Genomic Selection

Elisabeth Jonas, Freddy Fikse, Lars Rönnegård, and Elena Flavia Mouresan

Abstract Prediction of phenotypes is not only used for selection and breeding in animal and plant populations but also for the assessment of specific phenotypes, especially predisposition to diseases and disorders in human populations. The use of genetic markers has been shown to be useful for prediction and selection for phenotypic traits. The concept of using genetic markers for prediction of breeding values or phenotypes was suggested many decades ago, but applications of marker-assisted selection were limited due to the low number of markers that could be genotyped and the low number of confirmed quantitative trait loci (QTL) that could be selected upon. Genomic selection, in contrast, utilizes dense genetic markers across the whole genome for the prediction of phenotypes as all QTL can be assumed to be in linkage disequilibrium with at least one marker. Genomic selection allows thereby choosing the genetically best individuals without the need to confirm QTL. The concept of genomic selection, proposed in 2001, has since been further developed and applied. Nowadays, genomic selection is widely applied in breeding populations of plants and animals for the selection of future breeding individuals. The chapter introduces the general concept of genomic selection. It further discusses relevant prerequisites for the application of genomic selection, including genotyping platforms and reference populations. Some of the methods applied today as well as suggested advancements of methods are introduced. The final part of the chapter describes briefly applications in animal, plant, and human populations (status when writing this chapter), before concluding with some general notes on genomic selection.

E. Jonas (✉) · F. Fikse · E. F. Mouresan
Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
e-mail: elisabeth.jonas@slu.se

L. Rönnegård
Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

Statistics Unit, School of Technology and Business Studies, Dalarna University, Falun, Sweden

Om P. Rajora (ed.), *Population Genomics: Concepts, Approaches and Applications*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_11, © Springer International Publishing AG 2018

Keywords Breeding · Genotyping · Parametric methods · Prediction · Reference population · Selection candidates · Semiparametric methods

1 Introduction: Genomic Selection in a Nutshell

The possibility to change the distribution of a trait in animal and plant populations by means of selection has been developed tremendously over the last 100 years. We have gained insight into principles of population genetics and have been able to formulate these principles in terms of statistical models. Quantitative genetic theory is based on the principles of Mendelian inheritance and explains how selection of individuals affects the development of a population in future generations and thereby connects genetics on an individual level to population changes.

Models for genomic selection can be interpreted using quantitative genetic theory, connecting changes on a population level to the set of genotypes observed on an individual level. The idea is that most genes have some effect on a trait and that the sum of all gene effects for an individual can be predicted as genomic breeding values (GEBVs) using markers in linkage disequilibrium with the causative genes. In practice, this is done by first estimating the combined genetic effects for each individual of a reference population and subsequently using this information to predict GEBVs for the selection candidates. This requires that extensive genotype information is available both for the reference population and the selection candidates, which has only been possible for the past two decades.

Modern genotyping technologies enable genotyping of many individuals for a larger number of genome-wide markers at affordable cost. These advances in genotyping technologies have been exploited in genomic selection to compute GEBVs (Fig. 1). Hence, a reference population needs to consist of individuals with both genotype and phenotype information. Marker effect estimates from the reference population are combined with genotypes of selection candidates to predict the genetic potential of the selection candidates. Animals with the most desirable genetic potential are kept for breeding to become parents of the next generation of individuals.

The progress in the field of medium- and high-throughput genotyping platforms along with decreased costs for marker detection via sequencing technologies enhanced the use of genomic information in breeding (Davey et al. 2011). In 2001, Meuwissen et al. (2001) published their landmark paper on the use of genome-wide selection or genomic selection, proposing a marker-based selection methodology that incorporates marker information of many (dense) markers in the prediction model. Only a decade later, this method was already employed and implemented in (dairy cattle) breeding programs, and estimated breeding values (EBVs) based on genomic data (GEBVs) were officially published in a number of countries (Patry 2011). The use of genomic selection has also been a major interest for the breeding of crop species (Heffner et al. 2009; Cabrera-Bosquet et al. 2012). A number of methods for the estimation of effects have been suggested (Meuwissen et al. 2001; Habier et al. 2007; de los Campos et al. 2009; Zhong et al. 2009; Kizilkaya et al. 2010), and further developments are on the way, some of which will be detailed later.

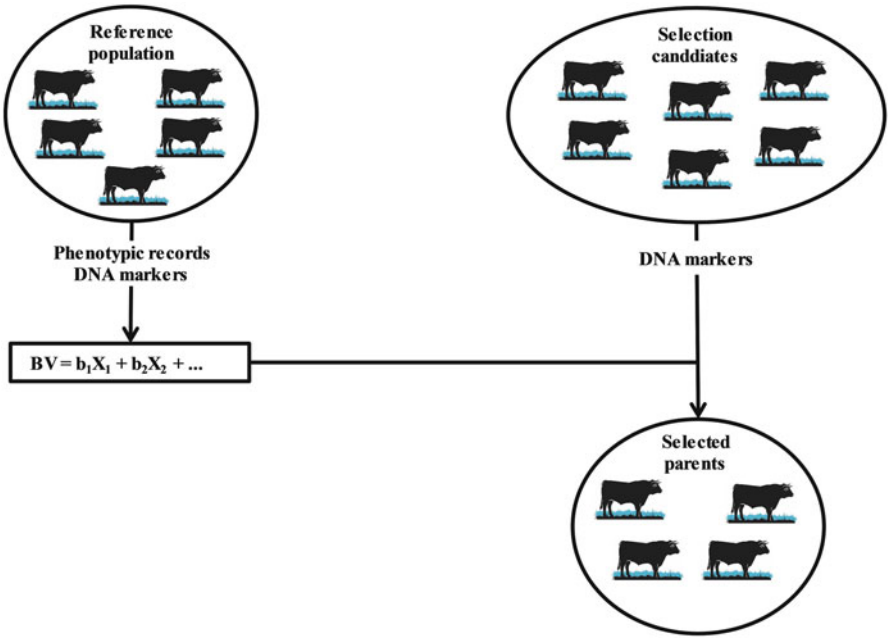

Fig. 1 Schematic overview of the concept of genomic selection including the reference population with both genotype and phenotype individuals and the selection candidates with genotype information. Information on breeding values (BV) will be used to select parents from the population of selection candidates

1.1 How Genomic Selection Really Works

Meuwissen et al. (2001) argue that linkage disequilibrium between markers and quantitative trait nucleotides (QTNs) is the driving force behind genomic prediction. Observing nearby genetic markers supplies information about the QTN, if there is a close association between a marker and the QTN. Given the linkage disequilibrium between QTN and markers as driving force, many expectations and speculations about the behavior of genomic selection have been put forward (e.g., about benefits of whole-genome sequence, across-breed genomic prediction), but more often than not, those expectations and speculations have not been consistent with real data.

Understanding of the mechanisms of genomic prediction became clearer when Habier et al. (2007) showed that accuracies of genomic breeding values were substantially larger than zero even if markers and QTNs were in linkage equilibrium. Genetic markers can capture family relationships and thereby contribute to the accuracy of estimating genomic

breeding values. Habier et al. (2013) described this and investigated the contribution of three information sources to the accuracy of genomic breeding value estimation: markers capturing additive genetic relationships, co-segregation, and linkage disequilibrium.

When the training population is small, much of the accuracy of genomic breeding values is due to markers describing family relationships. A consequence is that the predictive ability rapidly decays over generations. Only if training populations are large, marker estimates reflect more of the effect of actual QTNs nearby instead of additive genetic relationships, and predictions are persistent for more generations.

2 Background

2.1 History of Human-Introduced Genetic Changes to Populations

Selective breeding in both plant and animal species started many thousands of years ago. Genetic change over time was achieved by selecting the best-fit individuals as parents, such that the next generation of individuals was, on average, superior to the parent generation. Selection based on phenotypic criteria has been performed since the domestication of species (Rosenberg and Nordborg 2002; Morrell et al. 2012). While early selection was based on the observation of phenotypes within a group of individuals, more sophisticated tools are used for selection in plant and livestock populations as they are used in farming today. At the beginning of the eighteenth century, Robert Bakewell (1725–1795) established modern breeding by introducing systematic and structured selective breeding (Sweeney and McCouch 2007).

The demonstration of inheritance and the discovery of basic rules of inheritance by Gregor Johann Mendel in the middle of the nineteenth century established the beginning of modern genetics. The complexity of phenotypes and their inheritance could then be explained by their genotypes via rules of allele sharing across generations. Thereafter this had a major impact on animal and plant breeding.

Many of the basic statistical tools used in quantitative genetics were developed in the late nineteenth and early twentieth century by Francis Galton and Karl Pearson. In 1918, Ronald Aylmer Fisher used statistical models to demonstrate the resemblance between relatives and introduced the analysis of variance (Walsh 2001). Further milestones of breeding were laid by Jay Laurence Lush from the 1930s and Charles Roy Henderson from the 1970s and their suggestion of the use of statistical models (Lush 1933, 1947; Henderson 1975a, b). During the twentieth century, many statisticians, quantitative geneticists, and breeders contributed to the development and implementation of different breeding schemes in livestock and plants based on knowledge of trait inheritance and statistical approaches.

2.2 Basic Quantitative Genetics Relevant for Breeding

The basic concept of quantitative genetics, applied in breeding, is that an individual's phenotype P is determined by its genotypic value G and its environment E (Walsh 2001):

$$P = G + E$$

The genotypic value can be decomposed into additive (A), dominance (D), and epistatic (I) values in which A accounts for the average effects, D for the interaction between alleles at one locus, and I for the interaction between alleles at different loci:

$$G = A + D + I$$

A relevant measure applied in quantitative genetics is the narrow-sense heritability, h^2, which is the proportion of the total phenotypic variance P due to the additive genetic effects A:

$$h^2 = \mathrm{Var}(A)/\mathrm{Var}(P)$$

The heritability can be used to describe the phenotypic similarity between relatives or trait variation due to additive genetic effects.

The heritability is also used for the prediction of the response to selection, in the so-called breeder's equation. This equation describes the change of the population mean over one generation or the response to selection ΔZ, with an applied selection differential S (Falconer and Mackay 1996; Lynch and Walsh 1998; Xu and Hu 2010):

$$\Delta Z = h^2 S$$

When the heritability of a trait is close to zero, the response will be very little even if there is strong selection on that trait.

Quantitative genetic analyses initially focused on decomposition of phenotypic variance into underlying components (like A, E) for quantitative traits (i.e., traits involving many genes and influenced by environment). More recently, the possibility to genotype individuals for DNA markers allowed the attention to shift to identification of chromosomal regions with (large) effects on quantitative traits: quantitative trait loci (QTL). Single nucleotide polymorphisms (SNPs) that are causative of the QTL are hereinafter referred to as quantitative trait nucleotides (QTNs).

2.3 Examples of Breeding Programs and Selection Decisions

Breeding programs aim to change certain traits toward a breeding goal in a population. The duration until the genetically improved individuals are available for breeding is expressed as the generation interval. The generation interval is relevant for the genetic

and economic gains in a breeding program: a shorter interval means that improvement can be achieved earlier. Quantitative genetic theory has underpinned the design of selection schemes of plant and livestock populations for many decades. The molecular genetic background of traits has been integrated as a selection tool more recently. The application of these tools in breeding programs has aided selection of the individuals with the best genetic merit for the traits of interest. The true genetic merit, i.e., the true breeding value, of an individual is mostly unknown, and estimated breeding values can be used to predict how well offspring will perform.

Breeding programs aim to identify the best individuals for breeding to produce the next and improved generation. Differences across species regarding the reproduction capacity and the breeding goal traits influence the design of a breeding program. Crossing of lines can be used to create lines with a new combination of characteristics. Crossbreeding often exploits hybrid vigor, or heterosis. Heterosis effects are difficult to predict and are mainly realized in the first generation of crossbreeding. Overcoming cross-incompatibility is relevant in some species in order to allow new trait combinations. The final goal of crossing is to produce a generation with superior traits from each of the parental lines. A conflict exists between the need for diversity within the core breeding population and at least some degrees of uniformity within the production. Nucleus populations or diversity panels can be used to ensure the existence of diverse lines. These nucleus populations are, in many species, kept centrally by few breeding organizations, which define the breeding goals and design the breeding schemes (Figs. 2 and 3).

The breeding goal describes which traits are important for genetic improvement and their relative importance. But not all traits are measured on the selection candidates; some traits are only collected from relatives, and this information can be used for prediction of breeding values for selection candidates based on basic principles of population genetics. The more accurate a trait can be measured, the more accurate selection is. Especially when traits are influenced by other factors, such as the environment, the accuracy of selection might be negatively affected. Evaluation of performance in a controlled environment is one option to reduce the impact of environmental variation on accuracy of selection, and an alternative option is the evaluation of traits under various environmental conditions.

Plant lines with improved traits are still mainly selected based on their phenotypic appearance. However, many generations are needed to produce cultivars with the desired characteristics through conventional breeding, as well as (multilocation) testing (Sharma et al. 2002). There are several constraints in plant breeding including varying outdoor conditions to be accounted for. Also the modes of reproduction influence the possibilities of plant breeding as crossbreeding is restricted in some species.

Improvement schemes for livestock populations are often organized around a nucleus herd. Own performance and performance of relatives (offspring, sibs, parents) are evaluated for the identification of individuals with the best genetic merit. The pedigree plays a major role for developing breeding schemes and mating decisions in animal

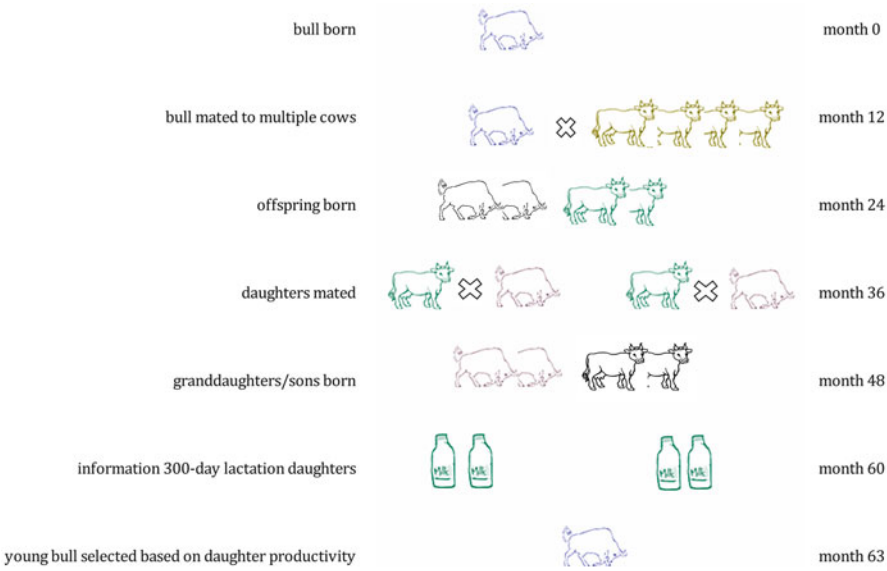

Fig. 2 Example of a traditional breeding scheme in dairy cattle (*Bos taurus*) with the time frame on the right side and the different stages of the breeding cycle on the left. Young bulls for selection are born in month 0 and in month 12 mated to cows. Daughters are born in month 24 and mated in month 36. The granddaughters of the bulls for selection are born, and data on relevant traits (e.g., milk yield, fertility, disease resistance) of the bull's daughters is collected. Information on these traits for one lactation is then available in month 60, and bulls can be selected for breeding based on their daughter's first lactation performance

breeding. The genetics of sires can be distributed widely via artificial insemination. The evaluation of the genetic merit targets therefore mainly the sires in livestock populations, as they have a great genetic impact on their population (Gerrits et al. 2005; Funk 2006).

