et al., 2004, 2006).

Although the five taxa are supported by our molecular marker data, it should be noted that the they were, to a large extent, collected from different countries (Table 1); geographic differentiation may thus have contributed to the pattern with comparatively well-delineated taxa in this subsection. Moreover, the *R. villosa* samples could derive from naturalised material since the natural occurrence of this species may be restricted to the Alps.

4.4. Support for taxa in subsection Rubigineae

The shape of the leaflets divides the subsection *Rubigineae* into two groups: taxa with slender leaflets and a wedge-shaped base, such as *R. agrestis*, *R. inodora*, and *R. elliptica*, and taxa with broad leaflets and a well-rounded base: *R. rubiginosa*, *R. micrantha*, and *R. gremlii*. Within each group, different taxa are characterised by the theoretically well-defined L- and D-type differences (taxa of the L-type exhibit a lax growth habit, deciduous sepals and a narrow hip orifice (diameter below 1 mm) while taxa of the D-type show a dense and compact growth habit, have persistent sepals and a wide hip orifice). These differences can, however, be very subtle in the field.

Rosa inodora appears as a well-defined cluster in the genetic structure analysis (Fig. 4i) while *R. agrestis* and *R. elliptica* overlap strongly. Again, a geographic differentiation may have exaggerated the species delimitations, since *R. inodora* was collected in Sweden and Denmark, and the other two taxa in The Netherlands, Belgium and France.

A group with *R. micrantha*, *R. rubiginosa*, *R. gremlii* and putative hybrid plants (*R. rubiginosa* × *R. gremlii*) were analysed together (Fig. 4j). Although *R. micrantha* itself has been regarded as a hybrid (Ritz and Wissemann, 2011), this taxon showed some separation from the remainder indicating that it is easier to identify and also a genetically better supported taxon than *R. rubiginosa* and *R. gremlii* which overlapped more. Individuals combining characters of both *R. rubiginosa* and *R. micrantha*, and displaying transitional forms, are usually described as the intermediate species *R. gremlii* (syn. *R. columnifera* Henker, 2000) or as the subspecies *R. rubiginosa* subsp. *columnifera* (Wissemann, 2003). Alternatively, Graham and Primavesi (1993) described the descendant of *R. micrantha* × *R. rubiginosa* as *R.* × *bigeneris*. It can be assumed that hybridisations have occurred in mixed populations of *R. micrantha* and *R. rubiginosa* since a long time. However, in some areas like Thuringia in Germany, *R. gremlii* often occurs by itself and is more common than, e.g., *R. rubiginosa* (Zündorf et al., 2006). Therefore, we suggest assigning all the presumed *R. rubiginosa*, *R. gremlii*, and *R. micrantha* individuals to the same species complex regardless of how these taxa may be involved in more or less recent hybridization events.

4.5. The origin of the different subsections

Sequencing of the nrITS region in a large number of rose species has revealed five major groups of sequences (Ritz et al., 2005). One of these groups (C-type sequence) occurred only in the dogroses, whereas three other groups were represented in both dogroses and species of the other sections, suggesting an allopolyploid origin of the section *Caninae*. In another study, five different rDNA gene families were identified from extensive cloning of the ITS-1 region (Kovařík et al., 2008). The beta family in the study by Kovařík et al. (2008) proved to be identical to the C-type sequence previously described by Ritz et al. (2005). Moreover, this gene family occurred in all three investigated dogrose species whereas the other four gene families occurred in only 1–2 species each (Kovařík et al., 2008). In addition, analysis of pollen-derived DNA samples showed that the C-type is overrepresented in the pollen compared to in genomic DNA, suggesting that it occurs mainly in the bivalent-forming chromosomes (Kovařík et al., 2008) while the other types, present on univalent-forming chromosomes, are found also in other sections. Presumably a 'proto-*canina*' taxon carrying the C-type has thus hybridised to species in several other sections of *Rosa* to form the dogrose complex (Wissemann, 2002; Kovařík et al., 2008; Khaitová et al., 2010). Some data based on chloroplast DNA studies may instead indicate of a multiple origin (Wissemann and Ritz, 2005; Bruneau et al., 2005) but these results are difficult to interpret properly in a hybridogenous complex.

