Dear Author

Here are the proofs of your article.

- You can submit your corrections **online,** via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For **fax** submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
- Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL:

`http://dx.doi.org/10.1007/s11105-015-0966-7`

If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information, go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us, if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

Please note: Images will appear in color online but will be printed in black and white.

1	Article Title	**Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers**	
2	Article Sub- Title		
3	Article Copyright - Year	**Springer Science+Business Media New York 2015 (This will be the copyright line in the final PDF)**	
4	Journal Name	Plant Molecular Biology Reporter	
5	Corresponding Author	Family Name	**Charles-Eric**
6		Particle	
7		Given Name	**Durel**
8		Suffix	
9		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
10		Division	
11		Address	42 rue Georges Morel, Beaucouzé cedex 49071
12		e-mail	charles-eric.durel@angers.inra.fr
13	Author	Family Name	**Ludivine**
14		Particle	
15		Given Name	**Lassois**
16		Suffix	
17		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
18		Division	
19		Address	42 rue Georges Morel, Beaucouzé cedex 49071
20		Organization	University of Liege, Gembloux Agro-Bio Tech, Forest Resources Management
21		Division	BIOSystem Engineering (BIOSE) Department
22		Address	Passage des Déportés 2, Gembloux 5030
23		e-mail	
24	Author	Family Name	**Caroline**
25		Particle	
26		Given Name	**Denancé**
27		Suffix	

28		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
29		Division	
30		Address	42 rue Georges Morel, Beaucouzé cedex 49071
31		e-mail	
32		Family Name	**Elisa**
33		Particle	
34		Given Name	**Ravon**
35		Suffix	
36	Author	Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
37		Division	
38		Address	42 rue Georges Morel, Beaucouzé cedex 49071
39		e-mail	
40		Family Name	**Arnaud**
41		Particle	
42		Given Name	**Guyader**
43		Suffix	
44	Author	Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
45		Division	
46		Address	42 rue Georges Morel, Beaucouzé cedex 49071
47		e-mail	
48		Family Name	**Rémi**
49		Particle	
50		Given Name	**Guisnel**
51		Suffix	
52	Author	Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
53		Division	
54		Address	42 rue Georges Morel, Beaucouzé cedex 49071
55		e-mail	
56		Family Name	**Laurence**
57		Particle	
58		Given Name	**Hibrand-Saint-Oyant**
59	Author	Suffix	
60		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
61		Division	

62		Address	42 rue Georges Morel, Beaucouzé cedex 49071
63		e-mail	
64	Author	Family Name	**Charles**
65		Particle	
66		Given Name	**Poncet**
67		Suffix	
68		Organization	Plateforme Gentyane, INRA UMR1095 Genetics, Diversity and Ecophysiology of Cereals
69		Division	
70		Address	63100, Clermont-Ferrand
71		e-mail	
72	Author	Family Name	**Pauline**
73		Particle	
74		Given Name	**Lasserre-Zuber**
75		Suffix	
76		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
77		Division	
78		Address	42 rue Georges Morel, Beaucouzé cedex 49071
79		Organization	INRA, UMR 1095 Genetics, Diversity and Ecophysiology of Cereals
80		Division	
81		Address	Clermont-Ferrand 63100
82		Organization	UBP, UMR 1095 Genetics, Diversity and Ecophysiology of Cereals
83		Division	
84		Address	Clermont-Ferrand 63100
85		e-mail	
86	Author	Family Name	**Laurence**
87		Particle	
88		Given Name	**Feugey**
89		Suffix	
90		Organization	Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV
91		Division	
92		Address	42 rue Georges Morel, Beaucouzé cedex 49071
93		e-mail	
94	Schedule	Received	
95		Revised	

96		Accepted
97	Abstract	In-depth characterization of apple genetic resources is a prerequisite for genetic improvement and for germplasm management. In this study, we fingerprinted a very large French collection of 2163 accessions with 24 SSR markers in order to evaluate its genetic diversity, population structure, and genetic relationships, to link these features with cultivar selection date or usage (old or modern, dessert, or cider cultivars), and to construct core collections. Most markers were highly discriminating and powerful for varietal identification, with a probability of identity $P_{(ID)}$ over the 21 retained SSR loci close to 10^{-28}. Pairwise comparisons revealed 34 % redundancy and 18.5 % putative triploids. The results showed that the germplasm is highly diverse with an expected heterozygosity H_e of 0.82 and observed heterozygosity H_o of 0.83. A Bayesian model-based clustering approach revealed a weak but significant structure in three subgroups (F_{ST} = 0.014–0.048) corresponding, albeit approximately, to the three subpopulations defined beforehand (Old Dessert, Old Cider, and Modern Cultivars). Parentage analyses established already known and yet unknown relationships, notably between old cultivars, with the frequent occurrence of cultivars such as "King of Pippin" and "Calville Rouge d'Hiver" as founders. Finally, core collections based on allelic diversity were constructed. A large dessert core collection of 278 cultivars contained 90 % of the total dessert allelic diversity, whereas a dessert subcore collection of 48 cultivars contained 71 % of diversity. For cider apples, a 48-cultivar core collection contained 83 % of the total cider allelic diversity.
98	Keywords separated by ' - '	*Malus* × *domestica* - SSR - Diversity - Genetic structure - Parentage analysis - Core collection
99	Foot note information	The online version of this article (doi:10.1007/s11105-015-0966-7) contains supplementary material, which is available to authorized users.

Electronic supplementary material

ESM 1
List of the 2163 accessions considered in the present study with their accession code (AcceNumber), name (AcceName), the name of the partner who furnished the leaf sample, their accession usage (mainly Dessert or Cider apple) and cultivar selection date (Old or Modern = before or after 1950), their ploidy level determined according to the occurrences of three alleles per locus (see text), their duplicate code according to the SSR profile (see text), their subgroup assignment inferred by the STRUCTURE analysis with the highest membership probability (qI), their group prior to the STRUCTURE analysis (subpopulation: 1: Modern cultivars, 2: Old Cider cultivars, 3: Old Dessert cultivars), their involvement in statistical analyses, and their selection in the dessert and cider core collections (see text). (XLS 476 kb)

Plant Mol Biol Rep
DOI 10.1007/s11105-015-0966-7

ORIGINAL PAPER

Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers

Lassois Ludivine[1,5] · Denancé Caroline[1] · Ravon Elisa[1] · Guyader Arnaud[1] · Guisnel Rémi[1] · Hibrand-Saint-Oyant Laurence[1] · Poncet Charles[2] · Lasserre-Zuber Pauline[1,3,4] · Feugey Laurence[1] · Durel Charles-Eric[1]

Abstract In-depth characterization of apple genetic resources is a prerequisite for genetic improvement and for germplasm management. In this study, we fingerprinted a very large French collection of 2163 accessions with 24 SSR markers in order to evaluate its genetic diversity, population structure, and genetic relationships, to link these features with cultivar selection date or usage (old or modern, dessert, or cider cultivars), and to construct core collections. Most markers were highly discriminating and powerful for varietal identification, with a probability of identity $P_{(ID)}$ over the 21 retained SSR loci close to 10^{-28}. Pairwise comparisons revealed 34 % redundancy and 18.5 % putative triploids. The results showed that the germplasm is highly diverse with an expected heterozygosity H_e of 0.82 and observed heterozygosity H_o of 0.83. A Bayesian model-based clustering approach revealed a weak but significant structure in three subgroups (F_{ST} = 0.014–0.048) corresponding, albeit approximately, to the three subpopulations defined beforehand (Old Dessert, Old Cider, and Modern Cultivars). Parentage analyses established already known and yet unknown relationships, notably between old cultivars, with the frequent occurrence of cultivars such as "King of Pippin" and "Calville Rouge d'Hiver" as founders. Finally, core collections based on allelic diversity were constructed. A large dessert core collection of 278 cultivars contained 90 % of the total dessert allelic diversity, whereas a dessert subcore collection of 48 cultivars contained 71 % of diversity. For cider apples, a 48-cultivar core collection contained 83 % of the total cider allelic diversity.

Keywords *Malus* × *domestica* · SSR · Diversity · Genetic structure · Parentage analysis · Core collection

Introduction

Apples (*Malus* × *domestica* Borkh.) constitute the main fruit crop in temperate regions (Velasco et al. 2010). Very early on, apple cultivars were selected and multiplied using grafting. Until the beginning of the twentieth century, the vast majority of these cultivars were grown from "chance seedlings" with unknown parentage, and most of today's well-known cultivars are still those chance seedlings discovered during the nineteenth century ("Jonathan," "Cox's Orange Pippin," "Granny Smith," "Red Delicious," "Golden Delicious," etc.) (Way et al. 1990). It was only in the second half of the twentieth century that cultivars from controlled hybridization such as "Idared," "Elstar," "Gala," "Jonagold," and "Cripps Pink" became fixtures on the apple market (Way et al. 1990). Most of these newly bred cultivars were obtained from a reduced number of founders and from some mutants and thus exhibit a high level of relationship (Noiton

Electronic supplementary material The online version of this article (doi:10.1007/s11105-015-0966-7) contains supplementary material, which is available to authorized users.

✉ Durel Charles-Eric
charles-eric.durel@angers.inra.fr

1 Institut de Recherche en Horticulture et Semences – UMR 1345, INRA, SFR 4207 QUASAV, 42 rue Georges Morel, 49071 Beaucouzé cedex, France

2 Plateforme Gentyane, INRA UMR1095 Genetics, Diversity and Ecophysiology of Cereals, 63100, Clermont-Ferrand, France

3 INRA, UMR 1095 Genetics, Diversity and Ecophysiology of Cereals, 63100 Clermont-Ferrand, France

4 UBP, UMR 1095 Genetics, Diversity and Ecophysiology of Cereals, 63100 Clermont-Ferrand, France

5 BIOSystem Engineering (BIOSE) Department, University of Liege, Gembloux Agro-Bio Tech, Forest Resources Management, Passage des Déportés 2, 5030 Gembloux, Belgium

and Alspach 1996). As a consequence, despite the high genetic diversity available, apple production worldwide is currently based on a limited number of cultivars, leading to a dramatic loss of diversity all over the world. For example, 70 % of the 2012 European Union production was based on only ten varieties, and "Golden Delicious" alone represented 35 % of the French production in 2011 (data from World Apple and Pear Association, 2013). In view of this situation, the preservation of apple genetic resources is essential to avoid the irretrievable loss of a high degree of diversity. Genetic material must be included in a germplasm bank for its conservation and further agronomical evaluation. The invaluable work of conserving apple genetic resources in France is carried out by many amateur associations and various governmental, regional, and local authorities. All of the cultivars conserved on French territory constitute a valuable biodiversity resource with an important link to inheritance.

Studying the genetic diversity of germplasm resources is not only significant for the protection of species, but also necessary for the development and utilization of germplasm resources for crop improvement. There is a growing interest at this time to understand the genetic bases of complex traits and to discover new germplasm characteristics in order to better take advantage of them for efficient breeding. Indeed, the tremendous apple allelic diversity should be used to face existing and future biotic and abiotic constraints with respect to sustainable production in the context of global change (Zeigler 2013). Furthermore, because the phenotypic description of the agronomic traits and the full genotyping of a large apple collection are costly and time consuming, working on a reduced germplasm collection is considered as a helpful mean to better evaluate and use plant germplasm (Upadhyaya et al. 2010). The core collection concept, i.e., a representative sample of the whole collection minimizing repetitiveness and maximizing genetic diversity, applied in many crop genebanks, was first proposed by Frankel and Brown (1984). Its use is recommended by the Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (FAO, 1996) as a way to improve the use of plant genetic resources.

Preliminary steps of genetic characterization and core collection constitution will focus on estimating genetic diversity and determining the genetic relationships among the germplasm accessions. Molecular markers have become an efficient way to address these issues by creating a fingerprint of each individual tree. Among the molecular markers proposed over the last 20 years, microsatellites or simple sequence repeat (SSR) markers are highly polymorphic, neutral, abundant, reliably reproducible, codominant, and quite cheap advantages that make them relevant for plant genetic analyses. SSRs have been successfully used to identify cultivars and germplasm accessions in many fruits such as grape (*Vitis* sp.) (Cipriani et al. 2010), sweet cherry (Mariette et al. 2010), citrus (Gulsen and Roose 2001), peach (Aranzana et al. 2010), and kiwifruit (*Actinidia* Lindl.) (Zhen et al. 2004). These markers have proved advantageous for diversity studies on apple (Garkava-Gustavsson et al. 2008; Gasi et al. 2010; Gross et al. 2014; Hokanson et al. 2001; Liang et al. 2015; Moriya et al. 2011; Pereira-Lorenzo et al. 2008; Song et al. 2006; Urrestarazu et al. 2012; van Treuren et al. 2010), and several hundred SSR markers have been developed and genetically mapped across the 17 linkage groups of the apple genome (Gianfranceschi et al. 1998; Liebhard et al. 2002; Silfverberg-Dilworth et al. 2006).

