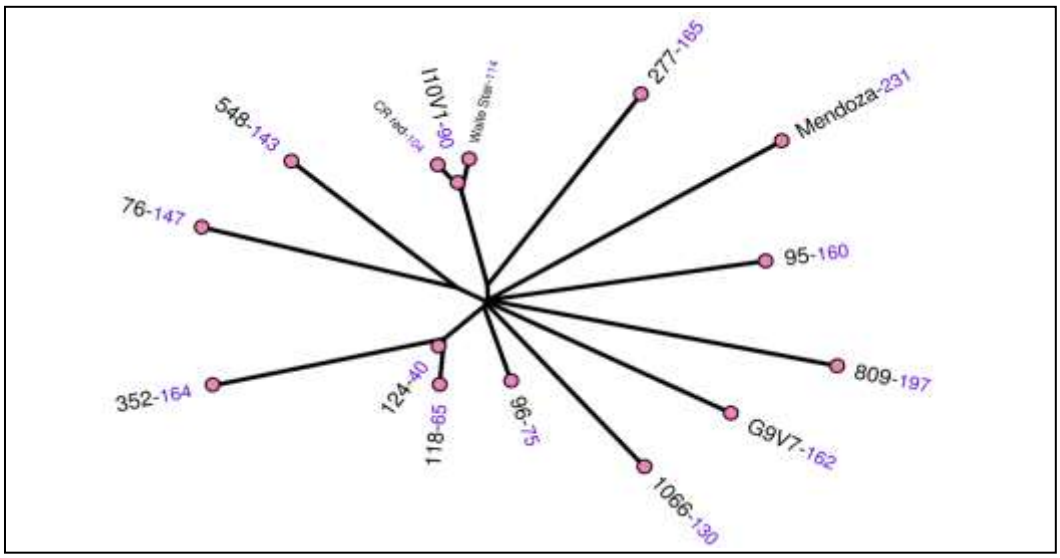
Australian Government
Wine Australia

The Australian Wine Research Institute

# Characterising genomic diversity in Australia's grapevine germplasm




## 1. FINAL REPORT TO WINE AUSTRALIA

Project number: **AWRI 1701-4.3.1**

Principal investigators:
**Dr Simon Schmidt and Dr Anthony Borneman**

Research organisation:
**The Australian Wine Research Institute**

Date: **9 October 2019**

**Project title:** Characterising genomic diversity in Australia's grapevine germplasm

**Authors:** Simon Schmidt, Michael Roach, Anthony Borneman

**Date:** 9 October 2019

**Address:** The Australian Wine Research Institute, Wine Innovation Central Building, Hartley Grove, cnr Paratoo Rd, Urrbrae (Adelaide), SA 5064

**Disclaimer/copyright statement:**

# Table of contents

# 1. Abstract

There are numerous clones of Chardonnay available that exhibit differences in key viticultural and oenological traits; however, the genetic variation that underlies these differences remains largely unknown. To address this knowledge gap, a Chardonnay reference genome was produced and used to compare sequence data from 15 different Chardonnay clones. A total of 1620 markers were identified that distinguish the 15 clones. The parentage of Chardonnay was explored by mapping sequence data from its parents (Gouais Blanc and Pinot Noir) against the Chardonnay reference genome. The analysis of genome sequence data indicates that Pinot Noir and Gouais Blanc share an extremely high degree of kinship.

# 2. Executive summary

Two pre-existing data sets, a high-quality reference genome for Chardonnay I10V1 and high-coverage whole-genome sequence data for 15 Chardonnay clones, were used to identify 1620 high-quality inter-clone nucleotide variants. There were limited shared somatic mutations among the Chardonnay clones, especially outside the highly related I10V1 group. Markers were reliably detected at coverages as low as 9.8-fold. Most of the identified clone-specific markers were also recovered from independently sourced clonal material.

Access to a diploid reference genome for Chardonnay also provided the means to unravel the detailed genetic ancestry of this cultivar and its parents, Pinot Noir and Gouais Blanc. Chardonnay matches both haplotypes of Pinot Noir across approximately one-fifth of its genome and these areas include large tracks of both homozygous and heterozygous variation. Mapping of the Gouais Blanc genome that was also generated through this project to the Chardonnay genome revealed an overlap between regions that mapped to Pinot Noir and regions that mapped to Gouais Blanc. This confirms that Gouais Blanc does indeed have a very close kinship relationship with Pinot Noir.

# 3. Background

The identification of grapevine varieties is complex due to the legacy of many centuries of cultivation in many countries, and the significant movement of material between regions and countries. Using microsatellite and single nucleotide polymorphism (SNP) DNA markers, a reliable identification of grapevine varieties can be achieved. However, high quality reference genomes are required to understand and to map the genomic variation in the grapevine, which is responsible for phenotypical and functional differences between varieties and clones.

The enabling technology for reference genome construction in the grapevine is in rapid development, and this project was established to build on know-how and tools developed as part of previous AWRI projects with Australian and international partners. Briefly, for any particular grapevine variety or species lacking an existing reference sequence, sequencing can be performed using long-read technology, and the resulting data can be *de novo* assembled to form reference-quality assemblies. For varieties/species with a known reference genome, re-sequencing can be performed to map genetic diversity. This information will enable the development of genetic tests capable of identifying clones of wine-grape cultivars.

In cases where genetic data are required by future projects for other clones or varieties, there are opportunities to build on the fundamental knowledge gained through assembly of a diploid genome of Chardonnay and evaluation of the genetic diversity extant within the clones of this cultivar.

# 4. Project aims

This project, together with others underway in the Australian grape and wine community, contributed to the following broad aim:

- The sector has an improved understanding of the genetic resource pool available in Australia's grapevine germplasm.

