Building international partnerships for the collation of historical data to study the environmental stability of genomic predictions in sweet cherry

C. Hardner[1], C. Peace[2], J. Quero-Garcia[3], S. Vanderzande[2], D. Giovannini[4], T. Barreneche[3], J.A. Campoy[3], G. Charlot[5], A. Liverani[4], N. Oraguzie[6], M. Villamil-Castro[1], L. Cai[7], M. Ordidge[8], A. Wünsch[9], T. Sotiropoulos[10], K. Kazantzis[10], A. Patocchi[11], S. Perren[11], D. Graetz[12] and A. Iezzoni[7]

[1]University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St. Lucia, Australia; [2]Department of Horticulture, Washington State University (WSU), Pullman, USA; [3]UMR 1332 BFP, Institut National de la Recherche Agronomique (INRA), Université Bordeaux, Villenave d'Ornon, France; [4]Council for Agricultural Research and Economics (CREA), Research Centre for Olive, Citrus and Tree Fruit, Forlì, Italy; [5]Centre Technique Interprofessionnel des Fruits et Légumes (Ctifl), Bellegarde, France; [6]Irrigated Agr. Res. & Extens. Ctr., Washington State University, Prosser, USA; [7]Department of Horticulture, Michigan State University, East Lansing, USA; [8]School of Agriculture, Policy and Development, University of Reading, Whiteknights Reading, United Kingdom; [9]Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Instituto Agroalimentario de Aragón – IA2 (CITA-Universidad de Zaragoza), Zaragoza, Spain; [10]Hellenic Agricultural Organization 'Demeter' (HAO), Institute of Plant Breeding and Plant Genetic Resources, Department of Deciduous Fruit Trees in Naoussa, Naoussa, Greece; [11]Agroscope Research Station, Wädenswil, Switzerland; [12]South Australian Research and Development Institute (SARDI), Plant Research Centre, Adelaide, Australia.

Abstract

Here we describe progress on the development of a RosBREED-led international collaboration to collate data on sweet cherry individuals (breeding progenies, selections, and cultivars) that have been phenotyped for traits of interest (maturity date, fruit size, firmness, acidity, and sweetness) and genotyped using the Illumina Infinium 6K cherry SNP array. Our hypotheses are that: i) individuals evaluated at various sites can be utilised to sample the phenotypic response of genome-wide alleles to varying environments, and ii) SNP array data can be used to characterise variation in QTL alleles as well as minor-effect alleles (genetic background) such that data from these different environments can be combined into a multi-environment analysis. This approach is an extension of genomic selection methods and exploits replication among environments at the allelic level, such that clonal replication of individuals among environments is not required. Collaborators have provided access to SNP array data for 762 unique individuals and phenotypic data for these individuals from 19 locations (1 USA, 13 France, 1 Italy, 1 United Kingdom, 1 Switzerland, 1 Greece, 1 Australia). The data were used in genome-wide selection models to demonstrate that rankings of individuals for fruit maturity among 2 locations in France, 1 in Italy, and 1 in the USA are very stable, and thus performance of individuals tested in only a single location can be predicted across other locations (e.g., USA to Europe). We will genotype several hundred additional individuals in 2017, and welcome additional partners to help widen the extent of environmental variation sampled. The larger the phenotypic dataset from environments and individuals, the greater the opportunity to identify genotype-by-environment patterns of QTL and polygenetic effects, generate and test hypotheses of factors driving those patterns, and predict the performance of individuals in multiple environments. We invite additional collaborators to join this international effort.

Keywords: genomic prediction, genotype-by-environment interaction, RosBREED, QTL

INTRODUCTION

Sweet cherry is a widely cultivated premium tree fruit crop (Bujdosó and Hrotkó, 2017) that is produced across a large number of countries. Turkey, USA, Iran and Italy, Spain,

Chile, Ukraine, and the Russian Federation currently produce the largest volumes, with smaller amounts from France, Australia, Greece and other regions in Europe, Asia, and Canada.

