The identification of apple pedigree information on a large diverse set of apple germplasm and its application in apple breeding using new genetic tools

N. P. Howard[1,2], D. C. Albach[1] and J. J. Luby[2]

Abstract

*An understanding of cultivar relatedness and population structure can be very useful information in breeding and germplasm collection and characterization efforts. Newly developed tools and techniques have made it possible to more easily identify previously unknown pedigree relations in cultivated apple (*Malus × domestica*). Recently, a large-scale international collaborative project has been initiated to discover and characterize apple germplasm pedigree relationships. For the purpose of developing organic apple cultivars, this information can be used to gain a better understanding of which old or underutilized cultivars that could be of interest to organic breeding may have desirable or undesirable alleles at important trait loci (e.g. disease resistance) based on existing literature and shared identical by descent haplotypes at underlying loci. Findings will also be used to aid germplasm collections and to provide historical information about cultivars.*

Keywords: Apple breeding, genetics, SNP array, pedigree analysis

Introduction

An understanding of variety relatedness is useful in several aspects of breeding efforts. Recently, newly developed tools and techniques have made it possible to more easily identify previously unknown pedigree relations in cultivated apple (*Malus × domestica*) using DNA markers. These tools include new SNP arrays (Chagné et al, 2012; Bianco et al, 2014; 2016), new, dense, and high quality integrated genetic maps made from SNPs on these arrays (Di Pierro et al, 2016; Howard et al, 2017), and the publication of a vastly improved genome draft (Daccord et al, 2017). Since apple cultivars are asexually propagated, extensive pedigree relationships can be identified that link modern cultivars to old and heirloom cultivars that are still in cultivation in antique orchards or preserved in germplasm collections. A recent example of this approach elucidated the extended pedigree of 'Honeycrisp' (Howard et al, 2017). Subsequent genetic analyses investigating scab resistance (Clark, 2014), the storage disorders soft scald and soggy breakdown (Howard et al, 2018), and the physiological disorder zonal leaf chlorosis (Manuscript submitted) in 'Honeycrisp' used this new pedigree information to gain an improved understanding of the genetic architecture of important traits for this cultivar.

Using the example of how the elucidation of the 'Honeycrisp' pedigree informed several subsequent genetic analyses, we have initiated an international collaborative project to identify pedigree relations within a large and diverse set of apple germplasm. The purpose of this research is to identify the progenitors of commonly grown modern apple cultivars, to provide historical information about apple cultivars, and to gain a better understanding of apple relatedness and diversity that will benefit germplasm collections and breeding programs. As this project is in the formative stages, the purpose of this contribution is to introduce the project to the wider apple genetics community, especially the methods that will

---
[1] Nicholas P. Howard and Dirk C. Albach, Institute for Biology and Environmental Science, Carl von Ossietzky Universität, Oldenburg, Germany

[2] Nicholas P. Howard and James J. Luby, Department of Horticultural Science, University of Minnesota, Saint Paul, USA

be employed, and to briefly discuss some of the ways in which the findings from this study will be used. For the purposes of this conference and for developing organic apple cultivars, findings from this study can be used to gain a better understanding of which old or underutilized cultivars that could be of interest to organic breeding may have desirable or undesirable alleles at important trait loci (e.g. disease resistance) based on existing literature and shared identical by descent haplotypes at underlying loci. This information will also be used to help avoid inbreeding and preserve a broad genetic basis for breeding in apple.

Materials and Methods

A diverse set of apple material was collected in 2017 from several germplasm collections in the US and Europe. This collection effort aimed to encompass the most commercially relevant cultivars, their known and likely relatives, historically relevant cultivars, and a broad array of heirloom cultivars. Other research institutes that have additional genetic data on apple SNP arrays were contacted regarding collaborating on this project. Their data has been or will be combined with existing data we have collected. SSR data was available for some germplasm and was used to aid in collection efforts, particularly to ensure likely pedigree relation candidates were collected and to avoid genotyping of genetic duplicates within collections.

All samples have been or will be run on the Illumina Infinium® 20K SNP array (Bianco et al, 2014), with the exception of some shared material that will be run on the Affymetrix Axiom® 480K SNP array (Bianco et al, 2016). The 480K SNP array includes the SNPs from the 20K SNP array. ASSIsT (Di Guardo et al, 2015) and Genomestudio™ software will be used to call SNP alleles and a SNP data curation protocol developed through the RosBREED project (manuscript in preparation) will be employed to create a highly curated marker dataset to ensure SNP calls are accurate. Parent-offspring relationships will be identified by evaluating Mendelian-inconsistent errors in comparisons between cultivars (similar to the process described in Howard et al, 2017). Following the identification of parent-offspring relationships, marker data is phased with FlexQTL™ software and shared extended haplotype length statistics will be generated between comparisons of phased marker data for cultivars to identify further pedigree relationships. Segregation data, scientific literature, and historical documents will be consulted to help order and validate pedigree relationships. These relationships will then be organized, documented, and shared with the research community. The pedigree information and SNP dataset will then be used for subsequent analyses alongside existing literature on marker-trait associations to gain further insights into the ancestral sources of important alleles and which cultivars possess specific functional alleles at important trait loci.

