Translation

Modern molecular genetics technologies in fruit tree breeding Amanov S., Seisenova A., Oleichenko S.

Traditional breeding of fruit trees is difficult due to their long generation time, large plant size, long juvenile phase, and the need to wait for the plant to reach physiological maturity in order to obtain marketable products. Recent advances in genomic technologies have provided new opportunities to improve the efficiency of plant breeding. Genome-wide association studies (GWAS) and genomic selection (GS) are useful, especially in fruit tree breeding. This review shows the expansion of modern breeding capabilities using genomics. New approaches based on genomics significantly expand the possibilities for overcoming barriers in traditional fruit tree breeding. Molecular marker systems and whole genome sequence data that can be used in fruit tree breeding are presented here. Methods for mapping linkage loci and quantitative traits (QTL), as well as GWAS and GS, are presented. Studies on QTL mapping, GWAS, and GS conducted on fruit trees are also presented. Future prospects for fruit tree breeding using genomics are shown, as well as problems that need to be overcome in the breeding process.

Keywords: genome-wide association study (GWAS), genomic selection (GS), statistical model, generation advancement technology.

Introduction

Recent advances in genomic analysis technologies enable genome-wide genotyping of DNA polymorphisms at low cost, high throughput, and high density (Davey et al. 2011). These technological advances have opened up new opportunities for improving the efficiency of plant breeding (Chia and Ware 2011, He et al. 2014, Varshney et al. 2014). Genome-wide association studies (GWAS) allow the detection and identification of quantitative trait loci (QTLs) and genes controlling phenotypic variation in germplasm collections and germplasm (Kooke et al. 2016, Hamblin et al. 2011). Genomic selection (GS; Meuwissen et al. 2001) allows the selection of the best genotypes based on genomic breeding values (GEBV) derived from genome-wide DNA polymorphism information (Heffner et al. 2009, Jannink et al. 2010, Lorenz et al. 2011). These new genomics-based approaches have evolved from traditional biparental QTL mapping and marker-assisted selection (MAS), but they have a much broader range of applications and greater potential than traditional methods. Advances in genomic analysis technologies are leading to the active use of genomics-based approaches for plant breeding and genetic research.

Genomics-based approaches can be particularly useful in the breeding of fruit trees (Chia and Ware 2011, Meneses and Orellana 2013, Myles 2013, van Nocker and Gardiner 2014). Traditional breeding of fruit trees is difficult due to the long generation time, large size, long juvenile phase of seedlings, and marketable appearance that cannot be assessed until the seedling is physiologically mature (Luby and Shaw 2001, Rikkerink et al.

2007). Numerous biotic and abiotic factors affecting fruit quality and quantity during the pre- and post-harvest periods also hinder genetic improvement (Rikkerink et al. 2007). For fruit trees, genomics-based approaches have great potential to overcome these barriers. For example, GWAS allows researchers to estimate the location and effect of QTLs/genes using existing varieties/lines without preparing a segregation population (Khan and Korban 2012).

When phenotypic data are already available for existing varieties/lines, the location and effects of QTLs/genes can be estimated without conducting field experiments. Markers that have a significant association with target traits can be used in MAS programs. GS improves selection accuracy and genetic gain per unit of time spent on genetic improvement of fruit quality and yield traits (Kumar et al. 2012b). Selection in the juvenile phase accelerates the selection process and screens out a significant portion of the progeny that proceeds to evaluation in field trials, which accelerates the selection of fruit trees by reducing selection cycles and increasing selection intensity (Luby and Shaw 2001, Ru et al. 2015). Thus, GWAS and GS are promising methods for improving the efficiency of fruit tree selection.

Molecular markers of fruit trees

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

Simple sequence repeat (SSR) markers are generated from both transcriptional and genomic sequences 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, multiallelic, reproducible, and often amplified between species (Miah et al. 2013). Due to this interspecies transferability, SSR markers have been used for comparative mapping and genome synteny analysis, for example, in Rosaceae species (Dirlewanger et al. 2004, Fan et al. 2013, Gasic et al. 2009).

Single nucleotide polymorphism (SNP) markers are also being developed using

using expressed sequences and genomic sequences. These markers are cost-effective in terms of the 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 trees array (Chagné et al. 2012, Myles et al. 2010). To date, the apple SNP array has been used to create linkage maps (Troggio et al. 2013) and to facilitate GS (Kumar et al. 2012b) and GWAS (Kumar et al. 2013).

Next-generation sequencing (NGS) technologies were used for SNP genotyping using genotyping-by-sequencing (GBS) (Elshire et al. 2011) and restriction site-associated DNA sequencing (Baird et al. 2008) methods. NGS-based genotyping methods allow simultaneous detection of thousands of SNPs across the genome in mapping populations. The advantage of NGS-based genotyping over SNP arrays is that there is no need for SNP discovery and array development. For new target species or populations, SNP array development is a very labor-intensive and costly process. Since NGS-based genotyping methods can be used without a reference genome (Catchen et al. 2011), they are useful for a wide range of organisms, including minor fruit trees. Another advantage of NGS-based genotyping methods is that they are less susceptible to systematic error, which can affect GWAS and GS (McTavish et al. 2015, Heslot et al. 2013). Although NGS-based genotyping typically yields a relatively large number of SNPs, most SNPs have numerous data gaps in the samples (Gardner et al. 2014). One solution to this problem is to use an imputation method to fill in missing genotypes, such as the TASSEL-GBS pipeline (Glaubitz et al. 2014) and Beagle (Browning and Browning 2007).

Fruit genetic databases

Fruit tree genome databases containing genomic, genetic, and breeding resources provide an effective platform for breeding programs. It is important for fruit crop researchers to understand how this data can be used to solve problems in fruit production. Genetic map, marker, and QTL data contribute to the development of markers for genomics—aiding in breeding and the discovery of genes underlying important agricultural traits.

