

Population Genomics Along With Quantitative Genetics Provides a More Efficient Valorization of Crop Plant Genetic Diversity in Breeding and Pre-breeding Programs



Peter Civan, Renaud Rincent, Alice Danguy-Des-Deserts,
Jean-Michel Elsen, and Sophie Bouchet

Abstract The breeding efforts of the twentieth century contributed to large increases in yield but selection may have increased vulnerability to environmental perturbations. In that context, there is a growing demand for methodology to re-introduce useful variation into cultivated germplasm. Such efforts can focus on the introduction of specific traits monitored through diagnostic molecular markers identified by QTL/association mapping or selection signature screening. A combined approach is to increase the global diversity of a crop without targeting any particular trait.

A considerable portion of the genetic diversity is conserved in genebanks. However, benefits of genetic resources (GRs) in terms of favorable alleles have to be weighed against unfavorable traits being introduced along. In order to facilitate utilization of GR, core collections are being identified and progressively characterized at the phenotypic and genomic levels. High-throughput genotyping and sequencing technologies allow to build prediction models that can estimate the

P. Civan

GDEC, INRAE, Université Clermont Auvergne, Clermont-Ferrand, France

R. Rincent

GQE-Le Moulon, INRAE, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Clermont-Ferrand, France

A. Danguy-Des-Deserts

GDEC, Université Clermont Auvergne, Clermont-Ferrand, France

GenPhySE (Génétique Physiologie et Systèmes d'Élevage), Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France

J.-M. Elsen

GenPhySE (Génétique Physiologie et Systèmes d'Élevage), Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France

S. Bouchet (✉)

INRAE, Université Clermont Auvergne, Clermont-Ferrand, France

e-mail: sophie.bouchet@inrae.fr

Om P. Rajora (ed.), *Population Genomics: Crop Plants*,

Population Genomics [Om P. Rajora (Editor-in-Chief)],

https://doi.org/10.1007/13836_2021_97, © The Author(s) 2021

genetic value of an entire genotyped collection. In a pre-breeding program, predictions can accelerate recurrent selection using rapid cycles in greenhouses by skipping some phenotyping steps. In a breeding program, reduced phenotyping characterization allows to increase the number of tested parents and crosses (and global genetic variance) for a fixed budget. Finally, the whole cross design can be optimized using progeny variance predictions to maximize short-term genetic gain or long-term genetic gain by constraining a minimum level of diversity in the germplasm. There is also a potential to further increase the accuracy of genomic predictions by taking into account genotype by environment interactions, integrating additional layers of omics and environmental information.

Here, we aim to review some relevant concepts in population genomics together with recent advances in quantitative genetics in order to discuss how the combination of both disciplines can facilitate the use of genetic diversity in plant (pre) breeding programs.

Keywords Genomic predictions · Genotype by environment interaction · Long-term genetic gain · Mating design · Multi-trait · Plant genetic resources · Population genomics · (Pre-)breeding · QTLs · Selection signatures

1 Introduction

The challenge in agriculture today is to produce enough food for an increasing population, using less land, water, fertilizer, and pesticides to limit ecological impacts. Global environmental changes, rainfall variability, nitrogen cycle alteration, higher temperature, and atmospheric CO₂ concentration strongly impact crop plants phenology (Jagadish et al. 2016), resistance to pathogens/insects outbreaks (Deutsch et al. 2018), and yield (Brisson et al. 2010). As a consequence, genetic gain for stress tolerance has become one of the most important targets in plant breeding. In this context, genetic variants present in modern varieties, traditional local varieties (i.e., landraces), and wild relatives may be of interest for crop plant breeders (McCouch et al. 2013).

For about 10,000 years humans have been exerting selection pressure, both consciously (by selecting the “best” seeds or animals to contribute to the next generation) and involuntarily (through farming practices and expansion of the natural distribution range), gradually changing domesticated plants and animals to suit their needs. This piecemeal process of selection gradually morphed into breeding, and first commercially successful plant breeding emerged at the end of the nineteenth century. The “best” (that was selected) has been covering many different criteria (Allard 1999) in different species, times, countries, environments, and now depends on the end-users/markets targeted. Multi-trait indexes have been empirically or economically built from phenotypic observations and expert opinion. This is generally called phenotypic selection (PS). Rapid genetic gain has been secured by breeders by selecting only the highest-performing parental individuals in order to

ensure high mean performance of progeny. The classical strategy in crop plants is to cross performant lines for different traits, aiming to obtain some recombinant lines in the progeny that cumulate a maximum of chosen criteria. However, continuous application of truncation selection (selection of the bests) without regular re-introduction of new alleles in the germplasm leads to a rapid loss of genetic diversity around loci under selection by hitch-hiking effects and all along the genome by drift. This can have negative consequences for loci not monitored in the process and reduce the long-term potential of the program. Additionally, truncation selection overlooks favorable alleles that only occur in lines that are not highly performing. The most famous examples of the negative impact of reduced diversity in cultivated plants concern disease resistance. All cultivated potato (*Solanum tuberosum*) varieties cultivated in Ireland were susceptible to late blight, leading to the Great famine in 1845–1849 (Mizubuti and Fry 2006). Similarly, maize (*Zea mays*) varieties that all contained the common male sterile genetic background were susceptible to southern leaf blight that caused 15% losses in 1970–1971 in the USA (Ullstrup 1972).

Through times, cultivated plants have experienced various genetic bottlenecks through selection and drift that accompanied domestication, migrations, and subsequent local adaptation (Spillane and Gepts 2001). These bottlenecks explain the reduction of genetic diversity compared to wild relatives or local traditional varieties referred to as landraces. Only a few studies have focused on long-term changes in genetic diversity in breeding programs, for instance in maize, *Zea mays* (Labate et al. 1999; Feng et al. 2006; van Inghelandt et al. 2010; Gerke et al. 2015; Allier et al. 2019d) and soybean, *Glycine max* (Bruce et al. 2019). There is evidence that modern breeding further reduced genetic diversity (Simmonds 1962; Cooper et al. 2001; Fu 2006, 2015) and changed its geographical distribution because of large open breeding systems. Such impacts of modern breeding are dramatic in the case of bread wheat, *Triticum aestivum*. Although landrace diversity is composed of two major genetic groups, Europe and Asia, Asian alleles are almost absent in worldwide modern lines. Note however that some extrinsic (from related species) DNA segments were introgressed into elite lines by breeders creating neo-diversity in bread wheat, *Triticum aestivum* (Balfourier et al. 2019), maize, *Zea mays* (Hufford et al. 2012), barley, *Hordeum vulgare* (Brown and Clegg 1983), soybean, *Glycine max* (Doyle 1988; Hyten et al. 2006; Han et al. 2016; Sedivy et al. 2017) or peanut, *Arachis hypogaea* (Fonceka et al. 2012), to list a few examples. Fu (2006) showed that genome-wide reduction of crop genetic diversity was minor but allelic reduction at some major QTLs was important. Directional selection actually tends to fix favorable alleles at some QTLs and neighboring regions by linkage drag (Maynard-Smith and Haigh 1974). So, there is an urgent need for efficient methodologies to monitor local and global diversity in breeding programs in order to maintain short-term and long-term genetic gain. It has been actually shown that a large genetic base of elite germplasm not only at known QTLs of agronomic interest but also all along the genome would assure long-term genetic gain and increase resilience of crop plants to biotic and abiotic stresses in unpredictable environmental conditions (Malézieux et al. 2009).

The way breeders rank selection candidates have changed through times. Genome-wide molecular markers and derived tools can now guide their decisions to complement the phenotyping information. With exponential capacities of genotyping and sequencing, improvements in computing and data storage, methodological and statistical developments, genomic selection (GS, Meuwissen et al. 2001) is becoming an essential tool to not only improve accuracy of selection, accelerate genetic gain using rapid cycles, optimize resource allocation, but also to better manage/introduce genetic diversity in breeding programs by optimizing parental contribution and cross design. Recurrent selection schemes (Hallauer and Sears 1972) for population improvement can also be re-visited with the help of GS to cumulate a maximum of favorable alleles in pre-breeding lines that can be integrated in breeding programs. Moreover, any useful information about loci controlling the variation of traits of agronomic interest (allele effects, genomic annotation) or subjected to historical evolutionary constraints (selection by environment or human), about genitors (genetic group, passport, and environmental data), can be used individually or as a covariate in prediction models to optimize selection process or cross designs. Therefore, population genomics in combination with quantitative genetics can provide relevant tools to evaluate, manage, and introduce GR in (pre)-breeding programs.

In this chapter, we first discuss how population genomics helps to assess genetic diversity, identify genes under selection, select candidate genes, and manage genetic diversity in crop plants. Then, we review the methodologies developed in genomic prediction and quantitative genetics to manage long-term genetic gain and genetic diversity in breeding programs. Finally, we present some future perspectives to optimize diversity valorization.

2 Population Genomics of Crop Plants Genetic Resources

2.1 Genetic Diversity in Crop Fields and Genebanks

Only a few crop species are widely cultivated around the world. Four crop species (wheat, maize, rice, and soybean) cover half of all land harvested worldwide – (<http://www.fao.org/faostat/en/#data/QC>). Moreover, most of the widely-grown species are represented by very few varieties in the fields. Such limited genetic variation in elite germplasm increases vulnerability to market and environmental changes. Mitigation of this situation through introduction of genetic innovations relies on GR that are maintained in around 1700 genebanks worldwide. However, only a small proportion of available GR has been explored and used so far, and it is believed that their comprehensive genomic characterization could enhance their utilization in breeding.

It is estimated that 24% of allelic diversity has been lost in maize compared to teosinte (Vigouroux et al. 2005), 70% in wheat compared to wild emmer (Haudry et al. 2007), and 30% in yam (*Dioscorea alata*) compared to its wild relatives

(Akakpo et al. 2017). This reduction of genetic diversity commonly observed in most crops is attributed to domestication and selection. Domestication corresponds to subsampling of wild progenitor species and results in what is called “the domestication bottleneck” (Goodman 1999, 2005; Meyer and Purugganan 2013; Allaby et al. 2019). Some observations suggest that the domestication bottleneck was not a rapid process associated with the dawn of agriculture, but rather a gradual genetic diversity loss that occurred during millennia (Allaby et al. 2019). The recent selection associated with modern plant breeding had a comparatively smaller impact on genetic diversity (van Heerwaarden et al. 2012), which has been well documented in maize and wheat (Reif et al. 2005; Glémin and Bataillon 2009; Meyer and Purugganan 2013). However, even the fraction of genetic diversity that has been retained in modern crops may not be effectively utilized today (Tenaillon and Charcosset 2011; Balfourier et al. 2019).

While breeding schemes rarely include diverse Genetic Resources (GR), the importance of collecting and characterizing genetic resources is widely recognized (McCouch et al. 2012). Genebank collections are an invaluable reservoir of favorable alleles that are not present in the cultivated gene pool. Examples of traits that have been successfully introgressed from GR into elite cultivars and had significant impact on crop production are numerous. In wheat for instance, dwarfing genes (reduced height loci *Rht-B1* and *Rht-D1*) and genes conferring durable resistance against a wide spectrum of insects and diseases were introgressed by Norman Borlaug during the Green Revolution. The Sorghum Conversion Program in the USA introgressed dwarf and photoperiod-insensitive alleles into African sorghum landraces to adapt them to temperate environments (Klein et al. 2008). The Germplasm Enhancement of Maize project (GEM) (Goodman et al. 2000) enabled massive introgression of GR alleles into the elite germplasm. Introgression lines have been massively produced for peanut as well, using wild relatives (Foncéka et al. 2009). Apart from the genes that control phenology (dwarfing genes, photoperiod insensitivity), great achievements include major genes of disease resistance, such as a resistance gene against grassy stunt virus introgressed from wild rice *Oryza nivara* (Plucknett 1987), leaf rust resistance genes in bread wheat introgressed from *Aegilops* (Kuraparthi et al. 2007) or other relatives (Steffenson et al. 2007; Ellis et al. 2014), and other genes providing resistance to biotic and abiotic stresses (Huang et al. 2016). There are also a few examples proving that wild gene pools contain genetic variants that can improve quality and yield, e.g. in tomato, *Lycopersicon esculentum* (Gur et al. 2004), wheat, *Triticum aestivum* (Uauy et al. 2006), maize, *Zea mays* (Ribaut and Ragot 2007) and rice, *Oryza sativa* (Imai et al. 2013).

Since comprehensive phenotyping and genomic characterization of all the GR is beyond the capacities of genebanks or other interested parties, “core collections” are often identified with the objective to represent most of the genetic diversity according to available information (passport and/or genotypes). These core collections are being intensively phenotyped on national levels (e.g., French initiatives Breedwheat <https://breedwheat.fr> and Amaizing <https://amaizing.fr>), or within

international initiatives, such as the Seeds of Discovery platform (<https://seedsofdiscovery.org>) for wheat and maize.

2.2 *Detection of Selection Signatures*

From the breeders' perspective, the value of genetic resources is given by the presence of agronomically interesting phenotypes that can be introgressed into elite germplasm. However, given the large number of genebank accessions multiplied by the number of potentially useful traits, phenotypic information is rarely available for genetic resources. Moreover, genetic determinants of many important traits are still poorly characterized. Among these knowledge gaps, genomicists explore crop genomes with the “bottom-up” approach (Ross-Ibarra et al. 2007) aiming to identify gene variants beneficial for crop production without phenotyping. This approach assumes that positive selection is the central force in the process of domestication and adaptation, and it is therefore possible to identify domestication and adaptive genes by screening signatures of selection along the genome. The general methodology is based on comparing genomically local diversity measures in the target population to some reference values, which can be modeled under the assumption of neutrality, or estimated from a genome-wide average in the population or orthologous regions of distinct populations. Although results are usually evaluated under some statistical framework to distinguish effects of selection and other evolution forces, specificity and sensitivity of these tests remain problematic: However, identification of genomic regions with limited genetic variability has double utility in the context of breeding. On the one hand, it helps to discover genes responsible for domestication and adaptation traits, and on the other hand, it points out to loci where re-introduction of lost diversity can boost resilience to environmental challenges.

Strong positive selection on a genetic variant with low initial population frequency results in a “selective sweep” (Maynard-Smith and Haigh 1974), the fixation of one haplotype around the selected allele. The following sections contain a brief description of the major genomic signatures of selection, together with a non-exhaustive list of available software tools. It needs to be emphasized that all these tools suffer to various extent from imperfect power (the ability to find real selection signals), specificity (the ability to filter out false positives), and resolution (the ability to identify the causal loci within long sweeps), as observed in association studies. Our ability to detect signatures of positive selection in a sample of genomes depends on the time elapsed since the selection episode, its strength and duration, the mutation and the recombination rates that break up haplotypes, as well as demographic events (intensity of bottlenecks, migration, differentiation, expansion...), which can actually create diversity patterns that resemble selection signatures. The resolution of the methods mostly depends on the extent of LD.

These factors need to be considered when interpreting genome-wide signatures of selection, and robust statistical thresholds are important to identify outlier loci, either

with respect to the rest of the genome or to another real or simulated population that was not under selection. In practice, multiple statistics need to be collected, and the more tests converge on the same result, the higher is the confidence that the identified locus is truly under selection. As in association studies, the identified loci need to be treated as “candidates” until phenotypes are established and the role of the genes is confirmed experimentally to avoid false positives (Pavlidis and Alachiotis 2017).

2.2.1 Decrease of Genetic Diversity

The most prominent signature of positive selection is the decrease of genetic diversity. As the frequency of the selected variant increases in the population, linked variation diminishes due to the genetic hitch-hiking effect (Maynard-Smith and Haigh 1974). This decrease of variation is easily detectable by comparing the nucleotide diversity (P_i ; H_e) of the studied population (e.g., a population from a specific environment, or a crop as a whole) to a reference population (a population from a different environment, or a wild progenitor). A major difficulty is to distinguish selection-related decrease of diversity from stochastic variation resulting from demographic processes. In practice, several-fold decrease of genetic diversity with respect to the reference population is regarded as a sign of selection. Outlier loci are identified based on a distribution of the values across the genome.

The decrease of nucleotide diversity in the vicinity of a selected variant is mainly due to a change of allelic frequencies on linked sites, rather than to a decrease of the total number of polymorphic sites. This shift in the Site Frequency Spectrum (SFS) toward high- and low-frequency derived variants is another signature of selection (Braverman et al. 1995) and is attributed to the fact that neutral variants that are initially linked with the beneficial allele increase in frequency, while newly-emerging neutral variants hitchhike with the selected allele, and therefore remain in the population. This shift in the SFS can be measured by a summary statistics Tajima's D (1989), where the average number of nucleotide differences between pairs of sequences (P_i ; π) is compared to the total number of polymorphic sites scaled by the sample size (Watterson's Θ ; θ_W). The lack of medium-frequency variants in the vicinity of the selected allele causes a decrease in π while the total number of polymorphic sites may remain unaffected, and this pattern is reflected by negative values of Tajima's D .

The statistical basis for the identification of the SFS shifts as signatures of selection was improved by the introduction of a Composite Likelihood Ratio (CLR) test (Kim and Stephan 2002). The CLR test compares the probability of the observed polymorphism data emerging under a standard neutral model with the probability of the data emerging under a selective sweep model. Nielsen (2005) introduced SweepFinder, a modification of the CLR test where the standard neutral model is replaced with an empirical SFS of the entire data set, which increases the robustness of the test under different demographic scenarios (e.g., mild bottlenecks). SweeD (Pavlidis et al. 2013) is another implementation of the CLR test that is numerically more stable and faster when analyzing large numbers of genomes.

2.2.2 Increase of LD

Local Increase of LD

A variety of tools that detect signatures of selection rely on the observation that haplotypes (stretches of DNA sequence uninterrupted by recombination) of recently selected genes extend much further than expected under neutrality. Extended Haplotype Homozygosity (EHH) (Sabeti et al. 2002), Integrated Haplotype Score (iHS) (Voight et al. 2006), Cross-population Extended Haplotype Homozygosity (XPEHH) which measures the reduction in haplotype diversity in cross-population comparisons (Sabeti et al. 2007), and nSL (Ferrer-Admetlla et al. 2014) are all based on the model of a hard selective sweep, where a de novo adaptive mutation arises on a haplotype that quickly sweeps toward fixation, reducing genetic diversity around the locus. If selection is strong enough, this occurs faster than recombination or mutation can act to break up the haplotype, and thus a signal of high haplotype homozygosity can be observed extending from an adaptive locus. These statistics, nSL in particular, retain some power to detect soft sweeps as well. They are implemented in Selscan that has been optimized for large datasets (Szpiech and Hernandez 2014).

Apart from the extended haplotypes, positive selection creates another specific pattern of LD. As the frequency of a beneficial mutation increases, together with the frequency of linked neutral variants, recombinations sometimes occur on either side of the selected mutation. Since recombinations on the two sides are independent, and double recombinations are much less likely, pairs of variants on each side of the beneficial mutation show elevated LD, but pairs of variants compared across the beneficial mutation show lower LD. This pattern can be measured by the ω -statistics (Kim and Nielsen 2004) that has been implemented in OmegaPlus (Alachiotis et al. 2012). Since the ω -statistics can be assessed at each interval between two SNPs, this method has the potential, at least in theory, to identify the locus under selection very precisely. However, it should also be noted that ω is only applicable when haplotypes are known, either on phased data or inbreds (e.g., in self-pollinating species).

All three aforementioned signatures of selection – decrease in nucleotide diversity, shift in the SFS, and a specific LD pattern – can be assessed simultaneously by RAI_{SD}, a tool introduced by Alachiotis and Pavlidis (2018). On modeled data, this composite evaluation test outperforms tools that measure those signatures of selection individually. However, unlike other methods, RAI_{SD} assumes that polymorphisms are sampled evenly across the genome, and this assumption may be severely violated (e.g., in exome data).

Global Increase of LD

In breeding programs, the detection of inbreeding is also relevant. Runs of homozygosity (ROH) are lengths of contiguous homozygous segments due to

transmission of identical haplotypes by parents in heterozygotes. It is the percentage of genome that is identical by descent. Individuals that have undergone recent inbreeding will exhibit long runs of homozygosity (MacLeod et al. 2009; Peripolli et al. 2017). ROH was adapted by Allier et al. (2019d) for inbreds and named ROHe (Runs of Expected Homozygosity).

Some of the causal factors behind the occurrence of ROH are population phenomena, such as genetic drift, population bottlenecks, inbreeding, and intensive artificial selection. Consequently, the identification and characterization of ROH can provide insights into how a population has evolved over time in the past, and additionally, into how a population has to be managed in the future in long-term breeding programs. Intense selection regimes in livestock populations have already alerted the scientific community about the need for strategies to preserve populations, characterize and monitor inbreeding, and manage the genetic diversity by optimizing genetic contributions and mating (See Sects. 3.3.7 and 3.3.8).

2.2.3 Extreme Differentiation

Local adaptation can also be indicated by extreme differentiation of allelic frequencies between genetic groups, especially in contrasted environments. As allelic differentiation can also result from demographic events, it is important to interpret outlier loci cautiously, especially when hierarchical population structure is detected. A relevant strategy to identify significant outliers relies on building a distribution of expected values of the tested statistics in the absence of selection, using neutral coalescent simulations (Bellucci et al. 2014).

