Cell PRESS

Improving fruit and wine: what does genomics have to offer?

Sean Myles

Department of Plant and Animal Sciences, Faculty of Agriculture, Dalhousie University, Truro, Nova Scotia, Canada

Will we still be drinking wines made from Pinot Noir and eating McIntosh apples in the 23rd century? Elite grape and apple cultivars, vegetatively propagated for centuries, are highly susceptible to evolving pathogens. In response, growers continually expand their agrochemical weaponry at enormous environmental costs. By contrast, breeders are seeking disease-resistant, tastier alternatives to the handful of dominant cultivars by exploring genetic diversity in these fruits. However, this is a formidable task because consumers cling to ancient cultivars, and breeding long-lived woody perennials is laborious and expensive. Although genomics tools may not solve the former sociocultural dilemma, they can help overcome the latter practical obstacles. Screening seedlings for desirable genetic profiles using molecular techniques reduces the time and high costs associated with growing plants to maturity and evaluating fruit. Such screening is currently in its infancy in apples and grapes, but the adoption of modern DNA sequencing technologies and statistical approaches promises to accelerate cultivar improvement significantly. Here, I describe standard approaches for molecular breeding in apples and grapes, and some of the challenges associated with the collection and analysis of next-generation DNA sequence data. In addition, I urge breeders to establish populations specifically designed for a future of inexpensive genome sequencing.

What is wrong with our fruit?

Many wine drinkers consider the complex aromatic character of a fine Pinot Noir the pinnacle of human achievement. As a result, grape growers have preserved Pinot Noir for a millennium through the use of vegetative propagation [1], which immortalizes a cultivar by generating genetically identical copies from its tissue. The recent widespread celebrations of the 200th birthday of the McIntosh apple provide another demonstration of our quest to preserve perceived perfection. Indeed, humans often expect the same cultivars to be available for centuries and look upon ancient 'heritage' cultivars with veneration.

However, freezing woody perennial crops in genetic space for centuries through vegetative propagation has perhaps done more harm than good. Perpetual propagation threatens the very existence of elite cultivars because pathogens continue to evolve while these cultivars remain unchanged, necessitating the increasing use of agrochemicals to make up for static host defense systems. The problem is particularly acute in wine grapes: we drink from only a small fraction of their gene pool, and they have been suffering from a severe lack of sex. Grape growers in Bordeaux, for instance, rely almost exclusively on only three ancient grape cultivars for red wine production, ignoring the immense amount of potentially useful diversity in the genus *Vitis*. The apple (genus *Malus*) has experienced a similar, albeit less severe, plight. Ultimately, our intense attachment to ancient apple and grape cultivars has prevented us from asking a basic question: why not replace them with equally flavorful, if not even tastier, cultivars that require less chemical input to grow?

We need to free our fruits from the shackles of perpetual propagation and shuffle their genomes through sexual reproduction. A future with sustainable and safe food production relies on the generation of new food through breeding [2]. This is a unanimously held conviction among growers of most annual crops where seed improvements from year to year are expected and relied upon. Halting the evolution of our fruits by perpetual propagation and allowing a few cultivars to dominate the international market for extended periods of time not only robs us of access to potentially useful traits and fascinating flavors, but is also short sighted and unsafe. Breeding is not optional, it is a necessity. It is essential that we begin to recognize fully the threat posed by the gradual depletion of allelic richness in

Glossary

Background selection: the selection of offspring from a cross based on genetic ancestry estimates of the offspring.
Foreground selection: the selection of offspring from a cross based on the presence and/or absence of a particular allele in the offspring.
Genomic selection (GS): a form of MAS in which genetic markers covering the whole genome are used to explain phenotypic variation. The number of genetic markers should be high enough that all QTL are in linkage disequilibrium with at least one marker.
Imputation: the substitution of some value for missing data. Genotype imputation involves 'filling in' missing genotypes that were not directly assayed in a sample of individuals. This is achieved by matching the patterns of genotypes neighboring the missing genotype to patterns in other genotyped samples.
Introgression: the movement of a gene, locus, or allele from one species or population into the gene pool of another, generally by repeated backcrossing of an interspecific hybrid to one of the parental species.
Marker-assisted breeding (MAS): a breeding strategy that involves selecting offspring that carry a particular genetic marker that is linked to a trait of interest.
Reduced representation library (RRL): a library of DNA molecules used for sequencing that is generated by digesting DNA samples with one or more restriction enzymes. DNA sequences are only produced from digested DNA fragments within a specific size range so that data are only generated from a distinct subset (i.e., a reduced representation) of the genome.

Corresponding author: Myles, S. (sean.myles@dal.ca).
Keywords: apple; grape; marker-assisted selection; next-generation DNA sequencing.

 0168-9525/$ – see front matter http://dx.doi.org/10.1016/j.tig.2013.01.006 Trends in Genetics, April 2013, Vol. 29, No. 4

our food and the pivotal role played by breeders in the security of our future food supply.

