Scientia Horticulturae 160 (2013) 251–263

Contents lists available at SciVerse ScienceDirect

Scientia Horticulturae

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

Structuration of the genetic and metabolite diversity among Prince Edward Island cultivated wild rose ecotypes

Kaushik Ghose[a,b], Jason McCallum[a], Sherry Fillmore[c], Chris Kirby[a], Kevin Sanderson[a], Simon Joly[d], Anne Bruneau[d], Bourlaye Fofana[a,*]

[a] Crops and Livestock Research Centre, Agriculture and Agri-Food Canada, 440 University Avenue, Charlottetown, C1A 4N6 PEI, Canada
[b] Department of Biology, University of Prince Edward Island, 550 University Avenue, Charlottetown, C1A 4P3 PEI, Canada
[c] Atlantic Food and Horticulture Research Centre, Agriculture and Agri-Food Canada, 32 Main Street, Kentville, B4N 1J5 NS, Canada
[d] Institut de recherche en biologie végétale, Jardin botanique de Montréal et Université de Montréal, 4101 Sherbrooke East, Montréal, Québec, Canada H1X

ARTICLE INFO

Article history:
Received 15 August 2012
Received in revised form 4 January 2013
Accepted 23 January 2013

Keywords:
Wild roses
Ecotypes
SSR
SNP
Field cultivation
Breeding
Prince Edward Island

ABSTRACT

Wild roses are widespread throughout North America but it is only recently that commercial interests have increased worldwide for rosehip products. Rosehip production under field conditions is new for North American wild roses and the cultivated materials are not well characterized. Despite recent phylogenetic studies on North American roses, micro-evolutionary studies at regional level in terms of genetic and metabolite diversity are scarce. We have characterized 30 wild rose ecotypes collected from different regions Prince Edward Island using microsatellite (SSR) markers, SNP of the chloroplast *trnaL* intron and the nuclear gene *GAPDH*, ploidy determination as well as the metabolite profiling for major phenolics and fatty acids. A total of 244 SSR alleles, with an average of 12.2 per marker were observed. The polymorphic information content (PIC) ranged between 0.00 and 0.97, with an average of 0.70. Cluster analysis showed three main clusters, cluster 3 showing the greatest diversity. Within each cluster, the ecotypes also showed large metabolite diversity, reflective of their genetic diversity. The data suggests the potential for the cultivation of a given ecotype for its unique metabolite profile or it incorporation in breeding programs. The ploidy and phylogenetic studies showed that all 30 ecotypes were tetraploid and that the collection is consisted of *Rosa virginiana* and its natural hybrids with *Rosa carolina*.

1. Introduction

The genus *Rosa* (*Rosaceae*) originated in the temperate regions of the northern hemisphere, but it is now widespread all over the globe (Werlemark and Nybom, 2010). The genus *Rosa* comprises more than 150 shrub species, although the species boundary delimitation remains a challenge for taxonomists and molecular biologists (Bruneau et al., 2007; Joly and Bruneau, 2007; Kimura et al., 2006; Macphail and Kevan, 2009). Recent molecular phylogenetic studies proposed a phylogenetic relationship between species of *Rosa* Sect. *Cinnamomeae* from North America. Using molecular markers and by conducting a haplotype network analysis through statistical parsimony, genealogical and multivariate analysis of 25 morphological characters (including ploidy determination based on stomata guard cell lengths) Joly et al. (2006) and Joly and Bruneau (2007) deciphered the evolutionary history of the polyploids and delineated the species boundaries within *Rosa* Sect. *Cinnamomeae* (*Rosaceae*) in eastern North America. Within the section *Cinnamomeae*, the North American *Rosa carolina* complex comprises five diploid species in the East of the Rocky Mountains and is composed of 2 diploid groups: one group consists of *R. blanda* and *R. woodsii* (which are indistinguishable from each other), while the other group is consisted of *R. foliolosa, R. nitida*, and *R. palustris* (Joly and Bruneau, 2006). The complex also includes 3 tetraploid species (*R. carolina* L., *R. virginiana* Mill., and *R. arkansana* Porter), and one hexaploid/octaploid species (*R. acicularis* Lindl.) (Lewis, 1957). There is also evidence of hybridization between the 3 tetraploid species, thus creating confounding effects. Joly et al. (2006) also proposed a list of diploids located east of the Rocky Mountains that are involved in the origins of the polyploid species. According to Joly et al. (2006), *R. arkansana* derived from the *blanda-woodsii* group, *R. virginiana* originated from the *foliolosa-nitida-palustris* group, and *R. carolina* is derived from a hybrid between the two diploid groups. Whereas the taxonomic problems are still well solved at the diploid level, where some species are morphologically difficult to distinguish, these problems are even more acute at the polyploid levels. *R. carolina* which is widespread east of the Mississippi river not only hybridizes with *R. arkansana* in the western part of its distribution (Erlanson MacFarlane, 1966) but also with *R. virginiana* in the

* Corresponding author. Tel.: +1 902 367 7602; fax: +1 902 566 7468.
E-mail address: bourlaye.fofana@agr.gc.ca (B. Fofana).

0304-4238/$ – see front matter © 2013 Crown Copyright and Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scienta.2013.01.016

eastern part. Moreover, the morphological similarity cuts across the ploidy levels and no single morphological character can be used to distinguish one species from another (Joly et al., 2006). In addition to this complex diversity at the species level within a large distribution range, a more complex and confounding polymorphism is observed at the micro-evolutionary level. Indeed, at a given location or site across a narrow distribution range such as the province of Prince Edward Island (Canada), different generations of siblings and parental genotypes grow in sympatry and are dispersed alongside with naturalized species such as *R. rugosa* or other dogrose species of the *Rosa* Sect. *Caninae* (Fofana, Personal observation). Thus, for wild rose species cultivation by propagation and commercial production purposes in the Canadian Maritimes, a careful species determination and a precise identification of genotypes that constitute a working collection are of critical importance to ensure genetic purity, stability and traceability. Despite extensive taxonomic and phylogenetic studies reported at the North American level (Watson, 1885; Crepin, 1896, 1889; Rehder, 1940, Erlanson MacFarlane, 1961, 1966; Bruneau et al., 2007), micro-evolutionary studies in terms of genetic and metabolite diversity at the regional level are scarce. Moreover, it is only recently that commercial interests have increased worldwide for wild rosehip products due to their nutriceutical and natural health product properties (Werlemark and Nybom, 2010). Thus, rosehips production from eastern North American native wild roses under field condition is new and is just emerging with the development of a collection and agronomic practices (Barry et al., 2008; Sanderson and Filmore, 2010). Although phylogenetic studies have established boundaries within the eastern North American *Rosa* section *Cinnamomeae* (Joly and Bruneau, 2007), the species status of the field propagated plant materials is unknown, nor is the genetic and metabolite diversity among the cultivated selections.

Microsatellite (SSR) and single nucleotide polymorphism (SNP) markers are commonly used for genetic diversity studies due to their high level of polymorphism, reliability and reproducibility (Gupta and Varshney, 2000; Deulvot et al., 2010). SSR markers have been successfully used to infer haplotype diversity (Kimura et al., 2006), QTL mapping (Dugo et al., 2005), genetic relationships (Rusanov et al., 2005; Scariot et al., 2006) in flower roses and in wheat germplasm (Fofana et al., 2008). SNPs occur frequently in plants (Zhu et al., 2003; Ching et al., 2002; Jing et al., 2007) and are useful genetic markers for applications such as genetic diversity studies or genetic mapping. SNP markers have been successfully used in genotyping and mapping studies of pea (*Pisum sativum*) (Deulvot et al., 2010), wheat (Akhunov et al., 2009), soybean (Hyten et al., 2008), and roses (Mercure and Bruneau, 2008).

The purpose of this study was to evaluate the extent of genetic and metabolite diversity among 30 wild rose selections collected throughout Prince Edward Island using SSR and SNP markers, assess whether plants grown in replicated field trials were true clones or not and to study the evolutionary history and species status for each entry.

Our data showed a high genetic and metabolite diversity within the working collection consisting only of tetraploid *R. virginiana* and its natural hybrids with *R. carolina*. The study showed a high potential for the selection and cultivation of a given ecotype having a unique metabolite profile for rosehip production or its consideration in breeding programs.

2. Materials and methods

2.1. Plant materials and field trials

The plants used for this study were the same studied by Sanderson and Fillmore (2010). They were originally taken as stem cuttings from plants provisionally identified as *Rosa virginiana* or *R. carolina* using standard floral keys but they may have included hybrids between *R. virginiana and R. carolina* as previously reported (Sanderson and Fillmore, 2010). The collection consisted of thirty wild rose selections collected at different locations (or sites) across Prince Edward Island, and referred to as ecotypes (Table 1). At planting time, all plants were well rooted, actively growing, and relatively homogenous in height (15–20 cm) and appearance with only one shoot and a few branches. The planting stock appeared to be free of pests and disease. The experiment was established at the Agriculture and Agri-Food Canada (AAFC) Crops and Livestock Research Centre in Harrington (46°22′N, 63°14′W), Prince Edward Island, PEI in 2005. The physical characteristics of the field and the agronomic practices are described in Sanderson and Fillmore (2010). The experimental design was a replicated lattice design with three replicates of each selection, and 30 different selections per block. A guard row planted with *Rosa rugosa* was established on each side of the entire trial to eliminate potential edge effects on measured plants. No data were collected from these guard rows. Each plot consisted of a single row of five plants, with 1 m between each plant and 4 m between plots. In June 2008, plants in each plot within replicate 1 (block 1) were morphologically similar in appearance. However, plant #3 in plot 143 (replicate 2, Block 2) and plants #4 and 5 in plot 188 (replicate 3, Block 3) were clearly morphologically different from the other plants in the same plot. These 3 plants had rather a similar morphology to each other. Leaf and mature rosehip (fruits) tissues were collected from 3 to 5 plants in each plot in June 2008 and bulked into a composite sample. Leaf and rosehips from plant #3 (plot 143) and plants #4 and 5 (plot 188) were collected separately from other plants of the same plot and maintained as a separated samples or as a bulked of 143–188 (plant #3–5). Samples were stored at −80 °C until processing.

