

# Improving Genetic Gain with Genomic Selection in Autotetraploid Potato

Anthony T. Slater,\* Noel O.I. Cogan, John W. Forster, Benjamin J. Hayes, and Hans D. Daetwyler

## Abstract

Potato (*Solanum tuberosum* L.) breeders consider a large number of traits during cultivar development and progress in conventional breeding can be slow. There is accumulating evidence that some of these traits, such as yield, are affected by a large number of genes with small individual effects. Recently, significant efforts have been applied to the development of genomic resources to improve potato breeding, culminating in a draft genome sequence and the identification of a large number of single nucleotide polymorphisms (SNPs). The availability of these genome-wide SNPs is a prerequisite for implementing genomic selection for improvement of polygenic traits such as yield. In this review, we investigate opportunities for the application of genomic selection to potato, including novel breeding program designs. We have considered a number of factors that will influence this process, including the autotetraploid and heterozygous genetic nature of potato, the rate of decay of linkage disequilibrium, the number of required markers, the design of a reference population, and trait heritability. Based on estimates of the effective population size derived from a potato breeding program, we have calculated the expected accuracy of genomic selection for four key traits of varying heritability and propose that it will be reasonably accurate. We compared the expected genetic gain from genomic selection with the expected gain from phenotypic and pedigree selection, and found that genetic gain can be substantially improved by using genomic selection.

## Core Ideas

- Progress in conventional potato breeding can be complex and slow.
- Genomic selection can be applied to autotetraploid potato.
- A number of factors that will affect genomic selection need to be considered.
- Genomic selection will accelerate genetic gain in potato.

**P**OTATO is a significant and highly versatile human food crop produced in nearly 150 countries (Bradshaw and Ramsey, 2009) and is the major source of carbohydrate in the diets of hundreds of millions of people. Potatoes also supply significant amounts of protein, vitamins, and minerals (Bradshaw and Ramsey, 2009; Storey, 2007). These attributes have seen potato rapidly rise to be the fourth most important human food crop on a global basis, although it has only been 500 yr since its emergence from its center of origin in South America.

Since emerging, potato has been selected and bred for higher levels of local adaption to various environments in production areas. This outcome was achieved relatively quickly because of the high genetic diversity of potato, allowing identification of genotypes that perform

Published in Plant Genome  
Volume 9. doi: 10.3835/plantgenome2016.02.0021

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A.T. Slater, N.O.I. Cogan, J.W. Forster, B.J. Hayes, and H.D. Daetwyler, AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Vic. 3083, Australia; J.W. Forster, B.J. Hayes, and H.D. Daetwyler, La Trobe University, Bundoora, Vic. 3086, Australia. Received 23 Oct 2015. Accepted 23 March 2016. \*Corresponding author (tony.slater@ecodev.vic.gov.au).

**Abbreviations:** BLUP, best linear unbiased prediction; DR, double reduction; GBS, genotyping-by-sequencing; LD, linkage disequilibrium; MAS, marker-assisted selection;  $N_e$ , effective population size; QTL, quantitative trait locus; SNP, single nucleotide polymorphism

better under various conditions. Such genetic heterogeneity is a result of both the obligate outbreeding habit of their progenitor species and the additional genotypic variation arising from the autotetraploid chromosomal constitution of cultivated varieties (Slater et al., 2014a). Despite this, genetic gain in complex traits, such as yield, has been slow to nonexistent (Jansky, 2009), especially when compared to other crops such as maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), and rice (*Oryza sativa* L.) (Fischer and Edmeades, 2010). The development of superior potato cultivars is demanding, as breeders must consider a large number of traits during cultivar development. A number of these traits are under simple genetic control, whereas others are under the influence of several to a large number of genetic effects (Slater et al., 2014a). Breeding for improvements in the simple traits is relatively straightforward but obtaining improvements in the more complex traits is more difficult.

A significant amount of effort has focused on developing genomic resources to improve potato breeding, which has culminated in a draft potato genome (Potato Genome Sequencing Consortium, 2011). This has enabled the identification of over 39,000 genes, including a number that control biotic stress resistance (Bakker et al., 2011; Jupe et al., 2012; Jupe et al., 2013) and quantitative trait loci (QTLs) for quality traits (D'hoop et al., 2014; Uitdewilligen et al., 2013). However, for traits such as yield, QTLs of large effect are unlikely, even though these traits exhibit reasonable levels of heritability (Slater et al., 2014b). A likely genetic architecture for these traits is many mutations of small effect, as observed for yield in rice and maize (Huang et al., 2010; Laurie et al., 2004). Genomic selection, where genome-wide SNPs that are in linkage disequilibrium (LD) with the many QTLs affecting yield and other traits are used to derive genomic estimated breeding values for these traits (Meuwissen et al., 2001), is now possible in potato, given the large number of SNPs discovered by genome sequencing (Uitdewilligen et al., 2013). Genomic selection is being successfully applied in a number of animal and plant improvement programs (Crossa et al., 2010; Daetwyler et al., 2010a; Grattapaglia et al., 2011; Lin et al., 2014; Resende et al., 2012; Riedelsheimer et al., 2012; VanRaden et al., 2009; Wiggans et al., 2011; Wolc et al., 2011, 2015). In this review, we investigate the feasibility of applying genomic selection to potato, including the extent of increase in genetic gain expected as a result of using genomic selection in novel breeding schemes.

## Improving Potato Breeding Models

Potato breeding is challenging, as ~40 traits are considered during new cultivar development (Gebhardt, 2013). These traits can be broadly divided into yield-related and tuber quality characteristics, as well as tolerances to biotic and abiotic stresses (Slater et al., 2014a). Knowledge of the genetic control of each trait and the degree of environmental influence on the expression of the target traits is important and will influence which method should be

used for selecting superior phenotypes and genotypes. Some traits are controlled by single genes, whereas others are under more complex control (Slater et al., 2014a). Breeding is made more challenging in potatoes than in some grain and forage crops, not only because of potato's highly heterozygous and autotetraploid nature but also because more market-specific traits are considered. Many target traits such as yield, tuber number, tuber size, specific gravity, and processing quality can be highly affected by the growing environment, which can vary significantly (Jansky, 2009). Consequently, a conventional breeding scheme will involve screening genotypes across a number of clonal generations and in a number of appropriate locations for a range of desirable characters, which can take over 10 yr (Jansky, 2009).

Over recent years, significant advances have been made in understanding the genetics of potato to improve breeding for more rapid genetic gain. A conventional breeding program creates a large population, then employs phenotypic recurrent selection over a number of generations, using a progression of selection pressures to reduce the population size while concurrently increasing the number of plants under evaluation of each genotype (Bradshaw and Mackay, 1994; Jansky, 2009). To reduce the size of a breeding population, the majority of programs practice early-generation visual selection to enable a more thorough assessment of fewer clones in later trials. Prior to the mid-1980s, most potato breeding programs planted a very large number of seedlings and then used intensive selection to reduce the population to a manageable size. Since then, several studies have concluded that this practice is ineffective, leading to the elimination of superior as well as inferior genotypes (Anderson and Howard, 1981; Bradshaw and Mackay, 1994; Brown et al., 1988, 1984, 1987; Tai and Young, 1984), largely because the expression of visual traits is strongly influenced by seed tuber weight (Maris, 1986). Some programs that still practice intense phenotypic selection (Haynes et al., 2012) have failed to obtain any improvement in yield over time, despite 150 yr of breeding (Jansky, 2009), which is likely to be a result of intense selection for visual attributes that are not highly correlated with final field performance.

To overcome the low heritability of these visual traits, progeny tests have been employed to determine better parental combinations (Bradshaw, 2007). Progeny testing of all members of a full-sib family will determine the value of the parent for these traits, without knowledge of the location or number of genes that regulate the expression of the trait, which will be caused by the accumulation of many minor genetic effects. Progeny tests for multiple traits (visual preference, late blight tuber and foliar resistances, white potato cyst nematode resistance, and fry color for processing families) have been used by a Scottish breeding program (Bradshaw, 2007; Bradshaw et al., 2003), resulting in the identification of superior clones and parental types with improved disease resistance, as well as a 14% increase in yield (Bradshaw et al., 2009).

More recently, the use of pedigree information in the analysis enabled the phenotypic values from all relatives (including full-sibs, half-sibs, and any other partial sibling) to maximize the amount of derived information to obtain a more accurate genotypic value. This allowed the estimation of breeding values for the selection of candidates on the basis of their genetic merit using best linear unbiased prediction (BLUP) (Henderson, 1975) to exploit the additive genetic variance. The use of BLUP estimated breeding values in potato breeding clearly demonstrated the advantage of using estimated breeding values over progeny means in cross-generation prediction of progeny performance, particularly for traits with low heritability (Slater et al., 2014b).

The use of molecular genetic markers provides the opportunity to improve breeding significantly by reducing both the duration and costs of a breeding cycle. Marker-assisted selection (MAS) has the ability to select for traits several years earlier in a program than would be practical using conventional screening methods. Marker-assisted selection will be effective for qualitative traits that are controlled by major genes but it may also be valuable for quantitative characteristics, if QTLs of large effect contribute to the measured trait. Although a substantial number of genetic markers linked to genes for important traits have been identified, only a few reports have been made of their use in commercial potato breeding programs (Dalla Rizza et al., 2006; Ortega and Lopez-Vizcon, 2012; Ottoman et al., 2009; Schultz et al., 2012). For potato breeders to adopt MAS, the use of markers must be cost-effective compared to conventional screening, as has been shown to be the case for pest and disease resistance screening (Slater et al., 2013).

As MAS can be cost-effectively applied at the second field generation (Slater et al., 2013) and estimated breeding values can be calculated for a number of more complex traits at the same stage (Slater et al., 2014b), combined use of the two methods could reduce the breeding cycle from >10 to 4 yr (Slater et al., 2014a). These advances will greatly accelerate the breeding cycle and therefore increase genetic gain over conventional breeding methods. When used together with a weighted selection index, they will ensure that improvement is made in all measured traits, from those under simple genetic control to those under far more complex control. Further reductions in the cycle time should be possible with genomic selection.

