Proposition of an SNP set to replace SSRs for standardized cultivar identification in apple

Article

Accepted Version

Muranty, H., Denancé, C., Petiteau, A., Howard, N., Micheletti, D., García-Gómez, B.E., Aranzana, M.J., Confolent, C., Poncet, C., Vanderzande, S., López-Girona, E., Chagné, D., Ordidge, M. ORCID: https://orcid.org/0000-0003-0115-5218, Peace, C. and Durel, C. E. (2024) Proposition of an SNP set to replace SSRs for standardized cultivar identification in apple. Acta Horticulturae, 1412. pp. 25-32. ISSN 2406-6168 doi: 10.17660/ActaHortic.2024.1412.4 Available at https://centaur.reading.ac.uk/115898/

Published version at: https://www.ishs.org/ishs-article/1412_4
Identification Number/DOI: 10.17660/ActaHortic.2024.1412.4
<https://doi.org/10.17660/ActaHortic.2024.1412.4>

Publisher: International Society for Horticultural Science

Proposition of a SNP set to replace SSRs for standardized cultivar identification in apple

H. Muranty[1], C. Denancé[1], A. Petiteau[1], N. Howard[2], D. Micheletti[3], B.E. García-Gómez[4], M.J. Aranzana[4,5], C. Confolent[6], C. Poncet[6], S. Vanderzande[7], E. López-Girona[8], D. Chagné[8], M. Ordidge[9], C. Peace[10] and C.E. Durel[1]

[1]Univ Angers, INRAE, Institut Agro, IRHS, SFR QUASAV, F-49000 Angers, France; [2]Fresh Forward Breeding and Marketing B.V., Huissen, the Netherlands; [3]Research and Innovation Center, Edmund Mach Foundation, 38010 S. Michele all'Adige, Italy; [4]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain; [5]Institute of Agrifood Research and Technology (IRTA), Barcelona, Spain; [6]UCA, INRAE, Genotyping and Sequencing Platform Gentyane, UMR GDEC, F-63000 Clermont-Ferrand, France; [7]Plant Breeding, Wageningen University and Research, Wageningen 6708PB, the Netherlands; [8]Plant & Food Research, Palmerston North 4442, New Zealand; [9]Department of Crop Science, School of Agriculture, Policy and Development, University of Reading, Reading RG6 6EU, UK; [10]Department of Horticulture, Washington State University, Pullman, WA, USA

Abstract

Apple is one of the most important fruit crops grown in temperate regions. Ex situ conservation of apple genetic resources mainly relies on grafted trees maintained in orchards, which can be space- and labor-intensive. To help collection managers to compare their germplasm with others' at both the national and international levels, Malus UNiQue genotype codes (MUNQ) were proposed for all accessions sharing the same genetic marker profile. These genetic profiles were initially obtained using a standardized set of 16 SSRs. However, an impending difficulty in SSR analysis is expected and many genetic studies nowadays use SNP markers. SNP-based MUNQ assignment is expected to be more streamlined, but just like for SSRs, it will be essential that a consistent marker set is used across studies worldwide to enable comparisons among them without genotyping each new unique individual with a more expensive array. For this reason, we developed a set of 96 SNP markers that can unequivocally distinguish cultivars in apple collections and detect redundancy faster at low cost. To support this approach, we started from genome-wide SNP genotypic data obtained with the apple 20K and/or 480K arrays for 2120 unique individuals, including 2036 with a MUNQ code assigned through SSR markers. A total of 182 SNPs were tested using the KASP technology in nanofluidic Integrated Fluidic Circuit (IFC) to finally choose a set of 96 SNPs ensuring that each pair of individuals among all 2120 was distinguished by at least 6 SNPs in the absence of missing data. A comparison of MUNQ code assignment for 754 newly genotyped accessions using the standardized set of 16 SSRs and the new set of 96 SNPs was performed. It was possible to group accessions with essentially unique profiles using SNP data, and the groups obtained were almost always consistent with groups obtained with SSR data. Thus, a transition from the SSR-based MUNQ system to a SNP-based MUNQ system will be possible.