2.2. Genotyping and genetic diversity

Genomic DNA was extracted from the 90 leaf samples using the Qiagen DNeasy 96 Plant Kit (Qiagen, Mississauga, Ontario). The microsatellite markers used in this study consisted of 53 pairs of SSR markers developed by Kimura et al. (2006) and Hibrand-Saint Oyant et al. (2008). The forward primer for each SSR marker was tailed with M13 primer (CACGACGTTGTAAACGAC) labeled with either FAM, HEX, or NED (ABI, Foster City, California), as described by Schuelke (2000) for the fluorescence detection of the SSR PCR products. Microsatellite PCR reactions were carried out in a 384-well thermocycler using a modified method of Röder et al. (1998) and as described by Fofana et al. (2009), with the thermal cycling performed at 95 °C/2 min; 30 cycles of 94 °C/1 min (−0.5 °C/s to 58/55 °C), 58/55 °C/50 s (+0.5° C/s to 72 °C), 72 °C/1 min, 1 cycle 72 °C/5 min. The internal molecular weight standard for the ABI3100 was Genescan 500-ROX (Applied Biosystems Inc., Foster City, California). The fluorescent-labeled PCR data were collected using capillary electrophoresis on an ABI 3100 genetic analyzer and converted to a gel-like image using Genographer (http://www.hordeum.oscs.montana.edu/genographer). Microsatellite alleles were sized using the internal molecular weight as standard and scored manually for presence or absence. To test the clonal propagation of the ecotypes across the three replicates and to infer the genetic relationship among the 30 ecotypes, two distinct genotypic datasets were created: one large dataset consisting of all 90 plots and one reduced dataset including the 30 plots of replicate 1.

Genotypic data were loaded into PowerMarker version 3.25 (Liu and Muse, 2005) for the estimation of the genetic diversity parameters such as gene diversity and polymorphic information content (PIC). The gene diversity is defined as the probability that two randomly chosen alleles from the population are different whereas the PIC is defined as the probability that the marker genotype of a given

Table 1
Site no., replicate 1 plot no., location, habitat and GPS location of all the 30 PEI ecotypes.

Site	Plot	Location	Habitat	GPS Coordinates
36	101	Meadowbank Rd. (in ditch)	Roadside/Dry	N46.19845; W63.24641
102	F1	Forestry	Forestry	
18	103	York Trail	Trailside/Dry	N46.31645; W63.09975
17	104	York Trail	Trailside/Dry	N46.31645; W63.09975
50	105	Hampton	Scrubland	N46.21162; W3.46317
19	106	York Trail	Trailside/Dry	N46.31645; W63.09975
55	107	Greenwich (civic # 1524)	Roadside/Dry	N46.4990; W62.64389
142	108	McEwen Creek Rd (Rte 352)	Trail/Roadside	N46.38706; W62.81619
139	109	Clyde River	Roadside	N46.20764; W63.26867
26	110	Brackley Beach	Riparian	N46.42286; W63.20035
24	111	St. Peter's	Trailside/Dry	N46.40991; W62.53864
33	112	Mt. Herbert (first)	Roadside/Dry	N46.23622; W63.04497
67	113	Cavendish (National Park 2)	Roadside/Dry	N46.49602; W63.40649
57	114	Opposite Midgell Cemetery	Roadside/Dry	N46.41410; W62.63442
29	115	Bristol	Roadside/Dry	N46.421768; W62.731220
61	116	Prince town Rd.	Hedgerow	N46.53819; W63.71207
25	117	Greenwich	Shore	N46.46553; W62.62265
141	118	Cavendish Beach, East	Field	N46.49875; W63.37981
F2	119	Forestry	Forestry	
137	120	Maximeville	Roadside	N46.42087; W64.13802
30	121	St. Andrews	Trail/Dry	N46.38226; W62.82985
28	122	Bristol - St. Peters harbor rd.	Roadside/Dry	N46.41746; W62.73174
122	123	New Dominion	Roadside/Dry	N46.17381; W63.24399
138	124	Rice Point	Field	N46.14117; W63.27130
32	125	Pisquid River (Rte 21)	Riparian	N46.33812; W62.85810
63	126	Fullerton's Marsh	Riparian	N46.23822; W63.05351
140	127	New London (Rte 224)	Roadside	N46.2819; W63.38529
22	128	Five Houses, Confed. Trail	Trailside/Dry	N46.40897; W62.53109
143	129	Naufrage	Roadside/Edge	N46.46373; W62.43458
68	130	Desable	Riparian	N46.2990; W63.425

offspring will allow deduction, in the absence of crossing-over, of which of the two marker alleles of the affected parent it received (Botstein et al., 1980). The PIC value is often used to measure the informativeness and discriminatory power of a genetic marker for linkage studies (Guo and Elston, 1999; Delgado-Martinez et al., 2012). The PIC value of an *n*-allele locus can be calculated as:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - 2\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p_i^2 p_j^2\right] \quad (1)$$

where p_i is the frequency of the *i*th allele, and *n* is the number of alleles (Botstein et al., 1980).

Pairwise genetic distances based on Nei's method (Nei, 1983) were estimated among observed operational taxonomic units (OTUs), herein referred to as rosehip ecotypes, and trees were constructed using neighbor joining and UPGMA methods. The 3 samples of *R. rugosa* were used as out group to root the tree. The output trees were visualized in TreeView v1.6.6 (Page, 1996).

2.3. Metabolite diversity

Frozen bulked rosehip flesh tissue samples, consisting of fruit from 3 to 5 individual plants in a plot, were ground into fine powder in an 80:20:0.1 (methanol:water:trifluoroacetic acid (TFA)) solvent mix using a polytron mixer. After centrifugation, the pellet was resuspended in the extraction solvent, extracted twice more and all supernatants were pooled and filtered. Phenolic compounds were separated, identified and quantitated using an Agilent 1100 series, quaternary pump HPLC system (Agilent Technologies, Mississauga, ON, CA) equipped with in-line degassing, diode array detector (DAD), robotic autosampler, sample and column temperature controls. A tertiary solvent system consisting of water:TFA (100:0.1), water:methanol:TFA (50:50:0.1), and acetonitrile was used for HPLC–MS analysis. UV–Vis spectra were recorded from 190 to 900 nm. Anthocyanins were monitored at 520 nm, flavonols at 360 nm, and tannins (catechin-based oligomers + hydrolysable gallotannins) at 280 nm. Reported units are based on absorbance integrals (mAU*s), and are normalized on a per g starting tissue basis, facilitating direct comparisons between sites.

The fatty acids were extracted from 1 g of seed powder with deoxygenated hexane using ASE 150 Accelerated Solvent Extractor. The extract was collected in pre-weighed glass vials and the solvent evaporated using a rotary evaporator. The weight of the oil after evaporation was recorded and an aliquot of each oil sample was diluted with 600 μL of $CDCl_3$ and transferred to 5 mm NMR tubes and the relative amount of individual fatty acid determined as described by Fofana et al. (2011).

2.4. Correlating the genetic and metabolite diversity

To infer the relationship between the genotypic data and the metabolite profiles obtained from each of the rose selection, a similarity matrix of Euclidian distances was created for each set of data. The metabolite data was run against the SSR genotypic data obtained from the rose selections using a principal component analysis (PCA). A biplot was used to display the relationships between the rose selections and their relationships with the metabolite profiles. A ranked correlation test was performed using 100 times permutation of the two similarity matrices for a valid Mantel test.

2.5. Phylogenetic analysis using chloroplast DNA trna-Leu SNP variation

Genomic DNA samples used for this marker were the same as in the SSR genotyping. However, since the species status of the collection was initially unknown, DNA samples of known *R. blanda*, *R. nitida*, *R. carolina*, and *R. virginiana* specimens from Joly et al. (2006), were included as references species (Table 2). Universal chloroplast primers C (Forward) 5′-CGAAATCGGTAGACGCTACG-3′ and D (Reverse) 5′-GGGGATAGAGGGACTTGAAC-3′ (Taberlet et al., 1991) were used to amplify the intronic region of the *trnaL* from each of the 30 ecotypes and the 5 reference wild rose species

Table 2
Twenty-two (22) reference *GAPDH* allele sequences of *Rosa* species obtained from NCBI GenBank database. For each sequence, the voucher information, accession numbers, alleles sampled, collector numbers, location, and approximate co-ordinates are provided.

	Species	Accession	Alleles[a] (NCBI acc. #)	Collector(s)	Province[b]	Lat., Long
	R. blanda	161	–	Joly&St. 410	NB	45°57′43.7″N; 67°22′26.1″W
cp-trnaL	*R. blanda*	819	–	Joly, S. 1011-1	Quebec	46°22′54.7″N; 75°01′21.9″W
	R. carolina	626	–	Joly & St. 967	NB	47°22′36.2″N; 66°04′42.2″W
	R. nitida	604	–	Joly & St. 941	NB	45°56′29.2″N; 64°52′07.3″W
	R. virginiana	182	–	Joly & St.431	NB	45°05′00.4″N; 67°03′01.1″W
	R. virginiana	195	–	Joly & St. 444	Maine	44°30′56.7″N; 68°11′14.6″W
	R. nitida	812	–	Joly, S. 1010-1	Quebec	46°22′45.3″N; 75°00′20.6″W
	R. carolina	241	–	Joly & St. 491	PA (USA)	40°24′29.4″N; 75°19′11.3″W
	R. rugosa	613	–	Joly & St. 950	NB	46°28′24.5″N; 64°43′10.7″W
	R. rugosa	218	–	Joly & St. 467	MA (USA)	42°41′05.2″N; 70°45′56.3″W
	R. rugosa	316	–	Joly & St.572	Quebec	
	R. blanda	160	A, B (DQ091014.1, DQ091015.1)	Joly& St. 409	NB	45°57′43.7″N; 67°22′26.1″W
GAPDH	*R. blanda*	621	A, B (DQ091028.1; DQ091029.1)	Joly et al., 962	NB	47°22′32.5″N,; 66°04′30.5″W
	R. carolina	626	A, B (DQ091139.1; DQ091140.1)	Joly et al., 967	NB	47°22′36.2″N; 66°04′42.2″W
	R. palustris	168	A, B (DQ091056.1; DQ091057.1)	Joly & St. 417	NB	45°33′43.2″N; 67°25′31.2″W
	R. virginiana	182	A, B, C (DQ091145.1; DQ091146.1; DQ091147.1)	Joly& St. 431	NB	45°05′00.4″N; 67°03′01.1″W
	R. virginiana	587	A, B, C, D (DQ091162.1; DQ091163.1; DQ091164.1; DQ091165.1)	Joly et al., 924	NS	45°43′09.7; 61°53′56.3″W
	R. virginiana	684	A, B, C (DQ091169.1; DQ091170.1; DQ091171.1)	Brou. 03-60-1	NF	–
	R. nitida	604	A, B (DQ091050.1; DQ091051.1)	Joly et al., 941	NB	45°56′29.2″N, 64°52′07.3″W
	R. nitida	675	A, B (DQ091052.1; DQ091053.1)	Brou. 03-55-1	NF	-

[a] Allele NCBI accession numbers are in brackets.
[b] NB, New Brunswick; NF, Newfoundland; NS, Nova Scotia.

and sequenced for SNP analysis. Five published *trna*L sequences (*Rosa rugosa* DQ778884.1, *R. virginiana* DQ778888.1, *R. carolina* DQ778840.1, *R. blanda* DQ778858.1 and *R. nitida* DQ778873.1) were also included as reference sequences. SNPs analysis was performed by multiple alignments using ClustalW and the phylogenic tree was constructed using MEGA4 (Tamura et al., 2007).