### Considerations for Genomic Selection in Potato

Genomic selection was first proposed by Meuwissen et al. (2001), and uses genome-wide marker effects estimated in a phenotyped and genotyped reference population to predict the genetic merit (future phenotypes) of otherwise uncharacterized selection candidates. Genomic selection differs from MAS in that it jointly analyzes all marker data and can therefore capture all the genetic variance, whereas MAS will only capture the variance from a limited number of QTLs. In addition, genomic

selection does not experience the problems associated with genome-wide association studies and QTL analysis, such as the overestimation of marker effects (Beavis, 1998). With the completion of the potato genome sequence and the identification of a large number of genome-wide SNPs (Uitdewilligen et al., 2013), application of genomic selection in potato can be confidently expected in the near future, although the heterozygous nature of potato (Slater et al., 2014a) will require some considered strategies.

### Dense Marker Maps, LD, and Marker Numbers

Genomic selection requires a large number of markers that are spread across the entire potato genome. A number of potato studies have addressed this issue, with successive studies increasing the number of markers identified across the genome for the development of genome-wide marker maps (Bonierbale et al., 1988; Dong et al., 2000; Gebhardt et al., 1991, 1989; Milbourne et al., 1998; Tanksley et al., 1992). This process culminated in the development of a dense genetic linkage map populated with 10,000 amplified fragment length polymorphism markers (Van Os et al., 2006), which was used to construct a physical map to assist the assembly of the potato genome sequence.

The completion of the genome sequence allowed the identification of multiple SNPs in any set of sequenced genotypes that were aligned to it (Uitdewilligen et al., 2013). These SNPs can be used as a densely spaced set of molecular markers. The SNP frequency in potato has been estimated to be approximately 1 per 24 bp within exons (Uitdewilligen et al., 2013), illustrating the extent of sequence diversity within potato. The assembly of the potato genome sequence has also led to the development of an 8303-feature SNP chip ([http://solcap.msu.edu/potato\\_infinium.shtml](http://solcap.msu.edu/potato_infinium.shtml), accessed 13 June 2016) (Felcher et al., 2012; Hamilton et al., 2011). Application of this tool has been limited, potentially because of the limited number of markers that it assays, as SNP chips in many other species have recorded a large percentage of unusable data as a result of missing data, SNPs not segregating in the populations under investigation, or through issues of polyploid SNP calling (~50% usable data only; Jan et al. (2016)). Problems with SNP cluster calling in polyploids, particularly of the heterozygous genotypic classes, have been partly addressed through the development of custom software packages (Voorrips et al., 2011; [http://www.illumina.com/Documents/products/technotes/tech\\_note\\_genomestudio\\_polypld\\_genotyping.pdf](http://www.illumina.com/Documents/products/technotes/tech_note_genomestudio_polypld_genotyping.pdf), accessed 13 June 2016) that now allow the calling of Infinium (Illumina, San Diego, CA) data into the five genotypic classes (Pembleton et al., 2013). The genome sequence will also provide a reference for resequencing for the delivery of large numbers of molecular genetic markers in a rapid timeframe. As the price of sequencing continues to reduce, low-coverage whole-genome sequencing will become an option in the near future.

Although the SNP chips will allow identification of genes with a large effect through genome-wide association studies, they may not capture the entire allele frequency spectrum (i.e., ascertainment bias) and may therefore not detect some relevant effects. Single nucleotide polymorphism chip systems will potentially be superseded in the near future by genotyping-by-sequencing (GBS) methods, which could deliver genome-wide SNP profiles at significantly reduced cost (Elshire et al., 2011; Xu et al., 2012). The genome-wide profiles will allow the computational analysis of partial or whole genomic SNP complements for their combined effect on complex phenotypic traits, thus enabling genomic selection applications and strategies in potato. To achieve this, the main issues faced with GBS methods currently applied are the number of features assayed, the volume of missing data that have to be compensated for, and the number of dominant marker types that are included in the data. The large number of fragments assayed is a positive attribute when sequencing costs are not considered; however, if a limited budget or a price point needs to be reached to be economically viable and beneficial, fewer sequence reads will be generated and there will be a higher volume of missing data. Handling datasets with up to 90% missing data is complex and challenging and relies heavily on imputation, the accuracy of which with high missing genotypes can be variable (Fu, 2014). An alternative option is to target specific portions of the genome through a capture-based system (Uitdewilgen et al., 2013). However, issues arising from sequence depth may incorrectly identify the specific allelic class present. The data presented in Uitdewilgen et al. (2013) indicate a high sequence depth is needed to identify the correct genotypic class accurately (60–80× coverage leading to 98.4% accuracy in genotypic calls). Empirical data from a genomic potato breeding program are needed to model the effect of miscalling the allelic status, including double reduction (DR), and defining the reduction in prediction accuracies as a result. The reduction in prediction accuracy caused by miscalling a number of allelic classes must also be compared to and contrasted with the increase in genome coverage and the increase in SNP marker numbers achieved by implementing such an approach.

The advantage of a large number of SNPs is that this provides coverage of the entire genome, ensuring that all QTL are in LD with at least one marker and therefore capturing the majority of the genetic variance. This would entail thousands of markers for those crops in which LD decays slowly but millions for those with rapid LD decay (Xu et al., 2012). The key question for genomic selection is the extent of LD in potato breeding populations and, in turn, how many SNPs are needed to capture the majority of QTLs through LD with markers. Calus et al. (2008) proposed that a mean LD of 0.25 between neighboring markers was sufficient for successful genomic selection in their simulation study. In potato, high LD has been shown to exist at distances less than 1 cM and to decay quickly to less than 0.2 at inter-marker distances greater than 1 cM (D'hoop et al., 2010). This

would imply that approximately four SNPs per cM would suffice to achieve adequate genomic selection accuracy. Thus, at least 8000 SNPs would be required, assuming a genome map length of 18.72 M (Sharma et al., 2013). Another way to approximate the number of markers that are required has been proposed by Meuwissen (2009) through the estimation of the effective population size ( $N_e$ ) of a breeding population and calculation of the number of markers needed using the formula  $10N_eL$ , where  $L$  is the genome length in Morgans. We have estimated the  $N_e$  of one potato breeding population to be 79 (details given below), which would require 14,789 SNPs for successful genomic selection. These marker estimates should be understood as the numbers needed when the genotype call rate is very high and all SNPs are segregating in the target population. If genotype call rates are only moderate or populations are more diverse, the number of SNPs required will increase.

These studies on LD in potato indicate that the worldwide population is genetically diverse. However, it is the diversity within specific breeding populations that is of most direct relevance to genomic selection. This diversity has often been greatly reduced by resource induced bottlenecking events (such as selection of relatively few parents). In such populations, the moderate SNP numbers described above should be sufficient.

## Genomic Selection Approaches in Autopolyploids

Working with the cultivated autotetraploid form of potato adds a degree of complexity, as the genotyping method ideally needs to determine the dosage of the different marker alleles, where the possible genotypes are AAAA, AAAB, AABBB, AB BBB, and BBBB. Uitdewilgen et al. (2013) and Ashraf et al. (2014) described a method based on GBS, whereas Gidskehaug et al. (2011) described a method of quantifying allele dosage in autotetraploids using intensity information from genotyping on SNP arrays.

One model for genomic selection (ignoring fixed effects other than the mean) is:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\mathbf{b} + \mathbf{e}, \quad [1]$$

where  $\mathbf{y}$  is a (number of phenotypes  $\times$  1) vector of phenotypes,  $\mathbf{1}_n$  is a vector of ones,  $\mu$  is the population mean,  $\mathbf{X}$  is a matrix of genotypes,  $\mathbf{b}$  is a vector of marker effects, and  $\mathbf{e}$  is the vector of random normal deviates. There are at least two possible assumptions regarding the effect of marker allele dosage on phenotype for genomic selection. One assumption would be a pseudodiploid model, where all heterozygous genotypes have an equal effect on the genotype, and that the effects of the heterozygotes is at the midpoint of the two homozygotes. In this instance, there is one effect per marker ( $\mathbf{b}$  is a number of markers ( $m$ )  $\times$  1 vector) and the  $\mathbf{X}$  matrix has the dimensions number of phenotypes ( $p$ )  $\times$   $m$ , coded as in Table 1.



**Table 1. Possible coding of the genotype design matrix for autotetraploid potato.**

Cultivar genotype	Pseudo diploid	Additive autotetraploid	Full autotetraploid, including nonadditive				
Number of effects per marker	1	1	1	2	3	4	5
AAAA	0	0	1	0	0	0	0
AAAB	1	1	0	1	0	0	0
AABB	1	2	0	0	1	0	0
ABBB	1	3	0	0	0	1	0
BBBB	2	4	0	0	0	0	1

This model could be adapted to estimate the additive marker effect by accounting for tetraploid allele dosage. In this case,  $\mathbf{X}$  still has the dimensions  $p \times m$ , but is now coded 0, 1, 2, 3, and 4, for AAAA, AAAB, AABB, ABBB, and BBBB, respectively (Table 1).

An alternative approach, which accounts for additive and nonadditive effects in a full autotetraploid model, is to assume that each genotype has its own effect. In this case, there are five possible effects per marker, assuming that the markers are fitted as random effects, so  $\mathbf{b}$  is a vector with the dimension  $5m \times 1$  and  $\mathbf{X}$  is a matrix with  $p$  rows and  $5m$  columns (Table 1).

These  $\mathbf{X}$  matrices can be used to implement genomic selection in either BLUP methods (assuming that all marker effects are derived from the same normal distribution) or a method that assumes nonlinear distributions (e.g., BayesB, BayesA, and BayesR) (Erbe et al., 2012; Meuwissen et al., 2001). A similar approach to the full autotetraploid model as described in Table 1 has been used to estimate haplotype effects (Pryce et al., 2010). It is noteworthy that the pseudodiploid model could also be used for a genotyping system designed for diploids (where only AA, AB, and BB genotypes are called).

Fitting the above models results in either a single effect for each marker or five effects (one for each genotype), which can be then used to calculate the genomic estimated breeding values for selection candidates as:

$$\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{b}}, \quad [2]$$

where GEBV is the genomic estimated breeding value,  $\mathbf{X}$  is the design matrix for selection candidates constructed from their genotypes, and  $\hat{\mathbf{b}}$  is the estimated genotype effects.