A number of statistics are available (Cruickshank and Hahn 2014), with F_{ST} (single site divergence index) (Wright 1931) being the most common. Outlier differentiation methods rely on a hypothesis that under certain conditions (migration-drift equilibrium under a neutral island model with spatially uniform migration and gene flow), population differentiation of allele frequencies across a large number of loci can be used to infer the process of selection acting on a subset of loci (Lewontin and Krakauer 1973). Loci with F_{ST} values (or other genetic distance measure) significantly greater than the genome-wide distribution of the statistics are presumed to be under diversifying selection or linked to those under selection. FDIST(2) (Beaumont and Nichols 1996) implemented in LOSITAN (Antao et al. 2008) assumes a finite island model to generate null F_{ST} distribution and can deal with heterozygosity. ARLEQUIN (Excoffier and Lischer 2009) adds hierarchical genetic structure to the inference. BayeScan (Foll and Gaggiotti 2008) uses a Bayesian method to estimate the relative probability that each locus is under selection. FLK (Bonhomme et al. 2010) uses a modified version of the Lewontin and Krakauer (1973) test for selection by comparing allele frequencies of different populations in a neighbor-joining tree constructed using a matrix of Reynolds's genetic distances (Reynolds et al. 1983). Its extension hapFLK (Fariello et al. 2013) calculates haplotype-based frequency differentiation index among hierarchically structured populations. It is robust with respect to bottlenecks and migration

and can detect incomplete sweeps. OutFLANK (Whitlock and Lotterhos 2015) does not invoke any specific demographic model and uses a modified version of the Lewontin and Krakauer method to infer a null F_{st} distribution. $X^T X$ (Günther and Coop 2013) implemented in Baypass or Bayenv2 (Coop et al. 2010, p. 2010), employs a Bayesian hierarchical model to test individual SNPs against a null model generated by the covariance in allele frequencies between populations from the entire set of SNPs.

PCAdapt (Luu et al. 2017) assumes that genes under selection are outliers with respect to the prevalent population structure. It calculates z-scores that measure the relatedness of each SNP to the first K principal components of genome-wide variation in a population. The computation uses Mahalanobis distance, which is robust in the presence of hierarchical population structure. Comparisons on simulated data revealed that the false discovery rate of PCAdapt is around 10%, similarly to HapFLK and OutFLANK (Whitlock and Lotterhos 2015), but much lower compared to Bayescan (40%), which is negatively impacted by admixture. PCAdapt and HapFLK are the most powerful tools in scenarios of population divergence and range expansion.

Differentiation among populations can also be detected by comparing site frequency spectra. Selection does not only shift the SFS in the vicinity of the beneficial mutation toward high- and low-frequency variants (see Sect. 2.2.1), but it also causes multilocus allele frequency spectra to differ between two populations. A CLR test can be used to assess whether such local genetic differentiation departs from the expectation under neutrality, as implemented in XP-CLR (Chen et al. 2010) (<https://reich.hms.harvard.edu/software>).

Additional methods for identification of loci involved in local adaptation exist, but may not be applicable on large data sets (Hoban et al. 2016).

2.2.4 Specific Cases of Genetic Differentiation

Domestication/Selection

Genome scans have been performed in order to detect signatures of selection in most major crops where whole or partial genome sequence data is available for at least 100 accessions. The scans usually aim to distinguish domestication signatures (obtained by comparing traditional landraces to wild progenitors) from genetic improvement signatures (obtained by comparing landraces and modern cultivars).

Hufford et al. (2012) detected 484 loci showing signatures of domestication and 695 loci showing signatures of genetic improvement (with 23% overlap) in maize, using differentiation indexes (F_{ST} , XP-CLR), diversity indexes (π , ρ , Tajima's D and normalized Fay and Wu's H), and haplotype lengths in each genetic group. These results suggest that some traits are of continuous agronomic importance since domestication and additional traits became of interest during the breeding era. In total, 6–11% of the identified loci had no annotation and could correspond to regulatory regions. Some identified candidates of domestication genes controlled

flowering time, nitrogen metabolism, thousand kernel weight, phyllotaxy, and seed germination. Some genetic improvement candidates were involved in the biosynthesis of a plant growth hormone gibberellin, or in drought tolerance pathway. According to a gene expression survey, the greatest changes in expression (presence or absence of expression) were observed in the domestication genes, i.e. between wild and cultivated lineages. Expression of the candidate targets of selection in cultivated lines was more homogeneous, with subtle variations, perhaps indicating the importance of fine-tuned expression, as opposed to “on and off” variability. This observation suggests that while the domestication period mostly selected particular gene variants, selection during the improvement period acted predominantly on cis-acting regulatory elements. This information is of interest for private breeding programs (Allier et al. 2019d) that intend to monitor global and local losses and gains of diversity over time in their germplasm using genetic and genomic indicators.

In bread wheat, regions that lost diversity during domestication (Haudry et al. 2007), improvement (Reif et al. 2005; Cavanagh et al. 2013), or both (Pont et al. 2019) have also been investigated. By examining genetic differentiation (PCAdapt) and diversity patterns (Reduction Of Diversity π Index, Tajima's D), Pont et al. (2019) confirmed selection signatures for domestication genes conferring brittle rachis (*Br1*), tenacious glume (*Tg*), homoeologous pairing (*Ph*) and non-free-threshing character (*Q*); improvement genes controlling photoperiod sensitivity (*Ppd*), vernalization (*VRN*), reduced height (*Rht*), glutenins (*Glu*) and gliadins (*Gli*), frizzy panicle (*FSP*), grain number (*GNS*), waxy proteins (*Wx*), and plant architecture (*uniCULm*). Through F_{ST} screening among landraces and cultivars, Cavanagh et al. (2013) found introgression patterns surrounding phenology genes (*Rht-B1*: dwarfing, *Ppd-B1*: photoperiod insensitivity, *Vrn1*: flowering time) and the *Sr36* gene involved in resistance to stem rust.

In tetraploid wheat, Maccaferri et al. (2019) used genetic diversity differentiation indexes (F_{ST} , hapFLK, XP-CLR, and XP-EHH) between wild and domesticated emmer (*T. turgidum* ssp. *dicoccoides*, *T.t.* ssp. *dicoccum*, respectively), durum landraces, and cultivars. They confirmed selection on domestication genes (two brittle rachis regions, a glume QTL controlling threshability) and improvement genes, some of which are associated with disease resistance (e.g., *Sr13* and *Lr14*) and grain yellow pigment content loci (e.g., *Psy-B1*). They also identified *TdHMA3-B1* as the best candidate involved in phenotypic variation of Cd accumulation in the grain. The non-functional *TdHMA3-B1b* allele could be responsible for a reduction in root vacuolar sequestration of Cd and Zn. It was suggested that under Zn-limiting conditions, this allele increases the pool of Zn available for transport to the shoot, thereby sustaining shoot growth.

Genetic diversity and differentiation indexes (Tajima's D and F_{st} , respectively) screened on wild and domesticated yam aided the detection of root development (SCARECRFOW-LIKE gene), starch biosynthesis (*Sucrose Synthase 4* and *Sucrose Phosphate Synthase 1*), and photosynthesis related genes (Akakpo et al. 2017) that likely facilitated habitat change during domestication (from cultivation under trees to open field cultivation).

Adaptive Introgression

Screening for past introgressions, i.e. DNA fragments that were absent in the cultivated gene pool until some point in time and appeared in recent material from crosses with related species is another way to identify candidates of agronomic interest (Hufford et al. 2013; Racimo et al. 2015; Schaefer et al. 2016; Rochus et al. 2018). From a basic point of view, detecting gene flow or gene introgression from a distinct population or a different species can help understanding adaptation to various environments and evolution (Anderson 1953; Rieseberg and Wendel 1993). When it increases fitness, it is referred to as “adaptive introgression.” It can also help reconstructing speciation processes.

In principle, genetic introgression can be implied when genealogy of a locus in a population or species (i.e., “local ancestry”) does not match the “global ancestry” estimated based on genome-wide variation. This approach has been used, for example, to reveal the portion of loci introgressed from *japonica* rice into the *indica* cultivar 93–11 (Yang et al. 2011). Since it is impractical to quantify alternative gene tree topologies on a genome-wide scale using more than a few individuals (the number of possible rooted trees grows exponentially with the increasing number of tips), other methods are necessary to study introgressions on a population level.

Several approaches can be employed to detect past admixtures that concern the whole genome without resolving the ancestry of individual loci. They quantify fractions of genomes associated with distinct populations. These include multivariate analyses, such as PCA (Patterson et al. 2006; Jombart et al. 2009), or model-based clustering algorithms like STRUCTURE, NewHybrids, ADMIXTURE, and sNMF (Pritchard et al. 2000; Anderson and Thompson 2002; Alexander et al. 2009; Frichot et al. 2014).

Different algorithms have been proposed to detect individual introgressed fragments. A model-based inference implemented in HAPMIX (Price et al. 2009) uses phased data (i.e., known haplotypes) and known ancestral populations (assuming only two contributing populations). The central idea of the method is to view haplotypes of each admixed individual as being sampled from the reference populations. At each position in the genome, HAPMIX estimates the likelihood that a haplotype from an admixed individual originated from one reference population or the other. A Hidden Markov Model (HMM) is used to combine these likelihoods with information from neighboring loci, to provide a probabilistic estimate of ancestry at each locus (Fig. 1).

HAPMIX, LAMP-LD, and RFMix packages for local ancestry inference were developed to provide accurate results on human data and recent admixture events but may be difficult to parameterize for crop species. A recently published open-source software Loter does not require any biological parameters and can be applied to a wide range of species (Dias-Alves et al. 2018). Performance testing on simulated datasets revealed that HAPMIX is severely impacted by imperfect haplotype reconstruction, and Loter is the least impacted by increased time since admixture. Loter was used to infer local ancestries in aromatic rices that originated millennia ago through an admixture event between *japonica* rice and Indian *aus*-like rice (Civán

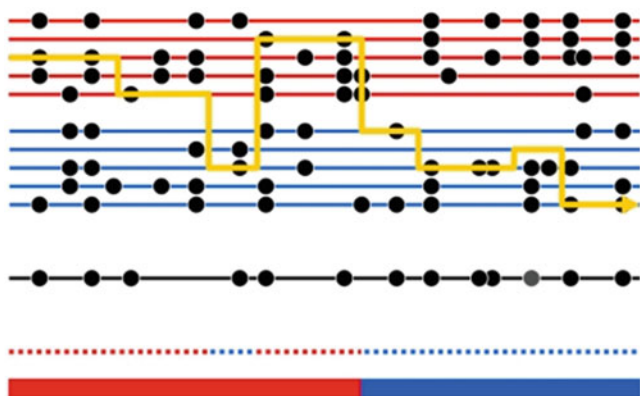


Fig. 1 Schematic representation of the Markov model used for ancestry inference. The black lower line in this figure represents a chromosomal segment from an admixed individual, carrying a number of typed mutations (black circles). The underlying ancestry is shown in the bottom color bar and reveals an ancestry change from the first population (red) to the second population (blue). The admixed chromosome is modeled as a mosaic of segments of DNA from two sets of individuals drawn from different reference populations (red and blue horizontal lines, respectively) closely related to the donors in the admixture event. The yellow line shows how the admixed chromosome is reconstructed with respect to this mosaic. The dotted line above the bottom color bar shows the reference population being copied from along the chromosome – note that at most positions, this is identical to the true underlying ancestry, but with occasional “miscopying” from the other population (blue dotted segment occurring within red ancestry segment). Reproduced from Price et al. (2009)

et al. 2019). However, the authors noted that the local ancestry inference was affected by sample size and missing data in simulations.

Another methodological approach mostly used to provide global estimates of gene flow is based on quantification of shared derived variation among non-sister taxa or populations (Kulathinal et al. 2009; Patterson et al. 2012; Peter 2016). In the absence of gene flow, a correct and rooted four-taxon tree will have the two most-recently diverged taxa sharing statistically equal amounts of derived variants with their non-sister taxon. Deviations from this expectation indicate gene flow. Popular implementations of this concept (often called ABBA-BABA) include the D-statistics (Green et al. 2010; Durand et al. 2011) and f-statistics (Reich et al. 2009, 2012). However, ABBA-BABA is based on a neutral evolution model, and its robustness in the presence of selection has not been tested. Since selection can increase local similarity among non-sister groups similarly to introgression, Civián and Brown (2018) argue that only variants demonstrably absent in their ancestral population should be counted toward the introgression signal.

Numerous cases of spontaneous introgression from wild relatives to cultivated species have been described (Hajjar and Hodgkin 2007; Guarino and Lobell 2011; Dempewolf et al. 2017; Burgarella et al. 2019). In maize, adaptive *mexicana* alleles were incorporated into the cultivated gene pool during the expansion of maize agriculture to the highlands of central Mexico (Matsuoka et al. 2002). Some of

these introgressed alleles have been functionally validated and shown to provide adaptations to altitude, biotic and abiotic stresses (Hufford et al. 2013; Fustier et al. 2019). In potato (*Solanum tuberosum*), the origin of tuberization under long days was traced to an introgression event from *Solanum microdontum* (Hardigan et al. 2017). Introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood) was detected through Tajima's D, F_{ST} and LD scans of local admixture, and complemented by analyses of gene expression (Suarez-Gonzalez et al. 2016). The team identified the *Populus PSEUDORESPONSE REGULATOR5* (*PRR5*) to be a strong candidate improving biomass, as well as cold, drought, and salinity tolerance. This gene was shown to work as a transcriptional regulator important for the circadian clock mechanism in Arabidopsis (Nakamichi et al. 2010, 2016, 2020). In poplar, it is upregulated at the onset of short days and it may play a crucial role in the timing of the onset of bud dormancy (Ruttink et al. 2007; Ko et al. 2011). A second candidate gene identified by Suarez-Gonzalez et al. (2016) is *COMT1* (*CAFFEIC ACID 3-O-METHYLTRANSFERASE 1*) that could be involved in lignification and/or pathogen defense (Barakat et al. 2011). In recent polyploids, such as bread wheat, authors consider PAV (presence-absence variation) or CNVs (copy number variation) identified from re-sequencing data as signals of putative introgressions (Balfourier et al. 2019; Cheng et al. 2019). Cheng et al. (2019) measured F_{ST} and π ratio between wild and cultivated lines and found 79 segments supposedly introgressed from wild relatives, co-localizing with 124 QTLs (grain yield, disease resistance, plant height).

The case of aromatic rice offers an example of how disentangling local ancestry and introgression history could aid the breeding process. It has been revealed that Basmati-like aromatic varieties of rice (Glaszmann 1987) originated from hybridizations between cultivated *japonica* rice (29–47% of the genome) introduced to the Indian subcontinent millennia ago, and local wild lineages of the Himalayan foothills similar to the present-day *aus* varieties (Civán et al. 2019). They possess some characteristics highly valued by consumers (fragrance, texture of cooked rice, grain elongation after cooking). Rice stickiness and texture after cooking is mainly determined by starch synthesis pathways, and particularly, the ratio of amylose and amylopectin. While *japonica* rice is often sticky (or glutinous) due to low amylose content, *aus* and *indica* varieties are generally nonglutinous. Aromatic varieties have intermediate amylose content and medium gel consistency and Civán et al. (2019) showed that they have mixed ancestry at the two genes, *Waxy* (Olsen et al. 2006) and *ALK* (Gao et al. 2011), responsible for these characteristics. Many aromatic landraces produce grain of superior quality in terms of fragrance and cooking properties, but are tall, lodging susceptible and low-yielding. Unfortunately, breeding efforts to cross aromatic landraces with high-yielding elite cultivars or introduce dwarfing genes were met with limited success. This is mainly attributed to cross incompatibility between aromatic and *indica* varieties, and high sterility in crosses (Singh et al. 2000). Local ancestry inference (Civán et al. 2019) revealed that most, but not all aromatic varieties carry a *japonica*-derived variant of the *S5* gene responsible for *japonica-indica* hybrid sterility (Chen et al. 2008). Identification of

high-quality aromatic landraces carrying non-*japonica* variants of *S5* could therefore be the first step in successful breeding of elite aromatic cultivars.

Environmental Differentiation: Landscape Genomics

Landscape genomics investigates associations of genetic variants with environmental variables, such as temperature, precipitation, altitude, and latitude gradients (Balkenhol et al. 2019). The goal is to identify candidate genes under selection, possibly indicating local adaptation, using outlier differentiation methods (see Sect. 2.2.3) and genetic-environment association (GEA) tests. Landscape genomics should not be confused with landscape genetics (Manel et al. 2003), which focuses on the effects of landscape variables on gene flow and population structure.

GEAs require environmental data such as WorldClim (<http://www.worldclim.org>, 2015). Bayenv2 tests for large allele frequency differences across environmental gradients by comparing observed allele frequency differences to transformed normal distribution of underlying population allele frequencies. Latent factor mixed models (LFMM) (Frichot et al. 2013) include population structure as latent (or hidden) variables to limit false positive signals. Spatial generalized linear mixed models (SGLMMs) (Guillot et al. 2014) are a computationally more efficient extension. Samβada (Stucki et al. 2017) is a multivariate analysis method that accounts for population structure with estimates of spatial autocorrelation in the data. When the trait of interest or the climatic gradient is correlated to the population structure, PCAdapt can also be used. Some other methods exist and are summarized in Rellstab et al. (2015).

Although most landscape genomics studies concern non-cultivated species, e.g. *Arabidopsis* (Hancock et al. 2011), associations with climatic data have also been investigated in forest trees (e.g., Sork et al. 2016; Rajora et al. 2016; Collevatti et al. 2020). Landscape genomics has been studied extensively in poplar (Suarez-Gonzalez et al. 2018), and it has been shown that introgressed genomic regions are enriched for disease resistance genes (TIR and LRR domains gene ontology terms) (Suarez-Gonzalez et al. 2016). In common bean (botanical name), 26 loci with selection signatures were found (Rodriguez et al. 2016), some of them involved in responses to environmental stress, such as drought response, cold acclimation or chilling susceptibility, and adaptation to different conditions of light and temperature. Four of these loci were also found to be under selection during domestication.

For sorghum, Lasky et al. (2015) showed that genome-environment associations can predict adaptive traits. Bellis et al. (2020) looked at correlation between allelic frequencies and *Striga* pressure. They demonstrated that local adaptation to this parasitic plant was partially controlled by the *LOW GERMINATION STIMULANT 1* (*LGS1*) gene. Wang et al. (2020) found some loci that may control seed mass adaptation to precipitation gradients.

In wild pearl millet (*Pennisetum glaucum*), Berthouly-Salazar et al. (2016) focused on climate gradients in Mali and Niger and collected genotype data from 11 populations, together with RNAseq data from a subset of four populations.

Looking at the genetic diversity patterns within populations (Tajima's D), differentiation among populations (F_{ST} , Bayescan), and correlation with environmental variables (centered loadings outliers using a PCA approach for each gradient), they found contigs displaying consistent signatures of selection among populations. Two of these contigs were associated with abiotic and biotic stress responses.

Time series data (that track samples over time) can be very informative for detecting genetic regions under selection. The factors involved in selection may be unknown.

Variety names of pearl millet, their phenotypes and climate data for a period from 1976 to 2003 were collected from a region of Niger that suffered from recurrent drought (Vigouroux et al. 2011). The research showed that an allele of the PHYC locus responsible for earliness increased in frequency over time at a rate that exceeds possible effect of genetic drift and sampling. A correlation between phenology and rainfall suggested that selection of this gene had a direct effect on earliness under shorter rainy seasons. It is noteworthy that this short-term adaptation was not due to introduction of new varieties, but due to within-variety selection, highlighting the importance to conserve within-variety diversity in genebanks.

Time series of a private breeding program (Allier et al. 2019d) can also be of interest to investigate genomic regions under selection or drift. Note that underlying population genetic structure and demographic history, when not properly accounted for, can generate many false positives. For instance, serial population bottlenecks occurring during founder effects of small populations migrating to new areas can result in fixed allelic differences due to genetic drift (Excoffier and Lischer 2009). Also, recent population range expansions from refugia can generate correlations between allele frequencies and environmental variables that are not due to selection.

Genome scan analyses are biased to detect loci with large effects, because power to detect small-effect loci is generally low. Since most phenotypic traits are likely to be polygenic, and thus governed by many loci of small effect, genome scans probably miss most of the loci involved in local adaptation (Stephan 2016; Rajora et al. 2016). Recently, Bayesian and other multilocus approaches have been developed (Rajora et al. 2016; Gompert et al. 2017). Some nonlinear functions have also been proposed to model the importance of environmental variables in explaining turnover of allele frequencies (Fitzpatrick and Keller 2015).

In reality, very few candidates for agronomically important genes revealed through genomic scans have been experimentally validated. Such validation usually requires validation by phenotyping in a controlled experimentation, an association study demonstrating a link between the genotypes and phenotypes (Saïdou et al. 2014) on a large panel of individuals, and a transformation experiment proving the function of the identified gene. All these experiments are costly, laborious, and technically difficult, but essential to convince breeders to monitor those genes in their germplasm. When validated, beneficial alleles can be introgressed into elite germplasm through backcrossing using diagnostic markers (Dempewolf et al. 2017), flanking markers, and sometimes genome-wide markers to minimize linkage drag and introduction of undesirable traits.

3 Population Genomics and Quantitative Genetics Assisted Infusion of Genetic Diversity in Breeding and Pre-breeding Programs

In elite germplasm, genetic diversity is generally limited compared to ancestral diversity. In that context, genebanks are a reservoir of underexploited favorable alleles. In case of one single favorable allele to introgress from genebanks to elites, flanking molecular markers can accelerate the process. But it takes a long time to validate QTLs and allele effects in different genetic backgrounds and design diagnostic markers to monitor the favorable allele in breeding programs. For example, it took 50 years (1960–2010) from the discovery of submergence-tolerant rice landraces to the successful release of submergence-tolerant rice varieties. It necessitated fine mapping and molecular characterization of the *SUBMERGENCE 1* (*SUB1*) locus and an introgression process (Bailey-Serres et al. 2010). This explains why only little use has been made of GR (Goodman 1999; Glaszmann et al. 2010; Wang et al. 2017).