Keywords: DNA-based diagnostics, genetic markers, germplasm collection, *Malus domestica*

INTRODUCTION

Apple (*Malus domestica* Borkh.) is an important fruit crop in temperate regions. There are thousands of cultivars conserved worldwide in either institutional, regional, or amateur collections (Bramel and Volk 2019). These collections traditionally consist of vegetatively propagated (grafted) trees that maintain the genetic combinations which arose in the initial seedling selections. Labeling errors can occur when propagating the material to rejuvenate a

collection and, given that some cultivars have been clonally distributed for hundreds of years, many variations in naming, caused by translations, mislabeling, and local re-naming are known to have occurred between collections. Morphological identification is complicated by genotype × environment interactions and is usually limited to a subset of well-described cultivars. On the contrary, characterization using genetic markers enables propagation errors and naming variations to be identified quite rapidly on a large scale and genotypic data can be compiled into datasets that are searchable and can be compared at an international level, noting that morphological comparisons might still be required to distinguish clonal mutants and ”sports”.

To help collection managers to compare their germplasm with others’ at the national and international levels, Malus UNiQue genotype codes (MUNQ) were previously proposed and assigned to accessions sharing the same genetic marker profile (Muranty et al. 2020; Durel et al. 2023). These genotype codes were initially obtained using a common set of SSR markers. However, SSR analysis is expected to become difficult when capillary sequencers will not be maintained anymore. Additionally, SSR marker profiles are time-consuming to produce because scoring is not fully automatic and requires harmonization between labs. Indeed, harmonization of SSR data requires a careful calibration of allele sizes, often using reference accessions, and sometimes requires re-genotyping of subsets of samples to confirm allele size calibration.

SNP-based MUNQ assignment is predicted to be more streamlined because SNP genotyping is more amenable to high-throughput automation. However, it will remain essential that a consistent marker set is used across studies to enable comparisons among them. With this in mind, our objective was to define a set of 96 SNP markers that could be used to distinguish cultivars in apple collections.

MATERIAL AND METHODS

Test data

To maximize alignment with published genotypic datasets, we chose SNPs that were common and compatible (Howard et al. 2021b) between the previously established 20K (Bianco et al. 2014) and 480K apple genotyping arrays (Bianco et al. 2016). A dataset involving 2120 unique individuals genotyped with either the 20K or the 480K arrays (or both) was assembled from previous works (Vanderzande et al. 2019; Muranty et al. 2020; Howard et al. 2021a, b, 2022, 2023). It comprised 1947 diploid and 173 triploid individuals. We considered a total of 7711 SNPs present on both the 20K and 480K arrays and used in the apple pedigree reconstruction project (Howard et al. 2018) as candidates for the fingerprinting set. SNP markers with a minor allele frequency < 0.001 or > 10% missing data were rejected, which reduced the number of SNPs considered to 7666.

A SNP-set for cultivar identification had previously been developed and used at Washington State University (WSU). To develop this set, another dataset involving 200 founder cultivars genotyped with the 8K (Chagné et al. 2012) or 20K array was assembled from data curated in Vanderzande et al (2019). Out of the 3350 SNPs from the 8K array that had good clustering and less than 5% call errors (classified as Type 1 in Table S8 of Vanderzande et al (2019)), 2418 that were also present on the 20K array were considered as candidates for this work. Only 1607 of these were however overlapped with the 7666 SNPs of the first dataset.

The total number SNPs used for the first dataset, the second dataset or both was 8477.

Choice of SNP marker sets

An R script was written, inspired by the method described in Winfield et al (2020), to identify sets of SNPs capable of differentiating all 2120 unique individuals within in the test data. As the dataset contained a mix of diploid and triploid individuals, a special rule was set for

diploid to triploid comparison: we accepted that heterozygote (AB) diploid genotypes would not be differentiated from either AAB or ABB genotypes of triploids. To maximize the robustness of the system, five independent sets of SNPs were initially chosen, each set sufficient to differentiate all individuals. The first three sets were chosen without accounting for the SNP genomic positions. Subsequently, all SNPs in linkage disequilibrium (r^2) > 0.3 with any SNP in the first three sets were excluded before the choice of the fourth set. The fifth set was then identified exclusively from SNPs on five chromosomes that were poorly represented in sets one to four. From the 7666 SNPs considered, five sets were selected, each containing 14 SNPs.
The WSU set was built following a similar approach as sets one to three above but using the second dataset of 2418 SNPs for 200 founder cultivars. In this attempt, five SNPs were also included because they were considered informative for some traits (fruit skin overcolor and fruit acidity). However, one was only on the 8K array and one only on the 50K array of Rymenants et al. (2020). This resulted in a set of 20 SNPs.