As described by Habier et al. (2007), VanRaden (2008), and Hayes et al. (2009c), an alternative implementation of the BLUP genomic selection is to first construct the genomic relationship matrix  $\mathbf{G}$  among individuals using the SNP data, then fit the model in Eq. [3]:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}, \quad [3]$$

where  $\mathbf{Z}$  is a design matrix mapping phenotypes to individuals and  $\mathbf{g}$  is a vector (individuals  $\times$  1) of genomic breeding values. The genomic breeding values are assumed to be distributed as:

$$\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2), \quad [4]$$

For the pseudodiploid model,  $\mathbf{G}$  can be constructed via Eq. [5] (following Yang et al. (2010)) for the elements  $jk$ , which indicates the genomic relationship between the individuals  $j$  and  $k$ ,

$$G_{jk} = \frac{1}{M} \sum_i \frac{(x_{ji} - 2p_i)(x_{ki} - 2p_i)}{2p_i(1 - p_i)}. \quad [5]$$

Alternatively,  $\mathbf{G}$  could be constructed as Eq. [6] for the diagonal element for the individual  $j$ :

$$G_{jj} = 1 + \frac{1}{M} \sum_i \frac{x_{ji}^2 - (1 + 2p_i)x_{ji} + 2p_i^2}{2p_i(1 - p_i)}, \quad [6]$$

where  $M$  is the number of markers,  $x_{ji}$  and  $x_{ki}$  are elements of the design matrix  $\mathbf{X}$  for the pseudodiploid model described above, and  $p_i$  for the  $i^{\text{th}}$  marker is obtained as:

$$\frac{4n_{bbbb} + 3n_{abbb} + 2n_{aabb} + n_{aaab}}{4N}, \quad [7]$$

where  $N$  is the number of individuals and  $n_{bbbb}$ ,  $n_{abbb}$ ,  $n_{aabb}$ , and  $n_{aaab}$  indicate the number of individuals carrying the BBBB genotype, the ABBB genotype, the AABB genotype, and the AAAB genotype respectively.

For the full autotetraploid model,  $\mathbf{G}$  can be constructed as shown in Eq. [8] for the elements  $jk$ , which indicates the genomic relationship between the individuals  $j$  and  $k$ :

$$G_{jk} = \frac{1}{M} \sum_i \frac{(x_{ji} - p_i)(x_{ki} - p_i)}{p_i(1 - p_i)}, \quad [8]$$

where  $M$  is now the number of markers  $\times$  5. For the diagonal element for the individual  $j$ ,  $\mathbf{G}$  can be constructed as:

$$G_{jj} = 1 + \frac{1}{M} \sum_i \frac{x_{ji}^2 - (2p_i)x_{ji} + p_i^2}{p_i(1 - p_i)}, \quad [9]$$

where  $p_i$  is now the frequency of each genotype (e.g., for the first marker  $n_{bbbb}N^{-1}$ , where  $n_{bbbb}$  is the number of individuals carrying the BBBB genotype,  $n_{aaab}$  is the number of individuals carrying the AAAA genotype and so on).

The model which best fits the data may be determined on the basis of maximum likelihood. For example,

the model shown in Eq. [3] could be fitted to the data and the maximum likelihood of the pseudodiploid  $G$  would be compared to the full autotetraploid  $G$  [ASReml, (Gilmore et al., 2009) could be used to fit these models]. The log-likelihood of tetraploid BLUP models with varying degrees of DR were compared in Slater et al. (2014b), where the rate of DR with the highest model likelihood was variable across traits.

The genomic relationship matrices described above are based on identity-by-state and simply measure the similarity of alleles between individuals. A more complex problem is the modeling of identical-by-descent probabilities within a genomic relationship context, with identical-by-descent measuring the likelihood that two alleles are the same and have originated from a common ancestor. In cattle (diploid), one comparison of genomic relationship matrices based on identity-by-state and identical-by-descent failed to yield a significant difference in the accuracy of genomic prediction (Luan et al., 2012). In autopolyploids, this issue may be more important due to the phenomenon of DR (Huang et al., 2015, 2014), which allows the combination of alleles from sister chromatids in the same gamete and adds additional complexity. Slater et al. (2014b) compared heritability estimates in potato across a range of traits, with a relationship matrix derived from pedigrees that either considered or ignored DR and concluded that the estimated values were very similar in both cases. However, it would be valuable to repeat the construction of  $G$  when considering DR. Genetic markers should make it possible to identify accurately when DR has occurred, provided that all three possible heterozygote genotypes can be accurately distinguished.

### Potential Reference Populations and Trait Heritability

D'hoop et al. (2010) used a substantial number of markers distributed across the potato genome to investigate the population structure within a large germplasm collection of tetraploid cultivars and progenitor clones. These data were analyzed using three methods, which all obtained very similar results. They determined that the population resolved into six groups that correlated with historic breeding aims and activities. The three main groups contained cultivars that separately related to the starch, processing, and fresh market sectors and exhibited a graduation in average dry matter content from very high to high then medium, respectively. The smallest group (called SH) contained three diploid clones. A group called 'Ancient' contained cultivars bred between 1850 and 1950; the last group, termed 'the rest' contained progenitor clones used for introgressing disease resistance genes, a diploid cultivar, and miscellaneous European cultivars. Previous studies had not been able to resolve a population structure but had been limited by a low number of markers (Gebhardt et al., 2004; Simko et al., 2004, 2006).

This population structure in potato would assist in the identification of the germplasm that should be

represented in the reference or training populations for genomic selection. As breeding programs have undertaken directional selection toward their individual breeding targets, germplasm from each of these targets should form the nucleus of the reference population. This would entail different reference populations for the starch, crisp (potato chip) processing, French fry processing, and fresh market classes. These reference populations would also need to be grown in appropriate conditions, such as tropical, temperate, Mediterranean, and hot, dry production environments, for the identification of appropriate locally adapted genetic backgrounds. The ultimate reference population could include germplasm from all of these subpopulations for each target, as well as germplasm from related species that could be used for the introgression of genes of interest. However, this latter reference population would require a very large number of markers to assess the extent of genetic variation, because of the rapid LD decay in potato across such populations.

The diversity of the target population will largely dictate the size of the reference required to achieve adequate genomic prediction accuracy. More diverse populations will require a larger reference population, as well as denser markers, given the lower extent of LD. Programs specific to particular breeding targets could be based on smaller reference populations than those programs aimed at multiple end uses. The reference population should include the desired phenotypic variation (Jannink et al., 2010).

The heritability of the traits of interest will also affect the size required of a reference population to achieve accurate genomic prediction. When breeding populations were analyzed allowing for tetrasomic inheritance, as compared to the (incorrect) assumption of disomic inheritance, the two analyses produced very similar results (Slater et al., 2014b) and therefore patterns of inheritance should not be a major confounding factor. A more influential effect will be the heritability of the traits themselves, which were found to vary from highly heritable (0.83) to those with low heritability (0.21). Those traits with low heritability will require a much larger reference population to obtain meaningful prediction accuracies, although population sizes of only 952 to 1137 individuals were large enough to improve the accuracy using the estimated breeding values derived from pedigrees (Slater et al., 2014b).

### Accuracy of Genomic Selection and Expected Genetic Gain

The accuracy of genomic selection can be predicted using population and trait parameters (Daetwyler et al., 2010b, 2008; Erbe et al., 2013; Goddard, 2009). We have used the formula of Daetwyler et al. (2010b) (Eq. [10]) to estimate the accuracy of genomic selection ( $r$ ) in this potato pedigree (Table 2):

$$r = \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}}, \quad [10]$$

**Table 2. Number of lines in each generation of potato, estimates of cumulative inbreeding, rate of inbreeding, and effective population size ( $N_e$ ) for 0.0 and 0.25 double reduction (DR).**

Generation	Number of lines	Inbreeding		Inbreeding rate		$N_e$	
		No DR	0.25 DR	No DR	0.25 DR	No DR	0.25 DR
1	97	NA†	0.25	NA	NA	NA	NA
2	196	0	0.319	0.000	0.092	NA	NA
3	1797	0.029	0.366	0.029	0.069	17.2	7.2
4	1218	0.034	0.376	0.005	0.016	97.1	31.7
5	396	0.042	0.384	0.008	0.013	60.4	39.0
Total	3704						

† NA, Not available

where  $N_p$  is the number of phenotyped and genotyped individuals in the reference population,  $h^2$  is the heritability of the trait, and  $M_e$  is the number of independent chromosome segments. There are many potential derivations of  $M_e$  and it is unclear which is most suitable. The two derivations most often applied are Eq. [11] Hayes et al., 2009b) and Eq. [12] (Goddard, 2009):

$$M_e = 2N_e L; \quad [11]$$

$$M_e = \frac{2N_e L}{\ln(4N_e L)}, \quad [12]$$

where  $L$  is the length of the genome in Morgans. There is evidence that the first derivation is more appropriate for populations with a high degree of relatedness (Wientjes et al., 2013), whereas the second is better suited to less related populations (Clark et al., 2012). Results are only presented for the Eq. [11] approximation, which is expected to result in more conservative estimates of the accuracy than are achievable with genomic selection (Table 2). The assumed genome length ( $L$ ) was double that of the 9.36-Morgans diploid linkage map reported in Sharma et al. (2013) (i.e., 18.72 Morgans) in order to account for the constitution of autotetraploid populations.