2.4 Selective Breeding Using Molecular Markers

Selection of breeding stock and lines based on phenotype and pedigree data allowed to improve many breeding populations. But traits are based on changes at the level of the DNA. It has, therefore, been suggested that using genetic information based on the inherited part of the individual, its DNA, will allow a better prediction of the genotypic value and, therefore, the phenotypic value of an individual or its real breeding value. As first suggestions for the use of (molecular) marker information in breeding programs (Dekkers and Hospital 2002), marker-assisted selection (MAS), marker-assisted recurrent selection (MARS), and marker-assisted breeding (MAB) were discussed. For a successful implementation of DNA information in selection scheme information, such

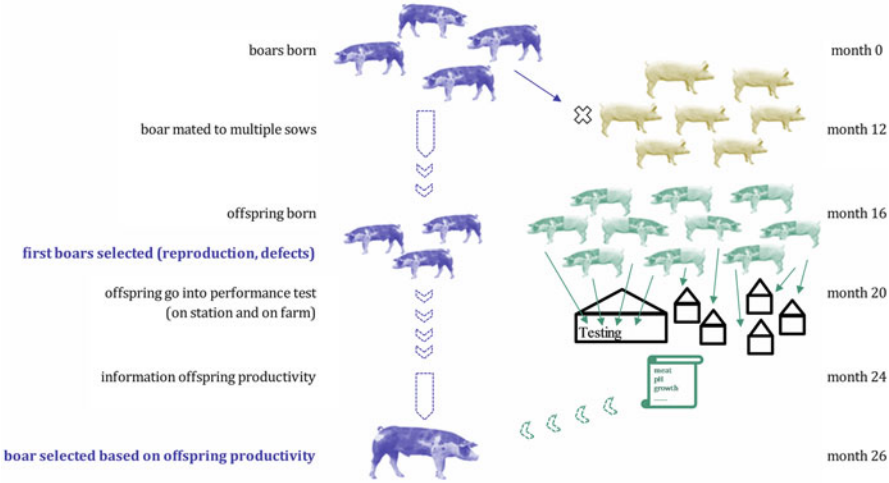

Fig. 3 Example of a traditional breeding scheme in pigs (*Sus scrofa*) with the time frame on the right side and the different stages of the breeding cycle on the left. Young boars from paternal lines for selection are born in month 0 and at month 12 mated to sows from maternal lines. Crossbred offspring are born in month 16 and tested in performance test stations or on farms. The offspring of the boars for selection are slaughtered, and the information from fattening period and slaughterhouse are collected. Information from reproduction and production is available at approximately month 26, and boars can be selected for breeding based on breeding values predicted using the offspring performance

as identification of markers, successful genotyping and validation of genotype and allele frequency in a large amount of individuals are required. Additionally, validated association of the genetic marker with the trait of interest and an assessment of the marker effect in a breeding program are needed. Furthermore, any potential negative effects on other economically important traits need to be excluded.

Selection using genetic markers has been suggested as a preferred method for traits for which phenotypic selection is more difficult, such as traits with low heritability. Other examples are traits for which the assessment of the phenotypes is difficult and cost-extensive or can only be done late in life. It has also been shown that MAS is particularly effective for traits with (one or few) major QTL effects (Gupta et al. 2010; Cabrera-Bosquet et al. 2012). Not only is the increased accuracy of selection due to the use of markers in breeding programs the main advantage of MAS but also the minimization of phenotyping (Bernardo and Yu 2007). In plant breeding, MAS can also be applied in year-round breeding nurseries or greenhouses where phenotypic data are less meaningful as their correlation to field data is low. Marker information can, in such settings, allow a prediction of the phenotypes (Lorenzana and Bernardo 2009). But costs for the identification of useful genetic markers and the implementation of such markers are relatively high compared to the gain achieved. Some markers

are also population- or family-specific. It has been, for this and other reasons, stated that MAS is not well suited for the improvement of crops (Jannink et al. 2010). The identification of useful markers is time-consuming, and as many traits are influenced by multiple genes, also a higher number of markers would be needed for MAS for each single trait (Gupta et al. 2010). Only relatively few causative mutations have been identified (in livestock) and implemented as a routine component in breeding programs.

3 Genomic Selection Designs and Strategies

3.1 Designing Genotyping Platforms

One of the prerequisites of the application of genomic selection is the availability of information from genetic markers across the genome. Many genetic markers were identified using methods such as Sanger sequencing and their information, for example, collected in the NCBI database (https://www.ncbi.nlm.nih.gov/). When next-generation sequencing (NGS) was accessible to a larger number of researchers, the amount of information increased significantly as whole-genome sequences were available. This allowed the discovery of many potential markers, such as SNPs. The time required for the sequencing of the full genome of an individual and the costs for it have been reduced significantly during the last decades (Goodwin et al. 2016). The genome of many species has been sequenced, and genotyping arrays have been developed based on the sequence information (Table 1), but the progress on the assembly of a good reference genome, the availability of full genome sequence information from multiple lines and varieties, and the development of SNP arrays capturing the complex and repetitive genome of plants is slow (Somers et al. 2003; Ganal and Roeder 2007; Trebbi et al. 2011). While genome-wide high-throughput genotyping platforms are available for many livestock species, the progress in the development of such platforms for plants has been slower.

The gene density differs widely between livestock and plant species. However, it is mainly the extent of linkage disequilibrium that plays a major role for the application of molecular genetic tools in breeding and MAS (Chao et al. 2010). Linkage disequilibrium is strongly related to the population history, especially resulting from evolutionary history, mating system, population size, admixture, recombination rate, and selection (Heffner et al. 2009). The decay of linkage disequilibrium varies not only with increasing physical distance of loci between species but also between populations of the same species and across chromosomes (Remington et al. 2001; Maccaferri et al. 2005; Chao et al. 2007; Mather et al. 2007; Tenaillon et al. 2008). Structures of linkage disequilibrium depend also on the breeding scheme; plant breeders, for example, often use full-sib families created from crosses of inbred parents, and linkage disequilibrium will be extensive within each family (Zhong et al. 2009). The marker density required for genomic selection will therefore depend on the population and breeding structure.

Table 1 Genome structure and available information on the genome of selected livestock and plant species including examples of options for genotyping (table restricted from updates in the communities)

Species	Ploidy	Genome structure	Genotyping arrays	Genebuild started/ released[a]	Draft/reference sequence
Barley (*Hordeum vulgare*)	Diploid ($n = 7$)	~5.7 Gb (1.36 GB assembled), complex, repeat-rich, ~1 SNP per 50–189 bases; 79, 379 genes called, 24,287 coding genes; 268 pseudogenes	9K (7,842 SNPs) iSelect SNP[b] chip; planned iSelect 50K SNP[b] chip; DArT markers	Started in 2014	(International Barley Genome Sequencing Consortium et al. 2012; Mascher et al. 2017)
			GoldenGate/BeadArray: OPA1, 1,536 or OPA2, 3,072 SNPs[b]		
Wheat (*Triticum aestivum*)	Hexaploid $n = 7$ (A, B, D)	~13.43 Gb; highly repetitive and duplicated; ~1 SNP per 335 bp; 103, 539 coding genes	9K SNP[b] array, 90K array (in development by consortium), wheat HD (817,000 SNPs)[c] genotyping array; genotyping by sequencing	Started in 2015	(Brenchley et al. 2012)
			Custom chips: GoldenGate[b], iSelect, TaqMan®, KASPar[d]		
Wheat (*Triticum durum*)	Tetraploid ($n = 7$)	Highly repetitive (~85–90%); duplicated	GoldenGate 275 SNPs[b]	No	
Rye (*Secale cereale*)	Tetraploid ($n = 7$)	~7.92 Gb; highly repetitive (~92% repetitive); ~36,000 genes	DArT markers (around 5,000); Rye5K SNP array[b]	No	(Bauer et al. 2017)
Potato (*Solanum tuberosum*)	Tetraploid ($n = 12$)	~844 Mb; 727 Mb assembled; highly heterozygous, inbreeding depression; ~39,021 coding genes	SolSTW array-20K SNP array[b]; potato 12K SolCAP array (in development); Infinium 8303 Potato Array[b]	Started in 2011	(Potato Genome Sequencing Consortium 2011)
Rice (*Oryza sativa L.* ssp. *indica*)	Diploid ($n = 12$)	466 Mb; 412 Mb assembled; 46,022 to 55,615 genes; 40,745 coding genes	GoldenGate, 1,536 or 768 SNPs[b]; BeadXpress, 384 SNPs[b]; GeneChip Rice 44K (44,100 SNPs)[c] array; KASP™ arrays[e]	2010	(Yu et al. 2002)
Rice (*Oryza sativa* ssp. *japonica*)	Diploid ($n = 12$)	~420 Mb; ~1 SNP per 10 kb; 32,000 to 50,000 genes; 35,679 coding genes	GoldenGate[b], 1,536 or 768 SNPs; BeadXpress, 384 SNPs[b]; GeneChip Rice 44K (44,100 SNPs)[c] array; KASP™ arrays[e]	Yes	(Goff et al. 2002)
Maize (*Zea mays*)	Diploid ($n = 10$)	~2.3 Gb; recent genome extension; historical genome duplication; ~1 SNP per 28–214 bp; 37,000 to 63,000 genes; 39,324 coding genes	MaizeSNP50 (56,110 SNPs) BeadChip[b], KASP™ arrays[f]; Maize genotyping array (616,201 variants including 609,442 SNPs)[c]	2015	(Schnable et al. 2009)

Cattle (*Bos Taurus*)	Diploid ($n = 29 + XY$)	~2.65 Gb; 3,000,000 SNPs identified; ~19,994 protein-coding genes; 797 pseudogenes	BovineSNP50 (54,609) SNP[b] BeadChip, BovineHD (777,962 SNPs) BeadChip[b]; BovineLD array[b]; 640,000 SNP[c] array; GGP Bovine LD (v3) (26,000 SNPs) genotyping array[b]	2011	(Elsik et al. 2009)
Pig (*Sus scrofa*)	Diploid ($n = 18 + XY$)	~3.02 Gb; 510,000 SNPs identified; ~21,630 protein-coding genes; 568 pseudogenes	PorcineSNP60 (64,232 SNPs) BeadChip[b]; Porcine genotyping array (658,692 SNPs)[c]	2012	(Groenen et al. 2012)
Chicken (*Gallus gallus*)	Diploid ($n = 33 + ZY$)	~1.29 Gb; 16,362 protein-coding genes; 50 pseudogenes; 1,800,000 SNPs identified	Chicken genotyping[c] array (580,000 SNPs)	2016	(Rubin et al. 2010)
Atlantic salmon (*Salmo salar*)	Diploid ($n = 29$)	~6,000 Mb; many chromosomal rearrangements; 33,709 genes	iSelect Atlantic Salmon 16,500 SNPs[b]; Salmon genotyping array (130,000 SNPs)[c]	No	(Lien et al. 2016)
Sheep (*Ovis aries*)	Diploid ($n = 26 + XY$)	~2.53 Gb; 20,921 protein-coding genes; 290 pseudogenes	OvineSNP50 (64,232 SNPs) BeadChip[b]; OvineHD (700,000 SNPs) chip[g]	2013	(Jiang et al. 2014)
Horse (*Equus caballus*)	Diploid ($n = 31 + XY$)	~2.43Gb; 20,436 protein-coding genes; 4,400 pseudogenes	EquineSNP50 (54,602 SNPs) BeadChip[b]; Axiom® Equine (670,769 SNPs) Genotyping Array[c]	2008	(Wade et al. 2009)
Dog (*Canis lupus familiaris*)	Diploid ($n = 38 + XY$)	~2.39 Mb; 19,856 protein-coding genes; 950 pseudogenes; >2.5 million SNPs; less lineage-specific repeat sequence compared to human; specific haplotype structure due to two bottlenecks	CanineHD (172,115 SNPs), Canine Array Sets A (460,000 SNPs) and B (670,000 SNPs) Axiom genotyping array[c]	2012	(Lindblad-Toh et al. 2005)
Loblolly pine (*Pinus taeda L.*)	Diploid ($n = 12$)	23.2 Gb; 50,172 genes annotated; average intron lengths >2.7 Kb up to 100 Kb		Yes	(Neale et al. 2014; Zimin et al. 2014)
Eucalyptus (*Eucalyptus grandis*)	Diploid ($n = 11$)	Estimated genome size, 640 Mb; approximately 36,376 protein-coding genes (of which 34% in tandem duplications); around 44.5% retrotransposons	DArT markers; EuCHIP60K (60,904 SNPs)[b]	Yes	(Myburg et al. 2014)
Eucalyptus (*Eucalyptus globulus*)	Diploid ($n = 11$)	Estimated genome size, 530 Mb	DArT markers; EuCHIP60K (60,904 SNPs)[b]	Yes	(Myburg et al. 2014)

(continued)

Table 1 (continued)

Species	Ploidy	Genome structure	Genotyping arrays	Genebuild started/ released[a]	Draft/reference sequence
Poplar (*Populus trichocarpa*)	Diploid ($n = 19$)	Estimated genome size, 423 Mb; approximately 36,393 genes		Drafts	(Tuskan et al. 2006)
Domesticated apple (*Malus*×*domestica* Borkh.)	Diploid ($n = 17$)	Estimated genome size, 742.3 Mb; unassembled part of the genome is 98% repetitive (138.4 Mb); diverse ploidy levels within genus and species (two to five times)	IRSC apple 8K (7,867 SNPs) SNP array v1[b]; Axiom_Apple480 (487,249 SNPs)[c]	2015	(Velasco et al. 2010)
European pear (*Pyrus*)	Diploid ($n = 17$)	Estimated genome size, 527–577 Mb depending on species; approximately 43,000 genes; highly heterozygous due to self-incompatibility	Pear Infinium® II 9K SNP array[b]	Draft 2013 for *Pyrus communis*	(Wu et al. 2013; Chagné et al. 2014)
Rubber tree (*Hevea brasiliensis*)	Diploid ($n = 18$)	~1.47 Gb (estimated 2.15 Gb haploid genome); ~43,792 predicted protein-coding genes; repeat-rich (around 71–78% of the genome length)	GBS	Drafts	(Tang et al. 2016)
Norway spruce (*Picea abies*)	Diploid ($n = 12$)	~9.6 Gb; ~29,000 predicted genes	311 iSelect SNP[b] chip for validation	No	(Nystedt et al. 2013)
Black spruce (*Picea mariana*)	Diploid ($n = 12$)		4267 iSelect SNP[b] chip	No	
White spruce (*Picea glauca*)	Diploid ($n = 12 + 1B$)	~20.8 Gb	7338 and 9559 iSelect SNP[b] chip	No	(Birol et al. 2013)

bp base pairs, K, in 1000 base pairs, *MB* megabases, *SNP* single nucleotide polymorphism, KASPar, DArT markers
[a]Date of gene build release on Ensembl (http://www.ensembl.org/index.html), Gramene (http://www.gramene.org/), GDR (https://www.rosaceae.org/), PineRefSeq (http://pinegenome.org/pinerefseq/), or Dendrome (https://dendrome.ucdavis.edu/)
Companies offering the arrays: [b]Illumina (some specific arrays for GeneSeek); [c]Affymetrix; [d]KBioscience; [e]developed for the Generation Challenge Programme and the Integrated Breeding Platform; [f]developed for CIMMYT; [g]International Sheep Genomics Consortium (ISGC)

Table 2 Information on genomic prediction and selection applied in different organisms

Species	Applications	Strategies	Advantages using genomic selection	Challenges
Cattle	Application in purebred Holstein Friesian dairy cattle populations; genomic breeding values used for selection; applications in smaller dairy breeds; test and application in beef cattle	Reduced generation interval and earlier selection of breeding bulls; reduced size of breeding programs with fewer bulls for testing; international collaborations	Reduction of generation interval; reduced costs for keeping bulls (fewer bulls)	Selection for low heritable traits; predictions in crossbred populations such as beef cattle; size of reference population (small cattle breeds); larger effective population in some populations (weaker linkage disequilibrium)
Pig	Studies using simulated and field data; collaborations between public and private sector; some lines on the market	Tests of application in purebred and crossbred populations; most studies within the frame of existing breeding programs	Additional genetic gain expected depending on trait and population	Costs for genotyping relatively high; already short generation interval; applications and accuracies in crossbred populations
Chicken	Advances in private sector in collaboration with public sector in layers and broilers, some lines on the market	Use of a reduced reference population and shorter generation interval	Increased genetic gain depending on trait; advancement especially if no information from relatives is available; improved male selection in layers	Costs for genotyping; refinement of breeding programs; adjustments of breeding goals; restore some variability in the highly specialized commercial lines
Atlantic salmon	Studies using simulated and real data close to applied breeding programs; applications in the private sector	Preselection of candidates for genotyping and within-family selection	Increased genetic gain in within-family selection, generally higher genetic gain with genomic selection	High costs in conventional breeding programs due to large number of selection candidates
Sheep	Some suggestions and applications in larger breeding populations	Test in different breeding populations and resource flocks	Some genetic gain estimated depending on trait (e.g., meat and wool traits); choice of prediction models relevant; possible reduction of generation interval	Value of each animal lower; organization levels of breeding often low; size of the reference population relevant
Horse	Few studies published Some application	International collaborations; selection mainly in sport horses, aim to reduce generation interval	Predicted improvement due to shorter generation interval; improved selection of female horses and young horses; advantages with exchanges genetics	Restricted breeding goals; international collaboration required; little literature

