

Review

Genomics-assisted breeding in fruit trees

Hiro Yoshi Iwata*¹⁾, Mai F. Minamikawa¹⁾, Hiromi Kajiya-Kanegae¹⁾, Motoyuki Ishimori¹⁾
and Takeshi Hayashi²⁾

¹⁾ Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

²⁾ Agroinformatics Division, NARO Agricultural Research Center (NARC), 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8666, Japan

Recent advancements in genomic analysis technologies have opened up new avenues to promote the efficiency of plant breeding. Novel genomics-based approaches for plant breeding and genetics research, such as genome-wide association studies (GWAS) and genomic selection (GS), are useful, especially in fruit tree breeding. The breeding of fruit trees is hindered by their long generation time, large plant size, long juvenile phase, and the necessity to wait for the physiological maturity of the plant to assess the marketable product (fruit). In this article, we describe the potential of genomics-assisted breeding, which uses these novel genomics-based approaches, to break through these barriers in conventional fruit tree breeding. We first introduce the molecular marker systems and whole-genome sequence data that are available for fruit tree breeding. Next we introduce the statistical methods for biparental linkage and quantitative trait locus (QTL) mapping as well as GWAS and GS. We then review QTL mapping, GWAS, and GS studies conducted on fruit trees. We also review novel technologies for rapid generation advancement. Finally, we note the future prospects of genomics-assisted fruit tree breeding and problems that need to be overcome in the breeding.

Key Words: genome-wide association study (GWAS), genomic selection (GS), pseudo-testcross strategy, statistical model, generation advancement technology.

Introduction

Recent advancements in genomic analysis technologies allow cost-effective, high-throughput, and high-density genotyping of genome-wide DNA polymorphisms (Davey *et al.* 2011). These technological advancements have opened up new avenues to promote the efficiency of plant breeding (Chia and Ware 2011, He *et al.* 2014, Varshney *et al.* 2014). Genome-wide association studies (GWAS) enable the detection and identification of quantitative trait loci (QTLs) and genes controlling phenotypic variations in a collection of cultivars and germplasm accessions (Brachi *et al.* 2011, Hamblin *et al.* 2011). Genomic selection (GS; Meuwissen *et al.* 2001) enables the selection of superior genotypes based on genomic estimated breeding values (GEBV) derived from the information of genome-wide DNA polymorphisms (Heffner *et al.* 2009, Jannink *et al.* 2010, Lorenz *et al.* 2011). These novel genomics-based approaches evolved from traditional biparental QTL mapping and marker-assisted selection (MAS), but they have a much broader range of applica-

tion and greater potential than the traditional methods. The advancements in genomic analysis technologies are leading to the active use of genomics-based approaches for plant breeding and genetics research.

Genomics-based approaches can be especially useful in fruit tree breeding (Chia and Ware 2011, Meneses and Orellana 2013, Myles 2013, van Nocker and Gardiner 2014). Conventional breeding of fruit trees has been hampered by their long generation time, large size, extended juvenile phase for seedlings, and a marketable product that cannot be assessed until a seedling is physiologically mature (Luby and Shaw 2001, Rikkerink *et al.* 2007). Numerous biotic and abiotic factors that affect both the quality and quantity of fruits during the pre- and post-harvest periods also complicate genetic improvement (Rikkerink *et al.* 2007). In fruit tree species, genomics-based approaches have great potential to break through these barriers. For example, GWAS enables researchers to estimate the positions and effects of QTLs/genes using existing cultivars/lines without preparing a segregating population (Khan and Korban 2012). When phenotypic data are already available for the existing cultivars/lines, the positions and effects of QTLs/genes can be estimated without performing field experiments. Markers that have significant association with

Communicated by T. Yamamoto

Received September 10, 2015. Accepted January 12, 2016.

*Corresponding author (e-mail: aiwata@mail.ecc.u-tokyo.ac.jp)

target traits can then be used in MAS programs. GS increases the selection accuracy and/or genetic gain per unit time spent in the genetic improvement of fruit quality and yield traits (Kumar *et al.* 2012b). Selection during the juvenile phase speeds the selection process and filters out a substantial proportion of progeny that will proceed to evaluation in field trials, which accelerates fruit tree breeding via shortened breeding cycles and increased selection intensity (Luby and Shaw 2001, Ru *et al.* 2015). Thus, GWAS and GS are promising methods for promoting the efficiency of fruit tree breeding.

In this article, we describe the potential of genomics-assisted breeding using novel genomic technologies in fruit tree species. Biparental QTL mapping and MAS of major QTLs will not completely give way to GWAS and GS, and it is important to use the right methods in the right places. We first introduce molecular marker systems and whole-genome sequence data that are available for fruit tree breeding. Next we introduce the statistical methods for biparental linkage and QTL mapping, GWAS, and GS that underpin genomics-assisted breeding, as well as computer programs available to implement the statistical methods. We then review QTL mapping, GWAS research, and GS studies conducted on fruit trees and some other species. Because trees have a long generation time, MAS and GS alone cannot accelerate the breeding of fruit trees. Therefore, we review novel technologies for rapid generation advancement that can be used to accelerate fruit tree breeding. Finally, we note the future prospects of genomics-assisted fruit tree breeding and problems that need to be overcome.

Molecular markers available for fruit trees

Molecular markers are indispensable for genomics-assisted breeding, and various marker systems have been developed. The availability of fruit tree genome sequences enables researchers to develop genome-wide markers for high-throughput genotyping and to construct high-density genetic maps (Myles 2013). It also improves our understanding of the molecular mechanisms underlying fruit traits and provides important clues concerning the evolution of their complex genomes.

Simple sequence repeat (SSR) markers are designed through both transcript sequences and genomic sequences by using bioinformatics programs; SSR markers are useful tools for gene/QTL mapping, MAS, and diversity analysis (Segura *et al.* 2008). SSR markers have several advantages as genetic markers: they are codominant, multi-allelic, reproducible, and often amplifiable across species (Miah *et al.* 2013). Because of this cross-species transferability, SSR markers have been used for comparative mapping and genome synteny analysis in Rosaceae species (e.g., Dirlwanger *et al.* 2004, Fan *et al.* 2013, Gasic *et al.* 2009).

Single nucleotide polymorphism (SNP) markers are also being developed using expressed sequence tags and the genome sequence. These markers are cost-effective in terms

of cost per marker and allow for higher throughput screening and higher density mapping compared to SSR markers. SNP arrays such as the Illumina Infinium II system have been developed for fruit tree species (Chagné *et al.* 2012, Myles *et al.* 2010). To date, the apple SNP array has been used to create linkage maps (Antanaviciute *et al.* 2012) and to assist GS (Kumar *et al.* 2012b) and GWAS (Kumar *et al.* 2013).

Next-generation sequencing (NGS) technologies have been used for genotyping SNPs via genotyping by sequencing (GBS) methods (Elshire *et al.* 2011) and restriction site-associated DNA sequencing (Baird *et al.* 2008). The NGS-based genotyping methods enable the simultaneous detection of thousands of SNPs throughout the genome in mapping populations. An advantage of NGS-based genotyping over SNP arrays is that there is no need for SNP discovery and array design. For newly targeted species or populations, SNP array development is very time-consuming and costly. Because NGS-based genotyping methods can be used without a reference genome (Catchen *et al.* 2011), they are useful for a range of organisms, including minor fruit trees. Another advantage of the NGS-based genotyping methods is that they are less affected by the ascertainment bias that may influence GWAS and GS (Albrechtsen *et al.* 2010, Heslot *et al.* 2013). Although NGS-based genotyping generally produces a relatively large number of SNPs, most SNPs have numerous missing data across samples (Gardner *et al.* 2014). One solution for this problem is to use an imputation method to fill in the missing genotypes, such as the TASSEL-GBS pipeline (Glaubitz *et al.* 2014) and Beagle (Browning and Browning 2007).

Fruit tree genome resources on the Web

Fruit tree genome databases that house genomic, genetic, and breeding resources provide an effective platform for breeding programs. It is important for fruit researchers to understand how these data can be used to solve problems in fruit production. Genetic map, marker, and QTL data facilitate the development of markers for genomics-assisted breeding and the discovery of genes underlying important agricultural traits.

Sequencing projects for fruit genomes have been reported at Phytozome (www.phytozome.net), which provides well-controlled microsynteny and gene family evolution data. Tree fruit Genome Database Resources (www.tfgdr.org) is a collection of Rosaceae, *Citrus*, and *Vaccinium* bioinformatics resources and software tools (Wegrzyn *et al.* 2012). The Genome Database for Rosaceae (www.rosaceae.org) is a central repository of genetics and breeding data and analysis tools for Rosaceae. It has a publicly available Breeders' Toolbox that is integrated with existing genomic and genetic data (Jung *et al.* 2014). The Citrus Genome Database (www.citrusgenomedb.org) is a community database providing access to citrus genomics, genetics, and breeding research (Jung *et al.* 2008). A genome database for *Vaccinium*

(www.vaccinium.org) is being developed to house and integrate genomic, genetic, and breeding data for blueberry, cranberry, and other *Vaccinium* species. The Plant Genome Duplication Database (chibba.agtec.uga.edu/duplication/) is a Web service providing intra-genome or cross-genome syntenic relationships (Lee *et al.* 2013), which helps to identify conserved genes between species. The development of genetic and genomic resources will facilitate genomics research and genomics-assisted breeding applications.

Methods for genetic analyses of fruit trees

Linkage analysis

The objective of linkage analysis is to estimate the recombination frequency between markers or between markers and loci affecting a trait; the analysis includes construction of a linkage map and mapping of genes that determine simply inherited traits, using multi-generation families. In fruit trees, a full-sib family consisting of two parental cultivars (hereafter T_1 and T_2) and their F_1 progeny is commonly used for linkage analysis (e.g., Cristofani *et al.* 1999, Doucleff *et al.* 2004, Fernández-Fernández *et al.* 2012, Hemmat *et al.* 1994, Yamamoto *et al.* 2002) because the inclusion of more than two generations is hindered by the long generation time.

In a two-generation family, recombination events occur during gametogenesis in each of the parents and the recombination frequency is estimated based on gametes transmitted from the parents to F_1 individuals. Consequently, linkage analysis is conducted separately for T_1 and T_2 , regarding the F_1 progeny as a hypothetical half-sib progeny produced from each T_1 and T_2 and another parent as an anonymous mating partner. Linkage analysis in a half-sib design is carried out in a similar way as for a backcross population derived from crossing two inbred lines, referred to as a BC population. For a half-sib family consisting of T_1 and the F_1 progeny, regarding T_2 as an anonymous mating partner of T_1 , heterozygous markers or loci in T_1 can be used for estimating recombination frequencies. In applying the backcross analysis to this half-sib family, T_1 and T_2 are treated as the F_1 individual and recurrent parent in the backcross design, respectively.

