A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine

Shanshan Yang1, Jonathan Fresnedo-Ramirez2, Minghui Wang2, Linda Cote3, Peter Schweitzer3, Paola Barba4, Elizabeth M Takacs1, Matthew Clark5, James Luby5, David C Manns6, Gavin Sacks7, Anna Katharine Mansfield8, Jason Londo8, Anne Fennell9, David Gadoury9, Bruce Reisch1, Lance Cadle-Davidson10 and Qi Sun7

Marker-assisted selection (MAS) is often employed in crop breeding programs to accelerate and enhance cultivar development, via selection during the juvenile phase and parental selection prior to crossing. Next-generation sequencing and its derivative technologies have been used for genome-wide molecular marker discovery. To bridge the gap between marker development and MAS implementation, this study developed a novel practical strategy with a semi-automated pipeline that incorporates trait-associated single nucleotide polymorphism marker discovery, low-cost genotyping through amplicon sequencing (AmpSeq) and decision making. The results document the development of a MAS package derived from genotyping-by-sequencing using three traits (flower sex, disease resistance and acylated anthocyanins) in grapevine breeding. The vast majority of sequence reads (≥99%) were from the targeted regions. Across 380 individuals and up to 31 amplicons sequenced in each lane of MiSeq data, most amplicons (83 to 87%) had <10% missing data, and read depth had a median of 220–244. Several strengths of the AmpSeq platform that make this approach of broad interest in diverse crop species include accuracy, flexibility, speed, high-throughput, low-cost and easily automated analysis.

Horticulture Research (2016) 3, 16002; doi:10.1038/hortres.2016.2; Published online: 17 February 2016

INTRODUCTION

Marker-assisted selection (MAS) is now commonly employed in perennial crop breeding programs to pursue the acceleration of cultivar development.1–3 In particular, MAS has been shown to provide advantages for selection during the juvenile phase;4,5 for pyramiding disease resistance genes;6,7 and for replacing expensive, time-consuming or technically difficult traits.8,9 Simply inherited traits with Mendelian or near-Mendelian segregation patterns are major targets for MAS. Examples of MAS have been reported for seedlessness and flower sex in grape, and disease resistance in apple, grape and tomato breeding.10–11 Markers have also been applied to quantitatively inherited traits, especially those with major quantitative trait loci (QTL) effect, including fruit acidity in peach,12 fruit size in tomato,13 peach and cherry,14 grain yield in rice15 and drought tolerance in chickpea.9

The development of molecular markers requires the detection of association between target traits and genotypes. Two approaches are often used to detect such associations: (a) QTL analysis with structured families, and (b) genome-wide association study, which takes advantage of linkage disequilibrium (LD) in diverse germplasm to capture the linkage between markers and causal genes.16,17 However, for highly heterozygous and diverse crops, such as grape, genome-wide association study has limitations.18–20 LD decays rapidly in species of Vitis, in which the square of the correlation coefficient ($r^2$) declines to 0.1 within 2.7 cM, making genome-wide association study an unsuitable method for genetic mapping with current genotyping platforms, such as single nucleotide polymorphism (SNP) microarrays.21,22 QTL analysis in mapping families has been a more effective method for genetic mapping in species with diverse backgrounds.21

Multiple genotyping platforms and molecular marker types have been utilized for MAS in highly heterozygous crops, including simple sequence repeats (SSR) and single-locus or multi-locus SNP assays. SSRs are particularly well-suited for MAS because of their multi-allelic nature and high transferability among distinct species or genera.23–25 which enables the analysis of complex crossing involving progenitors with multiple inter-specific hybridizations. However, SSR as a genotyping platform has its own disadvantages including low-throughput, labor-intensive and time-consuming.20 Low density of SSR markers could cause loss of linkage between markers and causal genes, or lack of segregation in certain families.26–28 SNP microarrays are closed platforms suffering from ascertainment bias,31 resulting in poor flexibility and poor transferability across diverse germplasm.32 In addition, the cost of microarray pre-design is still a major obstacle in adopting SNP arrays in horticultural breeding programs.33,34
To fill the gaps between the above technical and economic considerations for breeding and actual implementation of MAS,\(^{35}\) next-generation sequencing (NGS) technology offers a potential opportunity for unbiased genotyping with high-throughput and low per-sample cost. Genotyping-by-sequencing (GBS), with its simultaneous marker discovery and genotyping approach, delivers many benefits including availability of flanking DNA sequence information, high-sample throughput and scalability (multiplexing), and high resolution.\(^ {36}\) It has been successfully applied in marker discovery for many self-pollinating crops as well as outcrossing species.\(^ {37}\) However, successful implementation of GBS for MAS has not been reported in any heterozygous crop breeding programs. Many reasons may have contributed to the lack of adoptions. From a technical perspective, missing data, genotyping errors and heterozygote under-calling are common in GBS results due to uneven sequencing depth across sites and high level of sample multiplexing.\(^ {38}\)–\(^ {41}\) Rapid LD decay, large-scale genome structure variation coupled with lack of haplotype information makes it impractical to do genotype imputation in heterozygous perennial species with diverse backgrounds. From a practical perspective, the long turn-around time from sample collection to data analysis makes it difficult to fit into most breeding timeframes. Moreover, computational challenges in data processing further hinder breeders’ interest in implementing GBS. In summary, GBS is currently impractical for MAS in heterozygous crops. This study presents a novel and efficient strategy for molecular marker development and practical implementation in MAS, based on amplicon sequencing (AmpSeq). The semi-automated pipeline incorporates a machine learning model for primer design and uses Illumina’s Nextera dual-barcoding and sequencing platforms for genotyping (Illumina, San Diego, CA, USA). After detecting a SNP from GBS, the strategy starts from the design of primers using the GBS sequence tags. The converted amplicon markers can then be used for genotyping through NGS. The design involves multiplexing of both samples and markers. As a case study, we document the use of AmpSeq in grapevine breeding programs for three traits including flower sex, powdery mildew (PM) resistance and acylated anthocyanins. We chose these three traits to represent a Mendelian trait and two QTL with differing effects on phenotypic variance: (a) a single gene for flower sex with three alleles;\(^ {42}\) (b) a QTL with moderately high \(R^2\) (acylation of anthocyanins, initially reported here); and (c) a QTL with relatively low \(R^2\) (Ren2 locus for PM (Erissyphe necator) resistance from Vitis cinerea (Engelm. ex A. Gray) Engelm. ex Millard accession B9 (V. cinerea B9))\(^ {35}\)\(^ {36}\)\(^ {42}\). All three loci are located on different chromosomes, and we were able to test the AmpSeq approach for flower sex (male versus female versus hermaphrodite) across four different families where the male flower allele descends from V. c. the hermaphrodite allele directly from V. vinifera L. and the female allele descends from an unidentified North American Vitis species. Two of the three traits chosen for analysis would take 2–4 years to analyze phenotypically due to the time it takes for a seedling to produce flowers and fruit. We also report here the development of a pipeline package with tools for AmpSeq marker design and decision support.