(continued)

Table 2 (continued)

Species	Applications	Strategies	Advantages using genomic selection	Challenges
Dog	Studies done using data from dog breeding originations and simulated data	Test especially for disease traits in pedigree dogs	Accurate selection especially at young age before onset of diseases; genetic gain	Population management; effects of diversity loss; large enough reference populations needed; definition of phenotypes
Barley	Trials and application in the private sector, many suggestions from the public sector	Cross-validation using two generations; inclusion of generations in the greenhouse	Shorter generation interval; selection of strategy will determine if long-term or short-term genetic gain	Right choice of reference population; capture of nonadditive effects such as genotype by environment interaction
Wheat	Trials and application in private sector, many suggestions from public sector using real and simulated data	Test of different strategies including selection schemes different from traditional schemes	High accuracies of prediction but limited to the same environment; improved prediction compared to phenotypic or marker-assisted selection	Selection for disease traits; high costs for genotyping and phenotyping due to large number of breeding lines; inclusion of nonadditive effects; predictions in biparental families
Rice	Suggestions for specific traits, especially drought tolerance; genotyping platform available	Based on marker-assisted approaches, integration of markers for specific traits	Improved yield under drought (markers identified), yield under biotic and abiotic stress	Recording of good phenotypes, understanding the trait architecture; development of new cultivars
Maize	Trials in private sector, many suggestions from public sector using real and simulated data	Different strategies tested to define best reference population, mostly based on hybrid or double haploid lines, aim to reduce phenotyping	Improved selection of diverse lines for crossing; genetic gain depending on traits and models; reduction of generation interval; use of generations in the greenhouse	Improvements needed for multiple family approaches; inclusion of environmental effect especially for disease traits as well as genotype by environment interactions
Perennial ryegrass	Tested in simulations with varying accuracies depending on trait	Changes to current breeding programs to implement genomic selection	Shorter breeding cycle, thereby higher genetic gain	Collection of reliable phenotypes over multiple years to create a good reference population; issues like increased inbreeding and decreased accuracies to be solved

Pine	Use of simulated data and data from structured breeding populations including clonally replicated field trials	Tests of inclusion of different models	Increased accuracy in progeny for some traits, accuracies match phenotypic selection but more efficient due to time decrease	Genotyping costs; complex genome; need of better control of dominance effects; structure of linkage disequilibrium; lacking gene maps
Eucalyptus	Some suggestions made using diverse data (seedling and orchard designs, breeding populations, simulations)	Test especially for diversity and increase genetic gain; test of relationship reference and selection populations; size of reference population and heritability	Accuracies match phenotypic selection; opportunities to decrease generation interval	Genotyping costs; complex genome; accuracy of generation difference between selection candidates and reference population
Spruce	Tested using data from different spruce populations; relevance of genomic selection in recently domesticated/undomesticated species	Single-step methods for breeding programs with simple mating structures and shallow pedigrees	Potential larger gains per unit time compared to traditional breeding; selection of superior individuals within large full-sib families as a short-term goal	Marker densities still low; establishment of reference populations with deep pedigrees; need of solution for open-pollinated mating schemes

Shown is an overview of the applications, strategies, advantages, and challenges in different livestock and crop species and two examples of tree species based on the literature used in the chapter

3.2 Designing Reference Populations

A reference population, also called as discovery or training set, describes the breeding stock for which information on the relevant traits, including data in multiple environments if relevant, is available. The individuals of the reference population are genotyped, and information on their pedigree or relationship is available. Two approaches are generally offered for the reference population: (1) an established reference population before the start of the next breeding cycle (e.g., in multiple-stage selection) and (2) the prediction using a reference set from the same generation as the selection candidates (e.g., in one-stage selection) (Marulanda et al. 2016). The decision on the structure of the reference population and the relationship to the population of selection candidates is relevant in a genomic selection breeding scheme. Deterministic functions to predict the accuracy of genomic breeding values include the size of reference population (Daetwyler et al. 2008): a larger reference population leads to a higher accuracy. When genomic selection was firstly applied and tested in dairy cattle, only bulls were utilized in the reference population. For these bulls, information on the tested progeny were available (as reviewed by VanRaden 2008; Hayes et al. 2009a). As it was assumed that larger training populations result in more reliable predictions, initiatives to pool reference populations across countries emerged, such as the European initiative EuroGenomics (Lund et al. 2011) or a collaboration between the USA and Canada (VanRaden et al. 2009a, b; Muir et al. 2010). It had been pointed out in a review that such international collaborations are desirable (Dürr and Philipsson 2012), because they result in reference populations of tens of thousands progeny-tested dairy bulls. Similar approaches have also been taken in the wheat breeding community to develop universal training populations by merging large phenotype dataset (e.g., by the Wheat Initiative's Expert Working Group on Wheat Breeding Methods and Strategies) (Bassi et al. 2016). Such international connections of data are still less advanced in beef cattle (Berry et al. 2016) and other livestock. The size of the reference population is often restricted by the costs. The increase of the reference population might lead to a shift away from the collection of phenotypes, but collaborations might allow the elaboration of more phenotypes in a larger reference population. A reduction of testing with fewer locations or replications in exchange of more genotyped and phenotyped lines in the reference population has been suggested in plant breeding to balance limited resources when increasing the reference population (Longin et al. 2015).

Implementation of genomic selection led to changes in dairy cattle breeding programs, with less emphasis on progeny testing and selection of fewer bulls, and genotyping of females has become a necessary complement to maintain and update reference populations. It has also been a concern in other breeding schemes that the introduction of genomic selection will reduce the phenotypic evaluation and might have potential drawbacks in the future.

The size of the reference population depends on the resources, and this will determine the accuracy of genomic prediction. Also the structure of the reference population and the relationship to the selection candidates influence the required size of the reference population. One of the first observations of genomic prediction applied to the real data

in dairy cattle was that the accuracy of genomic breeding values was dependent on whether or not the sire of the selection candidate was in the reference population. This observation was further sustained by studies on the distance between reference population and population of selection candidates (e.g., Habier et al. 2010), reporting that accuracy of genomic breeding values decreased with decreasing additive genetic relationship between bulls in reference population and selection candidates. These observations are also influenced by the structures of the populations, such as the linkage disequilibrium and QTL effects. The accuracy of prediction is lower when reference and selection populations are less related. Adding individuals to the reference population will not always lead to gains, as it will largely depend on the relationship to the selection candidates (Calus 2016), thus the ability to cover the linkage disequilibrium between markers and QTL of the selection candidates. Furthermore, if the genetic diversity or the allele frequencies in the selection candidates change, an update of the training population is needed (Bassi et al. 2016). The degree of relatedness within a reference population was also shown to affect the prediction accuracy in livestock, where low relationships among animals in the reference population result in the highest accuracy of genomic breeding values (Pszczola et al. 2012). Such a strategy probably ensures the widest range of possible genotypes present in the reference population. Especially in dairy cattle, where genomic selection has been widely applied, discussions on the actual optimum size of the reference population are ongoing. An example using cows in the reference population suggested that an initial size of 2000 cows would still require that information from 600 cows have to be added every year to keep the accuracies constant (Pszczola and Calus 2015). In Holstein Friesian dairy cattle, the size of the reference population exceeds today more than 30,000 bulls worldwide.

Differences in the design of the reference population in plant breeding do also depend on the mating system of plants. A study using F6 wheat lines showed that a reference population of 700 lines allowed the highest predictive abilities. The tested lines were derived from three different crossing and selfing schemes each based on 60 parental lines (Cericola et al. 2017). Inbreeding plants have higher levels of linkage disequilibrium compared to the population-wide linkage disequilibrium in outbreeding plants. The size of the reference populations has to be larger in outbreeding plants, unless genomic prediction is performed only within families (Lin et al. 2014). Such difficulties can be aligned to multi-breed populations in livestock breeding. The design of the reference population has to follow the criteria stated above also in multi-breed populations. Discussions on the size of the reference population will therefore seldom be concluded in a single number, but the general statement of "the more the better" will be relevant. Bassi et al. (2016) reported that the size of the reference population varies in plant breeding scenarios and can vary from 60 to 10,000 individuals. They also concluded that the size should be as big as possible but that other criteria such as relatedness and trait heritability have to be taken into account (Bassi et al. 2016).

4 Methods and Models Applied in Genomic Selection

A lot of efforts have been devoted to the development of models for genomic prediction, and this section presents an overview of the methods. The methods first proposed, and still commonly used, assume a linear relationship between the phenotype on the one hand and genotypes on the other hand. More recently, nonparametric approaches have been proposed that are less dependent on assumptions like linearity and multivariate normality, among others.

Comparisons of genomic prediction methods have in most cases only identified small differences in the predictive performance from the empirical data, but the differences are expected to increase with larger reference populations. There are several reasons why there might still be small differences in performance between prediction methods. Firstly, the genetic architecture of majority of the traits considered for genomic prediction points toward a polygenic mode of inheritance (i.e., many QTN with relatively small effects), and only a few traits are influenced by a smaller number of QTN with a large effect. Secondly, the validation horizon is often short; methods relying on markers tracing genetic relationships perform well to predict breeding values in the next generation, and advantages of methods exploiting linkage disequilibrium are small. Predictive ability tends to decrease if there are more generations between the selection candidates and the reference population, more so for methods that rely on markers tracing genetic relationships than for methods exploiting linkage disequilibrium. Thus, larger differences between methods can be observed with a longer validation horizon. In addition, we should expect larger differences in performance between prediction methods as the sizes of the reference populations increase in the future.

4.1 Parametric Methods

Consider the linear regression equation where a phenotype is modeled as the sum of additive marker effects:

$$y_i = \mu + \sum_{j=1}^{p} Z_{ij} u_j + e_i \tag{1}$$

Here y_i is the phenotypic observation for individual i, μ is the population mean (ignoring any systematic fixed effects to keep notation simple), p is the number of markers, Z_{ij} is the genotype coding for individual i for marker j, u_j is the additive marker effects, and e_i is the residual effect. The equation is written in matrix form as

$$\boldsymbol{y} = \mu + Z\boldsymbol{u} + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{y}$ is a vector of observations (length n), Z is a matrix with genotypes, $\boldsymbol{u}$ is a vector of marker effects, and $\boldsymbol{e}$ is a vector of random residuals. Genomic breeding values for selection candidates are estimated as

$$\widehat{a_s} = Z_s \widehat{u}, \tag{3}$$

where Z_s is the matrix with genotypes for the selection candidates and $\widehat{u}$ the estimated marker effects.

Treating marker effects as fixed effects yields the ordinary least squares model considered by Meuwissen et al. (2001), but the predictive performance of this model was poor. The number of markers is usually much higher than the number of observations, and the challenge is to obtain estimates of marker effects that yield good predictive performance of genomic breeding values. This can be achieved by elaborated prior distributions of marker effects (i.e., treating marker effects $\boldsymbol{u}$ as random effects) and/or choice of estimation method. Estimation methods that have been evaluated rely on variable selection, shrinkage, or a combination of both. A short summary of the shrinkage methods is given in the sections below; an extensive treatment of all different approaches is outside the scope of this text, and readers are referred to reviews (e.g., de los Campos et al. 2013a; Gianola 2013; Kärkkäinen and Sillanpää 2013).

Shrinkage methods attempt to balance goodness of fit and predictive value by minimizing an objective function consisting of a measure of lack of fit (e.g., residual sum of squares or log likelihood) and a penalty term that causes estimates to be shrunk toward zero. Several options exist for the penalty term: in ridge regression, the penalty is proportional to the sum of squares of estimates of $\boldsymbol{u}$ (L2 norm), and in LASSO (least absolute shrinkage and selection operator), the penalty is proportional to the sum of absolute values of $\boldsymbol{u}$ (L1 norm). The elastic net algorithm uses a weighted combination of the sums of squares and sums of absolute values of $\boldsymbol{u}$ as penalty.

The choice of penalty term corresponds to assuming a specific distribution for the marker effects $\boldsymbol{u}$. For instance, application of ridge regression is equivalent to best linear unbiased prediction (BLUP) of marker effects when marker effects are assumed to follow a normal distribution with mean zero and a variance that is the same for all markers (VanRaden 2008). Other prior distributions for marker effects that have been considered in the context of genomic prediction are the Student distribution (Bayes A) and the Laplace distribution (Bayesian LASSO).

Variable selection methods exploit the assumption that only a small proportion of explanatory variables affect the outcome. The motivation to employ variable selection methods in genomic prediction is that not all genetic markers will be associated with a QTN. The expected effect of markers not associated with the QTN would then be zero. Meuwissen et al. (2001) proposed Bayes B, a variable selection approach where a large portion (π) of the markers was expected to have a zero effect and the remaining proportion ($1-\pi$) an effect drawn from a Student distribution. In their approach, the parameter π had to be specified a priori, but other solutions have been put forward to estimate this parameter from the data (e.g., Habier et al. 2011). The

number of components in the mixture is not restricted to two, and prior distributions consisting of multiple mixtures have been applied [e.g., Bayes R (Erbe et al. 2012)].

Equation (2) is referred to as the SNP model because it models the SNP effects $\boldsymbol{u}$ (length of $\boldsymbol{u}$ is equal to the number of markers, p). Interestingly, the SNP model can be reparametrized by substituting $Z\boldsymbol{u}$ with a vector of genomic breeding values $\boldsymbol{a}$ (length equal to the number of individuals, n). Hence, Eq. (2) can be written as

$$\boldsymbol{y} = \mu + \boldsymbol{a} + \boldsymbol{e} \tag{4}$$

If marker effects are normally distributed $\left(\boldsymbol{u} \tilde{} N\left(0, I\sigma_{\mathrm{u}}^2\right)\right)$, the distribution of $\boldsymbol{a}$ reduces to

$$\boldsymbol{a} \sim N\left(ZZ'\sigma_{\mathrm{u}}^2\right) = N\left(0, G\sigma_{\mathrm{a}}^2\right), \tag{5}$$

where G can be regarded as the realized genetic relationships between individuals. So, the element on row i and column j in the matrix $G\sigma_{\mathrm{a}}^2$ is the covariance between phenotypes of individual i and j.

This approach is commonly referred to as GBLUP. The advantage of this reparameterization is that genomic breeding values can be predicted using models and software similar to those used for pedigree-based breeding value estimation (with the pedigree-based relationship matrix replaced by a genomic relationship-based matrix). Furthermore, since the number of individuals in the reference population is typically much smaller than the number of markers, the computational demands are much lower.

The variance components σ_{u}^2 and σ_{a}^2 are the same if $ZZ' = G$. By scaling all columns in Z to have zero mean and variance $1/p$, the variance σ_{a}^2 can be interpreted as the additive genetic variance for the trait, and G is the genomic relationship matrix.

The effectiveness of GBLUP depends on how well the genomic relationship (derived from markers) reflects the actual relationships at QTN. This finding (de los Campos et al. 2013b) motivates studies on other approaches to construct the genomic relationship matrices. These differ, for example, in the definition of the base population (Meuwissen et al. 2011), in the age of the relationships they trace (e.g., Sun et al. 2016), or in the weight that is given to chromosomal segments (e.g.., Shen et al. 2013).

4.2 Semiparametric Methods

In this section, we present GBLUP models where the genomic relationship matrix G can either be smoothed, which will decrease the difference in genetic correlations between individuals, or G can be made more rugged to increase the differences in genetic correlations between individuals. These models can be advantageous because they tend to remove noise in the G matrix and give better genomic predictions when, for instance, marker interaction effects are simulated.

The first models presented are geostatistical kriging and reproducing kernel Hilbert space models. The term kriging is used in the geostatistical literature and is equivalent to empirical BLUP. The aim of kriging in geostatistics is to model the correlation between observations located on a map. The pair-wise correlations depend on the distance between the positions where the observation was recorded. By modeling each position as a random effect, the values at positions without any observation can be predicted. This is similar to genomic prediction, but instead of having relatedness (based on genetic markers), the distances between geographical positions are used.