It should be noted, however, that in a half-sib family it is necessary to distinguish the recombinant and non-recombinant types of gametes transmitted from T_1 to F_1 progeny with respect to two linked markers because the linkage phase in T_1 is usually unknown for the two markers. Consider two linked markers, A and B, and assume that the genotype of T_1 is $A_1A_2B_1B_2$, where A_1 (B_1) and A_2 (B_2) are two different alleles of A (B). Taking the linkage phase into consideration, there are two possible diplotypes for $A_1A_2B_1B_2$, A_1B_1/A_2B_2 and A_1B_2/A_2B_1 , where a diplotype is a representation of a genotype including the information of linkage phase. If an F_1 individual receives a gamete with haplotype A_1B_1 from T_1 , this gamete is classified as non-recombinant or recombinant according to the two possible diplotypes of T_1 , A_1B_1/A_2B_2 and A_1B_2/A_2B_1 , respectively.

The diplotype of T_1 is inferred based on the number of gametes with haplotypes A_1B_1 or A_2B_2 , n_1 , and that of gametes with A_1B_2 or A_2B_1 , n_2 , transmitted from T_1 to F_1 individuals. For example, the diplotype of T_1 is inferred as A_1B_1/A_2B_2 when $n_1 > n_2$ and as A_1B_2/A_2B_1 otherwise. Given the diplotype of T_1 for linked markers, linkage analysis is performed based on the genotypes of the F_1 individuals in the same way as in a backcross design.

This strategy of linkage analysis for a half-sib family of one parent and the F_1 progeny included in a full-sib family is called the pseudo-testcross strategy (Grattapaglia and Sederoff 1994). For linkage analysis of the gametes from T_2 , the same process is applied. A two-way pseudo-testcross entails conducting pseudo-testcross linkage analyses for each of two parents in turn (Grattapaglia and Sederoff 1994).

Backcross analysis is implemented in most software for linkage analysis, such as MAPMAKER (Lander *et al.* 1987), AntMap (Iwata and Ninomiya 2006), CARTHAGENE (De Givry *et al.* 2005), and JoinMap (van Ooijen 2006). Therefore, construction of a linkage map based on the pseudo-testcross design is allowed with these software packages for a full-sib family. In linkage analysis with a backcross design, the genotype of a marker for an individual in the BC population is recorded as “A” or “H” depending on which one of the two parental inbred lines is the origin of the allele transmitted by F_1 to the BC individual. In a pseudo-testcross design, T_1 is treated as the F_1 of the backcross design, but the parental origins of the two alleles possessed by T_1 are usually unknown. Therefore, the following procedure is used for pseudo-testcross analysis with the backcross option in these software packages.

For each marker, the genotypic data of F_1 progeny are recorded in two different ways. Assuming that the genotype of T_1 at a marker is A_1A_2 , for example, the genotype of an F_1 individual is recorded for the marker as “A” if it receives the A_1 allele from T_1 and as “H” otherwise, considering the genotype A_1A_2 as A_1/A_2 , where the left side of the slash is the maternal allele and the right side is the paternal allele. Subsequently, for the same marker, the genotype of an F_1 individual is once again recorded, but this time exchanging “A” and “H” in the original record of the genotype by regarding A_1A_2 as A_2/A_1 . Two genotype datasets for a marker thus obtained are treated as those of two different markers, such that the number of markers treated is duplicated. When identifying linkage groups for these duplicated markers, a set of linkage groups consisting of pairs of equivalent linkage groups is obtained. Within a linkage group, the records of genotypic data of markers suitably reflect the diplotype of T_1 . One linkage group is selected from the pair of equivalent linkage groups to construct the map of markers for the linkage group. By replicating this process for each pair of linkage groups in turn, the linkage maps for all linkage groups are produced.

In applying the two-way pseudo-testcross strategy to a full-sib family, we obtain two separate linkage maps for female and male parents. By using multi-allelic markers such

as SSR markers with more than two alleles segregating in F_1 individuals, markers heterozygous in both parents can be mapped in both linkage maps, which can be used for aligning female and male linkage maps (Maliepaard *et al.* 1998). By using the software JoinMap (van Ooijen 2006), two parental linkage maps can be integrated.

QTL analysis

Many fruit tree traits of economic importance, such as those related to fruit quality and productivity, are quantitative traits, and loci affecting such traits are called QTLs. A statistical method that combines linkage analysis with a statistical model of phenotypic values of a trait, referred to as QTL analysis, is applied for mapping QTLs and estimating their effects. As in linkage analysis, a full-sib family consisting of two parents, T_1 and T_2 , and their F_1 progeny, is commonly used as the analyzed population in QTL analysis of fruit trees (e.g., Ban *et al.* 2014, Kuniyama *et al.* 2014, Siviero *et al.* 2006, Weber *et al.* 2003, Yamamoto *et al.* 2014). Because the procedure of QTL analysis is partly based on that of linkage analysis, the pseudo-testcross strategy can also be used for the analysis of QTLs in a full-sib family (Grattapaglia *et al.* 1995), with QTL analysis conducted separately for T_1 and T_2 and with the F_1 progeny regarded as a half-sib progeny obtained from each T_1 and T_2 . When pseudo-testcross analysis is applied to both parents consecutively, this process of QTL analysis is referred to as a two-way pseudo-testcross (Grattapaglia *et al.* 1995), as in linkage analysis. In pseudo-testcross QTL analysis, heterozygous QTLs in the parent are targets of QTL detection and, using the linkage map of the parent, the analysis is conducted as in the backcross design of QTL analysis in inbred species, except that the haplotypes for marker genotypes of the parent must be inferred. Because interval mapping (Lander and Botstein 1989) is commonly applied to QTL analysis of a BC population in inbred species, we explain the application of this method for pseudo-testcross design.

In interval mapping, any position on a linkage map is tested for the presence of a QTL affecting the trait. A putative QTL, Q , is located at a tested position between two markers, A and B , on a linkage map and the effect of Q is evaluated. Consider a pseudo-testcross strategy of QTL analysis for T_1 in a full-sib progeny derived from crossing T_1 and T_2 . We assume that the QTL genotype of T_1 at Q is Q_1Q_2 , with Q_1 and Q_2 being the different alleles of Q . In F_1 progeny including n individuals, the phenotypic value of a trait for the i th F_1 individual, y_i , is expressed as a linear model:

$$y_i = \mu + u_i a + e_i, \quad (1)$$

where μ is the intercept of the model, u_i is a covariate indicating the allele of Q transmitted to the i th F_1 individual from T_1 (taking values 1 and 0 for alleles Q_1 and Q_2 , respectively), a is the difference in allelic effects between Q_1 and Q_2 , and e_i is a residual. The allele the F_1 individual receives

from T_1 (i.e., Q_1 or Q_2) is inferred from the alleles of the flanking markers A and B transmitted from T_1 to the F_1 individual, where it is assumed that the haplotype of T_1 with respect to A and B is known. The model parameters, including a and μ , are estimated such that the parameter values fit well the data of phenotypic values of n F_1 individuals under the model. The presence of a QTL at a position of interest is judged by comparing the goodness of model fit between two hypotheses, $H_1: a \neq 0$ and $H_0: a = 0$, representing the presence and absence of a QTL, respectively (Lander and Botstein 1989). Test statistics such as an LOD score are used for testing H_0 versus H_1 . The effects of QTLs on regions other than a tested position can be included in model (1) to enhance the power of QTL detection and the precision of estimation of QTL position and effect (Jansen 1993, Zeng 1994).

The interval mapping method is implemented in computer programs widely used for QTL analysis, including Mapmaker/QTL (Lincoln and Lander 1990), QTL Cartographer (Wang *et al.* 2010), R/qtl (Broman *et al.* 2003), and MapQTL (van Ooijen 2009). Pseudo-testcross analysis of QTLs is performed with the backcross option in these software packages using the data of phenotypes and marker genotypes of the F_1 progenies recorded following the haplotype of one of the parents in a full-sib family.

Based on an integrated linkage map constructed from two parental maps, a full-sib analysis of QTLs can be performed to investigate the presence of QTLs heterozygous in one or both parents. Full-sib QTL analysis with interval mapping is implemented in MapQTL (van Ooijen 2009). Assuming that the genotypes of T_1 and T_2 at a putative QTL, Q , located at a tested position are Q_1Q_2 and Q_3Q_4 , respectively, the model of interval mapping for full-sib analysis is:

$$y_i = \mu + u_i a + v_i b + e_i, \quad (2)$$

where μ , u_i , a , and e_i are the same terms as in the pseudo-testcross model (1), v_i is a covariate indicating which of the alleles at Q (Q_3 or Q_4) is transmitted to the i th F_1 individual from T_2 , and b is the difference in the allelic effects between Q_3 and Q_4 . The full-sib analysis with model (2) requires marker haplotypes to be known for T_1 and T_2 . The procedure of inferring marker haplotypes can be incorporated in the QTL analysis (Hayashi and Awata 2004, Knott *et al.* 1996).

In breeding of fruit trees such as apple, multiple full-sib families obtained from crosses among existing cultivars are established as a breeding population. The information of QTLs obtained from QTL analysis conducted individually for each of the multiple full-sib families can be integrated to increase the reliability of the estimated QTL positions and effects. Such an integrated QTL analysis was performed for fruit quality traits in apple by Costa (2015), where four full-sib families were individually analyzed with model (2), and then QTLs detected in the individual analyses were projected on a consensus map of the four families to elucidate reliable genomic regions of the QTLs. This integrated analysis

using the QTL information for multiple families is called MetaQTL analysis (Costa 2015, Veyrieras *et al.* 2007).

The parental cultivars used for crosses to establish multiple full-sib progenies may be genetically related due to their common ancestral cultivars. By adding such ancestral cultivars, multiple full-sib families are treated as a large complex pedigree. The analysis of such a large pedigree allows the QTL positions and effects to be estimated more precisely.

Bink *et al.* (2014) performed QTL analysis on a pedigree including 27 full-sib families derived from crosses among 33 parental cultivars, containing a total of 1300 individuals, and all of the ancestral cultivars of the parental cultivars. The ancestral cultivars for which both parents were unknown were regarded as founders in this large pedigree. The authors assumed a biallelic QTL with alleles Q and q and applied a multiple-QTL model, which is expressed for the phenotype of the j th individual of the i th full-sib family, y_{ij} , as:

$$y_{ij} = \mu + \sum_{l=1}^N u_{ijl} a_l + e_{ij}, \quad (3)$$

where μ is the intercept of the model, N is the number of QTLs included in the model, u_{ijl} is the covariate indicating the genotype of the l th QTL for the individual, a_l is the effect of the l th QTL, and e_{ij} is a residual. Bayesian estimation was applied to estimate the model parameters, where the prior probabilities for the QTL number N , the QTL positions and effects, and the QTL genotypes of the founders were taken into consideration. Given the QTL genotypes of the founders, possible allele transmissions from the founders to the parental cultivars through all of the ancestors were simulated. The posterior distributions of the model parameters including N , QTL positions, and QTL effects as well as the QTL genotypes of the founders were obtained through Markov chain Monte Carlo iterations (Bink *et al.* 2014).