The promise of genomics in apple and grape breeding

Apples and grapes are the focus here, but the principles presented are widely applicable to other long-lived perennial agricultural species. Consumer attachment to particular cultivars and the convenience offered by vegetative propagation are clearly important sociocultural factors that have contributed to a lack of breeding activity, but the development of new cultivars is also hindered because breeding them is laborious, time consuming, and expensive. Apples and grapes have a large plant size, long productive periods, an extended juvenile phase, and a final product that cannot be assessed until fruit bearing begins, which can take 3–7 years. A large proportion of the seedlings generated from a cross are culled within the first decade of fruit evaluation. The small proportion that shows promise is propagated into replicated trials, and an even smaller proportion of these may become commercial cultivars. For example, an apple-breeding program in Dresden-Pillnitz, Germany began with 52 000 seedlings and, after 26 years of evaluation, released only three of them as commercial cultivars [3].

These characteristics of apples and grapes also mean that they stand to benefit most from the genomics revolution. The ability to distinguish between a desirable and an undesirable genetic profile enables seedlings to be screened using molecular tools while they are small, before large amounts of time, space, and money have been allocated to growing them into mature plants to be evaluated. Moreover, the parents of each cross can be selected based on their genotypes and on knowledge of the inheritance pattern of important traits. This screening process is known as marker-assisted selection (MAS; see Glossary), and it is widely recognized to hold particular promise in woody perennial fruit crops that are traditionally expensive to breed [4–9].

There are several ways that markers can be used to improve the efficiency of breeding (Box 1). MAS is being increasingly used in apples and grapes, focusing on selecting traits that segregate in a Mendelian or near-Mendelian fashion and for which single genetic markers have been identified that account for all or most of the variance in the trait [6,10]. Often, MAS is used to select for resistance to pathogens at the seedling stage. In apple, for example, MAS is being used to select seedlings resistant to apple scab (*Venturia inaequalis*) [7,11,12], powdery mildew (*Podosphaera leucotricha*) [13], and fire blight (*Erwinia amylovora*) [14,15]. In grape, MAS is also used to breed for disease resistance [16,17], but seedlings can additionally be screened for several traits, including berry color [18], flower sex [19,20], seedlessness [17,21], and Muscat aroma [22]. The future challenge for MAS in grapes, apples, and most other crops is to move beyond simply inherited qualitative traits and enable selection for complex traits controlled by many loci. Although this can be achieved without genetic mapping (see below), MAS often relies first on the establishment of robust marker–trait associations, or quantitative trait loci (QTL), through genetic mapping.

Hunting for causal loci

Almost all of the known marker–trait associations in apples and grapes have been generated using linkage mapping, in which controlled crosses are made to generate a family of known relatedness and genetic markers are identified that co-segregate with phenotypes within this family. The power of linkage mapping in apples and grapes is limited compared with many other crops because of the prohibitive cost of growing large numbers of plants, their long generation times, and the inability to generate inbred lines due to self-incompatibility (apples) and inbreeding depression (grapes). Genome-wide association (GWA) mapping offers an attractive alternative because it makes use of diverse panels of individuals and can be applied to existing germplasm collections, thereby circumventing the time-consuming process of generating controlled crosses. GWA has revolutionized human disease mapping [23] and is increasingly being applied to agricultural crops [24–27]. Enthusiasm for GWA, however, must be tempered by a consideration of its main limitation.

GWA is useless when a trait is perfectly correlated with relatedness (e.g., when the trait is present in all members of a wild species but is absent in the domesticated species). This is because many genetic markers across the genome will correlate perfectly with such a trait because markers will often simply distinguish the two species from each other [28]. Grapes and apples are members of diverse genera with dozens of species that contain these types of trait, which geneticists aim to map and breeders aim to introgress from wild species into elite cultivars. For example, single-gene apple scab resistance has been introgressed from a wild Japanese crabapple, *Malus floribunda*, into the background of the domesticated apple, *Malus domestica* [11], and a long-term breeding program has introduced Pierce's disease resistance from a wild North American vine, *Vitis arizonica*, into domesticated grapes, *Vitis vinifera* [16]. The importance of these wild species should not be underestimated: traits provided by wild relatives contribute over US$100 billion to the world economy annually [29]. It is only by making controlled crosses that markers associated with these traits can be discovered and used in MAS.

Although breeders often target key traits found exclusively in wild relatives, it is arguably complex traits that vary quantitatively among elite domesticated cultivars that constitute the bulk of the targets of selection during breeding (e.g., fruit quality, plant architecture, and yield) [30]. To identify markers associated with these traits, the plant genomics community should move beyond biparental crosses and begin performing GWA using diverse collections of elite germplasm [28,31,32]. To genetically map all types of segregating phenotypes, however, a combination of both linkage and GWA mapping is optimal [33].

More so than in most crops, the greatest challenge for genetic mapping studies of any kind in grapes and apples is controlling the nongenetic, or 'environmental', variance. The importance of nongenetic variance is perhaps most evident in the world of wine. For example, the French concept of 'terroir' reflects the observation that the flavor profile of a wine will reflect the geographic area in which it was grown [34]. Simply tasting several bottles of Sauvignon

Box 1. Marker-assisted selection strategies

In apples and grapes, breeders often cross elite varieties from the cultivated species [e.g., Golden Delicious (*Malus domestica*) or Pinot Noir (*Vitis vinifera*)] with wild species that have evolved independently from the cultivated species for tens of millions of years (e.g., *Malus robusta* or *Vitis riparia*). These wild species are often highly unsuitable for commercial cultivation for many reasons, including their unpleasant taste profiles. Generally, the purpose of this breeding strategy is to generate fruit that are as similar as possible to elite cultivated varieties while having only a particular desirable trait from the wild species. For example, resistance to Pierce's disease is achieved by crossing elite *V. vinifera* cultivars with the wild species *Vitis arizonica* [16]. If known markers are linked to the trait of interest, MAS can be used to select offspring at the seedling stage that carry the marker and, therefore, will express the trait as adult plants, thereby making the breeding process less laborious and more cost effective.