Use of improved germplasm that reduces production costs, increases consumer demand, and supports sustainable production methods is an important strategy for industry profitability and growth. Cherry breeding probably started around the early 1800s in its native range in Europe (Quero-García et al., 2017; Iezzoni et al., 2017) by sampling open-pollinated seed from preferred cultivars. Crossing of selected parents commenced around the end of the 19th century/early 20th century in the USA and Canada followed by the United Kingdom, Russia, and Ukraine, with expansion to other European countries around the mid-20th century. In Australia, breeding commenced in the early 1980s, and around the turn of the 21st century in Chile.

Use of DNA-based tools to support breeding have gained momentum in the early 21st century, beginning with the use of DNA tests for alleles at the self-compatibility locus (reviewed in Herrero et al., 2017). More recently, quantitative trait loci (QTLs, i.e., genomic regions associated with large phenotypic effects) were identified for fruit colour (Sooriyapathirana et al., 2010), fruit maturity (Dirlewanger et al., 2012), bloom time (Castède et al., 2014), fruit size and firmness (Campoy et al., 2015) and other important traits in bi-parental families, and DNA tests subsequently developed (e.g., Sandefur et al., 2016). In the last 5 years, a 6K SNP array for cherry has been developed (Peace et al., 2012) and is being extended to include an additional 9K (Vanderzande, pers. commun.). These arrays are being used to support QTL discovery and explore opportunities for genomic prediction of background genetic effects. Much of this work has been supported through the large USDA-funded RosBREED project initiated in 2009 (Iezzoni et al., 2016, 2010) and continued in 2014 (Peace, 2017).

Interaction of genetic performance with environment (G×E) might be an important issue in breeding (Allard and Bradshaw, 1964), but there is limited information on G×E for sweet cherry traits. Multi-environment trials (METs), which are commonly used to detect G×E, can be expensive, particularly in tree crops such as cherry due to large size of each experimental unit, long juvenile period, and annual fluctuations in production. Lack of knowledge of patterns in G×E can reduce the confidence in predicted performance of introduced germplasm.

Here we describe progress on the development of a RosBREED-led international collaboration to collate historical data collected from local germplasm-testing programs for individuals that have also been genotyped with the latest SNP array technology. Our hypotheses are: i) that data on individuals evaluated in a particular environment (in the general sense, e.g., climate, weather, soil, management) can be utilised as a sample of the phenotypic response of genome-wide alleles to that environment, and ii) genome-wide markers can be used to characterise variation in QTL alleles as well as minor-effect alleles (genetic background) such that data from different environments can be combined into a single analysis.

METHODS

Potential collaborators were initially identified through established networks and published literature. A letter was sent to each potential collaborator outlining the proposed analysis, the process for the management of intellectual property (IP), and benefits to collaboration as well as requesting access to potentially appropriate data.

Characteristics of data requested from potential collaborators were:

i) at least 20 non-interspecific cultivars, selections, and/or multiple progenies at a single location;
ii) assessed for at least one of the traits of interest (fruit maturity, fruit size, firmness, sweetness, and powdery mildew resistance); and
iii) genotyped with the Illumina cherry 6K SNP array.

Protection of collaborators' IP was recognised as a fundamental issue to support contribution of data. The name of any commercially sensitive individuals was recoded.

Pedigree information was only used for data management, was not required in the analysis, and was not essential for data interpretation. Use of contributed data was restricted to the proposed data analysis, and where this overlapped with specific interests (such as diversity analyses) publication agreements were put in place. Data were only used by members of the immediate analysis team and are not shared. Contributors were offered the opportunity to participate in co-authorship of any publications in which their data are used.

A data curation pipeline was developed to manage contributed genotypic and phenotypic data (Figure 1). Initially, the format of names of individuals used in each dataset were standardised for matching across sources. Names were converted to lower case, with no special characters, and underscore used for spaces. Formatted names were then compared to a dictionary of existing original names. Potential synonyms were identified using full and partial matching. New unreconciled names were added to the dictionary.