Fruit tree genome sequencing projects have been presented on the Phytozome website (www.phytozome.net), which provides well-curated data on microsatellite density and gene family evolution. The Fruit Tree Genome Database (www.tfgdr.org) is a collection of bioinformatics resources and software tools for the Rosaceae, Citrus, and Vaccinium families (Wegrzyn et al. 2012). The Rosaceae Genome Database (www.rosaceae.org) is a central repository of genetic and breeding data and analysis tools for the Rosaceae family. It has publicly available tools for breeders, integrated with existing genomic and genetic data (Jung et al. 2014). The Citrus Genome Database (www.citrusgenomedb.org) is a public database providing access to research in the field of citrus genomics, genetics, and breeding (Jung et al. 2008). The Vaccinium Genome Database (www.vaccinium.org) is being developed to store and integrate genomic, genetic, and breeding data on blueberries, cranberries, and other Vaccinium species. The duplication duplication of (chibba.agtec.uga.edu/duplication/) is a a web service that provides information on intragenomic or intergenomic syntenic relationships (Lee et al. 2013), which helps identify conserved genes between species. The development of genetic and genomic resources will contribute to genomic research and the application of genomics as an aid in breeding.

Note added ([A1]): Synteny—the possession of a common chromosome sequence. Models of common synteny or damage synteny
can also be used as symbols to reveal phylogenetic relationships between several species, and even to reveal the genome of extinct organisms of ancestral species.
A qualitative distinction is sometimes made between macrosyntony, preserved synteny in large regions of a chromosome, and microsyntony, preserved synteny of only a few genes at a time

Methods of genetic analysis of fruit trees

QTL Analysis

Many economically important traits of fruit trees, such as fruit quality and yield, are quantitative traits, and the loci that influence these traits are called QTLs. To map QTLs and evaluate their influence, a statistical method is used that combines linkage analysis with a statistical model of phenotypic trait values, called QTL analysis. As in linkage analysis, in QTL analysis of fruit trees, a full-sib family consisting of two parents, T1 and T2, and their F1 progeny is typically used as the population under analysis (e.g., Ban et al. 2014, Kunihisa et al. 2014, Siviero et al. 2006, Weber et al. 2003, Yamamoto et al. 2014). Since the QTL analysis procedure is partly based on the linkage analysis procedure, the pseudo-testcrossing strategy can also be used for QTL analysis in a full-sib family (Grattapaglia et al. 1995), with QTL analysis performed separately for T1 and T2, and F1 offspring considered as half-siblings derived from each T1 and T2. When pseudo-testcross analysis is applied to both parents sequentially, this process analysis QTL is called a two-sided pseudo-testcross.

(Grattapaglia et al. 1995), as in linkage analysis. In pseudo-testcross QTL analysis, heterozygous QTLs in the parent are the targets of QTL detection, and using the parent's linkage map, the analysis is performed in the same way as in the backcross design for QTL analysis in inbred species, except that the haplotypes for the parent's marker genotypes must be derived. Since interval mapping (Lander and Botstein 1989) is commonly used for QTL analysis of BC populations in inbred species, we will explain the application of this method to the pseudo-testcross design.

In interval mapping, any position on the linkage map is tested for the presence of a QTL affecting the trait. The putative QTL, Q, is located at the tested position between two markers, A and B, on the linkage map, and the effect of Q is estimated. Consider a pseudo-testcross strategy for QTL analysis for T1 in a full-sib progeny derived from crossing T1 and T2. Suppose that the QTL genotype T1 in Q is Q1Q2, where Q1 and Q2 are different alleles of Q. In the F1 progeny, which includes n individuals, the phenotypic value of the trait for the i-th F1 individual, yi, is expressed as a linear model:

$$yi = \mu + uia + ei, \quad (1)$$

where μ is the model intercept, ui is the covariate indicating the Q allele, transmitted to the i-th individual F1 from T1 (taking values 1 and 0 for alleles Q1 and Q2, respectively), a is the difference in allelic effects between Q1 and Q2, and ei is the remainder. The allele that individual F1 receives from T1 (i.e., Q1 or Q2) is derived from the alleles of flanking markers A and B passed from T1 to individual F1, where it is assumed that the T1 haplotype with respect to A and B is known. The model parameters, including a and μ, are estimated so that the parameter values fit the data of the phenotypic values of n F1 individuals well within the model. The presence of a QTL at the position of interest is assessed by comparing the goodness of fit of the model between two hypotheses, H1: $a \neq 0$ and H0: $a = 0$, representing the presence and absence of a QTL, respectively (Lander and Botstein 1989). Test statistics such as the LOD score are used to test H0 against H1. The influence of QTLs on regions other than the tested position can be included in model (1) to increase the power of QTL detection and the accuracy of QTL position and effect estimation (Jansen 1993, Zeng 1994).

The interval mapping method has been 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

Note added ([A2]): In statistics, a covariate is an additional variable that may influence the dependent variable in a study, in addition to those that are the main objects of study. It is used to control or account for the influence of these additional variables in order to more accurately assess the effect of the factors under study.

MapQTL (van Ooijen, 2009). Pseudo-testcross QTL analysis is performed using the backcross option in these software packages using F1 offspring phenotype and marker genotype data recorded according to the haplotype of one of the parents in a full-sib family.

Based on an integrated linkage map constructed from the two parental maps, QTL analysis of full-sib offspring can be performed to investigate the presence of heterozygous QTLs in one or both parents. QTL analysis of full-sib families with interval mapping is implemented in MapQTL (van Ooijen 2009). Assuming that the T1 and T2 genotypes in the putative QTL, Q, located at the tested position are Q1Q2 and Q3Q4, respectively, the interval mapping model for the analysis of full-sib families is as follows:

$$yi = \mu + uia + vib + ei, \quad (2)$$

where μ, ui, a, and ei are the same terms as in the pseudo-test cross model (1), vi is a covariate indicating which of the alleles in Q (Q3 or Q4) is transmitted to the i-th F1 individual from T2, and b is the difference in allele effects between Q3 and Q4. Analysis of full siblings using model (2) requires knowledge of the marker haplotypes for T1 and T2. The procedure for inferring marker haplotypes can be incorporated into QTL analysis (Hayashi and Awata 2004, Knott et al. 1996). In the breeding of fruit trees such as apple trees, several full-sib families obtained by crossing existing varieties are created as the breeding population. QTL information obtained from QTL analysis performed individually for each of several full-sib families can be integrated to improve the reliability of QTL location and effect estimates. Such an integrated QTL analysis was performed for Costa apple fruit quality traits (2015), where four full-sib families were individually analyzed using model (2), and then the QTLs detected in the individual analyses were projected onto a consensus map of the four families to identify reliable genomic regions of QTLs. This integrated analysis using QTL information for multiple families is called MetaQTL analysis (Costa 2015, Veyrieras et al. 2007).