2.6. Ploidy determination

Two healthy young leaves were collected from each of the 5 individual plants per plot for each of the 29 ecotypes, for a total of 10 leaves per ecotype. One ecotype was missing as a result of winter kill at the time of this study. *R. rugosa* was used as a diploid reference genotype. Using a Carl Zeiss microscope at 40× magnification, the length of five stomatal guard cells was measured per leaf, for a total of 50 guard cell measurements for each ecotype.

2.7. GAPDH amplification, allele sampling and haplotype analysis

GAPDH was used as a target DNA for PCR amplification. Similar to the chloroplast SNP analysis, DNA from the 30 ecotypes was used as template. Moreover, a total of 22 *GAPDH* allelic forms from *R. blanda*, *R. nitida, R. virginiana, R. carolina*, and *R. palustris* were obtained from NCBI database and used as reference sequences (Table 2). These sequences were selected based on the origin of their source plant materials. Sequences from plants collected from Atlantic Canadian Provinces (New Brunswick, Nova Scotia and Newfoundland) were selected with the assumption that they may be more related to PEI wild roses than those from elsewhere. The 22 sequences were aligned using ClustalW and *GAPDH* forward (*GAPDH*-F 5′-CCATGTGAGATATATGAATG-3′) and reverse (*GAPDH*-R 5′-AATTCACCACTGCAACTTCAAGA-3′) primers were designed from the conserved consensus sequence. These primers were used to amplify the genomic region corresponding to *GAPDH* in the 30 ecotypes of the collection. PCR reactions were performed in a final volume of 50 μL containing 1× PCR reaction buffer, $MgCl_2$ concentration of 1.5 mM, 0.2 μM of each primer, 200 μM of each dNTP, 2 U Taq polymerase (Life Technologies, Burlington, Ontario, Canada), and 100 ng genomic DNA. PCR reactions were performed on a G-storm thermal cycler (VWR International, Mississauga, Ontario, Canada) using an initial denaturation at 95 °C for 3 min, followed by 30 cycles consisting of a denaturation step at 95 °C for 30 s, an annealing step at 55 °C for 30 s, and an extension step at 72 °C for 1 min. The final extension step was run at 72 °C for 7 min. The 725 bp PCR products were then purified using Qiawich PCR product purification kit (Qiagen, Mississauga, Ontario, Canada) following the manufacturer's protocol.

2.8. Allele sampling

GAPDH allelic forms were sampled in each of the 30 ecotypes. Because it is easy to miss polymorphic sites from direct sequencing when four alleles may be present in the genome, the *GAPDH* PCR amplicon from each ecotype was cloned using the TOPO TA cloning kit (Life Technologies, Burlington, Ontario, Canada) to capture all allelic forms present in the tetraploid species. Positive clones were confirmed by colony PCR using gene specific primers and plasmids were extracted from *E. coli* using the QIAprep Spin kit (Qiagen, Mississauga, Ontario, Canada) and sequenced (Genome Quebec, Montreal, QC). The resulting sequences were trimmed to remove vector sequences and filtered through the DNA Baser software to eliminate bad sequences and to assign quality value to each of the nucleotides. Assuming that there is no PCR bias between alleles and that all alleles are cloned with equal probability, the binomial distribution calculates the probability of sampling all alleles in an individual according to the following formula:

$$P = \left[1 - \left(\frac{t-1}{t}\right)^{n}\right]^{t}, \quad (2)$$

where t is the number of alleles in the individual and n is the number of clones sequenced (Joly et al., 2006). For a tetraploid with four alleles, 15 clones are required in order to get a 95% probability that all alleles have been sampled. If only three alleles are present in the tetraploid, 11 clones would be required. At least fourteen clones were sequenced from each polyploid ecotype, with additional clones sequenced in all cases where the alleles resulting in polymorphisms detected in direct sequencing were not recovered. Taq-induced PCR errors were identified and removed from analyses by comparing the sequence of the cloned amplicons to one another and to the initial sequences obtained from direct sequencing. Henceforth, it was assumed that all alleles were retrieved from each individual even if there is a non-zero probability that some alleles were not sampled in some individuals.

Table 3
Gene specific, insertion allele-specific, and multiplex MAMA PCR primers used in the study. The Tm, position of the mutation within the 725 bp *GAPDH* amplicon and the haplotype identified and confirmed by MAMA multiplex PCR are shown. The underlined nucleotide is the additional mismatch located at the MAMA primer's penultimate 3′nucleotide. The bold nucleotide is the mutant allele specific nucleotide.

Primer name	Sequence of the primer (5′-3′)	Tm (C)	Mutation position	Haplotype
MAMA 5R	CCTGTTGAGACAGAATTAAAGT**C**	54	519	Haplotype 2
MAMA 6R	CCACTGCAACTTCAAGATGAC**G**	53	697	Haplotype 1
MAMA 8F	TCAGTTGTTGACCTCACTG**T**	54	383	Haplotype 7
MAMA 9R	GATGACCACTAATCCAAACA**A**	53	683	Haplotype 3
INSERTION Forward	GACAGCAGGTAAATGAATT**T**	53	634	Haplotype 5
GAPDH Forward	CCATGTGAGATATATGAATG	55	–	–
GAPDH Nested Forward	CCAGATCAAGGCTGCTATCA	55	–	–
GAPDH Reverse	AATTCACCACTGCAACTTCAAGA	55	–	–

2.9. Haplotype analysis

On average 10–14 good sequences were obtained from each ecotype, for a total of 361 cloned sequences. These sequences, along with the 22 *GAPDH* sequences from *Rosa* reference species reported in NCBI were incorporated into MEGA4, aligned and clustered using ClustalW 1.6 (Thompson et al., 1994) using default parameters (Tamura et al., 2007). From a total of 361 sequences, a reduced dataset of 84 sequences including at least one representative sequence per ecotype for each of the 6 main clusters obtained was created and used to build a phylogenetic tree using neighbor joining and maximum composite likelihood methods, with 100 bootstrapping replicates. To further define affinity between haplotypes and to infer putative origins of the ecotypes present in the collection, 208 (at least two representative clones per ecotype) including all potential allelic variants were selected from the 361 GADPH sequences and introduced along with known reference species haplotypes for haplotype analysis using software Collapse version 1.2. A genetic affinity analysis of haplotype network for 30 unknown ecotypes and known reference species was performed using statistical parsonimy network (gap = missing) along with haplotypes from Joly et al. (2006) to look for the affinities of the haplotypes and using maximum likelihood tree (phyml – GTR + I + G) analysis (Guindon and Gascuel, 2003; Joly and Bruneau, 2006) to determine with which PEI rose *GAPDH* alleles the known haplotypes clustered when the statistical parsimony haplotype network was ambiguous.

2.10. Confirmation of detected SNPs and mutations by MAMA PCR-based genotyping

Because Taq polymerase and sequencing errors can mislead in allele identification based on SNP discovery through direct sequencing, a mismatch amplification mutation assay PCR (MAMA PCR) genotyping (Li et al., 2004, 2008) was performed using species-specific allele to (1) confirm the *GAPDH* allele sampling data, (2) identify any missing alleles that might be present in the genome, and (3) to predict the genotype (combination of different alleles present in the genome) of each ecotype. The MAMA PCR primers were designed based on the identified mutations (Table 3). The primers were designed in such a way that the 3′ nucleotide is complementary to the mutant and the penultimate nucleotide from 3′ end mismatches to both mutant and wild type DNA. The position of the mutations, primer sequences, annealing temperature and PCR product size are presented in Table 3. Using these MAMA primers, multiplex PCR were performed with one additional *GAPDH* nested forward primer and the previously used *GAPDH* primers. PCR reactions were performed using standard condition described above for *GAPDH* amplification but with a slight modification by including 5 touchdown cycles (55–50 °C) after the initial denaturation. PCR products were then run on 1% agarose gel containing ethidium bromide to check the presence/absence of mutation specific band and image captured.

3. Results

3.1. Genetic and metabolite diversity

The genetic diversity among 30 wild rose ecotypes was assessed using microsatellite markers. Twenty SSR markers showing good amplification and informative alleles were used for the genetic diversity studies. Using the reduced genotypic dataset obtained from the 30 wild rose ecotypes of replicate bloc 1 (replicate #1), the gene diversity and the genetic relationships between ecotypes were assessed. A total of 244 alleles, with an average of 12.2 per marker were detected. The observed polymorphic information content (PIC) ranged from 0.00 to 0.97 for Rw62D8 and Rw35C24 SSR markers, respectively, with an average of 0.70 (Table 4). To infer the genetic relationship between ecotypes, a frequency-based distance analysis was performed using the distance methods described by Nei (1983). Based on the pairwise matrix of distance, the output neighbor-joining tree showed three major clusters (Fig. 1A). Nineteen (19) of the ecotypes grouped together in cluster 3, whereas 8 and 3 of the ecotypes grouped together in cluster 2 and 1, respectively. The cluster 3 was most diverse with 4 sub-clusters (3-1 to 3-4) whereas cluster 2 had 2 sub-clusters (2-1 and 2-2).

To determine whether the observed genetic diversity is also reflected by a metabolite diversity, the metabolite profiling of the 30 ecotypes showed a large diversity among ecotypes for major metabolites such as phenolics, flavonols, anthocyanins, tiliroside, tannins, and fatty acids (Fig. 1B–F). Fourteen of the ecotypes showed an omega-3 fatty acid 18:3 content higher than 35% with ecotype 28=R122 showing the greatest level (41.2%) of omega-3 fatty acid 18:3. Ecotypes 17=R104, 19=R106, 57=R114, and 68=R130 showed the highest content for both total flavonol and tiliroside, a metabolite of special interest due to its anti-obesity activity (Ninomiya et al., 2007). The relationships between rose selections and the metabolites they produced were established using PCA analysis (Fig. 2). The metabolites that are correlated appeared close together. For example, flesh anthocyanin (anthFlesh) and flesh flavonol (flavFlesh) were found to be correlated. Similarly, saturated fatty acids and 18:1 appeared correlated but negatively correlated with 18:3, 18:2 which were found to be correlated with each other. Total tannin (TanTot) was found negatively correlated with fuzz flavonol (flavFuzz) and fuzz tiliroside (tiliFuzz). The whole group of metabolites consisting of seed flavonol (FlavSeed), seed tiliroside (tiliSeed), total flavonol (flavTot), total tiliroside (tiliTot), fuzz tiliroside (tiliFuzz) and fuzz flavonol (flavfuzz) were correlated to some degree. The rose selections producing a particular metabolite were clustered around this metabolite and their content decreased toward the center of the PCA graph. For example, the selection R120 was found to be high in flesh

Table 4
Gene diversity and polymorphic information contents for 20 microsatellite markers selected from Kimura et al. (2006) and Hibrand-Saint Oyant et al. (2008).