Before reaching a final list that was used in a full-scale experiment, three rounds of tests with the KASP approach (see below) were performed, each testing 96 or 95 SNPs, with partial overlap of the SNPs included between rounds of tests. Over all three rounds of tests, a total of 182 SNPs were tested. After testing the SNPs, some SNPs had to be removed either because of excessive missing data or because of discordances with 20K or 480K data (see results for thresholds). Each of the six sets, including the WSU set, was then compensated with other SNPs to restore the capability of differentiating all individuals in the test data, using the dataset of 2120 unique individuals and 7666 SNPs. When choosing these additional SNPs, SNPs on chromosomes with the lowest representation were considered first to provide a more uniform distribution across the genome.

Plant material
To test the chosen SNPs with the KASP approach, a total of 168 accessions in the INRAE germplasm collection were chosen, each round of test using 95 or 94 accessions, with partial overlap of the accessions included between rounds of tests. For these accessions, only SSR data were available for the INRAE sample, but 20K SNP data were predicted to be available in corresponding samples from other collections, based on accession name. In most cases, the corresponding accession was lacking SSR data, so no confirmation of matching by MUNQ was possible. Among these accessions were 139 diploid and 23 triploid accessions across the three test runs. Besides, some accessions were from non-domestica *Malus* species, or hybrids between these species and *M. domestica*.

For a full-scale experiment, 754 accessions with no prior genotype data (752 submitted by French fruit enthusiasts and 2 accessions from the INRAE collection) were genotyped with both the finally chosen 96-SNP set (KASP technology) and the standard 16-SSR set to compare the discrimination power of the two approaches.

Genotyping
Genomic DNA was extracted from leaf tissues using the NucleoMag plant kit (Macherey Nagel,). SNP assay design followed the approach described in Petiteau et al. (2023) for apple. Master mixes, sample DNA, and primer triplets were prepared and pipetted into integrated fluidic circuit (IFC) chips (Standard BioTools Inc.). Loading and cycling was performed in a Juno system. Fluorescence was measured using a Biomark system. A total of 6 accessions were genotyped in two replicates to assess reproducibility of the genotyping. Each plate included one well with water instead of DNA as control. 95 samples which had many missing or "invalid" calls after a first run were re-genotyped after diluting their DNA. SSR genotyping was performed with the 16 SSR markers used by Urrestarazu et al (2016) and with the same protocol.

Concordance with 20K/480K data

Fluorescence levels were analyzed using the Fluidigm® SNP Genotyping Analysis Software v3 to obtain genotype calls of the form XX, XY, and YY. These genotype calls were then converted in the AA, AB, and BB forms, according to the base revealed by the first and second forward primers (A and T = A; C and G = B). Two kinds of failed genotype calls were obtained: "missing data", which resulted from too low fluorescence or points between clusters, and "invalid calls", resulting from a technical failure in loading the reaction chambers in the IFC chip. Genotype calls obtained with arrays for the 2120 individuals in the test data but also from other ongoing studies, gathered in the apple pedigree reconstruction project (Howard et al. 2018) and curated following Vanderzande et al (2019), were used to evaluate concordance between array-based and KASP genotyping. Genotypes of triploid individuals were simplified to ease comparison with genotypes obtained with the KASP approach, with AAA transformed to AA, AAB and ABB to AB, and BBB to BB. Additionally, to streamline comparison, null alleles identified in SNP array data were also simplified by changing the null allele score to that of the other allele present (e.g., A/null was converted to AA) and setting homozygous "null" genotypes to missing. During the SNP testing stage, candidate matching pairs of accessions genotyped with the KASP approach and with array data were built based on their names. We subsequently evaluated the concordance between results obtained with both methods. SNPs with more than 10 missing data or more than 10 non-concordant calls were removed. The number of non-concordant calls for candidate-matching pairs of accessions was then computed for the remaining SNPs, and the number of non-concordant calls per SNP was recomputed after removing candidate matching pairs of accessions with more than 10 non-concordant calls. Then, a graphical representation of the fluorescence levels and of the corresponding obtained and expected genotype calls for each SNP was used to further refine the SNP choice for those presenting the highest number of non-concordant calls below 10.

With the results of the full-scale experiment, each accession genotyped with the KASP approach was compared to all available unique individuals genotyped with arrays and to all other accessions genotyped with the KASP approach, to identify candidate matching pairs. Matching pairs were accepted for further evaluation when less than five non-concordant calls were observed and more than 40 SNPs could be considered for the comparison, after excluding missing data.