The parental varieties used for crossing to produce multiple full-sib offspring may be genetically related due to their common ancestral varieties. Adding such ancestral varieties, several

full-sib families are considered as a large complex pedigree. Analysis of such a large pedigree allows for a more accurate assessment of QTL positions and effects.

Bink et al. (2014) performed a QTL analysis on a pedigree comprising 27 full-sib families derived from crosses between 33 parental varieties containing a total of 1,300 individuals and all ancestral varieties of the parental varieties. Ancestral varieties for which both parents were unknown were considered 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 jth individual of the ith family of full siblings, $y_{(ij),}$as:

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

where μ is the model intercept, N is the number of QTLs included in the model, $u_{ijl}$is the covariate indicating the genotype of the l-th QTL for an individual, $a_l$is the effect of the l-th QTL, and $e_{ij}$is the residual. Bayesian estimation was used to estimate the model parameters, taking into account the prior probabilities for the number of QTL N, the positions and effects of QTL, and the genotypes of the founder QTL. Considering the genotypes of the founder QTL, possible transfers of alleles from the founders to the parental varieties through all ancestors were modeled. Post-distributions of model parameters, including N , QT L positions and effects, and founder QTL genotypes, were obtained using Markov-Monte Carlo chain iterations (Di Guardo et al. 2017).

Linkage analysis

The purpose of linkage analysis is to estimate the frequency of recombination between markers or between markers and loci affecting a trait; the analysis involves constructing a linkage map and mapping genes that determine simple inherited traits using multigenerational families. In fruit trees, a full-sib family consisting of two parental varieties (hereinafter T1 and T2) and their F1 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), since the inclusion of more than two generations is difficult due to the long generation time. In a two-generation family, recombination events occur during gametogenesis in

each parent, and the recombination frequency is estimated based on gametes passed from parents to F1 individuals. Consequently, linkage analysis is performed separately for T1 and T2, considering the F1 offspring as hypothetical offspring of half-siblings obtained from each T1 and T2 and the other parent as an anonymous mating partner. Linkage analysis in the half-sib design is performed similarly to the analysis for a backcross population obtained by crossing two inbred lines, called a BC population. For a half-sib family consisting of T1 and F1 offspring, considering T2 as an anonymous mating partner of T1, heterozygous markers or loci in T1 can be used to estimate the recombination frequency. When applying backcross analysis to this half-sib family, T1 and T2 are considered as the F1 individual and the recurrent parent in the backcross design, respectively.

It should be noted, however, that in a half-sib family, it is necessary to distinguish between recombinant and non-recombinant types of gametes transmitted from T1 to F1 offspring with respect to the two linked markers, since the linkage phase in T1 is usually unknown for these two markers. Consider two linked markers, A and B, and assume that the T1 genotype is A1A2B1B2, where A1 (B1) and A2 (B2) are two different alleles of A (B). Given the linkage phase, there are two possible diplotypes for A1A2B1B2: A1B1/A2B2 and A1B2/A2B1, where a diplotype is a genotype that includes information about the linkage phase. If an F1 individual receives a gamete with haplotype A1B1 from T1, this gamete is classified as non-recombining or recombining according to the two possible diplotypes of T1, A1B1/A2B2 and A1B2/A2B1, respectively.

Diplotype T1 is determined based on the number of gametes with haplotypes A1B1 or A2B2, n1, and gametes with haplotypes A1B2 or A2B1, n2, transmitted from T1 to F1. For example, the T1 diplotype is determined as A1B1/A2B2 when $n1 > n2$, and as A1B2/A2B1 otherwise. Given the T1 diplotype for linked markers, linkage analysis is performed based on the genotypes of F1 individuals in the same way as in the backcross scheme.

This linkage analysis strategy for a family of half-sibs from one parent and F1 progeny included in a family of full siblings is called the pseudo-testcrossing strategy (Grattapaglia and Sederoff 1994). The same process is used for linkage analysis of gametes from T2. Bilateral pseudo-testcrossing involves performing pseudo-testcrossing linkage analysis for each of the two parents in turn (Grattapaglia and Sederoff 1994).

Backcross analysis is implemented in most linkage analysis programs

such as MAPMAKER (Lander et al. 1987), AntMap (Iwata and Ninomiya 2006), CARTHAGENE (De Givry et al. 2005), and JoinMap (van Ooijen 2006). Thus, these software packages can be used to construct a linkage map based on a pseudo-testcross design for a family of full siblings. When analyzing linkage using a backcross design, the marker genotype for an individual in the BC population is recorded as "A" or "H" depending on which of the two parental inbred lines is the source of the allele transmitted to the F1 individual BC. In a pseudo-testcross design, T1 is considered to be the F1 of the backcross, but the parental origin of the two alleles possessed by T1 is 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 the F1 offspring are recorded in two different ways. Assuming that the T1 genotype for a marker, for example, A1A2, the genotype of an F1 individual is recorded as "A" for the marker if it receives allele A1 from T1, and as "H" otherwise, considering the genotype A1A2 as A1/A2, where the left side of the slash is the maternal allele and the right side
is the paternal allele. Then, for the same marker, the genotype of the F1 individual is recorded again, but this time "A" and "H" in the original genotype record are swapped, considering A1A2 as A2/A1. The two sets of genotype data for the marker obtained in this way are considered as data for two different markers, so that the number of markers considered is doubled. When identifying linkage groups for these duplicate markers, a set of linkage groups consisting of pairs of equivalent linkage groups is obtained. Within a linkage group, the genotype data records for the markers correspondingly reflect the T1 diplotype. One linkage group is selected from the pair of equivalent linkage groups to construct a marker map for the linkage group. By repeating this process alternately for each pair of linkage groups, linkage maps are obtained for all linkage groups.