The first reason why introgression process is so long is that elite lines contain much more favorable alleles than GR in general. It takes several generations of recurrent backcrosses with elites and selection to fill up this performance gap between GR and elites. The challenge is to break only unfavorable allelic associations while not breaking the favorable ones when crossing elites and GR. We may find co-adapted alleles in a cluster of genes (tall plant and late flowering alleles, high yield potential, and low protein content for instance) that have been selected by natural selection, creating local epistasy. No recombinants exist even in experimental populations because genes are too close. The recombination that would be desirable for agronomic purposes (small plant and late flowering, high yield potential and high protein content) may be difficult to obtain for ecophysiological or mechanical (low recombining regions, no diversity) reasons (Mayr 1954). The major unfavorable alleles of plant GR to eliminate are involved in phenology and local adaptation (e.g., flowering time, photoperiod sensitivity, height...) because they may not suit the targeted environment.

Genomic predictions (see Sect. 3.3) could help diversity infusion. Predicting GR genetic values using models that are trained on GR core collections is feasible when core collections are phenotyped in (and adapted to) targeted environments. But predicting which elite by GR crosses are of interest remains a statistical challenge because marker effects may depend on genetic backgrounds (Rio et al. 2019). We may need to first produce and evaluate recombinant lines between different genetic groups we want to cross to get an accurate marker effect estimation (GR alleles in elite genetic background in our case).

3.1 Production and Evaluation of Elite × GR Recombinants

Multi-parental crosses between elites and GR may be a good option to combine QTL detection for multiple-trait, identification of favorable GR alleles, selection of pre-breeding lines that could be introduced in breeding programs and training prediction models. Multi-parental Advanced Backcross (AB-QTL) populations (Narasimhamoorthy et al. 2006), pyramidal Multiparent Advanced Generation Inter-Cross (MAGIC) populations (Cavanagh et al. 2008; Leung et al. 2015), Nested Association Mapping (NAM) populations connected by one common parent for US maize (Buckler et al. 2009), European maize (Bauer et al. 2013), US sorghum (Bouchet et al. 2017), Back-Cross-NAM for sorghum (Jordan et al. 2011) have been developed for that purpose. It has been shown that the connection between populations by one or several common parents improves power of QTL detection. According to simulations, Stich (2009) proposed the triple round robin design connected by donors as the most efficient multi-parental design to maximize power of QTL detection as well as maximize genetic gain. But the production of this type of population remains long and laborious. The optimal connection design is not straightforward to predict from a statistical point of view. The choice of parents is often based on empirical information from different sources, the recipient parents being chosen for performance and GR for specific criteria that breeders want to improve, such as disease resistance for instance.

3.2 Marker Assisted Selection

Marker Assisted Selection (MAS) is promising to accelerate and optimize introgression process (Charmet et al. 1999; Servin et al. 2004). It has been successful for the introgression of maize earliness (Simmonds 1979; Smith and Beavis 1996), flowering time and yield under drought (Ribaut and Ragot 2007) as well as many disease resistance genes (Sanz-Alferez et al. 1995; Thabuis et al. 2004). But it becomes very demanding when multiple genes need to be pyramided simultaneously. Very large population sizes are actually necessary to get a reasonable certainty that an individual with the target genotype can be identified. Gene pyramiding strategies using marker-assisted introgression have been proposed (Hospital and Charcosset 1997; Servin et al. 2004; Canzar and El-Kebir 2011; Xu et al. 2011; Beukelaer et al. 2015). If all genes cannot be fixed in a single step of selection, it is necessary to cross again selected individuals with individuals having the favorable alleles that are missing using a marker-based recurrent selection (Charmet et al. 1999; Bernardo and Charcosset 2006). To cumulate more loci in a single genotype, Hospital et al. (2000) proposed a marker-based recurrent selection (MBRS) method using a QTL complementation strategy in a randomly mating population, which is feasible only in open-pollinated species. More recently, Valente et al. (2013) developed the software Optimas and Han et al. (2017) proposed the

Predicted Cross Value (PCV) algorithm to select at each generation crosses that maximize the likelihood of pyramiding desirable alleles in their progeny. A forward variable selection model can be used to select QTLs that explain significant genetic variance (Jansen 1993; Segura et al. 2012) instead of using arbitrary statistical thresholds. Note that Hospital and Charcosset (1997) advised that all QTLs should be given the same weight in the cross molecular score estimation to avoid rapid fixation of main QTLs and loss of small-effect alleles in the process. Control of genetic background with a few molecular markers was proposed in plants by Hospital and Charcosset (1997). With the same idea, the Genotype-Assisted Selection (GAS) concept was introduced by Meuwissen and Sonesson (2004) in animals to control polygenic background genes while selecting favorable alleles at QTLs. They proposed a multi-generation optimization of optimum contribution selection (GAOC: Genotype-Assisted Optimum Contribution) (see Sects. 3.3.7 and 3.3.8) while increasing the frequency of the positive QTL allele to increase genetic gain.

For complex traits that are controlled by a large number of genes, such as yield, MAS is often associated with substantial linkage drag, i.e., introduction of linked unfavorable alleles along with the target favorable allele (Peng et al. 2014) and often was a failure (Simmonds 1993; Hospital and Charcosset 1997; Ribaut and Ragot 2007). An approach using Genomic Selection (GS) addressing this problem in introgression schemes has been proposed (Ødegaard et al. 2009), who demonstrated that backcrosses assisted by genomic selection in fish is the best strategy compared to synthetic production or phenotypic selection to simultaneously select for elite productivity and donor disease resistance for instance. In wheat, Heffner et al. (2010, 2011a, b) came to a similar conclusion by comparing a breeding scheme including MAS with 20 QTLs or MAS followed by GS. Heffner et al. (2010) actually showed that expected annual genetic gain from GS exceeded that of Marker Assisted Recurrent Selection (MARS) for complex traits by about threefold for maize and twofold for winter wheat using analytical simulations of rapid cycles by skipping some phenotyping steps (Bernardo 2009), in a pre-breeding process in particular.

3.3 Genomic Predictions

First predictions in animals, human, and plants were based on pedigrees (Henderson 1975; Falconer et al. 1996; Bijma and Woolliams 1999). Then Lande and Thompson (1990) proposed to estimate the molecular score of an individual by adding its allelic effects at QTLs involved in trait variation. It was later shown that allele effects were overestimated in QTL detection (Beavis et al. 1994; Beavis 1998) and that the significance threshold to select the list of QTLs could be questionable. As the infinitesimal model (Fisher 1918) considering that traits are controlled by many loci of small effects was the best model to explain the variation of many complex traits such as yield, Whittaker et al. (2000) and Meuwissen et al. (2001) proposed to use all available independent markers (hundreds to millions) to build a prediction model that estimates the genetic value of unphenotyped candidates based on a

related training population that is phenotyped and genotyped. Considering that genotyping is dense enough, each QTL should be in linkage disequilibrium with at least one marker. The model is thus able to capture more genetic variance than including significant associations only. The principle is to regress phenotypic values on all markers considered as random effects using a linear model. The critical difference with the Lande and Thompson (1990) approach is that we do not set a significance threshold for the loci selected for trait prediction, but we use them all. This molecular score is called Genomic Estimated Breeding Value (GEBV) or genetic value and is an estimation of the additive effects of all loci.

The first to implement GS was the US dairy industry (VanRaden et al. 2007; VanRaden 2008). It doubled genetic gain (Schaeffer 2006; García-Ruiz et al. 2016) for this species particularly well suited for the implementation of GS. It is now applied to many other animal species (Hayes et al. 2009). Daetwyler et al. (2008) showed how to use genomic prediction for analyzing the genetic risk of human diseases. In plants, Bernardo and Yu (2007) and Heffner et al. (2011a) showed the first promising results using simulations, and Lorenzana and Bernardo (2009) using empirical bi-parental data.

More details on genomic selection and prediction models are presented in another chapter of this book (Andres et al. 2020). Here we discuss, how genomic predictions could be used to optimize re-introduction of genetic diversity in plant breeding and pre-breeding programs.

3.3.1 Selection of a Relevant Training Set

Assuming that the number of markers and the training population size is optimal, the accuracy of the calibration model strongly depends on congruence between the allelic composition of the training population (to build the prediction model) and the allelic composition of the candidates whose performance is to be predicted (Habier et al. 2007). When the prediction uses unrelated populations to train the prediction equations, prediction accuracy actually becomes negligible (Cossa et al. 2014). Different ways of estimating prediction accuracy of a training population were developed and have been reviewed (Brard and Ricard 2015). Methods to optimize the composition of the calibration set prior to phenotyping have been proposed based on the Prediction Error Variance or on the Coefficient of Determination (Laloë 1993) of contrasts in unstructured (Rincent et al. 2012) or in structured populations (Rincent et al. 2017b). The algorithm of Rincent et al. (2012) has also been extended to optimize the training population for multiple correlated traits using a criterion called CDMulti (Ben-Sadoun et al. 2020). Other approaches based on spatial sampling (Bustos-Korts et al. 2016), or kinship coefficients (Rincent et al. 2012, 2017b) potentially taking genetic architecture into account (Mangin et al. 2019) were also developed.

3.3.2 Genomic Predictions Assisted Introgression

Using simulations, genomic predictions were shown to be efficient for rapid introgression of GR alleles when implementing 3 cycles per year in maize (Bernardo 2009; Combs and Bernardo 2013). Among 100 simulated QTLs, the adapted inbreds had the favorable allele at 50 or more QTLs and the GR at 50 or less QTLs. They compared 1 year of phenotypic selection versus 3 cycles of genomic selection. The results indicated that a useful strategy should involve 7–8 cycles of genomic selection (2–3 years). They showed that genetic gain was higher when starting from an F_2 population rather than a backcross population, even when the number of favorable alleles was substantially larger in the adapted parent than in the GR parent. Note that they used random mating in their simulations. This procedure would require only 3 years to get some progenies that could be integrated in the breeding program. Allier et al. (2019b) showed, using simulated data and optimal parental contribution method (see Sects. 3.3.6 and 3.3.7), that in a context of multiple allele introgression from a donor into one or several elites, three-way crosses and backcrosses were more adapted than two-way crosses when donors underperform the elite population. They demonstrated that three-way crosses should be preferred because they produce more progeny variance and combine alleles from more parents. This supports the strategy adopted in the Germplasm Enhancement of Maize project (Goodman et al. 2000). Two-way crosses were actually more adapted when donors outperform the elite population for the targeted trait.

3.3.3 Predictions of Accessions' Genetic Values Conserved in Genebanks

Using genotypes and phenotypes of a representative set of genebank accessions, we can build a model to predict the GEBV of the rest of the collection (Yu et al. 2016). As genotyping is less expensive than phenotyping, we can identify this way supplementary GR of interest (Crossa et al. 2016; Brauner et al. 2018, 2019). For instance, in maize, Allier et al. (2019c) calibrated a prediction model on a population, assembling a continuum from old dent accessions to elite iodent material, including founders of breeding pools, elite material released into public domain, and elite material from different private breeders. Yield predictive ability between the calibration population and RAGT2n company germplasm was 0.404 and allowed to detect landraces of agronomic interest to be introduced in the breeding program. But this strategy is possible only if the divergence is not too large between landraces and elite material and the predictive ability is sufficient. It is also necessary that the traits can be evaluated for the landraces in targeted environments. It turned out to be an appropriate approach for biomass sorghum (Yu et al. 2016) and dent maize. But for many species, the presence of some major genes involved in phenology may hinder good quality phenotyping of landraces, because of incapacity to flower, to mature on time or lodging. In that case, unadapted accessions may carry interesting favorable

alleles but cannot reveal their potential in the targeted environment. Good quality phenotypes may necessitate to first convert GR by eliminating major phenology unfavorable alleles or to phenotype elite x GR hybrids instead of GR (Longin and Reif 2014). If we consider dominance effects of favorable over deleterious alleles for those major genes involved in phenology, heterozygous hybrids between elites and GR are expected to get favorable major alleles from elites that annul or at least reduce the effects of deleterious alleles from GR. But this assumes that hybrids are technically easy to produce which may be a challenge for autogamous species, at least laborious and expensive.

3.3.4 Optimization of the Allocation of Resources

Thanks to resource allocation optimizations some budget can be saved in evaluation of major traits (yield in general) and be transferred to

1. increase the size of the germplasm (the number of progenitors, crosses, and progenies per cross), leading to an increased genetic variance and a higher chance of creating outstanding individuals.
2. evaluate new traits (such as quality).

Different strategies have been proposed:

1. skip some field evaluation steps, which is relevant in long-lived species such as trees, or when trait values are expensive and/or become available late in the cycle (Hayes et al. 2009),
2. optimize the experimental design, i.e., minimize the number of lines or replicates evaluated in each environment,

Lorenz (2013) showed that it was advantageous to maximize population size at the expense of replication in a breeding program. Endelman et al. (2014), Heslot and Feoktistov (2017), and Akdemir (Akdemir and Isidro-Sánchez 2019) proposed efficient strategies to optimize field evaluation (sparse design) using genomic predictions. Ben-Sadoun et al. (2020) showed that it was possible to reduce budget by 25% for a fixed accuracy of French Bread Making Score by phenotyping it in a reduced number of environments. The idea is to evaluate all alleles in all environments, not all individuals.

3. accelerate cycles: speed (pre)breeding (2 cycles per year for winter wheat, 3 cycles for maize, up to 6 cycles for spring wheat) using adequate growth chambers and greenhouses protocols (Christopher et al. 2015; Hickey et al. 2017; Ghosh et al. 2018) to increase the rate of development,
4. diminish the cost of evaluation of an expensive trait using indirect measurements and optimize phenotyping of both correlated traits (Ben-Sadoun et al. 2020). The strategy is called Trait-Assisted genomic selection (TA) (Fernandes et al. 2018).

It is possible to predict two correlated traits simultaneously using multivariate best linear unbiased prediction (BLUP) (Henderson and Quaas 1976). Those models

benefit from information contained in both genetic correlation between traits and genetic relationship among individuals (Calus and Veerkamp 2011). The training population is genotyped and phenotyped for both traits. Each training individual is phenotyped for at least one trait. If the candidate population is genotyped but not phenotyped for any of the traits, the strategy is called Multi-Trait genomic prediction (MT). If some of the candidates are phenotyped for the secondary trait, the strategy is called Trait-Assisted genomic selection (TA) (Fernandes et al. 2018).

As for single trait prediction, under a major QTL genetic architecture, Jia and Jannink (2012) found that Bayesian multivariate models (BayesA or BayesC π) performed better than multi-trait GBLUP model. But for polygenic genetic architecture, multi-trait GBLUP model was equal to the Bayesian multivariate models. Note that Jiang et al. (2015) developed Bayesian multivariate models that consider correlated SNP effects. Montesinos-López et al. (2016) extended the model to a Bayesian multi-trait and multi-environment genomic prediction model (BMTME) that takes into account the correlation between traits and the three-way interaction term (Trait \times Genotype \times Environment). More recently, multi-trait deep learning (MTDL) models have been developed to reduce the computational resources (Montesinos-López et al. 2018, 2019). MT models can actually suffer from a high computational demand, time, and some convergence problems (Michel et al. 2018). Obviously, genetic correlation between traits is a key factor determining the MT advantage over single trait (ST) methods (Calus and Veerkamp 2011; Jia and Jannink 2012; Hayashi and Iwata 2013; Guo et al. 2014). Although MT models improve the predictive ability when the targeted trait has a low heritability and the secondary trait has higher heritability, the advantage of MT models to predict high heritability traits is low (Jia and Jannink 2012; Hayashi and Iwata 2013; Iwata et al. 2013; Guo et al. 2014). Studies using experimental data demonstrated that advantage of MT models to predict individuals which have not been phenotyped either for the trait of interest or the correlated trait was small or null in pine tree, *Pinus taeda* (Jia and Jannink 2012), soybean, *Glycine max* (Bao et al. 2015), rye, *Secale cereale* (Schulthess et al. 2016), maize, *Zea mays* (dos Santos et al. 2016), bread wheat, *Triticum aestivum* (Michel et al. 2018; Schulthess et al. 2018; Lado et al. 2018), and sorghum, *Sorghum bicolor* (Fernandes et al. 2018). Several studies using experimental data demonstrated that TA models perform better than ST and MT models in terms of accuracy. The TA models using high-throughput phenotyping, for instance, improved the prediction accuracy of bread wheat grain yield by up to 70% (Rutkoski et al. 2016; Sun et al. 2017; Crain et al. 2018). TA models also improved bread wheat baking quality-related parameters using protein content (Michel et al. 2018) or dough rheological traits (Lado et al. 2018) as correlated traits. Measuring dough strength (W) instead of French Bread Making Score for 75% of the population maintains accuracy by reducing budget of phenotyping by up to 65% (Ben-Sadoun et al. 2020). For a fixed budget, it can increase predictive ability by up to 0.14. Predictive ability of Fusarium head blight severity in hybrid bread wheat was improved using plant height and heading date as correlated traits (Schulthess et al. 2018). Fernandes et al. (2018) showed that TA models increased prediction accuracy by up to 50% when using plant height as correlated trait to predict yield in sorghum.

Robert et al. (2020) proposed a new TA approach, in which the secondary trait is not phenotyped for the selection candidates, but predicted with crop-growth models. The advantage is that it is not necessary to sow the selection candidates, as only the genotypic information is used.

3.3.5 Mating Optimization

The breeder's goal is to obtain "transgressive" individuals (with extreme genetic values) for at least one trait, cumulating as many favorable alleles as possible, putatively coming from different parents. While animal breeders optimize the choice of males, plant breeders may want to optimize mating between two or more parents. Cross design is essential but without accurate tools to guarantee its performance, breeders often select highest-performing parents to ensure high mean performance of progeny, and may focus on one or two traits. The problem is that highest-performing individuals may present similar sets of alleles and may actually produce less genetic variance in progeny than parents that have less but complementary favorable alleles. Because it is not feasible to evaluate all possible crosses in the field, it would be valuable to predict the value of a cross or a global cross design before it is made. Instead of focusing on the performance of parents, the idea is to estimate a proxy of the value of top progenies, i.e., the predicted mean and variance of the progeny. Attempts have been made using distances between parents based on phenotypes (Souza and Sorrells 1991a, b; Utz et al. 2001), genetic distance based on molecular markers (Bohn et al. 1999; Hung et al. 2012), molecular scores (summing QTL effects), or GEBV (summing marker effects estimated by ridge regression) (Tiede et al. 2015), but they were not really successful.

In a pre-breeding program context, it is even more obvious that the interest of a donor for a recipient elite individual depends on its genetic value (which can predict mean performance of progeny) but also its originality at QTLs (which will contribute to increase genetic variance and long-term genetic progress). A first approach was to count the proportion of favorable alleles and complementarity of parents at QTLs (Dudley 1984, 1988; Bernardo 2014). Van Berloo and Stam (1998) discriminated among crosses using a marker score from QTL flanking marker genotypes weighted by their effects.

The idea of Genomic Mating (GM) strategies is to use genomic predictions to optimize complementation of parents to be mated (Akdemir and Isidro-Sánchez 2016). As progeny genetic variance is generated by randomly sampling parental chromosomes during meiotic division, then recombination between those chromosomes, if we can accurately estimate marker effects as well as recombination rates between markers, we can optimize mating such as maximizing the probability to get individuals that cumulate a maximum of favorable alleles. In theory, the value of a cross, or the Usefulness Criterion (UC) of a cross (Schnell and Utz 1975) is the expected genetic value of the selected fraction (the bests) of the progeny

$$UC = \mu + ih\sigma_A$$

with μ the population mean, i the selection intensity, h the square root of heritability of the trait, and σ_A the additive genetic standard deviation among progeny.

Between Two Parents for Biparental Populations

To calculate UC of crosses, we need marker effect and recombination rates estimates. In plants, meiotic recombination maps are usually estimated from bi-parental populations. Note that we can use some other types of populations (F_2 , BC, HD), using adapted transformation to get the meiotic recombination rate. Several unconnected or connected populations can also be analyzed together to build consensus or composite maps cumulating more recombination information. A higher resolution method is to infer historical recombination maps from landraces or wild populations (Choi and Henderson 2015; Petit et al. 2017; Danguy des Deserts et al. 2021).