Marker	Sample size	Allele frequency	Allele no	Gene diversity	PIC
RA003a	31	0.419	8	0.768	0.746
RW54N22	31	0.581	6	0.610	0.573
H24D11	31	0.387	3	0.662	0.588
RW35C24	31	0.065	30	0.966	0.965
RW55E12	31	0.097	25	0.953	0.951
RW16E19	31	0.097	23	0.947	0.944
RW59A2	31	0.065	28	0.961	0.960
RW20I17	31	0.226	13	0.868	0.855
RW25J16	31	0.194	14	0.891	0.881
RW34L6	31	0.323	8	0.774	0.742
RW52D4	31	0.323	11	0.805	0.783
RW62D8	31	1.000	1	0.000	0.000
C139	31	0.226	20	0.907	0.902
CL2845	31	0.742	4	0.418	0.384
CTG21	31	0.516	11	0.703	0.687
C172	31	0.258	14	0.876	0.866
CL2996	31	0.452	5	0.687	0.636
CTG623	31	0.387	10	0.801	0.785
CL2002	31	0.968	2	0.062	0.060
H1F03	31	0.355	8	0.770	0.738
Mean	31	0.384	12	0.722	0.702

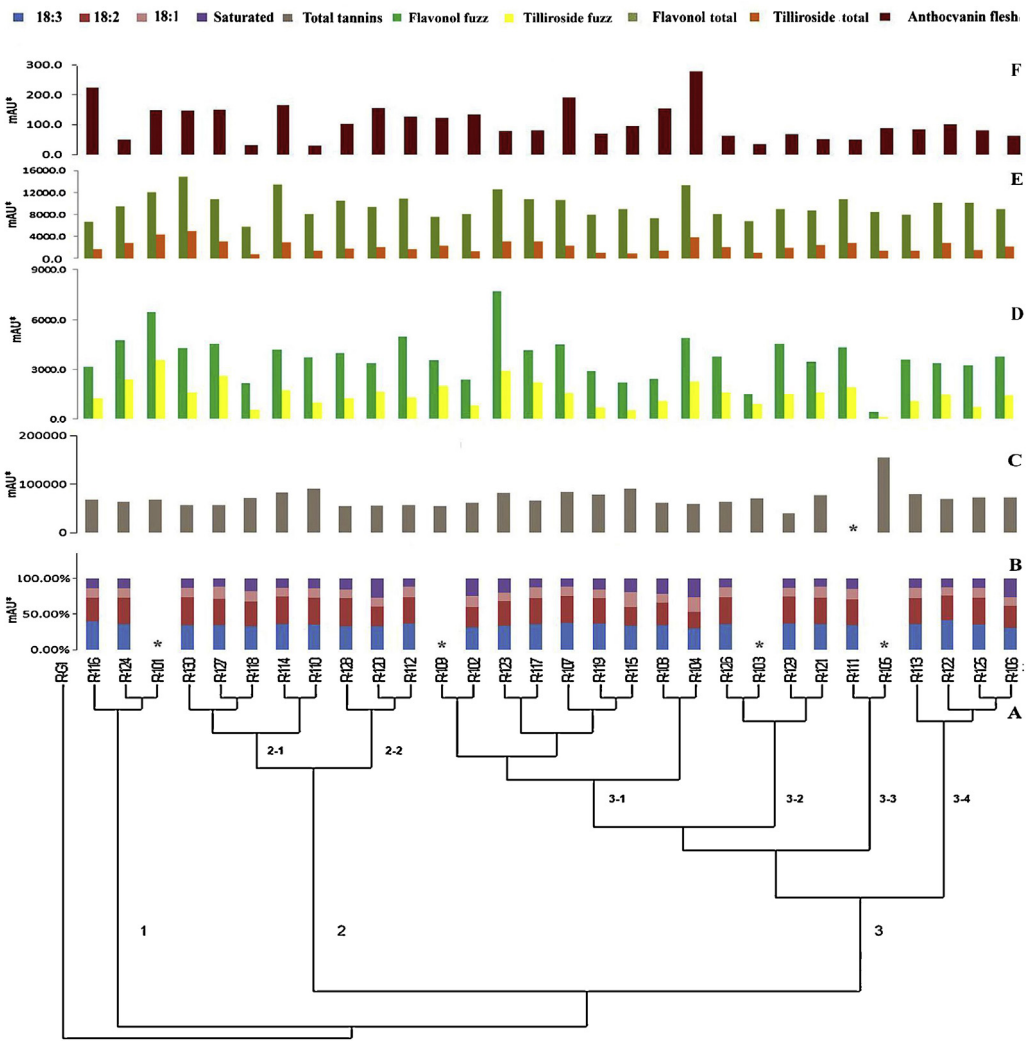

Fig. 1. Genetic (A) and metabolite (B–F) diversity in 30 wild rose ecotypes originating from Prince Edward Island. (B) Seed fatty acids (18,3, 18,2, 18,1, and 18,0 +16,0); (C) total tanins; (D) fuzz flavonol and tiliroside; (E) total flavonol and tiliroside; (F) flesh anthocyanins. Missing data are represented by an asterix.

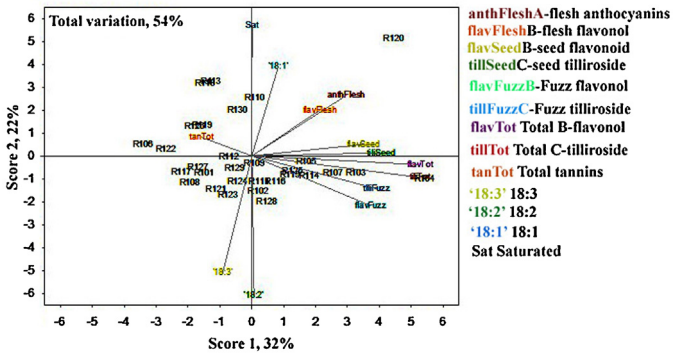

Fig. 2. Correlation between the genetic and metabolite diversity as revealed by principal component analysis. Scores 1 and 2 represent the largest amount of variation explained by the PCA for a total of 54%. The metabolites that are correlated appear close together. For example, anthFlesh and flavFlesh appeared correlated. Similarly, saturated fatty acids and 18:1 were correlated but were negatively correlated with 18:3, 18:2 which were found correlated with each other.

anthocyanins (anthFlesh) and flesh flovonol (flavFlesh) whereas rose ecotypes R122, R106, R119, R126, R113, R118, R130 and R110 were high in total tanins (tanTot). The Mantel test showed an original correlation of $r^2 = 0.03878$ between the two matrices.

To insure that plants in the three replicated blocks were propagated as clones, the genetic relationship among 90 rose samples representing 30 ecotypes in 3 replicates was evaluated using the 20 microsatellite markers. Using the large genotypic dataset, the UPGMA dendrogram (Fig. 3) showed the plants from the three replicates for 18 ecotypes grouped together and, thus being most similar to each other if not clonally replicated. Likewise plants from two replicates for 10 ecotypes grouped together, the plants from one of the replicate standing as a singlet on the tree. In a third scenario, two ecotypes had plants from all three replicates as a singlet on the tree, and thus, these plants were not clones of each other (Fig. 3).

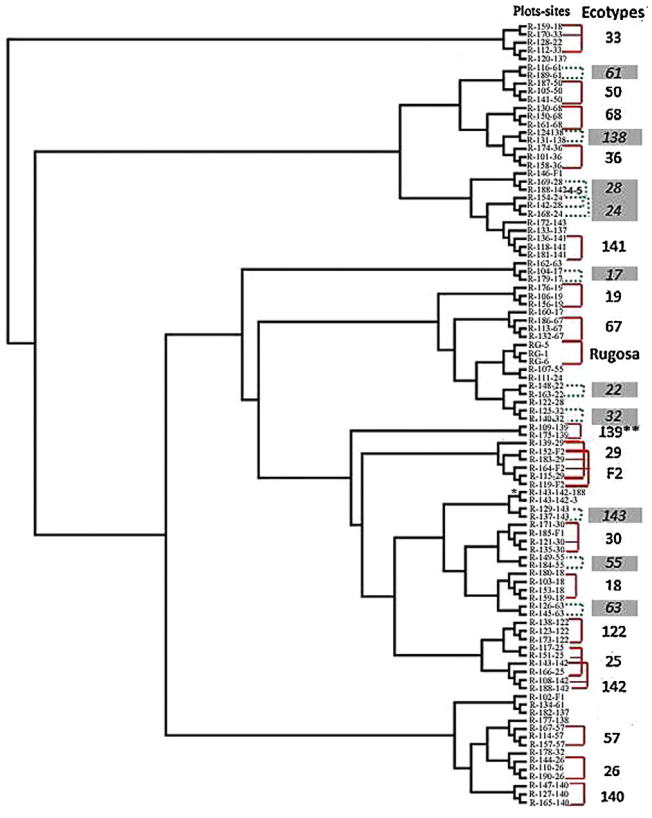

Fig. 3. UPGMA dendrogram based on the Nei (1983) matrix of distance and showing the grouping of stems cuttings grown in triplicated blocks. 18 ecotypes grouped and indicated with red solid lines were identical in all 3 replicates; 10 ecotypes grouped and indicated with green dashed lines had 2 replicates that are identical and one replicate different; 2 ecotypes (F1 and 137) were different in all three replications. The three plants within plots 143 and 188 that showed morphological variations from others but similar to each other are indicated by an asterisk. The double asterisk indicates that DNA sample was obtained only from two replicates for this ecotype. (For interpretation of the references to color in this artwork, the reader is referred to the web version of the article.)