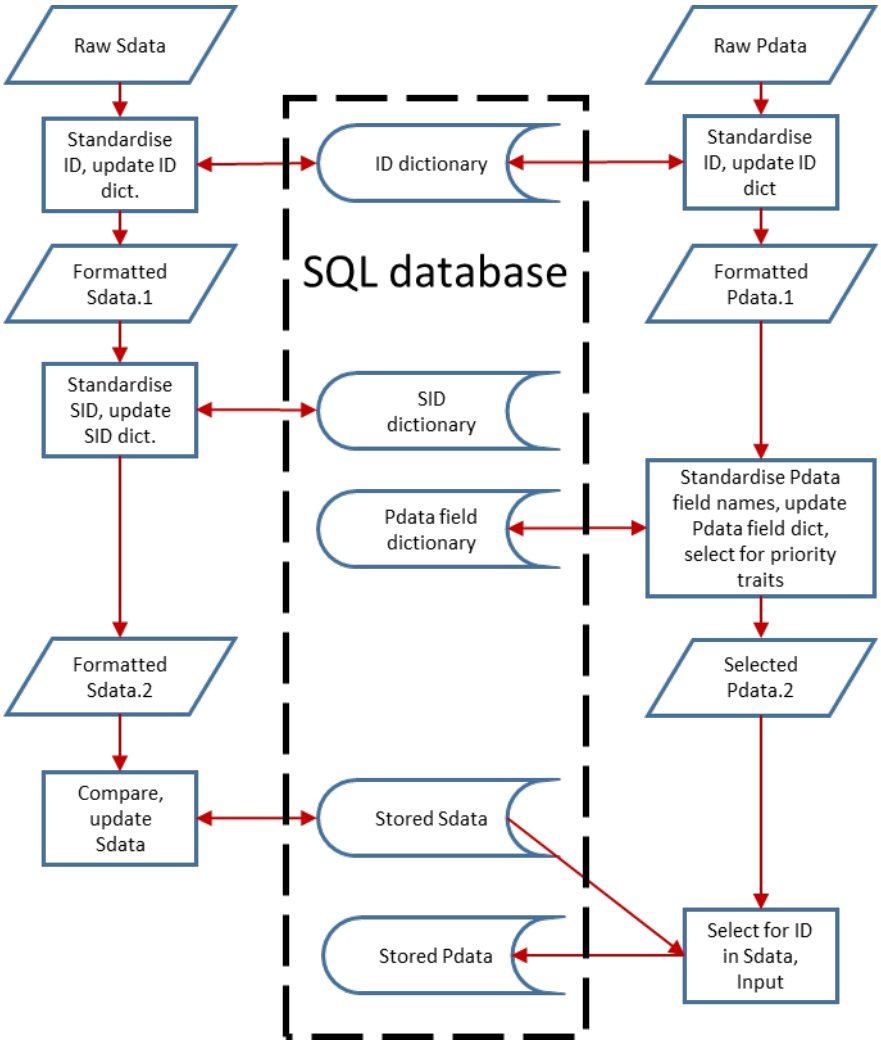

Figure 1. Process for management of genotypic and phenotypic data contributed by multiple sources. Parallelograms represent .csv files. ID refers to names of individuals and SID refers to NCBI SNP names. Pdata refers to phenotypic data and Sdata refers to genotypic (SNP array) data.

Following standardisation of individuals' names in the SNP data files, names of SNP loci in contributed data files were verified against the NCBI database. Where individuals were genotyped by more than one source, genotypic information from each source was compared to ensure SNP genotypes were consistent across the sources and to fill in missing

genotypic calls. Where there was a low frequency (i.e., 1-5 loci per sample) of inconsistent genotype calls among sources for the same individual, calls were set as missing (NA, not available).