MATERIALS AND METHODS

Plant materials

Four families were chosen for this study, all representing interspecific hybridization of diploid (2n = 38) Vitis species: V. vinifera ‘Chardonnay’ × V. cinerea B9; ‘Horizon’ (complex hybrid of V. vinifera, V. labrusca L., V. rupestris Schelle and V. aestivalis Michx.) × V. cinerea B9; ‘Horizon’ × Illinois 547-1 (V. rupestris B38 × V. cinerea B9) and MN1246 × MN1264 (both, complex hybrids with a genomic background including at least V. vinifera, V. riparia Michx., V. rupestris, V. labrusca, V. cinerea and V. aestivalis). The first three families were grown in research vineyards operated by Cornell University, Geneva, New York. The latter population was grown in research vineyards of the University of Minnesota, St Paul, located at Excelsior, Minnesota. All four populations segregated for flower sex (male, female and hermaphrodite) a trait controlled by a single locus with three alleles (reviewed in the study by Hyma et al.\(^ {42}\)). In addition, our ongoing research indicated that two populations segregated for acylation of anthocyanins, and the three populations descending from V. cinerea B9 segregated for Ren2 PM resistance.\(^ {42}\)\(^ {44}\)

AmpSeq marker development pipeline

The AmpSeq marker development procedure consists of four steps illustrated in Figure 1. First, GBS marker-trait associations were evaluated in TASSEL 4.3.13 (ref. 45). Genetic maps were constructed by the HetMappS strategy, and QTL were mapped in R/qtl ver. 1.37–11 as described by Hyma et al.\(^ {42}\) Marker phase and effect were estimated by a custom R script (Supplementary File 1) using the intermediate files ‘ phased’ and ‘Gmap’ from the HetMappS pipeline.\(^ {42}\) For the second step, the range of the haplotype block across the QTL region was defined by haplotype start and end markers with similar absolute value of marker effects. The anchor marker selected for each QTL was based on high association with trait within the 1.8-LOD (logarithm base 10 of odds) support interval of the QTL. A Perl script included in the package ‘calculated_LD_distribution.pl’ took in a text file ‘test_marker’ as input, which includes physical positions of the three markers: haplotype start, haplotype end and the anchor SNP. This script retrieved additional SNP markers from the GBS pipeline (pre-filtered by maximum of 5% missing data) that were within the haplotype range and verified to be in LD with the anchor marker. After the SNP markers were defined, a Perl script ‘parse_bam.pl’ was used to identify genomic regions suitable for designing PCR primers. The template for the primer sequences was based on the actual sequences from GBS data, rather than the reference genome sequence itself. This was to ensure that there was no mismatch between the primer sequences and their targeted alleles. The ‘parse_bam.pl’ script required three inputs: (1) the reference genome FASTA file; (2) a BAM file, converted from the SAM file generated by TASSEL GBS pipeline,\(^ {46}\) which contained GBS reads aligned to the reference genome; and (3) a cutoff for the LD P-value as calculated from the previous step. By default, the script required that at least 10 bp of the left and right flanking sequences are covered by the GBS read, and ≥90% alignment matching rate between the two alleles of up to 20 bp of each flanking sequences. Finally the script ‘primer3.pl’ was used to call the primer designing software Primer3 (ref. 46) and captured output properties from Primer3. The default primer sequence length from the pipeline was 22 bp. With a single SNP in the middle, the targeted amplicon size was 45 bp. From the output of the primer design pipeline followed by manual filtering, including removal of overlapping amplicons and primers with extreme annealing temperatures \(T_m\) outside of 47–79 °C, a total of 54 amplicons across the three loci were retained for testing: 19 for flower sex, 12 for PM resistance and 23 for acylated anthocyanins. The output file of the 54 amplicons is shown in Supplementary File 2. The documentation about the usage of the pipeline and the scripts are provided in Supplementary File 3.

Genotyping (PCR, library preparation and de-multiplex)

Two experiments were conducted to test three traits: flower sex, PM resistance due to the Ren2 locus and acylated-anthocyanin concentration. In Experiment 1, 19 primers for flower sex and 12 primers for Ren2 were pooled in one Illumina MiSeq lane (Illumina, San Diego, CA, USA) testing four 96-well plates (380 individuals and 4 blanks), each containing the
parents and a subsample of progeny from each grapevine breeding family. Experiment 1 consisted of four breeding families: ‘Horizon’ × V. cinerea B9, ‘Horizon’ × Illinois 547-1, and ‘Chardonnay’ × V. cinerea B9 and MN1246 × MN1264. The other 23 primers for acylated-anthocyanin concentration were determined in Pooled Experiment 1, testing two 96-well plates of ‘Horizon’ × Illinois 547-1 and two 96-well plates of V. rupestris B38 × ‘Horizon’ (380 individuals and 4 blanks).