Putative hybridization events that may have produced the three large subsections are, at present, very difficult to identify and may in each case have involved several different taxa, some of which can have become extinct. Presumably, the smaller subsections

have arisen by more recent hybridizations involving extant taxa. Based on sequence analysis, Wissemann (1999) thus suggested that the hexaploid *R. marginata* in subsection *Trachyphyllae* is a hybrid between a species from section *Caninae* (seed parent) and *R. gallica* (section *Gallicanae*; pollen parent). While our dendrogram placed *R. marginata* outside the dogrose cluster (Fig. 3c), the genetic structure analysis suggested a strong affinity between *R. marginata* and subsection *Caninae* (Fig. 4b) but not with *R. gallica* (Fig. 4e). We did, however, only have material of two German populations of the first-mentioned species, and a set of French populations of the second species, and these may not have been representative of the actual hybridization event.

Morphologically *R. glauca*, subsection *Rubrifoliae*, belongs to section *Rosa* (formerly *Cinnamomeae*) according to Wissemann (2003) but is treated in section *Caninae* since it has the *canina* meiosis (Wissemann, 2003). In our dendrogram, this species was placed outside the dogrose cluster (Fig. 3c) as well, showing strong affinity with both subsection *Caninae* and *Rubigineae* in the genetic structure analysis (Fig. 4b) and a weaker affinity with *R. majalis* and *R. pendulina* in section *Rosa* (formerly *Cinnamomeae*; Fig. 4a). A hybrid origin thus seems likely, presumably with a dogrose taxon as maternal parent.

The morphology of two taxa, *R. stylosa* and *R.* × *irregularis* (generally treated as a hybrid due to low fertility) indicates an influence of *R. arvensis* (section *Synstylae*) on the one hand, and possibly *R. canina*, *R. corymbifera*, and/or *R. balsamica* (De Cock et al., 2007; van der Mijnsbrugge et al., 2010) on the other hand. Henker (2000) and Wissemann (2003) placed *R. stylosa* within subsection *Caninae*, whereas Graham and Primavesi (1993) created a separate subsection, *Stylosae*. Phylogenetic analyses based on AFLP polymorphisms placed *R. stylosa* with subsection *Rubigineae* (Koopman et al., 2008). Our dendrogram is indecisive concerning the position of *R. stylosa* whereas *R.* × *irregularis* is placed among taxa in subsection *Caninae* (Fig. 3c). In our genetic structure analysis *R.* × *irregularis* clusters with subsection *Caninae* while *R. stylosa* shows affinities with both this subsection and subsection *Rubigineae* (Fig. 4a and d). When analysed together with *R. arvensis*, no affinity with this species and *R.* × *irregularis* could however be seen, but for *R. stylosa* an origin involving *R. arvensis* looks plausible (Fig. 4l).

4.6. Hybridization among subsections within section Caninae

The derivation of new dogrose species through hybridization between other dogrose taxa has been suggested in numerous papers on dogrose taxonomy and floristics. Due to the general lack of discriminatory characters between taxa, as well as the predominantly matroclinal inheritance, it is, however, very difficult to detect spontaneous *Caninae* hybrids (Ritz and Wissemann, 2003). Also in our DNA marker profiles the species within each of the subsections largely overlap. In contrast, hybridization involving different subsections might be easier to identify. Experimental studies have shown that reciprocal hybrid families between species from different subsections can be distinguished both from the parental species and from one another, based on morphological characterisation as well as on dominant and co-dominant DNA markers (Werlemark, 2000; Werlemark and Nybom, 2001; Nybom et al., 2004, 2006).