The project contributed to this aim by working toward the key deliverable of improving 'Understanding of clonal diversity in Chardonnay'. It achieved this through consolidating previously generated whole genome sequence data and using it to:

- Characterise genetic variation in at least 10 clones of Chardonnay
- Validate clone-specific genomic variants through the analysis of genome sequence data obtained from independently sourced plant material
- Investigate the parentage of Chardonnay through sequencing of Gouais Blanc
- Prepare publications and communications summarising the study of Chardonnay clonal variation

# 5. Method

The methods used to obtain genomic DNA from grapevine leaf material, generate whole genome sequence data from that extracted DNA, assemble a Chardonnay reference genome and use that reference genome to search for clone-specific single nucleotide variants are described in detail in two publications by Roach et al. (2018a, b). The Chardonnay clones used in this work are listed in Table 2 in the Results section of this report.

Leaf material for Gouais Blanc was sourced from the SARDI Research Station at Nuriootpa and DNA was extracted using a Qiagen DNeasy Plant Mini Kit. Gouais Blanc DNA was prepared for sequencing using a Nextera DNA Flex library prep and sequenced on an Illumina NextSeq500 platform using a 2 x 150 bp mid output run.

To identify the most likely parent for each phase-block pair, publicly available short-read sequencing data were obtained for three clonally derived variants of Pinot Noir: Pinot Blanc, Pinot Gris, and Pinot Meunier (BioProject: PRJNA321480). Data for Pinot Noir were not available at the time of analysis. To avoid potential issues with data from any single Pinot variety, pooled reads from all three were used for mapping. The sequencing data for Pinot, Chardonnay, and Gouais Blanc were mapped to previously generated Chardonnay reference genome primary contig and haplotig phase-block sequences using BWA-MEM v0.7.12 (Li 2013).

PCR duplicates and discordantly mapped reads were removed, and poorly mapping regions were masked using a window coverage approach. Heterozygous SNPs were called using VarScanv2.3 (Koboldt et al. 2012) (p-value < 1e−6, coverage > 10, alt reads > 30%) and identity-by-state (IBS) was assessed over 10 kb windows (5 kb steps) at every position where a heterozygous Chardonnay SNP was found.

Where the parent (Pinot or Gouais Blanc) was homozygous and matched the reference base, an IBS of 2 was called. Where the parent was homozygous and did not match the reference base, an IBS of 0 was called. Finally, where the parent and Chardonnay had identical heterozygous genotype, an IBS of 1 was called. The spread of IBS calls was used to assign 10 kb genome sequence windows as ‘Pinot’, ‘Gouais Blanc’, or ‘double-match’. The window coordinates were then transformed to chromosome-ordered scaffold coordinates and neighbouring identically called windows were chained together. Complementary Pinot/Gouais Blanc calls from the parent datasets were merged and clashing calls removed. For ease of visualisation, the ‘double-match’ calls from the Pinot dataset were merged with the Gouais Blanc calls (and vice versa). A SNP density track for the Chardonnay primary contigs was created over 5 kb windows from previously mapped Illumina reads.

# 6. Results/discussion

The project deliverable was to improve 'understanding of clonal diversity in Chardonnay'. The following section outlines the approaches undertaken to achieve this goal.

## 6.1. Genetic variation in clones of Chardonnay

As for many commercial grapevine varieties, there are currently many clones of Chardonnay, with each exhibiting a unique range of phenotypic traits. However, unlike varietal development, all of these genetic clones were established through the repeated asexual propagation of cuttings that presumably trace back to an original Chardonnay plant. It is therefore an accumulation of somatic mutations that has contributed to phenotypic differences that uniquely define each clone and which provide an avenue for the confirmation of a clone's identity.

Mapping of short-read re-sequencing data to a high-quality Chardonnay reference genome was used to define single nucleotide variation across 15 different Chardonnay clones (Table 2). The analysis of these highly related genomes (separated by a low number of true SNPs) was facilitated through applying a stringent kmer-based filter to remove false positives (including those calls due to sequencing batch or individual library size distribution at the expense of some false negative calls). After filtering, 1620 high-confidence marker variants, evenly distributed across the Chardonnay genome, were identified. Of these, 1479 were unique to individual clones, with the remainder (141) shared between different clone sets.

Unique clonal markers are an obvious foundation for the establishment of tests for clone identification. Shared markers are indicators of shared clonal heritage. The most prominent relationship apparent in this dataset is that between I10V1, CR Red and Waite Star. Both Waite Star and CR Red share all the markers of I10V1 (60 markers) indicating that these two Chardonnay variants are sports of Australia's most commonly planted Chardonnay clone. Similarly, clones 118 and 352 share almost all of the markers identified in clone 124 (24 markers), suggesting that both are derived from 124. Only two markers were identified that could be used to distinguish 124 from 118 and 352, both of which were recovered from low-confidence base calls.

The accumulation of SNPs can also lead to phenotypic differentiation that underlies the clonal selection process. A combination of Annovar (Wang et al. 2010) and Provean (Choi and Chan 2015) were used to annotate and predict the impact of clonal markers on the amino acid sequence of proteins for each of the clones. This approach correctly identified a previously characterised Muscat mutation (S272P) in DXS1 (Emanuelli et al. 2010) in the only Chardonnay clone in this study (clone 809) that is known to display the Muscat character, providing validation for the sequencing and data analysis strategy chosen for this project. In addition to this known Muscat mutation, an additional 55 marker mutations were identified that displayed a high chance of affecting protein function (Table 3). However, further work is required to investigate the links between known inter-clonal phenotypic variation and these specific mutations.