3.2. Characterization of the collection

3.2.1. Chloroplast DNA phylogeny

Using the 600 bp of chloroplast *trna*L sequence, a single SNP (C/T) at position 196 separated the 30 ecotypes into two clades on the

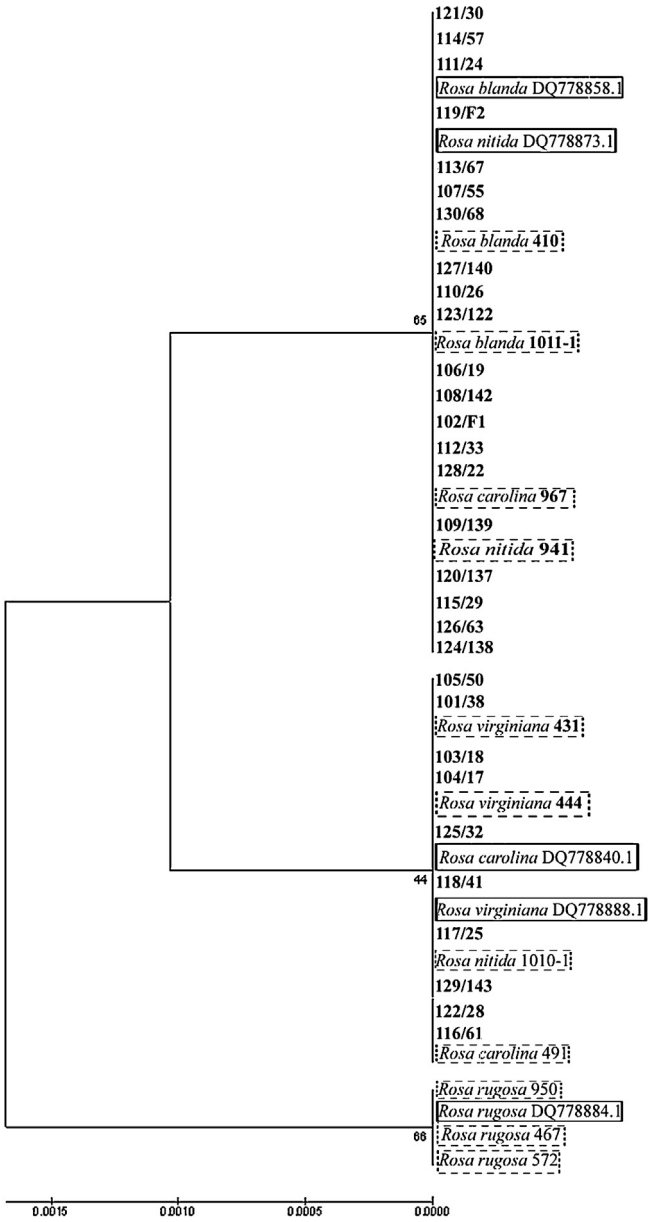

Fig. 4. Phylogenetic relationship between the 30 PEI ecotypes and the known diploid/tetraploid rose species. PEI rose ecotypes are indicated by plot-site numbers from which the sequences were obtained. Reference *Rosa* chloroplast *trna*L sequences obtained from GeneBank are indicated by their accession numbers boxed with a solid line. Chloroplast *trna*L amplified from reference wild *Rosa* species DNA are indicated by their collection numbers boxed with a broken line.

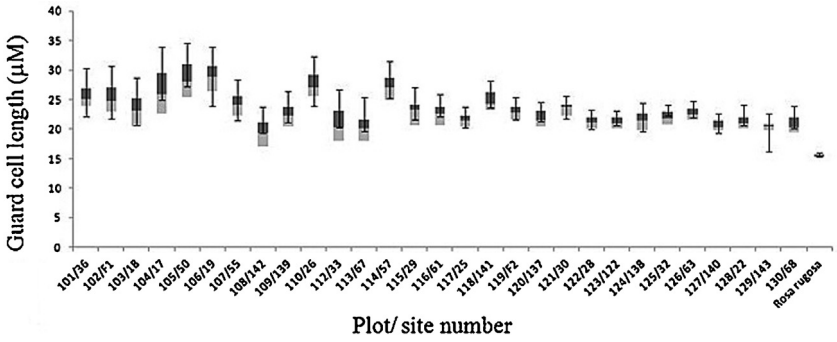

Fig. 5. Box plot diagram representing the range of length variation of guard cells for 30 rose ecotypes evaluated in field in PEI and compared to the diploid species *Rosa rugosa* taken as control. The ecotypes are indicated by plot-site numbers.

phylogenetic tree. Ten of the ecotypes were grouped together with the reference species *R. carolina*, *R. virginiana, R. nitida* whereas the remaining 20 were grouped with both the reference diploid species *R. blanda*, and *R. nitida*, and tetraploid *R. carolina* and *R. virginiana* (Fig. 4).

3.2.2. Ploidy

The guard cell measurement data showed that all the 29 ecotypes, except for the ecotype 142, had a guard cell length greater than 19.30 µm as compared to the diploid *R. rugosa* which had a guard cell length of 17.81 µm. The observed guard cell lengths were in the range of tetraploid species. Ecotype 142 showed a stomata guard cell length of 18.56 µm and thus lower than 19.18 µm. Compared to the diploid control *R. rugosa*, ecotype 142 had a stomata guard cell length larger than this well-known diploid species that showed a 17.81 µm as guard cell length (Fig. 5).

3.3. GAPDH allele sampling and sequence variation analysis

3.3.1. GAPDH alleles sampling and phylogeny

To assess the evolutionary history and species status for each ecotype, we conducted species-specific allele sampling, haplotype affinity and discrimination study using SNP analysis for the presence or absence of reference alleles in the studied population and to infer the relatedness between ecotypes within the collection and known wild rose species. All 30 ecotypes were shown to carry the *R. nitida*-specific allele 604B and the *R. virginiana*-specific alleles 684B and 587 D (haplotype V-o). Only 13 ecotypes showed the *R. blanda*

		Observed haplotypes										
Ecotype	Plot	Haplotype 1. (=Haplotype V-o) *R. nitida* allele 604 B and *R V*- specific 684B and 587D	Haplotype 2 (Haplotype IV-r) – *R. Blanda* allele 528 B and *R C*- specific 626B or RV 684C	Haplotype 3 (Haplotype ambiguous)	Haplotype 4 (Haplotype V-b) R. nitida allele 675A	Haplotype 5 (Haplotype ambiguous) Insertion	Haplotype 6 (Haplotype V-o sister) Deletion	Haplotype 7 (haplotype V-d)	No. of haplotype**	No. of alleles***	Mean guard cell length (µm)	Species status
		MAMA 6R*	MAMA 5R*	MAMA 9R*	-	-	-	MAMA 8F*				
36	101	+	-	-	-	-	-	-	1	1/1	25.5	Pure RV
F1	102	+	+	+	+	+	-	*	6	4/6	25.2	RV & RC Hybrid
18	103	+	-	-	+	-	-	*	3	3/3	23.3	Pure RV
17	104	+	+	+	+	-	-	*	5	4/5	27.4	RV & RC Hybrid
50	105	+	+	-	+	-	-	*	4	4/4	28.1	RV & RC Hybrid
19	106	+	+	-	+	-	-	-	3	3/3	28.4	RV & RC Hybrid
55	107	+	-	-	+	-	-	-	2	2/2	24.3	Pure RV
142	108	+	+	-	+	-	-	-	3	3/3	18.6	RV & RC Hybrid
139	109	+	-	-	+	-	+	*	4	4/4	22.9	Pure RV
26	110	+	+	+	-	-	-	*	4	3/4	27.7	RV & RC Hybrid
24	111	+	-	*	+	+	-	-	4	2/4	-	Pure RV
33	112	+	-	-	+	+	+	*	5	4/5	21	Pure RV
67	113	+	-	-	+	-	-	-	2	2/2	21.9	Pure RV
57	114	+	-	-	+	-	-	*	3	3/3	26.9	Pure RV
29	115	+	+	-	+	-	-	-	3	3/3	23.9	RV & RC Hybrid
61	116	+	+	-	+	+	-	-	4	3/4	22.5	RV & RC Hybrid
25	117	+	+	-	+	+	-	-	4	3/4	21.4	RV & RC Hybrid
141	118	+	-	-	+	-	-	*	3	3/3	24.5	Pure RV
F2	119	+	+	-	+	-	-	-	3	3/3	23	RV & RC Hybrid
137	120	+	-	-	+	-	-	-	2	2/2	22	Pure RV
30	121	+	-	-	-	+	-	-	2	1/2	23.3	Pure RV
28	122	+	-	-	-	-	-	-	1	1/1	21.3	Pure RV
122	123	*	+	+	-	+	-	*	5	3/5	21	RV & RC Hybrid
138	124	+	-	-	-	+	-	-	2	1/2	21.5	Pure RV
32	125	+	-	-	+	-	-	-	2	2/2	21.8	Pure RV
63	126	+	-	+	+	-	-	*	4	3/4	22.4	Pure RV
140	127	+	+	-	+	-	-	*	4	4/4	20.1	RV & RC
22	128	+	+	-	+	+	-	*	5	4'/5	21.7	RV & RC Hybrid
143	129	+	-	+	+	-	-	*	4	4/4	19.2	Pure RV
68	130	+	-	+	+	-	-	*	4	3/4	20.9	Pure RV

Fig. 6. *GAPDH* allelic variation and species status of PEI wild rose ecotypes as depicted by SNP, MAMA PCR, ploidy analysis. *Alleles confirmed and identified by MAMA PCR; **Number of haplotypes identified and confirmed by allele sampling and MAMA; ***Number of *GAPDH* alleles found in this ecotype among the 7 different haplotypes groups excluding ambiguous haplotypes following the allele sampling and MAMA PCR/number of *GAPDH* alleles found in this ecotype among the 7 different haplotypes groups including ambiguous haplotypes following the allele sampling and MAMA PCR; alleles identified only by MAMA PCR are indicated by an asterisk; RV, *Rosa virginiana;* RC, *Rosa carolina*.

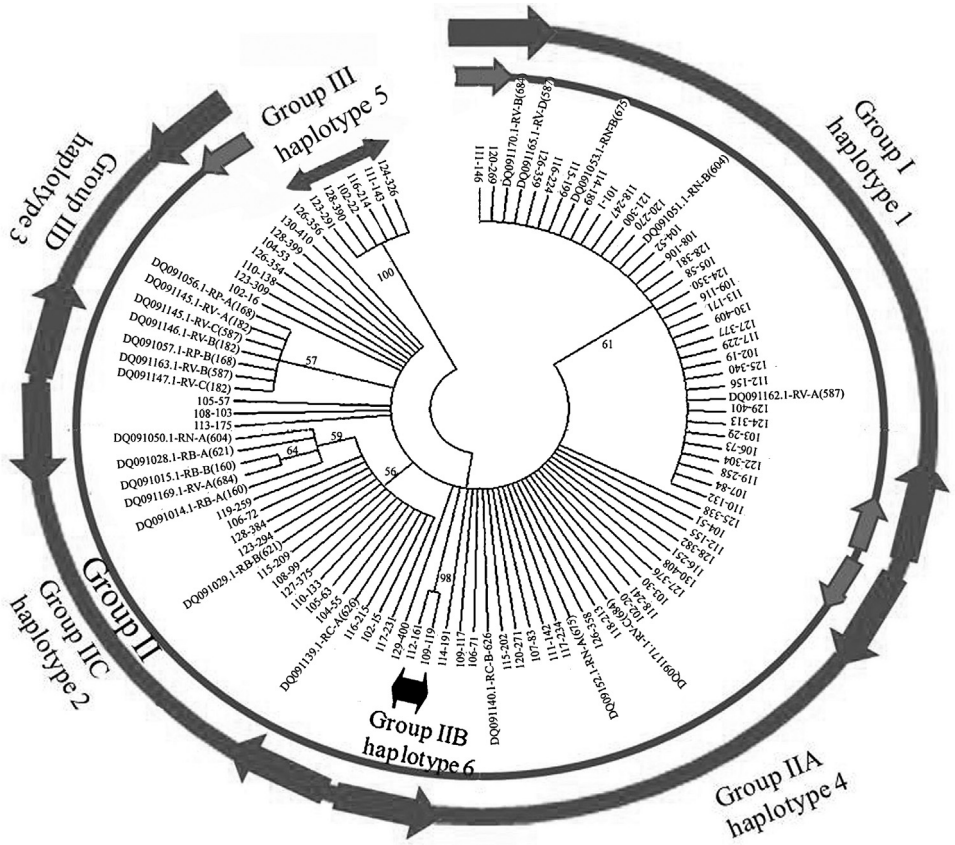

Fig. 7. Phylogenetic relationship of 84 *GAPDH* sequences derived from PEI rose ecotypes and 22 *GAPDH* reference sequences obtained from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances are in the units of the number of base substitutions per site. There were a total of 725 positions in the final dataset. RV, *Rosa virginiana*; RC, *Rosa carolina*; RB, *Rosa blanda*; RN, *Rosa nitida*; RP, *Rosa palustris*. PEI rose ecotypes are indicated by plot-clone numbers from which the sequences were obtained. Groups, subgroups and their corresponding haplotypes are indicated. Bootstrap value (%) indicating the power of clusters is shown next to the branches.