Error rate estimation

A total of six accessions were genotyped twice with the KASP approach. The number of non-concordant calls in the candidate matching pairs thus formed was computed using the same approach as when comparing genotypes obtained with the KASP approach to genotypes obtained with arrays.

RESULTS AND DISCUSSION

Selection of the most suitable SNPs

SNPs chosen for their capability of differentiating among 2120 unique individuals, as deduced from 20K or 480K-array data, were first tested by genotyping accessions from the INRAE germplasm collection using the KASP approach in three rounds of tests. At this SNP testing stage, a total of 182 SNPs were tested up to three times. Among these, 44 were rejected because of either too many missing data or too many non-concordant calls between KASP genotyping and SNP array genotyping, even after removing candidate matching pairs of individuals with more than 10 non-concordant calls.

Over the three rounds of test, 90% of genotype calls were obtained and the ~10% that failed were missing data for 9.1% and invalid calls for 1.4%. The SNPs that were kept for the full-scale experiment had up to 5 missing data in the first round, up to 10 missing data in the second round and up to 23 missing data or invalid calls in the third round. The threshold for keeping a SNP for the full experiment varied between rounds of tests because the number of

samples with less than 10 missing data varied: 93, 87, and 75, samples were considered in the first, second, and third round, respectively. Too many missing data, i.e. more than 10, were observed for two, seven, and 20 samples in the first, second, and third round of test, respectively. No invalid calls were obtained in the first and second round, while in the third round, invalid calls occurred in 94 SNPs and were concentrated in seven samples, with 14 to 91 invalid calls in these samples. Samples with many missing data were often from accessions of non-domestica *Malus* species, or hybrids between these species and *M. domestica*.
Thresholds were slightly more stringent regarding non-concordant calls between KASP genotyping and array genotyping, with up to three, five, and one accepted non-concordant calls in the first, second, and third round, respectively. After inspecting graphical representations of the fluorescence levels and the corresponding obtained and expected genotype calls for each SNP, 120 SNPs were found to fulfill all the criteria and 96 of them were chosen in order to build six subsets which, each, enable differentiation of the 2120 unique individuals of the test data in the absence of missing data. The 96 chosen SNPs were then used for the full-scale experiment

A robust SNP set for cultivar identification
Among the 96 SNPs that were finally chosen for the full-scale experiment, one had been tested three times, 83 had been tested twice and 12 had been tested once. As using the 96 x 96 format of the nanofluidic Integrated Fluidic Circuit (IFC) was intended, up to 96 SNP markers could be chosen to differentiate and identify all cultivars of the test set. Since the instrument is also available in a 48 x 48 format, we decided to choose six sets of 14-20 SNPs that each enables differentiation of the 2120 unique individuals of the test data in the absence of missing data. If the 48-SNP format is preferred, the combination of sets 2, 3 and 6 would be the best choice to enable also comparison with data obtained with only the 20 SNPs that initially constituted the WSU set, although cultivar identification with 48 SNPs would be more easily compromised by missing data.

Table 1. SNP set sizes.

Subset	**1**	**2**	**3**	**4**	**5**	**6**
# SNPs	16	14	14	17	15	20

The 96 finally chosen SNPs will be called the Apple-ID-SNP set. The sequences of the forward and reverse primers of these SNPs are available at https://doi.org/10.57745/HSAFLH.

Genotyping success in the full-scale experiment
In the full-scale experiment, 754 genotypes were tested with the selected 96 SNPs. Among the 95 samples that had been run twice, one at average DNA concentration and one with diluted DNA, 23 samples, were discarded either because they still had many missing or invalid data after dilution or too many discordant calls between the two runs. Additionally, one sample was discarded because it had many missing data in both KASP and SSR genotyping. Among the 736 remaining samples, depending on the SNP, 9 to 40 samples showed missing data, for a total of 1403 data points with missing data, i.e., 2% of the expected data points. The number of samples with invalid calls ranged from 0 to 27, for a total of 1063 data points with invalid calls, i.e., 1.5% of the expected data points. The choice of SNPs in the test stage was thus efficient in reducing the proportion of missing data in the full-scale experiment.

The number of SNPs that failed per sample was highly variable, with 578 samples having no failed SNP, 114 samples with 1-3 failed SNPs, five samples with 4-8 failed SNPs, 12 samples with 11-24 failed SNPs, and 27 samples with 25-79 failed SNPs. The sample with the highest number of failed SNPs was again an accession of mostly non-domestica *Malus* ancestry, 'Evereste'.

