Exploring and exploiting phenotypic and genetic diversity in peach: identification of major genes and QTLs by GWAS

M.J. Aranzana[1], J.R. Hernández Mora[1], D. Micheletti[2], S. Micali[3], N. Nazzicari[4], I. Pacheco[5], S. Foschi[6], T. Barreneche[7], B. Quilot-Turion[8], L. Wang[9], R. Ma[10], X. Li[11], I. Iglesias[12], J. Carbó[13], M. Troggio[2], E. Banchi[3], V. Aramini[3], M.T. Dettori[3], A. Caprera[4], C. da Silva Linge[5], T. Pascal[8], P. Lambert[8], Z. Gao[11], I. Verde[3], D. Bassi[5], L. Rossini[4,5], F. Laurens[14] and P. Arús[1]

[1]IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra (Cerdanyola del Vallès), 08193 Barcelona, Spain; [2]Research and Innovation Centre, Fondazione Edmund Mach (FEM), Via Mach 1, 38010 San Michele all'Adige (TN), Italy; [3]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CRA), Centro di Ricerca per la Frutticoltura, 00134 Roma, Italy; [4]Parco Tecnologico Padano, Via Einstein, 26900 Lodi, Italy; [5]Università degli Studi di Milano, DiSAA, Via Celoria 2, 20133 Milan, Italy; [6]Centro Ricerche Produzioni Vegetali, Via dell'Arrigoni 120, 47522 Cesena (FC), Italy; [7]INRA, UMR 1332 de Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France; [8]INRA UR1052 GAFL, Domaine Saint Maurice, 67 Allée des Chênes, CS60094, 84143 Montfavet, France; [9]Zhenzhou Fruit Research Institute, CAAS, 450009 Zhengzhou, China; [10]Horticultural Institute, Jiangsu Academy of Agricultural Sciences, Zhong-Lin Street 50, Nanjing, 210014, China; [11]Department of Horticulture, Zhejiang University, 310058 Hangzhou, China; [12]IRTA, Estació Experimental de Lleida, Parc de Gardeny, Edifici Fruitcentre, 25003 Lleida, Spain; [13]IRTA, Estació Experimental Mas Badi, La Tallada d'Empordà, 17134 Girona, Spain; [14]INRA, UMR 1345 Institut de Recherche en Horticulture et Semences, F49071 Beaucouzé, France.

Abstract

Genetic variability is a key requirement for breeding. Although new peach cultivars are released yearly to the market, the genetic pool of cultivated peaches is very limited. To evaluate the variability available in commercial but also in old local peach accessions we selected a panel of 1,580 accessions maintained and evaluated in four European and one Chinese germplasm collections. Phenotypic data collected over years following common protocols have been integrated in a database, generating a useful tool for breeders and researchers. These accessions were genotyped with the peach 9K SNP array v1. T. Genotypic data distributed the accessions in three main subpopulations (Occidental obtained in breeding programs, Occidental old local varieties and Chinese cultivars). Linkage disequilibrium (LD) was in agreement with previous studies reporting long extension. Phenotypic and genotypic data have been combined in a GWAS study allowing the design of markers for marker assisted selection (MAS). Preliminary analyses on quantitative traits are promising, while further analysis will be required to integrate all data in a single genome-wide association analysis.

Keywords: peach, SNP, population structure, genotyping

INTRODUCTION

Genetic variability is crucial for breeding gain and innovation. However peach accessions, especially those released from European and North American breeding programs, share an impoverished genetic pool (Scorza et al., 1985; Byrne, 1990; Aranzana et al., 2003) produced, in part, by the use of a reduced set of parents in the initial breeding programs. Diversity studies of cultivars adapted and grown in different regions of Africa, America, Europe and Asia have revealed a differential distribution of the variability, which correlates with the geographic location (Badenes et al., 2015). Most of this variability is contained in Chinese accessions (Li et al., 2013; Cao et al., 2014; Badenes et al., 2015).