Genome-wide association study and genomic selection

Association analysis is a method for finding an association between markers and loci affecting a trait by assessing the correlation between the genotypes of markers and phenotypes in an analyzed population. By using a large number of markers covering an entire genome, GWAS allows the confirming such an association for most genome segments. In GWAS, genome segments of analyzed individuals are discriminated based on the allelic states of markers located on the segments, whereas in QTL analysis those are discriminated based on their parental origins in the analyzed family. In GWAS, therefore, various populations, including a collection of individuals sampled from wild populations, germplasm, and breeding cultivars, are analyzed in addition to the multi-generation families derived from crossing parental cultivars, as used in QTL analysis.

In both GWAS and QTL analyses, the successful detection of genome regions associated with phenotypes relies on the availability of markers linked to the loci affecting a trait. The linkage between markers and the loci in the genomes of parents is weakened by recombination events occurring as

generations advance, but the decay of the linkage is subtle in QTL analysis due to the limited number of recombination events in a few generations. The existence of an association between markers and the loci in an analyzed population caused by linkage or some other factors, such as selection, over the history of the population is called linkage disequilibrium (LD). The length of a genome segment that can be covered by a marker depends on the degree of LD in the genome region in the population. In populations with a higher degree of LD, such as multi-generation families from crossing parental cultivars, the number of markers required for covering the entire genome is small, whereas numerous markers are required for covering the genome in populations with a lower degree of LD, such as a collection of distantly related individuals including germplasm collection. Mapping resolution is also influenced by the degree of LD, with finer resolution as the degree of LD decreases.

Recently, the information on several thousands to tens of thousands of SNPs distributed across a whole genome at high density has become available for fruit trees. Therefore, GWAS may be an effective method for mapping loci affecting a trait in a collection of cultivars without the need to establish experimentally crossed populations, which is hindered in most fruit trees due to the long generation time, thus replacing family-based QTL analysis (Khan and Korban 2012). Although GWAS is applied to a wide range of populations, factors such as population stratification and cryptic relatedness among individuals may cause an increased rate of false-positives, meaning spurious genotype-phenotype associations (Kumar *et al.* 2013). Thus, correction terms for population structure and kinship relationships must be included in statistical models of GWAS to control the false-positive rate.

A mixed linear model is a statistical model, which takes population structure and kinship relationship into consideration, and is suitable for GWAS (Yu *et al.* 2005). In this model, each marker is tested in turn for association with a trait, while incorporating population structure and kinship relationship as fixed and random effects, respectively. Consider a population consisting of n individuals with records of phenotypes and genotypes of p markers covering the entire genome, and denote the phenotypic value of the i th individual as y_i ($i = 1, 2, \dots, n$). The following model is used to test the l th marker in a mixed linear model:

$$y_i = \mu + z_{il} a_l + \sum_j x_{ij} c_j + g_i + e_i, \quad (4)$$

where μ is the intercept of the model, z_{il} is the covariate indicating the genotype of the i th individual at the l th marker (e.g., taking values 0, 1, and 2 for three possible genotypes when the marker is biallelic), a_l is the effect of the l th marker, x_{ij} is the covariate relating c_j (i.e., the effect of the j th non-genetic factor) to y_i , g_i is the genotypic value of the i th individual contributed by polygenic effect not captured by a tested marker, and e_i is a residual. In estimating the model parameters, a_l and c_j are treated as fixed effects and g_i is

treated as a random effect. The influence of the population structure is included in the model as non-genetic effects c_j for some j in the model.

The random polygenic effects of n individuals, collectively written as $\mathbf{g} = (g_1, g_2, \dots, g_n)'$, with the prime symbol representing transpose of a vector, are assumed to follow a multivariate normal distribution with a mean vector $\mathbf{0}$ and a variance-covariance matrix $\mathbf{A}\sigma_g^2$, where \mathbf{A} is a kinship matrix with the (j,k) th elements indicating the genetic relationship of a pair of the j th and k th individuals. Kinship relationships are considered in the construction of \mathbf{A} using pedigree information or marker genotypes. Assuming that \mathbf{g} is estimated for all markers, we can write $\mathbf{g} = \mathbf{Z}\mathbf{a}$, where \mathbf{a} is a vector of the effects of all markers and \mathbf{Z} is a design matrix relating \mathbf{a} to \mathbf{g} , with the (i,m) th element z_{im} being the covariate indicating the genotype of the i th individual at the m th marker. Therefore, \mathbf{Z} is regarded as the information of marker genotypes. Assuming that \mathbf{a} is a random effect following a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}\sigma_a^2$, with \mathbf{I} being an identity matrix, \mathbf{A} is expressed as $\mathbf{A} = \mathbf{Z}\mathbf{Z}'\sigma_a^2/\sigma_g^2$, indicating how marker genotypes are used for constructing \mathbf{A} (Goddard 2009). Because \mathbf{g} is the polygenic effect not captured by the l th marker in model (4), z_{il} and a_l might be excluded from \mathbf{Z} and \mathbf{a} . When the number of markers, p , is very large, as is the case for GWAS, the inclusion of a_l in \mathbf{g} provides almost the same result as GWAS with a_l excluded from \mathbf{g} .

Using the vector and matrix forms, model (4) is rewritten as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{z}_l a_l + \mathbf{g} + \mathbf{e}, \quad (5)$$

where \mathbf{b} is a vector of fixed effects containing the intercept (μ) and non-genetic effects (c_j), $\mathbf{z}_l = (z_{1l}, z_{2l}, \dots, z_{nl})'$ is a vector of the covariates indicating the genotypes of the l th marker for n individuals, \mathbf{g} is a vector of the random polygenic effects of n individuals as described above, and $\mathbf{e} = (e_1, e_2, \dots, e_n)'$ is a vector of residuals. Significance testing for the association of each tested marker with the phenotype is based on the estimates of a_l , \hat{a}_l , where a statistical test of hypotheses $H_0: a_l = 0$ versus $H_1: a_l \neq 0$ is conducted. In the usual procedure of GWAS with a mixed linear model, the P value of the estimate \hat{a}_l under H_0 is calculated and $-\log_{10}(P \text{ value})$ is plotted against each marker position to detect markers significantly associated with the phenotype; this plot is called a Manhattan plot. When H_0 is rejected, the association of a tested marker and phenotype is regarded as significant and a marker affecting a trait is detected. Estimation of parameters in mixed linear model (5) can be conducted with several software packages, including the rrBLUP package in R program (Endelman 2011) and TASSEL (Bradbury *et al.* 2007).

By substituting \mathbf{g} with $\mathbf{Z}\mathbf{a}$ and assuming that $\mathbf{z}_l a_l$ is included in \mathbf{g} in model (5), we obtain the modified model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}. \quad (6)$$

All markers are simultaneously fitted using model (6) for GWAS. The number of markers, p , is often much greater than that of analyzed individuals, n , when high-density SNP markers are used. For simultaneously estimating the effects of all markers, \mathbf{a} , under this situation of $p > n$, model (5) is handled as a mixed linear model by treating \mathbf{a} as random effects, where the method of best linear unbiased prediction (BLUP) is adopted for estimating \mathbf{a} , or managed by applying Bayesian estimation assigning a prior distribution for \mathbf{a} .

Iwata *et al.* (2013a) applied Bayesian estimation for GWAS in a population of Japanese pears using genome-wide SSR markers based on model (5) incorporating a variable selection procedure. To date, many types of Bayesian methods have been proposed (e.g., Bayes A, Bayes B, Bayes $C\pi$, Bayes $D\pi$, Bayesian LASSO), each depending on the prior distributions assumed for \mathbf{a} and other parameters and estimation procedures with or without variable selection. Several methods of Bayesian estimation for model (6) can be performed with the BGLR package in the R program (Pérez and de los Campos 2014).

GS is a method of individual selection based on prediction of the genotypic value of a target trait based on the genotypes of genome-wide markers (Meuwissen *et al.* 2001). From the point of view of artificial selection, the genotypic value is called the breeding value. Breeding values for individuals with marker genotypes in GS are predicted using the same statistical models as in GWAS, where genotypes and phenotypes of a collection of individuals, referred to as the training population, are used for the estimation of model parameters. Breeding values are predicted for selection candidates based only on their genotypes with the prediction model thus constructed.

Using the same notations as in GWAS, the objective of GS is to predict \mathbf{g} for selection candidates based on the information of marker genotypes \mathbf{Z} . The predicted value of \mathbf{g} , denoted as $\hat{\mathbf{g}}$, is referred to as the genomic estimated breeding value (GEBV). As we denote $\mathbf{g} = \mathbf{Z}\mathbf{a}$ with effects of all markers available (\mathbf{a}), GEBV is obtained as $\hat{\mathbf{g}} = \mathbf{Z}\hat{\mathbf{a}}$, with $\hat{\mathbf{a}}$ being the estimate of \mathbf{a} calculated from model (5) using the data of marker genotypes and phenotypes of the training population. Based on the model obtained by omitting the term $\mathbf{z}_l a_l$ from model (5), GEBV $\hat{\mathbf{g}}$ is directly calculated using a BLUP method without estimating \mathbf{a} , where kinship matrix \mathbf{A} is calculated with the marker genotypes \mathbf{Z} . The method of BLUP based on a kinship matrix obtained by marker genotypes is referred to as genomic BLUP (GBLUP; van Raden 2008). Bayesian methods are also applied to the estimation of breeding value \mathbf{g} , where some prior distributions are assigned to \mathbf{a} .

Kumar *et al.* (2012b) applied both BLUP and Bayesian LASSO to predict GEBV of fruit quality traits in apple. Instead of using statistical linear models such as model (6), some machine-learning algorithms for the discrimination of objects based on a large number of features, including support vector machine and random forest, also have been applied to predict GEBV in GS (Jannink *et al.* 2010).

QTL mapping and GWAS in fruit trees

Most of agronomic and horticulturally important traits, such as fruit quality, are quantitative and controlled by multiple, sometimes numerous, genes or QTLs. Many genes and QTLs for disease and pest resistance have been reported, including those for scab resistance in apple (Bus *et al.* 2010) and pear (Pierantoni *et al.* 2007, Terakami *et al.* 2006), plum pox virus resistance in apricot (Soriano *et al.* 2008), brown rot resistance in peach (Pacheco *et al.* 2014), and downy and powdery mildew resistance in grapevine (Riaz *et al.* 2011, van Heerden *et al.* 2014). Various QTLs controlling fruit quality traits (e.g., harvest time, fruit skin color, fruit weight, and sugar content) have also been identified in apple (Kenis *et al.* 2008, Kunihisa *et al.* 2014), pear (Yamamoto *et al.* 2014, Zhang *et al.* 2013), peach (Eduardo *et al.* 2011, Martínez-García *et al.* 2013), and grapevine (Correa *et al.* 2014).

QTL mapping is commonly performed using a single full-sib family, thus highlighting the instability of QTLs among different genetic backgrounds (Kenis *et al.* 2008). To overcome this limitation of QTL mapping, MetaQTL analysis is used; it is a novel tool that allows the results of multiple independent QTL mapping experiments to be integrated (Veyrieras *et al.* 2007). MetaQTL analyses have been performed for plum pox virus resistance in apricot (Marandel *et al.* 2009) and fruit quality traits in apple (Costa 2015).