A common family of correlation functions used for kriging is the family of Matérn covariance functions (named after the Swedish statistician Bertil Matérn). It depends on a couple of tuning parameters and the Euclidean distance between geographical positions. In the application for genomic prediction, the covariance function depends on the Euclidean distance between individuals in terms of their additive relationship.

Ober et al. (2011) showed that the kriging model gives better genomic predictions than the standard GBLUP model for simulated interaction effects. They also discuss the similarity and differences of the kriging model with the reproducing kernel Hilbert space (RKHS) approach of Gianola and van Kaam (2008). Similarly as for geostatistical kriging models, RKHS finds a correlation matrix that smooths the genomic correlation matrix (Morota and Gianola 2014).

Both spatial kriging and RKHS models have been shown to outperform GBLUP in genomic predictions when marker interaction effects are included in simulations. However, they are usually based on an additive specification of the marker data with the coding of marker genotypes being evenly spaced integers (such as 0, 1, and 2). Hence, they are not developed specifically for fitting nonadditive marker effects but are nonetheless much more flexible than the standard GBLUP model resulting in better genomic predictions.

4.3 Models Including Nonadditive Effects

Many animal and plant breeding schemes involve crossing of different breeds or lines or genotypes with the goal of harnessing the beneficial effects of breed complementarity and heterosis. The basis of heterosis are nonadditive effects like dominance or even interactions between loci (Falconer and Mackay 1996). It can be useful to include these effects in the statistical models, if these effects contribute substantially to the traits. The basic idea is to decompose the genotypic value into additive (A), dominance (D), and epistatic (I) values.

4.3.1 Models Including Dominance Effects

The SNP model that fits simultaneously additive and dominance effects of SNPs can be written as

$$\boldsymbol{y} = \mu + Z\boldsymbol{u} + X\boldsymbol{d} + \boldsymbol{e}, \tag{6}$$

where a vector of dominant SNP effects $\boldsymbol{d}$ is included for each of the p SNP markers and an element in the matrix X, x_{ij}, is the indicator variable for the heterozygous genotype of the jth SNP for individual i (Toro and Varona 2010).

In the standard SNP-BLUP, both additive and dominant effects are assumed to have normal distributions:

$$\boldsymbol{u} \sim N\left(0, I\sigma_{\mathrm{u}}^2\right), \quad \boldsymbol{d} \sim N\left(0, I\sigma_{\mathrm{d}}^2\right),$$

The equivalent GBLUP model is

$$\boldsymbol{y} = \mu + \boldsymbol{g} + \boldsymbol{e} \tag{7}$$

Here $\boldsymbol{g}$ is a vector of genomic breeding values (of length n) with
$V(\boldsymbol{g}) = G\sigma_{\mathrm{u}}^2 + D\sigma_{\mathrm{d}}^2$,
where G is the additive and D the dominance genomic relationship matrix.

4.3.2 Models Including Epistatic Effects

The SNP-BLUP model can be extended to include interaction effects between alleles at different loci:

$$\boldsymbol{y} = \mu + Z\boldsymbol{u} + Wv + \boldsymbol{e} \tag{8}$$

where v is the marker interaction effect, a normally distributed random effect, and the matrix W is constructed so that

$$W_j = Z_i \odot Z$$

with subscript giving column index with $j = (i - 1)p + i$ where p is the number of columns in Z and $\odot$ is the direct Hadamard product. Thus, W has n rows and $p \times p$ columns.

The equivalent GBLUP model is

$$\boldsymbol{y} = \mu + \boldsymbol{g} + \boldsymbol{e} \tag{9}$$

with $V(\boldsymbol{g}) = G\sigma_{\mathrm{u}}^2 + H\sigma_{\mathrm{v}}^2$ and $H = G \odot G$ is the epistatic relationship matrix.

However, the extensions of GBLUP in Eqs. (7) and (9) are expected to increase prediction accuracies only if training populations are large, so that marker estimates reflect more the effect of actual QTNs nearby, instead of additive genetic relationships.

5 Accuracies of Genomic Estimated Breeding Values

Accuracies of genomic estimated breeding values are used to quantify the predictive performance and how well can the model predict the real phenotypes of the selection candidates based on information from the reference population. Several attempts have been made to derive deterministic formula to predict accuracy of genomic breeding values (e.g., Goddard et al. 2011). The predictive performance is commonly summarized with two statistics: correlation between genomic breeding values and phenotypes and the coefficient of regression of phenotypes on genomic breeding values. Phenotypes can be actual observations, summary statistics like daughter/progeny deviations (VanRaden and Wiggans 1991), or de-regressed breeding values (Garrick et al. 2009).

Genomic selection studies commonly include an assessment of the predictive performance. This should avoid overfitting of the model that occurs easily as the number of marker effects to estimate is often much larger than the number of observations. Cross-validation is widely used as a technique to assess predictive performance. It divides the reference population into a training and a validation set, estimates marker effects in the training set, and then validates these. Various cross-validation designs have been applied: (1) two-generation scheme, (2) k-fold cross-validation, and (3) repeated subsampling validation. In the two-generation scheme, individuals are assigned to the training or test set based on their generation number or year of birth. The youngest individuals are included in the test set. In a k-fold cross-validation, individuals are divided into k disjoint sets of equal size. In each fold, one set is used for testing, and the other $k-1$ sets are used for training. This splitting is repeated until all sets have been used once for testing. In the repeated subsampling scheme, the reference population is randomly split into a large (e.g., 95%) training and a small testing (e.g., 5%) set. Again the splitting is repeated many times. All these cross-validation schemes have advantages and disadvantages, some of which are discussed by Morota and Gianola (2014), but there is no consensus about which one is the best. The two-generation scheme is the scheme closely resembling a practical genetic evaluation scenario and is the cross-validation scheme most often applied in genomic selection studies.

The purpose of genomic evaluation applied in practice is to predict the performance of future offspring. However, the offspring may be several generations separated from the reference population. In dairy cattle, for example, only the second or third generation of ancestors of selection candidates is included in the reference population in an efficient genomic selection scheme. Nevertheless, most cross-validation studies in dairy cattle have a validation horizon of at most one generation. This is similar in plant populations, especially in the discussed one-stage selection, which is the more common genomic selection scheme (Marulanda et al. 2016). Such a short validation horizon has two main consequences: (1) the predicted accuracy of selection is too optimistic, and (2) the comparison of models may not reflect the actual performance of the models. For example, models better at capturing linkage disequilibrium between markers and QTNs are expected to perform better for a longer validation horizon than models heavily relying on genetic

markers tracing family relationships. A design with a short validation horizon might be a problem in outbreeding plants, as population-wide linkage disequilibrium is large and predictions are more feasible using a family design.

As a concluding remark, the design of a cross-validation study needs to mimic the intended use of genomic breeding values, such that the estimated predictive ability is consistent with the actual application in mind, and several opportunities exist to improve cross-validation studies.

6 Further Advancements of Methods

The methods and models for the use in genomic predictions are continuously advanced. Some of the suggested extensions of the concept of genomic selection are described here. Such extensions are the inclusion of methods for the manipulation of genomes (genome editing), more detailed information (biological information, data on transcriptome or proteome), or improved genotyping tools (use of sequence information and the concept of genomic selection 2.0).

6.1 Integration of Genetically Engineered Individuals

Genetically engineered or genetically modified plants can be found in the food production chain, while the first genetically modified livestock species has only recently been approved for consumption by the FDA after an approximately 20-year approval period. Different techniques can be used for the modification of genomes including transformations, such as microinjection and electroporation. Transformations were the first modifications successfully applied in plant and livestock species. Other modifications include gene knockouts or knock-ins, which are more common in model species like mice, to test the functions and effects of genes. The inhibition of genes for a short time can be done using, for example, RNA interference (RNAi) employing short RNAs. Many of these methods, especially the modification of individuals using transformations, are rather unspecific, and multiple trials need to be done until the modification is successful. Examples of modifications include changes of the product composition [e.g., golden rice (*Oryza sativa*)], introduction of resistance/tolerance against pathogens [e.g., ringspot virus-resistant papaya (*Carica papaya*)], resistance/tolerance against insects [e.g., potato (*Solanum tuberosum*)], resistance/tolerance against herbicides [e.g., soybean (*Glycine max*)], abiotic stress tolerance [maize (*Zea mays*)], and pollination control system (e.g., maize) in plants as well as enhanced growth [AquAdvantage® salmon (*Salmo salar*)], enhanced production [alpha-lactalbumin pigs (*Sus scrofa*)], enhanced metabolism (EnviroPig®), and the production of human drugs [lysozyme goat (*Capra aegagrus hircus*)] in livestock (Forabosco et al. 2013).

A more recently developed method to modify parts of the genome is genome editing, which allows the targeted change of one or few nucleotides at a specified

position in the genome. Genome editing requires programmable nucleases, which were firstly identified in 1996. Methods commonly used for genome editing include zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and the Cas9-guide RNA system (CRISPR) (Gaj et al. 2013). Genome editing is of major interest in plants and also in livestock since modifications are more targeted and success rates are higher. While traditional modifications were applied in plants, their applications were more restricted in livestock. The opportunities offered by genome editing have therefore led to a huge interest of the livestock research community. More than 300 edited pigs, cattle (*Bos taurus*), and sheep (*Ovis aries*) have been developed since 2011, using nonhomologous end joining or homology-dependent repair (Tan et al. 2016). Edited animals were produced via zygotes or somatic cells. The technique can be used to produce animals as potential organ donors (pig), disease models (pig), bioreactors (cattle), and founder animals of genetic lines with enhanced productivity (cattle, sheep, goat) and to introduce disease resistance into populations (pig) (Proudfoot et al. 2015). Other traits of interest in livestock are especially the horn phenotype in cattle, mastitis resistance in dairy cattle, and resistance to the African swine fever in pigs. The selection for some of these traits cannot be achieved using other breeding methods since relevant alleles are not present in the population (e.g., resistance to African swine fever). The selection of other traits will require long selection periods with a high risk of inbreeding, for example, if the frequency of the alleles is too low to allow selection without loss of diversity (e.g., selection against horns).

The introduction of such new tools into livestock breeding programs will require that relationships between individuals are taken into account to decrease the risk of higher inbreeding. A simulation study suggested that the application of a combination of genomic selection and "promotion of alleles by genome editing" might lead to substantial improvements of response to selection (Jenko et al. 2015). However, one prerequisite of genome editing is that QTNs are identified. It was furthermore suggested that the breeding programs need to be adapted to avoid a rapid depletion of genetic variation in the population.

6.2 Inclusion of Biological Information

The initial and currently applied idea of genomic selection is that of a black box approach, where knowledge of the function of the markers used for selection is not considered. Nevertheless, incorporating genotypes from whole-genome SNP arrays into existing evaluation systems has been successful in increasing the accuracy of EBV of young animals for commonly recorded traits (Lôbo et al. 2011; Northcutt 2011; Wiggans et al. 2011). However, the applicability of these predictions is limited to selection within breeds as the prediction ability of the estimated marker effects is

highly dependent on the relationship between the reference population and the selection candidates (Boichard et al. 2016).

If the black box approach of genomic selection is overcome and additional biological information is available, genomic evaluation may become more accurate, especially for crossbreed predictions. One of the initiatives to improve accuracies is the "1000 bull genome project" (Daetwyler et al. 2014). The objective of this project is to make the sequence data of over 1000 influential sires available. This should improve imputation, genome-wide association studies (GWAS), and genomic prediction and, more importantly, promote the identification of causal variants.

The availability of accurately annotated genomes, both structural and functional, is essential for the biological insight into traits (Stein 2001). In order to relate markers to genes and phenotypes, a fully assembled genome with known gene locations and structures, information on noncoding RNA, regulatory and repetitive regions is required. Moreover, the functional annotation like gene ontology (GO) classification that describes products of eukaryotic cells in terms of molecular function, biological processes, and cellular components, as well as descriptions of metabolic and signaling pathways and gene regulatory networks, can provide valuable information. Currently, several such databases are available and updated continuously for a variety of species. Some examples are the GO browser agriGO (Du et al. 2010) that represents 45 agricultural species, including plant, fungi, insect pests, and livestock species, and the Reactome (Croft et al. 2011), MetaCyc (Caspi et al. 2014), and KEGG (Kanehisa et al. 2008) databases that integrate genomic, chemical, and systemic functional information.

The incorporation of biological information into the genomic evaluation can be done in various ways. A simple and straightforward approach is the selection of subsets of markers from the whole-genome SNP arrays that are associated with genes or metabolic pathways of interest. This could be extended to include a polygenic component, using pedigree relationships to account for the rest of the genome (Snelling et al. 2011). Moreover, the priors of Bayesian models could be shaped by biological knowledge and become more informative (MacLeod et al. 2016).

6.3 Transcriptome and Proteomic Assisted Selection

High-throughput technology is not only applicable to the information of the genome but also transcriptome, proteome, and metabolome. Information on the transcriptome, such as data from RNA sequencing, does provide information on mutations within the genome and adds knowledge of probable functionality as only expressed genes will contribute to the phenotype. High-throughput platforms such as expression arrays may further allow collecting expression information for many loci and individuals. High-throughput platforms do also exist for the analysis of the proteome (Chawade et al. 2016). Peptide-based selection using mass spectrometry might assist selection for certain phenotypes. Its application had been tested in plants and allowed the selection for traits for which no good genetic markers were available (Chawade et al. 2016). The analysis of metabolites and their variance was suggested as another (post-genomic) tool for

improved selection (Fernie and Schauer 2009). The use of metabolomics-assisted breeding, possibly in combination with sequencing and reverse genetics, might be useful for a number of traits including selection for resistance and tolerance traits in plants (Zamir 2001; Morandini and Salamini 2003; McCouch 2004; Takeda and Matsuoka 2008; Fernie and Schauer 2009). The feasibility of the use of expression profiles or protein signatures in future breeding systems has yet to be explored. A combination of tools based on traits might be a possible scenario for the improved selection especially for the improvement of complex traits.

6.4 Alternative Genotyping Methods

Genotyping of many individuals or lines is a prerequisite for genomic selection. The density of genotyping platforms required for a reliable prediction depends on the population of selection candidates and its genome structure. Genotyping arrays with various densities are available. The use of customized and population-specific arrays with lower marker density to genotype selection candidates and combining these with sequence data of influential ancestors of the selection candidates can reduce costs for genomic selection. Such approaches display alternatives to the use of high-density genotyping arrays and are applied in some breeding programs. The imputation of genotypes can additionally be used to increase the information content. The process of imputation implies that genotypes are predicted, which are not directly assayed in a sample. Ancestors will be genotyped using information on the full genome sequence or high-density marker arrays. If information on such dense genotyping is available, most haplotypes in the populations are covered and thus can be implied in individuals with information on fewer genetic markers (Marchini and Howie 2010). The increase of genotype information in the population might improve the accuracy of genomic selection (Druet et al. 2014). Imputation can also be used for the correction of genotyping errors. However, one essential step to allow accurate imputation is the correct phasing of the genomic information (Hickey 2013; Hickey et al. 2014).

Other alternatives to higher- or lower-density genotyping arrays exist. These should allow to reduce costs by skipping the need to develop genotyping arrays. The use of genotyping by sequencing (GBS) opens opportunities to fill the gap between highly explored lines of major interest and non-reference lines (Spindel et al. 2013; Williams et al. 2014). Genotyping by sequencing is especially of interest when little or no genomic information is available, no dense genotyping platform exists, or genetically highly diverse material is used. This approach is also useful for large genomes. It is therefore of interest especially in plant breeding. However, well-established bioinformatic infrastructures are required to fully explore genotyping by sequencing.

6.5 Genomic Selection 2.0

The term "genomic selection 2.0" was introduced based on the advance of tools for genotyping and sequencing (Hickey 2013). It was suggested that, while progress has been made using genomic selection in general, the large amounts of sequence data generated could be utilized even more. Genomic selection 2.0 is based on the use of big data and the availability of sequence data in combination with imputation methods and aims to integrate new methodologies for the integration of de novo mutations and variants different from SNPs. The genomic information of a huge data set in combination with phenotypic information should be powerful to identify QTNs for many traits. Genomic selection 2.0 intends to avoid sequencing at a high depth, which is not feasible for breeding derived from a large number of male ancestors, as in livestock breeding, or many potential breeding lines, as in plants. The choice of a lower-coverage sequencing for all individuals could assist to discover the haplotypes and allow the imputation to full sequences for all individuals. The power of methodologies will depend on the advancement of applications such as imputation algorithms, technologies for genotyping, and infrastructure of bioinformatic analysis. Furthermore a new generation of genomic selection is an important step to allow higher recombination in populations. It should allow the integration of de novo mutations, occurring from random events during recombination or being introduced by methods such as genome editing (Hickey 2013).