In this study, the analysis and quantification of the genetic diversity within 2163 accessions from the French apple germplasm were performed and allowed to check for possible redundancies and triploids. The population substructure of the entire collection was evaluated, and yet-unknown relationships have been inferred. Three core collections maximizing the genetic diversity both for dessert and cider apples have been established. The fine molecular characterization achieved will help to support conservation, management, and utilization of this large French germplasm which has never been previously molecularly assessed.

Materials and Methods

Plant Material

The germplasm included 2163 apple accessions: Old Dessert, Old Cider, and Modern Cultivars (containing only six modern cider cultivars, all others being dessert cultivars)—referred to below as "OD," "OC," and "MC." Cultivars bred after 1950 were considered as modern cultivars (MC). Among those 2163 apple accessions, 1049 originated from the INRA collection. The others 1114 accessions were gathered from several associations of amateurs, botanical gardens, and regional or national repositories covering the French territory (Fig. 1). These additional accessions were also well-diversified, based on pomological knowledge, with minimum overlap with the INRA collection. Some accessions with the same name from different collections were also analyzed to confirm (or not) the cultivars identity. Eight control samples corresponding to eight reference cultivars were included in this set ("Red Delicious," "Fiesta," "Worcester Pearmain," "Prima," "Michelin," "Malling 9" (rootstock), "*Malus floribunda #821*," and "*Malus robusta* 5"), as recommended by the European Collaborative Programme for Crop Genetic Resources (ECPGR) *Malus/Pyrus* working group (http://www.ecpgr.cgiar.org/working-groups/maluspyrus/), to allow both internal harmonization of data and further comparisons of results with other studies. The complete list and status of the evaluated accessions is available in the Online Resource ESM 1.

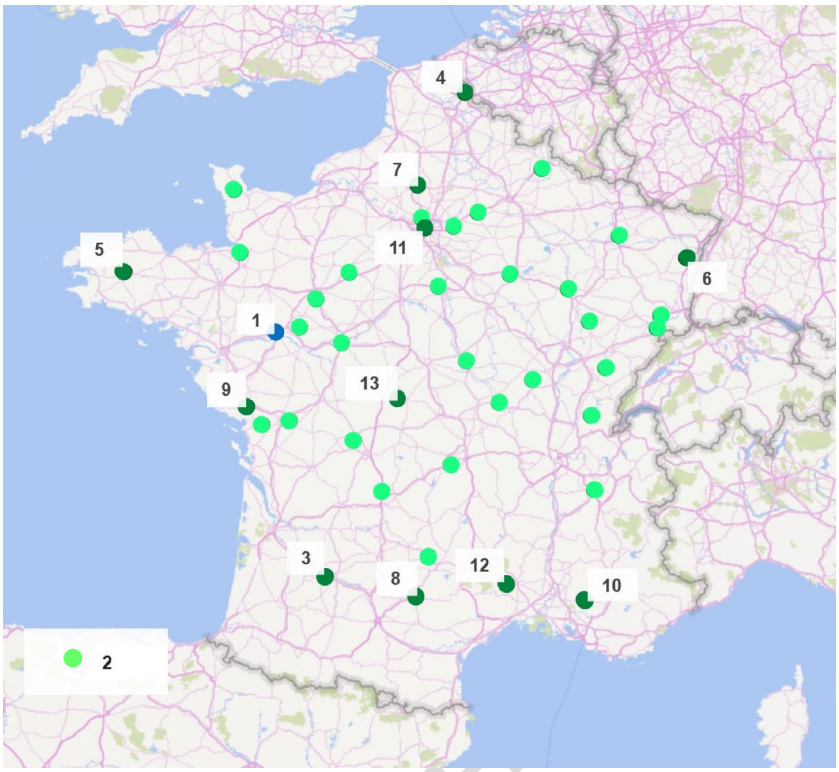

Fig. 1 Geographic distribution of the collections analyzed in the present study with indication of the sample sizes between brackets: *1*, INRA (1049); *2*, Les Croqueurs de Pommes (335); *3*, Conservatoire Végétal Régional d'Aquitaine (196); *4*, Centre Régional de Ressources Génétiques du Nord-Pas-de-Calais (149); *5*, Les Mordus de la Pomme (134); *6*, Confédération des Producteurs de Fruits d'Alsace (63); *7*, I z'on creuqué eun' pomm' (49); *8*, Conservatoire des Espèces Fruitières et de Vignes Anciennes (49); *9*, Verger Conservatoire de Pétré (34); *10*, Parc Naturel Régional du Lubéron (29); *11*, Jardin du Luxembourg (27); *12*, Fruits Oubliés Réseau (27); *13*, Société Pomologique du Berry (22). The INRA collection (1) is indicated in blue. The amateurs association "Les Croqueurs de Pommes" (2) is largely distributed over the French territory as visible with the 29 light-green dots

DNA Extraction and Quantification

Young leaf tissues (approximately 50 mg/sample) were collected and stored at −80 °C before to be reduced to a fine powder in liquid nitrogen by shaking using a Qiagen Tissue Lyser device. Total genomic DNA was isolated using a cetyltrimethylammonium bromide (CTAB) protocol according to Aldrich and Cullis (1993) with minor modifications. DNA samples extracted were quantified using a BMG Fluostar™ Omega fluorescence plate reader after Hoechst labeling and then adjusted to 5 ng/μL.

SSR Fingerprinting

A set of 24 SSR primers developed by different groups (Gianfranceschi et al. 1998; Guilford et al. 1997; Hokanson et al. 1998; Liebhard et al. 2002; Silfverberg-Dilworth et al. 2006; Vinatzer et al. 2004) was used to genotype the 2163 accessions (Table 1). These SSRs are distributed over the 17 apple linkage groups, and 15 of them are included in a former list of 17 SSR recommended by the ECPGR *Malus/Pyrus* working group (Urrestarazu et al. 2012). Forward primers were labeled with four different fluorescent dyes (6-FAM, VIC, NED, or PET) in order to be combined into six different multiplexed (MP_1–MP_6) reactions (Table 1). Polymerase chain reactions (PCR) for the six multiplex PCRs were performed in a final volume of 11 μL using 10 ng of DNA template, 0.18 μM of each primer except for some markers as described in Table 1, and 1× PCR Master mix of QIAGEN kit multiplex PCR (Qiagen, Hilden, Germany).

PCR cycling conditions were as follows: preincubation for 15 min at 95 °C, followed by 34 cycles, each consisting of 30 s denaturing at 94 °C, 90 s at annealing temperature, and 60 s elongation at 72 °C, the last cycle ending with a final 15-min extension at 72 °C. The following annealing temperatures were applied: 55 °C for MP_2 and MP_6, and 57 °C for MP_1, MP_3, MP_4, and MP_5. Furthermore, the MP_5 was amplified using an amplification program with, for the three first cycles, an annealing temperature reduced by 1 °C per cycle from 60 to 57 °C. SSR amplification products were analyzed with an ABI3730 XL sequencing system (Applied Biosystems, Foster City, CA, USA).

Fragment analysis and sizing were carried out using GeneMapper4.0 software (Applied Biosystems, Foster

Table 1 Characteristics of the 24 SSRs (microsatellites) used in this study with indication of the corresponding multiplex and dye

Locus	Linkage group number	Multiplex	Dye	Size range (bp)	Forward primer sequence 5′→3′	Reverse primer sequence 5′→3′	Primer concentration[f]
Hi02c07[a]	1	MP_3	VIC	104–150	agagctacggggatccaaat	gtttaagcatcccgattgaaagg	[0.09 μM]
CH-Vf1[e]	1	MP_6	VIC	133–185	atcaccaccagcagcaaag	gtttcttcatacaaatcaaagcacaaccc	[0.09 μM]
CH02c06[b]	2	MP_1	NED	204–282	tgacgaaatccactactaatgca	gtttgattgcgcgctttttaacat	[0.36 μM]
GD12[c]	3	MP_4	PET	150–200	ttgaggtgtttctcccattgga	gtttcttctaacgaagccgccatttcttt	[0.36 μM]
CH03e03[b]	3	MP_6	6-FAM	181–222	gcacattctgccttatcttgg	gtttaaaacccacaaatagcgcc	[0.18 μM]
NZ05g08[d]	4	MP_2	VIC	117–163	cggccatcgattatcttactctt	gtttcttggatcaatgcactgaaataaacg	[0.18 μM]
CH05d02[b]	4	MP_5	PET	201–235	aaactccctcacctcacatcac	gtttaatagtccaatggtgtggatgg	[0.18 μM]
CH05f06[b]	5	MP_2	PET	171–197	ttagatccggtcactctccact	gttttggaggaagacgaagaagaaag	[0.18 μM]
Hi04a08[a]	5	MP_6	NED	209–253	ttgaaggagtttccggtttg	gtttcactctgtgctggattatgc	[0.18 μM]
CH03d07[b]	6	MP_4	6-FAM	166–232	caaatcaatgcaaaactgtca	gtttggcttctggccatgatttta	[0.27 μM]
CH03d12[b]	6	MP_5	6-FAM	98–163	gcccagaagcaataagtaaacc	gtttattgctccatgcataaaggg	[0.18 μM]
CH04e05[b]	7	MP_2	6-FAM	178–232	aggctaacagaaatgtggtttg	gtttatggctcctattgccatcat	[0.18 μM]
CH01h10[b]	8	MP_1	VIC	92–146	tgcaaagataggtagatatatgcca	gtttaggagggattgtttgtgcac	[0.18 μM]
CH01f03b[b]	9	MP_1	6-FAM	141–200	gagaagcaaatgcaaaaccc	gtttctccccggctcctattctac	[0.18 μM]
CH02c11[b]	10	MP_3	NED	209–265	tgaaggcaatcactctgtgc	gtttttccgagaatcctcttcgac	[0.27 μM]
CH02d08[b]	11	MP_2	NED	209–262	tccaaaatggcgtacctctc	gtttgcagacactcactcactatctctc	[0.18 μM]
CH04g07[b]	11	MP_6	PET	153–217	ccctaacctcaatccccaat	gtttatgaggcaggtgaagaagga	[0.18 μM]
CH01f02[b]	12	MP_3	6-FAM	163–225	accacattagagcagttgagg	gtttctggtttgttttcctccagc	[0.18 μM]
GD147[c]	13	MP_3	PET	122–161	tcccgccatttctctgc	gtttaaaccgctgctgctgaac	[0.18 μM]
CH04c07[b]	14	MP_4	VIC	99–144	ggccttccatgtctcagaag	gtttcctcatgccctccactaaca	[0.18 μM]
CH01g05[b]	14	MP_5	NED	135–194	catcagtctcttgcactggaaa	gtttgacagagtaagctagggctaggg	[0.18 μM]
CH02c09[b]	15	MP_4	NED	227–261	ttatgtaccaactttgctaacctc	gtttagaagcagcagaggaggatg	[0.18 μM]
CH01h01[b]	17	MP_1	PET	95–161	gaaagacttgcagtgggagc	gtttggagtgggtttgagaaggtt	[0.18 μM]
Hi07h02[a]	17	MP_5	VIC	223–277	caaattggcaactgggtctg	gtttaggtggaggtgaagggatg	[0.18 μM]

The four first multiplexes (MP_1, MP_2, MP_3, and MP_4) correspond to SSR markers recommended by the ECPGR *Malus/Pyrus* working group (in bold), except CH01f03b which here replaces CH01h02 on LG9

[a] Silfverberg-Dilworth et al. (2006)

[b] Liebhard et al. (2002)

[c] Hokanson et al. (1998)

[d] Guilford et al. (1997)

[e] Vinatzer et al. (2004)

[f] Primer concentration within a given MP has been adjusted to get more homogeneous SSR marker amplification intensities

City, CA, USA), and the individual fragments were assigned as alleles. Chromatograms were independently read by two operators. The eight reference cultivars were used as control profiles.