**Table 1.** A summary of Chardonnay clonal marker variants

| SAMPLE GROUP | NUMBER OF SNPS AND INDELS |
|---|---|
| *MENDOZA* | 221 |
| *809* | 187 |
| *95* | 150 |
| *G9V7* | 143 |
| *277* | 137 |
| *76* | 137 |
| *548* | 133 |
| *352* | 121 |
| *1066* | 120 |
| *96* | 61 |
| *CR RED, WAITE STAR, I10V1* | 60 |
| *118* | 27 |
| *WAITE STAR* | 26 |
| *118, 352, 124* | 24 |
| *CR RED, WAITE STAR, I10V1, 277* | 18 |
| *CR RED* | 14 |
| *352, G9V7* | 11 |
| *76, 548* | 10 |
| *CR RED, WAITE STAR, I10V1, MENDOZA, G9V7, 95, 277, 352, 96, 1066, 124, 118, 809* | 8 |
| *118, 124, 96* | 4 |
| *CR RED, WAITE STAR, I10V1, MENDOZA, 809, 95, 277* | 2 |
| *CR RED, I10V1* | 2 |
| *118, 1066, 124, 96* | 2 |
| *124* | 2* |

*alternate base calls were very low for these two variants.

## 6.2. Validation of clone-specific genomic variants

To validate suitability of the markers for clonal identification, Chardonnay clones 76, 96, 277, 352, 548 and G9V7 were independently sourced from Foundation Plant Services, USA and a further three clones (76, 95, 96) were sourced from Mission Hill Family Estate, Quail's Gate and Burrowing Owl wineries in British Columbia, Canada. DNA from these samples was extracted and sequenced at a different location from that used for the generation of the original training data set (Michael Smith Genome Sciences Centre, British Columbia Cancer Research Centre, Vancouver, Canada) using different library preparation methods (Illumina Nextera) and sequenced on different platforms (Illumina HiSeq 2500 and MiSeq). The aim of this work was to ensure that any clone-specific markers identified would be robust toward the analytical method employed for their detection.

High-coverage (142- to 244-fold) sequencing was performed for three of these independently sourced clones (76, 95 and 96). Kmer analysis of these high-coverage samples identified between 55% (clone 96) and 83% (clone 76) of the expected markers for each sample (Figure 1). Missing markers are likely due to poor sequence coverage across the marker region. This can occur due to differences in sequencing library preparation. Some library preparation methods are known to introduce biases into sequence data sets. However, despite being the same clones, there was a significant proportion (14% to 44%) of the expected markers in each of the three samples that were not found in the independent material and which could not be attributed to insufficient marker loci coverage. This indicates that there may be some intra-clonal genetic variation that has accumulated during the independent passaging of clonal material. In summary, the markers shared between the discovery and validation datasets form validated clonal identification markers; however, the unshared markers may prove useful in assigning lineage or regionality in supplement of clonal identification.

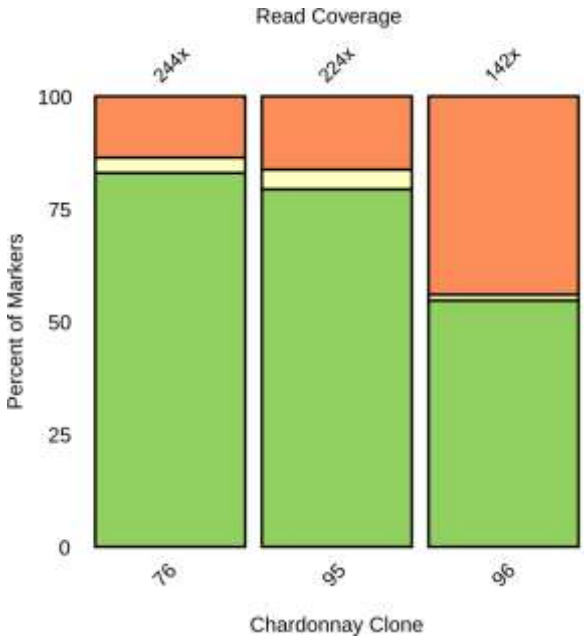



**Figure 1**. Recovery of marker SNPs from independently sourced material. Green – marker hits, yellow – insufficient coverage to make a marker call, orange – missing markers.

Marker discovery is typically undertaken at 100-fold coverage or greater. However, as the level of sequencing coverage ultimately affects the economics and robustness of clonal testing by sequencing, the impact of sequencing depth on marker identification was further assessed.

Data from the pooled results of two sequencing batches for independently sourced clone 95 was subsampled to a range of coverages and then screened for effectiveness of marker recovery. In the subsampled dataset, there was little difference in the number of recovered markers from 200-fold down to 25-fold coverage (79% to 63% respectively), and only a 6% decrease in markers confidently flagged as missing. At

12-fold coverage it was still possible to detect 35% of the markers for this clone, although a large deterioration in marker recovery due to poor coverage was evident (Figure 2).

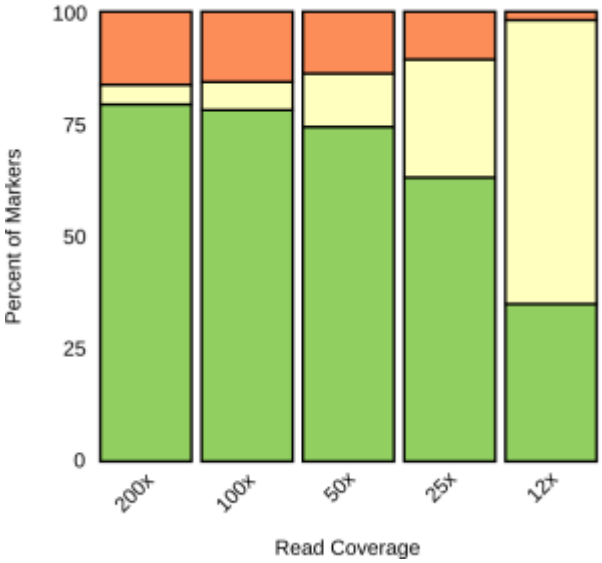



**Figure 2**. Recovery of marker SNPs from independently sourced clone 95 over a range of sequencing coverages. Green – marker hits, yellow – insufficient coverage to make a marker call, orange – missing markers.