A first strategy is to simulate progeny in silico (stochastic simulations) by randomly producing crossing-overs along parental gametes according to a recombination map (Bernardo and Charcosset 2006). The value of a cross is the mean of the top progeny genetic values, the number of individuals belonging to this top group depending on the intensity of selection (Iwata et al. 2013; Bernardo 2014; Lian et al. 2015; Mohammadi et al. 2015). Note that in plants, we observe a significant negative relationship between parental mean and progeny genetic variance (Mohammadi et al. 2015). This study also showed that mid-parent value explains 99.99% of mid-progeny value and 82–88% of top-progeny value. Mid-parent value and estimated progeny genetic variance explained 99.5 of top-progeny value. This demonstrates the usefulness to estimate genetic variance and not only mean of progeny to estimate cross value. The problem with large breeding programs is that stochastic simulations are compute-intensive. So, attempts are made to predict variance using mathematic formulas (analytically). The mean of a cross is predicted by the mid-parent value in self-pollinated species or the mean of testcross performance in a hybrid crop. Several formulas have been proposed to predict the progeny variance. A first way is to estimate the value of the best possible progeny. We can determine historical haplotype blocks along the genome based on linkage disequilibrium and consider that recombination occurs only between those blocks. The effect of one haplotype block is the sum of its individual allele effects. Daetwyler et al. (2015) defined the Optimal Haploid Value (OHV) of an heterozygous individual as the sum of the effects of the best allele at each haplotypic block, corresponding to the genetic value of the best theoretical gamete to pass on to the next generation. They demonstrated that for a wheat program using DH technology (i.e., getting homozygous lines by gamete cultivation and chromosome doubling using colchicine treatment), genetic gain was improved (up to 0.6 standard deviations) when estimating the value of a cross as the OHV of the corresponding F_1 heterozygous individual

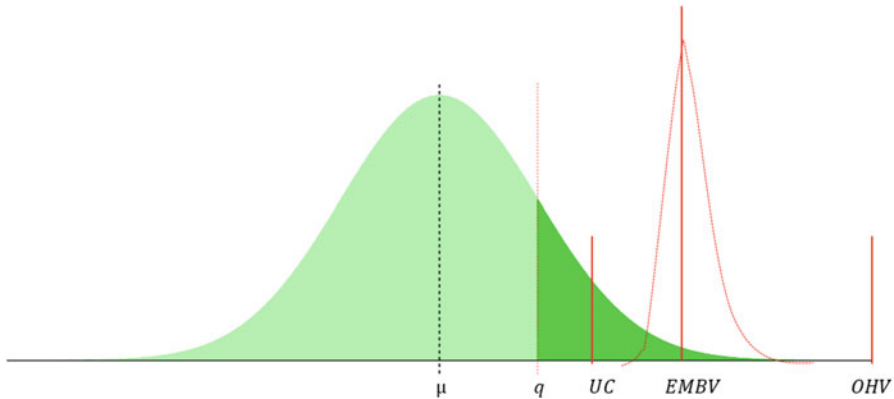


Fig. 2 Different indexes to estimate cross values. Along the distribution of a simulated progeny from one cross is indicated. μ : the mean of the simulated progeny, q : the 10% quantile of the top progenies, UC: the mean of the 10% top progenies, EMBV: the mean of the best progeny for 1,000 different simulations, OHV: the best possible progeny (if all recombinations are possible and population size is unlimited)

compared to standard GS. It also preserved a substantially greater amount of genetic diversity in the population. Müller et al. (2018) proposed the Expected Maximum haploid Breeding Value (EMBV) (Fig. 2). This is the expected GEBV of the best out of N DH lines produced by an F_1 using haplotypic blocks. Compared to OHV, EMBV actually takes into account the fact that the number of progenies produced is limited, the best theoretical progeny being impossible to reach. It can be estimated by stochastic simulations or using an analytical formula. Another analytical way to estimate the value of a bi-parental population explicitly includes the vectors of recombination rate between markers that are polymorphic between parents and marker effects (Zhong and Jannink 2007). This analytical formula estimates the probabilities of transmission of alleles at all QTLs from an F_1 individual (obtained by crossing two parents) to its gametes. In other words, the probability to get an outstanding progeny depends on the distribution of favorable alleles between parents and on the probability to break linkage between loci in repulsion phase and not to break linkage between loci in coupling. If we are interested in two genes that are genetically close to each other, if alleles are in the repulsion phase in the parental genotypes (neither parent has both favorable alleles), recombination widens the variance of the cross by providing extreme genotypes (you can get both favorable and unfavorable alleles in some progeny). On the opposite, if alleles are in the coupling phase in parents (you already observe the best and worse combinations), recombination reduces the variance (Zhong and Jannink 2007; Tiede et al. 2015) by getting combinations of intermediate effects. Formulas considering recombination rate between polymorphic markers were derived to calculate cross values for RILs and DH at generation k (Lehermeier et al. 2017b). The authors confirmed that predicting genetic variance in cross prediction increases genetic gain by 18% in maize compared to predicting the mean only. Formulas were also derived to optimize three- and

four-way cross designs (Allier et al. 2019b). The implementation is much faster compared to *in-silico* simulations. But note that in practice, we can only use analytical formulas to predict the next generation variance but not several generations ahead. We actually need to recover the parent genotypes at each cycle to estimate the variance of following generations.

Although Uemoto et al. (2015) suggested to filter MAF ($\geq 5\%$) to improve the prediction of GEBV, markers with low MAF should be kept for cross value prediction as they may be in greater linkage disequilibrium with low MAF QTLs and provide better predictions of the gametic variance (Santos et al. 2019).

Although the superiority of predictions of GEBV using haplotypes instead of single markers has not been demonstrated, Cole and VanRaden (2011) and Bonk et al. (2016) recommended the use of haplotypes to predict cross values in order to limit sampling errors when estimating individual marker effects. Another way to take into account local LD and uncertainty of markers estimates is to use Bayesian estimates of single marker effects (Sorensen et al. 2001; Lehermeier et al. 2017a). The idea is that combinations of alleles in haplotypic blocks may be better estimated (if present in the training population) than individual SNPs.

In a maize pre-breeding context, Allier et al. (2019c) compared different indexes to estimate cross values: the Modified Roger's Distance (MRD) between parents, the proportion of favorable alleles in donors (K) and recipients (J) (Bernardo 2014), OHV, genetic variance VarG in progeny and Lerhermeier's UC (Lehermeier et al. 2017b). They considered different selection rates in the progeny to calculate UC, 5% (UC1) and $10^{-8}\%$ (UC2). The main conclusion was that one might consider UC1 or OHV with large haplotypes for short-term genetic gain prediction, OHV with small haplotypes or UC2 with stringent selection for long-term genetic gain prediction. In other words, complementarity between parents is more important to consider for long-term genetic gain. Another conclusion was that in genetic diversity conservation programs, one might just want to maximize progeny variance (VarG) for the trait of interest, or the MRD between donor and recipient in the absence of trait-specific considerations.

At the Population Level

The long-term potential of a breeding program relies on the efficiency to combine favorable alleles scattered within many individuals (Goddard 2009; Jannink 2010). In a pre-breeding program where we want to increase the number of favorable alleles in a population, this can be optimized using Genotype-Building (GB) strategies. It is the founder population as a whole and not individuals or parents which must cumulate favorable alleles at a maximum number of haplotypic blocks. A parent (founder) is chosen for its complementarity with others. It may carry only a few rare but very favorable alleles and have a low individual genetic value. Considering that the best allele combination (ideotype) is known, there may be many possible cross

designs to get there. Because we cannot test all founder populations and cross combinations, the challenge is to build algorithms and solvers so that calculations are feasible and solutions are realistic.

The first proposed strategy was to select a subset of founders that possess altogether the best possible combination of haplotypic alleles along the genome. The Genotype-Building (GB) value of a subpopulation (Kemper et al. 2012) measures the GEBV of an ideal heterozygous progeny that would get the two best haplotype segments from two founders for each block. The Optimal Population Value (OPV) (Goiffon et al. 2017) is an extension of GB for inbreds. It measures the GEBV of the best possible homozygous progeny that can be produced, i.e., the value of the progeny that would get the best allele for each haplotypic segment in the founders. Note that it supposes an unlimited number of generations. The second extension is to consider time and resource constraints. Moeinizade et al. (2019) proposed the LAS (Look-Ahead Selection) algorithm where they improve the population for a few generations, starting with a subset of founders that maximizes OPV, and finally select for the best individuals in the last generation. They also consider a limited budget and vary the numbers of progenies produced from different crosses based on the genetic diversity of the parents: they spend more resources on those crosses that have wider predicted phenotypic distributions and thus higher probabilities of producing outstanding progenies. As for OHV, for GB, OPV, and LAS we assume adjacent markers are likely to segregate together and are grouped into representative haplotype blocks, recombination events occurring only between haplotype blocks.

3.3.6 The Theory of Contributions

According to the “breeder’s equation” (Lush 1937), genetic gain is limited by the initial additive genetic variance in breeder’s germplasm

$$\Delta\mu = ih\sigma_A$$

with $\Delta\mu$ (genetic gain) the expected change in mean performance per generation, i the selection intensity, h the square root of heritability of the trait, and σ_A the additive genetic standard deviation among progeny.

The level of diversity depends on the effective population size N_e (Fisher 1930; Wright 1931), which refers to the number of breeding individuals in an idealized panmictic population with absence of selection that would show the same amount of genetic diversity than the real population. Genetic diversity is generally measured by the expected heterozygosity He (Nei 1973). While the expected response to selection is proportional to the selection intensity, the number of reproducers and the corresponding effective population size is inversely proportional to the square of selection intensity on major QTLs (Sanchez et al. 2006). Consequently, maximizing selection intensity (using GS for instance) to maximize short-term genetic gain reduces the effective population size and long-term genetic gain.

The genetic gain is also proportional to the product of individuals' contributions (i.e., the number of offspring of each cross) and deviations from population mean (Woolliams and Thompson 1994; Woolliams et al. 1999). The rate of inbreeding, i.e. loss of diversity, is inversely proportional to the square of individuals' contributions (Robertson 1961; Wray and Thompson 1990).

3.3.7 Optimization of Contributions with Diversity Constraints

Based on the theory of contributions, the optimum contribution concept has been developed in animal breeding programs (James and McBride 1958; Wray et al. 1990; Wray and Goddard 1994; Brisbane and Gibson 1995; Meuwissen 1997; Woolliams et al. 2015) and tree breeding (Kerr et al. 1998; Hallander and Waldmann 2009a, b) to limit inbreeding. These methods have been recently adapted in crop breeding (Akdemir and Isidro-Sánchez 2016; Lin et al. 2016; Cowling et al. 2017; De Beukelaer et al. 2017; Akdemir et al. 2018; Gorjanc et al. 2018; Allier et al. 2019a, b, 2020).

The vector of parental contribution to the next generation is chosen at a predefined rate of population inbreeding (Wray and Goddard 1994; Meuwissen 1997), penalizing this way individuals that are too closely related and maintaining genetic diversity. The solution is a compromise between short- and long-term genetic gain, is heuristic, and can be optimized using different types of algorithms such as evolutionary algorithms, genetic algorithms in particular (Holland 1962; Goldberg and Holland 1988). When there is no explicit solution for complex problems (some objectives are not independent), simulated annealing and genetic algorithms are efficient to explore the solution space, obtain a pseudo-optimal solution, and limit the risk to get local minimum solutions. Simulated annealing (Metropolis et al. 1953) uses a Monte Carlo criteria, i.e., a probability of acceptance of a solution. New solutions are proposed until the algorithm converges, i.e., no new solution improves the objective functions. The number of iterations with different starting points for the decision variables and the choice of convergence criteria to decide to stop the algorithm are essential. Genetic algorithms (Holland 1962) work on a population of solutions instead of individual solutions: (1) The algorithm generates a population of possible solutions, each one with defined values for the decision variables. (2) The values defined for decision variables are considered as alleles at different loci (one locus = one decision variable and one allele = one value for the decision variable). (3) The algorithm creates new solutions from existing solutions by "reproduction" with mechanisms similar to genetic evolution (genetic recombination, mutation, selection). Those transition rules are probabilistic. Different reproduction (one point, uniform) and selection (roulette wheel, tournament, rank) operators exist to propose and choose solutions.

Allier et al. (2019b) combined optimum contribution with Usefulness Criterion (UC) (Lehermeier et al. 2017b) strategies in maize. They evaluated the interest of a multi-parental cross implying a donor and one or several elite recipients using the UCPC (Usefulness Criterion Parental Contribution) criterion. They simultaneously

predicted the full multivariate progeny distribution (mean, variance, and pairwise covariances) for the agronomic trait, genome-wide contribution of parents, and contributions at favorable alleles. They showed using this strategy that three-way crosses were more efficient for long-term genetic gain when donors are less performant than elites.

In animals, to maintain diversity, Bijma et al. (2020) proposed to produce a number of offspring that is proportional to the gametic variance of the reproducer to accelerate response to recurrent selection.

In addition to contributions, we can optimize mating in Optimal Cross Selection (OCS) approaches. It aims at identifying the optimal set of crosses maximizing the expected genetic value in the progeny under a constraint on genetic diversity in plants. It combines optimal contribution with optimal mating in a multi-objective problem that can be also optimized by heuristic algorithms. The classical OCS approach controls for genetic diversity in the total progeny. Allier et al. (2019a) applied OCS under a constraint on genetic diversity in the selected fraction of the progeny that is used as parents of the next generation accounting for within-family variance. They applied UCPC-based (Usefulness Criterion Parental Contribution) OCS in maize using a differential evolution algorithm (Storn and Price 1997; Kinghorn et al. 2009; Kinghorn 2011). They showed that OCS with constraints on UCPC and He was more efficient than classical OCS for long-term genetic gain with limited reduction of short-term genetic gain. Akdemir et al. (2018) maximized within-cross variance (Shepherd and Kinghorn 1998) and mating for multiple traits. It gives the list of parent mates that maximize gain, maximize cross variance, and minimize inbreeding. It is called Multi-Objective Optimized Breeding (MOOB). Compared to standard multi-trait breeding, the gains from multi-objective optimized parental proportions approaches were about 20–30% higher at the end of long-term simulations of breeding cycles.

The budget and the technical solutions are so different between species, private/public sector, that it is difficult to propose one algorithm that would handle conflicting objectives and satisfy the whole community (Wellmann 2019; Wellmann and Bennewitz 2019).

GS being more efficient at fixing major QTLs, it accelerates the loss of genetic diversity at QTLs according to simulation studies (Jannink 2010; Lin et al. 2016; Ben-Sadoun et al. 2020). Moreover, using RR-BLUP, the rare allele effects are shrunk toward zero, which increases the risk to lose individuals with rare favorable alleles and decreases long-term genetic gain (Goddard 2009; Jannink 2010; Habier et al. 2010; Pszczola et al. 2012). Several authors suggested to up-weight rare favorable alleles (Goddard 2009; Jannink 2010; Sun et al. 2014; Liu et al. 2015a) to select individuals for the next generation. They obtained encouraging results by simulation but did not propose stabilized rules to assign relevant weights to markers.

Considering computation time, several papers concluded the possibility for elite material to pre-select the population of eligible crosses based on parental mean genetic values before optimizing the cross design according to progeny genetic variance estimations (Zhong and Jannink 2007; Lehermeier et al. 2017b). They show that the genetic gain at the following generation is similar when considering

all possible crosses or when removing couples with lower mean genetic values. But the conclusion may be different in more diverse materials. In that case, crosses that have high variance but low mean may be interesting for long-term genetic gain, if we wait a sufficient number of generations to give a chance to rare favorable alleles to be selected in a pre-breeding population for instance.

3.3.8 Multiple Traits Optimization

The performance of new varieties often depends on multiple traits and/or constraints. The targeted ideotype can be a compromise between yield and quality for instance, with specific molecule concentrations for the industry. Breeding for multiple traits simultaneously is challenging because some traits are uncorrelated or unfavorably correlated due to linkage or pleiotropy. Bulmer effect (Bulmer 1971) actually mechanically creates negative correlations between traits under selection, yield and protein content in wheat, for instance. Moreover, the economic value of different traits may not be equally important.

In classical multi-trait selection where traits are not correlated or negatively correlated, we have several strategies: (1) tandem selection: we select each trait singly (at different steps or generations), (2) independent culling: we reject individuals that are not meeting required standards for all traits, (3) index selection: traits are combined, using different weights corresponding to economic value, into a score that is considered as a single trait. The problem is that it may exclude the best individuals for each trait and some beneficial alleles. And it does not control for inbreeding.

Although single-objective optimization problems may have a unique optimal solution, the chance to find the best solution to a multi-objective problem is very low. The solution will be a compromise, especially when traits are antagonistic. And there may be several interesting solutions depending on the ranking of objectives. Algorithms propose a multiplicity of compromise solutions called Pareto optimal solutions after judiciously scanning the decision space, i.e., different combinations of equality and inequality constraints. Population of solutions are classified into boundaries according to their level of dominance (see more explanations in Figs. 3 and 4 for two traits, Fig. 5 for three traits).

Note that at the end of a multi-objective optimization, the decision maker still has to select the preferred solution from the Pareto frontier using its own decision rules, i.e., ranking or weighing objective functions like in index strategies.

3.3.9 Production of Varieties Adapted to Local Constraints

The objective of plant breeders is to produce new varieties well adapted to target environments. For this purpose, they evaluate candidate lines for several years in multi-environment trials. Because phenotyping is expensive, only a limited number of lines are evaluated each year in a small number of environments.

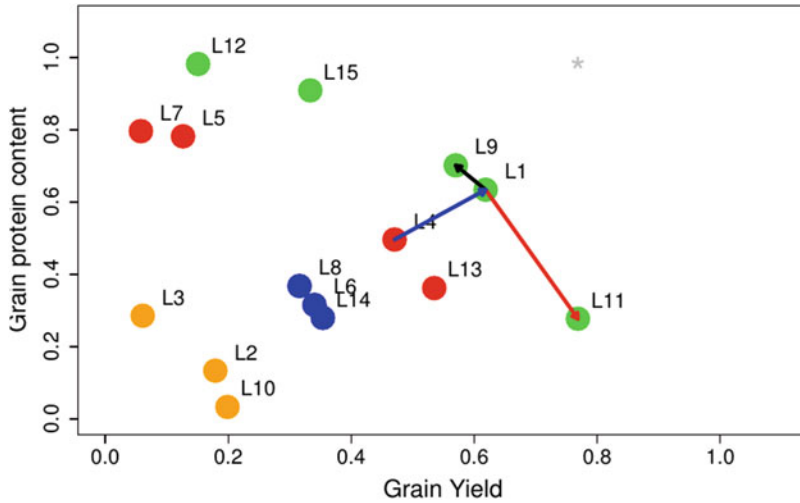


Fig. 3 Genetic values of grain yield and grain protein content for 15 genotypes. Different colors represent different levels of solutions (green: first level non-dominated, red: second level dominated by first level, blue: third level dominated by first and second level, orange: fourth level). Reproduced from Akdemir et al. (2018)

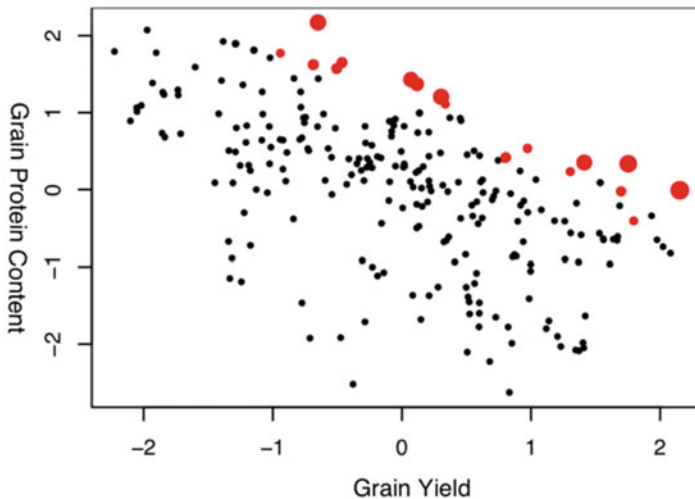


Fig. 4 Optimization of grain yield and grain protein content: Pareto frontier curve. Red points indicate the individuals that are selected by the algorithm. The size of the points is proportional to their contribution to the next generation. Reproduced from Akdemir et al. (2018)

Using genomic predictions accounting for Genotype by Environment Interactions (GEI), we can explore more combinations of genotypes and environments that we cannot afford observing in the field. We can use historical breeding databases

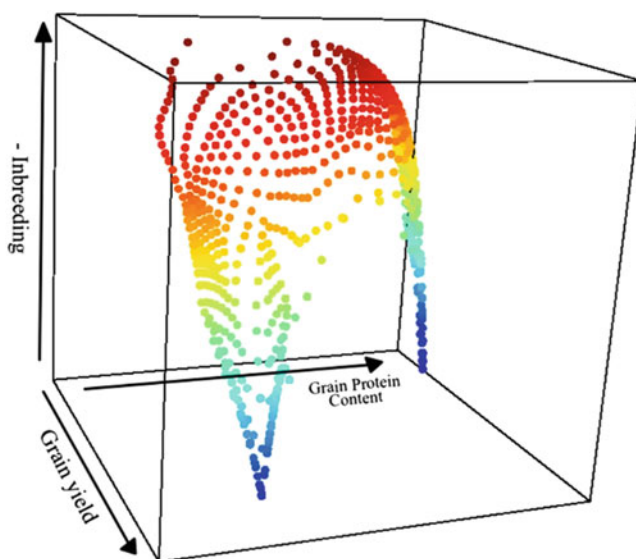


Fig. 5 Pareto optimal solutions for parental contributions (wheat data) obtained by solving the optimization problem for three parameters (three dimensions). The objective was to improve grain yield (GY) and grain protein content (GPC) while controlling group coancestry, i.e., maximizing GY, GPC, and the negative of inbreeding. The redness of the points indicates closeness to ideal solutions. Pareto optimal solutions are represented by a knee. Reproduced from Akdemir et al. (2018)

including numerous years and environment observations to calibrate those models. Different approaches have been proposed in the last decade.

Genotype by Environment Interactions (GEI) Predictions

While classical GS models rely on main effects and are not able to predict GEI (Crossa et al. 2010; Ly et al. 2013), those were adapted to predict environment-specific effects (Schulz-Streeck et al. 2013; Lopez-Cruz et al. 2015; Crossa et al. 2016; Bandeira Sousa et al. 2017), with possibly a genetic covariance between environments (Burgueno et al. 2012; Lado et al. 2016; Cuevas et al. 2017, 2018). These approaches, similar to multi-trait models, can increase prediction accuracy and can predict missing phenotypes of observed varieties (sparse testing) or unobserved varieties. They are more efficient in the sparse testing scenario in which information on a given variety can be shared between similar environments. Specific R packages were developed to fit simple GEI models with optimized computational properties (De Coninck et al. 2016; Granato et al. 2018). But these models cannot be used to make predictions of a genotype performance in new environments, as they rely on the phenotypic data to estimate the covariance between environments.