In Figure I, an example of MAS for a dominant Mendelian trait is provided. The first step involves foreground selection: a variety from a cultivated species is crossed with a wild relative to generate offspring, half of which are expected to carry the desired allele from the wild relative (indicated by the red bar in Figure I). By genotyping the desired allele, or markers linked to it, in the offspring at the seedling stage, an expected 50% of the offspring that do not carry the allele can be discarded, whereas the 50% that do carry the allele can be grown into adult plants and evaluated for commercial suitability.

The second step involves selection for the desired allele in combination with background selection. An elite cultivated variety is crossed with the F_1 hybrid that inherited one chromosome from the cultivated species and one chromosome from a wild species, which carries the desired allele that the breeder is aiming to introgress into the cultivated genetic background. The offspring from this cross will vary in the proportion of their ancestry derived from the wild and cultivated species. Genotype data can be used to infer the ancestry of each of the offspring, and breeders can then select offspring with the minimum amount of wild ancestry. In addition, the breeder can select for the desired allele in the offspring. In this manner, the breeding process is accelerated towards the goal of achieving a new cultivar that has the maximum amount of ancestry from the cultivated species, while carrying the desired allele from the wild species.

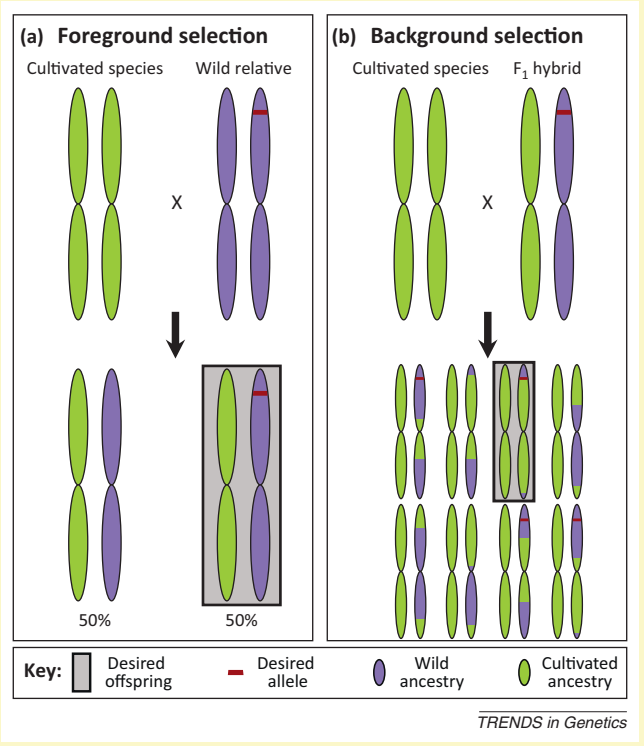

Figure I. Schematic of two marker-assisted selection strategies. **(a)** Foreground selection. **(b)** Background selection.

Blanc from around the world will give an appreciation of how genetically identical grapes can produce wines that range in aroma from fresh-cut grass to intense tropical fruit. Moreover, horticulturalists do not simply plant, wait, and then harvest. They choose particular rootstocks, training systems, pruning techniques, and postharvest treatments specific to each cultivar they grow and where they are growing it. Experimental designs and phenotyping protocols that minimize the effect of these nongenetic factors are therefore vital to the success of genetic mapping studies [27] and, thus, MAS in grapes and apples.

Molecular breeding without QTL

Genetic data are helpful in the breeding process even without having to perform genetic mapping. For example, genetic evaluations of two large grape collections in the USA and France have resolved pedigree relations and revealed numerous naming errors that aid with germplasm curation [35,36]. Inbreeding depression and gametic incompatibility can be avoided when the genetic relatedness among potential parents is known, which is particularly important in grape and apple breeding because many elite cultivars commonly used in breeding programs share close genetic relations [35,37].

Because desirable traits are often introgressed from wild apple and grape species, background selection can be used to choose individuals carrying the least wild ancestry in subsequent backcrosses, thereby enriching for the desired cultivated genetic background (Box 1). With the increasing availability of genome-wide polymorphism data, it is anticipated that the use of this approach will become more widespread. Assigning ancestry to each chromosomal segment in samples with ancestry from multiple species is not trivial, however, and the development of statistical methods for ancestry deconvolution is currently an active area of research (e.g., [38–40]).