Concordance of SNP profile for replicated DNA extraction
Among the six pairs of samples that were analyzed from different DNA extracts of the same leaves, 4 pairs had no differences in their SNP profiles, except for missing or invalid calls, and 2 pairs differed for one SNP (Table 2).

Concordance of unique profiles identified using SSR and SNP
A total of 492 unique SSR profiles, corresponding to 492 MUNQ, were identified among the 754 accessions of the full-scale experiment, with various numbers of accessions (from 1 to 13) gathered per SSR profile. Among the pairs of accessions with identical SSR profiles, 497 had no differences in their SNP profiles except for missing or invalid calls, 101 differed for one SNP, 18 differed for two SNPs, and 2 differed for three or four SNPs (Table 2). In the pair with four differing SNP calls, one of the samples had 25 missing data points, among which 24 were deemed invalid calls, possibly indicating low DNA quality. In the pair with three differing SNPs, both accessions were triploid. Among the pairs of accessions that had different SSR profiles, the minimum number of differing SNPs was 3, which was the case for nine pairs of accessions (Table 2). In addition, two pairs of accessions with different SSR profiles had only 4 differing SNPs. All of these eleven pairs involved triploid accessions that shared the same unreduced gamete-donating parent (Howard et al. 2022). Additionally, one accession that had 44 invalid calls was one of the members of all these eleven pairs, which limited the number of markers for which a difference could be computed. After removing this accession, the minimum number of differing SNPs for pairs of accessions that had different SSR profiles was 5. When considering data obtained with arrays for unique individuals sharing the same SSR profile as members of these eleven pairs, there were 14 or 17 differences between individuals. Consequently, grouping accessions with essentially unique profiles using the 96-SNP data could rely on a maximum of 2 SNPs with differing genotypes, while accessions with 5 or more differences would be considered as different without doubt. In between, i.e. for pairs of accessions with 3 or 4 SNP genotype differences, it would be recommended that the KASP assay be re-run in case of a high number of missing data or invalid calls, starting at the DNA extraction step, or additional SNPs or SSR markers should be used to distinguish the accessions or confirm identicalness.

Table 2. Number of differences between KASP SNP profiles for samples from the same leaves, with the same SSR profile (same MUNQ) or with different SSR profiles (different MUNQ).

	Number of different SNPs					
	0 (identical)	1	2	3	4	5-19
Same leaves	4	2	0	0	0	0
Same MUNQ	497	101	18	1	1	0
Different MUNQ	0	0	0	9	2	637

Concordance with 20K or 480K array calls
Among the 736 samples retained for the full-scale experiment an array-based SNP profile was deemed available for 464 samples. This large proportion (63%) of array-based matching SNP profiles illustrates that the chosen SNP set aligns well with existing array-based datasets, which is an improvement over the small set developed in Winfield et al. (2020) which could only align to 480K array-based datasets. In all cases where a MUNQ had been assigned to the array-genotyped accession, there was no discrepancy between the MUNQ assigned to the KASP-genotyped accession and the MUNQ assigned to the array-genotyped accession. However, two, three, or four non-concordant calls were observed between KASP data and array data for 12, six and three accessions, respectively (Table 3). These non-concordant calls often corresponded to poor genotyping using KASP, with points in between clusters or at the border of a cluster in KASP genotyping. A MUNQ was able to be newly assigned to 14 array-genotyped accessions with SNP profile matching a profile obtained with KASP data.

Table 3. Concordance between KASP SNP profiles and array SNP profiles

	Number of different SNPs				
	0 (identical)	1	2	3	4
Matching SNP profile in database	340	103	12	6	3

CONCLUSION

The set of 96 SNPs defined in this study, called the Apple-ID-SNP set, will enable the fingerprinting of apple accessions and the comparison of the obtained results with genotypic data obtained with the 20K or the 480K arrays or with these same 96 SNPs using the KASP technology in nanofluidic Integrated Fluidic Circuit (IFC). Using this technology will also enable to cope with the expected impeding difficulty in SSR analysis. By sharing genotyping results, e.g., via a public database, collection managers may continue to compare their germplasm with others' at both the national and international levels. A proper transition from the SSR-based MUNQ assignment to a SNP-based MUNQ assignment will require the SNP genotyping with the Apple-ID-SNP set of accessions representative of MUNQ for which no array data are yet available.