For each end of the small leaf (< 1 cm diameter) was harvested and placed in one tube of a Costar 96-well cluster tube collection plate (Corning, Corning, NY, USA), and DNA was isolated as described previously. Briefly, each 96-well plate received up to 91 unique samples plus two sets of duplicated individuals and a blank well to serve as quality controls. Frozen samples were ground using stainless-steel beads in a Geno/Grinder 2000 (OPS Diagnostics LLC, Lebanon NJ, USA) and processed using DNeasy 96-well DNA extraction kits (Qiagen, Valencia, CA, USA). The Qiagen AP1 lys buffer was amended with PVP-40 (2% w/v) prior to heating of the buffer, to improve DNA quality and quantity. Due to historically consistent yields of 25–50 ng µl⁻¹ within and between plates, DNA was used following twofold dilution without quantification.

AmpSeq uses two rounds of PCR: the 1st PCR to amplify a multiplex of markers and incorporate linker sequences, and the 2nd PCR to use the linker sequences to add a unique pair of indices, or barcodes, to each sample. The linker 5′-TCGTCGACGCTGTGTATGTCAGACAGACAG-3′ was added to the 5′ end of each forward primer to accommodate S5xx barcode adapters, and the linker 5′-GCTGCCGATCCCCAGGTGTTAAGACAGACACG-3′ was added to the 5′ end of each reverse primer to accommodate N7xx barcode adapters described in Supplementary File 4. For results presented here, samples were processed without previously testing the primers as a multiplex. Each combination of PCR were conducted on 384 wells per plate, processed as four 96-well plates, which were pooled after the 2nd PCR for Illumina sequencing.

For the 1st PCR, up to 31 primer pairs (62 total, 10 pmol each) were combined with Qiagen Multiplex PCR Plus Mix (Qiagen, Valencia, CA, USA), following the manufacturer's protocols calculated for 10 µl reaction volumes including 2 µl of DNA template. Thermocycling conditions for the 1st PCR were: 95 °C for 5 min; followed by 35 cycles of 95 °C for 30 s, 62 °C for 90 s and 72 °C for 30 s; followed by a final extension of 10 min at 68 °C. After the 1st PCR, 50 µl H₂O was added to each well prior to serving as a template for the 2nd PCR.

In the 2nd PCR, 384 unique combinations of 16 row and 24 column indices were incorporated in 10 µl reactions consisting of Platinum Taq (2 U per reaction, Invitrogen, Grand Island NY, USA), Platinum Taq buffer (1× final), MgCl₂ (1.5 mM final), dNTPs (0.2 mM final each), primers (10 pmol each) and 2 µl of each amplicon pool from the 1st PCR reaction. Thermocycling conditions for the 2nd PCR were: 95 °C for 5 min; followed by eight cycles of 95 °C for 30 s, 53 °C for 30 s and 72 °C for 30 s; followed by a final extension of 10 min at 68 °C. Equal volumes of all 384 wells were pooled into a single tube, then purified by Ampure (Beckman Coulter, Indianapolis IN, USA), following manufacturer's instructions.

The amplicon pool was sequenced on an Illumina MiSeq instrument with single-end reads of 31 bp with 8 bp dual index reads according to the manufacturer's instructions. Reads were de-multiplexed using the Illumina bcl2fastq pipeline software, allowing a single mismatch in the index reads.

Phenotyping and data analysis

The experimental design for phenotyping is described in Supplementary File 5 (refs 42, 47–49).

SNPs were called by a custom perl script 'run_gatk2.pl' (Supplementary File 6) from the de-multiplexed fastq file. Read counts were calculated in 500 bp sliding windows across the 12X.2 version of the PN40024 reference genome55 and BEDTools56 v2.2.21 to check the amplification specificity and efficiency of the AmpSeq markers. VCFtools52 v0.1.22 was applied to calculate missing rate per site (~missing-site) and per individual (~missing-ind), and mean depth per site (~site-mean-depth) and per individual (~depth). VCF file generated from 'run_gatk2.pl' was input to TASSEL 4.3.13 for association detection. Genotyping and phenotyping data of the two parents were used for phasing.

To detect family-based marker specificity, the squared correlation coefficients between AmpSeq marker frequencies were calculated per family for each trait by VCFtools–geno-r2 function, following a dendrogram construction in R57 v3.1.2. P-values of association between genotype and phenotype were calculated by a one-way analysis of variance co-segregation analysis General Linear Model (GLM) on a per-family basis and after pooling families with the same segregation pattern. The minor allele frequencies of the AmpSeq markers were also reported on a per-family and pooled basis.

Output from the primer design pipeline was used to classify primers as efficient and non-efficient based on their amplification and genotyping performance, and 4 statistical models (logistic regression, support vector machine, decision tree and random forest) were tested using the following cross-validation method for assessing 18 numerical parameters of each primer (num_tags, p-value, phase, reject_code, a1, a1_count, a2_seq, a1_seq, a2_count, a2_seq). The prediction accuracy of the cross-validation classification model was assessed using Receiver Operating Characteristic (ROC) curve analysis.54–57 The ROC curve is a commonly used diagnostic to assess the prediction power of different classification methods, by means of a graphical plot of the true positive rate (sensitivity) against the false positive rate (1—specificity) for each classification method that predicts a dichotomous outcome. The plot is showed at several thresholds, which are used to designate whether the prediction of a given method as positive. The area under this curve (AUC) is one of the most important performance metrics that can be applied for selecting the most adequate classification method because it represents the accuracy of the prediction. In practice, AUC values range from 0.50 to 1.00, and a higher AUC value indicates better prediction accuracy for the classification models: values below 0.60 should be considered poor, 0.60–0.70 moderate, 0.75–0.85 good and >0.90 are very good to excellent.57 A value of exactly 0.50 would indicate that the model is useless for testing, while a value of exactly 1.00 would indicate the model is a perfect test. The models were evaluated using packages caret, e1071, rpart and random forest implemented in R v3.1.2.