Applying a two-way pseudo-testcross strategy to a full-sib family yields two separate linkage maps for the female and male parents. Using multiallelic markers, such as SSR markers with more than two alleles segregating in F1 individuals, heterozygous markers in both parents can be mapped onto both linkage maps, which can be used to align the female and male linkage maps (Maliepaard et al. 1998).. Using JoinMap software (van Ooijen 2006), the two parental linkage maps can be integrated.

Genomic association study and genomic selection

Associative analysis is a method for finding connections between markers and loci that influence a trait by assessing the correlation between marker genotypes and phenotypes in the analyzed population. Using a large number of markers covering the entire genome, GWAS allows such connections to be confirmed for most segments of the genome. In GWAS, segments of the genome of the individuals being analyzed are differentiated based on the allelic states of markers located on the segments, whereas in QTL analysis, they are differentiated based on their parental origin in the family being analyzed. Thus, GWAS analyzes diverse populations, including a mixture of individuals selected from wild populations, germplasm, and breeding lines, in addition to multigenerational families derived from crosses of parental lines, as used in QTL analysis.

In both GWAS and QTL analysis, the successful detection of genomic regions associated with phenotypes depends on the availability of markers associated with loci that influence the trait. The linkage between markers and loci in parental genomes is weakened by recombination events occurring with each generation, but in QTL analysis, the weakening of linkage is negligible due to the limited number of recombination events in a few generations. The existence of a link between markers and loci in the analyzed population, caused by linkage or some other factors such as selection throughout the history of the population, is called linkage disequilibrium (LD). The length of the genome region 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 multigenerational families from crossing parental varieties, the number of markers needed to cover the entire genome is small, whereas a large number of markers are required to cover the genome in populations with lower LD, such as a collection of distantly related individuals, including a germplasm collection. Mapping resolution is also affected by the degree of LD: the lower the degree of LD, the higher the resolution.

Information on several thousand to tens of thousands of SNPs distributed throughout the genome at high density has recently become available for fruit trees. Therefore, GWAS can be an effective method for mapping loci that influence a trait in a collection of varieties without the need to create experimentally crossed populations, which is difficult in the case of most fruit trees due to their long generation time, thereby replacing family QTL analysis (Khan and Korban 2012). Although GWAS is applicable to a wide range of populations, factors such as population stratification and hidden relatedness between individuals can lead to an increase in the frequency of false positives, i.e., false associations between genotype and

phenotype (Kumar et al. 2013). Thus, GWAS statistical models need to include correction factors for population structure and relatedness to control the frequency of false positives.

A mixed linear model is a statistical model that takes into account population structure and kinship and is suitable for GWAS (Yu et al. 2005). In this model, each marker is tested sequentially for association with a trait, with population structure and kinship considered as fixed and random effects, respectively. Consider a population consisting of n individuals with phenotype and genotype records for p markers covering the entire genome, and denote the phenotypic value of the i-th individual as $y_{(i)}$ ($i = 1, 2, \ldots, n$). The following model is used to test the l-th marker in a mixed linear model:

$$y_i = \mu + z_{il} a_l + \sum x_{ij} c_i + g_i + e_i . \qquad (4)$$

where μ is the model intercept, z_{il} is the covariate indicating the genotype of the i-th individual for the l-th marker (e.g., taking values 0, 1, and 2 for the three possible genotypes when the marker is biallelic), a_l is the effect of the l-th marker, x_{ij} is the covariate linking c_j (i.e., the effect of the j-th non-genetic factor) with y_i, g_i is the genotypic value of the i-th individual due to a polygenic effect not captured by the tested marker, and e_i is the residual. When estimating the model parameters, $a_{(l)}$ and $c_{(j)}$ are considered as fixed effects, and $g_{(i)}$ is considered 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, denoted collectively as $g = (g_1, g_2, \ldots, g_{(n)})'$, where the symbol "prim" denotes a transposed vector, are assumed to follow a multivariate normal distribution with mean vector 0 and covariance matrix $A\sigma_{(g)}^{(2)}$, where A is a kinship matrix, whose elements (j,k) denote the genetic relationship between the j-th and k-th individuals. Kinship ties are taken into account when constructing A using information about pedigrees or marker genotypes. Assuming that g is estimated for all markers, we can write $g = Za$, where a is the vector of effects of all markers, and Z is the design matrix linking a with g, where the (i,m)-th element $z_{(im)}$ is the covariate indicating the genotype of the i-th individual for the m-th marker. Thus, Z is considered as information about marker genotypes. Assuming that a is a random effect following a multivariate normal distribution with mean 0 and covariance matrix $I\sigma_{(a)}^{(2)}$, where I is the identity matrix, A is expressed as

$A = ZZ'\sigma_a^2/\sigma_g^2$, which indicates how marker genotypes are used to construct A (Goddard 2009). Since g is a polygenic effect not captured by the lth marker in model (4), $z_{(il)}$ and $a_{(l)}$ can be excluded from Z and a. When the number of markers p is very large, as in GWAS, including $a_{(l)}$ in g gives almost the same result as GWAS with $a_{(l)}$ excluded from g.

Using vector and matrix forms, model (4) can be rewritten as: $y=Xb+z_l a_l$

$$+g+e, \qquad (5)$$

where b is a vector of fixed effects containing the intersection (μ) and non-genetic effects (cj), zl = (z1l, z2l,…, znl)' is a covariate vector indicating the genotypes of the l-th marker for n individuals, g is the vector of random polygenic effects of n individuals, as described above, and e = (e1, e2, …, en)' is the vector of residual values. Testing the significance of the association of each tested marker with the phenotype is based on estimates al, âl, where a statistical test of hypotheses H0: al = 0 against H1: al ≠ 0 is performed. In a typical GWAS procedure with a mixed linear model, the P-value of the estimate âl is calculated at H0, and –log10(P-value) is plotted as a function of the position of each marker to identify markers significantly associated with the phenotype; this plot is called a Manhattan plot. When H0 is rejected, the association between the tested marker and the phenotype is considered significant, and a marker influencing the trait is detected. Parameter estimation in the mixed linear model (5) can be performed using several software packages, including the rrBLUP package in R (Endelman 2011) and TASSEL (Bradbury et al. 2007).