7 Genomic Prediction Applied in Animal, Plant, and Human Populations

7.1 Examples of Genomic Selection in Livestock

7.1.1 Cattle (*Bos Taurus*)

The practical application of genomic selection is of immense interest in dairy cattle, where GEBVs are now common selection criteria across many countries (http://www.interbull.org 2013). This is not the case for other livestock populations, especially in populations where crossbreeding is used, such as pig or beef cattle. Also the lack of phenotypes and genotypes for a reliable prediction of a GEBV or the focus on traits with lower heritability, such as fertility in beef cattle, does restrict the use of genomic selection (Johnston et al. 2012). Genomic selection is expected to be more cost-efficient compared to traditional selection schemes as it allows the reduction of the expensive evaluation of phenotypes and allows an earlier selection of male animals for further breeding. A concern is often the cost for genotyping (of the reference population), but with the ongoing development of genotyping platforms, this might be less of a problem in the near future. Changes might need to be applied to the structure of the breeding industry, which will require a long-term planning (Johnston et al. 2012). But the question is how genomic selection is implemented

in different breeding programs, and which relevant aspects of the species and breeding program have to be taken into account?

Genomic selection is today a selection scheme in dairy cattle, especially Holstein Friesian, in many countries (Patry 2011). Decreased costs for genotyping had progressed genomic selection to a routine in some herds (Hayes et al. 2009a; Hayes et al. 2009b). Jannink et al. (2010) stated that the implementation of SNP information will reduce the costs for genetic evaluation. The cost for obtaining marker information can be equal to the costs to collect phenotypic information from 10 to 20 daughters per bull. Selection in dairy cattle is focused on bulls, and many breeding programs focus on genotyping solely bulls. Costs for genotyping can therefore be kept relatively low. The breeding decisions can be made by preselecting young bulls for further testing in the so-called preselection scheme. An alternative option to make use of the genotypes is the turbo scheme which allows the earlier selection of new breeding bulls (Pryce and Daetwyler 2012; Bouquet and Juga 2013). Additional genotyping of cows will allow a better assessment of additional traits and to increase the size of the reference population. One advantage of genomic selection in dairy cattle is the drop of the length of the generation interval from around 5 to 6 years in traditional dairy cattle breeding programs to around 1.5 years when using genomic selection (Pryce and Daetwyler 2012). The increase of the genetic gain in general might be attached to the risk of higher inbreeding as it might reduce the number of genetically superior breeding animals. Genetic markers should therefore also be used to avoid loss of diversity by carefully observing the remaining of haplotypes and the structure of the population (Young et al. 1988).

The implementation of genomic selection is also of major interest for beef cattle breeding, for which generation intervals are also long. A number of differences compared to dairy cattle exist, including the lower rate of the use of artificial insemination and the use of crossbreeding. The lower rate of artificial insemination, compared with dairy cattle, reduces the contribution of a selected individual to the genetic progress in the population at large and thereby reduces the amount of resources that can be invested in genotyping. The genetic makeup of populations is relevant, and genomic selection would probably be restricted to purebred operations. Predictions in crossbred populations are not as accurate as compared to those in purebred populations. Beef cattle populations are less uniform compared to dairy cattle populations, crossbreeding is common, and both *Bos taurus* and *Bos indicus* populations are a part of breeding schemes (Garrick 2011). The effective population size in many beef cattle population is low as is the number of bulls with reliable EBVs. This will restrict the reference population and the reliability of the estimated GEBVs (Johnston et al. 2012). The combination of data across countries and/or across breeds is an option to overcome the small size of reference populations, but higher-density marker panels might be required to reach reliable predictions when using such datasets (de Roos et al. 2009). Genotypes of cows can also be included to achieve larger reference populations, additionally allowing farmers to select superior cows (Saatchi et al. 2012). This would allow a better selection for fertility, one of the most important traits in cows, which has a low heritability. The inclusion of cows in the selection scheme would require changes to traditional breeding using progeny testing, which focusses largely on bulls. Genomic selection in beef cattle could thereby lead to a more balanced breeding goal via inclusion of animals and traits at the farm level.

Genomic Selection in Dairy Cattle

The seminal publication by Schaeffer (2006) illustrated that adoption of genomic selection could decrease the costs of running a breeding program and increase genetic progress, compared to progeny test schemes that had been in place for many years. The development of a high-density SNP array (Matukumalli et al. 2009) removed a last practical hindrance for the implementation of genomic selection.

The first official release of genomic breeding values was in 2009 in the USA (Wiggans et al. 2011). At that time, just over 5000 progeny-tested bulls were included in the reference population. Using an approach resembling Bayes A, reliabilities of genomic breeding values were on average 50% (VanRaden et al. 2009a, b). This meant an increase of 23% in reliability compared to the reliability of parent averages.

Dairy producers in the USA were quickly to adopt the technology, and by 2012 half of the Holstein service sires were genotyped as young bulls, i.e., bulls with just genotype information and no daughter information (Hutchison et al. 2014). Also breeding companies made changes to their breeding programs and started to use genotyped young bulls as sires of sons. As a consequence, Hutchison et al. (2014) and García-Ruiz et al. (2016) could observe a significant decrease in the generation interval.

Evidence of increased of genetic gain due to genomic selection was presented by García-Ruiz et al. (2016), who reported that the genetic gain for yield increased twofold after the introduction of genomic selection. For fertility, life span, and udder health, even larger increases in genetic gain were observed, in agreement with the prediction that genomic selection would be especially useful for traits with low heritability.

In recent years, focus has been expanded from genotyping predominantly males to genotyping females as well. In July 2017, a new milestone was reached in US dairy genetics with the submission of the two millionth genotyped animal to the US dairy database (Press release, 2017; https://queries.uscdcb.com/News/CDCB%20AGIL%20Two%20Million%20Genotype%20Mark.pdf). Genotyping females will allow commercial dairy farmers to make more informed breeding decisions in their own herd but also provides new opportunities for improved herd management.

7.1.2 Sheep (*Ovis aries*) and Goats (*Capra aegagrus hircus*)

Breeding of small ruminants, sheep and goats, varies as the size and structure of enterprises differ between countries. Small ruminants are especially part of the production system in low-income countries as the resource inputs are low. However, larger breeding cohorts exist in countries with options for higher input and selective

breeding based on performance information (van der Werf 2007). Differences in the management, structure, and size of the populations and breeding programs depend on the product (meat, wool, or milk) and also the location of the farm. The use of local breeds is more common for sheep and goat breeding; thus populations are small and breeding is more often country-specific. The integration of genomic selection into breeding programs is especially discussed for countries with large breeding populations, such as Australia, New Zealand, Great Britain, South Africa, or France. Reference populations have been established in some countries for the collection of reliable phenotypes (Swan et al. 2012). The main restriction of the application of genomic selection in small ruminants is the lack of the data/information on a large number of phenotypes, which is necessary for the creation of a reliable reference population. The shorter generation interval in small ruminants (compared to cattle) and the relatively high genotyping costs will restrict the genetic gain when using genomic selection. Effective population sizes are often large in small ruminants as these populations are usually more heterogeneous compared to other livestock populations. And finally natural service is still more common in many sheep breeding schemes, which will restrict the number of possible fertilizations from each ejaculate to one. More rams are required when natural service is used instead of artificial insemination. Other relevant points to be considered when applying genomic selection in small ruminants are the options for the size of the reference population (Shumbusho et al. 2013); population-specific factors in sheep, including seasonality of the production system; small-scale use of artificial insemination; and low value of individual animals, which will require different approaches of genomic selection compared to dairy cattle (Baloche et al. 2014). Predictions using crossbreed animals are also relevant in sheep. This should allow the application of genomic selection in a larger range of breeding populations and covering existing breed diversities in populations.

7.1.3 Pigs (*Sus scrofa*)

Separate selection schemes at the nucleus level, one for the paternal production-oriented breeds and one for the maternal reproduction-oriented breeds, exist in pig breeding. This structure needs to be considered when using genomic selection. Traditional selection has a larger focus on performance traits with a selection of superior sire lines for improved carcass and meat traits. While genomic selection in male lines can improve selection efficiency/effectiveness, phenotypes of relatives might be needed to increase the reference population (Tribout et al. 2012). Genomic selection could also take better care of the selection for maternal traits. The shift of the focus to maternal traits will especially be feasible when total costs of genotyping are reduced. Simulation studies have shown the improved accuracies of selection for economically important traits also in female purebred lines (Lillehammer et al. 2011; Tribout et al. 2012). However limitations exist for the prediction of performance of crossbred animals, which are usually used in the final stages of the production and often as maternal lines. Genetic correlations between traits in cross- and purebred

animals are less than 1 (Dekkers 2007). Suggestions to overcome the limitation of the less than unity genetic correlation between crossbred and purebred performance have been made, such as the integration of QTL information into breeding decision.

7.1.4 Poultry/Chicken (*Gallus gallus domesticus*)

Poultry has dual use and is therefore bred in different lines to allow for differential selection for egg and meat production. The generation interval in poultry is relatively short with 1 to 1.5 years, and the rate of genetic improvement in traditional breeding is more than double in chicken compared to cattle or pigs. The implementation of genomic selection could reduce the generation interval to only 6 months. But the population sizes need to be carefully evaluated to reduce costs of genotyping on the one hand while not reducing the effective population size. Adequate genotyping platforms have been recently developed, and first predictions showed the potential for genomic selection to increase genetic gain in poultry breeding (Preisinger 2012). However, the advantage and cost-efficiency over conventional breeding have to be proven before genomic selection could be applied as a selection tool in privately owned poultry breeding companies (Preisinger 2012). Hypothetical studies have also suggested the implementation of genomic selection in broiler lines. However, genotyping strategies need to be chosen carefully to reduce costs without the loss of important information on marker-phenotype relationship (Avendaño et al. 2010).

7.1.5 Aquaculture

The introduction of genomic selection has also been discussed in aquaculture, especially fish. Genotyping tools will enable to control inbreeding, but costs are currently the main inhibitor for a quick adoption of genomic selection in fish breeding schemes (Nielsen et al. 2011). Male and female fishes have many offspring; the contribution of male and female individuals to the breeding cohort is therefore high. If the use of genomic markers can improve selection in fish breeding, the expected genetic gain can be twice as high compared to traditional selection using BLUP. The use of genetic markers could also assist controlling inbreeding more effectively (Nielsen et al. 2011). Aquaculture breeding programs might need to be redesigned entirely to accommodate genomic selection. Such changes can be a reduction of number of families or reduced phenotypic evaluation (Sonesson and Meuwissen 2009; Nielsen et al. 2011). A combination of traditional BLUP estimation, preselection of candidates, and low-density genotyping arrays might be one possibility. It could reduce costs for genotyping many potential parents and thus reduce the expected genetic gain only slightly (Lillehammer et al. 2013).

7.2 *Examples of Genomic Selection in Companion Animals*

Estimated breeding values are used successfully for selection in some horse and dog populations, and genomic selection is discussed for further improvement. The lack of large enough reference populations is often the restricting factor for the implementation of genomic selection. Most dog breeds are based on a few founders, and the effective population size is relatively small. The use of genetic markers for selection is feasible. Genotyping arrays are available for both horses and dogs. Genetic markers should improve the predictive ability and lead to a more accurate selection. On the other hand, many of the traits used for selection are complex and not always measureable on a reasonable objective scale. The implementation of genomic selection will require a careful design of an appropriate reference population with reliable and relevant phenotypes.

7.2.1 Dogs (*Canis lupus familiaris*)

Only a few studies have investigated the application of genomic selection in dogs (Sánchez-Molano et al. 2015). Dog breeding is usually done based on pedigrees and phenotypic measurements. Breeding goals include improved health with traits aligned to the breed standard while avoiding inbreeding. Inherited disorders, such as hip dysplasia, heart problems, and certain kinds of cancer, put traditional dog breeding into negative lights and need to be taken into account in breeding programs. Since traits related to health often have a late onset, information from genetic markers using data from a large reference population would therefore be useful. The use of genomic selection or prediction models would also allow a better correction for environmental factors (among which the influence of the breeder). Problems to overcome in dog breeding before genomic selection could be applied are the need of collecting data from many dogs for developing a reference population of appropriate size and the need for reliable phenotypes and for continued phenotype collection after the introduction of genomic breeding values. Genomic selection is a potential tool to improve selection and especially traits related to welfare (such as health or inherited defects) of pedigree dogs.

7.2.2 Horses (*Equus ferus caballus*)

The sport horse industry aims for a more accurate selection to reach high genetic improvements. Generation intervals in horses are around 8–10 years, and earlier selection, for example, by using genetic markers (Haberland et al. 2012; Stock et al. 2016), could increase the rate of genetic improvement. Some of the traits of interest are also related to behavior and temperament, which are difficult to measure objectively. The establishment of international collaborations is not always straightforward. Limited exchange of genetic material leaves many small and (semi-)

isolated populations at risk of decreasing effective population sizes, increased inbreeding, and potential increase in prevalence of inherited diseases. Large reference populations with reliable phenotypes are needed to apply genomic selection with high accuracies. Genomic selection will improve accuracies achieved with EBVs in young animals and also horses imported from other countries for which only scarce information on relatives is available in the importing country. Genomic selection will be especially useful for traits with late onset and low heritability. In a comparison of several selection strategies against osteochondrosis, van Grevenhof (2011) found genomic selection to be a realistic option for the Dutch warmblood population. Similar to dog breeding, a very relevant aspect in horse breeding is the structure of many small-sized studs and fewer large enterprises, compared to livestock breeding, with its challenges to achieve the level of collaboration needed to put in place an organized scheme necessary for a successful implementation of genomic selection. Also international collaborations will be necessary as they have the potential to increase the reference population.

7.3 Examples of Genomic Selection in Crop Plants

The status of the implementation of genomic selection in different crop species varies. Genomic selection is of interest to the public and private crop breeding community. One main reason for the search of improved selection tools is the stable and high costs for phenotyping. As little can be done to reduce costs per line, the only option is a reduction of the number of lines to be phenotyped. The crop breeding community hopes for a significant reduction of costs for the development of new breeding lines when using genomic selection instead of traditional phenotypic selection (Heffner et al. 2009; Heffner et al. 2011; Resende et al. 2012b). But the implementation of genomic selection will depend on costs for genotyping and the availability of whole-genome sequencing and/or genotyping platforms. Crop breeding programs are versatile, and strategies for the implementation of genomic selection will need to be adjusted for each breeding program. Hybrid vigor or heterosis is important in many crop breeding populations, and models for genomic selection should also be able to take nonadditive effects into account (Duvick 1999).

Traditional selection is often based on phenotypic selection. Breeding of inbred lines for the production of hybrids and crossing of diverse parental lines for the production of new inbred lines in successive cycles of selfing are the two main strategies. Phenotypes might differ between the plant materials used for selection in early and advanced cycles of breeding because the number of tested lines in early cycles is often too large for a cost-effective collection of all relevant phenotypes. The use of phenotypes from the final cycles of breeding might therefore reflect more useful data as they will most accurately reflect the final product. The estimation of marker effects based on advanced cycles of selection (Zhao and Xu 2012) needs to be considered carefully. Nonadditive effects due to heterosis or inbreeding effects can change the prediction accuracies. The choice of the reference population has to

consider these effects. Also the structure of the reference populations is important, and one option for a high accuracy of prediction is that individuals in reference and validation subpopulations should show a close relationship (Asoro et al. 2011).

An additional difficulty in crop species is the impact of genotype by environment (GxE) effects on performance of lines (Lorenzana and Bernardo 2009; Heffner et al. 2011). Advanced generation populations in traditional breeding are therefore tested across different environments. The adaptability of crop lines is a relevant criterion for the successful production in the field. The evaluation of genotyped lines across different environments can increase the gain and cost-efficiency of genotyping (Bertin et al. 2010; Xu and Hu 2010; Morrell et al. 2012). Sequencing of selected lines in combination with the repeated collection of phenotypic data has been another suggestion to overcome an inaccurate estimation of genomic breeding values caused by genotype by environment effects (Morrell et al. 2012).