RESULTS

Converting GBS-derived SNPs to amplicon sequencing (AmpSeq) markers

The procedure for the marker development and amplicon sequencing primer design pipeline is illustrated in Figure 1. Three traits were chosen for study: flower sex, Ren2 PM resistance43,44 and acylated anthocyanins (that is, anthocyanins esterified to hydroxycinnamic acids). The percentage of phenotypic variance explained by each QTL (R² value) confirmed these represent a qualitative trait (93%), a quantitative trait with a moderate QTL (13%), and a quantitative trait with a major QTL (54%), respectively (Table 1). For each locus, two flowering markers indicating the haploblock boundary, and one anchor marker having high association with the phenotype from the curated genetic maps are listed in Table 1. The primer design pipeline and curation resulted in 54 AmpSeq markers physically located between the two flowering markers and genetically linked to the anchor markers, for further testing (Supplementary File 2).

In Experiment 1, 12,677,206 reads were generated for the target QTL region on chromosome 2 (chr2) associated with flower sex, and 4,537,611 reads were generated for the target QTL region on chr14 associated with Ren2 PM resistance. These 17,214,817 reads at the two target loci comprise 99.99% of the total reads, while only 2,237 reads aligned to off-target regions. Similar results were obtained in Experiment 2, with 14,927,137 reads (98.82%) aligning to the target QTL region on chr 3 associated with acylated anthocyanins.

The sequencing depth per individual and missing rate were evaluated for each amplicon (Tables 2–4). For both experiments, a majority of amplicons (80% in Experiment 1 and 83% in Experiment 2) had >50-fold coverage per individual, in a relatively narrow range from 50–284×. Most amplicons (87% in Experiment 1 and 83% in Experiment 2) had <10% missing data. Thus, in Experiment 1, the 380 individuals had 220× median coverage per marker and 168× mean coverage per marker. Similarly in Experiment 2 for 23 acylated-anthocyanin AmpSeq
markers, 380 individuals had genotyping data with 244× median coverage per marker and 188× mean coverage per marker. One amplicon per experiment failed because no polymorphism was identified during the SNP calling procedure. The amplification specificity, stable sequencing depth and consistent missing rate in the two experiments indicated the marker development strategy and amplicon primer design pipeline were effective for all three loci, with no obvious bias during amplification, pooling or sequencing.

Evaluation and validation of AmpSeq markers in grapevine breeding families

**Flower sex.** To evaluate the 19 AmpSeq markers for flower sex, genotypic data were obtained for four breeding families in Experiment 1. This trait is reportedly controlled by a single locus with three alleles, with male (M) dominant over hermaphrodite (H), which is dominant over female (f). Three families segregated 1:1 for M/H—‘Horizon’×Illinois 547-1, ‘Horizon’×V. cinerea B9 and ‘Chardonnay’×V. cinerea B9. One family segregated 1:1 for H/f—MN1246×MN1264. Parents of the four families represent three phenotypes: ‘Horizon’ (H), ‘Chardonnay’ (H) and MN1246 (H) produce hermaphrodite flowers; Illinois 547-1 (M) and V. cinerea B9 (Mf) produce male flowers; and MN1246 (f) produces female flowers. Amplicon data were merged with flower sex field ratings for a one-way analysis of variance co-segregation analysis. For the pooled association test using families with the same ratings for a one-way analysis of variance co-segregation analysis. The results confirmed that AmpSeq markers for Ren2 PM resistance worked in a separate but related breeding family, even when evaluated by a different phenotyping method. A majority of the amplicons (9/12) can be used to track resistant alleles in phase, meaning that the minor allele is associated with PM resistance. AmpSeq markers with similar P-values and R² patterns clustered together (Figures 3e and f).

**Acylated anthocyanins.** To evaluate the 22 AmpSeq markers for acylated-anthocyanin content, association tests were executed on 182 individuals including: 82 progeny from ‘Horizon’×Illinois 547-1, a subset of the original family for marker development, and 100 progeny from V. rupestris B38×‘Horizon’ to evaluate marker transferability. In a pooled analysis, 1 amplicon had a non-significant association, and 15 of 22 amplicons had a −log₁₀(P-value) > 10 (Figure 2e, Table 4). Individually, these 15 AmpSeq markers explained an average of 33% of the phenotypic variance, comparable to the QTL contributing 54% of the phenotypic variance. As with the flower sex dendrogram, AmpSeq markers with similar P-values and R² patterns clustered together (Figures 3g and h).

**Criteria of efficient AmpSeq markers**

To reduce costs for primer synthesis and testing, we tested models to determine post hoc the relative importance of AmpSeq marker parameters to predict the most effective markers. Four statistical models (logistic regression, support vector machine, decision tree and random forest) were tested using quantified parameters of each primer to classify the primer output from the primer design pipeline into efficient and non-efficient categories. In all cases, regardless of the training set size, the random forest model (with over 90% predictive accuracy) outperformed the other three models (Figure 4). Thus, we developed a decision support tool based on the random forest model (Supplementary File 8), which can be used to facilitate AmpSeq marker selection following the primer design pipeline. The most important parameters in this model included the P-value of LD with the anchor marker, the primer annealing temperatures, the rejection code and the length of the primer.