Descriptive Statistics of Genetic Diversity

The genetic uniqueness of each accession was determined using pairwise comparison of locus profile results. Accessions were considered as duplicates when they presented identical SSR fingerprints. An allelic difference was tolerated for a maximum of two SSR loci assuming that some genotyping errors and/or spontaneous SSR mutations could occur. Redundant accession profiles were further removed from the dataset to avoid bias in genetic analyses. An accession was declared as a putative triploid when at least three of the 24 SSR loci were characterized with three distinct alleles.

Basic statistics were computed with the CERVUS software package, version 3.0 (Kalinowski et al. 2007; Marshall et al. 1998) (http://www.fieldgenetics.com), on the unique diploid genotypes. For each SSR locus, the number of alleles per locus (A_o) and the effective number of alleles per locus ($A_e = 1/\sum p_i^2$, where p_i is the frequency of the ith allele) were identified. The allelic frequencies made it possible to observe the allele distribution and to identify rare alleles (frequency <2 %). The observed (H_o) and expected (H_e) heterozygosity, the significance of a deviation from the Hardy-Weinberg equilibrium including a Bonferroni correction, and the estimated frequency of null alleles were also estimated using CERVUS software. The polymorphic information content (PIC) (Botstein et al. 1980) of each marker was determined using the following equation:

$$PIC_i = 1 - \sum_{j=1}^{n} {p_{ij}}^2$$

where p_{ij} is the frequency of the jth allele for marker i and the summation extends over n alleles. A fixation index F was calculated as follows: $F = 1 - H_o/H_e$ (Prat et al. 2006).

The probability of identity $P_{(ID)}$ was calculated as follows (Waits et al. 2001):

$$P_{(ID)} = \sum p_i^4 + \sum\sum \left(2 p_i p_j\right)^2$$

where p_i and p_j are the frequencies of the ith and jth alleles and $i \neq j$. A $P_{(ID)}$ among sibs $P_{(ID)\mathrm{sib}}$ was also calculated (Evett and Weir 1998). Finally, the ability of each marker to discriminate two random cultivars was estimated with the "power of discrimination" (PD) (Kloosterman et al. 1993).

The genetic diversity of subgroups or core collections (see below) was compared to the genetic diversity of the initial population (dessert and cider) by considering the heterozygosity parameters (H_o and H_e) and the allelic richness calculated using a rarefaction framework with the program ADZE 1.0 (Gross et al. 2014; Szpiech et al. 2008).

Analysis of Genetic Structure

Factorial correspondence analysis (FCA) was used to represent the genetic diversity of the unique diploid genotypes. GENETIX software, version 4.05.2 (Belkhir et al. 2004), was used to illustrate FCA results and to estimate the F_{ST} genetic differentiation indexes between groups. F_{ST} were computed either for the three a priori defined subpopulations ("OD," "OC," and "MC") or for subgroups identified using STRUCTURE software (see below). The significance of F_{ST} was assessed by 10,000 resamplings of the genotypic data.

The genetic diversity structure of the unique genotypes was also investigated with an alternative approach using the Bayesian model-based clustering algorithm of STRUCTURE software, version 2.3.3 (Pritchard et al. 2000) (http://www.pritch.bsd.uchicago.edu). To analyze diploids and triploids together, we used the recessive allele approach (Pritchard et al. 2000; Urrestarazu et al. 2012). We used the LOCPRIOR model since we considered that, for our dataset, the available prior information concerning "usage" (dessert or cider) and "cultivar selection date" (bred before or after 1950) of cultivars could be favorable for assisting the clustering. We also evaluated the potential genetic structure with LOCPRIOR model according the geographic origins of the accessions, when accurately known. France was divided into six regions: the north, the northwest, the northeast, southwest, southeast, and center. The mean r value calculated by STRUCTURE in the LOCPRIOR model parameterizes the amount of information carried by the prior information. STRUCTURE was run with different values of the number of clusters (K) varying from 1 to 10 under the admixture model for which the allelic frequencies were correlated. To verify the consistency of the results, we performed ten independent runs per K value with 500,000 Markov chain Monte Carlo iterations after a burn-in of 200,000 steps. K_{opt} was inferred from the formula established by Evanno et al. (2005). For K_{opt}, individuals were assigned to a subgroup according to the probability of their membership in this subgroup. The graphical results were obtained by STRUCTURE HARVESTER (Earl and vonHoldt 2012) (http://taylor0.biology.ucla.edu/structureHarvester/). CLUMPP software, v.1.1.2 (Jakobsson and Rosenberg 2007), was used to compute average individual assignment probabilities (qI) over replicated runs showing a similar mode. The graphical display of

the STRUCTURE results was generated using DISTRUCT software, version 1.1 (Rosenberg 2004) (http://www.stanford.edu/group/rosenberglab/distruct.html). Genotypes were assigned to the subgroup for which they had the highest membership coefficient, considering strong affinity when the assignment probability (qI) was ≥0.8 (Liang et al. 2015; Urrestarazu et al. 2012).

Parentage Analysis

Parentage analysis was conducted on unique diploid genotypes with CERVUS software (Kalinowski et al. 2007; Marshall et al. 1998). The parameters of the simulated genotypes were the following: "offspring" 100, 000; "candidate parents" 2100; "prop. sampled" 0.3; "prop. loci typed 0.8; and "prop. loci mistyped" 0.01. In order to reveal only robust parentages, we limited the study to the inferences of "two-parent offspring" relationships and did not consider inferences of "one-parent offspring" relationships where the lacking parent offers more flexibility but more fuzzy assignments as well. Two criterions were considered to establish strict parentage relationships: a confidence level of the LOD score and the Delta value both higher than 95 %. Finally, an additional constraint was added to strengthen the results by limiting the maximum number of tolerated loci mismatches to only two in an inferred two-parent offspring trio (Salvi et al. 2014).

Core Collection Constitution

Three core collections were constructed with DARwin software version 5.0.158 (Perrier et al. 2003; Perrier and Jacquemoud-Collet 2006) (http://darwin.cirad.fr/darwin) with the "max length sub tree" option for identifying the most unstructured neighbor-joining tree with maximum maintenance of allelic diversity (Perrier and Jacquemoud-Collet 2006). The core collections were primarily designed for association genetics studies recently engaged in our laboratory. Three criteria were taken into account in the accessions selection process: (i) putative triploids were excluded; (ii) for practical propagation reasons, the accessions were selected only among the genotypes available within the INRA collection; (iii) the size of each core collection was a priori fixed for technical reasons and to allow further linkage disequilibrium and genome-wide association studies. A core collection containing 278 diploid dessert apple accessions was first constructed. A nested subcore collection composed of 48 diploid dessert apple accessions was also selected. Similarly, a small core collection of 48 diploid cider apple accessions was constructed.

Results

Accession Identification

Five of the 2163 accessions collected did not show any amplification and were discarded from the analysis. Among the 2158 remaining accessions, pairwise comparison of all locus profiles revealed 373 groups of replicates (Online Resource ESM 1), leading to the removal of 737 redundant accessions for further analyses (34 % of redundancy). The number of accessions in each of these identical SSR profile groups varied from two to 18 accessions. Among the 737 redundant accessions, 607 accessions presented a strict identical profile to their membership group, whereas 103 presented an allelic difference in one locus and 14 accessions in two loci. Moreover seven accessions showed a difference in three loci and two accessions in four loci. However, these three and four loci differences were observed for the same SSR markers in the same multiplexed PCR. Since a contamination problem could be suspected, they were finally discarded as redundant accessions. Following these observations, the apple germplasm dataset was reduced to 1421 unique genotypes. Among these accessions, 263 showed a putative triploid profile, representing 18.5 % of the accessions. Interestingly, "OC" cultivars showed 18.2 % of putative triploids and "OD" 20.1 %, whereas "MC" consisted in only 5.1 % of putative triploids.

A preliminary FCA performed with GENETIX4.05.2 software revealed that several accessions were very far away from the global dot distribution and were considered as "extreme" genotypes (results not shown). These concerned: three Tunisian-related accessions: "Ajmi" (X2440), "Aziza" (X2941), and "Chahla" (X2940); three wild or ornamental apple genotypes: "*Malus floribunda* #821," "*Malus robusta* 5," and "Maypole" (X6027); a presumably Iranian accession: "Précoce de Karaj" (X0897); and a presumably Turkish accession: "Douce Rayotte" (X9253). These eight accessions as well as two rootstocks ("Malling 9" and "MM106" = "Malling-Merton 106") or their redundant accessions (corresponding to grafting errors) were eliminated from the collection for further analysis. Finally, to avoid too many missing data, which could be problematic in various analyses, only accessions that amplified at least 17 of the 24 SSR loci were conserved for genetic analysis (Online Resource ESM 1). The final dataset used for further analyses was then constituted of 1319 genotypes distributed as follows: 1084 diploids (188 "OC," 737 "OD," and 159 "MC") and 235 putative triploids (42 "OC," 185 "OD," and 8 "MC").

Genetic Diversity of the Collection

A preliminary analysis with CERVUS on the 1084 diploid genotypes highlighted that all the SSR loci amplified in this study were polymorphic. However, as presented in Table 2, three out of

Table 2 Genetic diversity parameters assessed for 24 SSR loci in the subset of 1084 unique diploid apple accessions of the French apple germplasm

Locus	N°$_{obs}$	Missing data (%)	A_o	A_e	Rare alleles (%)[b]	H_o	H_e	F = 1 − (H_o/H_e)	PD	PIC	HW	$F_{(null)}$	$P_{(ID)}$ unrelated	$P_{(ID)sib}$
CH-Vf1	1022	5.6	19	3.5	13 (68.4)	0.76	0.72	−0.050	0.88	0.68	***	−0.03	0.121	0.421
CH01h10	1076	0.6	17	3.6	9 (52.9)	0.70	0.72	0.025	0.89	0.69	NS	0.01	0.111	0.420
Hi02c07	1076	0.6	16	3.6	10 (62.5)	0.74	0.72	−0.017	0.89	0.69	NS	−0.01	0.106	0.415
GD12	1074	0.8	15	3.6	7 (46.7)	0.74	0.72	−0.019	0.90	0.70	NS	−0.01	0.099	0.414
CH04e05	1073	0.9	20	3.9	12 (60.0)	0.75	0.74	−0.016	0.91	0.71	NS	−0.01	0.094	0.403
CH01f03b	1077	0.6	11	4.2	4 (36.4)	0.78	0.76	−0.025	0.91	0.73	NS	−0.01	0.091	0.393
Hi04a08	1066	1.6	11	4.4	4 (36.4)	0.77	0.77	0.009	0.92	0.75	NS	0.01	0.079	0.384
GD147	1072	1.0	16	5.0	8 (50.0)	0.80	0.80	0.000	0.94	0.78	*	0.00	0.065	0.366
CH03d12	1043	3.7	30	5.7	22 (73.3)	0.83	0.83	−0.001	0.95	0.81	NS	0.00	0.046	0.349
CH02d08	1073	0.9	20	6.2	11 (55.0)	0.84	0.84	−0.002	0.96	0.82	NS	0.00	0.043	0.341
CH03d07	1044	3.6	25	6.3	16 (64.0)	0.86	0.84	−0.018	0.96	0.82	NS	−0.01	0.043	0.340
CH02c09	1064	1.8	14	6.6	6 (42.9)	0.84	0.85	0.012	0.96	0.83	NS	0.01	0.041	0.336
CH01g05	1041	3.9	20	6.8	11 (55.0)	0.87	0.85	−0.014	0.96	0.84	NS	−0.01	0.038	0.333
CH04c07	1058	2.3	20	7.2	11 (55.0)	0.89	0.86	−0.036	0.97	0.85	NS	−0.02	0.033	0.328
CH05f06	1075	0.7	13	7.6	5 (38.5)	0.87	0.87	0.001	0.97	0.86	NS	0.00	0.031	0.324
CH01f02	1075	0.7	24	8.0	15 (62.5)	0.89	0.88	−0.013	0.97	0.86	NS	−0.01	0.028	0.319
CH01h01	1071	1.1	22	8.4	13 (59.1)	0.88	0.88	0.001	0.97	0.87	NS	0.00	0.026	0.316
Hi07h02	1028	5.1	24	8.6	13 (54.2)	0.89	0.88	−0.008	0.98	0.87	NS	0.00	0.024	0.314
CH02c06	1052	2.9	28	8.7	17 (60.7)	0.87	0.89	0.018	0.98	0.88	NS	0.01	0.023	0.313
CH04g07	1063	1.8	26	9.4	14 (53.8)	0.90	0.90	−0.008	0.98	0.89	NS	0.00	0.020	0.308
CH02c11	1071	1.1	19	10.0	7 (36.8)	0.89	0.90	0.008	0.98	0.89	*	0.00	0.019	0.305
Mean[a]	1061.6	2	19.5	6.2	10.86	0.83	0.82	−0.01	0.94	0.80		0.00	$1.3\ 10^{-28}$	$3\ 10^{-10}$
Total			410	131.2	228 (55.6)									
SSR with estimated frequency of null allele >0.1														
NZ05g08	1064	1.8	16	3.5	7 (43.7)	0.46	0.72	0.355	0.90	0.70	ND	0.22	0.10	0.42
CH05d02	956	11.7	17	8.7	7 (41.1)	0.51	0.89	0.428	0.98	0.87	ND	0.27	0.02	0.31
CH03e03	926	14.5	13	5.5	7 (53.8)	0.42	0.82	0.485	0.94	0.79	ND	0.32	0.06	0.36