In the frame of the European project FruitBreedomics we have studied a large panel of peach varieties, containing landraces as well as improved cultivars obtained in breeding programs. The objectives of the work presented here are i) to explore the genetic variability contained in European as well as in Chinese orchards and repositories and ii) conduct a

genome wide association analysis to identify and design molecular markers to be applied for marker assisted selection (MAS).

MATERIALS AND METHODS

Plant material and phenotypic information

We selected a panel of 1,580 accessions of *Prunus* from five germplasm collections mantained in four countries: Centro di Ricerca per la Frutticoltura (CRA-FRU, Roma); National Institute of Agronomic Research (INRA) of Avignon (France) and Bordeaux (maintained in the Prunus Genetic Resources Center, France); Institut de Recerca i Tecnologia Agroalimentàries (IRTA) stations of La Tallada d'Empordà and Gimenells (Spain); University of Milan together with CRPV (Italy); and Zhejiang University (Hangzhou, China). Phenotypic data for qualitative and quantitative traits measured over years in the respective locations were put together in the FruitBreedomics database developed for this purpose. Additionally, flowering date (2-3% flowers in stage F), sugar content (as average Brix degrees in 10 fruits), titrable acidity (meq 100 mL^{-1} in the juice of 10 fruits) and fruit weight (average on 10 homogenious fruits) were measured also in 2012 and 2013.

DNA extraction and SNP genotyping

Genomic DNA was extracted from young leaves using the DNeasy 96 Plant Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Accessions were genotyped using the International Peach SNP Consortium (IPSC) peach 9K SNP array v1 (Verde et al., 2012). SNP genotypes were scored with the Genotyping Module of the GenomeStudio Data Analysis software (Illumina, Inc.) using the default parameters and filtered for quality. SNPs considered for the analyses were those with GenTrain higher than 0.4, GeneCall 10% higher than 0.2 and identifying all three possible genotypes (AA, AB, BB) with less than 5% of No Call (failed genotyping).

Population structure and LD analysis

Population structure was studied by PCA analysis using the "prcomp" function of R and with the Structure v.2 (Pritchard et al., 2000) software, run under the admixture model assumption with correlated allele frequencies, using 100,000 interactions after a burn-in of 10,000 for a value of K ranging from 2 to 20. Linkage disequilibrium (LD) was calculated using the squared correlation based on genotypic allele counts as implemented in PLINK (Purcell et al., 2007).

Genome-wide association analysis

Population structure and kinship (K) matrix were computed using the Flapjack software (Milne et al., 2010). Association mapping was performed using the GLM and MLM procedure in TASSEL (Bradbury et al., 2007).

RESULTS AND DISCUSSION

Genetic diversity

The panel of 1,580 accessions was genotyped with the IPSC peach 9K SNP array v1 (Verde et al., 2012). Only those polymorphisms meeting the quality parameters previously described in the materials and methods section were used for the analysis (4,271 SNPs, 52.4%), corresponding to an average of 1 SNP every 53.8 kb (for a peach genome size of 230 Mbp) and approximately equivalent to 8.2 SNPs cM^{-1} of the reference *Prunus* map (519 cM as in Dirlewanger et al., 2004). Although after filtering for quality the number of SNPs (and consequently its density) was reduced approximately to its half, it should be still adequate for genome-wide association analysis accordingly with the high linkage disequilibrium (LD) reported for peach (Aranzana et al., 2010; Li et al., 2013).

A total of 1,240 unique accessions were analyzed with the 4,271 SNPs. The observed mean heterozygosity (H_o) $individual^{-1}$ was 0.286, ranging from 0.003 to 0.68. SNP markers

are less variable than SSRs, which explain the lower heterozygosity detected by SNPs in comparison with the one reported with 48 SSRs (Ho=0.44) in previous studies including also a large panel of peaches (Li et al., 2013). Here only 10.8% of accessions showed more than 44% of loci heterozygous. Similarly, the average observed heterozygosity per SNP (0.286) was lower than that for SSR (0.47). Most of the markers showed deviation from Hardy-Weinberg equilibrium, with an average expected heterozygosity (H_e=0.39) higher than the one observed.