---

**Table 1. QTL parameters for the development of AmpSeq markers**

| Trait Population                  | Trait Population                  | Trait Population                  | Trait Population                  | Trait Population                  | Trait Population                  | Trait Population                  | Trait Population                  |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| **Flower sex (M/H)**             | 'Chardonnay'×V. cinerea B9 (CC)   | 'Horizon'×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |
|                                  | 'Chardonnay'×V. cinerea B9 (HC)   | 'Horizon’×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |
| **Flower sex (M/H)**             | 'Chardonnay'×V. cinerea B9        | 'Horizon’×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |
| **Powdery mildew in controlled environment** | 'Horizon’×V. cinerea B9 (HC)      | 'Horizon’×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |
| **Acylated Anthocyanins**        | 'Horizon’×V. cinerea B9 (HC)      | 'Horizon’×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |
|                                  | 'Horizon’×V. cinerea B9 (HC)      | 'Horizon’×V. cinerea B9          | 'Hermaphrodite'×Illinois 547-1 (H) | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    | 'Horizon’×Illinois 547-1 (Hi)    |

Abbreviations: QTL, quantitative trait loci. R² (coefficient of determination): percentage of phenotypic variance explained by the QTL.
Table 2. AmpSeq marker-trait associations for flower sex

| Chr Position | Family of | AMPSeq design | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) | Flower sex (M/H) |
|--------------|-----------|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2             |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 4245542       | MN1246 × MN1264 (47) | B9 (82) | 0.12 | 0.7714 | 0.24 | 1.64E−27 | 0.8696 | 0.24 | 2.99E−40 | 1.38E−21 | 0.7714 |
| 4448335       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 4457222       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 4560461       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 4726521       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 5005636       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 5064267       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 5181254       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 5219969       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |
| 5220041       |          |               |                 |                 |                 |                 |                 |                 |                 |                 |                 |

**DISCUSSION**

This study had two main goals: (1) to develop a cost-effective and robust genotyping platform for MAS in heterozygous crops, and (2) to generate marker sets for MAS in grapevine breeding. A semi-automated primer design pipeline was developed to convert GBS tags to AmpSeq markers. The primer design procedure was facilitated by a decision support tool to predict which primers would perform appropriately in terms of predictability based on a pre-trained random forest model. For grapevine breeding, AmpSeq markers were developed for three economically important and representative traits with high breeding value: flower sex, PM resistance and acylated-anthocyanin concentration. The strategy was effective for all three traits given the diverse background and high heterozygosity of grapevine. In the process, 54 markers were tested on a total of 760 individuals in 6 breeding families involving 7 *Vitis* species across 2 breeding programs. The results indicated that the majority of AmpSeq markers have potential for accurate trait predictions, and a MAS package can be implemented for interspecific hybrid families.

Technical considerations of the AmpSeq strategy

Two key decision points for the usage of the AmpSeq primer design pipeline were the selection of an anchor marker and the definition of haploblock boundaries of each QTL. The optimal parameters depended on the nature of the trait and mapping family, and here we describe what worked for these three traits. First, each anchor marker was selected from our final GBS linkage map, after the HetMappS pipeline and genetic map curation removed SNPs that had segregation bias, low genotyping rate, putative sequencing error, redundancy or aligned to repetitive genome regions. Thus, the anchor marker was selected from high-quality, mapped SNPs. Second, the anchor marker had a high P-value for marker-trait association. Third, the anchor marker was within the 1.8 LOD interval of the QTL to ensure tight genetic linkage. The anchor marker was used to retrieve SNPs in LD from an un-filtered VCF file within the haploblock defined by the two boundary markers, to increase marker density under the QTL region and provide alternative alternative models for AmpSeq marker design. The range of the haploblock can be defined by the marker effect as suggested in ‘Materials and methods’, or by other haplotyping software specific for heterozygous species, like iXora or SHAPEIT2 (ref. 59).

Two custom options can be set for the amplicon primer design pipeline: one is P-value threshold of the LD test, and the other is the primer size, which was specified as 22 in this study to accommodate the 50 bp sequencing length and data analysis pipeline. To optimize the P-value of the LD test, all the parameters output from the pipeline were explored by four models (logistic regression, support vector machine, decision tree and random forest), in which the random forest model was the most useful to predict high-quality markers. In the random forest model, four primer parameters primarily contributed to the decision: the P-value of LD with the anchor marker, the reaction code and the annealing temperatures ($T_m$) of the forward and reverse primers. Supported by the split value of the decision tree model (Supplementary File 9) and general guidelines for PCR based on experimental experience and thermocycling conditions used here, the recommended criteria for efficient amplicon markers includes: (1) the default P-value of LD with the anchor marker should be 1e−25; (2) $T_m$ of primers should be between 52 and 70°C to obtain high depth and avoid missing data; (3) the rejection code should be 0, representing exactly two alleles identified in GBS tags; (4) extremely high read depths of GBS tags, which may imply
The notion of the predictive power is still attractive in breeding, when lower, which is consistent with theoretical framework. However, predictive accuracy, as they can explain 100% of phenotypic variance, the association between genotypes and phenotypes is of the QTL plays a major role. For a Mendelian trait such as effective markers and spurious conclusions in practice. Here, AmpSeq markers should be taken into consideration when segregation distortion, indicating the segregation pattern of family (Table 2). Segregation distortion, defined as observed minor allele frequency deviating from expected minor allele frequency of 0.5. Minor allele frequencies were calculated of each AmpSeq marker for each segregating family (Table 2). Segregation distortion, defined as observed minor allele frequency deviating from expected minor allele frequency of 0.25, was observed for some AmpSeq markers in all analyses. Similar distortion has been reported for SSR markers in grape, for GBS markers in rice and for sequence-tagged microsatellite site in chickpea. Although simulations claim that the presence of marker segregation distortion has little effect on linkage map construction and QTL analysis, distortion has resulted in less effective markers and spurious conclusions in practice. Here, AmpSeq markers with high P-values typically had extreme segregation distortion, indicating the segregation pattern of AmpSeq markers should be taken into consideration when choosing efficient markers.