Loci carrying null alleles at estimated frequencies >0.1 are listed at the bottom

*Significant at the 5 % level, **significant at the 1 % level, ***significant at the 0.1 % level

$N^o{}_{obs}$ number of observed accessions (Ntotal = 1084), A_o number of alleles, A_e effective number of alleles, H_o observed heterozygosity, H_e expected heterozygosity, *F* fixation index, *PD* power of discrimination, *PIC* polymorphic information content, *HW* exact test of departure from Hardy-Weinberg equilibrium, *NS* not significant, *ND* not done, $F_{(null)}$ estimated frequency of null alleles, $P_{(ID)}$ probability of identity

[a] In the column "$P_{(ID)}$ unrelated" and "$P_{(ID)\ sib}$," the mean is substituted with cumulative $P_{(ID)}$, which is the product of the $P_{(ID)}$ of individual loci

[b] Rare alleles correspond to frequency <0.02

24 SSR loci showed an estimated frequency of null allele $F_{null} > 0.1$ and a fixation index (F) value far from 0. It was then decided to remove them for further analyses in order to avoid a bias. The concerned SSR loci were NZ05g08, CH05d02, and CH03e03. Furthermore, two of them exhibited a high level of missing data (11.7 % for CH05d02 and 14.5 % for CH03e03). Among the remaining 21 loci, 18 were in Hardy-Weinberg equilibrium, whereas three were not (CH-Vf1, GD147, and CH02c11). The results of basic statistics on the 1084 unique diploid genotypes are presented in Table 2. SSR markers were classified according their PD, which ranged from 0.88 to 0.98, with a mean of 0.94. Four markers exhibited very high power of discrimination (CH02c06, CH02c11, CH04g07, and Hi07h02), whereas the three markers CH01h10, CH-Vf1, and Hi02c07 were comparatively less powerful for genotype discrimination. The number of missing data ranged between 0.6 % for CH01h10, Hi02c07, and CH01f03b, and 5.6 % for locus CH-Vf1, with a mean of 2.0 % per locus. A total of 410 alleles was revealed by the set of 21 SSR markers, leading to a mean number of alleles per locus of 19.5 (ranging from 11 for CH01f03b and Hi04a08, to 30 for CH03d12), whereas the mean effective number of alleles/locus was 6.2 (range 3.5–10.0). A total of 228 rare alleles (frequency <2 %) were identified, representing 55.6 % of the global allelic diversity (410 alleles); 41 alleles (10 % of the

total allelic diversity) out of this set were observed in only one accession ("unique alleles"). The mean value for expected heterozygosity (H_e) was 0.82 (range 0.72–0.90), which was very close to the value of 0.83 (range 0.70–0.90) for observed heterozygosity (H_o). The mean PIC value was 0.80 (range 0.68–0.89). The probability of identity $P_{(ID)}$ calculated for individual loci ranged from 0.019 for the most discriminating locus CH02c11 to 0.121 for the least discriminating locus CH-Vf1. The cumulative $P_{(ID)}$ over all 21 loci was 1.3×10^{-28} for unrelated genotypes and 3×10^{-10} for full sibs.

Structure Identification

First, a FCA was conducted with the 21 SSR data on the 1084 diploid genotypes by differentiating six geographic origins of the accessions (north, northwest, northeast, southwest, southeast, and center). No genetic trend could be highlighted (results not shown). The use of STRUCTURE software with LOCPRIOR model according the same geographic origins confirmed the absence of genetic differentiation at this geographic scale (results not shown). Second, a FCA was conducted by differentiating three subpopulations beforehand: "OD," "OC," and "MC." "OD" dots covered almost the entire graph, whereas "MC" and "OC" were concentrated into two distinct groups on the FCA graph (Fig. 2), suggesting a weak structure. Inertia values were 1.69 and 1.46 % for coordinate axes 1 and 2 of the graph. Pairwise F_{ST} comparisons confirmed a weak structure between these three subpopulations. The strongest F_{ST} values were observed between "MC" on one side and "OC" (0.048; $p_{value} = 0$) or "OD" on the other side (0.031; $p_{value} = 0$). A lower F_{ST} value of 0.014 ($p_{value} = 0$) was observed between "OC" and "OD."

Finally, the genetic structure of the 1319 unique diploid and triploid apple genotypes was also analyzed with the model-based clustering algorithm implemented in STRUCTURE software. The structure signal obtained by the STRUCTURE default mode was very weak (results not shown), and the LOCPRIOR model was successfully used with a mean r value of 0.96, indicating that the prior information is informative. For all K_{opt}, memberships were consistent between all runs. The peak of ΔK for $K = 3$ corresponded to the presence of three main subgroups (Fig. 3). Divergence between the corresponding subgroups given by STRUCTURE results was evaluated by pairwise F_{ST} comparisons. A low structure was observed between subgroups 1 and 2

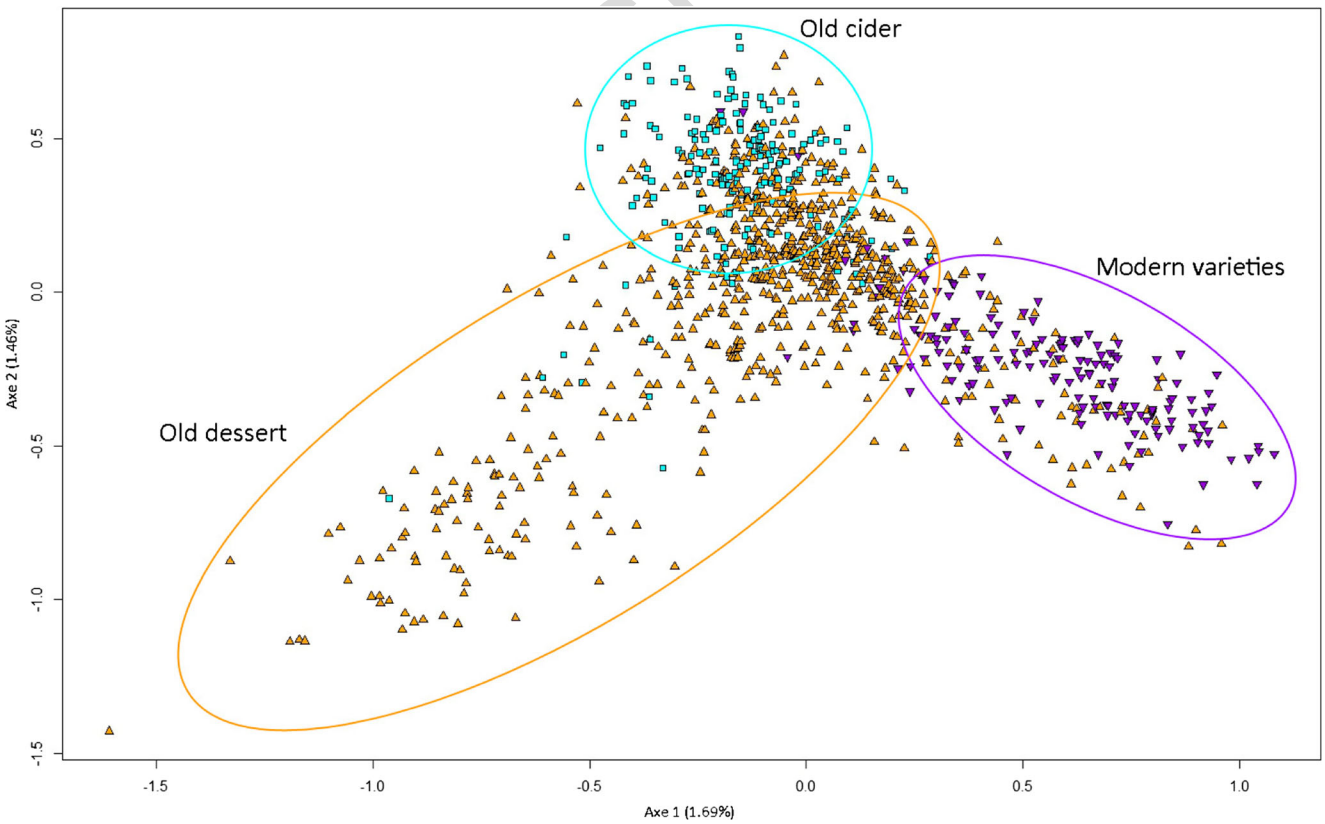

Fig. 2 Factorial correspondence analysis (FCA) of the 1084 unique diploid genotypes with GENETIX4.05.2 software for 21 SSRs. Assignment of genotypes to the Old Dessert, Old Cider, and Modern Cultivars subpopulations are depicted with *orange triangles*, *blue squares*, and *purple triangles*, respectively. Inertia values are 1.69 and 1.46 % for coordinate axes 1 and 2. *Circles* approximately group together the three a priori subpopulations with the respective colors

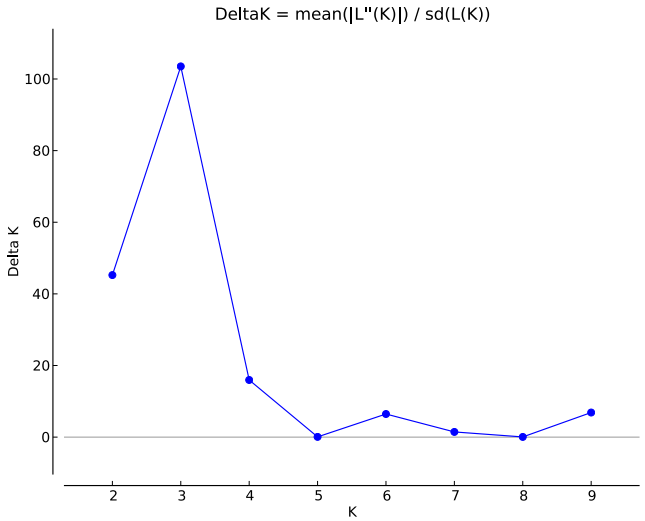

Fig. 3 Determination of K_{opt} according to the Evanno et al. (2005) method. The rate of change of the posterior probability of the data given the number of sungroups is plotted against K. The first peak (K = 3) corresponds to the optimum number of subgroups. Computation was performed for the 1319 unique apple accessions genotyped with 21 SSR markers

(F_{ST} = 0.040, p_{value} = 0) and between subgroups 1 and 3 (F_{ST} = 0.026, p_{value} = 0). F_{ST} was slightly higher between subgroups 2 and 3 (F_{ST} = 0.060, p_{value} = 0), leading up to the conclusion that a moderate genetic differentiation existed. Substructures were searched for in each of these three STRUCTURE subgroups but the subsequent results did not provide additional relevant conclusions (results not shown).