We have estimated the effective population size of our potato breeding population, which was breeding for fresh, crisp, and French fry processing (to calculate the size of the reference population required for genomic selection) by calculating the inbreeding level per generation in a five-generation pedigree going back to the early 1900s. A numerator relationship matrix (*NRM*) was calculated with a proportion of DR of 0.0 or 0.25 caused by autotetraploidy, as in Slater et al. (2014b). Inbreeding ( $F$ ) was approximated from the diagonal of this relationship matrix as  $F = NRM_{jj} - 1$ , for individual  $j$ . The effective population size was then estimated from the rate of inbreeding ( $dF$ ) from one generation to the next (Falconer and Mackay, 1996):

$$N_e = \frac{1}{2 \times dF}. \quad [13]$$

Inbreeding levels and  $N_e$  estimates per generation are listed in Table 2. There are five generations in the pedigree, which overlapped because of the use of parents in multiple generations; therefore, our separation into discrete generations only resulted in approximations of  $dF$  and  $N_e$ . There is a wealth of published studies on the estimation of  $N_e$  in pedigree and genetic marker data (Charlesworth, 2009; Hayes et al., 2003; Leroy et al., 2013). The relationships and degree of inbreeding of the foundation germplasm were unknown; subsequent generations showed increasing inbreeding, although the rate slowly decreased between Generations 3 and 5. This indicates that the greatest breeding bottleneck occurred in Generation 2 and that subsequent selections produced smaller effects on levels of diversity. Accounting for 0.25 DR with a tetraploid numerator relationship matrix dramatically increased inbreeding but the rate of inbreeding also followed a decreasing trend after Generation 3. Estimates of  $N_e$  were lowest in Generation 3 and highest in Generations 4 and 5, depending on the level of DR. As expected from the higher inbreeding levels obtained with the 0.25-DR scenarios, the  $N_e$  in these instances was lower than when assuming no DR. We could have used the harmonic  $N_e$  mean across all three generations ( $N_e \sim 36$ ) for predictions. However, to be conservative in our predictions of accuracy, we chose the mean of Generation 4 and 5 of the no-DR scenarios in our equation ( $N_e \sim 79$ ). Although the ancestral  $N_e$  of potato was large, the recent  $N_e$  in this population was much smaller and comparable to that of cattle breeds (e.g., Holstein, Jersey) in which genomic selection has been successfully implemented. It has been shown that the recent  $N_e$  is the most relevant indicator of diversity for genomic prediction accuracy (Hayes et al., 2009b; Luan et al., 2012).

The predicted accuracy using Eq. [11] ranged from 0.19 to 0.77, depending on  $h^2$  and reference population size (Table 3). The accuracies identified were encouraging, even for relatively small reference populations and traits of low heritability. Moderate accuracy, though insufficient to select a small number of elite cultivars, would be capable of distinguishing the top 50% of selection candidates and substantially increase the likelihood of success. The level of prediction accuracy would allow for a gradual progression toward a full-scale genomic selection program, in which the general breeding pool would be initially selected on the basis of genomic predictions from relatively small reference populations that leverage historical cultivar-specific information. Although this historical phenotypic information would provide good information on each cultivar, it would not indicate the superior additive genetics that could be identified from progeny testing or breeding values. The reference populations could be further enhanced through additional targeted genotyping and phenotyping to increase accuracy and, in turn, permit prediction of elite parents.

**Table 3. Predicted accuracy of genomic selection for four traits of potato, in which BVP is breeders' visual preference,  $h^2$  is narrow-sense heritability, Acc BLUP is the prediction accuracy of pedigree best linear unbiased prediction, and  $N_p$  is the reference population size.**

Trait	$h^2$ †	Acc BLUP‡	$N_p = 500$	$N_p = 1000$	$N_p = 2000$	$N_p = 5000$
BVP	0.23	0.33	0.19	0.27	0.37	0.53
Yield	0.56	0.19	0.29	0.40	0.52	0.70
Boiling color	0.73	0.44	0.33	0.44	0.57	0.74
Maturity	0.86	0.78	0.36	0.47	0.61	0.77

† Data from Slater et al. (2014b).

‡ Data from Slater et al. (2014b), reflecting the fact that only a subset of the pedigree was phenotyped for some traits.

## Strategies for Including Genomic Selection in Potato Breeding

### Rapid Improvement of Breeding Germplasm for Currently Phenotyped Traits

Genomic selection would readily fit with current potato breeding activities and could be described in terms of three distinct activities, being the tasks required to establish genomic selection, the genomic selection cycle itself, and routine phenotypic testing (including that for maintaining the accuracy of genomic selection) for variety development (Fig. 1).

The establishment activities include assembly of a reference population of the genotyped germplasm with phenotypes for the target traits that are relevant to the breeding program. The lowest risk option would be to use relevant historical cultivars and potential parents to ensure that the prediction accuracies would be high. This strategy has worked remarkably well in breeds of dairy cattle, which exhibit similar levels of population diversity to the potato population shown above. This strong focus will also reduce the size of the reference population that is required (e.g., Table 3). As a potato breeding program has access to a large amount of historic phenotypic data on cultivars, which are propagated clonally, these existing cultivars would simply require genotypic analysis to develop the algorithms to correlate the genotypic and phenotypic data. Further phenotyping could be undertaken, when necessary, to improve the phenotypic information and to increase prediction accuracy. New phenotyping would be on individuals rather than progeny tests, which is more cost-effective in most traits unless  $h^2$  is low.

The algorithms could then be used to predict which cultivars should be used as parents to create the next breeding populations.

The routine genomic selection cycle would see crosses made, seed germinated, and the seedlings grown in the glasshouse. Leaf samples from the seedling population could then be taken for genomic analysis to identify which seedlings should then be used as parents. As potato breeding programs typically create a number of full-sib families within the breeding population, genomic information from parents, full-sibs, and part-sib families

will also help to rapidly identify the superior 10% of individuals to use as parents for the next cycle. This genomic data from the first year could be used with the prediction equations to predict a future phenotype that would otherwise take a number of years to attain through conventional phenotyping trials. The best crossing strategy for the selected parents that yields the maximum genetic gain should be investigated using stochastic simulation of breeding options.

If the analysis can be conducted fast enough, the superior seedlings could be potted up to allow these plants to flower in their first growth cycle. If not, it would be necessary to wait for a second growth cycle before these genotypes could be used as parents. As a potato generation can be completed rapidly in the glasshouse, two cycles could easily be completed within 1 yr. This rapid cycling would therefore allow rapid enrichment of desirable genes for the 40+ traits that are currently the focus of breeding programs, while using the same marker resource. This early prediction of all desirable genetic effects would permit the identification of the best combination for the majority of traits. The strategy avoids the disadvantage of early conventional phenotyping or MAS screening for only a few traits, which neglects the possibility that seedlings may also harbor other very beneficial characteristics. Therefore, the prediction of data would not just be for the genes of major effect, such as those controlling qualitative disease resistance, but also for the traits that are controlled by a large number of small effects, such as yield. Therefore, genomic selection should concentrate on the improvement of superior cultivars for a range of traits.

The genomic data would also identify which seedlings should be taken to the field for routine phenotypic testing and selection. The superior seedlings would proceed to the field, albeit in smaller numbers (20–50%), as only the predicted superior seedlings need to proceed, thus providing savings to the program through the ability to conduct smaller field trials (Table 4). Selection rates in the G1 stage would then be higher, only eliminating very poor phenotypes, resulting in ~80% retention for better phenotypic evaluation at the G2 stage. They would then continue for successive cycles of routine phenotypic assessment and confirmation of the superior cultivars. Although the phenotypic assessment may be through conventional methods, selection of superior cultivars should be much faster, as the conventional data will be supported by the genomic data. This phenotypic assessment could then be used to refine the algorithms correlating the phenotypic and genomic data (Fig. 1). If the field evaluation reveals that a superior potato line has been missed when selecting parents in the rapid genomic selection cycles, it can still be added as a parent in the next cycle.

As the adoption of new techniques requires additional financial outlay by a breeding program, we have modeled two genomic selection programs to determine when genomic selection could be cost-effectively applied. As the genomic selection algorithms would identify the



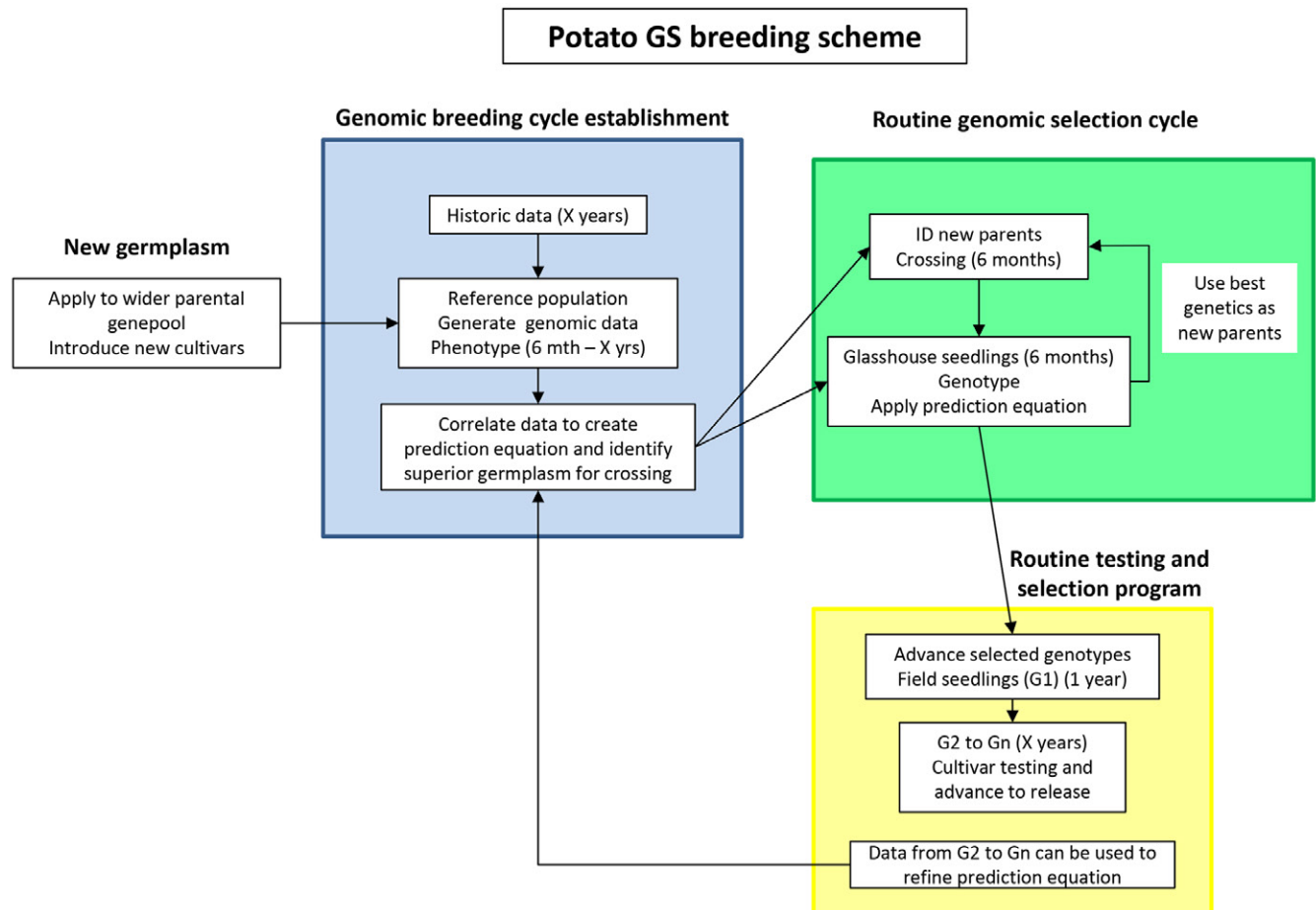


Fig. 1. Genomic selection (GS) scheme for potato illustrating the establishment activities, the routine genomic selection cycle, the routine testing and selection program, and how new germplasm would be introduced into the scheme.