Advantage of AmpSeq genotyping platform
The most significant advantage of AmpSeq for MAS is the ability to harness the high-resolution GBS or other NGS techniques during marker development, which provides resiliency against rapid LD decay in species with high heterozygosity and diversity, while nearly eliminating issues of missing data and heterozygote under-calling common to GBS. An average density of 55, 218 and 133 kb per AmpSeq marker was obtained for flower sex, PM resistance and acylated anthocyanins, respectively. In grapevine, SSR and indel markers reported to predict flower sex have failed in progenies resulting from the cross between complex North American hybrids (Bl Reisch and JJ Luby, personal communication), which may be due to the loss of linkage between the causal gene and markers. The situation has been similar for MAS of PM resistance. The closest pairs of Ren2 SSR markers were 550-kb apart in the 12X.2 reference genome. Four amplicon markers were generated by the AmpSeq strategy within the 550-kb region, targeting a lower probability of recombination between causal gene and markers, and improved ability to detect recombinations near the locus. AmpSeq gains efficiency via the NGS capacity to multiplex more markers per trait, more traits and more individuals. Our approach results in the genotyping of a haplotype block in one test for MAS decision making, which showed higher accuracy in cattle breeding and potential application in heterozygous crops such as cocoa and cotton, as well as for flowering time, cluster width and berry size in grapevine. While biallelic SNPs were traditionally criticized by limited polymorphisms per marker, the pooling of multiple SNP markers provides flexibility that overcomes this limitation. Further, multiple traits can be genotyped simultaneously for pyramiding, maximizing the application of MAS in contrast to phenotype screening. For example, anthocyanin pigments acylated to hydroxycinnamic acids are of interest to pigments acylated to hydroxycinnamic acids are of interest to-

| Chr | Position | Family of AmpSeq design | Mean depth | CV | Missing individuals | 'Horizon' × V. cinerea B9 (78) | 'Horizon' × Illinois 547-1 (91) |
|-----|----------|-------------------------|------------|----|--------------------|--------------------------------|--------------------------------|
|     |          |                         |            |    |                    | P-value | marker R² | Minor allele frequency | P-value | marker R² | Minor allele frequency |
| chr14 | 26205721 | HC                      | 65         | 38 | 2                  | 1.56E−02 | 0.1050 | 0.20               | 1.56E−05 | 0.1940 | 0.27               |
| chr14 | 26954543 | HC                      | 207        | 40 | 0                  | 9.23E−03 | 0.1174 | 0.27               | 2.24E−03 | 0.1309 | 0.47               |
| chr14 | 27052970 | HC                      | 1          | 0  | 380                | NaN      | 0       | 0.00               | NaN      | 0       | 0.00               |
| chr14 | 27118778 | HC                      | 52         | 21 | 1                  | 3.72E−03 | 0.1054 | 0.18               | 1.13E−07 | 0.2776 | 0.26               |
| chr14 | 27226381 | HC                      | 232        | 5  | 2                  | 2.42E−03 | 0.1148 | 0.19               | 1.45E−08 | 0.3101 | 0.26               |
| chr14 | 27742679 | HC                      | 244        | 2  | 1                  | 1.34E−03 | 0.1274 | 0.18               | 2.14E−08 | 0.3041 | 0.27               |
| chr14 | 28408093 | HC                      | 248        | 2  | 0                  | 6.12E−04 | 0.1440 | 0.19               | 3.82E−06 | 0.2165 | 0.27               |
| chr14 | 28415565 | HC                      | 248        | 2  | 0                  | 6.12E−04 | 0.1440 | 0.19               | 9.53E−06 | 0.2007 | 0.26               |
| chr14 | 28448817 | HC                      | 51         | 48 | 42                 | 3.78E−03 | 0.1399 | 0.34               | 1.05E−02 | 0.1205 | 0.37               |
| chr14 | 28686983 | HC                      | 248        | 2  | 0                  | 8.90E−03 | 0.0866 | 0.20               | 8.29E−06 | 0.2031 | 0.27               |
| chr14 | 28791319 | HC                      | 50         | 38 | 2                  | 1.32E−02 | 0.1089 | 0.25               | 6.44E−05 | 0.2010 | 0.34               |
| chr14 | 28824733 | HC                      | 101        | 49 | 2                  | 1.32E−02 | 0.0782 | 0.20               | 4.53E−05 | 0.1749 | 0.28               |

Abbreviations: CV, coefficient of variation of read depth across individuals; HC, 'Horizon' × V. cinerea B9; QTL, quantitative trait loci. Number of progeny analyzed is presented in parentheses. NaN means no association detected.
| Chr | Position | Family of AmpSeq design | Mean depth | CV | Missing | Acylated anthocyanins | Acylated anthocyanins | Acylated anthocyanins | Pooled (182) | 'Horizon' × Illinois 547-1 (82) | V. rupestris B38 × 'Horizon' (100) |
|-----|----------|-------------------------|------------|----|---------|-----------------------|-----------------------|-----------------------|--------------|--------------------------------|--------------------------------|
| chr3 | 14851567 | HI 250 0 0 | 1.07E-16 0.3370 0.33 | 1.07E-16 0.3370 0.33 | chr3 | 15171738 | HI 247 2 0 | 2.05E-17 0.3491 0.28 | 2.05E-17 0.3491 0.28 | chr3 | 15828002 | HI 250 0 127 | 1.76E-10 0.3102 0.47 | 1.76E-10 0.3102 0.47 |
| chr3 | 16007996 | HI 245 3 0 | 1.26E-16 0.3357 0.26 | 1.26E-16 0.3357 0.26 | chr3 | 16008017 | HI 244 3 0 | 1.26E-16 0.3357 0.26 | 1.26E-16 0.3357 0.26 | chr3 | 16162132 | HI 249 1 0 | 3.64E-17 0.3256 0.24 | 3.64E-17 0.3256 0.24 |
| chr3 | 16163517 | HI 249 1 0 | 5.22E-07 0.0042 0.05 | 5.22E-07 0.0042 0.05 | chr3 | 16246446 | HI 241 5 0 | 5.22E-07 0.0042 0.05 | 5.22E-07 0.0042 0.05 | chr3 | 16375440 | HI 249 1 0 | 3.64E-17 0.3256 0.24 | 3.64E-17 0.3256 0.24 |
| chr3 | 16402674 | HI 113 39 1 | 1.61E-06 0.1385 0.42 | 1.61E-06 0.1385 0.42 | chr3 | 16402706 | HI 14 6 2 | 2.15E-15 0.3143 0.33 | 2.15E-15 0.3143 0.33 | chr3 | 16473116 | HI 249 1 0 | 1.67E-16 0.3336 0.43 | 1.67E-16 0.3336 0.43 |
| chr3 | 16516083 | HI 4 4 200 | 1.80E-05 0.2417 0.35 | 1.80E-05 0.2417 0.35 | chr3 | 16520617 | HI 249 1 0 | 8.04E-17 0.3390 0.24 | 8.04E-17 0.3390 0.24 | chr3 | 16623313 | HI 73 53 3 | 1.85E-16 0.3329 0.30 | 1.85E-16 0.3329 0.30 |
| chr3 | 16774466 | HI 249 1 0 | 1.06E-17 0.3356 0.26 | 1.06E-17 0.3356 0.26 | chr3 | 16787032 | HI 4 4 200 | 1.80E-05 0.2417 0.35 | 1.80E-05 0.2417 0.35 | chr3 | 16873164 | HI 88 46 0 | 1.06E-17 0.3356 0.26 | 1.06E-17 0.3356 0.26 |
| chr3 | 17225240 | HI 20 27 2 | 9.37E-14 0.2861 0.22 | 9.37E-14 0.2861 0.22 | chr3 | 17225412 | HI 149 0 31 | 8.45E-03 0.0413 0.06 | 8.45E-03 0.0413 0.06 | chr3 | 17225412 | HI 149 0 31 | 8.45E-03 0.0413 0.06 | 8.45E-03 0.0413 0.06 |
| chr3 | 17225585 | HI 210 14 0 | 2.28E-17 0.3483 0.49 | 2.28E-17 0.3483 0.49 | chr3 | 17774559 | HI 62 151 124 | 1.97E-02 0.0649 0.31 | 1.97E-02 0.0649 0.31 | chr3 | 17774559 | HI 62 151 124 | 1.97E-02 0.0649 0.31 | 1.97E-02 0.0649 0.31 |