SNP data were used to study the pattern of LD and the division of the accession panel into subpopulations. As previously reported in literature (Aranzana et al., 2003, 2010), accessions derived from breeding programs were separated from local landraces. Additionally genome-wide variability correlated with the geographic origin of the genotypes, i.e. Occidental peaches shared a genetic pool differentiated from that of Asian peaches. LD extension was also analyzed per subpopulation, both genome-wide and at chromosome level, confirming its slow decay with distance (an average of 1.8 Mbp fragment conservation), especially in the Chinese collection.

Phenotypic diversity

Phenotypic records available before this study were transformed to meet common scales and were entered in a database developed within the project. In most cases only few of the accessions were phenotypically evaluated per year at each site, consequently data were combined in a large sized sample panel. Correlation between phenotypic data recorded in several years was evaluated at IRTA and at CRA-FRU collections, both with enough number of accessions presenting at least two years of records. Correlations were moderate for flowering time (r^2=0.69 for IRTA and r^2=0.58 for CRA-FRU), on the contrary, maturity data were more stable between years, with r^2=0.94 at both, the Italian and Spanish orchards. The correlation between data obtained in different years could be also assessed for fruit size, sugar and acidity content in the IRTA collection. Values were similar for fruit size and acidity (r^2=0.79 and r^2=0.76, respectively) and lower for sugar content (r^2=0.69).

Genome-wide association analysis (GWAS)

Genome-wide association studies on the following monogenic traits: gland shape (reniform/globular), flower type (showy/non-showy), fruit skin pubescence (peach/nectarine), flesh color (white/yellow), fruit shape (flat/round), fruit texture (melting/non-melting) and acidity (sub-acid/acid), identified SNPs associated with each of the traits (see Micheletti et al., 2015 for more detail). Those SNPs (93 in total) were located in the regions where the traits had been previously mapped (Dirlewanger et al., 1998, 2006; Dettori et al., 2001; Peace et al., 2005; Ogundiwin et al., 2009; Falchi et al., 2013; Vendramin et al., 2014). The associated SNPs were organized in 570 haplotypes with an average length of 2.6 Mbp (ranging from 2.0 to 5.2 Mbp) which is consistent with the LD extension in peach. A detailed analysis of the haplotypes and of their association with the traits reduced the number of significantly associated haplotypes to 36 (an average of 6.3 trait^{-1}).

For those key traits considered in the selection process of breeding programs (fruit skin pubescence, flesh color, fruit shape and acidity) we scrutinized the haplotypes and selected few discriminatory SNPs (2-4 trait^{-1}) to be used for marker assisted selection in breeding programs.

GWAS on quantitative traits (flowering date, fruit size, sugar and acidity content) were conducted year^{-1} and site^{-1}. In general the QTLs were consistent between years and sites. However some QTLs were exclusively observed in one collection. For instance, Figure 1 shows the preliminary GWAS Manhattan plots for flowering time in the Chinese and Spanish collections. Three QTLs were distinguished in the Chinese germplasm (one in chromosome 2, one in chromosome 3 and one in chromosome 8) while in the Spanish collection associations were observed in chromosome 1, chromosome 2 and chromosome 6. In both populations the associations were weak. Further analysis will be required to refine the data, especially to remove extreme phenotypic records (outliers) detected that could have created spurious associations and to integrate all data from the different collections in a single

analysis, which will increase the accuracy and power of the method.