**Table 4. Comparison of the cost of potato breeding under three selection regimes with two genomic selection models.**

Generation	Details	Intense	Moderate	Mild	GS 5000¶	GS 2000¶
G0	No. of seedlings†	100,000	20,000	6,667	5,000	2,000
	Seedling tuber production	\$300,000	\$60,000	\$20,001	\$15,000	\$6,000
	G0 selection rate	0%	0%	0%	50%	50%
G1	No. of seedlings	100,000	20,000	6,667	2,500	1,000
	G1 cost‡	\$164,731	\$32,946	\$10,983	\$32,946	\$10,983
	G1 selection rate	2%	10%	30%	80%	80%
G2	No. of genotypes	2,000	2,000	2,000	2,000	800
	G2 cost§	\$119,552	\$119,552	\$119,552	\$119,552	\$47,821
Total cost		\$633,040	\$257,248	\$194,618	\$138,671	\$55,468
Cost savings with GS 5000		489,453	117,668	55,705	–	–
Breakeven genotyping cost per genotype		98	24	11	–	–
Cost savings with GS 2000		572,656	200,870	138,908	–	–
Breakeven genotyping cost per genotype		286	100	69	–	–

† The number of seedlings varies because the selection intensity applied at G1 to arrive at 2000 genotypes in the G2 generation apart from the GS 2000 model.

‡ The G1 trial cost (in Australian dollars) is based on the cost per hectare taken from Slater et al. (2013).

§ The G2 trial cost (in Australian dollars) is based on the cost per hectare taken from Slater et al. (2013).

¶ GS = genomic selection seedling population size.

parents for the crossing program, the creation of breeding populations would be smaller and more targeted. Savings would also be found through smaller phenotyping trials and reduce the need to repeat trials to confirm

results. These savings have allowed us to model the breakeven costs for genotyping, in order for genomic selection to be cost-effective (Table 4). Further savings that we did not model would be gained in the reduced

number of trials needed to identify superior cultivars. In addition, revenue would arise from the superior genetic gain identified through genomic selection.

Most genomic selection breeding programs have focused on additive effects, as these are readily passed to offspring. The widespread use of clonal propagation in potato makes the prediction of nonadditive effects such as dominance and epistasis attractive, because they are fixed in the clonal cultivar. Prediction of the final cultivar phenotype will be more accurate if it were to include nonadditive effects; this could reduce evaluation time substantially. Furthermore, crosses could be guided to maximize the additive and nonadditive genetic merit in potential offspring.

Establishing the reference population in different production environments may identify germplasm with adaptation to different local conditions, but could also lead to the identification of germplasm that performs well over different production environments. Analysis of genotype  $\times$  environment effects would permit the identification of widely adapted germplasm. This will allow a consistent supply of the same cultivar from various growers to processing factories, so that the factories do not need to change production parameters, and a consistent supply to fresh market outlets, so that the retailers can provide familiar cultivars to consumers. Compiling different reference populations in multiple locations will quickly increase cost. Therefore, to save expense, the same set of genotyped individuals would be phenotyped in multiple locations. Genomic prediction would explicitly model genotype  $\times$  environment effects and a specific prediction equation could be developed for each location (Crossa et al., 2010; Heslot et al., 2014; Jarquín et al., 2014; Resende et al., 2012). We have predicted the expected genetic gain per generation ( $dG$ ) using the equation  $dG = i \times acc \times SD(g)$ , where  $i$  is the selection coefficient,  $acc$  is the accuracy of selection, and  $SD(g)$  is the genetic SD (Falconer and Mackay, 1996). The accuracy of phenotypic selection was approximated as  $\sqrt{h^2}$ , whereas the accuracy of BLUP was taken directly from the analyses in Slater et al. (2014b) using the same pedigree that was used for estimation of  $N_e$ . This comparison may underestimate the value of BLUP because not all individuals were phenotypically assessed, leading to an accuracy of  $<\sqrt{h^2}$ . To predict genetic gain from genomic selection, the selection coefficient was assumed to be 1.755, which reflects the level of 10% for selection candidates to be chosen as parents (Falconer and Mackay, 1996). The genetic SD [ $SD(g)$ ] was assumed to be 1 for all traits, making them comparable. As a simplification, we assumed that the size of the reference population stayed constant over time. However, additional phenotyping would occur in the routine testing and selection program (Fig. 1) and these phenotypes would update the reference population, increasing the accuracy of genomic selection. However, there will be a delay, as phenotypic information will only be gathered after G2 (or later), which results in at least two (or more) genomic selection cycles occurring before

the feedback. This means that recombination events would have occurred, thereby slowly deteriorating information from older reference population entries and limiting accuracy. In Holstein cattle, with a similar  $N_e$  to our potato estimates, the reduction in accuracy from predicting progeny from great-grandparents rather than parents was  $\sim 0.1$  to  $0.2$  (Habier et al., 2010). We have accounted for this in our prediction of genetic gain by subtracting  $0.1$  from all genomic selection accuracies in Table 3. Yield is only reliably measured at the G3 stage and hence we discounted its predicted accuracy by  $0.2$ . It is clear that genomic selection can lead to higher levels of genetic gain than either BLUP or phenotypic selection. When the reference population size is lower (i.e.,  $N_p = 500$ ), the genetic gain per cycle of genomic selection is less than for BLUP and phenotypic methods (Fig. 2). However, the cycle time is greatly accelerated, leading to more genetic gain per year and over time. Factoring in an increase in the reference population over time may increase the superiority of genomic selection over BLUP. Genetic gain is expected to be higher for traits with higher heritability (Fig. 3). If a low-heritability trait (e.g., breeders' visual preference) is judged to be very important, it may, to a large extent, dictate the size of the reference populations that are required to achieve adequate gain.

We have used a linear prediction of genetic gain where the rate of gain stays constant over time. In the long-term, the rate of gain would slow because selected loci increase in frequency and slowly explain a smaller proportion of the genetic variance, as demonstrated in several simulation studies (Goddard, 2009; Jannink, 2010). This would also be the case in potato. Our linear predictions of genetic gain are deterministic and do not account for the change in allele frequencies caused by selection. The main way to guard against reduced long-term gain is to put more weight on rare alleles to drive them to intermediate frequencies, where they account for more of the genetic variance. This can be accomplished by explicitly up-weighting rare alleles using an index and by fitting a pedigree effect in genomic selection models to account for variations that are not captured by genetic markers (Goddard, 2009; Hayes et al., 2009a; Jannink, 2010). Additional research using stochastic simulation will be necessary to model all these factors in more detail.

### Rapid Introduction of New Parents and Genes from Wild Relatives

There are currently 110 wild species recognized in a recent classification of tuber-bearing *Solanum* species (Ovchinnikova et al., 2011; Spooner et al., 2009) and four cultivated species (Ovchinnikova et al., 2011; Spooner et al., 2007). These species are found in a wide diversity of geographic locations and climates. Some are found close to the vegetation limits in the high Andean regions, which will be subjected to freezing temperatures; others originate from semidesert regions that are subjected to high temperatures and drought; and others are found in temperate conditions or rainforests. These wild relatives