Abbreviations: CV, coefficient of variation of read depth across individuals; HI, 'Horizon' × Illinois 547-1; QTL, quantitative trait loci. Pooled includes all two families. Number of progeny analyzed is presented in parentheses. NaN means no association detected.
As mentioned above, marker transferability is another big concern in a breeding program with diverse germplasm spanning more than one species, and SNP microarrays for grapevine and apple are both reported to suffer from this constraint. Loss of polymorphism of SSR markers has also been reported when used in unrelated germplasm, as is the case for the marker gwm261 for the locus Reduced height 8 (Rht8) in wheat or Ren4 in grapevine. This can result from sampling bias in the original study, since a distinct allele may not be amplified for the marker.

The current work indicates that the amplicon markers are transferable in related families in grapevine. Among AmpSeq markers available in the current MAS package, dendrogram construction by genotype clustering could guide the choice of useful markers when transferring to new families without phenotypic data. In some, but not all cases, the markers most significantly associated with the trait clustered together (Figure 3), and in all cases markers not associated with the trait had poor correlations with the other markers. Still, because the amplicons are likely not genotyping the causal allele, re-training markers pools through validating, correcting and supplementing the current MAS package with markers developed in other breeding families by GBS or other genotyping platforms, following the same AmpSeq strategy, is recommended.

Finally, AmpSeq also provides advantages in terms of cost, time and ease of application. Actual cost to obtain AmpSeq data is comparable to single-locus SSRs and cheaper than most other multi-locus platforms. With the simple and straightforward PCR and library preparation protocol, the turn-around time from marker development through testing can be reduced to 1 month, which in our experience is quicker than other SNP platforms and SSRs. In addition, the data analysis pipeline can be automated to output results in a spreadsheet format, which may be attractive to some plant breeders, eliminating bioinformatic challenges typical of GBS data and removing the need for graphic interpretation typical of other genotyping platforms.

A case of MAS implementation in grape breeding

The three traits selected for testing are economically important and representative targets for North American hybrid winegrape cultivar development: flower sex, PM resistance and acylated anthocyanins. Hybridization with wild species takes place to introgress positive adaptive traits including disease resistance, pest resistance and cold hardiness, but fruit quality and yield from the cultivated species V. vinifera are also critical. The families used for testing are five winegrape breeding families from...
Figure 3. Marker transferability explored by dendrogram analysis with P-values of marker-trait association. (a) Flower sex for 'Horizon' × Illinois 547-1; (b) Flower sex for 'Horizon' × V. cinerea B9; (c) Flower sex for 'Chardonnay' × V. cinerea B9; (d) Flower sex for MN1246 × MN1264; (e) Powdery mildew resistance assessed in controlled environment for 'Horizon' × V. cinerea B9; (f) Powdery mildew resistance assessed visually in the field for 'Horizon' × Illinois 547-1; (g) Acylated-anthocyanin content for 'Horizon' × Illinois 547-1; (h) Acylated-anthocyanin content for V. rupestris B38 × 'Horizon.' Color key below indicates that shades of red color represent high correlation, while shades of blue color represent low correlation.
CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Michelle Schaub, Hema Kasinathan, Anna Nowogrodzki, Paige Appleton and Jackie Lillis for their technical support phenotyping PM resistance; Steve Luce and Peter Hemstad for helping to maintain the mapping families used in this study; Mani Awale for phenotyping flower sex at the South Dakota State University breeding program; Steve Bogdanowicz for helpful discussions when designing the wet lab techniques and barcoding strategy; and the Cornell University Biotechnology Resource Center for providing the necessary infrastructure and personnel to generate the raw genetic data analyzed and presented here. The US Department of Agriculture National Institute of Food and Agriculture Specialty Crop Research Initiative provided funding for this project (award no. 2011-51181-30635).