I believe that the most promising use of genetic marker data that does not involve genetic mapping is genomic selection (GS) [41]. GS makes use of the same populations used for GWA, but instead of identifying markers associated with a trait, it uses all of the marker data to predict the trait. GS is especially useful for predicting complex traits controlled by many small-effect loci. Both simulation and empirical work has shown that GS provides an attractive alternative to conventional selection for fruit quality traits in a New Zealand apple-breeding program [42,43]. For breeding purposes, it is likely that GS will become the MAS method of choice because most traits targeted by breeders are likely controlled by many small-effect loci [30] and GS circumvents the need to identify these loci and estimate their effects. For researchers who aim to understand the genetic causes of trait variation, however, GWA can be performed in the same populations used for GS. Although, in many cases, significant hurdles remain to

collect the required genotype and phenotype data to implement GS and GWA, I believe that these two approaches have the potential to improve significantly the efficiency of apple and grape improvement in the coming decades.

Genetic variation in apples and grapes: ubiquitous but elusive

Grapes and apples are extraordinarily genetically diverse and this tremendous diversity presents significant challenges in acquiring the genome-wide polymorphism data required for MAS and GS. For example, single nucleotide polymorphism (SNP) genotyping microarrays have been developed for grape and apple, but when tested on diverse germplasm, approximately one-third of the assayed SNPs failed to produce useful genotypes [44,45]. The performance of these arrays relies on a perfect match between each 50-bp probe sequence based on the reference genome and the DNA sequence from samples that are often highly divergent from the reference genome. In surveys of diverse collections of domesticated apples and grapes, it is not uncommon to discover a SNP every 50 bp [46,47]; therefore, noisy probe-sequence hybridization signals should be expected, especially if attempting to genotype wild relatives with even more divergent sequences. Efforts to reduce probe length may be helpful, but variation in genome content further complicates SNP calling: insertion and/or deletion polymorphisms (indels), copy number variants (CNVs), and presence–absence variants (PAVs) interfere with hybridization signals and are known to be ubiquitous in several high-diversity species, including maize [48,49] and soybean [50]. Although arrays may suffice for providing saturated genetic linkage maps