Following the results of coverage titration indicating high rates of marker recovery at lower sequencing coverage, low coverage (9.8- to 24.8-fold) datasets were obtained from independent material of six additional clones. Despite the combination of independent material and low coverage it was still possible to detect between 7.9% (clone 352) and 42% (clone 76) of the expected markers for each sample.

Given the large number of unique markers for the majority of the clones evaluated here, reduced sequencing coverage (25-fold) would appear to be an efficient means to achieve clonal identification by sequencing.

## 6.3. Insights into the parentage of Chardonnay

Chardonnay has previously been reported to be the result of a cross between Pinot Noir and Gouais Blanc (Bowers et al. 1999, Hunt et al. 2010). These parentage assignments, and indeed the determination of many relationships between wine-grape cultivars, are based on relatively small numbers (32) of simple sequence repeat (SSR) markers. Whole genome sequencing combined with highly contiguous phased genome assemblies provides distinct advantages when attempting to understand complex and entwined parentage relationships such as those of wine-grape cultivars. Instead of the 32 SSR markers using in microsatellite analysis, the highly heterozygous grapevine genome provides more than 2.4 million markers when comparing Chardonnay with Pinot Noir (for example). This massive increase in marker number enables high resolution mapping of genomic regions and allows parental inheritance to be resolved in unparalleled detail.

The high density of SNPs in a comparison of Pinot Noir and Chardonnay was exploited in an attempt to identify the parental origin of each allele in the diploid Chardonnay

assembly. Phase blocks were assigned across the genome by stringently aligning and trimming both the primary contigs and haplotigs into pairs of closely aligning syntenic sequence blocks (P and H alleles). This produced 1153 phase-blocks covering 270 Mb of the 490 Mb genome (71% of the haplotigs). Each pair of phase blocks should have one allele inherited from each parent. To assign likely genomic parentage within each phase block, short-reads from Gouais Blanc, and a merged dataset comprising sequencing reads from several different genetic clones of Pinot Noir (Pinot Blanc, Gris, and Meunier, hereafter referred to as Pinot) (Marroni et al. 2017) were mapped to the phase block sequences. The proportion of inherited nucleotide variation (using heterozygous variant loci) was then used to attribute the likely parentage of each block (Figure 3).

Using the above described approach, it was possible to confidently assign parentage to 197 Mb of the 244 Mb of chromosome-ordered phase-blocks. Interestingly, rather than a 1:1 ratio of [Gouais Blanc to Pinot] matches, which would be expected if Pinot and Gouais Blanc were unrelated, Pinot was shown to match a higher proportion of the phase blocks (49% versus 34%). Further complicating this imbalance was the observation that in the remaining 17% of assigned regions, the pattern of nucleotide variation across the two heterozygous Chardonnay haplotypes matched both haplotypes of Pinot, with one of these haplotypes also matching one of the Gouais haplotypes. These ‘double Pinot haplotype’ regions are in some cases many megabases in size and are indicative of a common ancestry between Pinot and Gouais Blanc.

An orthogonal kmer-based approach was used to call identity-by-state (IBS) over the primary contigs and haplotigs to ensure that the data were not biased by analysing only the phase blocks. There was generally excellent concordance between the SNP- and kmer-based IBS methodologies.

While the presence of the homozygous ‘double-Pinot Noir’ regions could be the result of a high number of large-scale gene conversion events early in Chardonnay’s history, the numerous heterozygous double-Pinot Noir regions are only possible if the haplotype inherited from Gouais Blanc was almost identical to the non-inherited allele of Pinot Noir. Gouais Blanc sequencing indeed confirmed that within these ‘double-Pinot Noir’ regions, one of the two Pinot Noir haplotypes is a match for an allele of Gouais Blanc.