Compared to QTL mapping, GWAS is more suitable for QTL detection in fruit trees because it does not require biparental populations. Generating segregating populations derived from biparental crosses of fruit trees is difficult and costly due to their long juvenile periods (Khan and Korban 2012, Rikkerink *et al.* 2007). There are few studies on genetic determination of quantitative traits in fruit trees using GWAS (Table 1). In apple, Kumar *et al.* (2013) conducted GWAS using 1200 seedlings of seven full-sib families to reveal significant associations of 2500 SNPs with six fruit

traits. Significant associations were found in all six traits in the genomic regions, some of which were coincident to known candidate genes. Iwata *et al.* (2013a) conducted GWAS using 76 cultivars of Japanese pear to found significant associations of 162 markers with nine agronomic traits. Significant associations were found for harvest time, black spot resistance, and spur number. In peach, Cao *et al.* (2012) conducted GWAS using 104 landrace accessions genotyped with 53 genome-wide SSR markers, and they found associated markers for 10 traits related to fruit and phenological period. Association studies with candidate genes have been conducted in apple, and contributed to finding a gene controlling fruit flesh firmness (Cevik *et al.* 2010). Significant overlaps between results of QTL mapping and association studies for disease resistance and fruit quality traits have been reported in apple (Cevik *et al.* 2010, Kumar *et al.* 2012b, 2013), pear (Iwata *et al.* 2013a, Yamamoto *et al.* 2014), and peach (Cao *et al.* 2012, Picañol *et al.* 2013).

Current status and perspectives of MAS

Although recent advances in genomics research have provided abundant genomics resources (e.g., genetic and physical maps, QTLs, and numerous molecular markers for many traits), there are few reports of MAS application (Folta and Gardiner 2009, Ru *et al.* 2015). Issues including technical and economic barriers have prevented the integration of MAS technology into conventional breeding. Choosing reliable markers from among robust markers is difficult for breeders. The LD between a marker and trait locus may differ in every population (Brescghello and Sorrells 2006), and the effect of QTLs may differ among genetic backgrounds and environmental conditions (Li *et al.* 2003, Liao *et al.* 2001). In addition, MAS will not always have greater cost-effectiveness than conventional breeding, because the cost of MAS depends on the target traits (e.g., monogenic or polygenic) and the ease of measuring them (Morris *et al.*

Table 1. Summary of genome-wide association studies (GWAS) in fruit trees

Species	Population size	Markers	Traits	Statistical models for GWAS	References
Apple (<i>Malus × domestica</i>)	1200 seedlings	2500 SNPs	Six traits (fruit firmness, weighted cortical intensity, internal browning, titratable acidity, fruit splitting, bitter pit)	Mixed linear model	Kumar <i>et al.</i> (2013)
Japanese pear (<i>Pyrus pyrifolia</i>)	76 cultivars	162 markers (155 SSRs, 4 RAPD-STS, 2 ACC synthase genes, 1 S-RNase gene)	Nine traits (harvest time, resistance to black spot, firmness of fresh, fruit size, fruit shape in longitudinal selection, acid content, total soluble solid content, number of spurs, vigor of tree)	Multilocus Bayesian model	Iwata <i>et al.</i> (2013a)
Peach (<i>Prunus persica</i>)	104 landraces	53 SSR markers	10 traits (flesh color around the stone, red pigment in the flesh, flesh texture, flesh adhesion, flesh firmness, fruit weight, chilling requirement, flowering time, ripening time, and fruit development period)	General linear model	Cao <i>et al.</i> (2012)

SNP, single nucleotide polymorphism; SSR, simple sequence repeat; RAPD-STS, random amplified polymorphic DNA-sequence tagged sites; ACC, 1-aminocyclopropane-1-carboxylate.

2003). To solve these issues, the RosBREED project funded by the USDA Specialty Crop Research Initiative was established in 2009 (Iezzoni *et al.* 2010). An eight-stage marker-assisted breeding pipeline was implemented to enable the use of marker-assisted breeding in rosaceous tree fruits. Consequently, MAS has been successfully applied in apple (Edge-Garza *et al.* 2010, Kellerhals *et al.* 2011, Peace 2013, Sebolt 2013) and sweet cherry (Haldar *et al.* 2010).

The application of MAS is increasing, but its technical limitations also have been revealed. MAS is effective for improving traits controlled by a small number of major genes and/or large-effect QTLs (e.g., pest and disease resistance), whereas it is difficult to use for traits controlled by a large number of minor genes, as is the case with many fruit quality traits (Jannink *et al.* 2010, Kumar *et al.* 2012a). The new technology of GS was proposed to overcome the limitations using whole-genome prediction models based on the genome-wide markers (Meuwissen *et al.* 2001).

Genomic selection in tree crops

Compared to MAS, GS is more suitable for selecting traits controlled by many minor genes, and there is no need to generate crossing populations to develop markers. In GS, selection decisions are based on the predicted GEBV of selection candidates. In recent years, new sequencing technologies have decreased the costs of SNP genotyping and resulted in greater availability of numerous markers, which will increase prediction accuracy. GS was first described in dairy cattle breeding programs (Calus 2010, Hayes *et al.* 2009) and has been used subsequently in crops (Heffner *et al.* 2009, Lin *et al.* 2014).

Because GS has great potential for streamlining plant breeding, it is being eagerly studied for its application to various plant species. The use of GS in plant breeding has been evaluated mainly from two perspectives: (1) deterministic or stochastic simulations; and (2) empirical data analysis. The GS studies of fruit trees have thus far included apple (Kumar *et al.* 2012a, 2012b), Japanese pear (Iwata *et al.* 2013a), and grapevine (Fodor *et al.* 2014); here, we also describe studies in other tree crops including forest trees and oil palm.

Grattapaglia and Resende (2011) conducted deterministic simulations and assessed the impact of the degree of LD, size of the training set, trait heritability, and number of QTLs on the accuracy of GS in forest tree breeding. The degree of LD, which is modeled by effective population size and marker density, had the largest impact on the accuracy. The accuracy of GS was comparable to that of traditional selection based on pedigree-based BLUP (PBLUP) even at a moderate marker density (2 markers/cM) when the effective population size was small (≤ 30); shortening the breeding cycle by 50% with GS provided an increase of at least 100% in selection efficiency, suggesting a promising effect of GS in tree breeding.

Iwata *et al.* (2011) performed stochastic simulations to

evaluate the efficiency of GS in forest tree breeding, and their conclusion was consistent with that of Grattapaglia and Resende (2011). That is, by using a base population derived from a limited number ($n=25$) of elite trees (i.e., small effective population size), GS was advantageous over phenotypic selection even for a low-heritability polygenic trait at a moderate marker density (1 marker/cM). The simulations also suggest that updating the prediction model is indispensable for attaining large genetic gain from GS breeding, because the pattern of LD changes with increasing selection cycles.

Denis and Bouvet (2013) performed stochastic simulations to assess the efficiency of GS in eucalyptus breeding, with findings similar to those of Grattapaglia and Resende (2011) and Iwata *et al.* (2011). That is, GS attained two or three times greater genetic gain per unit time than that of phenotypic selection, although the gain per cycle declined in later breeding cycles. The authors also compared the performance of GS models with and without dominance effects. The model with dominance effects performed better for clone selection when heritability was high and dominance effects were preponderant, but no improvement was detected for parent selection.

Fodor *et al.* (2014) performed stochastic simulations to assess the efficiency of GS for grapevine breeding. The authors simulated the domestication histories of three grapevine diversity groups and evaluated the accuracy of GS in simulated breeding populations. High accuracy levels were obtained using a core collection covering three diversity groups as a training population, and the highest prediction accuracy was attained with the combination of GWAS and GS.

Wong and Bernardo (2008) simulated GS breeding in oil palm and demonstrated the superiority of GS over marker-assisted recurrent selection and phenotypic selection in terms of genetic gain per unit cost and time. Although their simulation results are suggestive for tree breeding, the results might not be easily generalized because they simulated breeding populations derived from selfing a hybrid between two inbred lines, whereas actual oil palm breeding populations are more complex (Cros *et al.* 2015).

In fruit trees, empirical data analysis has been conducted in apple (Kumar *et al.* 2012b) and Japanese pear (Iwata *et al.* 2013a), as summarized in **Table 2**. Kumar *et al.* (2012b) assessed the accuracy of GS by analyzing seven full-sib families with 1120 individuals that were genotyped for 2500 SNPs on the International RosBREED SNP Consortium apple Infinium array v1. The accuracy of GS ranged from 0.68 to 0.89 and was higher than the accuracy of PBLUP for all traits. GS allows the modeling of the variation caused by random sampling of two possible alleles from each parent at each locus during meiosis (called Mendelian sampling), whereas PBLUP cannot take such variation into account. The higher accuracy of GS suggests that it could account for both family effects and Mendelian sampling. Two modeling methods, GBLUP and Bayesian LASSO, were compared,

Table 2. Summary of empirical studies of genomic selection in perennial tree species