Replacing g with Za and assuming that zlal is included in g in model (5), we obtain a modified model:

$$y = Xb + Za + e. \qquad (6)$$

All markers are simultaneously fitted using model (6) for GWAS. The number of markers, p, often greatly exceeds the number of individuals analyzed, n, when high-density SNP markers are used. To simultaneously estimate the effects of all markers, a, in this situation where p > n, model (5) is considered as a mixed linear model, where a is treated as random effects, and the best linear unbiased prediction (BLUP) method or Bayesian estimation with an a priori distribution for a is used to estimate a. Iwata et al. (2013a) applied Bayesian estimation for GWAS in a Japanese pear population using genomic SSR markers based on model (5) with the inclusion of a variable selection procedure. To date, many types of Bayesian methods have been proposed (e.g., Bayes A , Bayes B , Bayes Cπ, Bayes Dπ, Bayesian

LASSO), each of which depends on the prior distributions adopted for a and other parameters, as well as on estimation procedures with or without variable selection. Several Bayesian estimation methods for model (6) can be implemented using the BGLR package in R (Pérez and de los Campos 2014). GS is an individual selection method based on predicting the genotypic value of a target trait based on genomic marker genotypes (Meuwissen et al. 2001). From the perspective of artificial selection, genotypic value is called breeding value. Breeding values for individuals with marker genotypes in GS are predicted using the same statistical models as in GWAS, where the genotypes and phenotypes of a set of individuals, called the training population, are used to estimate the model parameters. Breeding values for selection candidates are predicted based solely on their genotypes using a prediction model constructed in this way.

Using the same notation as in GWAS, the goal of GS is to predict g for selection candidates based on information about the genotypes of markers Z. The predicted value of g, denoted as $\hat{g}$, is called the genomic estimated breeding value (GEBV). Since we denote g = Za taking into account the effects of all available markers (a), GEBV is obtained as $\hat{g} = Z\hat{a}$, where â is the estimate of a calculated from model (5) using the marker genotype and phenotype data of the training population. Based on the model obtained by excluding the term zlal from model (5), GEBV $\hat{g}$ is calculated directly using the BLUP method without estimating a, where the kinship matrix A is calculated using marker genotypes Z . The BLUP method based on the kinship matrix obtained using marker genotypes is called genomic BLUP (GBLUP; van Raden 2008). Bayesian methods are also used to estimate breeding values g , where a certain prior distributions are assigned to a . Kumar et al. (2012b) applied both BLUP and Bayesian LASSO to predict GEBV for apple fruit quality traits. Instead of using statistical linear models such as model (6) to predict GEBV, some machine learning algorithms were also applied in GS to discriminate objects based on a large number of traits, including the support vector method
and random forest were also applied to predict GEBV in GS
(Jannink et al. 2010).

QTL mapping and GWAS in fruit trees

Most traits important for agronomy and horticulture, such as fruit quality, are quantitative and controlled by several, sometimes numerous, genes or QTL. Many genes and QTL responsible for disease and pest resistance have been reported, including genes

Note added ([A3]): Random forest is a machine learning algorithm that consists of a set of separate decision trees, i.e., independent models.
trees, i.e., independent models.

responsible for resistance to apple scab (Bus et al. 2010) and pear scab (Pierantoni et al. 2007, Terakami et al. 2006), resistance to plum pox virus in apricot (Soriano et al. 2008), resistance to brown rot in peach (Pacheco et al. 2014), and resistance to downy mildew and powdery mildew 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 usually performed using a single full-sib family, which highlights the instability of QTLs among different genetic backgrounds (Kenis et al. 2008). To overcome this limitation of QTL mapping, MetaQTL analysis is used, a new tool that allows the integration of results from several independent QTL mapping experiments (Veyrieras et al. 2007). MetaQTL analysis was performed to determine apricot resistance to plum pox virus (Marandel et al. 2009) and apple quality traits (Costa 2015).

Compared to QTL mapping, GWAS is more suitable for detecting QTL in fruit trees because it does not require bi-parental populations. Creating segregating populations derived from bi-parental crosses of fruit trees is a difficult and costly task due to their long juvenile period (Khan and Korban 2012, Rikkerink et al. 2007). There are few studies on the genetic determination of quantitative traits in fruit trees using GWAS (Table 1). In apple trees, Kumar et al. (2013) conducted a GWAS using 1,200 seedlings from seven full-sib families to identify significant associations of 2,500 SNPs with six fruit traits. Significant associations were found for all six traits in genomic regions, some of which coincided with known candidate genes. Iwata et al. (2013a) conducted a GWAS using 76 Japanese pear varieties and found significant associations
162 markers with nine agronomic traits. There were significant associations significant associations for harvest time, resistance to black spot, and number of spurs. In peach, Cao et al. (2012) conducted a GWAS using 104 genotypes of local varieties genotyped with 53 genomic SSR markers and found associated markers for 10 traits related to fruit and phenological period. Studies of associations with candidate genes were conducted on apple trees and contributed to the discovery of a gene controlling fruit flesh firmness (Cevik et al. 2010). A significant overlap of QTL mapping results and association studies for disease resistance and fruit quality traits was

observed 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 prospects of MAS

Although advances in genome research have provided vast amounts of data (e.g., genetic and physical maps, QTLs, and numerous molecular markers for many traits), there are only a few reports on the application of MAS (Folta and Gardiner 2009, Ru et al. 2015). Problems, including technical and economic barriers, hinder the integration of MAS technology into traditional plant breeding. Selecting reliable markers from among reliable markers is a very difficult task for breeders. The LD between a marker and a trait locus can vary in each population (Breseghello and Sorrells 2006), and the effect of QTLs can vary depending on the genetic background and environmental conditions (Li et al. 2003, Liao et al. 2001). Furthermore, MAS is not always more cost-effective than traditional breeding, as the cost of MAS depends on the target traits (e.g., monogenic or polygenic) and the ease of their measurement (Morris et al. 2003).