allele 528 B and the *R. carolina*-specific 626B (which is also identical to *R. virginiana* allele 684C (haplotype IVc)). Ambiguous haplotypes (haplotypes 3 and 5) were observed in 14 ecotypes (Fig. 6). In general, the *GAPDH* allele sampling analysis provided an indication of the possible number of alleles that may be present in each ecotype. But in most cases, the number of the sampled alleles for all the ecotypes ranged between 1 and 4 (Fig. 6).

Using a reduced dataset of *GAPDH* sequences, a phylogenetic analysis showed three major groups and six sub-groups with bootstrap indices ranging from 56 to 100% (Fig. 6). Group I comprise representative alleles from 29 of the 30 PEI rose ecotypes. These alleles were 100% identical to allele D (accessions 587) and allele B (accession 684) of *R. virginiana*, and to alleles A and B of *Rosa nitida* (accessions 604 and 675). Ecotype 122 (plot 123) was not represented in this group and no allele from *R. carolina* was also found in group I. Group II showed the largest variation with 4 sub-groups. Group II-A includes 21 sequences from the collection, allele B from *R. carolina* (accession 626), allele C from *R. virginiana* (accession 684), and allele A from *R. nitida* (accession 675). Group II-B consists of only 2 sequences from the collection and no allele from the reference species. Group II-C includes 13 sequences from PEI collection, allele A from *R. carolina* (accession 626), alleles A and B from *R. blanda* (accessions 621 and 160), and allele A from *R. virginiana* (accession 684). Group II-D consists of 11 PEI rose sequences, alleles A, B, C (accession 182), alleles B, C (accession 587) from *R. virginiana*, and alleles A, B (accession 168) from *R. pisocarpa*. Group III consists only of 6 PEI sequences with no alleles from the reference sequences (Fig. 7).

3.3.2. Haplotype affinity

To further define the affinity between haplotypes and to infer putative origins of the ecotypes present in the collection, the genetic affinity network analysis of haplotypes, statistical parsimony network and maximum likelihood tree analyses showed that the 30 rose ecotypes were grouped into six different haplotypes (Fig. 8), confirming the six groups shown in Figure 7.

3.3.3. SNP characterization and location

ClustalW multiple sequence alignment of 1 representative *GAPDH* sequence for each haplotype group revealed the key mutations that distinguish the six haplotype groups from each other. Haplotype 1 and haplotype 2 diverged from other haplotypes due to a T/C transition SNP (position 697) and an A/G transition SNP

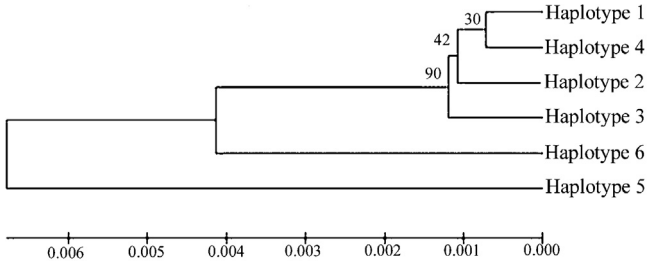

Fig. 8. Haplotype groups among PEI wild roses inferred by genetic affinity network analysis. Representative sequences from each haplotype groups were used for phylogenetic analysis using the UPGMA method. The bootstrap values (%) indicating the power of each cluster are shown next to the branches.

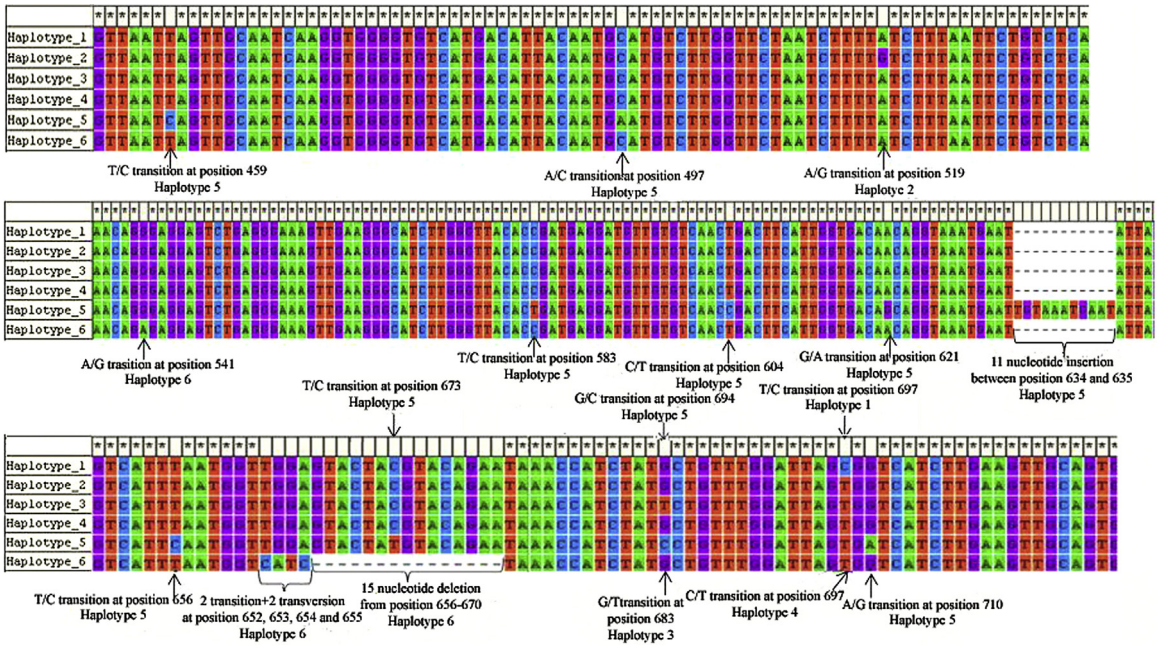

Fig. 9. Single nucleotide polymorphism (SNP), deletion and insertion events among 6 haplotype groups. ClustalW multiple sequence alignment obtained from representative alleles of the 6 haplotype groups depicting the single nucleotide polymorphisms, insertion and deletion responsible for separating the haplotypes is shown.

(position 519), respectively. Haplotype 3 separated as the result of a G/C/T tri-allelic SNP variation at position 683. Haplotype 4 is formed as the result of a C/T transition at position 697. Members of haplotype 5 carry an 11 nucleotide insertion (between position 634 and 635) and 9 SNPs (position 459, 497, 583, 604, 621, 656, 673, 694, and 710) whereas haplotype 6 is formed as a consequence of a 15 nucleotide deletion (from position 656 to 670), an A/G transition at position 541, and a tetra nucleotide (652–655) transition/transversion SNPs (Fig. 9).

3.3.4. SNP and insertion confirmation

To confirm the mutational events reported earlier (Fig. 9) MAMA PCR was performed. The 314 and 719 bp, 541 and 725 bp, 295 and 704 bp and, 363 and 725 bp amplicons depicted the SNPs that are specific to haplotype 1, 2, 3, and 7 respectively (Fig. 10A–D). The MAMA PCR confirmed the presence of the SNPs observed by allele sampling and further identified the targeted SNPs that were missed in some ecotypes by allele sampling, resulting in the observation of 1–4 alleles in each ecotype (Fig. 6). Haplotype 7 is a new haplotype

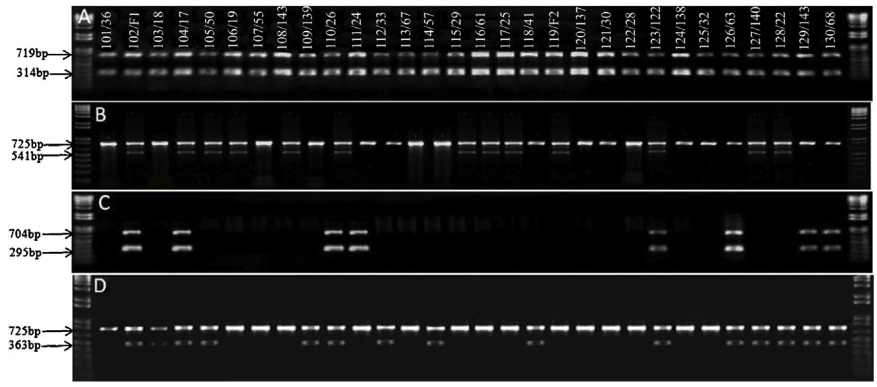

Fig. 10. Multiplex MAMA PCR amplifications depicting SNP-specific amplicons in PEI wild rose ecotypes. (A) *GAPDH* forward, nested *GAPDH* forward and MAMA 6R reverse primers' generated 719 bp and 314 bp amplicons that are specific for T/C transition at position 697, characterizing haplotype 1. (B) *GAPDH* forward and MAMA 5R reverse primer amplification of 541 bp amplicon specific for A/G transition at position 519 characterizing haplotype 2 and, the 725 bp wild type PCR band amplified by *GAPDH* forward and *GAPDH* reverse. (C) *GAPDH* forward, *GAPDH* nested forward and MAMA 9R reverse primers generated 704 bp and 295 bp amplicons specific for G/T transversion at position 683 characterizing haplotype 3. (D) *GAPDH* forward and *GAPDH* reverse primer generating the 725 bp wild type amplicon, and the MAMA 8F forward and *GAPDH* reverse primer amplifying the 363 bp band specific for C/T transition at position 383 characterizing haplotype 7. The ecotypes are labeled by plot-site numbers.