Species	Population type and size	Markers	Traits	Methods for building prediction models	Prediction accuracy	References
Apple (<i>Malus × domestica</i>)	1120 individuals (seedlings generated from a factorial mating design of four females and two male parents)	2500 SNPs	Six traits (fruit firmness, soluble solids, russet coverage, weighted cortex intensity, astringency, titratable acidity)	rrBLUP, Bayesian LASSO	0.68 to 0.89	Kumar <i>et al.</i> (2012b)
Japanese pear (<i>Pyrus pyrifolia</i>)	76 Japanese pear cultivars	162 markers (155 SSRs, 4 RAPD-STS, 2 ACC synthase genes, 1 S-RNase gene)	Nine traits (harvest time, resistance to black spot, firmness of flesh, fruit size, fruit shape in longitudinal section, acid content, total soluble solid content, number of spurs, vigor of tree)	BayesA, BayesB	-0.45 to 0.75	Iwata <i>et al.</i> (2013a)
Oil palm (<i>Elaeis guineensis</i>)	262 individuals (two parental populations involved in reciprocal recurrent selection with 131 individuals each)	265 SSR markers	Eight traits (bunch number, average bunch weight, fruit-to-bunch, pulp-to-fruit, kernel-to-fruit, oil-to-pulp ratios, number of fruits per bunch, average fruit weight)	GBLUP, Bayesian LASSO, Bayesian RR, BayesC π , BayesDr	-0.41 to 0.94	Cros <i>et al.</i> (2015)
Loblolly pine (<i>Pinus taeda</i>)	ca. 800 individuals (seedlings derived by crossing 32 parents in a circular mating design)	4825 SNPs	Two traits at multiple ages (3, 4, and 6 years old for diameter at breast height and 1, 2, 3, 4, and 6 years old for height)	rrBLUP	0.16 to 0.75 (0.63 to 0.75 at 6 years old)	M.F.R. Resende <i>et al.</i> (2012a)
Loblolly pine (<i>P. taeda</i>)	951 individuals (seedlings derived by crossing 32 parents in a circular mating design)	4853 SNPs	17 traits (four traits related to growth, six traits related to development, two traits related to disease resistance, five traits related to wood quality)	rrBLUP, BayesA, BayesC π , Bayesian LASSO, rrBLUP B	0.17 to 0.51	M.D. Resende <i>et al.</i> (2012)
Loblolly pine (<i>P. taeda</i>)	149 clones (full-sib progeny derived from 13 crosses)	3406 SNPs	Four traits (height, volume, lignin and cellulose contents)	rrBLUP including both additive and dominance effects	0.30 to 0.83 when 10% of the clones were selected within each cross; 0.56 to 0.68 when 10% of the clones were selected randomly across all crosses	Zapata-Valenzuela <i>et al.</i> (2012)
Loblolly pine (<i>P. taeda</i>)	165 clones (full-sib progeny from nine crosses)	3461 SNPs	Two traits (height, volume)	GBLUP	0.55 to 0.74 when 10% were selected within each cross	Zapata-Valenzuela <i>et al.</i> (2013)
Eucalyptus (<i>E. grandis</i> , <i>E. urophylla</i> , <i>E. globulus</i> and their F ₁ hybrids)	820 individuals for one population and 920 for the other	3564 SNPs for one population and 3129 for the other	Four traits (tree circumference, height growth, wood specific gravity, pulp yield)	rrBLUP	0.55 to 0.88	M.D. Resende <i>et al.</i> (2012)
White spruce (<i>Picea glauca</i>)	1694 individuals	6385 SNPs	12 traits (cell population, fiber coarseness, crystallite width, wood density, microfibril angle, wood stiffness, ring width, specific fiber surface, cell radial diameter, cell tangential diameter, cell wall thickness, 22-year height)	rrBLUP and rrBLUP with pedigree effects	0.33 to 0.44 when both training and testing data sets share individuals of the same families; 0.13 to 0.28 when training and testing data sets are made up of individuals of different families; -0.05 to 0.13 when families making up the validation data sets are from populations that are not represented in the training data sets	Beaulieu <i>et al.</i> (2014a)
White spruce (<i>P. glauca</i>)	1748 individuals	6932 SNPs	Four traits (average wood density, average microfibril angle, 17-year height, 17-year diameter at breast height)	Bayesian RR, Bayesian LASSO, and those with pedigree effects	0.52 to 0.79 with Bayesian RR when the validation sets were built with individuals within full-sib families and both training and testing sets were from the same breeding group; 0.29 to 0.59 with Bayesian RR when the validation sets were built with full-sib families and both training and testing sets were from the same breeding group	Beaulieu <i>et al.</i> (2014b)

SNP, single nucleotide polymorphism; RR, ridge regression; rrBLUP, random regression best linear unbiased prediction; LASSO, least absolute shrinkage and selection operator; SSR, simple sequence repeat; RAPD-STS, random amplified polymorphic DNA-sequence tagged sites; ACC, 1- aminocyclopropane-1-carboxylate; GBLUP, genomic best linear unbiased prediction.

but the difference in the accuracy between GBLUP and Bayesian LASSO was small.

Iwata *et al.* (2013a) conducted an empirical data analysis of GS in Japanese pear. The authors used 76 cultivars genotyped for 162 markers including 155 SSRs and phenotyped with nine agronomic traits. Because phenotypes were scored as ordinal categorical data, Bayesian methods for estimating latent continuous variation of phenotypes were employed. The level of accuracy of GS prediction was high (0.75) or medium (0.38–0.61) in seven of the nine traits. For fruit quality traits, no significant association was detected by using GWAS, but GS prediction showed medium levels of accuracy except for one trait. In traits for which a significant association was detected in GWAS, GS predictions were more accurate than those based on significant markers, suggesting the traits are determined by several minor and medium QTLs as well as major QTLs.

Cros *et al.* (2015) evaluated the prediction accuracy of GS in oil palm, using two parental populations involved in conventional reciprocal recurrent selection with 131 individuals each and genotyped for 265 SSRs (Table 2). The authors compared five GS modeling methods: GBLUP, Bayesian LASSO, Bayesian ridge regression, BayesC π , and BayesD π . The accuracy of GBLUP was significantly higher than that of PBLUP in three of eight traits in one parental population, but it was equal to that of PBLUP in the other parental population. These results suggest that GS could account for family effects and Mendelian sampling terms in the former population, but only family effects in the latter population. Fewer polymorphic markers and lower marker density may have impaired the advantage of GBLUP over PBLUP in the latter population. Differences in accuracy were small among the five modeling methods.

In forest tree species, empirical data analysis of GS has been performed for loblolly pine (M.F.R. Resende *et al.* 2012a, 2012b, Zapata-Valenzuela *et al.* 2012, 2013), eucalyptus (M.D. Resende *et al.* 2012), and white spruce (Beaulieu *et al.* 2014a, 2014b), as summarized in Table 2. In their pioneering study, M.F.R. Resende *et al.* (2012a) evaluated the accuracy of GS in loblolly pine, using 800 clonally replicated individuals grown at four sites and genotyped for 4825 SNPs. The accuracy of GS ranged from 0.64 to 0.74 and was comparable with or slightly less than that of selection based on PBLUP. The authors evaluated the accuracy of GS models across ages and environments. The model generated at early ages did not perform well in predicting phenotypes at later ages (6 years). The accuracy of models was highest at the sites where the models were generated but declined at different sites, suggesting that genotype \times environment (GE) interactions greatly affect the transferability of models across sites.

M.D. Resende *et al.* (2012) used two breeding populations to evaluate the accuracy of GS in eucalyptus. The two populations contained 43 and 75 families (738 and 920 individuals, respectively) sampled for GS. The accuracy of GS prediction ranged from 0.74 to 0.88 in one population and

from 0.55 to 0.73 in the other. The higher accuracy in the former population was expected because it had a smaller effective population size (11) than the latter one (51). The accuracy of GS was comparable to or greater than that of selection based on PBLUP. GS models showed poor predictability across populations, likely as a result of variable patterns of LD, inconsistent allelic effects, and GE interactions.

The prediction model of GS also can be used for predicting the segregation pattern of target traits in a progeny population. In cross breeding, it is important to select an optimal parental combination that has a high probability of generating offspring with desired characteristics. In fruit trees, the establishment of a segregating population and field evaluation of the population require much time and cost, therefore systematic planning for selecting an optimal combination becomes even more important. Iwata *et al.* (2013b) proposed a novel method for predicting the segregation of traits in a progeny population based on GS prediction models and applied the method to Japanese pear data in a proof-of-concept study. Empirical analysis using an actual breeding population and a simulation study based on real marker data suggested that the segregation of target traits can be predicted with reasonable accuracy, especially for a highly heritable trait. The proposed method can provide objective and quantitative criteria for choosing an optimal parental combination and sufficient breeding population size.

Generation advancement technologies in fruit trees

Most fruit tree crops have a long juvenile phase. For example, apple takes 5–12 years until the transition to the adult phase, whereas peach can start flowering within several years after germination (Hansche 1986, Visser 1964). To reduce the time and cost of fruit breeding, more effective methods for accelerating generation time are desired (van Nocker and Gardiner 2014). Recently, a fast-track breeding system, which is based on controlling the juvenile–adult phase transition by inducing a flowering gene or/and silencing a floral repressor combined with MAS, was developed to shorten the juvenile phase in fruit crops (Flachowsky *et al.* 2007, 2011, Wenzel *et al.* 2013).

In the model dicot *Arabidopsis thaliana*, many factors involved in floral induction have been identified over the last two decades (Amasino 2010). For fast-track breeding, the most promising gene may be *FLOWERING LOCUS T (FT)*. *AtFT* encodes a mobile regulator of flowering under the long-day condition, which matches the characteristics of a florigen (Corbesier *et al.* 2007). The introduction of *Citrus FT (CiFT)* to trifoliate orange drastically shortened the juvenile phase and resulted in flowering as early as 12 weeks in a greenhouse, whereas the first flowering normally occurs after at least several years (Endo *et al.* 2005). Similarly, overexpression of *MdFT1*, an apple *FT*-like gene, induced precocious flowering in apple (Kotoda *et al.* 2010). *FT*-like genes from poplar under the control of the heat-inducible

Gmhsp 17.5-E promoter from soybean were also used for floral induction in apple to avoid the disadvantages of constitutive expression (Wenzel *et al.* 2013).

Several other candidates for floral induction of fruit crops have been identified. *FT* activates various downstream genes, including a floral meristem identity gene, *APETALA1* (*API*; Abe *et al.* 2005, Wigge *et al.* 2005). Transgenic citrus (*Citrus sinensis* × *Poncirus trifoliata*) with constitutive expression of *API* produced fertile flowers and fruits in the first year (Peña *et al.* 2001). *LEAFY* (*LFY*) activates *API* and is also involved in the specification of floral meristem identity (Wagner *et al.* 1999, Weigel *et al.* 1992). Similar to *API* transformants, transgenic citrus expressing *LFY* also showed the early-flowering phenotype (Peña *et al.* 2001). *TERMINAL FLOWER 1* (*TFL1*) encodes an FT-like protein, an antagonist to *LFY* and *API*; it represses the conversion of inflorescence meristems to floral meristems (Liljegren *et al.* 1999). In apple and pear, the silencing of *TFL1*-like genes reduced vegetative growth and resulted in early flowering within several years even during *in vitro* cultivation. (Flachowsky *et al.* 2012, Freiman *et al.* 2012). *FRUITFULL* (*FUL*), a gene closely related to *API*, plays redundant roles in floral meristem identity (Ferrández *et al.* 2000, Litt and Irish 2003). Flachowsky *et al.* (2007, 2011) used *BpMADS4*, a *FUL*-like gene from silver birch, for floral induction of apple and sped up the generation cycle.

In addition to transgenesis, a technique for floral induction of host crops using a plant virus vector has been recently developed. An *Apple latent spherical virus* (ALSV) vector containing *AtFT* induced flowering of 30% of apple seedlings 1.5–2.0 months after inoculation (Yamagishi *et al.* 2011). Virus-induced silencing of *MdTFL1-1*, an apple homolog of *TFL1*, caused early flowering in approximately 10% of inoculated apple seedlings (Sasaki *et al.* 2011). Furthermore, simultaneous induction of *AtFT* and silencing of *MdTFL1* using an ALSV vector more stably induced flowering (90% of seedlings) and resulted in the completion of one life cycle within a year (Yamagishi *et al.* 2014). In early-flowering transgenic lines, transgenes can be eliminated by segregation before producing the final cultivars (van Nocker and Gardiner 2014). On the other hand, the ALSV vector does not influence the genotype and the floral character of the next generation, as ALSV is rarely detected in the successive progenies (Yamagishi *et al.* 2014). In the future, plant virus vector-induced transient induction may be applied to other fruit crops and become more promising for accelerating generation times.