REFERENCES

1. Myles S. Improving fruit and wine: what does genomics have to offer? Trends Genet 2013; 29: 190–196.
2. Ru S, Main D, Evans K, Peace C. Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. Tree Genet Genomes 2015; 11: 1–12.
3. van Nocker S, Gardiner SE. Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. Hort Res 2014; 1: 14022.
4. Bus VM, Esenmeniad D, Buck E, Laurens FApplication of genetic markers in rosaceous crops.Folta K, Gardiner S (eds). Genetics and Genomics of Rosaceae vol. 6. Springer International Publishing: New York, NY, USA, 2009 pp 563–599.
5. Kanaagac E, Vargas A, de Andrés M, Carreño J, Carreño JMarker-assisted selection for seedlessness in table grape breeding. Tree Genet Genomes 2012; 8: 1003–1015.
6. Elbac R, Zyprian E, Welter L, Topfer R. The use of molecular markers for pyramiding resistance genes in grapevine breeding. Vitis 2007; 46: 120–124.
7. Zini E, Raffeiner M, Di Gaspero G, Elbach R, Grande MS, Letschka T. Applying a defined set of molecular markers to improve selections of resistant grapevine accesses. Acta Hortic 2015; 1082: 73–78.
8. Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 2008; 363: 557–572.
9. Varshney RK, Thudi M, Nayak SN, Gaur PM, Kishorwani J, Krishnamurthy L et al. Genetic dissection of drought tolerance in chickpea (Cicer arietinum). Theor Appl Genet 2014; 127: 445–462.
10. Reisch B, Owens C, Cousins PG, Chatenet MNM, Byrne DH(eds). Fruit Breeding vol. 8. Springer International Publishing: New York, NY, USA, 2012, 225–262.
11. Shitasawa K, Hirakawa H. DNA marker applications to molecular genetics and genomics in tomato. Breeding Sci 2013; 63: 21–30.
12. Boudrehi K, Bendahmane A, Cardinet G, Tracotec C, Moing A, Dirlewanger E. Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. BMC Plant Biol 2009; 9: 59.
13. Alpert KB, Tanksdale SD. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2:2. A major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci USA 1996; 93: 15503–15507.
14. De Franceschi P, Stegemeg T, Cabrera A, van der Knep E, Rosyara UR, Sebolt AM et al. Cell number regulator genes in Prunus provide candidate genes for the control of fruit size in sweet and sour cherry. Mol Breed 2013; 32: 311–326.
15. Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S and others. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 2013; 126: 2699–2716.
16. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 2013; 9: 29.
17. Huang XH, Han B. Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 2014; 65: 531–551.
18. Fodor A, Segura V, Denis M, Neenschwander S, Fournier-Level A, Chatelat P et al. Genome-wide prediction methods in highly diverse and heterozygous species: Proof-of-concept through simulation in grapevine. PLoS One 2014; 9: e110436.
19. Khan MA, Korban SS. Association mapping in forest trees and fruit crops. J Exp Bot 2013; 64: 4045–4060.
20. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK et al. Genetic structure and domestication history of the grape. Proc Natl Acad Sci USA 2011; 108: 3530–3535.
21. Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E et al. Rapid genomic characterization of the genusVitis. PLoS One 2010; 5: e28219.
22. Cattanaro F, Testolin R, Scalabrin S, Morgante M, Gaspero GD. Genetic diversity in the grapevine germplasm: Tuberosa R, Graner A, Frison Ef., et. al, Genomics of Plant Genetic Resources. Springer International Publishing: New York, NY, USA, 2014 pp 683–704.
23. Arnold C, Rossetto M, McNally J, Henry RJ. The application of SSRs characterized for grape (Vitis vinifera) to conservation studies in Vitaceae. Am J Bot 2002; 89: 22–28.
24. Dirlewanger E, Denoys-Rothan B, Yamamoto T, Chagné D. Genomics tools across rosaceous crops. Tree Genes Genomes 2010; 6: 689–700.
25. Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. Biology 2012; 1: 460–483.
80 Lambert SG, Asenstorfer RE, Williamson NM, Iland PG, Jones GP. Copigmentation between malvidin-3-glucoside and some wine constituents and its importance to colour expression in red wine. Food Chem 2011; 125: 106–115.

81 Chagne D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C et al. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 2012; 7: e31745.

82 Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P et al. A SNP transferability survey within the genus Vitis. BMC Plant Biol 2008; 8: 128.

83 Gasperini D, Greenland A, Hedden P, Drees R, Harwood W, Griffiths S. Genetic and physiological analysis of Rht8 in bread wheat: an alternative source of semi-dwarfism with a reduced sensitivity to brassinosteroids. J Exp Bot 2012; 63: 4419–4436.

84 Mahanil S, Ramming D, Cadle-Davidson M, Owens C, Garris A, Myles S et al. Development of marker sets useful in the early selection of Ren4 powdery mildew resistance and seedlessness for table and raisin grape breeding. Theor Appl Genet 2012; 124: 23–33.

85 Fennell A. Genomics and functional genomics of winter low temperature tolerance in temperate fruit crops. Crit Rev Plant Sci 2014; 33: 125–140.

86 Fennell A, Mathiason K, Luby J. Genetic segregation for indicators of photoperiod control of dormancy induction in Vitis species. Acta Hortic 2005; 689: 533–539.

87 Poland JA, Rife TW. Genotyping-by-Sequencing for plant breeding and genetics. Plant Genome 2012; 5: 92–102.

88 Mathiason K, He D, Grimplet J, Venkateswari J, Galbraith DW, Or E et al. Transcript profiling in Vitis riparia during chilling requirement fulfillment reveals coordination of gene expression patterns with optimized bud break. Funct Integr Genomics 2009; 9: 81–96.

89 Sreekantan K, Mathiason K, Grimplet J, Schlauch K, Dickerson JA, Fennell AY. Differential floral development and gene expression in grapevines during long and short photoperiods suggest a role for floral genes in dormancy transitioning. Plant Mol Biol 2010; 73: 191–205.

90 Garris A, Clark L, Owens C, McKay S, Luby J, Mathiason K et al. Mapping of photoperiod-induced growth cessation in the wild grape Vitis riparia. J Am Soc Hortic Sci 2009; 134: 261–272.

Supplemental Information for this article can be found on the Horticulture Research website (http://www.nature.com/hortres).