Results

We have devised a pipeline to identify pedigree relationships within a large and diverse set of apple germplasm. Implementation of an early version of this pipeline led to a publication describing the elucidation of the extended pedigree of 'Honeycrisp' (Howard et al, 2017). These initial findings, coupled with successful testing of advanced methodologies on a set of available marker data from material in the RosBREED project, have provided the foundation for discussions with potential collaborators, as well as provided for confidence in the methodologies used. This project has grown to include many collaborators. The following groups are contributing germplasm: Seed Savers in Iowa, USA, the USDA apple collection in Geneva, the Temperate Orchard Conservancy in Oregon, USA, Ökowerk in Emden, Germany, the National Fruit Collection in Brogdale, UK. The following collaborators are contributing additional germplasm, marker data, and/or technical assistance: Carl von

Ossietzky Universität in Germany, the University of Minnesota in Minnesota, USA, the Fondazione Edmund Mach in Italy, the Institut national de la recherché agronomique (INRA) in France, Wageningen University and Research in The Netherlands, Washington State University in Washington, USA. At the time of this writing, the total number of samples either collected or pledged for which SNP data will be available is over 3000. Analyses on this material will begin in early 2018 when all of the SNP data becomes available.

Discussion

This project is in its early stages, but the initial findings and publications have been very promising for its prospective outcomes. The elucidation of the extended pedigree of 'Honeycrisp' was a useful finding itself, and it has informed interpretations of several marker-trait association studies showing how this data may be employed to characterize germplasm for breeding purposes. For example, a major locus for the post-harvest storage disorders soft scald and soggy breakdown was identified and 'Honeycrisp' was found to have two copies of an allele associated with increased incidence for these disorders (Howard et al, 2018). 'Honeycrisp' inherited these alleles from its maternal grandparent 'Northern Spy' and its paternal grandparent 'Golden Delicious'. These grandparents have a previously unidentified pedigree relationship other than parent-offspring and grandparent-grandchild that will be further characterized when the 20K SNP dataset becomes available. Other cultivars evaluated in this study possess this same allele, which demonstrated a similar detrimental effect on incidences of soft scald and soggy breakdown in offspring of these cultivars. Most of these cultivars inherited the allele through the same ancestors as 'Honeycrisp'. In another example, apple scab (*Venturia inaequalis*) resistance in 'Honeycrisp' was initially considered to be inherited through grandparent 'Frostbite', but new pedigree information demonstrated that the likely source of the resistance allele was from the previously unidentified grandparent of 'Frostbite', 'Duchess of Oldenburg' (Clark, 2014). This finding led to the identification of the same allele in other cultivars (ex. in 'Wildung', through parent 'Sharon', which has 'Duchess of Oldenburg' as one grandparent) and breeding selections in the UMN apple breeding program, all of which could trace the allele back to 'Duchess of Oldenburg' through previously unknown pedigree relationships. In a third example, 'Duchess of Oldenburg' was identified as the likely genetic source of zonal leaf chlorosis disorder in 'Honeycrisp' (Manuscript in preparation). In a similar fashion to these illustrative examples, new pedigree information identified from this present large-scale extended pedigree identification study will be used to characterize old or underutilized cultivars for allelic composition at important trait loci based on comparing existing literature and shared identical by descent haplotypes at these loci.

Finally, the collaborative nature of this project has resulted in involvement from several large germplasm collection organizations. They will benefit from learning whether their collections contain duplicate entries within and between germplasm collections and also whether their varieties are likely true-to-type based on pedigree identification. These results will help ensure that germplasm resources are accurately characterized and efficiently curated for user community of breeders and producers.

Acknowledgements

Durel, Hélène Muranty, and Caroline Denance from the Institut National de la Recherché Agronomique (INRA) in France, Detlef Stang from Ökowerk, Germany, Eric van de Weg from Wageningen University and Research, The Netherlands, and the Apfel:gut project in Germany.

References

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troggio M (2014). Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica* Borkh). PLoS One **9**(10):e110377.

Bianco L, Cestaro A, Linsmith G, Muranty H, Denancé C, Théron A, Poncet C, Micheletti D, Kerschbamer E, di Pierro EA, Larger S, Pindo M, van de Weg E, Davassi A, Laurens F, Velasco R, Durel CE, Troggio M (2016). Development and validation of the Axiom® Apple 480K SNP genotyping array. Plant J **86**(1):62–74.

Clark, Matthew Daniel. (2014). Characterizing the host response and genetic control in 'Honeycrisp' to apple scab (*Venturia inaequalis*). Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/172140.

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, van de Weg E, Gardiner SE, Bassil N, Peace C (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One **7**(2):e31745.

Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Muranty H, Micheletti D, Di Pierro EA, Gouzy J, Gaillard S, Velasco R, Durel CE, Rees DJG, Laurens F, Lespinasse Y, Quesneville H, Aubourg S, Weigel D, van de Weg E, Troggio M, Bucher E (2017). The apple genome: evolution and methylome dynamics of early fruit development. Nat Genet **49**(7):1099–1106.

Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H. J., Longhi, S., Costa, F., ... & van de Weg, E. W. (2015). ASSIsT: an automatic SNP scoring tool for in-and outbreeding species. Bioinformatics, **31**(23), 3873-3874.

Di Pierro AE, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJJ, Kruisselbrink JW, Longhi S et al (2016). A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res. **3**:16057.

Howard NP, van de Weg E, Bedford DS, Peace CP, Vanderzande S, Clark MD, Teh SL, Cai L, Luby JJ (2017). Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus × domestica*) pedigree-connected SNP data set. Hortic Res **4**:17003.

Howard, N. P., van de Weg, E., Tillman, J., Tong, C. B., Silverstein, K. A., & Luby, J. J. (2018). Two QTL characterized for soft scald and soggy breakdown in apple (*Malus × domestica*) through pedigree-based analysis of a large population of interconnected families. Tree Genom Genet, **14**(1), 2.