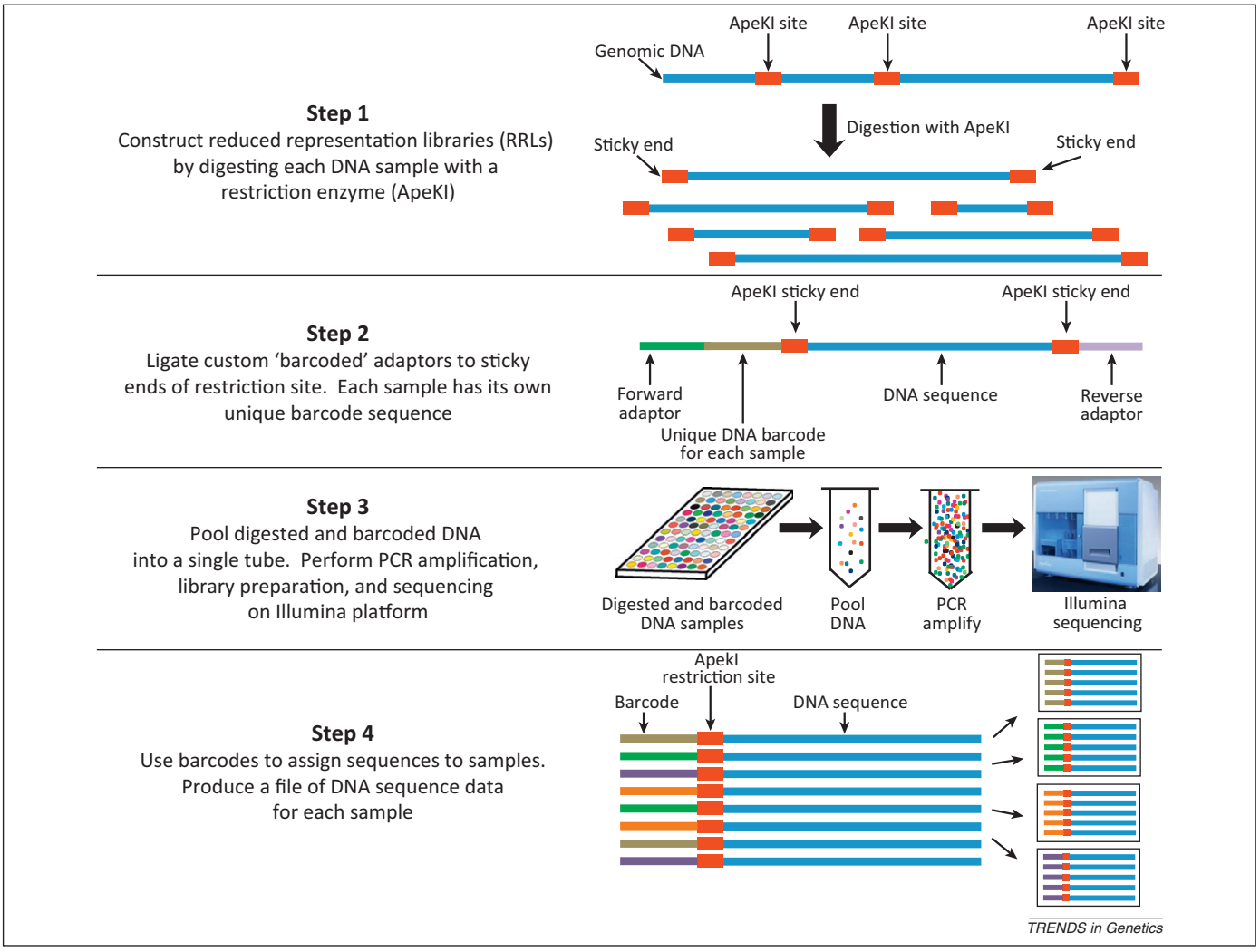

Figure 1. Genotyping-by-sequencing (GBS) [51]. In step 1, a genome reduction procedure is performed whereby each sample is digested with the restriction enzyme ApeKI to generate a reduced representation library (RRL). Because only digested fragments within a certain size range (approximately 100–400 bp) will be successfully sequenced using sequencing technology produced by Illumina (CA, USA), this results in DNA sequence reads being generated from only a small fraction of the genome. So that each sample is uniquely identifiable, a custom barcode adaptor is ligated to the sticky end produced by the restriction digestion in step 2. The Illumina standard forward and reverse adaptors are also ligated during this step. Step 3 involves pooling all of the samples together into a single tube. After pooling, the remaining steps of the Illumina standard library preparation procedure are performed and DNA sequencing takes place on a single lane of an Illumina next-generation sequencing (NGS) machine. The resulting large file of DNA sequence data contains a mix of sequences from all of the samples. In step 4, DNA sequences belonging to a particular sample are retrieved from the resulting file by identifying and gathering all of the sequence reads containing the unique barcode associated with that sample. In the end, a single file of DNA sequence data is obtained for each sample. These files can be individually aligned to a reference genome to generate measures of alignment confidence and quality associated with the bases of each read. A multi-sample genotype caller can then be applied to produce genotype calls and associated quality metrics (e.g., [61]). The application of various quality thresholds and heuristics results in a final table of genome-wide genotype data from the sequenced samples. It is worth noting that genotype callers that do not rely on a reference genome are also being developed {e.g., Universal Network Enabled Analysis Kit (UNEAK) [62]}.

under certain circumstances [46], the poor quality of the SNP data from these arrays casts serious doubt on the long-term utility of microarrays in high-diversity crops such as apples and grapes.

Although next-generation sequencing (NGS) was not designed to generate genome-wide polymorphism data from targeted sites across multiple individuals as done by microarrays, methods using reduced representation libraries (RRL) and multiplex barcoding are now routinely being used to generate these kinds of data from NGS [51–55]. Genotyping-by-sequencing (GBS) is an NGS-based method that is particularly attractive as an alternative to microarrays because it enables marker discovery and genotyping in one single step at a significantly lower per sample cost (Figure 1). In addition, anticipated improvements in NGS capacity in the future promise to further decrease the per-sample cost and increase the quality of the resulting genotype data.

However, as is the case with microarrays, high levels of genetic diversity also complicate genotype calling from NGS data. DNA sequences from samples distantly related to the reference genome will sometimes not align properly to the reference sequence and will result in incorrect or missing genotype calls (Box 2). In addition, GBS does not

Box 2. Analysis of GBS data

Although methods to generate dense marker maps from sets of diverse samples using NGS are promising, they are still in their infancy and there are several statistical hurdles to overcome before high-quality genotype data are routinely produced. This is especially the case in highly diverse and heterozygous species such as apples and grapes. Figure I provides a simple depiction of some of the challenges associated with calling genotypes from GBS data.

The reference genome sequence for a single locus is shown and the remaining rows depict the various sequences that one might encounter when obtaining GBS data from diverse sets of samples. In this case, samples are diploid and, therefore, there are two sequences shown for this locus for each sample. The restriction site on the left results from the first step of the GBS library production method (see Figure 1 in main text). The gray-shaded rectangle highlights a focal SNP, the genotypes of which are provided in the final two columns of the table. Checkmarks and Xs indicate whether a particular sequence read aligns successfully to the reference genome sequence.

Calling genotypes is simplest in cases such as sample A, where one of its sequences matches the reference genome perfectly and the other sequence differs by only a single nucleotide. With sufficient sequence coverage of each of these alleles, one can confidently call the genotype of the sample correctly as 'CT'. Samples B and C demonstrate that polymorphisms within the restriction site can result in false genotype calls. To generate sequence reads from the same genomic regions across all samples using GBS, the restriction site must be conserved across all samples. Depending on how much diversity one aims to capture in a GBS experiment and how many samples one processes, it may often be the case that restriction sites are not conserved across every sample in an experiment. Sequences carrying an allele that disrupts the restriction site will be excluded during the generation of the RRL, will not be part of the sequencing library, and thus will not get sequenced. This can result in false homozygous calls in heterozygous samples (B), or missing genotypes in samples homozygous for the allele that disrupts the restriction site (C).

Finally, sequences that are highly divergent from the reference genome may not align. In this case, sequences with more than three mismatches are discarded. As with the previous two examples, this can result in false homozygous calls in heterozygous individuals (D) and in missing genotype calls (E). In some cases, the presence or absence of reads at a locus can be interpreted as dominant markers [51,53], but this interpretation will be more difficult in highly heterozygous crops such as apples and grapes. These examples represent only a subset of the challenges associated with these data, because I have ignored the difficulties generated from other types of genetic diversity that are ubiquitous in many species, such as CNVs and PAVs [63,64].