To extend the predictions to new environments, Heslot et al. (2014), Jarquín et al. (2014), Malosetti et al. (2016), Millet et al. (2019), and Rincent et al. (2019) proposed to characterize environments with environmental covariates, like molecular markers are used to characterize varieties. These covariates are pedoclimatic characteristics supposed to affect the plants (precipitations, extreme temperature, radiation deficit) at the different developmental stages (Brancourt-Hulmel 1999). A crop model can be used to estimate the timing of the developmental stages, so that the covariates are estimated for a period during which they are supposed to impact plants. This work is inspired from the factorial regression methodology (Brancourt-Hulmel et al. 2000) in which a regression on a covariate explains the variability of the trait in presence of GEI. A generalization of factorial regression on a given covariate to the GBLUP mixed modeling context was proposed (Ly et al. 2017, 2018). In these studies, the covariate has a variety specific random effect with a variance/covariance matrix structured by the kinship. This allows predicting the sensitivity of new varieties to this covariate. It is important to note that the QTLs affecting main effects are not necessarily the same as the QTLs affecting GEI, and this can be taken into account in the statistical models at the marker level (Heslot et al. 2014) or at the kinship level (Rincent et al. 2019). These models involving environmental covariates are particularly useful in the context of climate change, because they can predict the behavior of various varieties in virtual prospective scenarios. If a relevant database exists to calibrate the GS model, it could be used to identify in-silico interesting combination of alleles to face given environmental conditions. If we consider that the genetic diversity available in the elite pool is not sufficient, the prediction models can also be used to screen genebanks for valuable GEI (Crossa et al. 2016; Yu et al. 2016).

Ecophysiological Modeling

The adaptation of plants to their environment has been long studied by ecophysiologicalists. Their research has allowed developing Crop Growth Models (CGM), which describe plant development using mechanistic relationships with physiological parameters and environmental covariates as inputs. In other words, the CGM simulates GEI by taking into account the specificities of the varieties (genetic parameter) and of the environments (environmental variables). Different ways of using CGM to predict GEI were proposed in the past.

The first application is to predict the developmental stages of the plants to estimate if stress appeared at critical stages. This strategy was applied in wheat and maize (Heslot et al. 2014; Jarquín et al. 2014; Malosetti et al. 2016; Ly et al. 2017; Millet et al. 2019; Rincent et al. 2019). Numerous studies indeed revealed that CGM were efficient to predict phenology even for new varieties (White and Hoogenboom 1996; Nakagawa et al. 2005; Yin et al. 2005; Messina et al. 2006). CGM can also be used to directly derive environmental covariates (Ly et al. 2017; Rincent et al. 2019). In Rincent et al. (2019), CGM SiriusQuality (Martre et al. 2006) was used to estimate dry matter stress index (DMSI) that directly relates to the

impacts of temperature, drought, and N deficit, alone or in combination, to daily biomass loss. The idea is to produce stress indexes as close as possible to what the crop experienced in the field. Such variables directly simulated by the CGM were shown to better capture GEI than basic pedoclimatic covariates.

The second application is much more ambitious: the genetic model and the CGM are fully integrated within the Gene-Based Modeling approach (GBM). In GBM, the CGM simulates the development of each variety by using variety specific genetic parameters as input. These genetic parameters (phyllochron, sensitivity to photoperiod) characterize the varieties independently from the environment and are thus supposed to be stable across environments. Once the genetic parameters are estimated for the calibration set, a GS model can be calibrated to predict the genetic parameters of new varieties. These predictions can then be used as input of the CGM to predict the target trait of the new varieties in various environments. The interest and feasibility of this approach coupling CGM and genetics have been validated for leaf elongation rate in maize (Reymond et al. 2003; Chenu et al. 2008), fruit quality (Quilot et al. 2005; Prudent et al. 2011), and phenology of various species (White and Hoogenboom 1996; Nakagawa et al. 2005; Yin et al. 2005; Messina et al. 2006; White et al. 2008; Uptmoor et al. 2012; Zheng et al. 2013; Bogard et al. 2014; Onogi et al. 2016; Rincet et al. 2017a). Recently, Technow et al. (2015), Cooper et al. (2016), and Messina et al. (2018) have illustrated the possibility of coupling CGM and GS models for predicting highly integrated traits such as grain yield. One major advantage of their approach and that of the work of Onogi et al. (2016) is that the genetic parameters and the marker effects are jointly estimated, and so information can be shared between individuals thanks to genotypic data. However, using GBM to predict such complex traits remain challenging, as numerous genetic parameters have to be phenotyped or estimated on the training population. More recently, Robert et al. (2020) proposed to combine GBM with a trait-assisted prediction approach. The GBM is used to predict a secondary trait (heading date) for the test set in all environments. This secondary trait is easy to predict, and its relationship to the target trait (yield) is environment specific and thus allows predicting environment-specific effects in bread wheat.

A last application of CGM is to help clustering environments with similar properties. The objective is to use the CGM to characterize the stressing conditions experienced by the plants in each environment, and then to group environments with similar scenarios. Taking pedoclimatic data and variety characteristics as input, CGM can indeed produce daily stress indexes from sowing to maturity. It has been shown that clustering based on stress scenarios identified by CGM was more relevant than clustering based on the experimental protocols (e.g., non-irrigated vs irrigated) and that it was efficient to capture GEI (Chenu et al. 2011; Touzy et al. 2019). For example, it can happen that in a multi-environment trial, an irrigated trial is more subjected to drought than a non-irrigated trial at another location. In contrast, the CGM is able to finely characterize each environment by taking into account the environmental conditions and the plant development. Once the CGM-based clustering is obtained, reference GS models (or GWAS) can be applied within each cluster, GEI being taken into account by the clustering.

Perspectives in the Field of GEI Prediction

Phenotyping is one of the main bottlenecks in plant breeding. GS models allow predicting new varieties in observed environments or new environments for observed varieties, but large phenotype databases are necessary to calibrate the GS models accurately. High-throughput phenotyping platforms and tools which allow phenotyping at the organ level, at the plant level, or at the plot/field level (Tardieu et al. 2017) constitute a great opportunity to calibrate GEI models. This observation can be used to calibrate CGM (Reymond et al. 2003) or as environment-specific proxies of the target trait (Amani et al. 1996). The systematic and wide use of sensors in the breeding programs will probably allow using deep learning approaches, supposed to be the most efficient when such large datasets are available. Note that in all the approaches described in this section, there were only two kinds of data involved in the model: genomic and phenotypic data. The introduction of other omics data such as transcriptomics, proteomics, and metabolomics in the models will probably allow a better understanding of how a given variety grows in various environments (see Sect. 4.2 below). The introduction of this information in “phenomic” prediction models or in Genomic-like Omics Based prediction models (GLOB) was proven to improve accuracy (Fu et al. 2012; Riedelsheimer et al. 2012; Rincet et al. 2018; Schrag et al. 2018). The combined use of phenomics and genomics is used in pre-breeding for yield potential in stressed environments under the International Wheat Yield Partnership (IWYP, <https://iwyp.org/>) (Reynolds et al. 2021). Once those tools are cost effective, they could be integrated routinely in breeding programs.

3.3.10 Application to Pre-breeding

When performance gap between donors and elites is too large, it may be judicious to improve a pre-breeding population before introducing GR in a breeding program. For a few generations, starting from relevant founders that bring complementary alleles and mating optimization, we can increase gradually the number of favorable alleles in the population. It is only after a sufficient number of generations that we start selecting individuals based on their genetic value to cross them with elites. Gorjanc et al. (2016) provided guidelines based on stochastic simulations. Starting from 3,000 genotyped maize landraces, they evaluated different pre-breeding programs that differed according to the population to initiate crosses: (1) the best landraces, (2) the best testcrosses, or (3) the best DH seeds derived from testcrosses. They tested different (1) sizes for the pre-breeding program, (2) levels of diversity within the 3,000 landraces, (3) trait heritabilities, (4) number of markers, (5) number of crosses and progeny size per cross, and (6) number of phenotypic observations. The highest genetic gain was achieved by initiation with testcrosses. But it was reconstructing the elite genome and not utilizing the landrace favorable alleles. The best compromise to start a pre-breeding program was to start from landraces. This process can be accelerated by using existing composite or recurrent selection

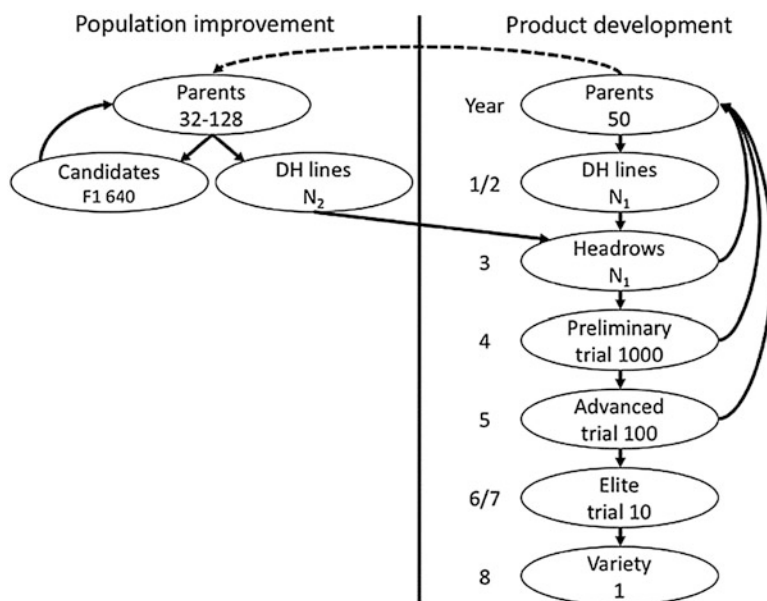


Fig. 6 Two-part breeding program. The population improvement component is based on recurrent selection that brings new genitors to the breeding program. Conventional strategy is based on variety development from elite parents. Reproduced from Gorjanc et al. (2018)

populations or inbred lines derived from local landraces. A recent initiative to characterize and use a part of the untapped variation in maize landraces is the Seeds of Discovery project (SeeD: <http://seedofdiscovery.org>). SeeD develops germplasm with 75% or more elite and 25% or less landrace genome to provide donors carrying new alleles.

Two-step breeding programs with an integrated pre-breeding program using rapid cycles (recurrent selection) (Gaynor et al. 2017; Gorjanc et al. 2018) is an efficient way to improve long-term genetic gain according to simulations (Fig. 6). An improvement population is produced by recurrent genomic selection with several cycles per year to increase the mean value of GR population in the pre-breeding program. A development population is produced using standard methods to develop new lines in the breeding program. It delivered about 2.5 times larger genetic gain compared to a conventional program for the same investment according to Gaynor et al. (2017) simulations. OCS increased long-term genetic gain by 15–78% depending on the number of parents.

Allier et al. (2020) proposed a strategy in three steps in case of a very large gap between elites and GR. They called base broadening phase (pre-breeding) the recurrent improvement of GR to decrease the performance gap with elites. It is kept independent from breeding programs until performance is satisfying. Best progenies are then crossed with elites to produce a bridging population. And the best bridging progenies can be parents in standard breeding programs. Allier et al.

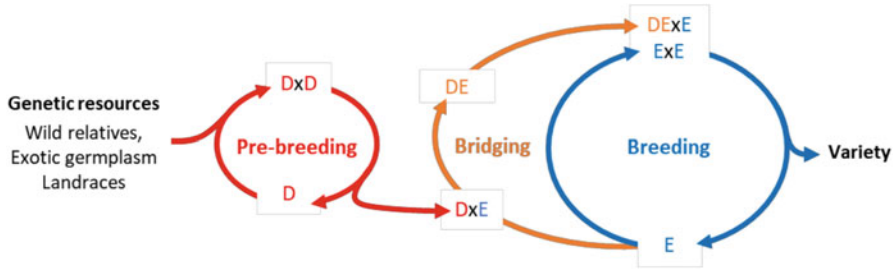


Fig. 7 Diagram illustrating the respective positioning of pre-breeding, bridging and breeding from genetic resources to variety release. Reproduced from Allier et al. (2020)

(2020) compared simulated breeding programs introducing donors with different performance levels. They observed that with recurrent introductions of improved donors, it is possible to maintain the genetic diversity and increase mid- and long-term performances with only limited penalty at short-term. When donors are already high-yielding, the bridging step could be skipped (Fig. 7).

From a practical point of view, several open-source software have been proposed. The R packages Rqtl (Broman et al. 2003), Popvar (Mohammadi et al. 2015), and software Alphasim (Faux et al. 2016) simulate bi-parental populations. The R Package Breeding Scheme Language (Yabe et al. 2017) simulates breeding programs. Multi-stage breeding schemes for hybrids using economic constraints are implemented in the R package Selectiongain (Mi et al. 2014, 2016).

To optimize mating for multiple traits, the R Package Genomic mating (Akdemir and Isidro-Sánchez 2016) and the software Alphamate (Gorjanc and Hickey 2018) have been proposed.

Forward stochastic simulations are proposed in python language in the software SeqBreed (Pérez-Enciso et al. 2020) and MoBPS (Pook et al. 2020), the last one implementing the optimum contribution method in an R environment.

To estimate the probability of getting the best progeny out of N with a specific cross, we can use the R package EMBV (Müller 2017). For qualitative traits controlled by major genes, the probability to cumulate a maximum of favorable alleles can be optimized using the software Optimas (Valente et al. 2013) or PCV (Han et al. 2017).

4 Future Perspectives

4.1 Improvement of Databases

We discussed above how diagnostic markers and genomic predictions can help the introduction of GR beneficial alleles from landraces or wild relatives in breeding populations. Operating procedures for conservation of those accessions have been in

place for decades in genebanks, but there is a lack of means and methodological results to optimize the discovery and transfer of beneficial alleles into modern varieties, especially for quantitative traits or multi-trait improvement (Mascher et al. 2019). What is essential to valorize those accessions is the existence of international databases with curated and standardized information (e.g., passport, curated phenotypes, validated GEBV, alleles at validated QTLs, introgressions, cloned genes, and site under ancient or recent selection pressure). There is actually no doubt that the better the database, the better the predictions and the integration of useful information to users. Many initiatives emerged to build national databases (<https://www.ars-grin.gov> GRIN-Global in the USA). Some national genebanks connect their database to regional (The European Search Catalogue for Plant Genetic Resources: EURISCO, <https://eurisco.ipk-gatersleben.de>) and international networks, such as the Global Gateway to Genetic Resources (Genesys, <https://www.genesys-pgr.org>). But not much information is shared beside the passport data. It is not straightforward to standardize experimental protocols, file formats and merge different databases. But this effort would facilitate integration of information and exchange of seeds among genebanks, plant geneticists, and breeders.

For plant phenotypic data management, the number of national initiatives multiplies for many species (Adam-Blondon et al. 2016), in particular in the phenomics context (Neveu et al. 2019). We can also cite the dataverse phenotypic database for CIMMYT wheat and maize trials (www.cimmyt.org/resources/data/). A multi-species integrative information system dedicated to plant and fungi pests called GNPIS has been developed in France, for instance (Pommier et al. 2019). It bridges genetic and genomic data, allowing researchers' access to both genetic information (e.g., genetic maps, quantitative trait loci, association genetics, markers, polymorphisms, germplasms, phenotypes and genotypes) and genomic data (e.g., genome sequences, physical maps, genome annotation and expression data). For genomic data and genome sequences in particular, transplant is an EU-funded project aiming at building hardware, software, and data infrastructure (Spannagl et al. 2016).

On the plant pathogen side, monitoring is generally organized at the national scale. The Australian cereal rust control program is estimated to save the industry \$289 million per year from resistance breeding, for instance. The European project Rustwatch (H2020 Sustainable Food Security-2017) tends to gather and standardize information about wheat cultivation surfaces, rust pressure, pathogen races, allelic composition of varieties and their bypass dates, in a standardized database to better understand the dynamic of bypass.

On the breeder side, from a pedigree and phenotype database in the UK, Fradgley et al. (2019) evaluated historical parental contributions in wheat and detected adaptation and selection signatures comparing genetic diversity levels with or without selection (experimental data vs simulated data, respectively) using gene dropping. Similar databases exist for oats, *Avena sativa* (Tinker and Deyl 2005) and rice, for instance (Bruskiewich et al. 2003).

An interesting initiative from NIAB is to propose a Toolbox to wheat breeders including evaluated wheat material introgressed with wild relatives (synthetic lines) (<https://triticeatoolbox.org>).

The university of California Davis (UC Davis) proposes a list of public wheat diagnostic markers online (MASwheat <https://maswheat.ucdavis.edu>).

For genomic selection, a project has started called Genomic Open-source Breeding informatics initiative (GOBii: <http://gobiiproject.org/>), funded by the Bill & Melinda Gates Foundation. The objective is to develop open-source data management, marker- and genomic-assisted breeding tools (PrAPI), for under-resourced breeding programs in particular, including trainings and workshops around the world (Selby et al. 2019).

The DivSeek project in the USA tends to bridge the gap between information requirements of genebank curators, plant breeders, and more targeted upstream biological researchers. They built a cooperative information platform for phenomics and genomics and gather a collaborative network of genebanks, breeders, scientists, database and computational experts for metadata curation. The objective is to share methodologies, open-source software and best practices related to genetic resources. For maize, the Seed project established a breeder's core of 4,000 landrace accessions that were genotyped and phenotyped, including testcross performance (<http://seedsofdiscovery.org>). For wheat, the Heat and Drought Wheat Improvement.

Consortium (HeDWIC, <http://www.hedwic.org/>) coordinated by CIMMYT aims at boosting heat and drought breeding using genomic and phenomic tools.

Then it is a long-term joint research goal to organize the conversion of information from population genomics and quantitative genetics to the development of some useful material for breeders. And public research may play an essential role in this activity, providing that means and findings are sufficient.

4.2 Integration of Omics to Better Decipher Genome/Phenome Relationship

Elite varieties have mainly been selected for production and post-harvest qualities with less attention to other features such as drought tolerance, nutrient use efficiency or durable pest and disease resistance. The effects of these factors have been mitigated by the use of treatments such as irrigation, fertilizers, and pesticides. Now that governments promote a more sustainable agriculture, breeding for stress tolerance may become common rules once the tools and methodologies are available. A better understanding of ecophysiological and expression determinants is essential to breed for stress tolerance. However large-scale phenotyping of physiological traits and generating data for population genomics and other “omics” aspects, for many varieties in different conditions with biological replicates, is still not affordable. But costs are likely to drop soon (Zivy et al. 2015).

4.2.1 Sequencing Fragments with Known DNA Patterns (Target Candidates)

Instead of sequencing the whole genome of accessions, we can target exome or specific domains such as LRR that are typical of resistance genes. Jupe et al. (2013), using Resistance gene enrichment Sequencing (RenSeq), reannotated the NB-LRR gene family and rapidly mapped resistance loci in segregating populations from hexaploid bread wheat. Arora et al. (2019), using R gene enrichment sequencing, a sequence capture bait library optimized for *Ae. tauschii* NLR domains and k-mer based association genetics (AgRenSeq) on a diverse panel (195 *Ae. tauschii* accessions), rapidly cloned four rust genes (*Sr33*, *Sr46*, *Sr45*, *SrTA1662*). Using mutagenesis coupled with exome capture and NLR-baits (MutRenSeq), Steuernagel et al. (2016) rapidly cloned *Sr22* and *Sr45* genes.

4.2.2 Population Transcriptomics

With the availability of Next Generation Sequencing (NGS) technologies, the possibility to directly sequence mRNA at relatively reduced cost becomes available.

Genomic predictions using whole-genome SNPs or GWAS are limited in capturing epistasis. Because mRNA, small RNA (sRNA) sequences and metabolic data are involved in transcriptional, translational, and post-translational processes, we expect them to provide such information. For instance, GWAS on transcripts allowed detecting candidate genes controlling oil content in maize, and their sequencing to detect polymorphisms and favorable alleles (Li et al. 2013). In grain maize, they evaluated the ability of this kind of data in parental lines to predict the performance of untested hybrids. They found that mRNA data are a superior predictor for grain yield and whole-genome SNP data for grain dry matter content, while sRNA performed relatively poorly for both traits. Combining mRNA and genomic data as predictors resulted in high predictive abilities across both traits and could contribute to more efficient selection of hybrid candidates in maize (Schrag et al. 2018).

RNA sequences can differentiate between isoforms of a gene family, a widespread phenomenon in complex crop genomes, which is difficult using DNA sequences. For example, in wheat, Oono et al. (2013) discovered this way phosphate starvation-responsive genes. Ramírez-González et al. (2018) showed differential expression of homoeolog genes due to epigenetic modifications and variation in transposable elements within promoters. The measurement of tissue and stress-specific co-expression networks throughout the development allows reconstructing regulatory networks. Some kernel component candidates were found using this strategy (Wen et al. 2016).

4.2.3 Population Proteomics

Carpentier et al. (2011) identified protein polymorphisms correlated to drought tolerance using shotgun approaches in banana and Grimaud et al. (2013) found cold-acclimation-related proteins in pea. Virilouvet et al. (2011) identified the *ZmASR1* gene under an abundance proteins QTL (pQTLs), candidate for drought tolerance in maize. The same gene was also associated in tomato, grape, lily, and banana (Maskin et al. 2001; Çakir et al. 2003; Wang et al. 2005; Henry et al. 2011).