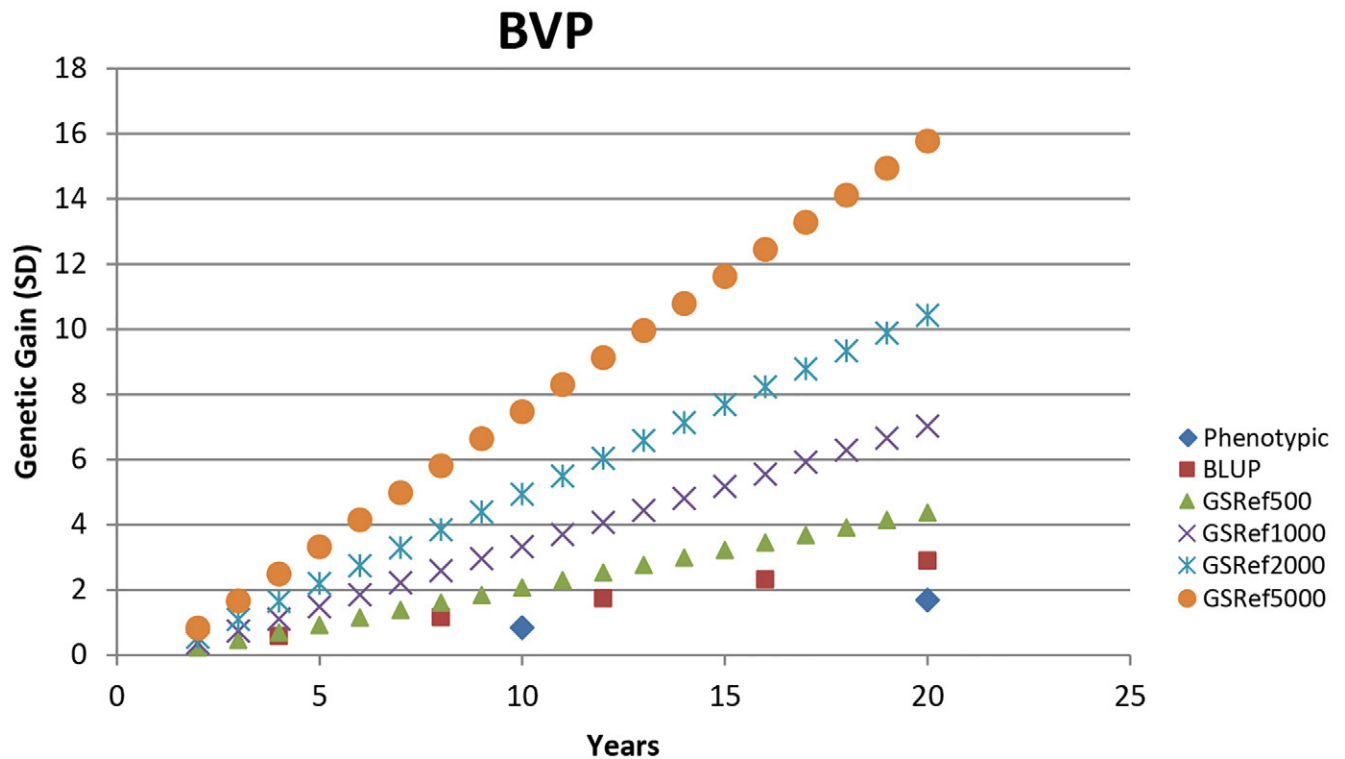


Fig. 2. Expected genetic gain in genetic SD of potato over a 20-yr period of phenotypic, best linear unbiased prediction (BLUP), and genomic selection for breeders' visual preference (BVP) when 10% of candidates are selected as parents for the next cycle. GS ref, genomic selection reference population size.

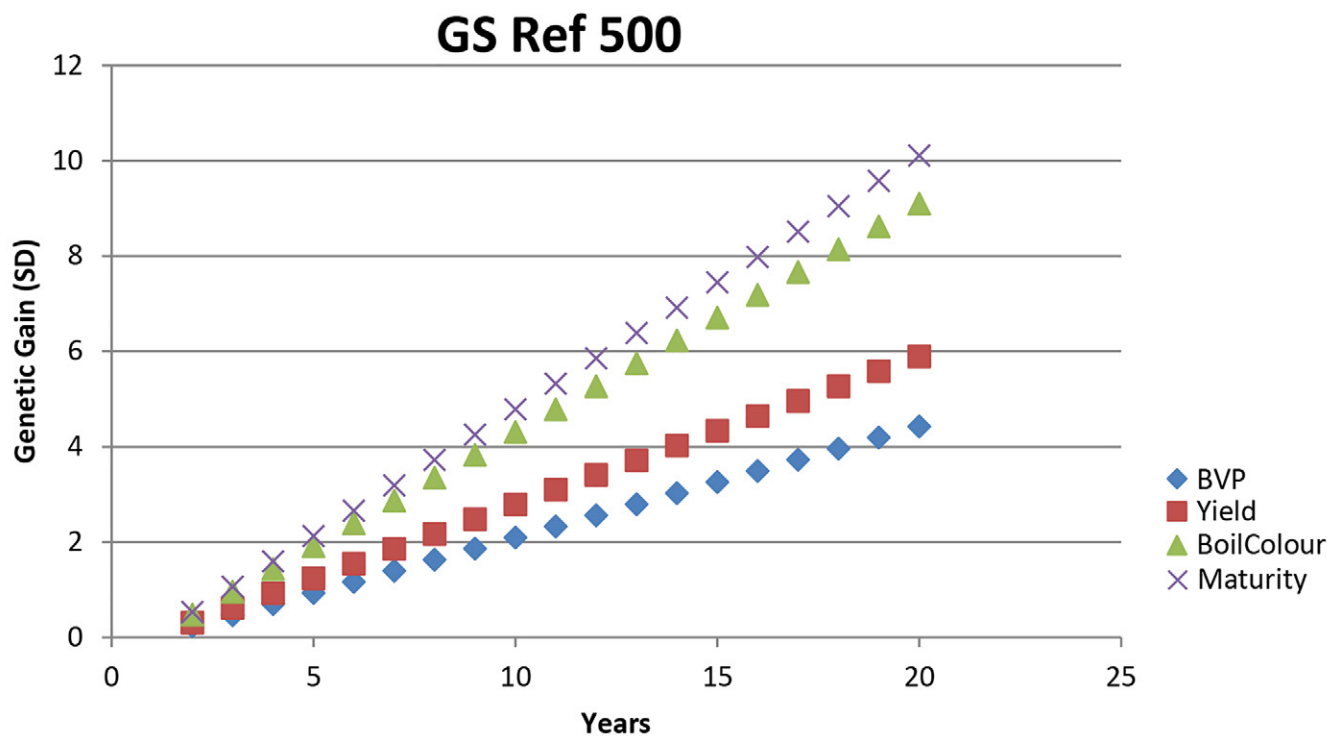


Fig. 3. Expected genetic gain in genetic SD in potato for breeders' visual preference (BVP), yield, boiling colour, and maturity using genomic selection with a reference population size of 500 over a 20-yr period.

will contain genes for numerous traits that are not present in commercial cultivars and represent a rich source for improvements in disease resistance, abiotic stress

resistance, and tuber quality characteristics (Hanneman, 1989; Jansky, 2009; Spooner and Bamberg, 1994) and are consequently a major resource for potato breeding

(Hawkes, 1994). The introgression of genes from wild species began in 1909 and has been limited, although significant, for the introduction of resistance genes against pests and pathogens such as late blight, viruses, and cyst nematodes (Bradshaw, 2007). However, introgression of desirable genes from wild species will require many generations of breeding to develop a commercial cultivar.

Similarly, potato breeding programs may have access to a wide array of potential parents with desirable characteristics from outside their program. The inclusion of both sources of variation into a genomic selection breeding program would be quite similar. In essence, the establishment phase needs to be completed for this new material, just as the initial reference population had to be evaluated to initiate genomic selection (Fig. 1). New material could be phenotyped and genotyped along with the reference population to identify whether desirable genetic variants could be introduced into the scheme (Fig. 1). Depending on the level of similarity to the reference population, it may be necessary to create seedlings from crosses to increase the number of plants that can be phenotypically evaluated and added to the reference population, allowing for more accurate prediction of new attributes. This would also allow for a statistical validation to ensure that the performance of new lines could be adequately predicted. Eventually, it should be possible to predict how effectively genomic selection can predict the performance of new lines from their genomic similarity to the reference population. If new commercial lines are moderately similar, it may be possible to initially predict them on their genotype alone. Addition of germplasm from a different breeding pool is expected to be simpler, as these lines would be relatively genetically similar to the reference population compared to wild germplasm.

The process for including attributes from wild relatives will differ depending on the genetic architecture of the trait. For quantitative traits, where genetic variation is contributed by many loci, the process described above for new cultivars would also apply. In such cases, rapid genomic selection cycling of seedlings would rapidly incorporate the desirable variation. If major QTLs or genes contribute a large proportion of the trait, more targeted and optimized introgression algorithms leading to low levels of linkage drag could be combined with genomic selection. Rather than taking many generations of ~10 yr in a conventional program, genomic selection would enable more accurate introgression in a much shorter time frame.

### *Increasing the Suite of Traits That Can Be Evaluated or Selected*

Genomic selection will also allow expansion of the range of traits that can be improved. These might include traits that are evaluated late in the breeding process or that have recently become a priority, such as low acrylamide levels, which have been the subject of recent attention (Bethke and Bussan, 2013) or they could be new abiotic

or biotic stresses including drought, heat, salinity, or specific diseases. The reference population could also examine traits for identification of more productive genotypes, such as efficient nutrient or water use. In this strategy, a reference population could be developed for a subset of genotypes to develop specific prediction equations for the new trait of interest. These specific predictions could then be incorporated within the larger set to develop a selection index that covers the expanded suite of traits. This expansion of desirable traits would also provide the advantage of being chosen from the combined set before the breeding pool is reduced by selection for a narrower set of traits. This strategy is attractive when the new trait is expensive to measure, as it allows for targeted phenotypic assessment of key lines to maximize prediction accuracy to apply to genotyping seedlings.

Genomic selection may also be enhanced by high-throughput phenomics, which are expected to result in a large number of indicator traits that are substantially correlated with the characteristics used in conventional potato breeding. This 'systems biology continuum' extends through the transcriptome for global RNA synthesis, the proteome for total protein content, the interactome for protein-protein interactions, the glycome and ionome for carbohydrates and small charged molecules respectively, the hormonome for phytohormone signaling, and the metabolome for metabolite analysis for understanding and measuring phenotypes (Mochida and Shinozaki, 2011). The continuum further extends into phenomics for whole plants, based on real-time measurements of plant growth and development using glasshouse-based imaging facilities (Furbank, 2009; Furbank and Tester, 2011; Houle et al., 2010) or through in-field use of imaging equipment (Araus and Cairns, 2014). Metabolomic methods would enable the sugar and starch profiles of tubers to be more closely examined, whereas high-throughput imaging will allow plant biomass accumulation to be monitored under various biotic and abiotic stress conditions. The development of high-throughput phenomic platforms will enable phenotype data to be captured at a rate to match the advances that have occurred with genomic selection (Cobb et al., 2013; Mackay et al., 2015), although there will be a requirement to optimize approaches for individual plant species. In this context, a focus on species-specific traits for potato, particularly subterranean stolon and tuber characters, will be essential.