When comparing the three STRUCTURE subgroups with the three a priori "OD," "OC," and "MC" subpopulations, it appears that the three subgroups highlighted by STRUCTURE corresponded, albeit approximately, to the three a priori subpopulations (Fig. 4). More precisely, the assignment proportions of each a priori subpopulation ("MC," "OC," "OD") to these three STRUCTURE genetic subgroups showed that 98 % of the "Modern Cultivars" were assigned to subgroup 3, whereas 98 % of the "Old Cider" cultivars were assigned to subgroup 1 (Table 3). In contrast, the "Old Dessert" cultivars were more largely distributed over the three subgroups, with a majority (66 %) assigned to subgroup 1, which is the largest one (Table 3). It is also worth mentioning that subgroup 2 contains fewer accessions than the other two, and that 97 % of these accessions are "OD." The same trend was observed when considering only accessions with a strong assignment probability (qI ≥ 0.8; data not shown). Consistently, most of the "MC" and "OC" cultivars were clearly assigned to STRUCTURE subgroups with, respectively, 92 and 93 % of the cultivars assigned with a probability ≥0.8, whereas only 42 % of "OD" cultivars showed a strong assignment (results not shown).

The genetic diversity of subgroups 2 and 3 was lower than for subgroup 1 based on H_e (Table 4). When considering only diploid genotypes with a high membership probability (qI ≥ 0.8), H_e was only 87 % for subgroup 2 in comparison with subgroup 1. Many private alleles could be observed in each of the three subgroups especially for genotypes with high qI (Table 4). By scaling down to subgroup 2 size, allelic richness was similar in subgroups 1 and 2 but smaller (~83 %) in subgroup 3 whatever all or high membership genotypes were considered (Table 4).

Parentage Analysis

Parent-offspring relationships in the 1084 unique diploid genotypes were explored by CERVUS software. A total of 46 putative trios (offspring and two inferred parents) were identified with high (95 %) confidence level consisting of 18 Modern and 28 Old cultivars (Table 5). The two parents of 14 Modern cultivars for which full parentage was already known were correctly inferred (e.g., "Alkmene" = "Doctor Oldenburg" × "Cox's Orange Pippin"). For two additional Modern cultivars ("Judor" and "Cidor") bred in the 1970s at INRA as juice and cider cultivars, the common known female parent ("Douce Moen") was correctly inferred and the initially unknown male parents were newly postulated as "Rouge de Trèves" and "Doux Joseph" (respectively) known as "OC" cultivars that were planted in the same orchard where the open pollinated progeny of "Douce Moen" was collected for breeding purposes. The parentage of the last two Modern cultivars ("Nabella" and "Deltana") was corrected since one of the already known parents was correctly inferred but the other was not (Table 5, see "Discussion"). The two parents of the remaining 28 Old cultivars were generally not known and thus newly inferred (Table 5). Accession "FRA1002," erroneously referred to as "Herrgottsapfel," was shown to exhibit the same SSR profile as "Astillisch" in another study (data not shown). Its paternity assignment fitted with the expected cross product from "Signe Tillish" × "Astracan rouge."

Core Collections

Based on H_e, the set of 278 INRA accessions selected to generate the "Old Dessert" core collection (Online Resource ESM 1) exhibited a genetic diversity similar to the set of 737 unique diploid 'OD' genotypes (Table 6). The mean number of alleles of the 'OD' core collection (16.4) was kept at 90 % of that of the overall dessert collection (18.1). Moreover, the allele frequencies of the "OD" core collection were very highly correlated to those observed for the overall dessert collection (R^2 = 0.99). For the nested sub-core collection of 48 dessert accessions, the mean number of alleles was lower (71 %) than in

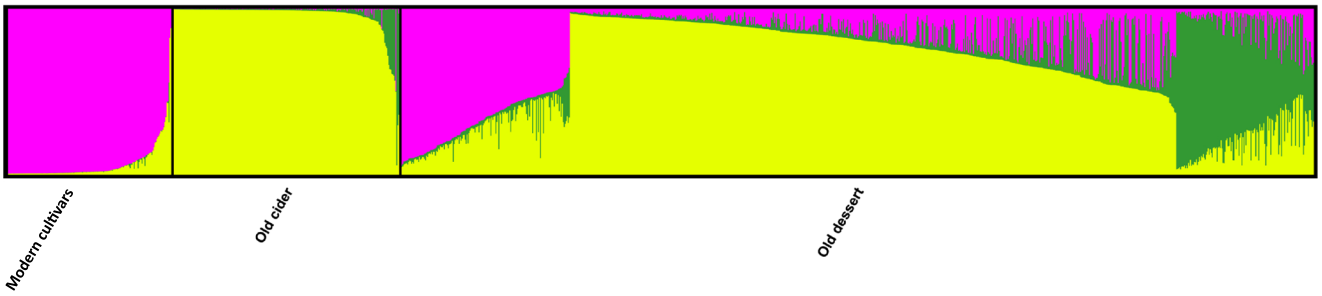

Fig. 4 Proportions of ancestry of 1319 unique apple genotypes for K = 3 ancestral gene pools ("subgroups") inferred with Structure2.3.3 software (Pritchard et al. 2000). Each genotype is represented by a vertical bar partitioned into K = 3 segments representing the amount of ancestry of its genome in three subgroups. The a priori classification concerning their usage (Dessert/Cider) and cultivar selection date (Old/Modern) is indicated. The three subgroups are depicted using the following color codes: *yellow* = subgroup 1; *green* = subgroup 2; *pink* = subgroup 3

the overall collection but H_e and the allelic richness remained higher (Table 6). In the "Old Cider" core collection of 48 INRA accessions (Online Resource ESM 1), the mean number of alleles was reduced to 83 % when compared to the overall set of 188 unique diploid "OC" genotypes, but H_e and the allelic richness remained higher (Table 6) with a very high correlation between allelic frequencies of the core and initial collections ($R^2 = 0.94$). In all cases, allelic richness was much higher in core collections than in an average sample of the overall collection of the same size (Table 6).

Discussion

Gene Pool Representativeness and Geographical Structure

Numerous diversity studies have been performed on apple germplasm (Garkava-Gustavsson et al. 2008; Gasi et al. 2010; Gross et al. 2014; Hokanson et al. 2001; Liang et al. 2015; Moriya et al. 2011; Pereira-Lorenzo et al. 2008; Song et al. 2006; Urrestarazu et al. 2012; van Treuren et al. 2010) but as far as we know, the present study is the largest one to be performed so far at a national level with such a large number of SSR markers. Apple genetic resources are conserved by many very active structures in France (Fig. 1) and the coordination of this conservation network is entrusted to INRA. Thanks to this network that covers all of the major repositories and pomological societies, the representativeness of the germplasm studied was excellent at the national level (Fig. 1). A large majority of the accessions could be considered to be of French origin, although a significant part of the germplasm studied was composed of foreign cultivars. The proportion of foreign cultivars was tentatively estimated at 15–25 % according to the different sources in the literature or to website resources and information derived from redundant accessions (data not shown), but it was extremely difficult to ensure such a value since many inconsistencies were observed over information sources and over duplicate groups affiliations. Moreover, geographic origin was not always documented and a typical French name of a so-called local cultivar could be attributed to an accession that was finally shown to be a duplicate of a well-known foreign cultivar. A similar situation was described both by Urrestarazu et al. (2012) and Liang et al. (2015) for Spanish and Italian accessions that finally turned out to be redundant with the well-known American cultivar "Rome Beauty." Finally, most of the robustly assigned foreign cultivars analyzed in the present study came from European countries and Russia.

In the same way, due to the lack of documentation about geographic origin and historical widespread exchanges of apple cultivars over geographic regions, it was extremely difficult to know the real French region of origin of many accessions studied. It could explain that no clear relationship between the geographical

Table 3 Proportions of membership of each pre-defined subpopulation (Modern Cultivars, Old Cider, Old Dessert) in each of the three subgroups as inferred by Structure2.3.3

A priori population	Subgroups inferred by Structure2.3.3			Number of individuals
	1	2	3	
Modern Cultivars	0.018 (3)	0.000 (0)	0.982 (164)	167
Old Cider	0.983 (226)	0.017 (4)	0.000 (0)	230
Old Dessert	0.664 (612)	0.151 (139)	0.185 (171)	922
Number of individuals	841	143	335	1319

Numbers in brackets represent the number of individuals in each group

t4.1 **Table 4** Descriptive information for the three subgroups of diploid genotypes identified by Structure analysis

		Subgroup	N	H_o	H_e	Number of alleles				Allelic richness
						Total	Private[a]	Unique[b]	Mean no. allele	
t4.4	All diploid genotypes	1	655	0.83	0.81	384	79	46	18.3	12.8
t4.5		2	124	0.83	0.77	268	8	51	12.8	(12.8)
t4.6		3	305	0.80	0.78	287	10	39	13.7	10.6
t4.7	Genotypes with qI > 0.8	1	396	0.83	0.81	323	90	52	15.4	8.5
t4.8		2	34	0.81	0.71	180	15	48	8.6	(8.6)
t4.9		3	188	0.79	0.76	224	25	23	10.7	7.3

Information is detailed either for all diploid genotypes or for genotypes with a membership probability >0.8. Summary statistics include the sample size (N), observed (H_o) and expected (H_e) heterozygosity, total, private, unique, average number of alleles per locus (mean no. allele). Allelic richness is scaled to the smallest subgroup (subgroup 2, $N = 124$ for all diploid genotypes and $N = 34$ for genotypes with qI > 0.8). For this subgroup 2, the average number of alleles is copied as the reference allelic richness (between brackets)

[a] Alleles detected only in that subgroup

[b] Alleles detected only in one accession

origin and the genetic structure was found in the studied French germplasm, as also reported by Cornille et al. (2012). For example, an accession collected in the north may actually have originated from another region because of unknown historical exchanges. Furthermore, the redundancy rate between regions highlighted in this study also reflects exchanges of plant material over geographic regions.

Choice of SSR Markers

Reliable and polymorphic SSR markers are essential to study the genetic diversity and structure of such a germplasm. As indicated, several apple diversity studies using SSR markers have already been published. Unfortunately, many different markers have been used with only limited occurrences of large overlapping between studies involving a large number of markers. This situation makes it difficult to accurately compare diversity parameters over studies except for some parameters such as H_e, which summarizes the fundamental genetic variation of a germplasm. Conversely, a more in-depth meta-analysis could be performed to classify SSR markers according to their PD or their PIC in order to mine the most informative ones over several germplasm collections. In the present study, we used a set of 16 SSR markers previously recommended by the ECPGR *Malus/Pyrus* working group (Table 1), plus eight additional markers, in order to reach a higher genome coverage for the genetic diversity and structure analyses. This ECPGR set is highly recommended for all new SSR diversity analyses since it will allow more accurate comparison of diversity and redundancy over germplasms worldwide. Three SSR markers (NZ05g08, CH05d02, and CH03e03) exhibited high frequencies of null alleles ($F_{null} > 0.1$) most probably overestimating the corresponding fixation indexes (F). They were discarded from further analyses as they may introduce a bias in both Hardy-Weinberg test computation and parentage analysis (Dakin and Avise 2004). It is noteworthy that NZ05g08 and CH05d02 both belong to the upper part of linkage group (LG) 4. The NZ05g08 marker was already identified as generating null alleles by Urrestarazu et al. (2012) and Pina et al. (2014) and should thus be replaced by another one in the ECPGR set.

All of the other 21 SSR loci analyzed displayed a high level of polymorphism with 11 to 30 alleles per locus and a mean of 19.5 alleles per locus. This result reveals more alleles per locus than previously observed on apple by Gharghani et al. (2009), Liang et al. (2015), and Urrestarazu et al. (2012) which revealed 17, 16.8, and 16.7 alleles per locus, respectively. This difference could be attributed to the higher number of accessions observed in our study compared to others. The mean of effective alleles/locus (*Ae*) was 6.2. The difference between the allele number and the effective allele number can be explained by the high number of loci showing rare alleles (55.6 % of the total allelic diversity) with low frequency (<2 %). Furthermore, 10 % of the total allelic diversity was present in only one accession. Such a level of rare or unique alleles indicates a substantial genetic diversity not used at this time for breeding. The level of rare alleles obtained in our study is comparable to that obtained by Urrestarazu et al. (2012) (63 %) in the Spanish apple germplasm.