Note added ([A4]): LD (linkage disequilibrium) between a marker and a trait locus is non-random co-inheritance of marker and trait alleles (e.g., disease) in a population, which arises due to their physical proximity on the chromosome. The closer the marker and trait locus are located, the higher the probability of their linkage and the stronger the LD.

To address these issues, the RosBREED project was established in 2009, funded by the USDA Specialty Crop Research Initiative (Iezzoni et al. 2010). The project implemented an eight-step marker-assisted selection pipeline, enabling marker-assisted selection for rosaceous crops. Subsequently, MAS was successfully applied to apple (Edge-Garza et al. 2010, Kellerhals et al. 2011, Peace 2013, Sebolt 2013) and cherry (Haldar et al. 2010).

The application of MAS is expanding, but its technical limitations are also becoming apparent. MAS is effective for improving traits controlled by a small number of major genes and/or QTL with large effects (e.g., pest and disease resistance), while 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). A new technology of genomic selection has been proposed to overcome these limitations using whole-genome prediction models based on whole-genome markers (Meuwissen et al. 2001).

Genomic selection in fruit crops

Compared to MAS, GS is more suitable for selecting traits controlled by many minor genes, and there is no need to create crossing populations to develop markers. In GS, selection decisions

selection decisions are based on the predicted GEBV of candidates for selection. In recent years, new sequencing technologies have reduced the cost of SNP genotyping and led to greater availability of numerous markers, which will improve prediction accuracy. GS was first described in cattle breeding programs (Calus 2010, Hayes et al. 2009) and has since been used in crop breeding (Heffner et al. 2009, Lin et al. 2014).

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

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

Iwata et al. (2011) conducted stochastic modeling to evaluate the effectiveness of GS in forest tree breeding, and their conclusion coincided with that of Grattapaglia and Resende (2011). That is, when using a base population derived from a limited number (=25) of elite trees (i.e., a small effective population size), GS had an advantage over phenotypic selection even for a polygenic trait with low heritability at moderate marker density (1 marker/cM). Modeling also shows that updating the prediction model is necessary to achieve large genetic gains from GS selection, as the nature of LD changes with increasing selection cycles. (2011). That is, GS achieved two to three times greater genetic gain per unit of time than phenotypic selection, although the gain per cycle decreased in later selection cycles. The authors also compared the effectiveness of GS models with and without dominance effects. The model with dominance effects showed better

results in clone selection when heritability was high and dominance effects prevailed, but no improvement was found in parent selection.

Fodor et al. (2014) conducted stochastic modeling to evaluate the effectiveness of GS in grapevine breeding. The authors simulated the domestication history of three groups of grape diversity and evaluated the accuracy of GS in simulated breeding populations. High levels of accuracy were obtained using the main collection covering the three diversity groups as the training population, and the highest prediction accuracy was achieved when GWAS and GS were combined.

Wong and Bernardo (2008) simulated breeding using GS in oil palm and demonstrated the superiority of GS over recurrent marker-assisted selection and phenotypic selection in terms of genetic gain per unit of cost and time. Although their modeling results provide a basis for tree selection, their results are not easily generalized because they modeled breeding populations resulting from self-pollination of a hybrid between two inbred lines, whereas real oil palm breeding populations are more complex (Cros et al. 2015).

With regard to fruit trees, empirical data on apple trees (Kumar et al. 2012b) and Japanese pears (Iwata et al. 2013a) were analyzed and summarized in Table 2. Kumar et al. (2012b) evaluated GS accuracy by analyzing seven full-sib families with 1,120 individuals whose genotypes were determined by 2,500 SNPs on the International RosBREED SNP Consortium apple Infinium array v1. GS accuracy ranged from 0.68 to 0.89 and was higher than PBLUP accuracy for all traits. GS allows modeling of variation caused by random sampling of two possible alleles from each parent at each locus during meiosis (so-called Mendelian sampling), whereas PBLUP cannot account for such variation. The higher accuracy of GS suggests that it can account for both family effects and Mendelian sampling. Two modeling methods, GBLUP and Bayesian LASSO, were compared, and Fodor et al. (2014) conducted stochastic modeling to evaluate the effectiveness of GS for grapevine breeding. The authors modeled the domestication history of three groups of grape diversity and evaluated the accuracy of GS in simulated breeding populations. High levels of accuracy were obtained using the main collection covering the three diversity groups as the training population, and the highest prediction accuracy was achieved when GWAS and GS were combined.

Wong and Bernardo (2008) modeled GS selection in oil palm and

Note added ([A5]): I did not insert the table; see the original in English.

demonstrated the superiority of GS over recurrent marker-assisted selection and phenotypic selection in terms of genetic gain per unit of cost and time. Although their modeling results provide a basis for tree breeding, their results are not easily generalized because they modeled breeding populations resulting from self-pollination of a hybrid between two inbred lines, whereas actual oil palm breeding populations are more complex (Cros et al. 2015).

With regard to fruit trees, empirical data on apple trees (Kumar et al. 2012b) and Japanese pears (Iwata et al. 2013a) were analyzed and summarized in Table 2. Kumar et al. (2012b) evaluated the accuracy of GS by analyzing seven full-sib families with 1,120 individuals, whose genotypes were determined by 2,500 SNPs on the International RosBREED SNP Consortium apple Infinium array v1 matrix. GS accuracy ranged from 0.68 to 0.89 and was higher than PBLUP accuracy for all traits. GS allows modeling variation caused by random sampling of two possible alleles from each parent at each locus during meiosis (so-called Mendelian sampling), whereas PBLUP cannot account for such variation. The higher accuracy of GS suggests that it can account for both family effects and Mendelian sampling. Two modeling methods, GBLUP and Bayesian LASSO, were compared, but the difference in accuracy between GBLUP and Bayesian LASSO was small.