## **Conclusion**

Genomic selection will enable accelerated genetic gain for a combination of traits, including those under complex genetic control. The  $N_e$  of a representative potato breeding program is comparable to that of Holstein cattle, for which genomic selection has been successfully implemented in many countries. Conservatively predicted genomic selection accuracies ranged from 0.2, under conditions of low  $h^2$  and small reference populations, to 0.8 in larger reference populations. Less than 20,000 SNPs would be required to achieve useful



predictions of genetic merit that could be applied in rapid breeding cycles to substantially increase genetic gain compared to phenotype- or even BLUP-based selection. Finally, many ancestral clones are still available, which greatly simplifies the mining of historical phenotype data for assembly of genomic selection reference populations. Taken together, these factors make genomic selection in potato thoroughly feasible and attractive.

## Acknowledgments

This work was supported by funding from the Victorian Department of Economic Development, Jobs, Transport, and Resources.

## References

- Anderson, J., and H. Howard. 1981. Effectiveness of selection in the early stages of potato breeding programmes. *Potato Res.* 24:289–299. doi:10.1007/BF02360366.
- Araus, J.L., and J.E. Cairns. 2014. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19:52–61. doi:10.1016/j.tplants.2013.09.008.
- Ashraf, B.H., J. Jensen, T. Asp, and L.L. Janss. 2014. Association studies using family pools of outcrossing crops based on allele-frequency estimates from DNA sequencing. *Theor. Appl. Genet.* 127:1331–1341. doi:10.1007/s00122-014-2300-4.
- Bakker, E., T. Borm, P. Prins, E. van der Vossen, G. Uenk, M. Arens, et al. 2011. A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theor. Appl. Genet.* 123:493–508. doi:10.1007/s00122-011-1602-z.
- Beavis, W.D. 1998. QTL analysis: Power, precision, and accuracy. In: A.C. Paterson, editor, *Molecular dissection of complex traits*. CRC Press, New York. p. 145–162.
- Bethke, P.C., and A.J. Bussan. 2013. Acrylamide in processed potato products. *Am. J. Potato Res.* 90:403–424. doi:10.1007/s12230-013-9321-4.
- Bonierbale, M.W., R.L. Plaisted, and S.D. Tanksley. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120:1095–1103.
- Bradshaw, J.E. 2007. Potato-breeding strategy. In: D. Vreugdenhil, J. Bradshaw, C. Gebhardt, et al., editors, *Potato biology and biotechnology: Advances and perspectives*. Elsevier, Amsterdam. p. 157–177.
- Bradshaw, J.E., M.F.B. Dale, and G.R. Mackay. 2003. Use of mid-parent values and progeny tests to increase the efficiency of potato breeding for combined processing quality and disease and pest resistance. *Theor. Appl. Genet.* 107:36–42.
- Bradshaw, J.E., M.F.B. Dale, and G.R. Mackay. 2009. Improving the yield, processing quality and disease and pest resistance of potatoes by genotypic recurrent selection. *Euphytica* 170:215–227. doi:10.1007/s10681-009-9925-4.
- Bradshaw, J.E., and G.R. Mackay. 1994. Breeding strategies for clonally propagated potatoes. In: J.E. Bradshaw and G.R. Mackay, editors, *Potato genetics*. CAB International, Wallingford. p. 467–497.
- Bradshaw, J.E., and G. Ramsey. 2009. Potato origin and production. In: J. Singh and L. Kaur, editors, *Advances in potato chemistry and technology*. Academic Press, New York. p. 1–26.
- Brown, J., P.D.S. Caligari, M.F.B. Dale, G.E.L. Swan, and G.R. Mackay. 1988. The use of cross prediction methods in a practical potato breeding programme. *Theor. Appl. Genet.* 76:33–38. doi:10.1007/BF00288828.
- Brown, J., P.D.S. Caligari, G.R. Mackay, and G.E.L. Swan. 1984. The efficiency of seedling selection by visual preference in a potato breeding programme. *J. Agric. Sci.* 103:339–346. doi:10.1017/S0021859600047298.
- Brown, J., P.D.S. Caligari, G.R. Mackay, and G.E.L. Swan. 1987. The efficiency of visual selection in early generations of a potato breeding programme. *Ann. Appl. Biol.* 110:357–363. doi:10.1111/j.1744-7348.1987.tb03266.x.
- Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos, and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178:553–561. doi:10.1534/genetics.107.080838.
- Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10:195–205. doi:10.1038/nrg2526.
- Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J.H.J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* 44:1–9. doi:10.1186/1297-9686-44-4.
- Cobb, J., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126:867–887. doi:10.1007/s00122-013-2066-0.
- Crossa, J., G.I. Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724. doi:10.1534/genetics.110.118521.
- D’hoop, B., P.C. Keizer, M.J. Paulo, R.F. Visser, F. Eeuwijk, and H. Eck. 2014. Identification of agronomically important QTL in tetraploid potato cultivars using a marker–trait association analysis. *Theor. Appl. Genet.* 127:731–748. doi:10.1007/s00122-013-2254-y.
- D’hoop, B., M. Paulo, K. Kowitwanich, M. Sengers, R. Visser, H. van Eck, et al. 2010. Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor. Appl. Genet.* 121:1151–1170. doi:10.1007/s00122-010-1379-5.
- Daetwyler, H.D., J.M. Hickey, J.M. Henshall, S. Dominik, B. Gredler, J.H.J. van der Werf, et al. 2010a. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim. Prod. Sci.* 50:1004–1010. doi:10.1071/AN10096.
- Daetwyler, H.D., R. Pong-Wong, B. Villanueva, and J.A. Woolliams. 2010b. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031. doi:10.1534/genetics.110.116855.
- Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi:10.1371/journal.pone.0003395.
- Dalla Rizza, M., F. Vilaró, D. Torres, and D. Maeso. 2006. Detection of PVY extreme resistance genes in potato germplasm from the Uruguayan breeding program. *Am. J. Potato Res.* 83:297–304. doi:10.1007/BF02871590.
- Dong, F., J. Song, S.K. Naess, J.P. Helgeson, C. Gebhardt, and J. Jiang. 2000. Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* 101:1001–1007. doi:10.1007/s001220051573.
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 123:307–326.
- Erbe, M., B. Gredler, F.R. Seefried, B. Bapst, and H. Simianer. 2013. A function accounting for training set size and marker density to model the average accuracy of genomic prediction. *PLoS One* 8:e81046. doi:10.1371/journal.pone.0081046.
- Erbe, M., B.J. Hayes, L.K. Matukumalli, S. Goswami, P.J. Bowman, C.M. Reich, et al. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95:4114–4129. doi:10.3168/jds.2011-5019.
- Falconer, D.S., and T.F.C. Mackay. (1996) *Introduction to quantitative genetics*. 4th ed. Pearson, Prentice Hall, Harlow, UK.
- Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton, R.E. Veilleux, et al. 2012. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One* 7:e36347. doi:10.1371/journal.pone.0036347.
- Fischer, R.A., and G.O. Edmeades. 2010. Breeding and cereal yield progress. *Crop Sci.* 50:S-85–S-98. doi:10.2135/cropsci2009.10.0564.
- Fu Y.B. (2014) Genetic diversity analysis of highly incomplete SNP genotype data with imputations: An empirical assessment. *G3 (Bethesda)* 4:891–900. doi:10.1534/g3.114.010942.
- Furbank, R.T. 2009. Plant phenomics: From gene to form and function. *Funct. Plant Biol.* 36:v–vi. doi:10.1071/FPv36n11\_FO.