ACKNOWLEDGEMENTS

This work was supported by the INVITE project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817970. SNP genotyping was performed at the Gentyane platform supported by France Génomique National infrastructure, funded as part of the "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (grant No. ANR-10-INBS-09). The involvement of CP and initial involvement of SV was supported by the USDA National Institute of Food and Agriculture Hatch project 1014919, Crop Improvement and Sustainable Production Systems (WSU reference 00011). The authors would like to thank the Biological Resource Center "RosePom - Pome Fruits and Roses" (https://www6.angers-nantes.inrae.fr/irhs/Ressources-mutualisees/Ressources-genetiques/CRB-Fruits-a-pepins-et-rosier)), INRAE, member of BRC4Plants, INRAE, 2022 (Biological Resource Centres for plants of AgroBRC-RARe; https://doi.org/10.17180/WN42-3J20), and associated staff, especially the UE HORTI, INRAE, 2018, Horticulture Experimental Facility, , for maintaining the plant material and associated datasets used in the present study.

Literature cited

Bianco L, Cestaro A, Linsmith G, et al (2016) Development and validation of the Axiom®Apple480K SNP genotyping array. Plant J 86:62–74. https://doi.org/10.1111/tpj.13145

Bianco L, Cestaro A, Sargent DJ, et al (2014) Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). PLoS ONE 9:e110377. https://doi.org/10.1371/journal.pone.0110377

Bramel P, Volk GM (2019) A GLOBAL STRATEGY FOR THE CONSERVATION AND USE OF APPLE GENETIC RESOURCES. Global Crop Diversity Trust, Bonn, Germany

Chagné D, Crowhurst RN, Troggio M, et al (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS ONE 7:e31745. https://doi.org/10.1371/journal.pone.0031745

Durel CE, Denancé C, Muranty H, et al (2023) MUNQ and PUNQ - A European and international apple and pear germplasm coding system. Acta Hortic in press

Howard NP, Albach DC, Luby JJ (2018) The identification of apple pedigree information on a large diverse set of apple germplasm and its application in apple breeding using new genetic tools. Foerdergemeinschaft Oekologischer Obstbau e. V. (FOEKO), Weinsberg, Germany

Howard NP, Micheletti D, Luby JJ, et al (2023) Pedigree reconstruction for triploid apple cultivars using single nucleotide polymorphism array data. PLANTS PEOPLE PLANET 5:98–111. https://doi.org/10.1002/ppp3.10313

Howard NP, Peace C, Silverstein KAT, et al (2021a) The use of shared haplotype length information for pedigree reconstruction in asexually propagated outbreeding crops, demonstrated for apple and sweet cherry. Hortic Res 8:1–13. https://doi.org/10.1038/s41438-021-00637-5

Howard NP, Troggio M, Durel C-E, et al (2021b) Integration of Infinium and Axiom SNP array data in the outcrossing species Malus × domestica and causes for seemingly incompatible calls. BMC Genomics 22:246. https://doi.org/10.1186/s12864-021-07565-7

Howard NP, van de Weg E, Luby JJ (2022) A new method to reconstruct the direction of parent-offspring duo relationships using SNP array data and its demonstration on ancient and modern cultivars in the outcrossing species malus × domestica. Hortic Res uhab069. https://doi.org/10.1093/hr/uhab069

Muranty H, Denancé C, Feugey L, et al (2020) Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. BMC Plant Biol 20:2. https://doi.org/10.1186/s12870-019-2171-6

Petiteau A, Denancé C, Muranty H, et al (2023) Marker-assisted breeding (MAB) on apple and pear and new approaches for QTLs and major gene genotyping involved in disease resistance. Acta Hortic 221–230. https://doi.org/10.17660/ActaHortic.2023.1362.30

Rymenants M, van de Weg E, Auwerkerken A, et al (2020) Detection of QTL for apple fruit acidity and sweetness using sensorial evaluation in multiple pedigreed full-sib families. Tree Genet Genomes 16:71. https://doi.org/10.1007/s11295-020-01466-8

Urrestarazu J, Denancé C, Ravon E, et al (2016) Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biol 16:130. https://doi.org/10.1186/s12870-016-0818-0

Vanderzande S, Howard NP, Cai L, et al (2019) High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. PLOS ONE 14:e0210928. https://doi.org/10.1371/journal.pone.0210928

Winfield M, Burridge A, Ordidge M, et al (2020) Development of a minimal KASP marker panel for distinguishing genotypes in apple collections. PLOS ONE 15:e0242940. https://doi.org/10.1371/journal.pone.0242940