Redundancy and Triploidy Level in the French Apple Germplasm

The extremely low probability ($P_{(ID)} = 1.3 \times 10^{-28}$) of matching by chance any two genotypes at all 21 loci gave us great confidence in the ability of our SSR marker set to accurately detect duplicated accessions. About one third (34 %) of

Table 5 Parentages of 46 modern and old apple cultivars inferred according to the maximum likelihood approach developed in CERVUS software (Kalinowsky et al. 2006)

Offspring AcceNumber	Offspring AcceName	Modern/ Old	First candidate parent AcceNumber	First candidate parent AcceName	Second candidate parent AcceNumber	Second candidate parent AcceName	Trio loci compared	Trio loci mismatch
X2437	Alkmene	M	FRA0943	Docteur Oldenburg	X1954	Cox's Orange Pippin	21	0
X3513	Cidor	M	X5101	Doux Joseph	X5124	Douce Moën	20	0
X8194	Deljuga	M	X4443	Delgollune	X4712	Gala	21	0
X6573	Delorina	M	X2775	Florina	X4002	Grifer	21	0
X9005	Delrouval	M	X2836	Akane	X2957	Delcorf	21	0
X8819	Deltana	M	X2775	Florina	X3069	Granny Smith	20	1
X6136	Discovery	M	X2321	Worcester Pearmain	X6468	Beauty of Bath	21	0
X2458	Empire	M	Delicious	Delicious	X0557	Mc Intosh	21	0
X2629	Estiva	M	X0153	Usta Gorria	X0557	Mc Intosh	21	0
X7843	Galmac	M	X2474	Jerseymac	X4712	Gala	21	0
X5072	Judaine	M	X2738	Reinette du Mans	X2997	Priam	21	1
X3593	Judor	M	X5124	Douce Moën	X5177	Rouge de Trèves	21	1
X7842	Mairac	M	X4712	Gala	X7192	Maigold	21	0
X0569	Milton	M	X0557	Mc Intosh	X7200	Transparente Blanche	21	1
X6632	Nabella	M	X1239	James Grieve	X6920	La Paix	20	0
X7193	Orange Suisse	M	X1954	Cox's Orange Pippin	X7194	Ontario	21	0
X6841	Rubinola	M	X2596	Prima	X6395	Rubin	21	0
X6024	Telamon	M	X0557	Mc Intosh	X0972	Golden Delicious	21	0
X2403	Avzena Blahova	O	X6920	La Paix	X9116	Falquette	20	0
X1846	Belle de Mleiev	O	X0557	Mc Intosh	X2640	Reine des Reinettes	20	0
X9260	Belle Fille Orange	O	FRA0950	Calville de Dantzig	FRA1005	Zuccalmaglio	21	2
FRA0936	Bittenfelder Sämling	O	FRA0968	Striesselapfel	X1071	Reinette de Caux	21	0
X1618	Calville Rouge du Mt Dore	O	X1229	Grand Alexandre	X1291	Calville Rouge d'Hiver	21	0
FRA0481	Cinq Côtes	O	FRA0190	Pomme de Faure	FRA0426	Pomme d'Ile (Duval)	21	1
FRA0967	Comte d'Orloff	O	X1344	Reinette de Landsberg	X8233	Petite Madeleine	21	0
FRA1024	Cramoisie de Croncels	O	X1206	Calville du Roi	X2086	Nouvelle Europe	21	0
X1307	Directeur Lesage	O	X7200	Transparente Blanche	X8933	Baguette d'Été	20	1
FRA0308	Du Vivier	O	X0421	Belle de Magni	X1077	Reinette Étoilée	20	0
FRA1094	Dülmener Rosenapfel	O	X1071	Reinette de Caux	X8233	Petite Madeleine	21	1
X8212	Elizon	O	FRA0475	Belle Cotelée	X1059	Fenouillet Gris	21	0
X8717	France Deliquet	O	FRA0390	Reinette Rousse	FRA0531	Glane	21	0
FRA0959	Geheimrat Wessener	O	X0691	Boiken	X2640	Reine des Reinettes	21	0
X8719	George Carpenter	O	X2640	Reine des Reinettes	X9418	Sans Pareil de Peasgood	21	0
FRA1002	false Herrgottsapfel (Astillisch)	O	FRA0949	Signe Tillish	FRA1095	Astrakan Rouge	21	0
X0554	Jubilee	O	X0557	Mc Intosh	X2529	Newton Pippin	20	0

t5.38 **Table 5** (continued)

Offspring AcceNumber	Offspring AcceName	Modern/ Old	First candidate parent AcceNumber	First candidate parent AcceName	Second candidate parent AcceNumber	Second candidate parent AcceName	Trio loci compared	Trio loci mismatch
X0695	La Nationale	O	X1291	Calville Rouge d'Hiver	X9085	Romarin	20	0
X5199	Muscadet Petit Orne	O	FRA0749	Hauchecorne	X3830	Rousse de la Sarthe	21	0
FRA0847	Ognon	O	FRA0827	Vernajoux	X2640	Reine des Reinettes	20	0
FRA0105	Pomme Violette Thomassine	O	FRA0209	Court Pendu Rouge	X6471	Api Noir	21	1
X6176	Rose d'Ajoie Blaser	O	FRA0709	Pomme Raisin	X1291	Calville Rouge d'Hiver	20	0
X7199	Rose de Berne	O	FRA0709	Pomme Raisin	X1291	Calville Rouge d'Hiver	20	0
FRA0387	Rouge à Longue Queue	O	Delicious	Delicious	FRA0705	Pomme Gros	21	1
FRA0790	Rouge Des Vergnes	O	FRA0763	Fromentoune	X2998	De L'Estre	21	0
X8416	Transparente de Bois Guillaume	O	X1646	Saint Germain	X7201	Transparente de Croncels	18	0
FRA1047	Verollot	O	FRA0531	Glane	X9267	Nez de Chat	21	1
X9124	Vierge du Pilat	O	X0972	Golden Delicious	X2086	Nouvelle Europe	20	0

The number of loci mismatches among the number of loci compared is given

redundancy was detected in the present germplasm, which level reflects the traditional exchanges of plant material through grafting over geographic regions that occurred for a very long time, as underlined in other studies (Liang et al. 2015; Pina et al. 2014). As an example, a redundancy case has been interestingly solved thanks to the expertise of a member of the association "Les Croqueurs de Pommes" who noticed that the accession "Belle Josephine de Brie" (FRA0824, from the "La Brie" region of France) was very similar to another accession known as "Marie-Louise" (FRA0932, from another French region, "Pays de Montbéliard," located almost 400 km apart). Both accessions were shown to be duplicates according to their SSR profile (Online Resource ESM 1), which was consistent both with their very similar pomological description and with their denomination since "Josephine" and "Marie-Louise" were the first names of the two successive wives of Napoleon. A high genetic level of redundancy between accessions has already been observed within apple germplasms (Gross et al. 2012; Liang et al. 2015; Urrestarazu et al. 2012; van Treuren et al. 2010) and their identification is a preliminary step before undertaking a detailed genetic characterization of the germplasm. Furthermore, duplicate identification makes it possible to rationalize germplasm management. Accessions with the same name from different collections were mostly confirmed as duplicates. But, many errors were also highlighted which will necessitate further analysis. For duplicates with different names, further pomological and passport data analyses will also be necessary to check for true synonym status (when not already known), to identify interesting phenotypic mutations not accounted for with SSR markers (Gross et al. 2012; Liang et al. 2015), and to discard false synonymy resulting from grafting errors or erroneous former pomological identification. Cipriani et al. (2010) showed that many duplicates identified by SSR in grapevine are phenotypically well differentiated from each other for several traits, probably due to punctual genetic mutations, genomic structural variations or even epigenetic modifications. For these reasons, accessions sharing the same SSR fingerprinting should be subjected to further morphological and agronomical evaluation before being considered as strict replicates and being eliminated from a collection. Finally, several cases of homonymy, i.e., accessions with the same name but different genetic profiles, were also highlighted (e.g., "Double Bon Pommier" or "Api Double Rose"; Online Resource ESM 1). Some of them could have been checked with passport data to identify which of them were mislabeled within each pair and renamed as "unknown" in the collection. Grafting failures were especially identified through duplicate status with known rootstocks (e.g., "MM106"). Others could not be differentiated with the available data and should be evaluated in the field in order to identify those that are inconsistent with the identity assigned to them.

Table 6 Descriptive information for the overall sets of unique diploid old dessert or cider apple cultivars and for the core collections defined in both sets

Population	*N*	H_o	H_e	Mean no. alleles	Allelic richness	
					N = 278	*N* = 48
Overall unique diploid Old Dessert cvrs	737	0.78	0.81	18.1	14.5	9.8
Core collection CC-dessert-278	278	0.77	0.81	16.4	(16.4)	9.9
Core Collection CC-dessert-48	48	0.74	0.84	12.9		(12.9)
Overall unique diploid Old Cider cvrs	188	0.78	0.81	13.5		9.5
Core collection CC-cider-48	48	0.76	0.83	11.2		(11.2)

Summary statistics include the sample size (*N*), observed (H_o) and expected (H_e) heterozygosity, average number of alleles/locus (mean no. alleles), and allelic richness. First, allelic richness is scaled to *N* = 278 for comparing the overall dessert collection to the CC-dessert-278 for which the average number of alleles is copied as the reference allelic richness (between brackets). Second, it is scaled to *N* = 48 for comparing: (i) the overall dessert collection and the CC-dessert-278 to the CC-dessert-48, and (ii) the overall cider collection to the CC-cider-48 278 for which the average number of alleles is copied as the reference allelic richness (between brackets)

The average rate of putative triploid accessions found in our germplasm was 18.5 %. It is noteworthy that Modern Cultivars showed a much lower rate of putative triploids (~5 %) compared to Old Dessert (~20 %) and Old Cider (~18 %) cultivars. This reflects that the empirical selection performed by farmers and gardeners in the past (until 1950) has been more efficient than modern selection for this characteristic, which is however frequently linked to a larger fruit size. Other authors also found even higher rates of triploids in their national or regional collections: 28 % (Pereira-Lorenzo et al. 2007), 24 % (Urrestarazu et al. 2012), and 21 % (van Treuren et al. 2010) of triploids. Checking the triploid status of the postulated accessions by flow cytometry should be performed in the near future for at least a part of the French apple germplasm.

Genetic Diversity and Structure Observed in the French Apple Germplasm

As expected, because apple is a self-incompatible cross-pollination species, both observed and expected heterozygosity values were high regardless of the SSR marker, suggesting that the collection was highly diverse. In comparison with other studies, the mean $H_e = 0.82$ observed in our study was similar to those reported on apple by Urrestarazu et al. (2012) ($H_e = 0.82$), Larsen et al. (2006) ($H_e = 0.78$), Gasi et al. (2010) ($H_e = 0.78$), Pereira-Lorenzo et al. (2007) ($H_e = 0.80$), Gharghani et al. (2009) ($H_e = 0.83$), Liang et al. (2015) ($H_e = 0.83$), and Coart et al. (2003) ($H_e = 0.72$ for wild apple populations and 0.77 for ornamental apple populations). Conversely, H_e was only 0.44 for grape (Cipriani et al. 2010), 0.69 for cacao (Motilal et al. 2009) and 0.04 for a self-pollinating species such as rice (Faivre-Rampant et al. 2011). The difference with the mean observed heterozygosity ($H_o = 0.79$) could be partly explained because genotypes showing a single peak at a given locus were considered as homozygous, leading to an underestimation of heterozygosity for loci with null alleles that occur at high frequency (Liang et al. 2015). Analysis of the global structuration over accessions showed that two types of weak but significant structures could be observed. On one hand, an a priori structure was found between "OD," "OC," and "MC." The highest F_{ST} value (0.048) was observed between "MC" and "OC," whereas the smallest one (0.014) was observed between "OD" and "OC." These results logically reveal that the "MC" group derives from founders that are not all fully and equally represented in the old cultivars subpopulations, thus generating a switch in allelic representation between modern and old subpopulations. The results are also consistent with the very weak differentiation between dessert and cider apples, as already shown by Cornille et al. (2012) on a partially redundant set of accessions. A recent study by our group also underlined the difficulties in finding loci involved in the dessert vs. cider differentiation at the genome level (Leforestier et al. 2015).

On the other hand, the use of STRUCTURE software showed that the French apple germplasm also had a significant structure between three subgroups, with F_{ST} values ranging from 0.026 to 0.060. It is noteworthy that these subgroups identified by STRUCTURE corresponded, albeit approximately, to the 3 a priori "OD," "OC," and "MC" subpopulations. However, it could be highlighted that "OC" and "MC" were mostly shared in separate subgroups 1 and 3 (respectively) identified by STRUCTURE and with a strong assignment probability, whereas "OD" was found in the three subgroups with a lower assignment probability. Coherently, all major founders of modern cultivars were assigned to subgroup 3 with high membership probabilities (qI ≥ 0.8 or close to 0.8). This was especially the case for "Golden Delicious," "McIntosh," "Jonathan," "Delicious," "Cox's Orange Pippin," "Rome Beauty," "James Grieve," "Worcester Pearmain," and "Granny Smith," each of these founding cultivars belonging to the "OD" subpopulation (Online Resource ESM 1). Many other well-known international cultivars were assigned to sugroup 3, such as "Reinette Dorée de Blenheim" (syn.