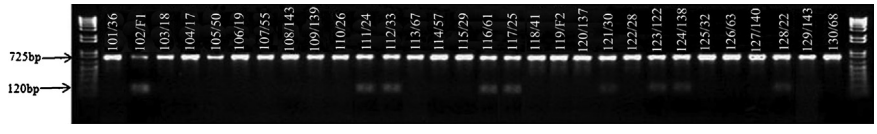

Fig. 11. Multiplex allele-specific PCR amplifications depicting the insertion-specific amplicons in PEI wild rose ecotypes. The *GAPDH* Forward, the insertion-specific Forward and the *GAPDH* Reverse primers were used in multiplex PCR. The 725 bp products are amplified by external *GAPDH* forward and *GAPDH* reverse primer in all the thirty ecotypes. The 120 bp is amplified by the insertion forward and *GAPDH* reverse primers in nine of the thirty ecotypes.

identified by MAMA PCR but not previously by any allele sampling analysis study. A 120 bp amplicon is characteristic for the insertion revealed by the insertion allele-specific forward and *GAPDH* reverse primer (Fig. 11). By considering all the 7 haplotypes (including the ambiguous haplotypes), 5 ecotypes (F1, 17, 33, 122, and 22) showed more than 4 alleles. The observed distribution of haplotypes among the 30 ecotypes based on allele sampling and MAMA PCR is shown in Fig. 6.

4. Discussion

The Biodiversity of the North American wild roses is now well known and described (Erlanson MacFarlane, 1966; Joly et al., 2006; Bruneau et al., 2007). However, no study has been conducted at the micro-evolutionary level. In this report, we conducted a micro-evolutionary study of 30 PEI native wild rose ecotypes selected and grown in a field setting by characterizing their genetic and metabolite diversity, phylogeny, and species status. Our data showed large genetic and metabolite diversity and indicated the presence of only *R. virginiana* and its natural hybrids with *R. carolina* within the population.

The gene diversity revealed by the 20 markers used in this study ranged between 0.00 and 0.97, with an average of 0.72. For polyploid organisms, gene diversity is an estimate of the probability that two randomly chosen genes from a population are different (Nei, 1987). The gene diversity observed among the 30 rose ecotypes was found to increase with the allele number, which is consistent with data reported by Huang et al. (2002). The PIC, a closely related measure of gene diversity (Botstein et al., 1980) varied from 0.00 to 0.97, with an average of 0.70. The PIC values reported here fit with that of previously reported haplotype diversity studies in apple (*Rosaceae*) (Jin et al., 2012).

The inference of genetic relationships between germplasm is a useful means to identify structure within a set of genotypes (Shin et al., 2006). Our study identified three main clusters, with cluster 3 showing the greater diversity and comprising 19 of the 30 ecotypes. Within each main cluster and sub-cluster, wild rose ecotypes also displayed large metabolite diversity, evidencing that the genetic diversity is also reflected by metabolite diversity. Since genetic diversity is the backbone for plant breeding programs (Manifesto et al., 2001; Shin et al., 2006), our data suggests that it is possible to select and propagate a given ecotype for its unique metabolite profile. Alternatively, the selected ecotype can be included in breeding programs for a targeted introgression of its unique metabolite composition into the desired lines.

For plant production in the field, rose plants are usually propagated by cuttings. However, many factors such as sample collection, labeling, manipulation during rooting and field transplantation can be source of errors and impacting on the genotype traceability. Indeed, collecting of rose stem samples from a site where many generations of rose genotypes grow together as a population in less than 1 m^2 can lead to contamination problems if extreme care is not observed. Using the SSR markers, our data showed a relatively low rate of mis-sampling during the establishment of the collection. Ninety-three percent (93.3%) of the plant materials grown in the field were clonally propagated in at least 2 replications, and 6.7% were not. Of the 93%, 60% of the plants were found to be clones in all 3 replications. The data reported here precisely identified groups of plants that are clones and those that are not. The results provide a guideline for the choice of plant materials from the collection and the use of genotype selections as elite lines and/or their use in breeding programs.

Although molecular and phylogenetic studies have established boundaries in *Rosa* sect. *Cinnamomeae* in eastern North American (Joly and Bruneau, 2007), the species status for the 30 PEI wild rose selections was unknown. SNP study of chloroplast DNA split the collection into 2 groups. Twenty ecotypes from the collection grouped with the wild diploid species *R. blanda* and *R. nitida* and tetraploid species *R. carolina* whilst 10 were grouped with the diploid species *R. nitida* and the tetraploid species *R. carolina* and *R. virginiana*. As chloroplast DNA is assumed to be maternally transmitted, ecotypes that grouped with a given known wild species are supposed to inherit their chloroplast haplotype from this species as maternal parent. *R. blanda* has been reported to be a parent to *R. carolina* together with *R. foliolosa, R. nitida* or *R. palustris* whereas *R. virginiana* is derived from the cross within *foliolosa-nitida-palustris* diploid group (Joly et al., 2006). The data reported here suggests that *R. nitida* and/or *R. blanda* are probably the maternal diploid progenitors for the majority of the PEI wild roses investigated and that *R. carolina* or *R. virginiana* seems to be the maternal progenitors for 1/3 of the collection. The later hypothesis may account for the hybrid phenotypes and confounding effects regularly reported by botanists (Bruneau et al., 2007). Nonetheless, this conclusion should be taken with caution considering the small size of diploid species used in this study. Alternatively, some diploid species may be present within the collection. To assess this assumption, a ploidy determination was conducted. The results showed that all ecotypes had a guard cell length in the range of tetraploid species. Thus, the collection may consist of *R. carolina*, *R. virginiana* or their natural hybrids. Measurement of the stomatal guard cell length has previously been used to determine the ploidy in different polyploid species (Omidbaigia et al., 2010; Keri and Sandra, 2007) including roses (Zlesak, 2009; Joly et al., 2006; Joly and Bruneau, 2007).

To define the specie status and derive firm conclusions on the origins for the polyploidy of ecotypes present in the collection, species-specific allele sampling, haplotype affinity and discrimination study using *GAPDH* SNP analyses showed that all 30 ecotypes carry the *R. nitida*-specific allele 604B and the *R. virginiana*-specific alleles 684B and 587D. In contrast, only 13 ecotypes showed the *R. blanda*-specific allele 528B and the *R. carolina*-specific allele 626B. Thus, the 17 ecotypes carrying only the *R. nitida*-specific allele 604B and the *R. virginiana*-specific alleles 684B and 587D were considered as true *R. virginiana*. Although *R. carolina*-specific allele 626B and the *R. virginiana* allele 684C are 100% identical, the 13 ecotypes carrying both the *R. virginiana*-specific alleles 684B and 587D on one hand, and the *R. blanda*-specific allele 528B and the *R. carolina*-specific allele 626B on the other hand were considered as hybrids. Indeed, single or low-copy nuclear gene such as glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) that had been subjected to independent evolutionary pressure with relatively rapid evolutionary rates is useful for inferring the phylogenetic relationship (Vaezi and Brouillet, 2009). The use of nuclear genes was particularly useful in this study because non-haploid organisms (except for clonal and apomictic taxa) receive one chromosome copy from each parent. Thus, nuclear genes retain information about the reticulate history of organisms, as opposed to maternally or paternally transmitted markers (Joly et al., 2006).

Phylogenetic analysis of the *GAPDH* sequences showed that the PEI ecotypes and the reference tetraploid species segregated into six groups. Haplotype analysis showed that ecotypes within group I display haplotype 1 which is identical to haplotype V-o in the network reported by Joly et al. (2006). Although allele sampling identified haplotype 1 in only 29 of the 30 ecotypes, subsequent multiplex MAMA PCR analysis using specific primers for this haplotype showed its presence in all the 30 ecotypes (Fig. 6). All sequences in group IIA carry a C/T transition mutation at position 697 and diverged from others. Members of this group have haplotype 4, which is identical to haplotype V-b in the haplotype network reported by Joly et al. (2006). *GAPDH* allele sampling identified haplotype 4 in 24 of the ecotypes. The SNP that characterizes haplotype 4 was shared between *R. virginiana*, *R. carolina*, *R. nitida* and 24 of the thirty ecotypes. This observation suggests the occurrence of

this SNP mutation prior to *R. virginiana* and *R. carolina* speciation and may originate from the diploid species *R. nitida*. Group IIB consists of 2 members that carry a 15 nucleotides deletion (position 656-670) and are grouped under haplotype 6. Haplotype analysis revealed that the 15 nucleotides deletion occurred in the genetic background of haplotype 1 (which is identical to haplotype V-o). Members of group IIB were identified as the sister to *R. nitida* allele 604B (haplotype V-o reported by Joly et al. (2006) and identical to haplotype 1 in this study) in the Maximum Likelihood tree and in the haplotype network. Group IIC members were found under haplotype 2 which is identical to Joly et al. (2006)'s haplotype IV-r, in turn identical to *R. blanda* allele 528 B. However, this haplotype IV-r is found in some *R. virginiana* from the Gaspé Peninsula (Eastern Quebec) where very little (or none) *R. carolina* are known (Joly, personal observations). Moreover, no distinct haplotype providing a clear evidence of a *R. blanda/woodsii* background was observed in the ecotypes investigated in this study. Nonetheless, since the occurrence of haplotype 2 is more likely to occur in *R. carolina* than in *R. virginiana* (Joly et al., 2006), one may reasonably assume that ecotypes carrying this haplotypes may have inherited it most likely from *R. carolina* or its progenitors rather than from *R. virginiana* or its progenitors. Members of group IID carry haplotype 3 which diverged from haplotype 4 due to a unique T/G transversion mutation at position 683 not yet observed in any previously published *Rosa GAPDH* sequences. Group IID merged with haplotype V-b in the haplotype network tree (Joly et al., 2006), but are not identical, and its position is somewhat ambiguous because of that unique mutation. Members of group III have haplotype 5 which is characterized by 9 SNPs and an 11 nucleotide insertion between positions 634 and 635. Haplotype 5 grouped with *gymnocarpa*/*pisocarpa* in the haplotype network tree. However, its position is ambiguous since the haplotype background was observed in both *R. carolina* and *R. virginiana.* Moreover, allele sampling identified this haplotype in 6 ecotypes whereas insertion allele specific PCR using insertion-specific primer confirmed the insertion in 3 additional ecotypes, for a total of 9 ecotypes. Taken together, group II showed the greatest diversity and included *R. virginiana*, *R. carolina* and their parental diploid species. It also comprised the majority of allelic variants encountered in PEI ecotypes classified in this study as natural hybrids involving *R. carolina* and *R. virginiana.* Thus, group II reflects the most the noise created by hybridization, incomplete lineage sorting, allelic segregation as previously reported (Joly et al., 2006).