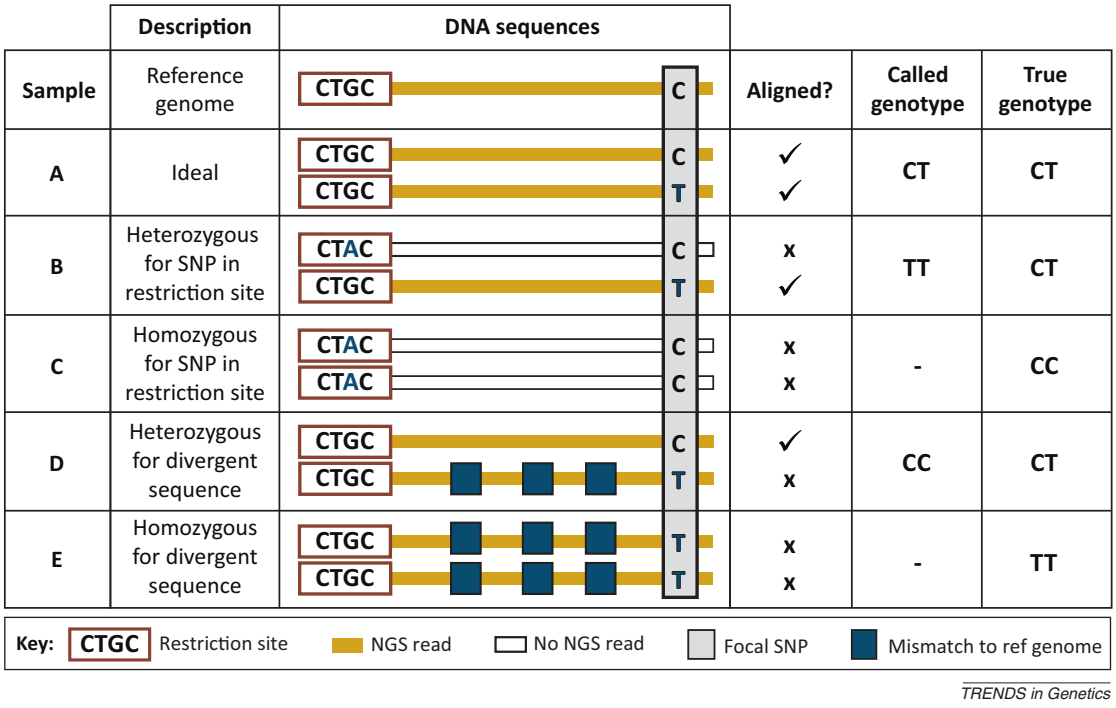

Figure I. Examples of the challenges associated with calling genotypes from genotyping-by-sequencing (GBS) data.

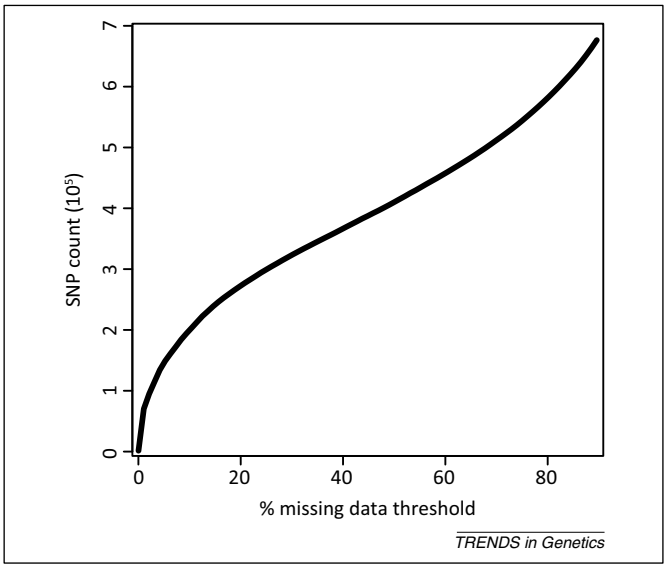

Figure 2. The nature of missing genotype data generated from genotyping-by-sequencing (GBS). GBS data from 95 diverse apple cultivars, including several wild *Malus* species, generated from a single lane of Illumina next-generation sequencing (NGS) resulted in the identification of >650 000 single nucleotide polymorphism (SNPs). However, the number of SNPs identified decreases rapidly as the % missing data threshold decreases (i.e., as it becomes more stringent). For example, if SNPs with >20% missing data are excluded from analysis (i.e., a 20% missing data threshold), more than half of the identified SNPs are discarded (Myles, unpublished data).

produce an equal amount of sequence for each sample or site in the genome and this results in large amounts of missing genotype data (Figure 2). Finally, heterozygous crops, such as apples and grapes, require more NGS data for accurate genotype calling compared with inbreds, because each SNP in a heterozygous species has three possible genotypes rather than only two. Thus, for every dollar spent on NGS, one should not only expect fewer genotypes from heterozygous species compared with inbreds, but also generally less accurate genotype calls.