"Blenheim Reinette"), "Borowitsky" (syn. "Charlamowsky" or "Duchesse of Oldenburg"), "Grand Alexandre" (syn. "Alexander" or "Aporta"), "Reine des Reinettes" (syn. "King of Pippins"), "Transparente Blanche/Jaune" (syn. "White/Yellow Transparente" or "Papirovka"), "Dülmener Rosenapfel," "Winter Banana," "Lady Hamilton" (Online Resource ESM 1). Conversely, subgroup 2 was almost only gathering accessions with typical French names. This subgroup also seemed to gather more cultivars from south of France (especially from "Parc Naturel Régional du Lubéron" [10], "Fruits Oubliés Réseau" [12], "Conservatoire des Espèces Fruitières et de Vignes Anciennes" [8], and "Conservatoire Végétal Régional d'Aquitaine" [3]; Fig. 1) whereas the collections from the north and northwest of France hardly contained accessions assigned to this subgroup (e.g., "Centre Régional de Ressources Génétiques du Nord-Pas-de-Calais" [4], "Confédération des Producteurs de Fruits d'Alsace" [6], "I z'on creuqué eun' pomm'" [7], "Les Mordus de la Pomme" [5], "Verger Conservatoire de Pétré" [9], or western sections of "Les Croqueurs de Pommes" [2]). Based on the allelic richness parameter, this subgroup 2 was as diverse as subgroup 1 but more diverse than subgroup 3. For subgroup 1, the large contribution of cider accessions with high membership probability may indicate that these accessions share a common genetic basis with the dessert accessions assigned to this subgroup, especially those with high membership probability. When focusing on accessions with $qI \geq 0.8$ in subgroup 1, a large proportion of dessert accessions came from collections from north and west of France (e.g., "Centre Régional de Ressources Génétiques du Nord-Pas-de-Calais," "I z'on creuqué eun' pomm'," "Les Mordus de la Pomme"), as did many of the cider accessions belonging to this subgroup.

All the F_{ST} observed were low or moderate, indicating a weak differentiation among subgroups. Generally, these low or moderate differentiations are expected for out-crossing species like apple tree and fit with the large gene flow observed both within domesticated apple population and between domesticated and wild apple populations, as described by Cornille et al. (2012). The F_{ST} values obtained in the present study are consistent with those observed on other apple germplasms. Gharghani et al. (2009) obtained a F_{ST} of 0.087 between subpopulations of Iranian apple germplasm; Pereira-Lorenzo et al. (2007) observed an F_{ST} of 0.058 between nonnative and local apple cultivars; Richards et al. (2009) observed a mean $F_{ST} = 0.05$ between sites for apples; Coart et al. (2003) observed an F_{ST} of 0.011 between wild and domesticated apples populations and 0.060 between wild and ornamental apples populations.

Parentage Analysis Within the French Apple Germplasm

The initially known parentage of 18 Modern cultivars was correctly inferred in all cases for at least one of the two parents, and in 77 % of the cases for the two parents. These expected results served as control and validated the parentage assignment obtained with the CERVUS software (Kalinowski et al. 2007; Marshall et al. 1998), indicating that the number and informativeness of SSR markers were sufficient. Furthermore, some inconsistencies with the expected parentage of two Modern cultivars could be documented. "Nabella," bred at the Research Institute of Pomology, Holovousy, Czech Republic, as "Nonnetit" (synonym = "Mother apple") × "Starking Delicious" (Blazek et al. 1995; Fischer et al. 2004), should be corrected as "Nonnetit = Mother apple = La Paix" × "James Grieve." Interestingly, the inferred female parent, "La Paix," is identified as a putative synonym of "Nonetti" (=FRA0918, collected by Croqueurs de Pommes de Lorraine), which is most probably a typing error of "Nonnetit." Also, Deltana, bred by the Delbard nurseries, Malicorne, France, as ["Golden Delicious" × "Grive Rouge"] × "Florina" was corrected as "Granny Smith" × "Florina."

Several interesting features could be observed such as the rather frequent occurrence of some cultivars as parents of old cultivars (4 × "King of Pippin" = "Reine des Reinettes," 4 × "Calville Rouge d'Hiver"), or the geographic convergence of parentage (e.g., "Ognon" and "Vernajoux" are both described as traditionally grown in the French "Haute-Vienne" department; "Verollot" and "Nez de Chat" are two cider cultivars from the "Pays d'Othe," another French region). Complete paternity assignment of some well-known old cultivars was proposed, including "Calville Rouge du Mont Dore," inferred as a cross between the Ukrainian cultivar "Alexander/Grand Alexandre" and the French cultivar "Calville Rouge d'Hiver." The German cultivar "Dülmener Rosenapfel" was inferred as a cross between "Reinette de Caux" (also known as "Dutch Mignonne" since it is thought to come from the Netherlands) and "Petite Madeleine" (with "St Jacques" and "Bouchon" as identified duplicates). It is thus not a seedling from "Gravenstein," as frequently reported. Another German cultivar "Bittenfelder Sämling" was also shown to result from a cross involving "Reinette de Caux/Dutch Mignonne." Intriguingly, "Reinette de Caux" was also indicated as a putative parent of the famous triploid Dutch cultivar "Belle de Boskoop" by Ramos-Cabrer et al. (2007). Interestingly, two "Rose" cultivars both originally from Switzerland ("Rose de Berne" and "Rose d'Ajoie Blaser") were inferred as full-sibs from the same cross between "Pomme Raisin" (synonym of "Sauergrauech") and "Calville Rouge d'Hiver." From a practical point of view, identifying cultivars that are frequently inferred as parents of other cultivars may indicate their particular interest as progenitors for new breeding purposes. Especially, they could be preferred for the purpose of using old germplasm to enlarge the genetic base of modern breeding programs. However, some preliminary evaluation is necessary since it may also be the case that

the higher frequency of parentage would reflect a higher frequency of geographic distribution of these particular cultivars in France in the past. Moreover, the empirical breeding goals of farmers and gardeners one or several centuries ago may be somewhat divergent from the present breeding goals of modern breeding programs. Finally, by combining this study with other germplasm analyses performes in other European countries (e.g., (Liang et al. 2015; Urrestarazu et al. 2012; van Treuren et al. 2010), more complete European-wide multi-generation pedigree networks could be searched for in our germplasm, as was done on old grapevine cultivars by Lacombe et al. (2013) or on recent apple cultivars with known pedigrees by Salvi et al. (2014).

Definition of INRA Diploid Core Collections for Association Genetics Studies

Three core collections were defined based on genetic diversity. Additional phenotypic information was not enough available to help building the core collections despite it can help to optimize further screening and analyses of agronomical traits (Nicolaï et al. 2013). In grapevine, Emanuelli et al. (2013) compared a phenotypic and a genetic core collection. They showed that the latter retained more genetic diversity while maintaining a similar phenotypic variability. In the present study, the core collections were based only on the SSR allelic diversity and should thus maximize the genetic variation. The results showed that only a small number of accessions is needed to retain the most frequent alleles since up to 71 % of the observed alleles were represented with only 48 conserved accessions of the dessert sub-core collection. The high level of heterozygosity in apple is the major factor contributing to the capture of a large part of the genetic diversity with such a small number of individuals.

These core collections are already used for various goals as exemplified by the study of the differentiation between dessert and cider apples (Leforestier et al. 2015). The dessert apple core collection is also currently being phenotyped for various agronomical traits and SNP genotyped within the framework of the European project FruitBreedomics (Laurens et al. 2010). These data will thus make it possible to perform genome-wide association studies to decipher the genetic architecture of important traits such as fruit quality and biotic or abiotic stress resistance.

Conclusion

This study is the largest one ever to be performed at the national level with such a large number of SSR markers. The representativeness of the French apple germplasm was excellent thanks to the strong involvement of all the major repositories and pomological societies. As already shown in various other studies, the genetic diversity is especially large in domesticated apple, which exhibits a high level of heterozygosity. SSR marker data helped to identify a large number of redundancies ("duplicates") both within and between collections, information that is extremely useful for curating the germplasm. Additional phenotypic and passport data checking is now necessary to solve pending identification questions. The overall diversity structure was shown to be rather weak and partially coincided with the cultivar selection date and the usage of the cultivars. Several unknown parentages were inferred, underlying the unaware preference of particular genotypes as parents of old cultivars during the empirical selection process performed in the past. Finally, core collections were established that will be used for further research projects aimed at gaining insight into genetic and functional bases of major agronomical traits in apple. To conclude, we highly recommend the use of the 16 SSRs proposed by the *Malus/Pyrus* ECPGR group for any future apple fingerprinting studies since it will allow the allelic adjustment of SSR data over countries, thus empowering future worldwide analyses and comparisons.

In the "Cultivar Usage" column, the asterisk (*) indicates an accession initially classified as Cider (resp. Dessert) that has finally been considered as Dessert (resp. Cider) in the statistical analyses because of more consistent information derived from its duplicate(s). In addition, seven accessions of the Dessert core collection ("CC-dessert-278" in the column "Core Collection") identified by "Cider!" were initially considered as Dessert cultivars but finally corrected as Cider cultivars thanks to additional information from partners; since the core collection was already vegetatively propagated and genotyped for further association studies, they were maintained in the Dessert core collection for contingency reasons.

In the "Old/Modern" column, the asterisk (*) indicates an accession initially classified as Old (resp. Modern) that has finally been considered as Modern (resp. Old) in the statistical analyses because of more consistent information derived from its duplicate(s).

In the "Subgroup" column, a bold number indicates that the highest subgroup membership probability (qI) is greater than 0.8.

In the "Analyzed/Excluded" column,

- A indicates an accession that has been considered in the statistical analyses.
- E indicates an accession that has been excluded from the statistical analyses (mostly because another duplicated accession has been retained; in that case, the subgroup membership and the qI probability has been imputed according to the analyzed duplicate accession).

- E (SSR) indicates an accession that has been excluded from the statistical analyses because of an exceedingly low number of SSR marker data (<17 SSR).
- E (Ext.) indicates an accession that has been excluded from the statistical analyses because of its extreme situation in a preliminary FCA.
- E (Rs) indicates an accession that has been excluded from the statistical analyses because of its rootstock status identified using the SSR profile (generally MM106 instead of the expected accession).

Acknowledgments The authors thank the Fondation pour la Recherche sur la Biodiversité (FRB), which contributed to the funding of this study (entitled "CorePom" AAP-IN-2009-14), and the French "Pays de la Loire" region for the post-doctoral grant of L. Lassois.

They greatly thank all of the following amateur groups, associations, governmental and non-governmental organizations for their kind contribution of apple leaf material for this study, in addition to their general support and collaboration:

• Association Nationale des Amateurs Bénévoles pour la Sauvegarde des Variétés Fruitières Régionales en voie de Disparition, dite "Les Croqueurs de Pommes" (especially J. Lefèvre, J. Marchand, J.C. Hénin, H. Fourrey, C. Scribe)
• Association "Les Mordus de la Pomme" (especially J.F. Aubert)
• Association pour la Sauvegarde des Variétés Fruitières du Terroir, dite "I z'on creuqué eun'pomm" (especially E. Verbrugghe)
• Association "Fruits Oubliés Réseau" (especially C. Sunt)
• Confédération des Producteurs de Fruits d'Alsace (especially P. Heitzler, D. Siegel)
• Conseil Général du Tarn - Service Biodiversité Végétale/Conservatoire des Espèces Fruitières et de Vignes Anciennes (CREFAV) (especially I. Calvière)
• Conservatoire Végétal Régional d'Aquitaine (CVRA) (especially E. Leterme)
• Espaces Naturels Régionaux/Centre Régional de Ressources Génétiques du Nord-Pas-de-Calais (CRRG) (especially J.B. Rey, R. Stievenard)
• Jardin du Luxembourg (especially G. Bachelier)
• Parc Naturel Régional du Luberon (especially J.P. Talichet)
• Société Pomologique du Berry (especially F. Moyse, D. Labouret)
• Verger Conservatoire de Pétré (especially S. Guérin)
• Union Pomologique de France (UPF) (especially D. Retournard)

The staffs of the ANAN genotyping platform of the SFR 149 QUASAV (Angers, France), the GENTYANE genotyping platform (INRA, Clermont-Ferrand, France) and the Unité Expérimentale Horticole (INRA, Angers, France) are warmly acknowledged, for their help in producing the genotyping data and for maintaining the apple germplasm collection in experimental plots, respectively.