7.3.1 Rice (*Oryza sativa/Oryza glaberrima*)

Reports on genomic selection in rice are rare, and pedigree breeding based on phenotypes is still the predominant breeding method (Li and Zhang 2013). Successes have, for example, been made in increase of yield, but yield potential needs further improvement in the future. Also selective breeding success stories to improve complex traits (such as drought tolerance) are limited. One of the reasons for this limitation is the lack of information on reliable phenotypes especially from hybrid breeding, which is increasingly common in rice (Yan et al. 2011). There is furthermore little genetic variation in the current breeding populations; genotyping might assist to adjust breeding strategies to avoid the loss of important genes due to a more narrow gene pool (Breseghello 2013). Genotyping can also assist to identify more diverse parental lines, which can then be used to achieve high heterosis effects in crossbred populations (Chen et al. 2013). But more research is needed to fully exploit the possibilities of genomic selection in rice.

7.3.2 Maize (*Zea mays*)

Large efforts are underway for the implementation of genomic selection in maize, another important crop in many countries around the globe. Significant improvements have been made since the domestication, and there is little resemblance between the original Balsas teosinte (*Zea mays* ssp. *parviglumis*) before domestication and modern maize plants today. Improvements are especially focused on tassel, ear, cob, and kernel characteristics, flowering traits, as well as resistance to drought and pathogens. Genomic selection will improve the breeding process further as it allows the prediction of untested lines, including testcrosses, in advanced breeding populations (Albrecht et al. 2011). One additional advantage of the application of genomic selection in maize is the reduction of the generation interval. Phenotypic evaluation is not required throughout the entire selection process when genomic selection is used, and generations of lines can be bred in greenhouses (Zhao et al.

2012). The design of the reference population requires good knowledge of the population structure and genetic relationships within and across relevant lines (Albrecht et al. 2011; Windhausen et al. 2012). Biparental or diversity panels and testcrosses are important/useful in maize breeding. Advanced breeding populations are often based on the performance in many testcrosses. But genotyping of all testcrosses will be expensive, and preselection is required. Genetic markers can be used to investigate genetic differences between lines to select more diverse individuals for crossing (Albrecht et al. 2011). Different strategies for genomic selection are being tested in maize. International centers, such as the International Maize and Wheat Improvement Center (CIMMYT), drive research in this sector.

7.3.3 Wheat (*Triticum aestivum*)

It has been shown that the implementation of genomic selection can lead to higher genetic gain per unit time and cost reduction compared to traditional pedigree-based selection in wheat (Burgueno et al. 2012). But the use of related populations in the reference and selection set is a major factor to achieve a reliable accuracy when using genomic selection in wheat (Crossa et al. 2013). Information from the reference population for predictions needs to be collected in environments which are similar or the same to those of the selection candidates as environments play a major role (Crossa et al. 2011, 2013). Accuracies across field trials can be increased when information based on many lines and different environments are included in the modeling of the genetic effects (Burgueno et al. 2012; Dawson et al. 2013). Wheat breeding focusses on a number of traits, including grain yield, quality traits, tolerance to abiotic stresses (drought and heat), and disease resistance, as listed in a review from the CIMMYT breeding scheme (Guzman et al. 2016). A good phenotypic recording of disease traits is important for the selection of more resistant lines. However, evaluation of infection traits is time-consuming and costly. Genomic selection can be used as a strategy to improve the gene pool for resistance (Rutkoski et al. 2011) and also other relevant agronomic traits. It allows the implementation of historic information and the prediction of many traits at the same time. No additional costs may incur if lines are selected based on predicted phenotypes from genomic information as shown in the example of the first wheat line traits of the CIMMYT breeding program (Guzman et al. 2016).

Genomic Selection in Spring Bread Wheat: CIMMYT's Breeding Efforts

The International Maize and Wheat Improvement Center (CIMMYT) has discussed the use of genomic selection for the improvement of their wheat and maize breeding programs early on. The spring bread wheat program is one of the examples in which genomic selection has been tested, and details of the

(continued)

program have been described in Battenfield et al. (2016), and a summary is provided here.

The F7 spring bread wheat lines were derived from F5 lines, which were tested and evaluated for quality traits for 1 year in Mexico. Superior lines from testing were chosen for advanced end-use quality testing. Five plants per lines were used for genotyping using genotyping by sequencing with further imputation of missing genotypes. Marker effects were calculated using a number of models including ridge regression best linear unbiased predictor, reproducing kernel Hilbert space, partial least squares regression, elastic net, and random forest. The efficiency of the models in predicting breeding values was tested using cross-validation on data across multiple years trained at 80% randomly selected data to predict 20% masked data, as well as forward prediction trained on all prior data. The data collection for the described spring bread wheat modeling started with trials harvested in 2010 and included a total of 47,817 lines in the yield trial, of which 7858 lines had been screened for quality. From a total of 5520 of these lines, phenotypes and genotypes were available until 2015.

When comparing the results of predictions using cross-validation and forward prediction, it was concluded that cross-validation will likely lead to an overestimation of the prediction ability of genomic selection. On the other hand, only small differences were observed between the predictive abilities of using different models for genomic selection. Correlations between the observed and predicted phenotypes differed for different traits and varied between years. The response to selection using phenotypic and genomic selection increased between 35% (test weight, kg h L^{-1}) and 147% (alveograph P, tenacity divided by L, extensibility, mm mm^{-1}). One main advantage when implementing genomic selection in the CIMMYT spring bread wheat program is the possibility to select for phenotypes, such as wheat quality, which will, in a phenotypic selection program, only be used as selection criteria during late stages of the breeding program. This advantage is common to many other crop species, most of which evaluate major traits of interest only late in the breeding pipeline. Accuracies from the tested models in the spring bread wheat program were high enough to allow the application of genomic selection and increased with larger training populations. Genomic selection will allow a reduced phenotypic evaluation, which currently requires more seed material and which represents a considerable cost factor. While genomic selection might not replace the collection of phenotypes, it will allow early selection of future breeding material. A 1.4 to 2.7 times greater gain from selection was further predicted when the number of selection candidates increases from 2000 to 10,000. The implementation of genomic selection in the CIMMYT spring bread wheat breeding program has started in 2012, and it is predicted that it will enable the selection for specific end-user traits.

7.3.4 Barley (*Hordeum vulgare*)

A shorter breeding cycle and thereby early selection gain is also the expectation when using genomic selection in addition to phenotypic evaluation in barley breeding. The accuracy of genomic selection is higher if correlations between reference and selection populations are high and/or trait heritabilities are low (Iwata and Jannink 2011). Typical traits used in the breeding goals are yield or yield-related traits (grain dry matter yield or thousand kernel weight), quality traits, and resistances against diseases. A carefully selected reference population may allow an improvement using genomic selection compared to phenotypic selection even in biparental crosses (Jannink et al. 2010). But the predictive ability depends often on the relatedness, the population structure needs therefore to be taken into account, and the number of markers required will depend on population structure and the linkage phase (Thorwarth et al. 2017). A reference population might use more inbred or highly replicated samples, more diverse samples, or lines different from the population used for phenotypic selection. Genomic selection might also help to improve decisions on crossbreeding (Bernardo 2010), if a reference population is well selected. This is especially relevant in self-pollinating plants, such as barley, where time-consuming crossing by hand is required in order to produce biparental crosses. However, some studies express concerns regarding the risks of lower genetic variation due to the loss of favorable alleles (Jannink 2010), especially when breeding cycles are shorter.

7.3.5 Other Crop Species

Other crop species, with complex breeding goals, are forage plants, for which the aim is to increase production as well as maximize perennial persistency. Perennial forage grass [mostly ryegrass (*Lolium perenne*)] plots should be used with a consistent quality and quantity over many years; deployment of hybrid breeding is, therefore, not applicable (Wilkins and Humphreys 2003). Genomic selection should especially improve the prediction when correlations between phenotypic evaluation and performance are low, such as for complex traits or traits that could be recorded only in advanced reproduction cycles. The use of genetic markers might assist the reduction of the lengthy periods for phenotypic selection (Hayes et al. 2013; Resende et al. 2014). However, data and sample management from parental lines, including recordings of pedigree information, need to be improved. Genomic selection should allow a focus on a few traits during the phenotypic evaluation and will enable to control that relevant alleles remain in the breeding cohort (Resende et al. 2014). The use of genetic markers might be more efficient for the introgression of specific genes compared to backcrossing (Wilkins and Humphreys 2003).

Application of genomic selection has, until now, been less discussed for other crop species including examples from the genus *Brassica* (Cowling et al. 2009; Cowling and Balazs 2010; Cullis et al. 2010; Wurschum et al. 2014), oats (*Avena*

sativa) (Asoro et al. 2011), potato (*Solanum tuberosum*) (Barrell et al. 2013), sugar beet (*Beta vulgaris*) (Hofheinz et al. 2012; Wurschum et al. 2013), sugarcane (*Saccharum officinarum*) (Gouy et al. 2013), or soybean (*Glycine max*) (Shu et al. 2013). Some restrictions are the availability of genotyping tools, sizes of possible reference populations, as well as the need for further improvements in evaluation of phenotypes.

It had been suggested that modifications to breeding programs (such as number of lines per breeding cycle, number of test staged in the program, more collaborations between breeders) might be needed to achieve economic gain via genomic selection (Cowling and Balazs 2010; Hayes et al. 2013). It is important to keep in mind that the selection unit is not a single plant but a heterogeneous line, variety or plot. Genomic selection needs to be adapted to address the traits and structure of the distributed product, breeding schemes which are used to produce seeds from inbred or hybrid lines for the use by farmers.

7.4 Examples of Genomic Selection in Trees

The generation interval, breeding cycle, and duration until phenotypes can be evaluated in tree breeding are long. The identification of better estimators for the quality seedlings for the production is therefore a major interest for the forest and fruit tree industry. Advantages of using genomic selection will arise mainly from the shorter selection cycles (Iwata et al. 2011).

7.4.1 Forest Trees

Testing different scenarios of genomic selection in eucalyptus breeding for height and diameter at multiple ages allowed the total breeding cycle to be halved (Resende et al. 2012a). Intensive progeny testing can be eliminated, and a second clonal trial will not be needed allowing for good economic returns (Resende et al. 2012b). Methods to reduce the maturity age (breeding duration) and speed up propagation are already implemented in tree breeding. However, emphasis should be put on reducing the testing phase if the total breeding interval needs to be reduced (Resende et al. 2012a). Even though it had been concluded that genomic selection will, alongside other reproductive methods, decrease the total time of a breeding cycle in conifers, it has also been seen that models predicted early during the breeding cycle, for example, in seedlings, have only limited applicability for the selection of older trees. Also the comparisons of predictions across locations did not lead to high accuracies for all scenarios (Resende et al. 2012a). Additionally, genetic regions explaining trait variation were often population-specific as shown using eucalyptus populations (Resende et al. 2012b). Older data and genotypes within the same breeding scheme including crossings of the same elite trees might therefore be

more useful to create a reference population aiming for high accuracies, as suggested in some of the scenarios in conifers (Iwata et al. 2011).

A number of studies had been conducted to identify markers associated with relevant traits (e.g., wood quality, wood formation, growth, hardiness, drought response, disease resistance) in trees, but not many of those markers are currently being used in breeding programs (Thavamanikumar et al. 2013). Genomic selection has been tested as a theoretic approach in forest trees using simulated datasets; however, many studies show the application of real data (e.g., Resende et al. 2012a; Beaulieu et al. 2014b). Genomic selection has especially been suggested as a useful tool in elite breeding programs where relatively low number of markers are adequate to cover structures of linkage disequilibrium (Thavamanikumar et al. 2013). But the rapid decay of linkage disequilibrium in most tree populations is one of the main problems identified in studies. It was suggested that this limitation could be avoided when using elite trees and thereby introducing a genetic bottleneck. A prediction model built on data from progeny of crosses between elite trees can additionally be used to select elite trees via genomic selection. A study compared estimated breeding values and genomic breeding values using cross-validation within clones from half-sib families of loblolly pine (*Pinus taeda*). Even though derived accuracies were relatively high, this was suggested to be due to family linkage rather than identified historic linkage disequilibrium as only few genetic markers were used (Zapata-Valenzuela et al. 2012). A study in maritime pine (*Pinus pinaster Ait.*) showed good predictive ability for different traits, despite the low marker coverage and low linkage disequilibrium (Isik et al. 2016). A more comprehensive breeding scheme was simulated for a population of conifers (Iwata et al. 2011) for which different scenarios were tested for a 60-year breeding program in a seed orchard. The use of genetic markers in a genomic selection scheme could also provide additional information on parentage since some of the traditional tree breeding programs, for example, in eucalyptus breeding, are open-pollinated (Zelener et al. 2005). Studies have also shown the potential of genomic selection to improve traits in spruce compared to traditional pedigree-based selection (Beaulieu et al. 2014a, b; Ratcliffe et al. 2015; Lenz et al. 2017). The potential of genomic selection over traditional breeding has been shown in recently domesticated or undomesticated populations of trees (e.g., white spruce) but has been suggested for within populations or families due to the low marker coverage (Beaulieu et al. 2014a, b).

Genomic Selection in *Eucalyptus*

Conventional tree breeding is typically characterized by long breeding cycles. Hybrids are often preferred in *Eucalyptus* breeding schemes as they are superior to their parents in the most relevant traits, including growth, wood quality, and biotic and abiotic stress resistance as they inherit relevant characteristics from each of the parents (Tan et al. 2017). The cycle of a conventional breeding scheme in *Eucalyptus* can take between 12 and 18 years;

(continued)

genomic selection does, therefore, offer new opportunities as it might reduce this cycle. However, when selecting superior tree clones in hybrid eucalypt breeding, both additive and nonadditive effects are relevant (Resende et al. 2017). Relatedness of selection and training population can additionally lead to over- or underestimation of the prediction accuracy. It was suggested that a high marker density will be advantageous in such situations (Resende et al. 2017).

Tan et al. (2017) and Resende et al. (2017) used the Illumina Infinium EuCHIP60K, which includes more than 45,000 SNPs to study controlled crossings of *E. urophylla* and *E. grandis* trees. The aim was to test genomic selection for the selection of superior F_2 individuals for traits height, volume, circumference at breast height, basic wood density, and screened pulp yield. Genomic best linear unbiased prediction, ridge regression best linear unbiased prediction, Bayesian LASSO, and reproducing kernel Hilbert space regression were tested in these studies. Predictive abilities of the genomic selection models differed based on the selection scheme, with the highest predictive abilities obtained from cross-validation in a between-family selection including full- and half-sib individuals (Resende et al. 2017). The mean accuracies varied between 0.34 and 0.54 depending on the traits and reached maximums of 0.73 to 0.87 in the best scenario based on relatedness. The predictive ability using different models varied from 0.27 to 0.274, but all models of genomic selection did outperform other pedigree-based predictions. Also this study showed that the relationship between training and selection candidates, as well as the size of the training population, had a large impact on the predictive ability (Tan et al. 2017).

It was concluded from both studies that (a) genomic selection will reduce the time until superior breeding lines are selected and (b) data obtained from genotyping provide additional information on the genomic relationship matrix and can be used for the estimation of heritability. However further issues need to be resolved, such as the selection across generations and environments. The inclusion of nonadditive effects and the estimation in hybrid breeding as purebred parents will not provide information for accurate predictions in hybrid offspring.