Working to overcome the statistical hurdles of interpreting NGS data, however, is still a more attractive option than reverting back to the use of microarrays. The development of microarrays is so costly and time consuming that it is generally undertaken only by large, well-funded consortia of researchers (e.g., [45,56,57]). Genotypes from NGS data, by contrast, can be obtained by small research groups with reduced budgets. Perhaps most importantly, unlike microarray data, NGS data collected today are likely to increase in value over time: anticipated improvements to reference genomes, alignment algorithms, and genotype callers should enable the extraction of more useful genotypes in the future than is currently possible.

In this respect, there is an urgent need for improved genotype calling and imputation algorithms that 'fill in' the missing genotypes produced by NGS in high-diversity, heterozygous crops. Imputation applied to low-coverage NGS data already produces sufficiently high-quality genotypes to map complex traits in inbred crops [25,26] and in humans [58] at a fraction of the cost of microarrays. However, most heterozygous crops are more diverse than are humans and imputation accuracy is expected to decline with increasing levels of genetic diversity. In addition, imputation in human studies makes use of reference panels of high-quality genotype data, such as the 1000 Genomes reference panel [59], and such panels do not exist for most crops. Concerted efforts to produce such publicly available panels for agricultural species should be encouraged because they can be leveraged by anyone to increase the number, and improve the quality, of their NGS-derived genotype data and, thus, accelerate crop improvement through MAS.

Looking beyond the genome

The extent to which genomics can significantly improve the breeding process depends crucially on the cost of obtaining measurements based on phenotype being greater than the cost of collecting the required genetic information [60]. As the cost of generating DNA sequences continues to fall and the statistical tools for analyzing them improve, it is inevitable that the acquisition of dense genome-wide polymorphism data will become inexpensive and routine. However, the cost of generating and evaluating mature apple trees and grapevines is bound to remain rather static. These considerations make MAS and GS particularly attractive for the future of apple- and grape-breeding programs.

The genome is finite, but the phenome is essentially infinite. The limitation to MAS and GS in a future of inexpensive DNA sequence data will not be the collection of genotype data, but the collection of extensive and reliable phenotype data. In apple and grape breeding, the gains in efficiency offered by genomics will depend strongly on improving the accuracy and throughput of phenotyping, and the ability to work with large germplasm collections. In anticipation of the future genomic data deluge, researchers and breeders committed to shuffling up new genomic combinations would be wise to focus now on increasing the throughput of their phenotyping strategies and on establishing populations suitable for genetic mapping, MAS, and GS. How consumers will respond to novel genomic combinations in their wine glass or apple pie remains unknown, but thinking decades ahead in this manner is arguably the best tactic to reap fully the benefits of the genomics revolution and evade the perils of perpetual propagation.

Acknowledgments

I thank Patrick J. Brown, Emilie de Rosenroll, Charlie Embree, Kyle Gardner, Andrew Jamieson, Satish Kumar, Chris Owens, and Marina Myra for editorial assistance, and Sharon Mitchell for graphics. This article was written, in part, thanks to funding from the Canada Research Chairs Program and Agriculture and Agri-Food Canada.

References

1 Viala, P. and Vermorel, V. (1901–1910) *Ampélographie, Volumes I – VII*, Masson (in French)
2 Collard, B.C.Y. and Mackill, D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 557
3 Ignatov, A. and Bodishevskaya, A. (2011) *Malus*. In *Wild Crop Relatives: Genomic and Breeding Resources Temperate Fruits* (Kole, C., ed.), pp. 45–64, Springer
4 Kellerhals, M. *et al.* (2000) Marker-assisted selection in apple breeding. *Acta Hort.* 521, 255–266
5 Dirlewanger, E. *et al.* (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9891–9896