Iwata et al. (2013a) conducted an empirical analysis of GS data for Japanese pear. The authors used 76 varieties genotyped for 162 markers, including 155 SSRs, and phenotyped for nine agronomic traits. Since phenotypes were evaluated as ordinal categorical data, Bayesian methods were used to estimate the latent continuous variation of phenotypes. The accuracy of GS predictions was high (0.75) or moderate (0.38–0.61) for seven of the nine traits. No significant associations were found for fruit quality traits using GWAS, but GS predictions showed moderate accuracy, except for one trait. For traits with significant associations in GWAS, GS predictions were more accurate than predictions based on significant markers, suggesting that these traits are determined by several minor and moderate QTLs as well as major QTLs.

Cros et al. (2015) evaluated the accuracy of GS prediction in oil palm using two parental populations involved in traditional reciprocal recurrent selection, each consisting of 131 individuals and genotyped for 265 SSRs. The authors compared five GS modeling methods: GBLUP, Bayesian LASSO, Bayesian ridge regression, BayesCπ, and BayesDπ. The accuracy of GBLUP was significantly higher than PBLUP for three of the eight

traits in one parental population, but equal to PBLUP in the other parental population. These results show that GS can account for family effects and Mendelian sampling terms in the first population, but only family effects in the second population. Fewer polymorphic markers and lower marker density may have reduced the advantage of GBLUP over PBLUP in the second population. Differences in accuracy between the five modeling methods were small.

With regard to forest tree species, empirical analysis of GS data was conducted 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. 2014, Chen et al. 2018), as shown in Table 2. In their pioneering study, M.F.R. Resende et al. (2012a) evaluated GS accuracy in loblolly pine using 800 clonally reproduced individuals grown at four sites and genotyped at 4825 SNPs. GS accuracy ranged from 0.64 to 0.74 and was comparable to or slightly lower than PBLUP-based selection accuracy. The authors evaluated the accuracy of GS models for different ages and environmental conditions. A model developed at an early age did not perform well in predicting phenotypes at a later age (6 years). The accuracy of the models was highest in the locations where they were created but decreased in other locations, indicating that genotype-environment (GE) interaction significantly affects the transferability of models between locations.

M.D. Resende et al. (2012) used two populations to evaluate GS accuracy in eucalyptus. The two populations contained 43 and 75 families (738 and 920 individuals, respectively) selected for GS. GS prediction accuracy 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 first population was expected, as it had a smaller effective population size (11) compared to the second (51). The accuracy of GS was comparable to or superior to that of PBLUP-based selection. GS models showed low predictability across populations, likely due to variable LD patterns, uncoordinated allelic effects, and GE interactions.

The GS prediction model can also be used to predict the segregation pattern of target traits in the progeny population. When crossing, it is important to select the optimal combination of parents that has a high probability of producing progeny with the desired characteristics. In the case of fruit trees, creating a segregating population and field evaluating the population is time-consuming and costly, so systematic planning to select the optimal combination becomes even more important. Iwata et al. (2013b) proposed a new method for predicting

segregation of traits in a progeny population based on GS prediction models and applied this method to Japanese pear data in a proof-of-concept study. Empirical analysis using a real breeding population and a simulation study based on real marker data showed that the segregation of target traits can be predicted with reasonable accuracy, especially in the case of traits with high heritability. The proposed method can provide objective and quantitative criteria for selecting the optimal combination of parents and a sufficient breeding population size.

Techniques for accelerating the growth of fruit trees

Most fruit trees have a long juvenile phase. For example, apple trees require 5–12 years to transition to the adult phase, while peaches can begin flowering within a few years after germination (Hansche 1986, Visser 1964). To reduce the time and cost of fruit crop breeding, more effective methods of accelerating generation time are needed (van Nocker and Gardiner 2014). An accelerated breeding system has recently been developed based on controlling the transition from the juvenile phase to the adult phase by inducing the flowering gene and/or suppressing the flowering repressor in combination with MAS, with the aim of reducing the juvenile phase in fruit crops (Flachowsky et al. 2007, 2011, Wenzel et al. 2013).

Over the past two decades, many factors involved in flowering induction have been identified in the model dicotyledonous plant Arabidopsis thaliana (Parcy F. 2005, Amasino 2010). For accelerated breeding, the most promising gene may be FLOWERING LOCUS T (FT). AtFT encodes a mobile regulator of flowering under long-day conditions, which corresponds to the characteristics of florigen (Corbesier et al. 2007). The introduction of Citrus FT (CiFT) into trifoliate orange sharply reduced the juvenile phase and led to flowering after only 12 weeks in a greenhouse, whereas the first flowering usually occurs no earlier than several years (Endo et al. 2005). Similarly, overexpression of MdFT1, an apple gene similar to FT, caused premature flowering in apple trees (Kotoda et al. 2010). FT-like genes from poplar under the control of the heat-inducible promoter Gmhsp 17.5-E from soybean were also used to induce flowering in apple to avoid the disadvantages of constitutive expression (Wenzel et al. 2013).

Several other candidates for the induction of flowering in fruit crops have been identified. FT activates various downstream genes, including the flower meristem identity gene APETALA1 (AP1; Blackman et al. 2010, Wigge et al. 2005). Transgenic citrus (Citrus sinensis × Poncirus trifoliata) with

Constitutive expression of AP1 resulted in fruiting flowers and fruits in the first year (Peña et al. 2001). LEAFY (LFY) activates AP1 and also participates in determining the identity of the floral meristem (Wagner et al. 1999, Weigel et al. 1992). Similar to AP1 transformants, transgenic citrus plants expressing LFY also exhibited an early flowering phenotype (Peña et al. 2001). TERMINAL FLOWER 1 (TFL1) encodes an FT-like protein, an antagonist of LFY and AP1; it suppresses the conversion of inflorescence meristems into floral meristems (Liljegren et al. 1999). In apple and pear, suppression of TFL1-like genes reduced vegetative growth and led to early flowering within a few years, even when cultivated in vitro. (Flachowsky et al. 2012, Freiman et al. 2012). FRUITFULL (FUL), a gene closely related to AP1, plays a redundant role in the identification of floral meristems (Ferrándiz et al. 2000, Litt and Irish 2003). Flachowsky et al. (2007, 2011) used BpMADS4, a gene similar to FUL, from silver birch to induce flowering in apple trees and accelerate the generation cycle.