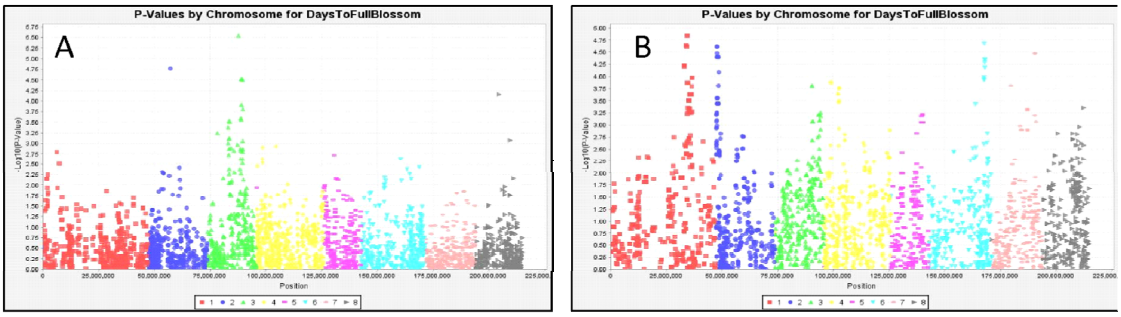

Figure 1. Genome wide association results for flowering time in the Chinese (A) and Spanish (B) peach germplasm collection. Chromosomes are marked with a different color on the horizontal axis of the Manhattan plots.

CONCLUSION

Peach germplasm is preserved in collections maintained and evaluated during years. Such accessions represent a repository of variability that can be introduced as desired in breeding programs. The use of the SNP array in the characterization of the accessions maintained, has allowed the identification of sports, synonyms and validation of the identity of the trees, and has also explored the available variability. Genotypic data, together with phenotypic information, will allow a more efficient management of the collections and of their use in breeding programs, and will provide a base for further coordination of germplasm collections around Europe.

We have proved the viability of genome-wide association in peach by identifying SNPs associated to major genes in the regions predicted by classical linkage mapping, which has allowed the design of SNP markers adequate for MAS. Preliminary results on quantitative characters, with more complex inheritance, are promising.

ACKNOWLEDGEMENTS

We acknowledge Zhi-jun Shen (Jiangsu Academy of Agricultural Sciences), Ke Cao (Zhengzhou Fruit Research Institute), Hui-juan Jia, Xian-qiao Meng (Zhejiang University), Ming Xie (Zhejiang Academy of Agricultural Sciences) and Jian-bao Tian (Shanxi Academy of Agricultural Sciences) for collecting materials and collecting phenotypic data; Martina Lama (CRPV, Cesena, Italy) for her participation in lab work and the technical staff of the "Biologie du Fruit et Pathologie" unit of the INRA, in particular Hélène Christmann, for her technical help in DNA extraction, and Jean Claude Barbot, Hélène Christmann, Anthony Bernard and Aude Mear for their active collaboration in the phenotyping of the INRA-Bordeaux peach collection. We thank the INRA's "Prunus Genetic Resources Center" for preserving and managing the peach collections and the Experimental Unit of INRA-Toulenne (UEA) for growing the trees. This work has been funded under the UE seventh Framework Programme, the views expressed in this work are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Literature cited

Aranzana, M.J., Carbó, J., and Arús, P. (2003). Microsatellite variability in peach [Prunus persica (L.) Batsch]: cultivar identification, marker mutation, pedigree inferences and population structure. Theor. Appl. Genet. *106* (*8*), 1341–1352. PubMed http://dx.doi.org/10.1007/s00122-002-1128-5.

Aranzana, M.J., Abbassi, K., Howad, W., and Arús, P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet. *11* (*1*), 69. PubMed http://dx.doi.org /10.1186/1471-2156-11-69.

Badenes, M.L., Cambra, M., López, M.M., Batlle, I., Iglesias, I., Garcia Brunton, J., Romeu, J., Aranzana, M.J., López-Girona, E., Vives, C., et al. (2015). A peach germplasm collection for increasing the genetic diversity in European

Breeding Programs. Acta Hortic. *1084*, 125–129 http://dx.doi.org/10.17660/ ActaHortic.2015.1084.15.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics *23* (*19*), 2633–2635. PubMed http://dx.doi.org/10.1093/bioinformatics/btm308.

Byrne, D.H. (1990). Isozyme variability in four diploid stone fruits compared with other woody perennial plants. J. Hered. *81* (*1*), 68–71 http://dx.doi.org/10.1093/oxfordjournals.jhered.a110927.

Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., Cheng, S., Zeng, P., Chen, C., Wang, X., et al. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. Genome Biol. *15* (*7*), 415 PubMed.

Dettori, M.T., Quarta, R., and Verde, I. (2001). A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. Genome *44* (*5*), 783–790. PubMed http://dx.doi.org/10.1139/g01-065.

Dirlewanger, E., Pronier, V., Parvery, C., Rothan, C., Guye, A., and Monet, R. (1998). Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. Theor. Appl. Genet. *97* (*5-6*), 888–895 http://dx.doi.org/10.1007/s001220050969.

Dirlewanger, E., Cosson, P., Howad, W., Capdeville, G., Bosselut, N., Claverie, M., Voisin, R., Poizat, C., Lafargue, B., Baron, O., et al. (2004). Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid–location of root-knot nematode resistance genes. Theor. Appl. Genet. *109* (*4*), 827–838. PubMed http://dx.doi.org/10.1007/s00122-004-1694-9.

Dirlewanger, E., Cosson, P., Boudehri, K., Renaud, C., Capdeville, G., Tauzin, Y., Laigret, F., and Moing, A. (2006). Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet. Genomes *3* (*1*), 1–13 http://dx.doi.org/10.1007/s11295-006-0053-1.

Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., Vizzotto, G., and Morgante, M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. Plant J. *76* (*2*), 175–187 PubMed.

Li, X.W., Meng, X.Q., Jia, H.J., Yu, M.L., Ma, R.J., Wang, L.R., Cao, K., Shen, Z.J., Niu, L., Tian, J.B., et al. (2013). Peach genetic resources: diversity, population structure and linkage disequilibrium. BMC Genet. *14* (*1*), 84. PubMed http://dx.doi.org/10.1186/1471-2156-14-84.

Micheletti, D., Dettori, M.T., Micali, S., Aramini, V., Pacheco, I., Da Silva Linge, C., Foschi, S., Banchi, E., Barreneche, T., Quilot-Turion, B., et al. (2015). Whole-Genome Analysis of Diversity and SNP-Major Gene Association in Peach Germplasm. PLoS ONE *10* (*9*), e0136803. PubMed http://dx.doi.org/10.1371/ journal.pone.0136803.

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W.T.B., Flavell, A.J., and Marshall, D. (2010). Flapjack–graphical genotype visualization. Bioinformatics *26* (*24*), 3133–3134. PubMed http://dx.doi.org/ 10.1093/bioinformatics/btq580.

Ogundiwin, E.A., Peace, C.P., Gradziel, T.M., Parfitt, D.E., Bliss, F.A., and Crisosto, C.H. (2009). A fruit quality gene map of *Prunus*. BMC Genomics *10* (*1*), 587. PubMed http://dx.doi.org/10.1186/1471-2164-10-587.

Peace, C.P., Crisosto, C.H., and Gradziel, T.M. (2005). Endopolygalacturonase: a Candidate Gene for Freestone and Melting Flesh in Peach. Mol. Breed. *16* (*1*), 21–31 http://dx.doi.org/10.1007/s11032-005-0828-3.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155* (*2*), 945–959 PubMed.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81* (*3*), 559–575. PubMed http://dx.doi.org/10.1086/519795.

Scorza, R., Mehlenbacher, S.A., and Lightner, G.W. (1985). Inbreeding and coancestry of freestone peach cultivars of the eastern united states and implications for peach germplasm improvement. J. Am. Soc. Hortic. Sci. *110*, 547–552.

Vendramin, E., Pea, G., Dondini, L., Pacheco, I., Dettori, M.T., Gazza, L., Scalabrin, S., Strozzi, F., Tartarini, S., Bassi, D., et al. (2014). A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. PLoS ONE *9* (*3*), e90574. PubMed http://dx.doi.org/10.1371/journal.pone.0090574.

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C.T., Gasic, K., Micheletti, D., Rosyara, U.R., Cattonaro, F., Vendramin, E., et al. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS ONE *7* (*4*), e35668. PubMed http://dx.doi.org/10.1371/journal.pone.0035668.