- Furbank, R.T., and M. Tester. 2011. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16:635–644. doi:10.1016/j.tplants.2011.09.005.
- Gebhardt, C. 2013. Bridging the gap between genome analysis and precision breeding in potato. *Trends Genet.* 29:248–256. doi:10.1016/j.tig.2012.11.006
- Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schöler. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breed.* 13:93–102. doi:10.1023/B:MOLB.0000012878.89855.df.
- Gebhardt, C., E. Ritter, A. Barone, T. Debener, B. Walkemeier, U. Schachtschabel, et al. 1991. RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor. Appl. Genet.* 83:49–57. doi:10.1007/BF00229225
- Gebhardt, C., E. Ritter, T. Debener, U. Schachtschabel, B. Walkemeier, H. Uhrig, et al. 1989. RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor. Appl. Genet.* 78:65–75. doi:10.1007/BF00299755
- Gidskehaug, L., M. Kent, B.J. Hayes, and S. Lien. 2011. Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics* 27:303–310. doi:10.1093/bioinformatics/btq673.
- Gilmore, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. (2009) ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257. doi:10.1007/s10709-008-9308-0.
- Grattapaglia, D., M.D. Vilela Resende, M.R. Resende, C.P. Sansaloni, C.D. Petrolí, A.A. Missiaggia, et al. 2011. Genomic selection for growth traits in *Eucalyptus*: Accuracy within and across breeding populations. *BMC Proc.* 5:1–2. doi:10.1186/1753-6561-5-S7-A1.
- Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. doi:10.1534/genetics.107.081190.
- Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:1–12. doi:10.1186/1297-9686-42-5.
- Hamilton, J., C. Hansey, B. Whitty, K. Stoffel, A. Massa, A. Van Deynze, et al. 2011. Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics* 12:302. doi:10.1186/1471-2164-12-302
- Hanneman, R.E.J. 1989. The potato germplasm resource. *Am. J. Potato Res.* 66:655–667. doi:10.1007/BF02853985
- Hawkes, J.G. 1994. Origins of cultivated potatoes and species relationships. In: J.E. Bradshaw and G.R. Mackay, editors, *Potato genetics*. CAB International, Wallingford. p. 3–42.
- Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009a. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443. doi:10.3168/jds.2008-1646
- Hayes, B.J., H.D. Daetwyler, P.J. Bowman, G. Moser, B. Tier, R. Crump, et al. 2009b. Accuracy of genomic selection: Comparing theory and results. *Proc. Assoc. Adv. Anim. Breed. Gen.* 17:352–355.
- Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009c. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91:47–60. doi:10.1017/S0016672308009981.
- Hayes, B.J., P.M. Visscher, H.C. McPartlan, and M.E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13:635–643. doi:10.1101/gr.387103.
- Haynes, K., D. Gergela, C. Hutchinson, G. Yench, M. Clough, M. Henninger, et al. 2012. Early generation selection at multiple locations may identify potato parents that produce more widely adapted progeny. *Euphytica* 186:573–583. doi:10.1007/s10681-012-0685-1.
- Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447. doi:10.2307/2529430.
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127:463–480. doi:10.1007/s00122-013-2231-5
- Houle, D., D.R. Govindaraju, and S. Omholt. 2010. Phenomics: The next challenge. *Nat. Rev. Genet.* 11:855–866. doi:10.1038/nrg2897
- Huang, K., K. Ritland, S. Guo, D.W. Dunn, D. Chen, Y. Ren, et al. 2015. Estimating pairwise relatedness between individuals with different levels of ploidy. *Mol. Ecol. Resour.* 15:772–784. doi:10.1111/1755-0998.12351.
- Huang, K., K. Ritland, S. Guo, M. Shattuck, and B. Li. 2014. A pairwise relatedness estimator for polyploids. *Mol. Ecol. Resour.* 14:734–744. doi:10.1111/1755-0998.12217.
- Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961–967. doi:10.1038/ng.695
- Jan, H.U., A. Abbadi, S. Lücke, R.A. Nichols, and R.J. Snowdon. 2016. Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS One* 11:e0147769. doi:10.1371/journal.pone.0147769.
- Jannink, J.-L. 2010. Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42:1–11. doi:10.1186/1297-9686-42-35.
- Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genomics* 9:166–177. doi:10.1093/bfpg/elq001.
- Jansky, S. 2009. Breeding, genetics and cultivar development. In: J. Singh and L. Kaur, editors, *Advances in potato chemistry and technology*. Academic Press, New York. p. 27–62.
- Jarquín, D., J. Crossa, X. Lacaze, P. Cheyron, J. Daucourt, J. Lorgeou, et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127:595–607. doi:10.1007/s00122-013-2243-1.
- Jupe, F., L. Pritchard, G. Etherington, K. MacKenzie, P. Cock, F. Wright, et al. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* 13:75. doi:10.1186/1471-2164-13-75
- Jupe, F., K. Witek, W. Verweij, J. Śliwka, L. Pritchard, G.J. Etherington, et al. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 76:530–544. doi:10.1111/tpj.12307.
- Laurie, C.C., S.D. Chasalow, J.R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, et al. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168:2141–2155. doi:10.1534/genetics.104.029686.
- Leroy, G., T. Mary-Huard, E. Verrier, S. Danvy, E. Charvolin, and C. Danchin-Burge. 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genet. Sel. Evol.* 45:1297–9686. doi:10.1186/1297-9686-45-1
- Lin, Z., B.J. Hayes, and H.D. Daetwyler. 2014. Genomic selection in crops, trees and forages: A review. *Crop Pasture Sci.* 65:1177–1191. doi:10.1071/CP13363.
- Luan, T., J. Woolliams, J. Odegard, M. Dolezal, S. Roman-Ponce, A. Bagnato, et al. 2012. The importance of identity-by-state information for the accuracy of genomic selection. *Genet. Sel. Evol.* 44:28. doi:10.1186/1297-9686-44-28
- Mackay, I., E. Ober, and J. Hickey. 2015. GplusE: Beyond genomic selection. *Food Energy Secur.* 4:25–35 doi:10.1002/fes3.52.
- Maris, B. 1986. The effect of seed tuber weight on characters in the first and the second clonal generation of potato populations. *Euphytica* 35:465–482. doi:10.1007/BF00021855.
- Meuwissen, T.H.E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:35. doi:10.1186/1297-9686-41-35
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Milbourne, D., R.C. Meyer, A.J. Collins, L.D. Ramsay, C. Gebhardt, and R. Waugh. 1998. Isolation, characterisation and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.* 259:233–245. doi:10.1007/s004380050809
- Mochida, K., and K. Shinozaki. 2011. Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol.* 52:2017–2038. doi:10.1093/pcp/pcr153.
- Ortega, F., and C. Lopez-Vizcon. 2012. Application of molecular marker-assisted selection (MAS) for disease resistance in a practical potato breeding programme. *Potato Res.* 55:1–13. doi:10.1007/s11540-011-9202-5.

- Ottoman, R., D. Hane, C. Brown, S. Yilma, S. James, A. Mosley, et al. 2009. Validation and implementation of marker-assisted selection (MAS) for PVY resistance (*Ry<sub>adg</sub>* gene) in a tetraploid potato breeding program. *Am. J. Potato Res.* 86:304–314. doi:10.1007/s12230-009-9084-0.
- Ovchinnikova, A., E. Krylova, T. Gavrilenko, T. Smekalova, M. Zhuk, S. Knapp, et al. 2011. Taxonomy of cultivated potatoes (*Solanum* section *Petota*: Solanaceae). *Bot. J. Linn. Soc.* 165:107–155. doi:10.1111/j.1095-8339.2010.01107.x.
- Pembleton, L.W., N.O.I. Cogan, and J.W. Forster. 2013. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13:946–952. doi:10.1111/1755-0998.12129.
- Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195. doi:10.1038/nature10158.
- Pryce, J.E., S. Bolormaa, A.J. Chamberlain, P.J. Bowman, K. Savin, M.E. Goddard, et al. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J. Dairy Sci.* 93:3331–3345. doi:10.3168/jds.2009-2893.
- Resende, M.F.R., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, et al. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190:1503–1510. doi:10.1534/genetics.111.137026.
- Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44:217–220. doi:10.1038/ng.1033.
- Schultz, L., N.O.I. Cogan, K. McLean, M.F.B. Dale, G.J. Bryan, J.W. Forster, et al. 2012. Evaluation and implementation of a potential diagnostic molecular marker for *HI*-conferred potato cyst nematode resistance in potato (*Solanum tuberosum* L.). *Plant Breed.* 131:315–321. doi:10.1111/j.1439-0523.2012.01949.x.
- Sharma S.K., D. Bolser, J. de Boer, M. Sønderkær, W. Amoros, M.F. Carboni, et al. (2013) Construction of reference chromosome-scale pseudo-molecules for potato: Integrating the potato genome with genetic and physical maps. *G3 (Bethesda)* 3:2031–2047. doi:10.1534/g3.113.007153.
- Simko, I., K.G. Haynes, E.E. Ewing, S. Costanzo, B.J. Christ, and R.W. Jones. 2004. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. *Mol. Genet. Genomics* 271:522–531. doi:10.1007/s00438-004-1010-z.
- Simko, I., K.G. Haynes, and R.W. Jones. 2006. Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173:2237–2245. doi:10.1534/genetics.106.060905.
- Slater, A.T., N.O.I. Cogan, and J.W. Forster. 2013. Cost analysis of the application of marker-assisted selection in potato breeding. *Mol. Breed.* 32:299–310. doi:10.1007/s11032-013-9871-7.
- Slater, A.T., N.O.I. Cogan, B.J. Hayes, L. Schultz, M.F.B. Dale, G.J. Bryan, et al. 2014a. Improving breeding efficiency in potato using molecular and quantitative genetics. *Theor. Appl. Genet.* 127:2279–2292. doi:10.1007/s00122-014-2386-8.
- Slater, A.T., G.M. Wilson, N.O.I. Cogan, J.W. Forster, and B.J. Hayes. 2014b. Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. *Theor. Appl. Genet.* 127:809–820. doi:10.1007/s00122-013-2258-7.
- Spooner, D.M., M. Ames, D. Fajardo, and F. Rodriguez. (2009) Species boundaries and interrelationships of *Solanum* sect. *Petota* (wild and cultivated potatoes) are drastically altered as a result of PBI-funded research. Botany & Mycology 2009 Annual Meeting, Snowbird, UT. 25–29 July. Botanical Society of America, Granville, OH.
- Spooner, D.M., and J.B. Bamberg. 1994. Potato genetic resources: Sources of resistance and systematics. *Am. Potato J.* 71:325–337. doi:10.1007/BF02849059.
- Spooner, D.M., J. Nunez, G. Trujillo, M. del Rosario Herrera, F. Guzman, and M. Ghislain. 2007. Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proc. Natl. Acad. Sci. USA* 104:19398–19403. doi:10.1073/pnas.0709796104.
- Storey, M. 2007. The harvested crop. In: D. Vreugdenhil, J. Bradshaw, C. Gebhardt, et al., editors, *Potato biology and biotechnology: Advances and perspectives*. Elsevier, Amsterdam. p. 441–470.
- Tai, G., and D. Young. 1984. Early generation selection for important agronomic characteristics in a potato breeding population. *Am. J. Potato Res.* 61:419–434. doi:10.1007/BF02852811.
- Tanksley, S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, et al. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141–1160.
- Uitdewilligen, J.G.A.M.L., A.-M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser, and H.J. van Eck. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One* 8:e62355. doi:10.1371/journal.pone.0062355.
- Van Os, H., S. Andrzejewski, E. Bakker, I. Barrera, G.J. Bryan, B. Caromel et al. 2006. Construction of a 10,000-marker ultradense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* 173:1075–1087. doi:10.1534/genetics.106.055871.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980.
- VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, et al. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24. doi:10.3168/jds.2008-1514.
- Voorrips, R.E., G. Gort, and B. Vosman. 2011. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12:1–11. doi:10.1186/1471-2105-12-172.
- Wientjes, Y.C.J., R.F. Veerkamp, and M.P.L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193:621–631. doi:10.1534/genetics.112.146290.
- Wiggans, G.R., P.M. Vanraden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* 94:3202–3211. doi:10.3168/jds.2010-3866.
- Wolc, A., C. Stricker, J. Arango, P. Settar, J. Fulton, N. O'Sullivan, et al. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet. Sel. Evol.* 43:5. doi:10.1186/1297-9686-43-5.
- Wolc, A., H. Zhao, J. Arango, P. Settar, J. Fulton, N. O'Sullivan, et al. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet. Sel. Evol.* 47:59. doi:10.1186/s12711-015-0133-5.
- Xu, Y., Y. Lu, C. Xie, S. Gao, J. Wan, and B. Prasanna. 2012. Whole-genome strategies for marker-assisted plant breeding. *Mol. Breed.* 29:833–854. doi:10.1007/s11032-012-9699-6.
- Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569. doi:10.1038/ng.608.