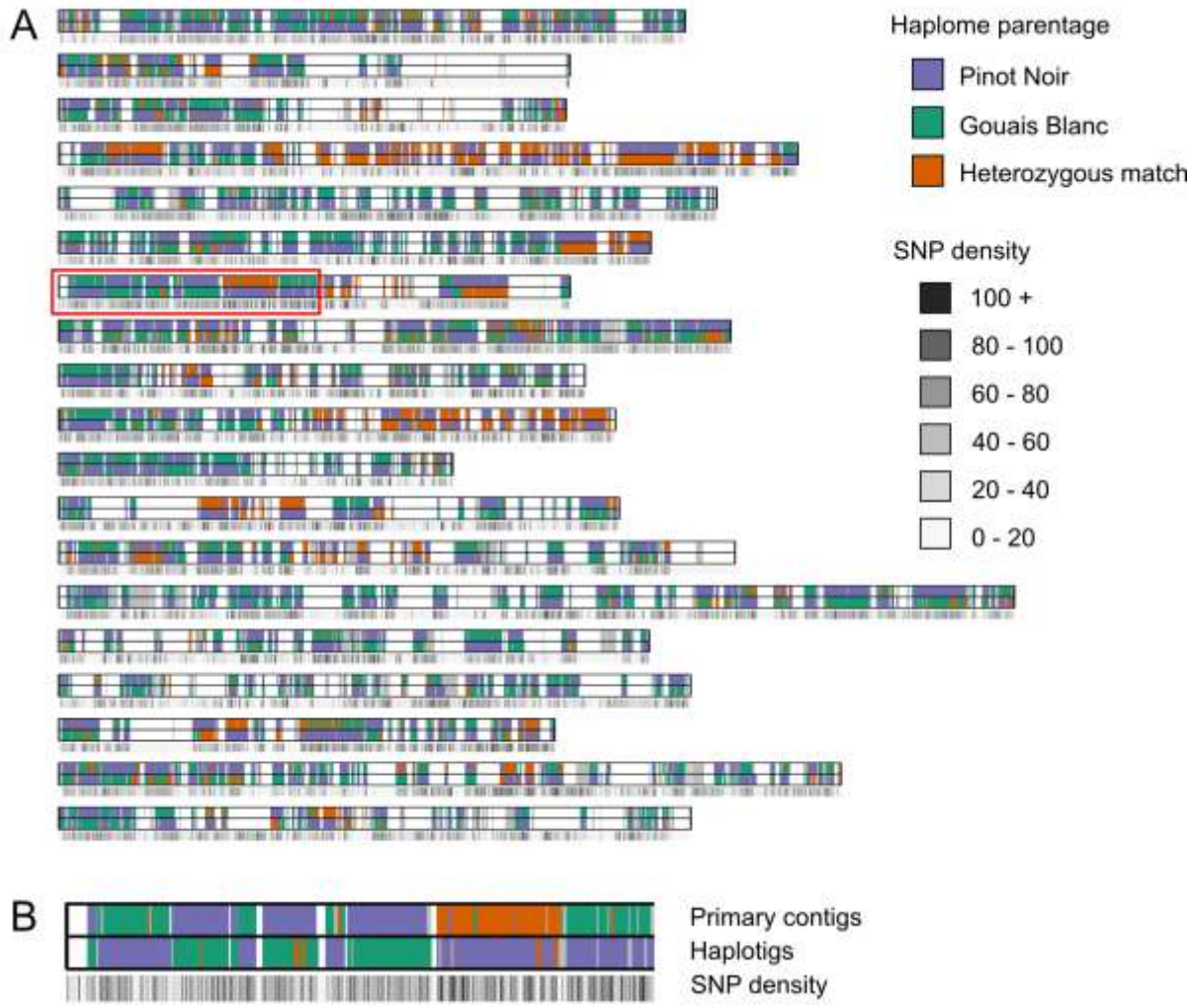



**Figure 3.** Parental architecture of the Chardonnay genome. A – representation of phased Chardonnay chromosomes coloured according to parental origin (in purple and green) and aligned with a SNP density track underneath each chromosome. B – an enlargement of *Vitis vinifera* chromosome 7, red box above.

# 7. Outcome/conclusion

## 7.1. Performance against planned outputs

The project deliverable was to improve 'understanding of clonal diversity in Chardonnay'. This was achieved through characterising single nucleotide polymorphisms in 15 Chardonnay clones and verifying their existence in an independent dataset. In addition, by comparing a reference genome of Chardonnay to Pinot Noir, a more detailed understanding of the relationships between Chardonnay and its parents was resolved. These achievements are summarised in more detail below.

### 7.1.1. Characterise genetic variation in at least 10 clones of Chardonnay

Work performed towards this output has provided a set of 1620 single nucleotide markers that together are capable of uniquely identifying 15 Chardonnay clones from an unknown grapevine-derived sample. The learnings gained from the performance of this work can readily be extended to other clones or wine-grape cultivars.

### 7.1.2. Validate clone-specific genomic variants through the analysis of genome sequence data obtained from independently sourced plant material

The markers identified in output 7.1.1 were successfully validated against material sourced from an independent international plant germplasm provider, allowing validation of marker sets. Not only were the markers validated against alternative primary material, but alternative sequencing and analytical approaches were also used to demonstrate the analytical robustness of the approach.

### 7.1.3. Investigate the parentage of Chardonnay through sequencing of Gouais Blanc

Genetic anomalies observed in the mapping of Pinot sequence data to the Chardonnay reference genome led to a hypothesis that Gouais Blanc must have a direct relationship with Pinot. This idea was tested by sourcing and sequencing an example of Gouais Blanc. Mapping of this Gouais Blanc genome to the Chardonnay genome revealed an overlap between regions that mapped to Pinot and regions that mapped to Gouais Blanc confirming that Gouais Blanc does indeed have a kinship relationship with Pinot. Understanding the exact nature of that relationship will require further work.

## 7.2. Practical implications

To date no reliable means of clonal identification or verification exists such that suppliers and purchasers of clonal material can provide/receive certifications of clonal authenticity. Similarly, managers of vineyards are unable to verify which varietal clone they are using to produce grapes. Results generated in this project provide the foundations for such a test for the wine-grape cultivar Chardonnay.

## 7.3. Benefits to the industry

Having generated the basis for a clone identification test for Chardonnay and by releasing that information into the public domain, it is now possible for that information to be used for clone identification. Such a test could enable purchasers of established vineyards to verify what they are buying; in addition, it would give confidence to suppliers as a means of quality control for material prior to importation or distribution. It would also underpin clonal research which may seek to better understand geographically based clone performance metrics.

# 8. Recommendations

## 8.1. Future directions

This project has established methods for the production of wine-grape reference genomes and methods of comparative genomics that enable the classification of clonal material based on genomic data. To date this has only been undertaken for key clones from a single cultivar, Chardonnay, although it is noted that similar work has been undertaken at the AWRI in less detail for Shiraz, in a separately funded Wine Australia project (SAR 1303).

Given the ongoing innovations in sequencing technology and sequencing capacity improvements in recent years, sequencing of all cultivars and all clones of each cultivar is possible within the timeframe and budget of individual research projects. Whether this occurs within Australia or not, it will be necessary to re-sequence Australian examples of key varieties and clones in order to validate markers used for clone identification here and abroad. Such an undertaking would also serve to catalogue Australia's own extensive and unique wine-grape heritage. A broader and more detailed understanding of Australia's wine-grape resources would enrich the story of Australia's grapegrowing and winemaking history, enable focused clonal selection or evaluation in-country, enable targeted importation of complementary germplasm, and help to cement the future of the grapegrowing industry in this country.