Future perspectives for genomics-assisted breeding in fruit trees

GWAS and GS will become increasingly important methods in future fruit tree breeding and genetics, in combination with the increased throughput and decreased cost of genome-wide SNP genotyping and the improved accuracy and power of statistical methods. GWAS can estimate the location

and effects of QTLs/genes related to target traits, and the estimates can be further used in MAS, GS (Foder *et al.* 2014), and their combinations (van Nocker and Gardiner 2014). GS is also useful for traits controlled by numerous small-effect loci. Even in traits for which no significant association is detected, GS predictions can attain some level of accuracy (Iwata *et al.* 2013a). In addition to GWAS and GS, generation advancement technologies that promote rapid-cycle breeding with short generation time and small plant size will help to streamline fruit tree breeding.

GWAS and GS require phenotypic data and marker genotypic data for analysis and modeling. Collecting phenotypic data for a large number of cultivars/lines, however, is costly because of the problems that have hampered fruit tree breeding, namely the long juvenile phase and large plant size. As suggested by Myles (2013), researchers need to think about how to increase the throughput of their phenotyping strategies and how to establish populations suitable for genetic mapping, MAS, and GS. One way to collect phenotypic data for numerous cultivars/lines is to use breeding populations, which are routinely developed and evaluated in breeding programs. If phenotypic and marker genotypic data can be routinely gathered in breeding programs, the resultant collection will boost the detection power of GWAS and the accuracy of GS. The data collected from breeding populations also will be useful for elucidating functional genomics in plants (Poland 2015). Another possible way is to use marker genotypic data for identifying the optimal subset of cultivars/lines to phenotype, because marker genotyping is less costly than phenotyping. Careful selection of the subset based on an optimization algorithm will increase the accuracy of GS (Cros *et al.* 2015). When there is a subpopulation structure in breeding materials, a core collection that includes representative individuals from all subpopulations can be a good training population to increase the accuracy of GS (Fodor *et al.* 2014). The development of a field-based high-throughput phenotyping system is also necessary for promoting the efficiency of phenotypic data collection in breeding programs (Araus and Cairns 2014, Deery *et al.* 2014, Fiorani and Schurr 2013, Poland 2015, White *et al.* 2012).

The relationships between phenotypes and marker genotypes are sometimes strongly affected by non-genetic factors. Fruit quality traits are controlled by complex genetic systems and are readily influenced by environmental conditions (Chagné *et al.* 2014). Cultivation management, such as rootstock selection, training systems, pruning techniques, and post-harvest treatments, also influence fruit quality and yield (Myles 2013). Thus, genotype × environment × management (GEM) interactions should be taken into account when GWAS and GS are applied to breeding populations. When these interactions have a profound effect, MAS and GS will not be without complications. For example, models generated at one site would not be functional at another site with different environmental conditions (M.F.R. Resende *et al.* 2012a). Fruit tree species may have an advantage from

this viewpoint: because most fruit tree species can be propagated clonally, multi-environmental trials can be performed using a set of the same clones to assess the influence of GE or GEM interactions on GWAS and GS in detail (e.g., M.F.R. Resende *et al.* 2012a). As noted by Heffner *et al.* (2009), however, the genotype of any individual is composed of alleles that have been evaluated in a large number of target environments, and thus it may be possible to keep the accuracy of GS high even in the presence of GE interactions if phenotypic data have been collected across many environments.

Another way to deal with GE and/or GEM interactions is to develop mathematical models describing the pattern of the environmental responses of cultivars/lines. An ecophysiological model is a method for simulating the response of plant phenotypes to environmental factors and for describing the pattern of environmental response of cultivars/lines with parameters of the models. These parameters are expected to reflect the genetic characteristics of cultivars/lines, and they can be used for QTL analysis to dissect the varietal variation in the environmental response. In apple, an ecophysiological phenology model was applied to simulate the flowering time and used to calculate the risk of damage by spring frost under climate change (Eccel *et al.* 2009). This type of approach will be useful for developing new cultivars that can adapt well to future climate change. Because fruit tree cultivars cannot be developed over a short time, it is necessary to develop such new varieties based on a long-term perspective.

The phenotypes of fruit trees are also affected by age-induced factors (i.e., ontogenetic factors) and their interactions with genetic factors. To clarify the genetic, ontogenetic, and environmental effects on phenotypes, it is necessary to separate confounded effects caused by consecutive years of growth (i.e., ontogenetic effect) and climatic years (i.e., environmental effect). A staggered-start experimental design (Loughin 2006) enables us to statistically separate the effects of genetic \times ontogenetic and genetic \times year interactions. The potential of the staggered-start design for fruit tree research was demonstrated in the genetic dissection of apple tree architecture (Segura *et al.* 2008).

Genomics-based approaches and genomics-assisted breeding using these methods will provide significant benefits to the genetic improvement of fruit trees, although there are some obstacles to overcome. Designing breeding programs that make optimal use of genomics-based approaches is an important task for plant breeders, because each species is unique and each requires a tailor-made solution (Lin *et al.* 2014). Stochastic simulations assuming a certain breeding program (e.g., Denis and Bouvet 2013, Iwata *et al.* 2011, Yabe *et al.* 2013, 2014) will be helpful for finding ways to optimize the use of genomics-based approaches in breeding programs under species-specific and/or breeding-program-specific restrictions.

Acknowledgments

This work was supported by a grant from the Ministry of Agriculture, Forestry, and Fisheries of Japan (Genomics-based Technology for Agricultural Improvement, NGB2010) and by Grants-in-Aid for Scientific Research (A), MEXT (No. 25252002), and JST/CREST.

Literature Cited

- Abe, M., Y. Kobayashi, S. Yamamoto, Y. Daimon, A. Yamaguchi, Y. Ikeda, H. Ichinoki, M. Notaguchi, K. Goto and T. Araki (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309: 1052–1056.
- Albrechtsen, A., F.C. Nielsen and R. Nielsen (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.* 27: 2534–2547.
- Amasino, R. (2010) Seasonal and developmental timing of flowering. *Plant J.* 61: 1001–1013.
- Antanaviciute, L., F. Fernández-Fernández, J. Jansen, E. Banchi, K.M. Evans, R. Viola, R. Velasco, J.M. Dunwell, M. Troggio and D.J. Sargent (2012) Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* Infinium whole genome genotyping array. *BMC Genomics* 13: 203.
- Araus, J.L. and J.E. Cairns (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19: 52–61.
- Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko and E.A. Johnson (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.
- Ban, Y., N. Mitani, T. Hayashi, A. Sato, A. Azuma, A. Kono and S. Kobayashi (2014) Exploring quantitative trait loci for anthocyanin content in interspecific hybrid grape (*Vitis labruscana* \times *Vitis vinifera*). *Euphytica* 198: 101–114.
- Beaulieu, J., T. Doerksen, S. Clément, J. MacKay and J. Bousquet (2014a) Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113: 343–352.
- Beaulieu, J., T.K. Doerksen, J. MacKay, A. Rainville and J. Bousquet (2014b) Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics* 15: 1048.
- Bink, M.C., J. Jansen, M. Madduri, R.E. Voorrips, C.E. Durel, A.B. Kouassi, F. Laurens, F. Mathis, C. Gessler, D. Gobbin *et al.* (2014) Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor. Appl. Genet.* 127: 1073–1090.
- Brachi, B., G.P. Morris and J.O. Borevitz (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12: 232.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- Breseghele, F. and M.E. Sorrells (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165–1177.
- Broman, K.W., H. Wu, S. Sen and G.A. Churchill (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.
- Browning, S.R. and B.L. Browning (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome associ-

- ation studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81: 1084–1097.
- Bus, V., H. Bassett, D. Bowatte, D. Chagné, C. Ranatunga, D. Ulluwishewa, C. Wiedow and S. Gardiner (2010) Genome mapping of an apple scab, a powdery mildew and a woolly apple aphid resistance gene from open-pollinated mildew immune selection. *Tree Genet. Genomes* 6: 477–487.
- Calus, M.P.L. (2010) Genomic breeding value prediction: methods and procedures. *Animal* 4: 157–164.
- Cao, K., L. Wang, G. Zhu, W. Fang, C. Chen and J. Luo (2012). Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. *Tree Genet. Genomes* 8: 975–990.
- Catchen, J.M., A. Amores, P. Hohenlohe, W. Cresko and J.H. Postlethwait (2011) Stacks: Building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* 1: 171–182.
- Cevik, V., C.D. Ryder, A. Popovich, K. Manning, G.J. King and G.B. Seymour (2010) A *FRUITFULL*-like gene is associated with genetic variation for fruit flesh firmness in apple (*Malus domestica* Borkh.). *Tree Genet. Genomes* 6: 271–279.
- Chagné, D., R.N. Crowhurst, M. Troggo, M.W. Davey, B. Gilmore, C. Lawley, S. Vanderzande, R.P. Hellens, S. Kumar, A. Cestaro *et al.* (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7: e31745.
- Chagné, D., D. Dayatilake, R. Diack, M. Oliver, H. Ireland, A. Watson, S.E. Gardiner, J.W. Johnston, R.J. Schaffer and S. Tustin (2014) Genetic and environmental control of fruit maturation, dry matter and firmness in apple (*Malus × domestica* Borkh.). *Hortic. Res.* 1: 14046.
- Chia, J.-M. and D. Ware (2011) Sequencing for the cream of the crop. *Nat. Biotechnol.* 29: 138–139.
- Corbesier, L., C. Vincent, S. Jang, F. Fornara, Q. Fan, I. Searle, A. Giakountis, S. Farrona, L. Gissot, C. Turnbull *et al.* (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* 316: 1030–1033.
- Correa, J., M. Mamani, C. Muñoz-Espinoza, D. Laborie, C. Muñoz, M. Pinto and P. Hinrichsen (2014). Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (*Vitis vinifera* L.). *Theor. Appl. Genet.* 127: 1143–1162.
- Costa, F. (2015) MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. *Tree Genet. Genomes* 11: 819.
- Cristofani, M., M.A. Machado and D. Grattapaglia (1999) Genetic linkage maps of *Citrus sunki* Hort. ex. Tan. and *Poncirus trifoliata* (L.) Raf. and mapping of citrus tristeza virus resistance gene. *Euphytica* 109: 25–32.
- Cros, D., M. Denis, L. Sánchez, B. Cochard, A. Flori, T. Durand-Gasselin, B. Nouy, A. Omoré, V. Pomiès, V. Riou *et al.* (2015) Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 128: 397–410.
- Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen and M.L. Blaxter (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12: 499–510.
- Deery, D., J. Jimenez-Berni, H. Jones, X. Sirault and R. Furbank (2014) Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 5: 349–379.
- De Givry, S., M. Bouchez, P. Chabrier, D. Milan and T. Schiex (2005) CARTHAGENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21: 1703–1704.
- Denis, M. and J.-M. Bouvet (2013) Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. *Tree Genet. Genomes* 9: 37–51.
- Dirlwanger, E., E. Graziano, T. Joobeur, F. Garriga-Calderé, P. Cosson, W. Howad and P. Arús (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. USA* 101: 9891–9896.
- Doucleff, M., Y. Jin, F. Gao, S. Riaz, A.F. Krivanek and M.A. Walker (2004) A genetic linkage map of grape, utilizing *Vitis rupestris* and *Vitis arizonica*. *Theor. Appl. Genet.* 109: 1178–1187.
- Eccel, E., R. Rea, A. Caffarra and A. Crisci (2009) Risk of spring frost to apple production under future climate scenarios: the role of phenological acclimation. *Int. J. Biometeorol.* 53: 273–286.
- Edge-Garza, D.A., C.P. Peace and Y. Zhu (2010) Enabling marker-assisted seedling selection in the Washington apple breeding program. *Acta Hortic.* 859: 369–373.
- Eduardo, I., I. Pacheco, G. Chietera, D. Bassi, C. Pozzi, A. Vecchietti and L. Rossini (2011) QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet. Genomes* 7: 323–335.
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.
- Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4: 250–255.
- Endo, T., T. Shimada, H. Fujii, Y. Kobayashi, T. Araki and M. Omura (2005) Ectopic expression of an *FT* homolog from *Citrus* confers an early flowering phenotype on trifoliolate orange (*Poncirus trifoliata* L. Raf.). *Transgenic Res.* 14: 703–712.
- Fan, L., M.-Y. Zhang, Q.-Z. Liu, L.-T. Li, Y. Song, L.-F. Wang, S.-L. Zhang and J. Wu (2013) Transferability of newly developed pear SSR markers to other Rosaceae species. *Plant Mol. Biol. Rep.* 31: 1271–1282.
- Fernández-Fernández, F., L. Antanaviciute, M.M. van Dyk, K.R. Tobutt, K.M. Evans, D.J.G. Rees, J.M. Dunwell and D.J. Sargent (2012) A genetic linkage map of an apple rootstock progeny anchored to the *Malus* genome sequence. *Tree Genet. Genomes* 8: 991–1002.
- Ferrándiz, C., Q. Gu, R. Martienssen and M.F. Yanofsky (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* 127: 725–734.
- Fiorani, F. and U. Schurr (2013) Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64: 267–291.
- Flachowsky, H., A. Peil, T. Sopanen, A. Elo and V. Hanke (2007) Overexpression of *BpMADS4* from silver birch (*Betula pendula* Roth.) induces early-flowering in apple (*Malus × domestica* Borkh.). *Plant Breed.* 126: 137–145.
- Flachowsky, H., P.-M. Le Roux, A. Peil, A. Patocchi, K. Richter and M.-V. Hanke (2011) Application of a high-speed breeding technology to apple (*Malus × domestica*) based on transgenic early flowering plants and marker-assisted selection. *New Phytol.* 192: 364–377.
- Flachowsky, H., I. Szankowski, S. Waidmann, A. Peil, C. Tränkner and M.-V. Hanke (2012) The *MdTFLL1* gene of apple (*Malus × domestica* Borkh.) reduces vegetative growth and generation time. *Tree Physiol.* 32: 1288–1301.
- Fodor, A., V. Segura, M. Denis, S. Neuenschwander, A. Fournier-Level, P. Chatelet, F.A.A. Homa, T. Lacombe, P. This and L.L. Cunff (2014) Genome-wide prediction methods in highly diverse and heterozygous species: Proof-of-concept through simulation in