7.4.2 Fruit Trees

Traits of interest for breeders of fruit trees are fruit quality (e.g., firmness, astringency, soluble solids, and acidity), precocity, yield, and disease resistance. The selection using traditional methods is difficult as most of these traits are polygenic or complex and controlled by many genes. Information using genetic markers may allow to identify relevant QTL, but methods like MAS are only applicable for traits with a few QTL with major effects, while genomic selection allows the prediction of the total genetic value or phenotype and is thus more applicable for complex traits

(Kumar et al. 2012a). Only a few studies have evaluated the potential of genomic selection in fruit trees, such as apple (*Malus domestica*), grapes (*Vitis vinifera*), or pear (*Pyrus*) (Kumar et al. 2012a; Kumar et al. 2012b; Iwata et al. 2013; Myles 2013). Most of the molecular markers used in apple breeding have focused on resistance traits and applied markers for marker-assisted selection. But such single-gene markers did not provide a method for long-time disease resistance selective breeding because pathogens or pests did develop new strategies to overcome such resistances (Kumar et al. 2012a). Genomic selection is suggested as a possibly better alternative as it incorporates multiple markers and might allow a selection including genes with smaller effects. Two alternative strategies are suggested: the use of genomic selection for parent selection (as in forest trees for the elite parent lines) or for the selection of future cultivars (Kumar et al. 2012a). Preliminary results in an apple and pear tree population have indicated that genomic selection will allow selection prior to expensive phenotypic evaluation and might have the potential to speed up the selection process. However cross-validation within the same generation of trees has been used to derive the accuracies (Kumar et al. 2012a; Iwata et al. 2013). The application of genomic selection in crossbred individuals is relevant in fruit trees. Crossbred scenarios will require the prediction of nonadditive effects. One additional point of consideration is the use of grafts. Full-sib families are commonly used in apple breeding programs, and seedlings are grafted onto clonal rootstocks, a strategy which differs from the cloning used for phenotypic evaluation in forest trees (Kumar et al. 2012a).

If genomic selection in tree breeding can provide similar accuracies as conventional breeding, it will be able to increase genetic gain and reduce sizes and costs for breeding programs significantly. But strategies need to be developed to allow either long-term effects with low decay of accuracy over several generations or options for a cost-efficient regular updating of the prediction model. It has yet to be shown how genomic selection will perform in crossbred situations and across multiple generations, as many of those studies apply their simulation in a single generation only (Grattapaglia and Resende 2011; Kumar et al. 2012a; Zapata-Valenzuela et al. 2012; Iwata et al. 2013).

7.5 Examples of Genomic Prediction Applied on Human (Homo sapiens) Populations

Genomic prediction has been suggested as a useful tool in assessing genetic predisposition for human diseases and personalized medicine (de los Campos et al. 2010; Makowsky et al. 2011). However, genomic prediction has not been successfully applied to any great extent in humans yet. Nevertheless, the models for genomic selection have been successful in human studies to estimate the heritability of complex traits (Yang et al. 2010).

The accuracy of genomic predictions in a test population with estimates based on trait values measured in a reference population depends largely on the variance in relatedness between pairs of individuals in the test and reference populations or equivalently the mean linkage disequilibrium over all pairs of loci (Goddard et al. 2011). In humans, linkage disequilibrium is small, and useful genomic prediction would therefore require a very large reference set. Consequently, genomic prediction has not been found to be as useful as in animal and plant populations with larger linkage disequilibrium.

The statistical models developed for genomic selection have been found to be extremely valuable in human genetics for heritability estimation. Separating genetic and environmental effects in humans has been notoriously difficult in the past because human populations generally consist of small families where relatives share many environmental factors. Yang et al. (2010) showed that by combining all SNP information from practically unrelated individuals (i.e., pair-wise genomic correlations between individuals typically smaller than 0.1) in a GBLUP, it is possible to estimate the heritability of complex traits. By using unrelated individuals, any possible confounding of genetic and environmental effects is eliminated.

8 Future Directions and Perspective

Many of the genomic selection research and development efforts focused on improving the accuracy of genomic breeding values, exploring a large range of parametric and nonparametric models for genomic prediction. While the application of genomic selection requires robust machinery for genomic prediction, it is important to realize that the real benefits of genomic prediction can only be harvested when accompanied by changes in the breeding program. Optimizations of breeding strategies that utilize genomic breeding values are thus far underexplored, and much gain can be expected from studies on novel and innovative breeding schemes. Synergies between genomic selection and reproduction techniques and/or genome editing are examples of components of such breeding schemes. Another example of an element to consider in the design of breeding schemes is that strategies for genotyping selection candidates can affect the composition of the future reference population, giving rise to a complex optimization problem if the aim is long-term genetic improvement.

Genotype information can also be used for population management. This relates not only to conservation of populations at risk but also for the maintenance of genetic variability in commercial populations. Genomic selection was believed to have a positive impact on rates of inbreeding, but the first indications of experience from the field report increased rates of inbreeding in genomic breeding schemes. However, there remains much scope for development of genomic tools that consider both genetic progress and maintenance of genetic diversity.

There is much potential to utilize genomic information for prediction of phenotypes of animals, plants, and trees, in order to tailor management, similar to utilizing genomic prediction for personalized medicine discussed in the context of human

genetics. For example, mating schemes can be optimized using genomic information to avoid inbreeding or to capitalize on hybrid vigor and other nonadditive genetic effects. Moreover, knowledge about the predisposition to certain diseases can be used to direct preventive measures to individuals with elevated risk.

In summary, genomic information can in the first place be used to enhance genetic gains and offers also opportunities for improved management, at various levels.

9 Conclusions

Since the first suggestion of genomic selection and prediction in 2001, the development of genotyping methods has allowed the introduction of this advanced selection tool across many populations. Breeders are hoping for an easier and more accurate selection tool, which allows an earlier selection of advanced lines or individuals. Early estimations based on information from dairy populations revealed that the application of genomic selection should increase the rate of genetic gain and that genomic selection has the potential to revolutionize animal breeding (Schaeffer 2006; Hayes et al. 2009a; Thornton 2010; Goddard 2012). Similar improvements have also been predicted for plant breeding. It has been shown in studies using empirical and simulated data that the use of genetic markers will accelerate breeding and reduce the generation interval/time for the development of new varieties (Rudi et al. 2010). Genomic selection in combination with high-throughput phenotyping might revolutionize the selection for complex traits (Cabrera-Bosquet et al. 2012). In Holstein Frisian dairy cattle, the implementation of SNP information was predicted to provide as much information as real data from phenotypes from 10 to 20 daughters per bull (Jannink et al. 2010). Available SNP information would thereby allow to collect phenotypic records from fewer offspring with no loss of accuracy. However, statistical models for different breeding scenarios have to be developed (Heslot et al. 2012). Inclusion of nonadditive effects, such as heterosis or genotype by environment interactions, will be relevant for some traits and in some populations. Improved phenotyping has to be established as the accuracy and throughput of phenotype measurements are currently the main limiting factors (Lorenzana and Bernardo 2009).

There is little doubt that genomic selection is a success in the main dairy cattle breed, Holstein Friesian. Genomic selection is also practiced in other dairy cattle breeds, but not as successful in terms of accuracy of selection as in the Holstein breed, and it remains unclear if the successes can be repeated in other species. Further advancements in technology are needed in situations with complex population compositions and genome structure. Massive sequencing at low coverage (genomic selection 2.0) and better use of biological knowledge as priors in genomic prediction are promising directions of future developments. Good knowledge on the functionality of mutations is imperative, to be able to target the right QTN in selection and avoid unwanted side effects.

The statistical models used for genomic selection in livestock have been proven useful to estimate heritabilities in human genetic studies. Genomic prediction has also been suggested as a tool to predict genetic predisposition of human health disorders, even though not many success stories are documented to date. Similar to prediction phenotypic in humans, genomic prediction has potential to be useful for management purposes at agricultural farms to optimize production processes. Sequencing data is currently used in breeding populations, but reliability of the data and the information acquired from the data have to be questioned: how complex can data be in order to be implemented in prediction models and how much background do we need on the inheritance of genome structures different from polymorphisms. There is little doubt that the inclusion of more information on genotypes will improve predictions. Whether the inclusion of information from molecular genetic markers will be advantageous to other phenotypic and environmental measures is probably a question of costs, rather than results.

The current advances of the methods, some of which introduced here, need further discussion. Methods and models will need to be tested from case to case, and different models might be needed for different traits. Much of the benefits from genomic selection arise from the possibility to determine the outfall of Mendelian sampling as soon as a DNA sample can be taken. The phenotypes can therefore be predicted with higher accuracy as exact genotypes are already known. It thus seems pertinent to determine the accuracy of Mendelian sampling deviations calculated from genomic breeding values, and to consider that statistics in the comparison of models and methods, apart from some exceptions (e.g., Rius-Vilarrasa et al. 2012), this is rarely done.

The validation of prediction models needs careful consideration. Accuracies based on cross-validation might not reflect accuracies of selection achievable in breeding schemes applied in practice. Many of the current selection schemes in plant breeding are based on phenotypes recorded during the first steps of selection, which may be different from those for the final breeding goal. The correlations to final breeding goal might therefore be low. Application of genetic markers will allow a better prediction of early selection. However, accuracies should be calculated based on models applicable to real breeding populations.

Despite the current pitfalls, the concept of genomic selection has led to a number of advances driven by the need for improved selection in plant and livestock populations. It has contributed to the fast application of genotyping and sequencing tools in nonhuman populations. It has also opened new opportunities and advanced options for methods for prediction models. Phenotyping has been put in the spotlight again, as reliable phenotypes are required for accurate predictions. The options of a better use of phenotypes have led to an extension of measurements and inclusion of complex traits, especially such related to health/welfare and sustainability, into selection schemes. While such progress is not solely based on the development of genomic selection, the new opportunity for the use of genome-wide marker sets for the prediction in populations has assisted such new opportunities.

References

http://www.interbull.org. Interbull. 2013. Accessed 31.01.2013.

Albrecht T, et al. Genome-based prediction of testcross values in maize. Theor Appl Genet. 2011;123:339–50.

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink JL. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome. 2011;4:132–44.

Avendaño S, Watson KA, Kranis A. Genomics in poultry breeding—from utopias to deliverables. In: 9th world congress on genetics applied to livestock production (WCGALP). Germany: Leipzig; 2010

Baloche G, et al. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J Dairy Sci. 2014;97:1107–16.

Barrell PJ, Meiyalaghan S, Jacobs JME, Conner AJ. Applications of biotechnology and genomics in potato improvement. Plant Biotechnol J. 2013;11:907–20.

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci. 2016;242:23–36.

Battenfield SD, Guzmán C, Gaynor RC, Singh RP, Peña RJ, Dreisigacker S, Fritz AK, Poland JA. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. Plant Genome. 2016;9(2). https://doi.org/10.3835/plantgenome2016.01.0005.

Bauer E, et al. Towards a whole-genome sequence for rye (Secale cereale L.). Plant J. 2017;89:853–69.

Beaulieu J, Doerksen T, Clement S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity. 2014a;113:343–52.

Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics. 2014b;15:1048.

Bernardo R. Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci. 2010;50:624–7.

Bernardo R, Yu J. Marker-assisted selection without QTL mapping: prospects for genome-wide selection for quantitative traits in maize. Maize Genet Cooperat Newslett 2007:26.

Berry DP, Garcia JF, Garrick DJ. Development and implementation of genomic predictions in beef cattle. Anim Front. 2016;6:32–8.

Bertin N, Martre P, Génard M, Quilot B, Salon C. Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. J Exp Bot. 2010;61:955–67.

Birol I, et al. Assembling the 20 Gb white spruce (Picea glauca) genome from whole-genome shotgun sequencing data. Bioinformatics. 2013;29:1492–7.

Boichard D, Ducrocq V, Croiseau P, Fritz S. Genomic selection in domestic animals: principles, applications and perspectives. C R Biol. 2016;339:274–7.

Bouquet A, Juga J. Integrating genomic selection into dairy cattle breeding programmes: a review. Animal. 2013;7:705–13.

Brenchley R, et al. Analysis of the bread wheat genome using whole genome shotgun sequencing. Nature. 2012;491:705–10.

Breseghello F. Traditional and modern plant breeding methods with examples in rice (Oryza sativa L.). J Agric Food Chem. 2013;61:8277–86.

Burgueno J, de los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. Crop Sci. 2012;52:707–19.

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL. High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol. 2012;54(5):312–20.

Calus MPL. Editorial: genomic selection with numerically small reference populations. Animal. 2016;10:1016–7.

de los Campos G, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics. 2009;182:375–85.
de los Campos G, Gianola D, Allison DB. Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat Rev Genet. 2010;11:880–6.
de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 2013a;193:327–45.
de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genet. 2013b;9:e1003608.
Caspi R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2014;42:D459–71.
Cericola F, et al. Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. PLoS One. 2017;12:e0169606.
Chagné D, et al. The draft genome sequence of European Pear (Pyrus communis L. "Bartlett"). PLoS One. 2014;9:e92644.
Chao S, Zhang W, Dubcovsky J, Sorrells M. Evaluation of genetic diversity and genome-wide linkage disequilibrium among US wheat (Triticum aestivum L.) germplasm representing different market classes. Crop Sci. 2007;47:1018–30.
Chao SM, et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). BMC Genomics. 2010;11:727.
Chawade A, Alexandersson E, Bengtsson T, Andreasson E, Levander F. Targeted proteomics approach for precision plant breeding. J Proteome Res. 2016;15:638–46.
Chen HD, He H, Zhou FS, Yu HH, Deng XW. Development of genomics-based genotyping platforms and their applications in rice breeding. Curr Opin Plant Biol. 2013;16:247–54.
Cowling WA, Balazs E. Prospects and challenges for genome-wide association and genomic selection in oilseed Brassica species. Genome. 2010;53:1024–8.
Cowling WA, Buirchell BJ, Falk DE. A model for incorporating novel alleles from the primary gene pool into elite crop breeding programs while reselecting major genes for domestication or adaptation. Crop Pasture Sci. 2009;60:1009–15.
Croft D, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011;39:D691–7.
Crossa J, et al. Genomic selection and prediction in plant breeding. J Crop Improv. 2011;25:239–61.
Crossa J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. 2013;112:48–60.
Cullis BR, Smith AB, Beeck CP, Cowling WA. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. Genome. 2010;53:1002–16.
Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One. 2008;3:e3395.
Daetwyler HD, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46:858–65.
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12(7):499–510. https://doi.org/10.1038/nrg3012.
Dawson JC, et al. The use of unbalanced historical data for genomic selection in an international wheat breeding program. Field Crop Res. 2013;154:12–22.
Dekkers JCM. Marker-assisted selection for commercial crossbred performance. J Anim Sci. 2007;85:2104–14.
Dekkers JCM, Hospital F. The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet. 2002;3:22–32.
Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity. 2014;112:39–47.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 2010. https://doi.org/10.1093/nar/gkx382.

Dürr J, Philipsson J. International cooperation: the pathway for cattle genomics. Anim Front. 2012;2:16–21.

Duvick DN. Heterosis: feeding. People and protecting natural resources. In: Coors JG, Pandey S, editors. The genetics and exploitation of heterosis in crops. Madison, WI: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.; 1999. p. 19–29.

Elsik CG, Tellam RL, Worley KC. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009;324:522–8.

Erbe M, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2012;95:4114–29.

Falconer D, Mackay T. Quantitative genetics. London, UK: Longman, Harrow; 1996.

Fernie AR, Schauer N. Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet. 2009;25:39–48.

Forabosco F, Lohmus M, Rydhmer L, Sundstrom LF. Genetically modified farm animals and fish in agriculture: a review. Livest Sci. 2013;153:1–9.

Funk DA. Major advances in globalization and consolidation of the artificial insemination industry. J Dairy Sci. 2006;89:1362–8.

Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 2013;31(7):397–405.

Ganal MW, Röder MS. Microsatellite and SNP markers in wheat breeding. In: Varshney RK, Tuberosa R, editors. Genomic assisted crop improvement: genomics applications in crops, vol. 2. Dordrecht: Springer; 2007. p. 1–24.

García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci U S A. 2016;113(28):E3995–4004. https://doi.org/10.1073/pnas.1519061113.

Garrick DJ. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet Sel Evol. 2011;43:–17.

Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol. 2009;41:55.

Gerrits RJ, et al. Perspectives for artificial insemination and genomics to improve global swine populations. Theriogenology. 2005;63:283–99.

Gianola D. Priors in whole-genome regression: the Bayesian alphabet returns. Genetics. 2013;194:573–96.

Gianola D, van Kaam JBCHM. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics. 2008;178:2289–303.

Goddard ME. Uses of genomics in livestock agriculture. Animal Production Science. 2012;52:73–7.

Goddard ME, Hayes BJ. Genomic selection. J Anim Breed Genet. 2007;124:323–30.

Goddard ME, Hayes BJ, Meuwissen THE. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. 2011;128:409–21.

Goff SA, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. 2002;296:92–100.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17:333–51.

Gouy M, et al. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor Appl Genet. 2013;126:2575–86.

Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7:241–55.

van Grevenhof I. Breeding against osteochondrosis. Wageningen: Wageningen University; 2011.

Groenen MAM, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491:393–8.

Gupta PK, Langridge P, Mir RR. Marker-assisted wheat breeding: present status and future possibilities. Mol Breed. 2010;26:145–61.
Guzman C, et al. Wheat quality improvement at CIMMYT and the use of genomic selection on it. Appl Transl Genom. 2016;11:3–8.
Haberland AM, König von Borstel U, Simianer H, König S. Integration of genomic information into sport horse breeding programs for optimization of accuracy of selection. Animal. 2012;6:1369–76.
Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol. 2010;42:–5.
Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.
Habier D, Fernando RL, Garrick DJ. Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics. 2013;194:597–607.
Hayes B, Bowman P, Chamberlain A, Goddard M. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009a;92:433–43.
Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol. 2009b;41:–51.
Hayes BJ, et al. Prospects for genomic selection in forage plant species. Plant Breeding. 2013;132:133–43.
Heffner EL, Sorrells ME, Jannink J-L. Genomic selection for crop improvement. Crop Sci. 2009;49:1–12.
Heffner EL, Jannink JL, Sorrells ME. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome. 2011;4:65–75.
Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975a;31:423–47.
Henderson CR. Use of all relstives in intraherd prediction of breeding values and producing abilities. J Dairy Sci. 1975b;58:1910–6.
Heslot N, Yang H-P, Sorrells ME, Jannink J-L. Genomic selection in plant breeding: a comparison of models. Crop Sci. 2012;52:146–60.
Hickey JM. Sequencing millions of animals for genomic selection 2.0. J Anim Breed Genet. 2013;130:331–2.
Hickey JM, et al. Sequencing millions of animals for genomic selection 2.0. In: Proceedings, 10th world congress of genetics applied to livestock production. Vancouver; 2014.
de Roos APW, Hayes BJ, Goddard ME. Reliability of genomic predictions across multiple populations. Genetics. 2009;183:1545–53.
Hofheinz N, Borchardt D, Weissleder K, Frisch M. Genome-based prediction of test cross performance in two subsequent breeding cycles. Theor Appl Genet. 2012;125:1639–45.
Hutchison JL, Cole JB, Bickhart DM. Short communication: use of young bulls in the United States. J Dairy Sci. 2014;97:3213–120. https://doi.org/10.3168/jds.2013-7525.
International Barley Genome Sequencing Consortium et al. A physical, genetic and functional sequence assembly of the barley genome. Nature. 2012;491:711–6.
Isik F, et al. Genomic selection in maritime pine. Plant Sci. 2016;242:108–19.
Iwata H, et al. Potential assessment of genome-wide association study and genomic selection in Japanese pear Pyrus pyrifolia. Breed Sci. 2013;63:125–40.
Iwata H, Hayashi T, Tsumura Y. Prospects for genomic selection in conifer breeding: a simulation study of Cryptomeria japonica. Tree Genet Genomes. 2011;7:747–58.
Iwata H, Jannink J-L. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 2011;51:1915–27.
Jannink J-L, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics. 2010;9:166–77.
Jannink JL. Dynamics of long-term genomic selection. Genet Sel Evol. 2010;42:–35.
Jenko J, et al. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. Genet Sel Evol. 2015;47:55.

Jiang Y, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344:1168–73.
Johnston DJ, Tier B, Graser HU. Beef cattle breeding in Australia with genomics: opportunities and needs. Animal Production Science. 2012;52:100–6.
Kanehisa M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36:D480–4.
Kärkkäinen HP, Sillanpää MJ. Fast genomic predictions via Bayesian G-BLUP and Multilocus models of threshold traits including censored Gaussian data. G3 (Bethesda). 2013;3:1511–23.
Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci. 2010;88:544–51.
Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagne D. Towards genomic selection in apple (Malus x domestica Borkh.) breeding programmes: prospects, challenges and strategies. Tree Genet Genomes. 2012a;8:1–14.
Kumar S, et al. Genomic selection for fruit quality traits in apple (Malus x domestica Borkh.). PLoS One. 2012b;7(5):e36674.
Lenz PRN, et al. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). BMC Genomics. 2017;18:335.
Li Z-K, Zhang F. Rice breeding in the post-genomics era: from concept to practice. Curr Opin Plant Biol. 2013;16:261–9.
Lien S, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016;533:200–5.
Lillehammer M, Meuwissen THE, Sonesson AK. Genomic selection for maternal traits in pigs. J Anim Sci. 2011;89:3908–16.
Lillehammer M, Meuwissen THE, Sonesson AK. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. Genet Sel Evol. 2013;45
Lin Z, Hayes BJ, Daetwyler HD. Genomic selection in crops, trees and forages: a review. Crop Pasture Sci. 2014;65:1177–91.
Lindblad-Toh K, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438:803–19.
Lôbo RB, et al. Implementation of DNA markers to produce genomically—enhanced EPDs in Nellore cattle. Acta Sci Vet. 2011;39(Suppl 1):s23–7.
Longin CFH, Mi X, Würschum T. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor Appl Genet. 2015;128:1297–306.
Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet. 2009;120:151–61.
Lund MS, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet Sel Evol. 2011;43:43.
Lush JL. Linebreeding. Iowa Agric Exp Sta Bull 1933:301.
Lush JL. Family merit and individual merit as bases for selection. Am Nat. 1947;81:241–61.
Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer; 1998.
Maccaferri M, Sanguineti MC, Noli E, Tuberosa R. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed. 2005;15:271–89.
MacLeod IM, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics. 2016;17:144.
Makowsky R, et al. Beyond missing heritability: prediction of complex traits. PLoS Genet. 2011;7: e1002051.
Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11:499–511.
Marulanda JJ, et al. Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. Theor Appl Genet. 2016;129:1901–13.

Mascher M, et al. A chromosome conformation capture ordered sequence of the barley genome. Nature. 2017;544:427–33.
Mather KA, et al. The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics. 2007;177:2223–32.
Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One. 2009;4:e5350. https://doi.org/10.1371/journal.pone.0005350.
McCouch S. Diversifying selection in plant breeding. PLoS Biol. 2004;2:1507–12.
Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.
Meuwissen THE, Luan T, Woolliams JA. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J Anim Breed Genet. 2011;128:429–39.
Morandini P, Salamini F. Plant biotechnology and breeding: allied for years to come. Trends Plant Sci. 2003;8:70–5.
Morota G, Gianola D. Kernel-based whole-genome prediction of complex traits: a review. Front Genet. 2014;5:363.
Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. Nat Rev Genet. 2012;13:85–96.
Muir B, Van Doormaal BJ, Kistemaker G. International genomic co-operation—North American perspective. In: Proceedings of the Interbull international workshop, Paris, France; 2010. pp 71–76.
Myburg AA, et al. The genome of Eucalyptus grandis. Nature. 2014;510:356–62.
Myles S. Improving fruit and wine: what does genomics have to offer? Trends Genet. 2013;29:190–6.
Neale DB, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. 2014;15:R59.
Nielsen HM, Sonesson AK, Meuwissen THE. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. J Anim Sci. 2011;89:630–8.
Northcutt SL. Genomic choices. American Angus Association®/AngusGenetics Inc. release. 2011. http://www.angus.org/AGI/GenomicChoice070811.pdf (posted July, 2011)
Nystedt B, et al. The Norway spruce genome sequence and conifer genome evolution. Nature. 2013;497:579–84.
Ober U, et al. Predicting genetic values: a kernel-based best linear unbiased prediction with genomic data. Genetics. 2011;188:695–708.
Patry C. Impacts of genomic selection on classical genetic evaluations. Jouy-en-Josas: Institut National de la Recherche Agronomique (INRA); 2011.
Potato Genome Sequencing Consortium, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95.
Preisinger R. Genome-wide selection in poultry. Animal Production Science. 2012;52:121–5.
Proudfoot C, et al. Genome edited sheep and cattle. Transgenic Res. 2015;24:147–53.
Pryce JE, Daetwyler HD. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim Prod Sci. 2012;52:107–14.
Pszczola M, Calus MPL. Updating the reference population to achieve constant genomic prediction reliability across generations. Animal. 2015;10:1018–24.
Pszczola M, Strabel T, van Arendonk JAM, Calus MPL. The impact of genotyping different groups of animals on accuracy when moving from traditional to genomic selection. J Dairy Sci. 2012;95:5412–21.
Ratcliffe B, et al. A comparison of genomic selection models across time in interior spruce (Picea engelmannii × glauca) using unordered SNP imputation methods. Heredity (Edinb). 2015;115 (6):547–55.
Remington DL, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A. 2001;98:11479–84.
Resende M, et al. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 2012a;193

Resende MDV, et al. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. 2012b;194:116–28.
Resende RMS, Casler MD, Resende MDV. Genomic selection in forage breeding: accuracy and methods. Crop Sci. 2014;54:143–56.
Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, Grattapaglia D. Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity (Original Article). 2017. https://doi.org/10.1038/hdy.2017.37.
Rius-Vilarrasa E, et al. Influence of model specifications on the reliabilities of genomic prediction in a Swedish–Finnish red breed cattle population. J Anim Breed Genet. 2012;129:369–79.
Rosenberg NA, Nordborg M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet. 2002;3:380–90.
Rubin C-J, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464:587–91.
Rudi N, Norton GW, Alwang J, Asumugha G. Economic impact analysis of maker-assisted breeding for resistance to pests and post harvest deterioration of cassava. Afr J Agr Res Econ. 2010;4:110–22.
Rutkoski JE, Heffner EL, Sorrells ME. Genomic selection for durable stem rust resistance in wheat. Euphytica. 2011;179:161–73.
Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. Genet Sel Evol. 2012;44:38.
Sánchez-Molano E, et al. Genomic prediction of traits related to canine hip dysplasia. Front Genet. 2015;6:97.
Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet. 2006;123:218–23.
Schnable PS, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5.
Sharma HC, Crouch JH, Sharma KK, Seetharama N, Hash CT. Applications of biotechnology for crop improvement: prospects and constraints. Plant Sci. 2002;163:381–95.
Shen X, Alam M, Fikse F, Rönnegård L. A novel generalized ridge regression method for quantitative genetics. Genetics. 2013;193(4):1255–68.
Shu YJ, et al. Genomic selection of seed weight based on low-density SCAR markers in soybean. Genet Mol Res. 2013;12:2178–88.
Shumbusho F, Raoul J, Astruc JM, Palhiere I, Elsen JM. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs1. J Anim Sci. 2013;91:3644–57.
Snelling WM, et al. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle12. J Anim Sci. 2011;89:1731–41.
Somers DJ, Kirkpatrick R, Moniwa M, Walsh A. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome. 2003;46:431–7.
Sonesson AK, Meuwissen THE. Testing strategies for genomic selection in aquaculture breeding programs. Genet Sel Evol. 2009;41:37.
Spindel J, et al. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet. 2013;126:2699–716.
Stein L. Genome annotation: from sequence to biology. Nat Rev Genet. 2001;2:493–503.
Stock KF, Jönsson L, Ricard A, Mark T. Genomic applications in horse breeding. Anim Front. 2016;6:45–52.
Sun XC, Fernando R, Dekkers J. Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. Genet Sel Evol. 2016;48(1):77.
Swan AA, Johnston DJ, Brown DJ, Tier B, Graser H-U. Integration of genomic information into beef cattle and sheep genetic evaluations in Australia. Animal Production Science. 2012;52:126–32.
Sweeney M, McCouch S. The complex history of the domestication of rice. Ann Bot. 2007;100:951–7.

Takeda S, Matsuoka M. Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet. 2008;9:444–57.
Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK. Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. BMC Plant Biol. 2017;17(1):110. https://doi.org/10.1186/s12870-017-1059-6.
Tan W, Proudfoot C, Lillico SG, Whitelaw CBA. Gene targeting, genome editing: from Dolly to editors. Transgenic Res. 2016;25:273–87.
Tang C, et al. The rubber tree genome reveals new insights into rubber production and species adaptation. Nat Plants. 2016;2:16073.
Tenaillon MI, Austerlitz F, Tenaillon O. Apparent mutational hotspots and long distance linkage disequilibrium resulting from a bottleneck. J Evol Biol. 2008;21:541–50.
Thavamanikumar S, Southerton SG, Bossinger G, Thumma BR. Dissection of complex traits in forest trees—opportunities for marker-assisted selection. Tree Genet Genomes. 2013;9:627–39.
Thornton PK. Livestock production: recent trends, future prospects. Phil Trans Roy Soc B-Biol Sci. 2010;365:2853–67.
Thorwarth P, et al. Genomic prediction ability for yield-related traits in German winter barley elite material. Theor Appl Genet. 2017;130(8):1669–83.
Toro MA, Varona L. A note on mate allocation for dominance handling in genomic selection. Genet Sel Evol. 2010;42:33.
Trebbi D, et al. High-throughput SNP discovery and genotyping in durum wheat (Triticum durum Desf.). Theor Appl Genet. 2011;123:555–69.
Tribout T, Larzul C, Phocas F. Efficiency of genomic selection in a purebred pig male line. J Anim Sci. 2012;90:4164–76.
Tuskan GA, et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science. 2006;313:1596–604.
van der Werf JHJ. Marker-assisted selection in sheep and goats. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD, editors. Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. Rome: Food and Agriculture Organization of the United Nations; 2007.
VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
VanRaden PM, Wiggans GR. Derivation, calculation, and use of national animal model information. J Dairy Sci. 1991;74:2737–46.
VanRaden PM, Wiggans GR, Van Tassell CP, Sonstegard TS, Schenkel F. Benefits from cooperation in genomics. In: Proceedings of the Interbull international workshop. Genomic information in genetic evaluations. Uppsala, Sweden; 2009a pp 67–72.
VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci. 2009b;92:16–24. https://doi.org/10.3168/jds.2008-1514.
Velasco R, et al. The genome of the domesticated apple (Malus [times] domestica Borkh.). Nat Genet. 2010;42:833–9.
Wade CM, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science. 2009;326:865–7.
Walsh B. Quantitative genetics. In: eLS. John Wiley & Sons Ltd., Chichester. 2001
Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United States: past, present, future. J Dairy Sci. 2011;94:3202–11.
Wilkins PW, Humphreys MO. Progress in breeding perennial forage grasses for temperate agriculture. J Agric Sci. 2003;140:129–50.
Williams AV, Nevill PG, Krauss SL. Next generation restoration genetics: applications and opportunities. Trends Plant Sci. 2014;19:529–37.
Windhausen VS, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda). 2012;2:1427–36.
Wu J, et al. The genome of the pear (Pyrus bretschneideri Rehd.). Genome Res. 2013;23:396–408.

Wurschum T, Abel S, Zhao Y. Potential of genomic selection in rapeseed (Brassica napus L.) breeding. Plant Breeding. 2014;133:45–51.

Wurschum T, Reif J, Kraft T, Janssen G, Zhao Y. Genomic selection in sugar beet breeding populations. BMC Genet. 2013;14:85.

Xu S, Hu Z. Methods of plant breeding in the genome era. Genet Res. 2010;92:423–41.

Yan ZB, Yan WG, Deren CW, McClung A. Hybrid rice breeding. B.R. Wells Rice Research Series—Arkansas Agricultural Experiment Station University of Arkansas, vol 591. 2011. pp 61–63.

Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.

Young CW, Bonczek RR, Johnson DG. Inbreeding of and relationship among registered Holsteins. J Dairy Sci. 1988;71:1659–66.

Yu J, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science. 2002;296:79–92.

Zamir D. Improving plant breeding with exotic genetic libraries. Nat Rev Genet. 2001;2:983–9.

Zapata-Valenzuela J, et al. SNP markers trace familial linkages in a cloned population of Pinus taeda-prospects for genomic selection. Tree Genet Genomes. 2012;8:1307–18.

Zelener N, Poltri SNM, Bartoloni N, Lopez CR, Hopp HE. Selection strategy for a seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for Eucalyptus dunnii. Tree Physiol. 2005;25:1457–67.

Zhao F, Xu S. An expectation and maximization algorithm for estimating G x E interaction effects. Theor Appl Genet. 2012;124:1375–87.

Zhao Y, et al. Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet. 2012;124:769–76.

Zhong SQ, Dekkers JCM, Fernando RL, Jannink JL. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics. 2009;182:355–64.

Zimin A, et al. Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics. 2014;196:875–90.