4.2.4 Population Metabolomics: Phenotypes Targeting Candidate Metabolic Pathways

Metabolomics can detect targeted primary (sugars, organic- and amino-acids. . .) and secondary metabolites (photosynthates necessary to biomass formation, flavonoids, sugar-phosphates, phytohormones, phytoalexins) without genome sequence information. But it is not yet possible to work on the entire metabolome. Doerfler et al. (2014) detected 15 metabolites QTLs (mQTLs) of the flavonoid-pathway for cold and light stress in *Arabidopsis thaliana*. Pathogen induced markers were identified for *Rhizoctonia solani* in potatoes (Aliferis and Jabaji 2012), fungal pathogens in soybean (Aliferis et al. 2014), and bacterial blight-resistance in rice (Wu et al. 2012). An aroma (mesifurane) candidate gene was detected in strawberry, *Fragaria x ananassa* (Zorrilla-Fontanesi et al. 2012). The use of metabolomics in breeding has been reviewed in Fernandez et al. (2021).

4.2.5 Population Epigenomics

Epigenomic variations are involved in the control of plant developmental processes and shaping phenotypic plasticity to the environment (Gallusci et al. 2017; Moler et al. 2019). The elucidation of epigenetic regulatory networks using DNA methylation information should improve crop models. For instance, we can predict lycopene accumulation during tomato fruit ripening (Liu et al. 2015b), anthocyanin accumulation in apple (El-Sharkawy et al. 2015), energy-use efficiency in canola lines (Hauben et al. 2009).

Concerning histone marks, as they are likely to be erased following meiosis, they are of little interest to breeding applications in sexually propagated crops. But they can be relevant for clonally propagated crops, for pathogen resistance, for instance (Jaskiewicz et al. 2011).

It is well known that DNA mutation, copy number variants or methylation, in genes, promoters or regulatory regions can affect gene expression, which modifies phenotypes in different environmental contexts. Many studies also showed that re-arrangements of loci on chromosomes, inversions, insertions of transposable elements, deletions can also lead to gene silencing. All those types of

polymorphisms/annotation could help improving genomic prediction models. Molecular markers at the vicinity of genes actually tend to link more to causal variants in maize (reference). QTL effects are higher in genic regions (Wallace et al. 2014), which is consistent with the fact that a large portion of variability of gene expression is attributed to cis polymorphisms in maize (Schadt et al. 2003). Taking into account the proximity of molecular markers to genes actually improves prediction of agronomic traits in diverse populations of hybrid maize (Ramstein et al. 2020).

To facilitate and optimize those models, we still need the development of generalized methods that integrate multiple data types.

4.2.6 Integration of Different Population “Omics” Information

The long-term objective is to be able to integrate all possible “omics” information on the same samples. We will be able to detect eQTLs, pQTLs, and mQTLs and look for co-localization with molecular marker-based QTLs (cis-QTLs), giving direct access to the genes, favorable alleles, and regulatory factors outside of the gene (trans-QTL). As skills are spread in different groups, a European network named COST project was organized to help building regulation networks from integrated databases. To make it useful to breeders, the first objective is to define traits of interest for specific climatic zones or constraints.

Then, cellular phenotyping (transcriptome/proteome/metabolome) will help building more realistic models to predict phenome in the field. Models taking into account non-additive effects, nonlinear relationships between enzyme concentrations and metabolic fluxes (Fiévet et al. 2010; Vacher and Small 2019) could actually explain even more genetic variance and improve predictions.

5 Conclusion

Integration of concepts and tools of population genomics and quantitative genetics can lead to a better valorization of genetic diversity in crop (pre)breeding programs.

Advances in population genomics offer a new dimension to quantitative genetics in the form of increasing data on genetic diversity and structure, identification of new candidate genes of agronomic interest associated with signatures of selection, associations with environmental covariates and phenotypes, and prediction of genetic values of various plant genetic resources.

Genomic predictions can detect germplasm of interest in genebanks without the need of phenotyping if the calibration population is relevant and the quality of phenotyping is satisfactory. Good quality phenotyping will actually always be a cornerstone to efficient plant breeding and predictions. Genomic predictions can help to optimize the time and cost of the breeding process, allowing a transfer of budget to test a larger number of genitors and crosses. It can accelerate recurrent

selection to produce pre-breeding and breeding lines that contain new favorable alleles. It can predict optimum parental contribution and mating in (pre)breeding programs to optimize short-term genetic gain but also assure long-term genetic gain by constraining germplasm diversity. Currently, the main methodological challenge here is a good estimation of marker effects and progeny variance.

Increasingly detailed multi-omic characterization of genetic resources (through genomics, transcriptomics, methylomics, proteomics) is expected to help understand and predict the genome-phenome relationship, and ultimately design ideotypes for particular growth conditions and uses. The hope is that additional layers of omics data will improve estimation of marker by environment effect. Currently, several technical hurdles are preventing industrial implementation of multi-omics approaches in the breeding process. On the fundamental level, effects of epigenetic variation on gene expression – on the background of nucleotide variation – are still difficult to detect, quantify, and generalize. Also, it remains to be seen whether genotype is a good predictor of methylome, transcriptome, and metabolome, i.e. whether training sets characterized with multi-omic data can improve genomic prediction of candidates that have been genotyped with SNPs, giving higher weights in prediction models to QTLs. Moreover, multi-omics approaches in the next generation of genomic prediction can only come with increased analytical complexity and cost. Nonetheless, recent years have witnessed an emergence and proliferation of methods designed for multi-omic data integration and analysis, and with the continuous drop of sequencing costs, multi-omics crop research will attract significant efforts in the immediate future. With a combination of multi-omic, agronomic, phenological and physiological data, supplemented with precise environment characterization (weather, soils, crop management) and targeted trialing, we are set on the path to decipher the complex GxE interactions and predict the performance of existing and new varieties in current and future environments.

For practical applications, it is necessary to integrate population genomics and other “omics” information with phenotypes in common public databases, so that robust methodologies and decision tools could be developed to convert this information into feasible protocols. In that context, one role of public research could be to develop and disseminate databases, new methodologies, and produce decision tools that could be validated by breeders in interactive projects. Public research could also coordinate the design, production, and evaluation of ready-to-use crop plant resources, pre-breeding genitors in particular.

References

- Adam-Blondon A-F, Alaux M, Pommier C, Cantu D, Cheng Z-M, Cramer GR, et al. Towards an open grapevine information system. *Hortic Res.* 2016;3(1):1–8.
- Akakpo R, Scarcelli N, Dansi A, Djedatin G, Thuillet A-C, Rhoné B, et al. Molecular basis of African yam domestication: analyses of selection point to root development, starch biosynthesis, and photosynthesis related genes. *BMC Genomics.* 2017;18(1):782.
- Akdemir D, Isidro-Sánchez J. Efficient breeding by genomic mating. *Front Genet.* 2016;7:210.

- Akdemir D, Isidro-Sánchez J. Design of training populations for selective phenotyping in genomic prediction. *Sci Rep.* 2019;9(1):1–15.
- Akdemir D, Beavis W, Fritsche-Neto R, Singh AK, Isidro-Sánchez J. Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity.* 2018;122(5):672–83.
- Alachiotis N, Pavlidis P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun Biol.* 2018;1(1):1–11.
- Alachiotis N, Stamatakis A, Pavlidis P. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics.* 2012;28(17):2274–5.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–64.
- Aliferis KA, Jabaji S. FT-ICR/MS and GC-EI/MS metabolomics networking unravels global potato sprout's responses to *Rhizoctonia solani* infection. *PLoS One.* 2012;7(8)
- Aliferis KA, Faubert D, Jabaji S. A metabolic profiling strategy for the dissection of plant defense against fungal pathogens. *PLoS One.* 2014;9(11)
- Allaby RG, Ware RL, Kistler L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol Appl.* 2019;12(1):29–37.
- Allard RW. Principles of plant breeding: Wiley; 1999.
- Allier A, Lehermeier C, Charcosset A, Moreau L, Teyssèdre S. Improving short and long term genetic gain by accounting for within family variance in optimal cross selection. *Front Genet.* 2019a;10:1006.
- Allier A, Moreau L, Charcosset A, Teyssèdre S, Lehermeier C. Usefulness criterion and post-selection parental contributions in multi-parental crosses: application to polygenic trait introgression. *G3: Genes, Genomes, Genetics.* 2019b;9(5):1469–79.
- Allier A, Teyssèdre S, Lehermeier C, Charcosset A, Moreau L. Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. *Theor Appl Genet.* 2019c:1–15.
- Allier A, Teyssèdre S, Lehermeier C, Claustres B, Maltese S, Melkior S, et al. Assessment of breeding programs sustainability: application of phenotypic and genomic indicators to a North European grain maize program. *Theor Appl Genet.* 2019d;132(5):1321–34.
- Allier A, Teyssèdre S, Lehermeier C, Moreau L, Charcosset A. Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics.* 2020;21:1–16.
- Amani I, Fischer RA, Reynolds MP. Canopy temperature depression association with yield of irrigated spring wheat cultivars in a hot climate. *J Agron Crop Sci.* 1996;176(2):119–29.
- Anderson E. Introgressive hybridization. *Biol Rev.* 1953;28(3):280–307.
- Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics.* 2002;160(3):1217–29.
- Andres RJ, Dunne JC, Samayoa LF, Holland JB. Enhancing crop breeding using population genomics approaches. In: *Population genomics.* Cham: Springer; 2020. p. 1–45.
- Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics.* 2008;9(1):323.
- Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, et al. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat Biotechnol.* 2019;37(2):139–43.
- Bailey-Serres J, Fukao T, Ronald P, Ismail A, Heuer S, Mackill D. Submergence tolerant rice: SUB1's journey from landrace to modern cultivar. *Rice.* 2010;3(2):138–47.
- Balfourier F, Bouchet S, Robert S, De Oliveira R, Rimbart H, Kitt J, et al. Worldwide phylogeography and history of wheat genetic diversity. *Sci Adv.* 2019;5(5):eaav0536.
- Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, et al. Landscape genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP, editor. *Population genomics: concepts, approaches and applications.* Cham: Springer Nature Switzerland AG; 2019. p. 261–322.

- Bandeirae Sousa M, Cuevas J, de Oliveira Couto EG, Pérez-Rodríguez P, Jarquín D, Fritsche-Neto R, et al. Genomic-enabled prediction in maize using kernel models with genotype x environment interaction. *G3: Genes, Genomes, Genetics*. 2017;7(6):1995–2014.
- Bao Y, Kurlle JE, Anderson G, Young ND. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Mol Breed*. 2015;35(6):128.
- Barakat A, Yassin NBM, Park JS, Choi A, Herr J, Carlson JE. Comparative and phylogenomic analyses of cinnamoyl-CoA reductase and cinnamoyl-CoA-reductase-like gene family in land plants. *Plant Sci*. 2011;181(3):249–57.
- Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, et al. Intraspecific variation of recombination rate in maize. *Genome Biol*. 2013;14(9):R103.
- Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. *Proc Biol Sci*. 1996;263(1377):1619–26.
- Beavis WD. QTL analyses: power, precision, and accuracy. *Molecular dissection of complex traits*. Boca Raton: CRC Press; 1998. p. 145–62.
- Beavis W, Smith O, Grant D, Fincher R. Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop Sci*. 1994;34(4):882–96.
- Bellis ES, Kelly EA, Lorts CM, Gao H, DeLeo VL, Rouhan G, et al. Genomics of sorghum local adaptation to a parasitic plant. *Proc Natl Acad Sci*. 2020;117(8):4243–51.
- Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al. Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell*. 2014;26(5):1901–12.
- Ben-Sadoun S, Rincint R, Auzanneau J, Oury FX, Rolland B, Heumez E, et al. Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor Appl Genet*. 2020;133:2197–212.
- Bernardo R. Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Sci*. 2009;49(2):419–25.
- Bernardo R. Genomewide selection of parental inbreds: classes of loci and virtual biparental populations. *Crop Sci*. 2014;54(6):2586–95.
- Bernardo R, Charcosset A. Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. *Crop Sci*. 2006;46(2):614–21.
- Bernardo R, Yu J. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci*. 2007;47(3):1082–90.
- Berthouly-Salazar C, Thuillet A-C, Rhoné B, Mariac C, Ousseini IS, Couderc M, et al. Genome scan reveals selection acting on genes linked to stress response in wild pearl millet. *Mol Ecol*. 2016;25(21):5500–12.
- Beukelaer HD, Meyer GD, Fack V. Heuristic exploitation of genetic structure in marker-assisted gene pyramiding problems. *BMC Genet*. 2015;16.
- Bijma P, Woolliams JA. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. *Genetics*. 1999;151(3):1197–210.
- Bijma P, Wientjes YC, Calus MP. Breeding top genotypes and accelerating response to recurrent selection by selecting parents with greater Gametic variance. *Genetics*. 2020;214(1):91–107.
- Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, et al. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. *J Exp Bot*. 2014;65(20):5849–65.
- Bohn M, Utz HF, Melchinger AE. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sci*. 1999;39(1):228–37.
- Bonhomme M, Chevalat C, Servin B, Boitard S, Abdallah J, Blott S, et al. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*. 2010;186(1):241–62.
- Bonk S, Reichelt M, Teuscher F, Segelke D, Reinsch N. Mendelian sampling covariability of marker effects and genetic values. *Genet Sel Evol*. 2016;48(1):36.

- Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, et al. Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics*. 2017;206(2):573–85.
- Brancourt-Hulmel M. Crop diagnosis and probe genotypes for interpreting genotype environment interaction in winter wheat trials. *Theor Appl Genet*. 1999;99(6):1018–30.
- Brancourt-Hulmel M, Denis JB, Lecomte C. Determining environmental covariates which explain genotype environment interaction in winter wheat through probe genotypes and biadditive factorial regression. *Theor Appl Genet*. 2000;100(2):285–98.
- Brard S, Ricard A. Is the use of formulae a reliable way to predict the accuracy of genomic selection? *J Anim Breed Genet*. 2015;132(3):207–17.
- Brauner PC, Müller D, Schopp P, Böhm J, Bauer E, Schön C-C, et al. Genomic prediction within and among doubled-haploid libraries from maize landraces. *Genetics*. 2018;210(4):1185–96.
- Brauner PC, Schipprack W, Utz HF, Bauer E, Mayer M, Schön C-C, et al. Testcross performance of doubled haploid lines from European flint maize landraces is promising for broadening the genetic base of elite germplasm. *Theor Appl Genet*. 2019;132(6):1897–908.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*. 1995;140(2):783–96.
- Brisbane JR, Gibson JP. Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Theor Appl Genet*. 1995;91(3):421–31.
- Brisson N, Gate P, Gouache D, Charmet G, Oury FX, Huard F. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crop Res*. 2010;119(1):201–12.
- Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 2003;19(7):889–90.
- Brown AHD, Clegg MT. Isozyme assessment of plant genetic resources. *Isozymes Curr Top Biol Med Res*. 1983;11:285–95.
- Bruce RW, Torkamaneh D, Grainger C, Belzile F, Eskandari M, Rajcan I. Genome-wide genetic diversity is maintained through decades of soybean breeding in Canada. *Theor Appl Genet*. 2019;132(11):3089–100.
- Bruskiewich RM, Cosico AB, Eusebio W, Portugal AM, Ramos LM, Reyes MT, et al. Linking genotype to phenotype: the international rice information system (IRIS). *Bioinformatics*. 2003;19(suppl_1):i63–5.
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. *Science*. 2009;325(5941):714–8.
- Bulmer MG. The effect of selection on genetic variability. *Am Nat*. 1971;105(943):201.
- Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, Billot C, et al. Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Front Plant Sci*. 2019;10
- 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.
- Bustos-Korts D, Malosetti M, Chapman S, Biddulph B, van Eeuwijk F. Improvement of predictive ability by uniform coverage of the target genetic space. *G3: Genes, Genomes, Genetics*. 2016;6(11):3733.
- Çakir B, Agasse A, Gaillard C, Saumonneau A, Delrot S, Atanassova R. A grape ASR protein involved in sugar and abscisic acid signaling. *Plant Cell*. 2003;15(9):2165–80.
- Calus MP, Veerkamp RF. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol*. 2011;43(1):1.
- Canzar S, El-Kebir M. A mathematical programming approach to marker-assisted gene pyramiding. In: *International workshop on algorithms in bioinformatics*. Berlin: Springer; 2011. p. 26–38.
- Carpentier SC, Panis B, Renaut J, Samyn B, Vertommen A, Vanhove A-C, et al. The use of 2D-electrophoresis and de novo sequencing to characterize inter- and intra-cultivar protein polymorphisms in an allopolyploid crop. *Phytochemistry*. 2011;72(10):1243–50.
- Cavanagh C, Morell M, Mackay I, Powell W. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol*. 2008;11(2):215–21.

- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci*. 2013;110(20):8057–62.
- Charmet G, Robert N, Perretant M, Gay G, Sourdille P, Groos C, et al. Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theor Appl Genet*. 1999;99(7):1143–8.
- Chen J, Ding J, Ouyang Y, Du H, Yang J, Cheng K, et al. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice. *Proc Natl Acad Sci*. 2008;105(32):11436–41.
- Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res*. 2010;20(3):393–402.
- Cheng H, Liu J, Wen J, Nie X, Xu L, Chen N, et al. Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biol*. 2019;20(1):136.
- Chenu K, Chapman SC, Hammer GL, Mclean G, Salah HBH, Tardieu F. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. *Plant Cell Environ*. 2008;31(3):378–91.
- Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC. Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J Exp Bot*. 2011;62(6):1743–55.
- Choi K, Henderson IR. Meiotic recombination hotspots – a comparative view. *Plant J*. 2015;83(1):52–61.
- Christopher J, Richard C, Chenu K, Christopher M, Borrell A, Hickey L. Integrating rapid phenotyping and speed breeding to improve stay-green and root adaptation of wheat in changing, water-limited, Australian environments. *Procedia Environ Sci*. 2015;29:175–6.
- Civán P, Brown TA. Role of genetic introgression during the evolution of cultivated rice (*Oryza sativa* L.). *BMC Evol Biol*. 2018;18(1):57.
- Civán P, Ali S, Batista-Navarro R, Drosou K, Ihejiro C, Chakraborty D, et al. Origin of the aromatic group of cultivated rice (*Oryza sativa* L.) traced to the Indian subcontinent. *Genome Biol Evol*. 2019;11(3):832–43.
- Cole JB, VanRaden PM. Use of haplotypes to estimate Mendelian sampling effects and selection limits. *J Anim Breed Genet*. 2011;128(6):446–55.
- Collevatti RG, dos Santos JS, Rosa FF, Amaral TS, Chaves LJ, Ribeiro MC. Multi-scale landscape influences on genetic diversity and adaptive traits in a neotropical savanna tree. *Front Genet*. 2020;11:259.
- Combs E, Bernardo R. Genomewide selection to introgress semidwarf maize germplasm into US Corn Belt inbreds. *Crop Sci*. 2013;53(4):1427–36.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. *Genetics*. 2010;185(4):1411–23.
- Cooper HD, Spillane C, Hodgkin T, Cooper H. Broadening the genetic base of crops: an overview. In: *Broadening the genetic base of crop production*. New York: CABI; 2001. p. 1–23.
- Cooper M, Technow F, Messina C, Gho C, Totir LR. Use of crop growth models with whole-genome prediction: application to a maize multi-environment trial. *Crop Sci*. 2016;56(5):2141–56.
- Cowling WA, Li L, Siddique KH, Henryon M, Berg P, Banks RG, et al. Evolving gene banks: improving diverse populations of crop and exotic germplasm with optimal contribution selection. *J Exp Bot*. 2017;68(8):1927–39.
- Crain J, Mondal S, Rutkoski J, Singh RP, Poland J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome*. 2018;11(1)
- Crossa J, Campos GDL, Perez P, Gianola D, Burgueno J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*. 2010;186(2):713–24.

- Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity*. 2014;112(1):48–60.
- Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, Saint-Pierre C, et al. Genomic prediction of gene bank wheat landraces. *G3: Genes, Genomes, Genetics*. 2016;6(7):1819–34.
- Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol*. 2014;23(13):3133–57.
- Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P, de los Campos G. Bayesian genomic prediction with genotype x environment interaction kernel models. *G3: Genes, Genomes, Genetics*. 2017;7(1):41–53.
- Cuevas J, Granato I, Fritsche-Neto R, Montesinos-Lopez OA, Burgueño J, e Sousa MB, et al. Genomic-enabled prediction Kernel models with random intercepts for multi-environment trials. *G3: Genes, Genomes, Genetics*. 2018;8(4):1347–65.
- Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3(10)
- Daetwyler HD, Hayden MJ, Spangenberg GC, Hayes BJ. Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics*. 2015;200(4):1341–8.
- Danguy des Deserts A, Bouchet S, Sourdille P, Servin B. Evolution of recombination landscapes in diverging populations of bread wheat. *bioRxiv*. 2021;13(8):evab152.
- De Beukelaer H, Badke Y, Fack V, De Meyer G. Moving beyond managing realized genomic relationship in long-term genomic selection. *Genetics*. 2017;206(2):1127–38.
- De Coninck A, De Baets B, Kourounis D, Verbosio F, Schenk O, Maenhout S, et al. Needles: toward large-scale genomic prediction with marker-by-environment interaction. *Genetics*. 2016;203(1):543–55.
- Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L. Past and future use of wild relatives in crop breeding. *Crop Sci*. 2017;57(3):1070–82.
- Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, et al. Increase in crop losses to insect pests in a warming climate. *Science*. 2018;361(6405):916–9.
- Dias-Alves T, Mairal J, Blum MGB. Loter: a software package to infer local ancestry for a wide range of species. *Mol Biol Evol*. 2018;35(9):2318–26.
- Doerfler H, Sun X, Wang L, Engelmeier D, Lyon D, Weckwerth W. mzGroupAnalyzer-predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PLoS One*. 2014;9(5)
- dos Santos JPR, de Castro Vasconcellos RC, Pires LPM, Balestre M, Von Pinho RG. Inclusion of dominance effects in the multivariate GBLUP model. *PLoS One*. 2016;11(4)
- Doyle JJ. 5S ribosomal gene variation in the soybean and its progenitor. *Theor Appl Genet*. 1988;75(4):621–4.
- Dudley JW. Theory for identification and use of exotic germplasm in maize breeding programs. *Maydica*. 1984;29:391–407.
- Dudley JW. Evaluation of maize populations as sources of favorable alleles. *Crop Sci*. 1988;28(3):486–91.
- Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. 2011;28(8):2239–52.
- Ellis JG, Lagudah ES, Spielmeier W, Dodds PN. The past, present and future of breeding rust resistant wheat. *Front Plant Sci*. 2014;5:641.
- El-Sharkawy I, Liang D, Xu K. Transcriptome analysis of an apple (*Malus x domestica*) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. *J Exp Bot*. 2015;66(22):7359–76.
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, et al. Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci*. 2014;54(1):48–59.
- Excoffier L, Lischer H. Arlequin ver 3.5 user manual; an integrated software package for population genetics data analysis. Swiss Institute of Bioinformatics. 2009.