As with the curation of SNP data, the first stage of curation of phenotypic data involved standardisation of names of individuals. Data files were then reorganised into a long format, where each record represented an individual observation. Original field names were used to identify experimental and sampling design factors and names were standardised, unless new factors were identified. In particular, a field representing "experimental unit" was created to identify the smallest unit to which a unique random genetic treatment was applied, as discussed in Brien and Demetrio (1998).

RESULTS

To date, we have collated genotypic data from the 6K Illumina cherry array for 762 individuals (Table 1) contributed by CITA, CREA, INRA, and WSU (see author affiliation for details). The genome-wide SNP data of 28 individuals genotyped by more than one source were identical. However, comparison of SNP data identified that an individual named "Noire de Meched" genotyped by INRA was different from an individual with the same name at CREA.

Table 1. Number of genotyped individuals by each (diagonal) and multiple (of-diagonal) contributors[1].

Contributor	Contributor				
	CITA	CREA	INRA	WSU	WBP
CITA	3	0	0	0	0
CREA		47	9	3	0
INRA			198	20	0
WSU				66	0
WBP					478

[1] All individuals are selections or cultivars except for WBP which are seedling offspring from the Washington State University sweet cherry breeding program.

Phenotypic data on fruit maturity timing, fruit size, firmness, and sweetness were contributed by eight sources, RosBREED/WSU, INRA, CREA (Italy), Ctifl (France), University of Reading (UK), HAO (Greece), Agroscope (Switzerland), and SARDI (Australia) for a total of 903 individuals evaluated across 19 locations (Table 2), of which 606 had been already genotyped.

DISCUSSION

This project has successfully developed an international collaboration to collate performance data for different sets of sweet cherry individuals grown at various locations. Individuals have also been SNP-genotyped, or will be. Key components of this activity are the development of trust among collaborators through developing and outlining IP strategies (Cifuentes Jara et al., 2015), involvement of all in publication of results derived from these data, and communicating benefits to collaborators. The RosBREED project is also collating similar data for a similar type of analysis in apple (focusing on the traits of fruit firmness, crispness, sweetness, acidity, and storage disorders), peach (fruit maturity time, texture type, firmness, acidity, sweetness, and bacterial spot incidence), and strawberry (fruit yield, cull proportion, size, firmness, sweetness, and resistance to powdery mildew).

One of the initial benefits from this project has been the elucidation of cryptic cultivar identities. Comparing genotypic data, we have demonstrated that 'Noire de Meched' in France is different from the individual with the same name in Italy. Genotypic data have also been used to identify potential synonyms.

Table 2. Numbers of individuals phenotyped at 19 locations for four traits and number of individuals genotyped (Sdata) with 6K cherry SNP-array.

Location[1]		Data								Total[4]	
		Trait[2] / Sdata[3]									
		FM/ Avail	FM/ NA	Firm/ Avail	Firm/ NA	Size/ Avail	Size/ NA	Sweet/ Avail	Sweet/ NA	Avail	NA
Lenswood	AU					38	52	38	50	38	68
Wadenswil	CH	10	13	10	13	10	13	10	13	10	13
Balandran	FR	115	7			114	7			116	7
Bordeaux	FR	123	10			123	9			123	10
Bourran	FR	193				193				193	
Cefel	FR					19				19	
Centrex	FR	12				12				12	
Etoile	FR	99	5			99	5			101	5
La Morinière	FR	15				15				15	
La Tapy	FR	31				39				39	
Sefra	FR	13				13				13	
Sefra_bozas07	FR	9				16				16	
Serfel	FR	24				24				24	
Sefra_SLdA69	FR					19				19	
Verexal	FR	12				12				12	
Naoussa	GR	30	1			50	1	34	1	50	1
Forli	IT	59	1			59	1			59	1
Brogdale	UK	41	223			41	223			41	223
Prosser	US	374	15							374	15
Total[4]		595	244	10	13	233	282	59	60	606	297

[1]AU: Australia, CH: Switzerland, FR: France, GR: Greece, IT: Italy, UK: United Kingdom, US: United States of America.
[2]FM = date of fruit maturity, firm = fruit firmness, size = fruit size, sweet = fruit sweetness.
[3]Avail = SNP data available, NA = SNP data not available.
[4]Note numbers are not additive as some individuals are planted at more than one location.