Based on morphological characters, *R. canina* or *R. corymbifera* (seed parent) and *R. rubiginosa* (pollen parent) may have given rise to *R. dumalis*, while the reciprocal hybrids *R. rubiginosa* × *R. canina* (or *R. corymbifera*) could be responsible for *R. micrantha* (Ritz and Wissemann, 2003; Wissemann and Ritz, 2007). On the basis of a microsatellite DNA study, Ritz and Wissemann (2011) claim that *R. micrantha* (at least the hexaploid types) could have arisen from such hybridisation, involving unreduced gametes. Our genetic structure analysis placed *R. micrantha* firmly inside subsection *Rubigineae* but with some affinity to *R. canina* which may be indicative of a past hybridization event (Fig. 4k). No influence of subsection *Rubigineae* was found, however, in the analysis of *R. dumalis*.

Another putative hybridization between subsection *Caninae* and *Rubigineae* concerns *R. balsamica*, with the subsection *Caninae* species suggested as maternal parent (Wissemann, 2000). Our genetic structure analysis however placed *R. balsamica* firmly within subsection *Caninae* (Fig. 4d).

5. Conclusions

Analysis of the dogrose samples by PCO generated one relatively well-defined cluster that contained most of the samples belonging to *Rubigineae*. When a taxonomic structure was *a priori* superimposed, three large subsections were clearly discernible; *Rubigineae*, *Vestitae*, and *Caninae* + *Tomentellae*. In addition, three smaller subsections, possibly derived by hybridization between a dogrose seed parent and a pollen parent from another section, could be delineated: *Trachyphyllae*, *Rubrifoliae* and *Stylosae* (the latter supporting Graham and Primavesi, 1993). For *R. stylosa* we accept a hybridogenic origin possibly involving *R. arvensis*; nevertheless, this taxon is clearly positioned within section *Caninae*. We propose to include *R. balsamica* within subsection *Caninae* (and to omit the subsection *Tomentellae*). Within the large subsections, much interspecific overlap was noted except for *Vestitae*. Since different species had often been collected in different countries, sampling procedures may have superimposed a pattern of geographic differentiation on the actual genetic differentiation among the species, thus exaggerating the taxonomic differentiation. Partitioning of genetic variability was very different in dogroses compared to other rose sections. Two species complexes with low to very low support for species taxa were identified, a large complex within subsection *Caninae* from which only *R. montana*, *R. stylosa* and *R. balsamica* could be clearly differentiated, and a smaller complex within subsection *Rubigineae*, containing *R. micrantha*, *R. rubiginosa* and *R. gremlii*. The population structure analysis that we performed here on a large number of plants turned out to be a suitable tool to study differentiation among and within taxa with various degrees of hybridisation.

Acknowledgements

This study has been carried out with financial support from the Netherlands Ministry of Economic Affairs, Agriculture, and Innovation, and from the Commission of the European Community (QLRT-2001-01278, Genetic evaluation of European rose resources for conservation and horticultural use (Generose)). This study does not necessarily reflect the Commission's views and in no way anticipates the Commission's future policy in this area.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.02.024.

References

Bruneau, A., Joly, S., Starr, J.R., Drouin, J.N., 2005. Molecular markers indicate that the narrow Québec endemics *Rosa rousseauiorum* and *Rosa williamsii* are synonymous with the widespread *Rosa blanda*. Can. J. Bot. 83, 386–398.

De Cock, K, 2008. Genetic diversity of wild roses (*Rosa* spp.) in Europe, with an in-depth morphological study of Flemish populations. PhD thesis. Faculty of Bioscience Engineering, Ghent University, Belgium.

De Cock, K., Vander Mijnsbrugge, K., Breyne, P., Nybom, H., Smulders, M.J.M., Van Slycken, J., De Riek, J., 2007. The diversity of autochthonous roses in Flanders (Belgium) in view of the European GENEROSE reference framework. Acta Hort 760, 621–628.

De Cock, K., Van Der Mijnsbrugge, K., Breyne, P., Van Bockstaele, E., Van Slycken, J., 2008. Morphological and AFLP-based differentiation within the taxonomical complex section *Caninae* (subgenus *Rosa*). Ann. Bot. 102, 685–697.

De Riek, J., Calsyn, E., Everaert, I., Van Bockstaele, E., De Loose, M., 2001. AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet. 103, 1254–1265.