MAMA PCR confirmed the presence of haplotype 2 in 13 ecotypes as also depicted by *GAPDH* allele sampling analysis (Figs. 6 and 10B). The insertion allele-specific PCR not only confirmed the 11 nucleotide insertion characterizing haplotype 5 and detected in 6 ecotypes after allele sampling but also identified this insertion in 3 additional ecotypes, for a total of 9 ecotypes. This highlights the usefulness and complementarity between MAMA PCR, insertion allele-specific PCR, and allele sampling. In general, the allele sampling and MAMA PCR confirmed the presence of up to 4 alleles in most of the ecotypes. Nonetheless, by considering all the 7 haplotypes (including the ambiguous haplotypes), 5 ecotypes (F1, 17, 33, 122, and 22) showed 5-6 alleles. Most of these ecotypes (except 33) were considered as hybrids in this study and all of them (except 17) carry one or both of the observed indels. The reason for this is unclear but using SNP genotyping, Deulvot et al. (2010) also observed a third allele in a diploid pea RIL population. The genotypes showing a third allele were reported to carry a 14 bp deletion as is the case in our study with 15 nucleotides deletion.

5. Conclusions

Using SSR and SNP markers, this study allowed us to better understand the genetic structure within a collection of 30 wild rose ecotypes collected from different regions of Prince Edward Island. The genetic diversity was reflected by metabolite diversity for important health promoting natural metabolites, which is particularly important for the growers, the health and nutriceutical industries. The collection comprises exclusively tetraploid species and includes 17 ecotypes tentatively characterized as authentic *R. virginiana* whereas 13 ecotypes were defined as natural hybrids derived from a spontaneous hybridization between *R. virginiana* and *R. carolina* The observed genetic and metabolite diversity could be of high interest for the selection of elite or breeding lines showing high yield and health promoting potentials.

Acknowledgements

This work was partly supported by an AAFC start-up fund to BF and by an AIF fund received from UPEI through the Innovative Canadian bioActives and Neutriceuticals (ICAN) project. The authors acknowledge David Main, Sylvia Wyand, Velma MacLean, and Lloyd Kerry (CLRC) for their technical assistance.

References

Akhunov, E., Nicolet, C., Dvorak, J., 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor. Appl. Genet. 119, 507–517.

Barry, R., Sanderson, K., Fillmore, S., 2008. Establishment of wild roses for commercial rose hip production in Atlantic Canada. Int. J. Fruit Sci. 8, 266–281.

Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

Bruneau, A., Starr, J.R., Joly, S., 2007. Phylogenetic relationships in the genus *Rosa*, new evidence from chloroplast DNA sequences and appraisal of current knowledge. Syst. Bot. 32, 366–378.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M., Rafalski, A.J., 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3, 19.

Crepin, F., 1889. Sketch of a new classification of roses. J. Royal Hortic. Soc. 11, 217–228.

Crepin, F., 1896. Rosae Americanae, I. Observations upon the genus *Rosa* in North America. Bot. Gazette 22, 1–34.

Delgado-Martinez, F.J., Amaya, I., Sanchez-Sevilla, J.F., Gomez-Jimenez, M.C., 2012. Microsatellite marker-based identification and genetic relationships of olive cultivars from the Extramedura region of Spain. Gen. Mol. Res. 11, 918–932.

Deulvot, C., Charrel, H., Marty, A., Jacquin, F., Donnadieu, C., Lejeune-Hénaut, I., Burstin, J., Aubert, G., 2010. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics 11, 468, http://dx.doi.org/10.1186/1471-2164-11-468.

Dugo, M.L., Satovic, Z., Millán, T., Cubero, J.I., Rubiales, D., Cabrera, A., Torres, A.M., 2005. Genetic mapping of QTLs controlling horticultural traits in diploid roses. Theor. Appl. Genet. 111, 511–520.

Erlanson MacFarlane, E.W., 1961. A self-polination mechanism and other Items in Rose species. America Rose Ann. 48, 188–193.

Erlanson MacFarlane, E.W., 1966. The old problem of species in *Rosa* with special reference to North America. America Rose Ann. 51, 150–160.

Fofana, B., Cloutier, S., Kirby, C.W., McCallum, J., Duguid, S., 2011. A well balanced omega-6/omega-3 ratio in developing flax bolls after heating and its implications for use as a fresh vegetable by humans. Food Res. Int. 44, 2459–2464.

Fofana, B., Humphreys, G., Rasul, G., Cloutier, S., Somers, D., 2008. Assessment of molecular diversity at QTL for preharvest sprouting resistance in wheat using microsatellite markers. Genome 51, 375–386.

Fofana, B., Humphreys, G., Rasul, G., Cloutier, S., Woods, S., Brûlé-Babel, A., Lukow, O.M., Somers, D., 2009. Mapping quantitative trait loci controlling pre-harvest sprouting resistance in a red × white seeded spring wheat cross. Euphytica 165, 509–521.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704, http://dx.doi.org/10.1080/10635150390235520.

Guo, X., Elston, R.C., 1999. Linkage information content of polymorphic genetic markers. Hum. Hered. 49, 112–118.

Gupta, P.K., Varshney, R.K., 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113, 163–185.

Hibrand-Saint Oyant, L., Crespel, L., Rajapakse, S., Zhang, L., Foucher, L., 2008. Genetic linkage maps of rose constructed with new microsatellite markers and locating QTL controlling flowering traits. Tree Genet. Genome 4, 11–23.

Huang, X.Q., Börner, A., Röder, M.S., Ganal, M.W., 2002. Assessing genetic diversity of wheat *Triticum aestivum* L. germplasm using microsatellite markers. Theor. Appl. Genet. 105, 699–707.

Hyten, D.L., Song, Q., Choi, I.Y., Yoon, M.S., Specht, J.E., Matukumalli, L.K., Nelson, R.L., Shoemaker, R.C., Young, N.D., Cregan, P.B., 2008. High-throughput genotyping

with the Golden Gate assay in the complex genome of soybean. Theor. Appl. Genet. 116, 945–952.
Jin, W., Zhang, Q., Liu, S., Wei, Q., Jin, W., Cheng, Z., Xue, X., Yang, T., 2012. Genetic diversity of 41 apple rootstocks based on simple sequence repeat markers. J. Am. Soc. Hortic. Sci. 137, 51–56.
Jing, R., Johnson, R., Seres, A., Kiss, G., Ambrose, M.J., Knox, M.R., Ellis, T.H., Flavell, A.J., 2007. Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics 177, 2263–2275.
Joly, S., Starr, J.R., Lewis, W.H., Bruneau, A., 2006. Polyploid and hybrid evolution in roses east of the Rocky mountains. Am. J. Bot. 933, 412–425.
Joly, S., Bruneau, A., 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Syst. Biol. 55, 623–636.
Joly, S., Bruneau, A., 2007. Delimiting species boundaries in *Rosa* Sect. Cinnamomeae *Rosaceae* in eastern North America. Syst. Bot. 32, 819–836.
Keri, D.J., Sandra, M.R., 2007. Analysis of ploidy level and its effects on guard cell length, pollen diameter and fertility in *Hydrangea macrophylla*. Hortscience 423, 483–488.
Kimura, T., Nishitani, C., Iketani, H., Ban, Y., Yamamoto, T., 2006. Development of microsatellite markers in rose. Mol. Ecol. Notes 6, 810–812.
Li, B., Kadura, I., Fu, D., Watson, D.E., 2004. Genotyping with TaqMAMA. Genomics 83, 311–320.
Li, L., Paulo, M.J., Strahwald, J., Lubeck, J., Hofferbert, H.R., Tacke, E., Junghans, H., Wunder, J., Draffehn, A., van Eeuwijk, F., Gebhardt, C., 2008. Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. Theor. Appl. Genet. 116, 1167–1181.
Liu, K., Muse, S.V., 2005. PowerMarker, integrated analysis environment for genetic marker data. Bioinformatics 21, 2128–2129.
Macphail, V.J., Kevan, P.G., 2009. Review of the breeding systems of wild roses *Rosa* spp. Floricult. Orn. Biotech. 3, 1–13.
Manifesto, M.M., Schlatter, A.R., Hopp, H.E., Suárez, E.Y., Dubcovsky, J., 2001. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci. 42, 682–690.
Mercure, M., Bruneau, A., 2008. Hybridization between the escaped *Rosa rugosa* (*Rosaceae*) and native *R. blanda* in eastern North America. Am. J. Bot. 95, 597–607.
Nei, M., 1987. Genetic distance and molecular phylogeny. In: Ryman, N., Utter, F. (Eds.), Population Genetics and Fishery Management. Washington University Press, Seatle, pp. 193–223.
Nei, M., 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19, 153–170.
Ninomiya, K., Matsuda, H., Kubo, M., Morikawa, T., Nishida, N., Yoshikawa, M., 2007. Potent anti-obese principle from *Rosa canina*, structural requirements and mode of action of *trans*-tiliroside. Bioorg. Med. Chem. Lett. 17, 3059–3064.
Omidbaigia, R., Mirzaeea, M., Hassanib, M.E., Moghadamc, M.S., 2010. Induction and identification of polyploidy in basil *Ocimum basilicum* L. medicinal plant by colchicine treatment. Intern. J. Plant Prod. 42, 87–98.
Page, R.D.M., 1996. TREEVIEW, an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Rehder, A., 1940. Manual of Cultivated Trees and Shrubs Hardy in North America. New York.
Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P., Ganal, M.W., 1998. A microsatellite map of wheat. Genetics 149, 2007–2023.
Rusanov, K., Kovacheva, N., Vosman, B., Zhang, L., Rajapakse, S., Atanassov, A., Atanassov, I., 2005. Microsatellite analysis of *Rosa damascena* Mill. accessions reveals genetic similarity between genotypes used for rose oil production and old Damask rose varieties. Theor. Appl. Genet. 111, 805–809.
Sanderson, K., Fillmore, S., 2010. Evaluation of native rose selections for rose hip production in Prince Edward Island. Intern. J. Fruit Sci. 10, 379–389.
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. Hortic. Sci. 131, 66–73.
Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18, 233–234.
Shin, J.H., Kwon, S.J., Lee, J.K., Min, H.K., Kim, N.S., 2006. Genetic diversity of maize kernel starch-synthesis genes with SNAPs. Genome 49, 1287–1296.
Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4, molecular evolutionary genetics analysis MEGA software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Vaezi, J., Brouillet, L., 2009. Phylogenetic relationships among diploid species of *Symphyotrichum (Asteraceae*, Astereae) based on two nuclear markers, ITS and *GAPDH*. Mol. Phylogenet. Evol. 51, 540–553.
Watson, S., 1885. Contributions to American botany, I. A history and revision of the roses of North America. PNAS 20, 324–352.
Werlemark, G., Nybom, H., 2010. Dogroses, botany, horticulture, genetics, and breeding. Hortic. Rev. 36, 199–255.
Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D., Cregan, P.B., 2003. Single-nucleotide polymorphisms in soybean. Genetics 163, 1123–1134.
Zlesak, D.C., 2009. Pollen diameter and Guard cell length as predictors of ploidy in diverse rose cultivars, species and breeding line. Floricult. Orn. Biotech. 3 (Special issue 1), 53–70.