- Falconer DS, Mackay TF, Frankham R. Introduction to quantitative genetics (4th edn). Trends Genet. 1996;12(7):280.
- Fariello MI, Boitard S, Naya H, San Cristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics. 2013;193(3):929–41.
- Faux A-M, Gorjanc G, Gaynor RC, Battagin M, Edwards SM, Wilson DL, et al. AlphaSim: software for breeding program simulation. Plant Genome. 2016;9(3):1–14.
- Feng L, Sebastian S, Smith S, Cooper M. Temporal trends in SSR allele frequencies associated with long-term selection for yield of maize. Maydica. 2006;51(2):293.
- Fernandes SB, Dias KO, Ferreira DF, Brown PJ. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet. 2018;131(3):747–55.
- Fernandez O, Millet EJ, Rincet R, Prigent S, Pétriacq P, Gibon Y. Chapter seven – plant metabolomics and breeding. In: Pétriacq P, Bouchereau A, editors. Plant metabolomics in full swing, Advances in botanical research, vol. 98. Cambridge: Academic Press; 2021. p. 207–35.
- Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 2014;31(5):1275–91.
- Fiévet JB, Dillmann C, de Vienne D. Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. Theor Appl Genet. 2010;120(2):463.
- Fisher SRA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52:399–433.
- Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930. 272 p.
- Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol Lett. 2015;18(1):1–16.
- Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180(2):977–93.
- Foncéca D, Hodo-Abalo T, Rivallan R, Faye I, Sall MN, Ndoeye O, et al. Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biol. 2009;9(1):103.
- Fonceca D, Tossim H-A, Rivallan R, Vignes H, Lacut E, de Bellis F, et al. Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. PLoS One. 2012;7(11):e48642.
- Fradgley N, Gardner KA, Cockram J, Elderfield J, Hickey JM, Howell P, et al. A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. PLoS Biol. 2019;17(2):e3000071.
- Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30(7):1687–99.
- Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014;196(4):973–83.
- Fu Y-B. Impact of plant breeding on genetic diversity of agricultural crops: searching for molecular evidence. Plant Genet Resour. 2006;4(1):71–8.
- Fu Y-B. Understanding crop genetic diversity under modern plant breeding. Theor Appl Genet. 2015;128(11):2131–42.
- Fu J, Falke KC, Thiemann A, Schrag TA, Melchinger AE, Scholten S, et al. Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. Theor Appl Genet. 2012;124(5):825–33.
- Fustier M-A, Martínez-Ainsworth NE, Aguirre-Liguori JA, Venon A, Corti H, Rousselet A, et al. Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude. PLoS Genet. 2019;15(12):e1008512.

- Gallusci P, Dai Z, Génard M, Gauffretau A, Leblanc-Fournier N, Richard-Molard C, et al. Epigenetics for plant improvement: current knowledge and modeling avenues. *Trends Plant Sci.* 2017;22(7):610–23.
- Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y, et al. ALK, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. *J Integr Plant Biol.* 2011;53(9):756–65.
- 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.
- Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, et al. A two-part strategy for using genomic selection to develop inbred lines. *Crop Sci.* 2017;57(5):2372–86.
- Gerke JP, Edwards JW, Guill KE, Ross-Ibarra J, McMullen MD. The genomic impacts of drift and selection for hybrid performance in maize. *Genetics.* 2015;201(3):1201–11.
- Ghosh S, Watson A, Gonzalez-Navarro OE, Ramirez-Gonzalez RH, Yanes L, Mendoza-Suárez M, et al. Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nat Protoc.* 2018;13(12):2944–63.
- Glaszmann JC. Isozymes and classification of Asian rice varieties. *Theoret Appl Genetics.* 1987;74(1):21–30.
- Glaszmann J, Kilian B, Upadhyaya H, Varshney R. Accessing genetic diversity for crop improvement. *Curr Opin Plant Biol.* 2010;13(2):167–73.
- Glémin S, Bataillon T. A comparative view of the evolution of grasses under domestication. *New Phytol.* 2009;183(2):273–90.
- Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica.* 2009;136(2):245–57.
- Goiffon M, Kusmec A, Wang L, Hu G, Schnable PS. Improving response in genomic selection with a population-based selection strategy: optimal population value selection. *Genetics.* 2017;206(3):1675.
- Goldberg DE, Holland JH. Genetic algorithms and machine learning. *Mach Learn.* 1988;3(2):95–9.
- Gompert Z, Egan SP, Barrett RD, Feder JL, Nosil P. Multilocus approaches for the measurement of selection on correlated genetic loci. *Mol Ecol.* 2017;26(1):365–82.
- Goodman MM. Broadening the genetic diversity in maize breeding by use of exotic germplasm. In: *Genetics and exploitation of heterosis in crops, ASA, CSSA, and SSSA books; 1999.* p. 139–48.
- Goodman MM. Broadening the US maize germplasm base. *Maydica.* 2005;50(3/4):203.
- Goodman MM, Moreno J, Castillo F, Holley RN, Carson ML. Using tropical maize germplasm for temperate breeding. *Maydica.* 2000;45(3):221–34.
- Gorjanc G, Hickey JM. AlphaMate: a program for optimizing selection, maintenance of diversity and mate allocation in breeding programs. *Bioinformatics.* 2018;34(19):3408–11.
- Gorjanc G, Jenko J, Hearne SJ, Hickey JM. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics.* 2016;17(1):30.
- Gorjanc G, Gaynor RC, Hickey JM. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet.* 2018;131(9):1953–66.
- Granato I, Cuevas J, Luna-Vázquez F, Crossa J, Montesinos-López O, Burgueño J, et al. BGGE: a new package for genomic-enabled prediction incorporating genotype × environment interaction models. *G3: Genes, Genomes, Genetics.* 2018;8(9):3039–47.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science.* 2010;328(5979):710–22.
- Grimaud F, Renaut J, Dumont E, Sergeant K, Lucau-Danila A, Blervacq A-S, et al. Exploring chloroplastic changes related to chilling and freezing tolerance during cold acclimation of pea (*Pisum sativum* L.). *J Proteomics.* 2013;80:145–59.
- Guarino L, Lobell DB. A walk on the wild side. *Nat Clim Change.* 2011;1(8):374–5.

- Guillot G, Vitalis R, le Rouzic A, Gautier M. Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spat Stat.* 2014;8:145–55.
- Günther T, Coop G. Robust identification of local adaptation from allele frequencies. *Genetics.* 2013;195(1):205–20.
- Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 2014;15(1):30.
- Gur A, Semel Y, Cahaner A, Zamir D. Real time QTL of complex phenotypes in tomato interspecific introgression lines. *Trends Plant Sci.* 2004;9(3):107–9.
- Habier D, Fernando R, Dekkers J. The impact of genetic relationship information on genome-assisted breeding values. *Genetics.* 2007;177(4):2389.
- Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol.* 2010;42(1):5.
- Hajjar R, Hodgkin T. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica.* 2007;156(1–2):1–13.
- Hallander J, Waldmann P. Optimization of selection contribution and mate allocations in monoeocious tree breeding populations. *BMC Genet.* 2009a;10(1):70.
- Hallander J, Waldmann P. Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theor Appl Genet.* 2009b;118(6):1133–42.
- Hallauer AR, Sears JH. Integrating exotic germplasm into corn belt maize breeding programs I. *Crop Sci.* 1972;12(2):203–6.
- Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, et al. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytol.* 2016;209(2):871–84.
- Han Y, Cameron JN, Wang L, Beavis WD. The predicted cross value for genetic introgression of multiple alleles. *Genetics.* 2017;205(4):1409–23.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science.* 2011;334(6052):83–6.
- Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci U S A.* 2017;114(46):E9999.
- Hauben M, Haesendonckx B, Standaert E, Kelen KVD, Azmi A, Akpo H, et al. Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proc Natl Acad Sci U S A.* 2009;106(47):20109–14.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, et al. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol.* 2007;24(7):1506–17.
- Hayashi T, Iwata H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics.* 2013;14(1):34.
- Hayes B, Bowman P, Chamberlain A, Goddard M. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci.* 2009;92(2):433.
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 2010;50(5):1681–90.
- Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 2011a;51(6):2597–606.
- Heffner EL, Jannink J-L, Sorrells ME. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome.* 2011b;4(1):65–75.
- Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics.* 1975;423–47.
- Henderson CR, Quaas RL. Multiple trait evaluation using relatives' records. *J Anim Sci.* 1976;43(6):1188–97.
- Henry IM, Carpentier SC, Pampurova S, Van Hoylandt A, Panis B, Swennen R, et al. Structure and regulation of the *Asr* gene family in banana. *Planta.* 2011;234(4):785.

- Heslot N, Feoktistov V. Optimization of selective phenotyping and population design for genomic prediction. *bioRxiv*. 2017;172064.
- Heslot N, Akdemir D, Sorrells ME, Jannink J-L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet*. 2014;127(2):463–80.
- Hickey LT, Germán SE, Pereyra SA, Diaz JE, Ziem LA, Fowler RA, et al. Speed breeding for multiple disease resistance in barley. *Euphytica*. 2017;213(3):64.
- Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat*. 2016;188(4):379–97.
- Holland JH. Outline for a logical theory of adaptive systems. *J ACM*. 1962;9(3):297–314.
- Hospital F, Charcosset A. Marker-assisted introgression of quantitative trait loci. *Genetics*. 1997;147(3):1469–85.
- Hospital F, Goldringer I, Openshaw S. Efficient marker-based recurrent selection for multiple quantitative trait loci. *Genet Res*. 2000;75(3):357–68.
- Huang L, Raats D, Sela H, Klymiuk V, Lidzbarsky G, Feng L, et al. Evolution and adaptation of wild emmer wheat populations to biotic and abiotic stresses. *Annu Rev Phytopathol*. 2016;54:279–301.
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 2012;44(7):808–11.
- Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenz MT, Ellstrand NC, Ross-Ibarra J. The genomic signature of crop-wild introgression in maize. *PLoS Genet*. 2013;9(5):e1003477.
- Hung HY, Browne C, Guill K, Coles N, Eller M, Garcia A, et al. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity*. 2012;108(5):490–9.
- Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci*. 2006;103(45):16666–71.
- Imai I, Kimball JA, Conway B, Yeater KM, McCouch SR, McClung A. Validation of yield-enhancing quantitative trait loci from a low-yielding wild ancestor of rice. *Mol Breeding*. 2013;32(1):101–20.
- Iwata H, Hayashi T, Terakami S, Takada N, Saito T, Yamamoto T. Genomic prediction of trait segregation in a progeny population: a case study of Japanese pear (*Pyrus pyrifolia*). *BMC Genet*. 2013;14(1):81.
- Jagadish SVK, Bahuguna RN, Djanaguiraman M, Gamuyao R, Prasad PVV, Craufurd PQ. Implications of high temperature and elevated CO₂ on flowering time in plants. *Front Plant Sci*. 2016;7
- James JW, McBride G. The spread of genes by natural and artificial selection in closed poultry flock. *J Genet*. 1958;56(1):55.
- Jannink J-L. Dynamics of long-term genomic selection. *Genet Sel Evol*. 2010;42(1):35.
- Jansen RC. Interval mapping of multiple quantitative trait loci. *Genetics*. 1993;135(1):205–11.
- Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet*. 2014;127(3):595–607.
- Jaskiewicz M, Conrath U, Peterhansel C. Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Rep*. 2011;12(1):50–5.
- Jia Y, Jannink J-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*. 2012;192(4):1513–22.
- Jiang J, Zhang Q, Ma L, Li J, Wang Z, Liu JF. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity*. 2015;115(1):29–36.
- Jombart T, Pontier D, Dufour A-B. Genetic markers in the playground of multivariate analysis. *Heredity*. 2009;102(4):330–41.
- Jordan D, Mace E, Cruickshank A, Hunt C, Hensell R. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Sci*. 2011;51(4):1444–57.

- Jupe F, Witek K, Verweij W, Śliwka J, Pritchard L, Etherington GJ, et al. Resistance gene enrichment sequencing (RESeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J*. 2013;76(3):530–44.
- Kemper KE, Bowman PJ, Pryce JE, Hayes BJ, Goddard ME. Long-term selection strategies for complex traits using high-density genetic markers. *J Dairy Sci*. 2012;95(8):4646–56.
- Kerr RJ, Goddard ME, Jarvis SF. Maximising genetic response in tree breeding with constraints on group coancestry. *Silvae Genet*. 1998;47(2):165–73.
- Kim Y, Nielsen R. Linkage disequilibrium as a signature of selective sweeps. *Genetics*. 2004;167(3):1513–24.
- Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*. 2002;160(2):765–77.
- Kinghorn BP. An algorithm for efficient constrained mate selection. *Genet Sel Evol*. 2011;43(1):4.
- Kinghorn BP, Banks R, Gondro C, Kremer VD, Meszaros SA, Newman S, et al. Strategies to exploit genetic variation while maintaining diversity. In: *Adaptation and fitness in animal populations*. Springer; 2009. p. 191–200.
- Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, et al. The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. *Crop Sci*. 2008;48(Supplement_1):S-12-S-26.
- Ko J-H, Prassinis C, Keathley D, Han K-H, Li C. Novel aspects of transcriptional regulation in the winter survival and maintenance mechanism of poplar. *Tree Physiol*. 2011;31(2):208–25.
- Kulathinal RJ, Stevison LS, Noor MAF. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet*. 2009;5(7).
- Kuruparth V, Sood S, Chhuneja P, Dhaliwal HS, Kaur S, Bowden RL, et al. A cryptic wheat–*Aegilops triuncialis* translocation with leaf rust resistance gene Lr58. *Crop Sci*. 2007;47(5):1995–2003.
- Labate JA, Lamkey KR, Lee M, Woodman WL. Temporal changes in allele frequencies in two reciprocally selected maize populations. *Theor Appl Genet*. 1999;99(7-8):1166–78.
- Lado B, Barrios PG, Quincke M, Silva P, Gutiérrez L. Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Sci*. 2016;56(5):2165–79.
- Lado B, Vázquez D, Quincke M, Silva P, Aguilar I, Gutiérrez L. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor Appl Genet*. 2018;131(12):2719–31.
- Laloë D. Precision and information in linear models of genetic evaluation. *Genet Sel Evol*. 1993;25(6):557–76.
- Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*. 1990;124(3):743–56.
- Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, et al. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv*. 2015;1(6).
- Lehermeier C, de Los Campos G, Wimmer V, Schön C-C. Genomic variance estimates: with or without disequilibrium covariances? *J Anim Breed Genet*. 2017a;134(3):232–41.
- Lehermeier C, Teyssière S, Schön C-C. Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. *Genetics*. 2017b;207(4):1651–61.
- Leung H, Raghavan C, Zhou B, Oliva R, Choi IR, Lacorte V, et al. Allele mining and enhanced genetic recombination for rice breeding. *Rice*. 2015;8(1):34.
- Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. 1973;74(1):175–95.
- Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*. 2013;45(1):43.

- Lian L, Jacobson A, Zhong S, Bernardo R. Prediction of genetic variance in biparental maize populations: genomewide marker effects versus mean genetic variance in prior populations. *Crop Sci.* 2015;55(3):1181–8.
- Lin Z, Cogan NO, Pemberton LW, Spangenberg GC, Forster JW, Hayes BJ, et al. Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. *Plant Genome.* 2016;9(1)
- Liu R, How-Kit A, Stamatiti L, Teyssier E, Rolin D, Mortain-Bertrand A, et al. A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci U S A.* 2015a;112(34):10804–9.
- Liu H, Meuwissen TH, Sørensen AC, Berg P. Upweighting rare favourable alleles increases long-term genetic gain in genomic selection programs. *Genet Sel Evol.* 2015b;47(1):19.
- Longin CFH, Reif JC. Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci.* 2014;19(10):631–6.
- Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink J-L, et al. Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model. *G3: Genes, Genomes, Genetics.* 2015;5(4):569–82.
- Lorenz AJ. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. *G3: Genes, Genomes, Genetics.* 2013;3(3):481–91.
- Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet.* 2009;120(1):151–61.
- Lush JL. *Animal breeding plans.* Ames: Collegiate Press, Inc; 1937.
- Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour.* 2017;17(1):67–77.
- Ly D, Hamblin M, Rabbi I, Melaku G, Bakare M, Gauch HG, et al. Relatedness and genotype \times environment interaction affect prediction accuracies in genomic selection: a study in cassava. *Crop Sci.* 2013;53(4):1312.
- Ly D, Chenu K, Gauffreteau A, Rincet R, Huet S, Gouache D, et al. Nitrogen nutrition index predicted by a crop model improves the genomic prediction of grain number for a bread wheat core collection. *Field Crop Res.* 2017;214:331–40.
- Ly D, Huet S, Gauffreteau A, Rincet R, Touzy G, Mini A, et al. Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *Field Crop Res.* 2018;216:32–41.
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, et al. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet.* 2019;51(5):885–95.
- MacLeod IM, Hayes BJ, Goddard ME. A novel predictor of multilocus haplotype homozygosity: comparison with existing predictors. *Genet Res.* 2009;91(6):413–26.
- Malézieux E, Crozat Y, Dupraz C, Laurans M, Makowski D, Ozier-Lafontaine H, et al. Mixing plant species in cropping systems: concepts, tools and models: a review. In: Lichtfouse E, Navarrete M, Debaeke P, Véronique S, Alberola C, editors. *Sustainable agriculture.* Dordrecht: Springer; 2009. p. 329–53.
- Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA. Predicting responses in multiple environments: issues in relation to genotype \times environment interactions. *Crop Sci.* 2016;56(5):2210–22.
- Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol.* 2003;18(4):189–97.
- Mangin B, Rincet R, Rabier C-E, Moreau L, Goudemand-Dugue E. Training set optimization of genomic prediction by means of EthAcc. *PLoS One.* 2019;14(2)
- Martre P, Jamieson PD, Semenov MA, Zyskowski RF, Porter JR, Tribot E. Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. *Eur J Agron.* 2006;25(2):138–54.