De Riek, J., Everaert, I., Esselink, D., Calsyn, E., Smulders, M.J.M., Vosman, B., 2007. Assignment tests for variety identification compared to genetic similarity-based methods using experimental datasets from different marker systems in sugar beet. Crop Sci. 47, 1964–1974.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol. Bioinformatics Online 1, 47–50.

Felsenstein, J, 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Graham, G.G., Primavesi, A.L., 1993. Rose of Great Britain and Ireland. B.S.B.I. Handbook No. 7 Botanical Society of the British Islas, London, p. 208. ISBN: 0901158224.

Henker, H., 2000. Hegi illustrierte flora van mitteleuropa. Parey Buchverlag Berlin 2000, 1–108.

Jacobs, M.M.J., Smulders, M.J.M., van den Berg, R.G., Vosman, B., 2011. What is in a name: genetic structure in *Solanum* section *Petota* studied using population-genetic tools. BMC Evol. Biol. 11, 42.

Jan, C.H., Byrne, D.H., Manhart, J., Wilson, H., 1999. Rose germplasm analysis with RAPD markers. HortScience 34, 341–345.

Khaitová, L., Werlemark, G., Nybom, H., Kovařík, A., 2010. Frequent silencing of rDNA loci on the univalent-forming genomes contrasts with their stable expression on the bivalent-forming genomes in polyploid dogroses (*Rosa* section *Caninae*). Heredity 104, 113–120.

Koopman, W.J.M., Vosman, B., Sabatino, G.J.H., Visser, D., Van Huylenbroeck, J., De Riek, J., De Cock, K., Wissemann, V., Ritz, C.M., Maes, B., Werlemark, G., Nybom, H., Debener, T., Linde, M., Smulders, M.J.M., 2008. AFLP markers as a tool to reconstruct complex relationships in the genus *Rosa* (Rosaceae). Am. J. Bot. 95, 353–366.

Kovařík, A., Werlemark, G., Leitch, A.R., Souckova-Skalicka, K., Lim, K.Y., Khaitová, L., Koukalova, B., Nybom, H., 2008. The asymmetric meiosis in pentaploid dogroses (*Rosa* section *Caninae*) is associated with a skewed distribution of rRNA gene families in the gametes. Heredity 101, 359–367.

Kurtto, A., Lampinen, R., Junikka, L. (Eds.), 2004. Atlas Florae Europaeae. Distribution of vascular plants in Europe. 13. Rosaceae (Spiraea to Fragaria, excl. Rubus). The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki. p. 320.

Lim, K.Y., Werlemark, G., Matyasek, R., Bringloe, J.B., Sieber, V., El Mokadem, H., Meynet, J., Hemming, J., Leitch, A.R., Roberts, A.V., 2005. Evolutionary implications of permanent odd polyploidy in the stable sexual, pentaploid of *Rosa canina* L. Heredity 94, 501–506.

Matsumoto, S., Kouchi, M., Fukui, H., Ueda, Y., 2000. Phylogenetic analyses of the subgenus *Eurosa* using the ITS nrDNA sequence. Acta Hort 521, 193–202.

Nilsson, Ö., 1967. Drawings of Scandinavian plants 1–8 *Rosa* L. Bot. Notiser 120, 1–408.

Nybom, H., Esselink, G.D., Werlemark, G., Vosman, B., 2004. Microsatellite DNA marker inheritance indicates preferential pairing between highly homologous genomes in polyploid and hemisexual dog-roses *Rosa* L. section *Caninae*). Heredity 92, 139–150.

Nybom, H., Esselink, G.D., Werlemark, G., Leus, L., Vosman, B., 2006. Unique genomic configuration revealed by microsatellite DNA in polyploid dogroses, *Rosa* section *Caninae*. J. Evol. Biol. 19, 635–648.

Olsson, Å.M.E., Nybom, H., Prentice, H.C., 2000. Relationships between Nordic dogroses (*Rosa* L. section *Caninae*, Rosaceae) assessed by RAPDs and elliptic Fourier analysis of leaflet shape. Syst. Bot. 25, 511–521.