## 8.2. Priorities for further R&D, extension and policy

- The availability of cultivar- and clone-level genome sequence datasets is still very limited. Research priorities could include an expansion of grapevine sequencing efforts to encompass all major Australian cultivars and clones.
- At the time of writing there are no diagnostic tests for wine-grape clones available either on the Australian market or internationally. Development priorities would begin with the evaluation of suitable technologies and approaches toward establishment of a diagnostic test for grapevine clones using the available Chardonnay dataset as a test case. The same test could also be used for cultivar authentication once a broader dataset becomes available.
- Genomic diagnostics for plants and pathogens are an inevitable part of the future of agriculture. Yet, the general understanding among grapegrowers and winemakers, and indeed the wider community, of the opportunities that may be derived from modern high-throughput sequencing technology and computational informatics is still limited. Extension activities could focus on education about the fundamentals of modern genomic technologies and the implications of technology developments for the future of the industry. Upskilling the workforce, or key individuals within it, will help to lay the foundations for a more informed use of these technologies in the future and will prepare industry for understanding and using the results of genetics-based diagnostic tests that are to come.

# 9. Appendix 1: Communication

## 9.1. Communication of the outcomes

Outcomes and knowledge generated during the progress of this project were communicated to peers through peer-reviewed publications and conference presentations.

Information was extended to industry stakeholders through presentations and workshops at the 17th Australian Wine Industry Technical Conference in July 2019.

## 9.2. Journal articles written during the project

Roach, M.J., Schmidt, S.A., Borneman, A.R. 2018a. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19: 460.

Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J.J., Jones, S.J.M., Pretorius, I.S., Schmidt, S.A., Borneman, A.R. 2018b. Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay*. PLoS Genet.* 14: e1007807.

# 10. Appendix 2: Intellectual property

Outputs arising from the project are freely available for industry, government agencies and research providers (Roach et al. 2018a, b). Raw sequence data is available via public repositories at BioProject accession PRJNA399599 at the National Center for Biotechnology Information in the USA.

# 11. Appendix 3: References


Anderson, M.M., Smith, R.J., Williams, M.A., Wolpert, J.A. 2008. Viticultural evaluation of French and California Chardonnay clones grown for production of sparkling wine. *Am. J. Enol. Vitic.* 59: 73–77.

Bettiga, L. 2003. Comparison of seven Chardonnay clonal selections in the Salinas Valley. *Am. J. Enol. Vitic.* 54: 203–206.

Boss, P.K., Davies, C., Robinson, S.P. 1996. Anthocyanin composition and anthocyanin pathway gene expression in grapevine sports differing in berry skin colour. *Aust. J. Grape Wine Res.* 2: 163–170.

Bowers, J., Boursiquot, J., This, P., Chu, K., Johansson, H., Meredith, C. 1999. Historical genetics: The parentage of Chardonnay, Gamay, and other wine grapes of northeastern France. *Science* 285: 1562–1565.

Choi, Y., Chan, A.P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31: 2745–2747.

Cirami, R., Ewart, A.J.W. 1995. Clonal selection, evaluation and multiplication in Australia. Wolpert, J., Walker, M.A., Roberts, D. (eds.). *Proceedings of the International Symposium on Clonal Selection June 20 & 21, Oregon Convention Center, Portland, Oregon*: American Society of Enology and Viticulture: 52–59.

Emanuelli, F., Battilana, J., Costantini, L., Le Cunff, L., Boursiquot, J.-M., This, P., Grando, M.S. 2010. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* 10: 241.

Hunt, H.V., Lawes, M.C., Bower, M.A., Haeger, J.W., Howe, C.J. 2010. A banned variety was the mother of several major wine grapes. *Biol. Lett.* 6: 367–369.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 22: 568–576.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997.

Longbottom, M.L., Dry, P.R., Sedgley, M. 2008. Observations on the morphology and development of star flowers of *Vitis vinifera* L. cvs Chardonnay and Shiraz. *Aust. J. Grape Wine Res.* 14: 203–210.

Marroni, F., Scaglione, D., Pinosio, S., Policriti, A., Miculan, M., Di Gaspero, G., Morgante, M. 2017. Reduction of heterozygosity (ROH) as a method to detect mosaic structural variation. *Plant Biotechnol. J.* 15: 791–793.

Nicholas, P. 2006. Grapevine clones used in Australia. SARDI. Available from: http://www.graftedvines.com.au/images/Grapevine%20Clones%202006.pdf

Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J.J., Jones, S.J.M., Pretorius, I.S., Schmidt, S.A., Borneman, A.R. 2018a. Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PLoS Genet.* 14: e1007807.

Roach, M.J., Schmidt, S.A., Borneman, A.R. 2018b. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19: 460.

Sweet, N.L. 2007. Chardonnay history and selections at FPS. FPS Grape Program Newsletter. Available from: https://ucanr.edu/sites/intvit/files/24489.pdf

Wang, K., Li, M., Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 38: e164–e164.

Wolpert, J., Kasimatis, A., Weber, E. 1994. Field performance of six Chardonnay clones in the Napa Valley. *Am. J. Enol. Vitic.* 45: 393–400.

# 12. Appendix 4: Staff

Simon A. Schmidt
Anthony R. Borneman
Michael J. Roach
Radka Kolouchova

# 13. Acknowledgements

This work was supported by Australia's grapegrowers and winemakers through their investment body Wine Australia with matching funds from the Australian Government. The AWRI is a member of the Wine Innovation Cluster in Adelaide, SA.

The authors would like to thank Bioplatforms Australia for their support throughout this project, including during the earlier establishment and data generation phases of the project. Thanks are also extended to SARDI, in particular Dr Mike McCarthy, for the provision of Gouais Blanc leaf material and to Nick Dry at Yalumba Nursery for his support.