In addition to transgenesis, a technique for inducing flowering in host plants using a plant viral vector has recently been developed. The Apple latent spherical virus (ALSV) vector containing AtFT induced flowering in 30% of apple seedlings 1.5–2.0 months after inoculation (Yamagishi et al. 2011). Virus-induced suppression of MdTFL1-1, the apple TFL1 homolog, caused early flowering in approximately 10% of inoculated apple seedlings (Sasaki et al. 2011). Furthermore, simultaneous induction of AtFT and suppression of MdTFL1 using the 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 until final varieties are obtained (van Nocker and Gardiner 2014). On the other hand, the ALSV vector does not affect the genotype and flowering characteristics of the next generation, as ALSV is rarely detected in subsequent offspring (Yamagishi et al. 2014). In the future, plant virus vector-induced temporary induction may be applied to other fruit crops and become more promising for accelerating generation time.

Prospects for fruit tree breeding using genomics

GWAS and GS will become increasingly important methods in the future of fruit tree breeding and genetics, combined with increased throughput and reduced cost of SNP genotyping, as well as improved accuracy and power of statistical methods. GWAS can estimate the location and effect of QTL/genes associated with target traits, and these 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 loci with small effects. Even for traits for which no significant associations are found, GS predictions can achieve a certain level of accuracy (Iwata et al. 2013a). In addition to GWAS and GS, generation acceleration technologies that promote selection with a fast cycle, short generation time, and small plant size will help optimize fruit tree breeding.

GWAS and GS require phenotypic data and marker genotype data for analysis and modeling. However, collecting phenotypic data for a large number of varieties/lines is costly due to problems that hinder fruit tree breeding, namely the long juvenile phase and large plant size. As suggested by Myles (2013), researchers should consider how to increase the throughput of their phenotyping strategies and how to create populations suitable for genetic mapping, MAS, and GS. One way to collect phenotypic data for numerous varieties/lines is to use breeding populations that are regularly developed and evaluated in breeding programs. If phenotypic and marker genotypic data can be collected regularly in breeding programs, the resulting collection will increase the efficiency of GWAS and the accuracy of GS. Data collected from breeding populations will also be useful for elucidating functional plant genomics (Poland 2015). Another possible approach is to use marker genotyping data to identify the optimal subset of varieties/lines for phenotyping, since marker genotyping is less costly than phenotyping. Careful selection of the subset based on an optimization algorithm will improve the accuracy of GS (Cros et al. 2015). If there is a subpopulation structure in breeding materials, the main collection, which includes representative individuals from all subpopulations, can be a good training population for improving GS accuracy (Fodor et al. 2014). The development of a field system for high-throughput phenotyping is also necessary to improve the efficiency of phenotypic data collection in breeding programs (Shakoor et al. 2017, Deery et al. 2014, Fiorani and Schurr 2013, Poland 2015, White et al. 2012).

The relationship between phenotypes and marker genotypes is sometimes strongly influenced by non-genetic factors. Fruit quality is controlled by complex genetic systems and is easily influenced by environmental conditions (Chagné et al. 2014). Cultivation management, such as rootstock selection, canopy training systems, pruning methods, and post-harvest handling, also affects fruit quality and yield (Myles 2013). Thus,

when applying GWAS and GS to breeding populations, genotype × environment × management (GEM) interactions should be taken into account. When these interactions have a profound effect, MAS and GS are not without complications. For example, models developed in one location will not function in another location with different environmental conditions (M.F.R. Resende et al. 2012a). From this perspective, fruit trees may have an advantage: since most fruit tree species can be propagated clonally, multi-environmental trials can be conducted using a set of identical clones to assess in detail the impact of GE or GEM interactions on GWAS and GS (e.g., M.F.R. Resende et al. 2012a). However, as noted by Heffner et al. (2009), the genotype of any individual consists of alleles that have been evaluated in a large number of target environments, and thus it is possible to maintain high GS accuracy even in the presence of GE interactions if phenotypic data have been collected in many environments.

Another way to solve the problem of GE and/or GEM interaction is to develop mathematical models describing the nature of the response of varieties/lines to environmental influences. An ecophysiological model is a method for simulating the response of plant phenotypes to environmental factors and describing the patterns of response of varieties/lines to the environment using model parameters. These parameters are expected to reflect the genetic characteristics of varieties/lines and can be used for QTL analysis to study varietal differences in response to the environment. In the case of apple trees, an ecophysiological phenological model was applied to simulate flowering time and used to calculate the risk of damage from spring frosts under climate change conditions (Eccel et al. 2009). This approach will be useful for breeding new varieties that can adapt well to future climate change. Since fruit tree varieties cannot be bred in a short period of time, it is necessary to develop such new varieties with a long-term perspective.

The phenotypes of fruit trees are also influenced by age-related factors (i.e., ontogenetic factors) and their interaction with genetic factors. To clarify the genetic, ontogenetic, and environmental influences on phenotypes, it is necessary to separate the mixed effects caused by successive years of growth (i.e., ontogenetic effect) and climatic years (i.e., environmental effect). A staggered-start experimental design (Loughin 2006) allows the effects of genetic × ontogenetic and genetic × annual interactions to be statistically separated. The potential of staggered-start designs for fruit tree research

fruit trees was demonstrated in a genetic analysis of apple tree architecture (Segura et al. 2008).

Genomics-based approaches and genomics-assisted breeding will bring significant benefits to the genetic improvement of fruit trees, although some obstacles need to be overcome. The development of breeding programs that make optimal use of genomics-based approaches is an important task for plant breeders, as each species is unique and requires an individual approach (Lin et al. 2014). Stochastic simulations based on a specific breeding program (e.g., Denis and Bouvet 2013, Iwata et al. 2011, Yabe et al. 2013, 2014) will be useful for finding ways to optimize the use of genomics-based approaches in breeding programs, taking into account species- and/or breeding program-specific constraints.

Acknowledgements

This work was supported by a grant from the Ministry of Agriculture, Forestry and Fisheries, Japan (Genomics-based technologies for improving agriculture, NGB2010) and research grants (A), MEXT (No. 25252002) and JST/CREST.

References