The primary purpose of the collation of genotypic and phenotypic data is for use in analyses to predict stability of major-effect QTLs and minor-effect background alleles among environments. Initial genomic prediction analyses undertaken with a subset of these data demonstrated that there was little interaction between genome-wide effects and environments at Prosser (USA), Bourran (France), Baladran (France), and Forli (Italy) for fruit maturity timing (Hardner et al., in prep.), demonstrating that performance of individuals in an environment in which they have not yet been tested may be predicted by leveraging the genetic correlation among relatives that has been assessed in that environment.

Outputs of this work can ultimately be used to improve the efficiency of breeding programs. Identifying groups of environments among which individuals are highly correlated is expected to increase confidence in the deployment of elite cultivars to untested environments and introduction of untested germplasm from exotic environments. In addition, the ability to predict performance using only genomic information can be used to implement marker-assisted selection.

Development of an expanded Illumina SNP array for cherry (Vanderzande et al., pers. commun.) is supporting the genotyping of at least another 672 cultivars, selections, and progenies. It is expected that these individuals will be included in future analyses, particularly if phenotypic data for any of the targeted traits are available. Involvement of additional collaborators is welcomed.

ACKNOWLEDGEMENTS

This work was supported by the USDA National Institute of Food and Agriculture – Specialty Crop Research Initiative project, "RosBREED: Combining disease resistance with horticultural quality in new rosaceous cultivars" (2014-51181-22378); Ministerio de Economía y Competitividad research project RTA2015-00027-00-00 and FEDER Funds. Collection of phenotypic data was supported by Defra, United Kingdom.

Literature cited

Allard, R.W., and Bradshaw, A.D. (1964). Implications of genotype-environment interaction in applied plant breeding. Crop Sci. *4* (*5*), 503–508 https://doi.org/10.2135/cropsci1964.0011183X000400050021x.

Brien, C.J., and Demetrio, C.G.B. (1998). Using the randomisation in specifying the ANOVA model and table for properly and improperly replicated grazing trials. Aust. J. Exp. Agric. *38* (*4*), 325–334 https://doi.org/10.1071/EA97046.

Bujdosó, G., and Hrotkó, K. (2017). Cherry production. In Cherries: Botany, Production and Uses, J. Quero-García, A. Iezzoni, J. Pulawska, and G. Lang, eds. (Wallingford, Oxon, UK: CAB International), p.1–13 https://doi.org/10.1079/9781780648378.0001.

Campoy, J.A., Le Dantec, L., Barreneche, T., Dirlewanger, E., and Quero-Garcia, J. (2015). New insights into fruit firmness and weight control in sweet cherry. Plant Mol. Biol. Report. *33* (*4*), 783–796 https://doi.org/10.1007/s11105-014-0773-6.

Castède, S., Campoy, J.A., García, J.Q., Le Dantec, L., Lafargue, M., Barreneche, T., Wenden, B., and Dirlewanger, E. (2014). Genetic determinism of phenological traits highly affected by climate change in *Prunus avium*: flowering date dissected into chilling and heat requirements. New Phytol. *202* (*2*), 703–715 https://doi.org/10.1111/nph.12658. PubMed

Cifuentes Jara, M., Henry, M., Réjou Méchain, M., Lopez, O.R., Wayson, C., Michel Fuentes, J.M., Castellanos, E., Zapata-Cuartas, M., Piotto, D., Alice Guier, F., et al. (2015). Overcoming obstacles to sharing data on tree allometric equations. Ann. For. Sci. *72* (*6*), 789–794 https://doi.org/10.1007/s13595-015-0467-8.