# 14. Appendix 5: Supplementary material

**Table 2**. Information relating to Chardonnay clones used in this study

| | Alternate name ** | Origin† | Yield* (kg/vine) | Ripening time | Comments |
|---|---|---|---|---|---|
| **CR red** | | Clare, South Aust. | | | Red-coloured berries |
| **I10V1** | FPS06, Heat treated Wente clone | UCD (Olmo68 - FPS05) | 8.7 | Early | |
| **Mendoza** | UCD1 | | 8.4, low | Late | Prone to millerandage |
| **G9V7** | FPS05, Olmo69 (Wente or Martini selections) | Stanley Lane Vineyards, Carneros [83] | 8.5, 12.9, 5.2 | | |
| **95** | FPS73,38 | Côte-d'Or, Dijon | 6.4 | late ripening | Sour rot in northern climate |
| **76** | FPS69, FPS76 | Côte-d'Or, Dijon | 5.6 | late | |
| **352** | FPS41 | Côte-d'Or, l'Espiguette | 6.4 | late | |
| **277** | FPS 42, 49 and 51 | Dijon | | late | |
| **96** | FPS70, FPS96 | Côte-d'Or, Dijon | 6.4 | late | |
| **1066** | | Côte-d'Or, ENTAV | Low | | Prone to millerandage |
| **124** | FPS84,98 | Côte-d'Or, Champagne | 5.9 | late | |
| **Waite Star** | | | | | Seedless, shiny hairless leaves |
| **118** | FPS104 | Côte-d'Or, Champagne | 5.9 | Early | |
| **809** | | Saône-et-Loire, ENTAV | Low – medium | Early | Muscat character |
| **548** | FPS548 | Saône-et-Loire, ENTAV | Low-medium | late | |

* From Cirami and Ewart (1995), Wolpert et al. (1994), Bettiga (2003) and Anderson et al. (2008).
** From Sweet (2007) and Nicholas (2006)
† From Boss et al. (1996), Sweet (2007) and Longbottom et al. (2008).

Table 3 Predicted effect of single nucleotide polymorphisms within open reading frames on protein function