Olsson, Å., Prentice, H.C., 2001. Morphometric diversity and geographic differentiation in six dogrose taxa (*Rosa* Sect. *Caninae*, Rosaceae) from the Nordic countries. Nord. J. Bot. 21, 225–241.

Paetkau, D., Waits, L.P., Clarkson, P.L., Craighead, L., Strobeck, C., 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. Genetics 147, 1943–1957.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rehder, A., 1940. Manual of Cultivated Trees and Shrubs. Macmillan, New York.

Ritz, C.M., Wissemann, V., 2003. Male correlated non-matroclinal character inheritance in reciprocal hybrids of *Rosa* section *Caninae* (DC.) Ser. (Rosaceae). Plant Syst. Evol. 241, 213–221.

Ritz, C.M., Wissemann, V., 2011. Microsatellite analyses of artificial and spontaneous dogrose hybrids reveal the hybridogenic origin of *Rosa micrantha* by the contribution of unreduced gametes. J. Hered. 102, 217–227.

Ritz, C.M., Schmuthz, H., Wissemann, V., 2005. Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. J. Hered. 96, 4–14.

Scariot, V., Akkak, A., Botta, R., 2006. Characterization and genetic relationships of wild species and old garden roses based on microsatellite analysis. J. Am. Soc. Hort. Sci. 13, 66–73.

Täckholm, G., 1922. Zytologische studien uber die gattung *Rosa*. Acta Hort. Berg. 7, 97–381.

Van Huylenbroeck, J., Smulders, M.J.M., Debener, T., Nybom, H., Gudin, S., Cox, P., Crespel, L., De Riek, J., 2005. GENEROSE: genetic evaluation of European rose resources for conservation and horticultural use. Acta Hort. 690, 119–123.

Vander Mijnsbrugge, K., De Cock, K., Cox, K., Breyne, P., 2010. Conservation measures for *Rosa arvensis* Huds. in Flanders (Belgium) based on congruent genetic and phenotypic population differentiation. Conserv. Genet. 11, 2243–2253.

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

Werlemark, G., 2000. Evidence of apomixis in hemisexual dogrose species, Rosa sect *Caninae*. Sex. Plant Reprod. 12, 353–359.

Werlemark, G., Nybom, H., 2001. Skewed distribution of morphological character scores and molecular markers in three interspecific crosses in *Rosa* section *Caninae*. Hereditas 134, 1–13.

Wissemann, V., 1999. Genetic constitution of *Rosa* section *Caninae* (*R. canina*, *R. jundzillii*) and section *Gallicanae* (*R. gallica*). J. Appl. Bot. 73, 191–196.

Wissemann, V., 2000. Epicuticular wax morphology and the taxonomy of *Rosa* (section *Caninae*, subsection *Rubiginosae*). Plant Syst. Evol. 221, 107–112.

Wissemann, V., 2002. Molecular evidence for allopolyploid origin of the *Rosa canina*-complex (Rosaceae, Rosoideae). J. Appl. Bot. 76, 176–178.

Wissemann, V., 2003. Conventional taxonomy (wild roses). In: Roberts, A.V, Debener, T., Gudin, S. (Eds.), Encyclopedia of Rose Science, Elsevier Academic Press, Oxford. p. 111–117.

Wissemann, V., Ritz, C.M., 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. Bot. J. Linn. Soc. 147, 275–290.

Wissemann, V., Ritz, C.M., 2007. Evolutionary patterns and processes in the genus *Rosa* (Rosaceae) and their implications for host-parasite co-evolution. Plant. Syst. Evol. 266, 79–89.

Zhang, J., Esselink, G.D., Che, D., Fougère-Danezan, M., Arens, P., Smulders, M.J.M., The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. J. Hortic. Sci. Biotech. **88**, 2013, 85-92.

Zündorf, H.-J., Günther, K.-F., Korsch, H., Westhus, W., 2006. Flora von Thüringen. Weissdorn Verlag, Jena.