Thanks are also due to Xavier Perrier, working at the CIRAD, Montpellier, France, for his valuable assistance in data analysis using DARwin software, to Tristan Marshall for his assistance in using CERVUS software, and to Laurana Serres-Giardi for her help with structure and duplicate analyses.

References

Aldrich J, Cullis C (1993) RAPD analysis in flax: optimization of yield and reproducibility using klenTaq 1 DNA polymerase, chelex 100, and gel purification of genomic DNA. Plant Mol Biol Rep 11:128–141. doi:10.1007/BF02670471

Aranzana M, Abbassi E-K, Howad W, Arus P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11:69

Belkhir K, Borsa P, Chikhi L (2004) GENETIX 4.05.02, logiciel sous WindowsTM pour la genetique des populations Laboratoire Genome, Populations, Interactions CNRS UMR 5000, Montpellier

Blazek J, Kloutvor J, Paprstein F, Vondracek J (1995) New apple cultivar 'Nabella'. Vědecké práce ovocnářské 14:119–125

Botstein D, White R, Skolnick M (1980) Davis R. Construction of a genetic linkage map in man using RFLP Am j hum genet 32:314–331

Cipriani G et al (2010) The SSR-based molecular profile of 1005 grapevine (Vitis vinifera L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theor Appl Genet 121:1569–1585. doi:10.1007/s00122-010-1411-9

Coart E, Vekemans X, Smulders MJM, Wagner I, Van Huylenbroeck J, Van Bockstaele E, Roldán-Ruiz I (2003) Genetic variation in the endangered wild apple (Malus sylvestris (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. Mol Ecol 12:845–857. doi:10.1046/j.1365-294X.2003.01778.x

Cornille A et al (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8:e1002703

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity 93:504–509

Earl D, VonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method Conservation. Genet Resour 4:359–361. doi:10.1007/s12686-011-9548-7

Emanuelli F et al (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Evett I, Weir B (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates Inc., USA

Faivre-Rampant O et al (2011) Assessment of genetic diversity in Italian rice germplasm related to agronomic traits and blast resistance (Magnaporthe oryzae). Mol Breeding 27:233–246. doi:10.1007/s11032-010-9426-0

Fischer C, Richter K, Blazek J (2004) Testing of Czech cultivars and advanced selections of apples for fire blight (Erwinia amylovora) resistance. Hortic Sci 31:8–11

Frankel O, Brown A (1984) Plant genetic resources today: a critical appraisal. In: Holden J, Williams J (eds) Crop genetic resources: conservation and evaluation. George Allen and Unwin, London, pp 249–257

Garkava-Gustavsson L, Kolodinska Brantestam A, Sehic J, Nybom H (2008) Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis. Hereditas 145:99–112. doi:10.1111/j.0018-0661.2008.02042.x

Gasi F, Simon S, Pojskic N, Kurtovic M, Pejic I (2010) Genetic assessment of apple germplasm in Bosnia and Herzegovina using microsatellite and morphologic markers. Sci Hortic 126:164–171. doi:10.1016/j.scienta.2010.07.002

Gharghani A et al (2009) Genetic identity and relationships of Iranian apple (Malus × domestica Borkh.) cultivars and landraces, wild Malus species and representative old apple cultivars based on simple sequence repeat (SSR) marker analysis. Genet Resour Crop Evol 56:829–842. doi:10.1007/s10722-008-9404-0

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. Theor Appl Genet 96:1069–1076. doi:10.1007/s001220050841

Gross BL, Henk AD, Richards CM, Fazio G, Volk GM (2014) Genetic diversity in Malus ×domestica (Rosaceae) through time in response to domestication. Am J Bot 101:1770–1779. doi:10.3732/ajb.1400297

Gross BL, Volk GM, Richards CM, Forsline PL, Fazio G, Chao CT (2012) Identification of "duplicate" accessions within the USDA-ARS National Plant Germplasm System Malus collection. J Am Soc Hortic Sci 137:333–342

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in Malus X domestica (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94:249–254. doi:10.1007/s001220050407

Gulsen O, Roose ML (2001) Chloroplast and nuclear genome analysis of the parentage of lemons. J Am Soc Hortic Sci 126:210–215

Hokanson SC, Lamboy WF, Szewc-McFadden AK, McFerson JR (2001) Microsatellite (SSR) variation in a collection of Malus (apple) species and hybrids. Euphytica 118:281–294. doi:10.1023/A:1017591202215

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a Malus×domestica borkh. core subset collection. Theor Appl Genet 97:671–683. doi:10.1007/s001220050943

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. doi:10.1093/bioinformatics/btm233

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Kloosterman A, Budowle B, Daselaar P (1993) PCR-amplification and detection of the human D1S80 VNTR locus. Int J Leg Med 105:257–264. doi:10.1007/BF01370382

Lacombe T, Boursiquot JM, Laucou V, Di Vecchi-Staraz M, Peros JP, This P (2013) Large-scale parentage analysis in an extended set of grapevine cultivars (Vitis vinifera L.). Theor Appl Genet 126:401–414. doi:10.1007/s00122-012-1988-2

Larsen A, Asmussen C, Coart E, Olrik D, Kjær E (2006) Hybridization and genetic variation in Danish populations of European crab apple (Malus sylvestris). Tree Genet Genomes 2:86–97. doi:10.1007/s11295-005-0030-0

Laurens F et al (2010) Review on apple genetics and breeding programs and presentation of a new initiative of a news European initiative to increase fruit breeding efficiency. Journal of fruit science 27:102–107

Leforestier D, Ravon E, Muranty H, Lemaire C, Giraud T, Durel C, Branca A (2015) Genomic basis of the differences between cider and dessert apple varieties. Evol Appl 8:650–661

Liang W, Dondini L, De Franceschi P, Paris R, Sansavini S, Tartarini S (2015) Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. Plant Mol Biol Rep 33:458–473

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (Malus x domestica Borkh.). Mol Breeding 10:217–241. doi:10.1023/A:1020525906332

Mariette S, Tavaud M, Arunyawat U, Capdeville G, Millan M, Salin F (2010) Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. BMC Genet 11:77

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655. doi:10.1046/j.1365-294x.1998.00374.x

Moriya S, Iwanami H, Okada K, Yamamoto T, Abe K (2011) A practical method for apple cultivar identification and parent-offspring analysis using simple sequence repeat markers. Euphytica 177:135–150. doi:10.1007/s10681-010-0295-8

Motilal L, Zhang D, Umaharan P, Mischke S, Boccara M, Pinney S (2009) Increasing Accuracy and Throughput in Large-Scale Microsatellite Fingerprinting of Cacao Field Germplasm Collections Tropical Plant Biol 2:23–37. doi:10.1007/s12042-008-9016-z

Nicolaï M, Cantet M, Lefebvre V, Sage-Palloix A-M, Palloix A (2013) Genotyping a large collection of pepper (Capsicum spp.) with SSR loci brings new evidence for the wild origin of cultivated C. annuum and the structuring of genetic diversity by human selection of cultivar types. Genet Resour Crop Evol 60:2375–2390. doi:10.1007/s10722-013-0006-0

Noiton D, Alspach P (1996) Founding clones, inbreeding, coancestry and status number of modern apple cultivars. J Am Soc Hortic Sci 121:773–782

Pereira-Lorenzo S, Ramos-Cabrer A, Díaz-Hernández M (2007) Evaluation of genetic identity and variation of local apple cultivars (Malus × domestica Borkh.) from Spain using microsatellite markers. Genet Resou Crop Ev 54:405–420. doi:10.1007/s10722-006-0003-7

Pereira-Lorenzo S, Ramos-Cabrer AM, González-Díaz AJ, Díaz-Hernández MB (2008) Genetic assessment of local apple cultivars from La Palma, Spain, using simple sequence repeats (SSRs) Sci Hortic 117:160–166 doi:http://dx.doi.org/10.1016/j.scienta.2008.03.033

Perrier X, Flori A, Bonnot F (2003) Data analysis methods Genetic diversity of cultivated tropical plants., pp 43–76

Perrier X, Jacquemoud-Collet J (2006) DARwin software

Pina A, Urrestarazu J, Errea P (2014) Analysis of the genetic diversity of local apple cultivars from mountainous areas from Aragon (Northeastern Spain) Sci Hortic 174:1–9 doi:http://dx.doi.org/10.1016/j.scienta.2014.04.037

Prat D, Faivre-Rampant P, Prado E (2006) Analyse du génome et gestion des ressources génétiques forestières., Savoir Faire

Pritchard J, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959

Ramos-Cabrer AM, Diaz-Hernandez M, Pereira-Lorenzo S (2007) Morphology and microsatelittes in spanish apple collections. J Hortic Sci Biotech 82:257–265

Richards C, Volk G, Reilley A, Henk A, Lockwood D, Reeves P, Forsline P (2009) Genetic diversity and population structure in Malus sieversii, a wild progenitor species of domesticated apple. Tree Genet Genomes 5:339–347. doi:10.1007/s11295-008-0190-9

Rosenberg NA (2004) distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. doi:10.1046/j.1471-8286.2003.00566.x

Salvi S, Micheletti D, Magnago P, Fontanari M, Viola R, Pindo M, Velasco R (2014) One-step reconstruction of multi-generation pedigree networks in apple (Malus × domestica Borkh.) and the parentage of Golden Delicious. Mol Breeding 34:511–524. doi:10.1007/s11032-014-0054-y

Silfverberg-Dilworth E et al (2006) Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome. Tree Genet Genomes 2:202–224. doi:10.1007/s11295-006-0045-1

Song Y, Zhai H, Yao Y-x, Li M, Du Y-p (2006) Analysis of Genetic Diversity of Processing Apple Varieties Agr Sci China 5:745–750 doi:http://dx.doi.org/10.1016/S1671-2927(06)60119-3

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24:2498–2504. doi:10.1093/bioinformatics/btn478

Upadhyaya HD, Yadav D, Dronavalli N, Gowda CLL, Singh S (2010) Mini core germplasm collections for infusing genetic diversity in plant breeding programs. Electronic Journal of Plant Breeding 1:1294–1309

Urrestarazu J, Miranda C, Santesteban L, Royo J (2012) Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. Tree Genet Genomes 8:1163–1180. doi:10.1007/s11295-012-0502-y

van Treuren R, Kemp H, Ernsting G, Jongejans B, Houtman H, Visser L (2010) Microsatellite genotyping of apple (Malus × domestica Borkh.) genetic resources in the Netherlands: application in collection management and variety identification Genet Resour Crop Evol 57:853–865 doi:10.1007/s10722-009-9525-0

Velasco R et al. (2010) The genome of the domesticated apple (Malus [times] domestica Borkh.) Nat Genet 42:833–839 doi:http://www.nature.com/ng/journal/v42/n10/abs/ng.654.html#supplementary-information

Vinatzer BA, Patocchi A, Tartarini S, Gianfranceschi L, Sansavini S, Gessler C (2004) Isolation of two microsatellite markers from BAC clones of the Vf scab resistance region and molecular characterization of scab-resistant accessions in Malus germplasm. Plant Breeding 123:321–326. doi:10.1111/j.1439-0523.2004.00973.x

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Way R et al. (1990) Apples. In: Moore J, Ballington J (eds) Genetic resources of temperate fruit and nut crops. International society for horticultural science, Wageningen, Netherlands, pp 1–62

Zeigler RS (2013) Food security, climate change and genetic resources. In: Jackson M, Ford-Lloyd B, Parry M (eds) Plant Genetic Resources and Climate Change. CAB International, UK, pp 1–15

Zhen Y, Li Z, Huang H, Wang Y (2004) Molecular Characterization of Kiwifruit (Actinidia) Cultivars and Selections Using SSR Markers. J Am So Hortic Sci 129:374–382

AUTHOR QUERY

AUTHOR PLEASE ANSWER QUERY.

Q1. Please check values here if presented correctly.