Dirlewanger, E., Quero-García, J., Le Dantec, L., Lambert, P., Ruiz, D., Dondini, L., Illa, E., Quilot-Turion, B., Audergon, J.M., Tartarini, S., et al. (2012). Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity (Edinb.) *109* (*5*), 280–292 https://doi.org/10.1038/hdy.2012.38. PubMed

Hardner, C., Hayes, B.J., Kumar, S., Vanderzande, S., Cai, L., Piaskowsi, J., Quero-Garcia, J., Campoy, J.A., Barreneche, T., Giovannini, D., Liverani, A., Charlot, G., Villamil-Castro, M., Oraguzie, N., and Peace, C. (in prep.). Prediction of genetic value for sweet cherry fruit maturity across environments using a 6K SNP array.

Herrero, M., Rodrigo, J., and Wünsch, A. (2017). Flowering fruit set and development. In Cherries: Botany, Production and Uses, J. Quero-García, A. Iezzoni, J. Pulawska, and G. Lang, eds. (Wallingford, Oxon, UK: CAB International), p.14–35 https://doi.org/10.1079/9781780648378.0014.

Iezzoni, A., Weebadde, C., Luby, J., Chengyan Yue, van de Weg, E., Fazio, G., Main, D., Peace, C.P., Bassil, N.V., and McFerson, J. (2010). RosBREED: enabling marker-assisted breeding in *Rosaceae.* Acta Hortic. *859*, 389–394 https://doi.org/10.17660/ActaHortic.2010.859.47.

Iezzoni, A., Weebadde, C., Peace, C., Main, D., Bassil, N.V., Coe, M., Fazio, G., Gallardo, K., Gasic, K., Luby, J., et al. (2016). Where are we now as we merge genomics into plant breeding and what are our limitations? Experiences from RosBREED. Acta Hortic. *1117*, 1–6 https://doi.org/10.17660/ActaHortic.2016.1117.1.

Iezzoni, A., Wünsch, A., Höfer, M., Giovannini, D., Jensen, M., Quero-Garcia, J., Campoy, J.-A., Vokurka, A., and Barreneche, T. (2017). Biodiversity, germplasm resources and breeding methods. In Cherries: Botany, Production and Uses, J. Quero-García, A. Iezzoni, J. Pulawska, and G. Lang, eds. (Wallingford, Oxon, UK: CAB International), p.36–59 https://doi.org/10.1079/9781780648378.0036.

Peace, C.P. (2017). DNA-informed breeding of rosaceous crops: promises, progress and prospects. Hortic Res *4*, 17006 https://doi.org/10.1038/hortres.2017.6. PubMed

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U.R., Stegmeir, T., Sebolt, A., Gilmore, B., Lawley, C., Mockler, T.C., et al. (2012). Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS ONE *7* (*12*), e48305 https://doi.org/10.1371/journal.pone.0048305. PubMed

Quero-García, J., Schuster, M., Lopez-Ortega, G., and Charlot, G. (2017). Sweet cherry varieties and improvement. In Cherries: Botany, Production and Uses, J. Quero-García, A. Iezzoni, J. Pulawska, and G. Lang, eds. (Wallingford, Oxon, UK: CAB International), p.60–94 https://doi.org/10.1079/9781780648378.0060.

Sandefur, P., Oraguzie, N., and Peace, C. (2016). A DNA test for routine prediction in breeding of sweet cherry fruit

color, Pav-R-f-SSR. Mol. Breed. *36* (*33*), 11 https://doi.org/10.1007/s11032-016-0458-y.

Sooriyapathirana, S.S., Khan, A., Sebolt, A.M., Wang, D.C., Bushakra, J.M., Lin-Wang, K., Allan, A.C., Gardiner, S.E., Chagne, D., and Iezzoni, A.F. (2010). QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). Tree Genet. Genomes *6* (*6*), 821–832 https://doi.org/10.1007/s11295-010-0294-x.