- Mascher M, Schreiber M, Scholz U, Graner A, Reif JC, Stein N. Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat Genet.* 2019;51(7):1076–81.
- Maskin L, Gudesblat GE, Moreno JE, Carrari FO, Frankel N, Sambade A, et al. Differential expression of the members of the Asr gene family in tomato (*Lycopersicon esculentum*). *Plant Sci.* 2001;161(4):739–46.
- Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J. Microsatellites in Zea – variability, patterns of mutations, and use for evolutionary studies. *Theor Appl Genet.* 2002;104(2-3):436–50.
- Maynard-Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974;23(1):23–35.
- Mayr E. Change of genetic environment and evolution. London: Allen and Unwin; 1954.
- McCouch SR, Kovach MJ, Sweeney M, Jiang H, Semon M. The dynamics of rice domestication: a balance between gene flow and genetic isolation. In: Biodiversity in agriculture: domestication, evolution, and sustainability. Cambridge University Press; 2012. p. 311–29.
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, et al. Feeding the future. *Nature.* 2013;499(7456):23–4.
- Messina CD, Jones JW, Boote KJ, Vallejos CE. A gene-based model to simulate soybean development and yield responses to environment. *Crop Sci.* 2006;46(1):456–66.
- Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M. Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). *Eur J Agron.* 2018;100:151–62.
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E. Simulated annealing. *J Chem Phys.* 1953;21(161-162):1087–92.
- Meuwissen THE. Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci.* 1997;75(4):934–40.
- Meuwissen TH, Sonesson AK. Genotype-assisted optimum contribution selection to maximize selection response over a specified time period. *Genet Res.* 2004;84(2):109–16.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819.
- Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet.* 2013;14(12):840–52.
- Mi X, Utz HF, Technow F, Melchinger AE. Optimizing resource allocation for multistage selection in plant breeding with R package. *Crop Sci.* 2014;54(4):1413–8.
- Mi X, Utz HF, Melchinger AE. Selectiongain: an R package for optimizing multi-stage selection. *Comput Stat.* 2016;31(2):533–43.
- Michel S, Kummer C, Gallee M, Hellinger J, Ametz C, Akgöl B, et al. Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet.* 2018;131(2):477–93.
- Millet EJ, Kruijer W, Coupel-Ledru A, Prado SA, Cabrera-Bosquet L, Lacube S, et al. Genomic prediction of maize yield across European environmental conditions. *Nat Genet.* 2019;51(6):952–6.
- Mizubuti ESG, Fry WE. Potato late blight. In: Cooke BM, Jones DG, Kaye B, editors. The epidemiology of plant diseases. Dordrecht: Springer; 2006. p. 445–71.
- Moeiniazade S, Hu G, Wang L, Schnable PS. Optimizing selection and mating in genomic selection with a look-ahead approach: an operations research framework. *G3: Genes, Genomes, Genetics.* 2019;9(7):2123.
- Mohammadi M, Tiede T, Smith KP. PopVar: a genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. *Crop Sci.* 2015;55(5):2068–77.
- Moler ERV, Abakir A, Eleftheriou M, Johnson JS, Krutovsky KV, Lewis LC, et al. Population epigenomics: advancing understanding of phenotypic plasticity, acclimation, adaptation and

- diseases. In: Rajora OP, editor. Population genomics: concepts, approaches and applications [Internet]. Cham: Springer International Publishing AG; 2019.
- Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge KM, et al. A genomic Bayesian multi-trait and multi-environment model. *G3: Genes, Genomes, Genetics*. 2016;6(9):2725–44.
- Montesinos-López OA, Montesinos-López A, Crossa J, Gianola D, Hernández-Suárez CM, Martín-Vallejo J. Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3: Genes, Genomes, Genetics*. 2018;8(12):3829–40.
- Montesinos-López OA, Montesinos-López A, Luna-Vázquez FJ, Toledo FH, Pérez-Rodríguez P, Lillemo M, et al. An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3: Genes, Genomes, Genetics*. 2019;9(5):1355–69.
- Müller D. embvr: computation of expected maximum haploid breeding values. Zenodo. 2017.
- Müller D, Schopp P, Melchinger AE. Selection on expected maximum haploid breeding values can increase genetic gain in recurrent genomic selection. *G3: Genes, Genomes, Genetics*. 2018;8(4):1173–81.
- Nakagawa H, Yamagishi J, Miyamoto N, Motoyama M, Yano M, Nemoto K. Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. *Theor Appl Genet*. 2005;110(4):778–86.
- Nakamichi N, Kiba T, Henriques R, Mizuno T, Chua N-H, Sakakibara H. PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the arabidopsis circadian clock. *Plant Cell*. 2010;22(3):594.
- Nakamichi N, Takao S, Kudo T, Kiba T, Wang Y, Kinoshita T, et al. Improvement of arabidopsis biomass and cold, drought and salinity stress tolerance by modified circadian clock-associated PSEUDO-RESPONSE REGULATORS. *Plant Cell Physiol*. 2016;57(5):1085–97.
- Nakamichi N, Kudo T, Makita N, Kiba T, Kinoshita T, Sakakibara H. Flowering time control in rice by introducing arabidopsis clock-associated PSEUDO-RESPONSE REGULATOR 5. *Biosci Biotechnol Biochem*. 2020;84(5):970–9.
- Narasimhamoorthy B, Gill BS, Fritz AK, Nelson JC, Brown-Guedira GL. Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. *Theor Appl Genet*. 2006;112(5):787–96.
- Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A*. 1973;70(12):3321–3.
- Neveu P, Tireau A, Hilgert N, Nègre V, Mineau-Cesari J, Bricchet N, et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytol*. 2019;221(1):588–601.
- Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005;39(1):197–218.
- Ødegaard J, Yazdi MH, Sonesson AK. Incorporating desirable genetic characteristics from an inferior into a superior population using genomic selection. *Genetics*. 2009;181(2):737–45.
- Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD. Selection under domestication: evidence for a sweep in the rice waxy genomic region. *Genetics*. 2006;173(2):975–83.
- Onogi A, Watanabe M, Mochizuki T, Hayashi T, Nakagawa H, Hasegawa T, et al. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. *Theor Appl Genet*. 2016;129(4):805–17.
- Oono Y, Kobayashi F, Kawahara Y, Yazawa T, Handa H, Itoh T, et al. Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC Genomics*. 2013;14(1):77.
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12)
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics*. 2012;192(3):1065.

- Pavlidis P, Alachiotis N. A survey of methods and tools to detect recent and strong positive selection. *J Biol Res.* 2017;24(1):7.
- Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol.* 2013;30(9):2224–34.
- Peng T, Sun X, Mumm RH. Optimized breeding strategies for multiple trait integration: I. Minimizing linkage drag in single event introgression. *Mol Breeding.* 2014;33(1):89–104.
- Pérez-Enciso M, Ramírez-Ayala LC, Zingaretti LM. SeqBreed: a python tool to evaluate genomic prediction in complex scenarios. *Genet Sel Evol.* 2020;52(1):7.
- Peripolli E, Munari DP, Silva M, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet.* 2017;48(3):255–71.
- Peter BM. Admixture, population structure, and F-statistics. *Genetics.* 2016;202(4):1485–501.
- Petit M, Astruc J-M, Sarry J, Drouilhet L, Fabre S, Moreno CR, et al. Variation in recombination rate and its genetic determinism in sheep populations. *Genetics.* 2017;207(2):767.
- Plucknett DL. Gene banks and the world's food: Princeton University Press; 1987.
- Pommier C, Michotey C, Cornut G, Roumet P, Duch E, et al. Applying FAIR principles to plant phenotypic data management in GnpIS. *Plant Phenomics.* 2019;2019:1671403.
- Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W, Armisen D, et al. Tracing the ancestry of modern bread wheats. *Nat Genet.* 2019;51(5):905–11.
- Pook T, Schlather M, Simianer H. MoBPS-modular breeding program simulator. *G3: Genes, Genomes, Genetics.* 2020;10(6):1915–8.
- Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 2009;5(6):e1000519.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59.
- Prudent M, Lecomte A, Bouchet J-P, Bertin N, Causse M, Génard M. Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. *J Exp Bot.* 2011;62(3):907–19.
- Pszczola M, Strabel T, Mulder HA, Calus MPL. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci.* 2012;95(1):389–400.
- Quilot B, Génard M, Lescourret F, Kervella J. Simulating genotypic variation of fruit quality in an advanced peach × *Prunus davidiana* cross. *J Exp Bot.* 2005;56(422):3071–81.
- Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet.* 2015;16(6):359–71.
- Rajora OP, Eckert AJ, Zinck JWR. Single-locus versus multilocus patterns of local adaptation to climate in eastern white pine (*Pinus strobus*, Pinaceae). *PLoS One.* 2016;11(7):e0158691.
- Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. *Science.* 2018;361(6403)
- Ramstein GP, Larsson SJ, Cook JP, Edwards JW, Ersoz ES, Flint-Garcia S, et al. Dominance effects and functional enrichments improve prediction of agronomic traits in hybrid maize. *Genetics.* 2020;215(1):215–30.
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature.* 2009;461(7263):489–94.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. *Nature.* 2012;488(7411):370–4.
- Reif JC, Hamrit S, Heckenberger M, Schipprack W, Peter Maurer H, Bohn M, et al. Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. *Theor Appl Genet.* 2005;111(5):906–13.
- Reilstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol.* 2015;24(17):4348–70.

- Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 2003;131(2):664–75.
- Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics.* 1983;105(3):767–79.
- Reynolds MP, Lewis JM, Ammar K, Basnet BR, Crespo-Herrera L, Crossa J, et al. Harnessing translational research in wheat for climate resilience. *J Exp Bot.* 2021;72(14):5134–57.
- Ribaut J-M, Ragot M. Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *J Exp Bot.* 2007;58(2):351–60.
- Riedelheimer C, Czedik-Eysenberg A, Grieder C, Lisek J, Technow F, Sulpice R, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet.* 2012;44(2):217–20.
- Rieseberg LH, Wendel JF. Introgression and its consequences in plants. *Hybrid Zones Evol Process.* 1993;70:109.
- Rincint R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, et al. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics.* 2012;192(2):715–28.
- Rincint R, Charcosset A, Moreau L. Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. *Theor Appl Genet.* 2017a;130(11):2231–47.
- Rincint R, Kuhn E, Monod H, Oury F-X, Rousset M, Allard V, et al. Optimization of multi-environment trials for genomic selection based on crop models. *Theor Appl Genet.* 2017b;130(8):1735–52.
- Rincint R, Charpentier J-P, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, et al. Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3: Genes, Genomes, Genetics.* 2018;8(12):3961–72.
- Rincint R, Malosetti M, Ababaei B, Touzy G, Mini A, Bogard M, et al. Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. *Theor Appl Genet.* 2019;132(12):3399–411.
- Rio S, Mary-Huard T, Moreau L, Charcosset A. Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. *Theor Appl Genet.* 2019;132(1):81–96.
- Robert P, Le Gouis J, Breadwheat Consortium T, Rincint R. Combining crop growth modelling with trait-assisted prediction improved the prediction of genotype by environment interactions. *Front Plant Sci.* 2020;11
- Robertson A. Inbreeding in artificial selection programmes. *Genet Res.* 1961;2(2):189–94.
- Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, Tosser-Klopp G, et al. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics.* 2018;19(1):71.
- Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, et al. Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol.* 2016;209(4):1781–94.
- Ross-Ibarra J, Morrell PL, Gaut BS. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc Natl Acad Sci.* 2007;104(suppl 1):8641–8.
- Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, et al. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics.* 2016;6(9):2799–808.
- Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, et al. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell.* 2007;19(8):2370.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature.* 2002;419(6909):832–7.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature.* 2007;449(7164):913–8.

- Saïdou A-A, Thuillet A-C, Couderc M, Mariac C, Vigouroux Y. Association studies including genotype by environment interactions: prospects and limits. *BMC Genet.* 2014;15(1):3.
- Sanchez L, Caballero A, Santiago E. Palliating the impact of fixation of a major gene on the genetic variation of artificially selected polygenes. *Genet Res.* 2006;88(2):105–18.
- Santos DJA, Cole JB, Lawlor TJ Jr, VanRaden PM, Tonhati H, Ma L. Variance of gametic diversity and its application in selection programs. *J Dairy Sci.* 2019;102(6):5279–94.
- Sanz-Alferez S, Richter TE, Hulbert SH, Bennetzen JL. The Rp3 disease resistance gene of maize: mapping and characterization of introgressed alleles. *Theor Appl Genet.* 1995;91(1):25–32.
- Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003;422(6929):297–302.
- Schaefer J, Duvernell D, Campbell DC. Hybridization and introgression in two ecologically dissimilar *Fundulus* hybrid zones. *Evolution.* 2016;70(5):1051–63.
- Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet.* 2006;123(4):218–23.
- Schnell FW, Utz HF. Bericht über die Arbeitstagung der Vereinigung österreichischer Pflanzenzüchter. Gumpenstein: BAL Gumpenstein; 1975.
- Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, et al. Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics.* 2018;208(4):1373–85.
- Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y. Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor Appl Genet.* 2016;129(2):273–87.
- Schulthess AW, Zhao Y, Longin CFH, Reif JC. Advantages and limitations of multiple-trait genomic prediction for *Fusarium* head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 2018;131(3):685–701.
- Schulz-Streeck T, Ogutu JO, Gordillo A, Karaman Z, Knaak C, Piepho H-P. Genomic selection allowing for marker-by-environment interaction. *Plant Breeding.* 2013;132(6):532–8.
- Sedivy EJ, Wu F, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol.* 2017;214(2):539–53.
- Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet.* 2012;44(7):825–30.
- Selby P, Abbeles R, Backlund JE, Basterrechea Salido M, Bauchet G, Benites-Alfaro OE, et al. BrAPI – an application programming interface for plant breeding applications. *Bioinformatics.* 2019;35(20):4147–55.
- Servin B, Martin OC, Mézard M. Toward a theory of marker-assisted gene pyramiding. *Genetics.* 2004;168(1):513–23.
- Shepherd RK, Kinghorn BP. A tactical approach to the design of crossbreeding programs. In: *Proceedings of the sixth world congress on genetics applied to livestock production.* Armidale: University of New England; 1998. p. 431–8.
- Simmonds NW. Variability in crop plants, its use and conservation. *Biol Rev.* 1962;37(3):422–65.
- Simmonds NW. Principles of crop improvement. London: Longman; 1979.
- Simmonds NW. Introgression and incorporation. Strategies for the use of crop genetic resources. *Biol Rev.* 1993;68(4):539–62.
- Singh RP, Huerta-Espino J, Rajaram S. Achieving near-immunity to leaf and stripe rusts in wheat by combining slow rusting resistance genes. *Acta Phytopathol Entomol Hung.* 2000;35(1/4):133–40.
- Smith S, Beavis W. Molecular marker assisted breeding in a company environment. In: *The impact of plant molecular genetics.* Springer; 1996. p. 259–72.
- Sorensen D, Fernando R, Gianola D. Inferring the trajectory of genetic variance in the course of artificial selection. *Genet Res.* 2001;77(1):83–94.

- Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *Am J Bot*. 2016;103(1):33–46.
- Souza E, Sorrells ME. Prediction of progeny variation in oat from parental genetic relationships. *Theor Appl Genet*. 1991a;82(2):233–41.
- Souza E, Sorrells ME. Relationships among 70 North American oat germplasms: I. Cluster analysis using quantitative characters. *Crop Sci*. 1991b;31(3):599–605.
- Spannagl M, Alaux M, Lange M, Bolser DM, Bader KC, Letellier T, et al. transPLANT resources for triticeae genomic data. *Plant Genome*. 2016;9(1).
- Spillane C, Gepts P. Evolutionary and genetic perspectives on the dynamics of crop genepools. In: Broadening the genetic base of crop production. CABI; 2001. p. 25–70.
- Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ. A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Aust J Agr Res*. 2007;58(6):532–44.
- Stephan W. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol*. 2016;25(1):79–88.
- Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat Biotechnol*. 2016;34(6):652.
- Stich B. Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics*. 2009;183(4):1525–34.
- Storn R, Price K. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optim*. 1997;11(4):341–59.
- Stucki S, Orozco-terWengel P, Forester BR, Duruz S, Colli L, Masembe C, et al. High performance computation of landscape genomic models including local indicators of spatial association. *Mol Ecol Resour*. 2017;17(5):1072–89.
- Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QCB, et al. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol Ecol*. 2016;25(11):2427–42.
- Suarez-Gonzalez A, Hefer CA, Lexer C, Cronk QCB, Douglas CJ. Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Mol Ecol*. 2018;27(7):1667–80.
- Sun C, VanRaden PM, Cole JB, O’Connell JR. Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. *PLoS One*. 2014;9(8)
- Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink J-L, Sorrells ME. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome*. 2017;10(2).
- Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. 2014;31(10):2824–7.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3):585–95.
- Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. Plant phenomics, from sensors to knowledge. *Curr Biol*. 2017;27(15):R770–83.
- Technow F, Messina CD, Totir LR, Cooper M. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One*. 2015;10(6):e0130855.
- Tenaillon MI, Charcosset A. A European perspective on maize history. *C R Biol*. 2011;334(3):221–8.
- Thabuis A, Palloix A, Servin B, Daubeze AM, Signoret P, Lefebvre V. Marker-assisted introgression of 4 *Phytophthora capsici* resistance QTL alleles into a bell pepper line: validation of additive and epistatic effects. *Mol Breed*. 2004;14(1):9–20.
- Tiede T, Kumar L, Mohammadi M, Smith KP. Predicting genetic variance in bi-parental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. *Mol Breed*. 2015;35(10):199.

- Tinker NA, Deyl JK. A curated Internet database of oat pedigrees. *Crop Sci.* 2005;45(6):2269–72.
- Touzy G, Rincet R, Bogard M, Lafarge S, Dubreuil P, Mini A, et al. Using environmental clustering to identify specific drought tolerance QTLs in bread wheat (*T. aestivum* L.). *Theor Appl Genet.* 2019;132(10):2859–80.
- Uauy C, Brevis JC, Dubcovsky J. The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. *J Exp Bot.* 2006;57(11):2785–94.
- Uemoto Y, Sasaki S, Kojima T, Sugimoto Y, Watanabe T. Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. *BMC Genet.* 2015;16(1):134.
- Ullstrup AJ. The impacts of the southern corn leaf blight epidemics of 1970-1971. *Annu Rev Phytopathol.* 1972;10(1):37–50.
- Uptmoor R, Li J, Schrag T, Stützel H. Prediction of flowering time in Brassica oleracea using a quantitative trait loci-based phenology model. *Plant Biol.* 2012;14(1):179–89.
- Utz HF, Bohn M, Melchinger AE. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. *Crop Sci.* 2001;41(5):1470–8.
- Vacher M, Small I. Simulation of heterosis in a genome-scale metabolic network provides mechanistic explanations for increased biomass production rates in hybrid plants. *NPJ Syst Biol Appl.* 2019;5(1):1–10.
- Valente F, Gauthier F, Bardol N, Blanc G, Joets J, Charcosset A, et al. OptiMAS: a decision support tool for marker-assisted assembly of diverse alleles. *J Hered.* 2013;104(4):586–90.
- van Berloo R, Stam P. Marker-assisted selection in autogamous RIL populations: a simulation study. *TAG Theor Appl Genet.* 1998;96(1):147–54.
- van Heerwaarden J, Hufford MB, Ross-Ibarra J. Historical genomics of North American maize. *Proc Natl Acad Sci.* 2012;109(31):12420–5.
- van Inghelandt D, Melchinger AE, Lebreton C, Stich B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet.* 2010;120(7):1289–99.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91(11):4414–23.
- VanRaden PM, Tooker ME, Cole JB, Wiggans GR, Megonigal JH Jr. Genetic evaluations for mixed-breed populations. *J Dairy Sci.* 2007;90(5):2434–41.
- Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, et al. An analysis of genetic diversity across the maize genome using microsatellites. *Genetics.* 2005;169(3):1617–30.
- Vigouroux Y, Barnaud A, Scarcelli N, Thuillet A-C. Biodiversity, evolution and adaptation of cultivated crops. *C R Biol.* 2011;334(5):450–7.
- Virlouvet L, Jacquemot M-P, Gerentes D, Corti H, Bouton S, Gilard F, et al. The *ZmASR1* protein influences branched-chain amino acid biosynthesis and maintains kernel yield in maize under water-limited conditions. *Plant Physiol.* 2011;157(2):917–36.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):e72.
- Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* 2014;10(12):e1004845.
- Wang H-J, Hsu C-M, Jauh GY, Wang C-S. A lily pollen ASR protein localizes to both cytoplasm and nuclei requiring a nuclear localization signal. *Physiol Plant.* 2005;123(3):314–20.
- Wang C, Hu S, Gardner C, Lübberstedt T. Emerging avenues for utilization of exotic germplasm. *Trends Plant Sci.* 2017;22(7):624–37.
- Wang J, Hu Z, Upadhyaya HD, Morris GP. Genomic signatures of seed mass adaptation to global precipitation gradients in sorghum. *Heredity.* 2020;124(1):108–21.
- Wellmann R. Optimum contribution selection for animal breeding and conservation: the R package *optiSel*. *BMC Bioinformatics.* 2019;20(1):1–13.

- Wellmann R, Bennewitz J. Key genetic parameters for population management. *Front Genet.* 2019;10:667.
- Wen W, Liu H, Zhou Y, Jin M, Yang N, Li D, et al. Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the Maize Kernel. *Plant Physiol.* 2016;170(1):136–46.
- White JW, Hoogenboom G. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron J.* 1996;88(3):416–22.
- White JW, Herndl M, Hunt LA, Payne TS, Hoogenboom G. Simulation-based analysis of effects of Vrn and Ppd loci on flowering in wheat. *Crop Sci.* 2008;48(2):678–87.
- Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F ST. *Am Nat.* 2015;186(S1):S24–36.
- Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. *Genet Res.* 2000;75(2):249–52.
- Woolliams JA, Thompson R. A theory of genetic contributions. In: *Proceedings of the 5th world congress on genetics applied to livestock production.* Guelph; 1994. p. 127–34.
- Woolliams JA, Bijma P, Villanueva B. Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics.* 1999;153(2):1009–20.
- Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE. Genetic contributions and their optimization. *J Anim Breed Genet.* 2015;132(2):89–99.
- Wray NR, Goddard ME. Increasing long-term response to selection. *Genet Sel Evol.* 1994;26(5):431.
- Wray NR, Thompson R. Prediction of rates of inbreeding in selected populations. *Genet Res.* 1990;55(1):41–54.
- Wray NR, Woolliams JA, Thompson R. Methods for predicting rates of inbreeding in selected populations. *Theor Appl Genet.* 1990;80(4):503–12.
- Wright S. Evolution in Mendelian populations. *Genetics.* 1931;16(2):97–159.
- Wu J, Yu H, Dai H, Mei W, Huang X, Zhu S, et al. Metabolite profiles of rice cultivars containing bacterial blight-resistant genes are distinctive from susceptible rice. *Acta Biochim Biophys Sin.* 2012;44(8):650–9.
- Xu P, Wang L, Beavis WD. An optimization approach to gene stacking. *Eur J Oper Res.* 2011;214(1):168–78.
- Yabe S, Iwata H, Jannink J-L. A Simple package to script and simulate breeding schemes: the breeding scheme language. *Crop Sci.* 2017;57(3):1347–54.
- Yang C, Sakai H, Numa H, Itoh T. Gene tree discordance of wild and cultivated Asian rice deciphered by genome-wide sequence comparison. *Gene.* 2011;477(1–2):53–60.
- Yin X, Struik PC, Tang J, Qi C, Liu T. Model analysis of flowering phenology in recombinant inbred lines of barley. *J Exp Bot.* 2005;56(413):959–65.
- Yu X, Li X, Guo T, Zhu C, Wu Y, Mitchell SE, et al. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat Plants.* 2016;2(10):1–7.
- Zheng B, Biddulph B, Li D, Kuchel H, Chapman S. Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. *J Exp Bot.* 2013;64(12):3747–61.
- Zhong S, Jannink J-L. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics.* 2007;177(1):567–76.
- Zivy M, Wienkoop S, Renaut J, Pinheiro C, Goulas E, Carpentier S. The quest for tolerant varieties: the importance of integrating “omics” techniques to phenotyping. *Front Plant Sci.* 2015;6:448.
- Zorrilla-Fontanesi Y, Rambla J-L, Cabeza A, Medina JJ, Sánchez-Sevilla JF, Valpuesta V, et al. Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. *Plant Physiol.* 2012;159(2):851–70.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