6 di Gaspero, G. and Cattonaro, F. (2010) Application of genomics to grapevine improvement. *Aust. J. Grape Wine Res.* 16, 122–130
7 Bus, V. *et al.* (2000) Marker-assisted selection for pest and disease resistance in the New Zealand apple breeding program. *Acta Hort.* 538, 541–547
8 Shulaev, V. *et al.* (2008) Multiple models for Rosaceae genomics. *Plant Physiol.* 147, 985–1003
9 Tartarini, S. and Sansavini, S. (2003) The use of molecular markers in pome fruit breeding. *Acta Hort.* 622, 129–140
10 Korban, S. and Tartarini, S. (2009) Apple structural genomics. In *Genetics and Genomics of Rosaceae* (Folta, K. and Gardiner, S., eds), pp. 85–119, Springer
11 King, G.J. *et al.* (1999) Introgression of the *Vf* source of scab resistance and distribution of linked marker alleles within the *Malus* gene pool. *Theor. Appl. Genet.* 99, 1039–1046
12 Hemmat, M. *et al.* (2003) Identification and mapping of markers for resistance to apple scab from 'Antonovka' and 'Hansen's baccata #2'. *Acta Hort.* 622, 153–161
13 Evans, K.E. and James, C.J. (2003) Identification of SCAR markers linked to *Pl-w* mildew resistance in apple. *Theor. Appl. Genet.* 106, 1178–1183
14 Khan, M.A. *et al.* (2007) Development of molecular markers linked to the Fiesta linkage group 7 major QTL for fire blight resistance and their application for marker-assisted selection. *Genome* 50, 568–577
15 Peil, A. *et al.* (2007) Strong evidence for a fire blight resistance gene of *Malus robusta* located on linkage group 3. *Plant Breed.* 126, 470
16 Riaz, S. *et al.* (2009) Using marker-assisted selection to breed Pierce's disease-resistant grapes. *Am. J. Enol. Viticult.* 60, 199–207
17 Mahanil, S. *et al.* (2012) Development of marker sets useful in the early selection of Ren4 powdery mildew penetration resistance and seedlessness for table and raisin grape breeding. *Theor. Appl. Genet.* 124, 23
18 This, P. *et al.* (2007) Wine grape (*Vitis vinifera L.*) color associates with allelic variation in the domestication gene *VvmybA1*. *Theor. Appl. Genet.* 114, 723–730
19 Dalbo, M.A. *et al.* (2000) A gene controlling sex in grapevines placed on a molecular marker-based genetic map. *Genome* 43, 333–340
20 Marguerit, E. *et al.* (2009) Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. *Theor. Appl. Genet.* 118, 1261–1278
21 Doligez, A.D. *et al.* (2002) Genetic mapping of grapevine (*Vitis vinifera L.*) applied to the detection of QTLs for seedlessness and berry weight. *Theor. Appl. Genet.* 105, 780–795
22 Emanuelli, F. *et al.* (2010) A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* 10, 241
23 Donnelly, P. (2008) Progress and challenges in genome-wide association studies in humans. *Nature* 456, 728
24 Zhao, K. *et al.* (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2, 467
25 Tian, F. *et al.* (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43, 159–162
26 Huang, X. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967
27 Atwell, S. *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631
28 Myles, S. *et al.* (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21, 2194–2202
29 Pimentel, D. *et al.* (1997) Economic and environmental benefits of biodiversity. *BioScience* 47, 747–757
30 Bernardo, R.N. (2010) *Breeding for Quantitative Traits in Plants*, (2nd edn), Stemma Press
31 Hamblin, M.T. *et al.* (2011) Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27, 98
32 Nordborg, M. and Weigel, D. (2008) Next-generation genetics in plants. *Nature* 456, 720–723
33 McMullen, M.D. *et al.* (2009) Genetic properties of the maize nested association mapping population. *Science* 325, 737–740
34 Robinson, J. (2006) *The Oxford Companion to Wine* (3rd edn), Oxford University Press
35 Myles, S. *et al.* (2011) Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3530–3535
36 Laucou, V. *et al.* (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor. Appl. Genet.* 122, 1233
37 Noiton, D.A.M. and Alspach, P.A. (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivar. *J. Am. Soc. Hort. Sci.* 121, 773–782
38 Bryc, K. *et al.* (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. U.S.A.* 107, 786–791
39 Tang, H. *et al.* (2006) Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* 79, 1–12
40 Wegmann, D. *et al.* (2011) Recombination rates in admixed individuals identified by ancestry-based inference. *Nat. Genet.* 43, 847–853
41 Heffner, E.L. *et al.* (2009) Genomic selection for crop improvement. *Crop Sci.* 49, 1–12
42 Kumar, S. *et al.* (2012) Towards genomic selection in apple (*Malus x domestica* Borkh.) breeding programmes: prospects, challenges and strategies. *Tree Genet. Genomes* 8, 1–14
43 Kumar, S. *et al.* (2012) Genomic selection for fruit quality traits in apple (*Malus×domestica Borkh.*). *PLoS ONE* 7, e36674
44 Myles, S. *et al.* (2010) Rapid genomic characterization of the genus *Vitis*. *PLoS ONE* 5, e8219
45 Chagné, D. *et al.* (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7, e31745
46 Micheletti, D. *et al.* (2011) Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. *Tree Genet. Genomes* 7, 857
47 Lijavetzky, D. *et al.* (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8, 424
48 Morgante, M. *et al.* (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat. Genet.* 37, 997–1002
49 Springer, N.M. *et al.* (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* 5, e1000734
50 Kim, M.Y. *et al.* (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22032–22037
51 Elshire, R.J. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379
52 Baird, N.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376
53 Poland, J.A. *et al.* (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7, e32253
54 Peterson, B.K. *et al.* (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7, e37135
55 Davey, J.W. *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499
56 Kijas, J.W. *et al.* (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4, e4668
57 Verde, I. *et al.* (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP Detection and validation in breeding germplasm. *PLoS ONE* 7, e35668
58 Pasaniuc, B. *et al.* (2012) Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat. Genet.* 44, 631–635
59 1000 Genomes Project Consortium *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061
60 Luby, J. and Shaw, D.V. (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience* 36, 872–879
61 McKenna, A. *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303
62 Lu, F. and Lipka, A.E. *et al.* (2013) Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genet* 9, e1003215
63 Morgante, M. (2006) Plant genome organisation and diversity: the year of the junk! *Curr. Opin. Biotechnol.* 17, 168–173
64 Morgante, M. *et al.* (2007) Transposable elements and the plant pan-genomes. *Curr. Opin. Plant Biol.* 10, 149