- grapevine. PLoS ONE 9: e110436.
- Folta, K.M. and S.E. Gardiner (2009) Genetics and genomics of Rosaceae. New York, Springer.
- Freiman, A., L. Shlizerman, S. Golobovitch, Z. Yablovitz, R. Korchinsky, Y. Cohen, A. Samach, E. Chevreau, P.-M. Le Roux, A. Patocchi *et al.* (2012) Development of a transgenic early flowering pear (*Pyrus communis* L.) genotype by RNAi silencing of *PcTFL1-1* and *PcTFL1-2*. *Planta* 235: 1239–1251.
- Gardner, K.M., P. Brown, T.F. Cooke, S. Cann, F. Costa, C. Bustamante, R. Velasco, M. Troggio and S. Myles (2014) Fast and cost-effective genetic mapping in apple using next-generation sequencing. *G3 (Bethesda)* 4: 1681–1687.
- Gasic, K., D.O. Gonzalez, J. Thimmapuram, L. Liu, M. Malnoy, G. Gong, Y.P. Han, L.O. Vodkin, H.S. Aldwinckle, N.J. Carroll *et al.* (2009) Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. *Plant Genome* 2: 23–38.
- Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun and E.S. Buckler (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9: e90346.
- Goddard, M. (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136: 245–257.
- Grattapaglia, D. and R. Sederoff (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.
- Grattapaglia, D., F.L. Bertolucci and R. Sederoff (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Theor. Appl. Genet.* 90: 933–947.
- Grattapaglia, D. and M.D. Resende (2011) Genomic selection in forest tree breeding. *Tree Genet. Genomes* 7: 241–255.
- Haldar, S., S. Haendiges, D.A. Edge-Garza, N.C. Oraguzie, J. Olmstead, C.P. Peace and A. Iezzoni (2010) Applying genetic markers for self-compatibility in the WSU sweet cherry breeding program. *Acta Hort.* 859: 375–380.
- Hamblin, M.T., E.S. Buckler and J.L. Jannink (2011) Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27: 98–106.
- Hansche, P.E. (1986) Heritability of juvenility in peach. *HortScience* 21: 1197–1198.
- Hayashi, T. and T. Awata (2004) Efficient method for analysis of QTL using F1 progenies in an outcrossing species. *Genetica* 122: 173–183.
- Hayes, J., P.J. Bowman, A.J. Chamberlain and M.E. Goddard (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92: 433–443.
- He, J., X. Zhao, A. Laroche, Z.X. Lu, H. Liu and Z. Li (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5: 484.
- Heffner, E.L., M.E. Sorrells and J.L. Jannink (2009) Genomic selection for crop improvement. *Crop Sci.* 49: 1–12.
- Hemmat, M., N.F. Weeden, A.G. Manganaris and D.M. Lawson (1994) Molecular marker linkage map for apple. *J. Hered.* 85: 4–11.
- Heslot, N., H.P. Yang, M.E. Sorrells and J.L. Jannink (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52: 146–160.
- Heslot, N., J. Rutkoski, J. Poland, J.L. Jannink and M.E. Sorrells (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* 8: e74612.
- Iezzoni, A., C. Weebadde, J.J. Luby, C. Yue, W.E. van de Weg, G. Fazio, D. Main, C. Peace, N.V. Bassil and J. McFerson (2010) RosBREED: Enabling marker-assisted breeding in Rosaceae. *Acta Hort.* 859: 389–394.
- Iwata, H. and S. Ninomiya (2006) AntMap: constructing genetic linkage maps using an ant colony optimization algorithm. *Breed. Sci.* 56: 371–377.
- Iwata, H., T. Hayashi and Y. Tsumura (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genet. Genomes* 7: 747–758.
- Iwata, H., T. Hayashi, S. Terakami, N. Takada, Y. Sawamura and T. Yamamoto (2013a) Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed. Sci.* 63: 125–140.
- Iwata, H., T. Hayashi, S. Terakami, N. Takada, T. Saito and T. Yamamoto (2013b) Genomic prediction of trait segregation in a progeny population: a case study of Japanese pear (*Pyrus pyrifolia*). *BMC Genet.* 14: 81.
- Jannink, J.L., A.J. Lorenz and H. Iwata (2010) Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9: 166–177.
- Jansen, R.C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.
- Jung, S., M. Staton, T. Lee, A. Blenda, R. Svancara, A. Abbott and D. Main (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. *Nucleic Acids Res.* 36: D1034–D1040.
- Jung, S., S.P. Ficklin, T. Lee, C.-H. Cheng, A. Blenda, P. Zheng, J. Yu, A. Bombarely, I. Cho, S. Ru *et al.* (2014) The Genome Database for Rosaceae (GDR): year 10 update. *Nucleic Acids Res.* 42: D1237–D1244.
- Kellerhals, M., L. Franck, I.O. Baumgartner, A. Patocchi and J.E. Frey (2011) Breeding for fire blight resistance in apple. *Acta Hort.* 896: 385–389.
- Kenis, K., J. Keulemans and M.W. Davey (2008) Identification and stability of QTLs for fruit quality traits in apple. *Tree Genet. Genomes* 4: 647–661.
- Khan, M.A. and S.S. Korban (2012) Association mapping in forest trees and fruit crops. *J. Exp. Bot.* 63: 4045–4060.
- Knott, S.A., J.M. Elsen and C.S. Haley (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor. Appl. Genet.* 93: 71–80.
- Kotoda, N., H. Hayashi, M. Suzuki, M. Igarashi, Y. Hatsuyama, S. Kidou, T. Igasaki, M. Nishiguchi, K. Yano, T. Shimizu *et al.* (2010) Molecular characterization of *FLOWERING LOCUS T*-like genes of apple (*Malus × domestica* Borkh.). *Plant Cell Physiol.* 51: 561–575.
- Kumar, S., M.C. Bink, R.K. Volz, V.G. Bus and D. Chagné (2012a) Towards genomic selection in apple (*Malus × domestica* Borkh.) breeding programmes: prospects, challenges and strategies. *Tree Genet. Genomes* 8: 1–14.
- Kumar, S., D. Chagné, M.C.A.M. Bink, R.K. Volz, C. Whitworth and C. Carlisle (2012b) Genomic selection for fruit quality traits in apple (*Malus × domestica* Borkh.). *PLoS ONE* 7: e36674.
- Kumar, S., D.J. Garrick, M.C. Bink, C. Whitworth, D. Chagné and R.K. Volz (2013) Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* 14: 393.
- Kunihisa, M., S. Moriya, K. Abe, K. Okada, T. Haji, T. Hayashi, H. Kim, C. Nishitani, S. Terakami and T. Yamamoto (2014) Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. *Breed. Sci.* 64: 240–251.

- Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln and L.A. Newberg (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
- Lander, E.S. and D. Botstein (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Lee, T.H., H. Tang, X. Wang and A.H. Paterson (2013) PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41: D1152–D1158.
- Li, Z.K., S.B. Yu, H.R. Lafitte, N. Huang, B. Courtois, S. Hittalmani, C.H.M. Vijayakumar, G.F. Liu, G.C. Wang, H.E. Shashidhar *et al.* (2003) QTL × environment interactions in rice. I. Heading date and plant height. *Theor. Appl. Genet.* 108: 141–153.
- Liao, C.Y., P. Wu, B. Hu and K.K. Yi (2001) Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. *Theor. Appl. Genet.* 103: 104–111.
- Liljegrén, S.J., C. Gustafson-Brown, A. Pinyopich, G.S. Ditta and M.F. Yanofsky (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* 11: 1007–1018.
- Lin, Z., B.J. Hayes and H.D. Daetwyler (2014). Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 65: 1177–1191.
- Lincoln, S.E. and E.S. Lander (1990) Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL. Technical Report. Whitehead Institute for Biomedical Research, Cambridge.
- Litt, A. and V.F. Irish (2003) Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics* 165: 821–833.
- Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells and J.L. Jannink (2011) Genomic selection in plant breeding: Knowledge and prospects. *Adv. Agron.* 110: 77–123.
- Loughin, T.M. (2006) Improved experimental design and analysis for long-term experiments. *Crop Sci.* 46: 2492–2502.
- Luby, J.J. and D.V. Shaw (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience* 36: 872–879.
- Maliapaard, C., F.H. Alston, G. van Arkel, L.M. Brown, E. Chevreau, F. Dunemann, K.M. Evans, S. Gardiner, P. Guilford, A.W. van Heusden *et al.* (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor. Appl. Genet.* 97: 60–73.
- Marandel, G., J. Salava, A. Abbott, T. Candresse and V. Decroocq (2009) Quantitative trait loci meta-analysis of *Plum pox virus* resistance in apricot (*Prunus armeniaca* L.): new insights on the organization and the identification of genomic resistance factors. *Mol. Plant Pathol.* 10: 347–360.
- Martínez-García, P.J., D.E. Parfitt, E.A. Ogundiwin, J. Fass, H.M. Chan, R. Ahmad, S. Lurie, A. Dandekar, T.M. Gradziel and C.H. Crisosto (2013) High density SNP mapping and QTL analysis for fruit quality characteristics in peach (*Prunus persica* L.). *Tree Genet. Genomes* 9: 19–36.
- Meneses, C. and A. Orellana (2013) Using genomics to improve fruit quality. *Biol. Res.* 46: 347–352.
- Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Miah, G., M.Y. Rafiq, M.R. Ismail, A.B. Puteh, H.A. Rahim, K.N. Islam and M.A. Latif (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *Int. J. Mol. Sci.* 14: 22499–22528.
- Morris, M., K. Dreher, J.M. Ribaut and M. Khairallah (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol. Breed.* 11: 235–247.
- Myles, S., J.M. Chia, B. Hurwitz, C. Simon, G.Y. Zhong, E. Buckler and D. Ware (2010) Rapid genomic characterization of the genus *Vitis*. *PLoS ONE* 5: e8219.
- Myles, S. (2013) Improving fruit and wine: what does genomics have to offer? *Trends Genet.* 29: 190–196.
- Pacheco, I., D. Bassi, I. Eduardo, A. Ciacciulli, R. Pirona, L. Rossini and A. Vecchietti (2014) QTL mapping for brown rot (*Monilinia fructigena*) resistance in an intraspecific peach (*Prunus persica* L. Batsch) F₁ progeny. *Tree Genet. Genomes* 10: 1223–1242.
- Peace, C.P. (2013) RosBREED by the numbers. *RosBREED Q Newsl* 4: 4.
- Peña, L., M. Martín-Trillo, J. Juárez, J.A. Pina, L. Navarro and J.M. Martínez-Zapater (2001) Constitutive expression of *Arabidopsis* *LEAFY* or *APETALA1* genes in citrus reduces their generation time. *Nat. Biotechnol.* 19: 263–267.
- Pérez, P. and G. de los Campos (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198: 483–495.
- Picañol, R., I. Eduardo, M.J. Aranzana, W. Howad, I. Batlle, I. Iglesias, J.M. Alonso and P. Arús (2013) Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. *Euphytica* 190: 279–288.
- Pierantoni, L., L. Dondini, K.-H. Cho, I.-S. Shin, F. Gennari, R. Chiodini, S. Tartarini, S.-J. Kang and S. Sansavini (2007) Pear scab resistance QTLs via a European pear (*Pyrus communis*) linkage map. *Tree Genet. Genome* 3: 311–317.
- Poland, J. (2015) Breeding-assisted genomics. *Curr. Opin. Plant Biol.* 24: 119–124.
- Resende, M.D., M.F. Resende, C.P. Sansaloni, C.D. Petrolí, A.A. Missiaggia, A.M. Aguiar, J.M. Abad, E.K. Takahashi, A.M. Rosado, D.A. Faria *et al.* (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* 194: 116–128.
- Resende, M.F.R. Jr., P. Muñoz, J.J. Acosta, G.F. Peter, J.M. Davis, D. Grattapaglia, M.D.V. Resende and M. Kirst (2012a) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol.* 193: 617–624.
- Resende, M.F.R. Jr., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, E.J. Jokela, T.A. Martin, G.F. Peter and M. Kirst (2012b) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190: 1503–1510.
- Riaz, S., A.C. Tenschler, D.W. Ramming and M.A. Walker (2011) Using a limited mapping strategy to identify major QTLs for resistance to grapevine powdery mildew (*Erysiphe necator*) and their use in marker-assisted breeding. *Theor. Appl. Genet.* 122: 1059–1073.
- Rikkerink, E.H.A., N.C. Oraguzie and S.E. Gardiner (2007) Prospects of association mapping in perennial horticultural crops. *In: Oraguzie, N.C., E.H.A. Rikkerink, S.E. Gardiner and H.N. Silva (eds.) Association mapping in plants.* New York, Springer, pp. 249–269.
- Ru, S., D. Main, K. Evans and C. Peace (2015) Current applications, challenges, and perspectives of marker-assisted seedling selection

- in Rosaceae tree fruit breeding. *Tree Genet. Genomes* 11: 8.
- Sasaki, S., N. Yamagishi and N. Yoshikawa (2011) Efficient virus-induced gene silencing in apple, pear and Japanese pear using *Apple latent spherical virus* vectors. *Plant Methods* 7: 15.
- Sebolt, A. (2013) Breeder profile: Kate Evans. *RosBREED Q Newsl* 4(2): 6–7.
- Segura, V., C. Cilas and E. Costes (2008) Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modelling of repeated spatial and temporal measures. *New Phytol.* 178: 302–314.
- Siviero, A., M. Cristofani, E.L. Furtado, A.A. Garcia, A.S. Coelho and M.A. Machado (2006) Identification of QTLs associated with citrus resistance to *Phytophthora gummosis*. *J. Appl. Genet.* 47: 23–28.
- Soriano, J.M., E.M. Vera-Ruiz, S. Vilanova, J. Martínez-Calvo, G. Llácer, M.L. Badenes and C. Romero (2008) Identification and mapping of a locus conferring plum pox virus resistance in two apricot-improved linkage maps. *Tree Genet. Genomes* 4: 391–402.
- Terakami, S., M. Shoda, Y. Adachi, T. Gonai, M. Kasumi, Y. Sawamura, H. Iketani, K. Kotobuki, A. Patocchi, C. Gessler *et al.* (2006) Genetic mapping of the pear scab resistance gene *Vnk* of Japanese pear cultivar Kinchaku. *Theor. Appl. Genet.* 113: 743–752.
- van Heerden, C.J., P. Burger, A. Vermeulen and R. Prins (2014) Detection of downy and powdery mildew resistance QTL in a ‘Regent’ × ‘RedGlobe’ population. *Euphytica* 200: 281–295.
- van Nocker, S. and S.E. Gardiner (2014) Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. *Hortic. Res.* 1: 14022.
- van Ooijen, J.W. (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.
- van Ooijen, J.W. (2009) MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, Netherlands.
- Varshney, R.K., R. Terauchi and S.R. McCouch (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12: e1001883.
- Veyrieras, J.B., B. Goffinet and A. Charcosset (2007) MetaQTL: A package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinformatics* 8: 49.
- Visser, T. (1964) Juvenile phase and growth of apple and pear seedlings. *Euphytica* 13: 119–129.
- Wagner, D., R.W.M. Sablowski and E.M. Meyerowitz (1999) Transcriptional activation of APETALA1 by LEAFY. *Science* 285: 582–584.
- Wang, S., C.J. Basten and Z.B. Zeng (2010) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.
- Weber, C.A., G.A. Moore, Z. Deng and F.G. Gmitter, Jr. (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* × *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. *J. Amer. Soc. Hort. Sci.* 128: 508–514.
- Wegrzyn, J.L., D. Main, B. Figueroa, M. Choi, J. Yu, D.B. Neale, S. Jung, T. Lee, M. Stanton, P. Zheng *et al.* (2012) Uniform standards for genome databases in forest and fruit trees. *Tree Genet. Genomes* 8: 549–557.
- Weigel, D., J. Alvarez, D.R. Smyth, M.F. Yanofsky and E.M. Meyerowitz (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* 69: 843–859.
- Wenzel, S., H. Flachowsky and M.-V. Hanke (2013) The Fast-track breeding approach can be improved by heat-induced expression of the *FLOWERING LOCUS T* genes from poplar (*Populus trichocarpa*) in apple (*Malus × domestica* Borkh.). *Plant Cell Tissue Organ Cult.* 115: 127–137.
- White, J.W., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, M.M. Conley, K.A. Feldmann, A.N. French, J.T. Heun, D.J. Hunsaker *et al.* (2012) Field-based phenomics for plant genetics research. *Field Crops Res.* 133: 101–112.
- Wigge, P.A., M.C. Kim, K.E. Jaeger, W. Busch, M. Schmid, J.U. Lohmann and D. Weigel (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* 309: 1056–1059.
- Wong, C.K. and R. Bernardo (2008) Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet.* 116: 815–824.
- Yabe, S., R. Ohsawa and H. Iwata (2013) Potential of genomic selection for mass selection breeding in annual allogamous crops. *Crop Sci.* 53: 95–105.
- Yabe, S., R. Ohsawa and H. Iwata (2014) Genomic selection for the traits expressed after pollination in allogamous plants. *Crop. Sci.* 54: 1448–1457.
- Yamagishi, N., S. Sasaki, K. Yamagata, S. Komori, M. Nagase, M. Wada, T. Yamamoto and N. Yoshikawa (2011) Promotion of flowering and reduction of a generation time in apple seedlings by ectopic expression of the *Arabidopsis thaliana FT* gene using the *Apple latent spherical virus* vector. *Plant Mol. Biol.* 75: 193–204.
- Yamagishi, N., R. Kishigami and N. Yoshikawa (2014) Reduced generation time of apple seedlings to within a year by means of a plant virus vector: a new plant-breeding technique with no transmission of genetic modification to the next generation. *Plant Biotechnol. J.* 12: 60–68.
- Yamamoto, T., T. Kimura, M. Shoda, T. Imai, T. Saito, Y. Sawamura, K. Kotobuki, T. Hayashi and N. Matsuta (2002) Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. *Theor. Appl. Genet.* 106: 9–18.
- Yamamoto, T., S. Terakami, N. Takada, S. Nishio, N. Onoue, C. Nishitani, M. Kunihisa, E. Inoue, H. Iwata, T. Hayashi *et al.* (2014) Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci.* 64: 351–361.
- Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland *et al.* (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38: 203–208.
- Zapata-Valenzuela, J., F. Isik, C. Maltecca, J. Wegrzyn, D. Neale, S. McKeand and R. Whetten (2012) SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genet. Genomes* 8: 1307–1318.
- Zapata-Valenzuela, J., R.W. Whetten, D. Neale, S. McKeand and F. Isik (2013) Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3 (Bethesda)* 909–916.
- Zeng, Z.B. (1994) Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.
- Zhang, R.-P., J. Wu, X.-G. Li, M.A. Khan, H. Chen, S.S. Korban and S.-L. Zhang (2013) An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus L.*). *Plant Mol. Biol. Rep.* 31: 678–687.