| Sample Group (clone1, clone2, …) | mutation_type | nucleotide substitution | predicted protein substitution | contig | start | provean score | top_UniprotKB_hits |
|---|---|---|---|---|---|---|---|
| ***CR red*** | stopgain | 118G>T | G40X | 000062F | 830894 | - | |
| | synonymous SNV | 1170T>C | P390P | 000538F | 151050 | - | Serine carboxypeptidase-like 25 |
| ***CR red, Waite Star, I10V1*** | nonsynonymous SNV | 1196G>A | R399K | 000010F | 1037878 | 0 | |
| | stopgain | 770G>A | W257X | 000013F | 1038978 | - | |
| | stopgain | 368C>G | S123X | 000068F | 903834 | - | |
| | nonsynonymous SNV | 940G>A | D314N | 000146F | 437038 | -0.843 | Amidophosphoribosyltransferase, chloroplastic |
| | nonsynonymous SNV | 1295A>G | E432G | 000595F | 131111 | -6.403 | RNA polymerase sigma factor SigB |
| | synonymous SNV | 189C>A | T63T | 000784F | 60239 | - | |
| ***CR red, Waite Star, I10V1, 277*** | nonsynonymous SNV | 722C>T | A241V | 000003F_009 | 189403 | -3.525 | AT-hook motif nuclear-localized protein 22 |
| | nonsynonymous SNV | 2448T>A | D816E | 000083F | 1029557 | -3.009 | Pentatricopeptide repeat-containing protein At3g18110, chloroplastic |
| | nonsynonymous SNV | 617G>A | R206Q | 000471F | 70949 | -0.878 | Retrovirus-related Pol polyprotein from transposon RE1 |
| ***Mendoza*** | synonymous SNV | 1044A>G | R348R | 000001F_011 | 2670208 | - | 2-methylbutanal oxime monooxygenase |
| | nonsynonymous SNV | 1099C>T | R367C | 000005F | 2813489 | 0.135 | CBS domain-containing protein CBSCBSPB3 |
| | nonsynonymous SNV | 451C>G | L151V | 000007F | 2513069 | 0 | Uncharacterised mitochondrial protein AtMg00820 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | nonsynonymous SNV | 1187C>T | A396V | 000020F | 1222181 | 0.317 | Protein ASPARTIC PROTEASE IN GUARD CELL 1 |
| | synonymous SNV | 102C>T | T34T | 000021F | 561968 | - | Dof zinc finger protein PBF |
| | frameshift deletion | 2767delA | K923fs | 000031F | 1031111 | - | 1-phosphatidylinositol 3-phosphate 5-kinase FAB1 |
| | nonsynonymous SNV | 970G>A | V324I | 000033F | 62273 | -0.867 | FAD synthase |
| | nonsynonymous SNV | 1694G>A | S565N | 000056F | 1631108 | -1.72 | Protein BRASSINOSTEROID INSENSITIVE 1 |
| | nonsynonymous SNV | 426T>A | F142L | 000057F | 292346 | -5.68 | ETHYLENE INSENSITIVE 3-like 5 protein |
| | frameshift deletion | 1424delT | L475fs | 000120F | 1049095 | - | Asparagine--tRNA ligase, cytoplasmic 1 |
| | nonsynonymous SNV | 970C>T | R324W | 000169F | 739108 | 0 | F-box protein At3g27290 |
| | nonsynonymous SNV | 1304C>T | S435L | 000195F | 454283 | -5.818 | Uncharacterized GPI-anchored protein At1g61900 |
| | nonsynonymous SNV | 667C>T | R223W | 000233F | 353481 | -7.518 | ABC transporter C family member 8 |
| | frameshift deletion | 224_225del | H75fs | 000261F | 205241 | - | UDP-glucose 6-dehydrogenase 1 |
| | nonsynonymous SNV | 961T>C | F321L | 000266F | 25197 | 0 | SCY1-like protein 2 |
| | nonframeshift insertion | 754_755insGAG | S252delinsRG | 000297F | 31292 | - | Polypyrimidine tract-binding protein homolog 1 |
| | synonymous SNV | 405C>A | L135L | 000321F | 296293 | - | Cell division control protein 2 homolog C |
| | nonsynonymous SNV | 836G>A | R279Q | 000328F | 430851 | 1.878 | Protein cereblon |
| | nonsynonymous SNV | 1097A>G | Y366C | 000433F | 293572 | -1.433 | |
| | nonsynonymous SNV | 1913A>G | H638R | 000681F | 14529 | -7.911 | Probable receptor-like protein kinase At5g61350 |
| ***352*** | nonsynonymous SNV | 1163C>T | T388I | 000020F | 784243 | -2.827 | Xyloglucan galactosyltransferase MUR3 |
| | synonymous SNV | 1233G>A | Q411Q | 000066F | 61306 | - | Transposon TX1 uncharacterized 149 kDa protein |
| | nonsynonymous SNV | 503C>T | P168L | 000075F | 234969 | -4.588 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | nonsynonymous SNV | 841G>C | A281P | 000081F | 299589 | 0.201 | WRKY transcription factor WRKY51 |
| | nonsynonymous SNV | 1124G>A | S375N | 000250F | 625996 | -0.35 | |
| | frameshift deletion | 1357delG | G453fs | 000271F_901 | 447191 | - | Cytochrome P450 CYP73A100 |
| | nonsynonymous SNV | 1358G>A | G453E | 000271F_901 | 447192 | -7.696 | Cytochrome P450 CYP73A100 |
| | nonsynonymous SNV | 35G>A | R12K | 000332F | 20443 | -0.144 | Probable leucine-rich repeat receptor-like serine/threonine-protein kinase At3g14840 |
| | synonymous SNV | 1158T>A | S386S | 000372F | 131896 | - | Transcription factor bHLH62 |
| | stopgain | 430C>T | Q144X | 000448F | 87079 | - | |
| | stopgain | 73G>T | G25X | 000601F | 138931 | - | |
| ***G9V7*** | nonsynonymous SNV | 530G>A | G177E | 000000F | 1062588 | 0 | Protease Do-like 10, mitochondrial |
| | nonsynonymous SNV | 235G>A | V79M | 000013F | 1775702 | -0.003 | D-tagatose-1,6-bisphosphate aldolase subunit GatY |
| | synonymous SNV | 516A>T | V172V | 000014F | 2557160 | - | Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B |
| | nonsynonymous SNV | 662T>G | L221R | 000017F | 960034 | -4.367 | 26S proteasome non-ATPase regulatory subunit 4 homolog |
| | nonsynonymous SNV | 2186T>A | L729H | 000018F | 2368434 | -4.898 | Probable alpha-galactosidase D |
| | nonsynonymous SNV | 773C>T | A258V | 000025F_901 | 348353 | -0.802 | KRR1 small subunit processome component homolog |
| | nonsynonymous SNV | 361G>T | D121Y | 000025F_901 | 423041 | -8.961 | Cyclin-A3-2 |
| | frameshift insertion | 396dupC | G132fs | 000037F | 166010 | - | Auxin-induced protein 22D |
| | synonymous SNV | 693C>T | F231F | 000102F | 633604 | - | Phosphatidate phosphatase PAH1 |
| | nonsynonymous SNV | 1897C>T | R633W | 000114F | 102534 | -7.844 | RNA polymerase I termination factor |
| | frameshift deletion | 178_179del | T60fs | 000118F | 496016 | - | Basic leucine zipper 24 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | nonframeshift insertion | 119_120insGAT | D40delinsEI | 000165F | 401445 | - | Retrovirus-related Pol polyprotein from transposon RE2 |
| | nonsynonymous SNV | 567G>C | M189I | 000205F | 632237 | -1.632 | Heavy metal-associated isoprenylated plant protein 3 |
| | nonsynonymous SNV | 416C>A | P139Q | 000256F | 97022 | -7.622 | E3 ubiquitin-protein ligase ATL41 |
| | nonsynonymous SNV | 122G>T | R41L | 000279F | 151434 | -1.754 | Chlorophyll a-b binding protein 91R, chloroplastic |
| | synonymous SNV | 225C>T | L75L | 000369F | 113383 | - | |
| ***95*** | nonsynonymous SNV | 1326G>A | M442I | 000004F | 208436 | 0.242 | Cellulose synthase-like protein B6 |
| | synonymous SNV | 3192T>C | S1064S | 000004F | 2879156 | - | Putative ribonuclease H protein At1g65750 |
| | nonsynonymous SNV | 32T>A | L11H | 000011F | 1884869 | -0.568 | |
| | synonymous SNV | 174G>A | T58T | 000061F | 484084 | - | Laccase-14 |
| | nonsynonymous SNV | 1274T>C | I425T | 000075F | 442492 | -4.443 | Mini-chromosome maintenance complex-binding protein |
| | nonsynonymous SNV | 1153G>A | V385I | 000099F | 572947 | -0.512 | Cytochrome P450 78A6 |
| | nonsynonymous SNV | 943C>T | R315C | 000099F | 1109334 | -0.292 | |
| | stopgain | 1963C>T | R655X | 000102F | 975331 | - | |
| | synonymous SNV | 348T>A | L116L | 000131F | 164100 | - | Protein Thf1 |
| | nonsynonymous SNV | 158C>T | S53F | 000131F | 650218 | -0.039 | Putative calcium-transporting ATPase 11, plasma membrane-type |
| | stopgain | 267G>A | W89X | 000148F | 255516 | - | |
| | synonymous SNV | 762C>T | Y254Y | 000335F | 213738 | - | Beta-glucosidase BoGH3B |
| | nonsynonymous SNV | 832G>A | A278T | 000342F | 106182 | -1.382 | Truncated transposon Ty1-A Gag-Pol polyprotein |
| | frameshift deletion | 279delA | S93fs | 000342F | 138479 | - | |
| | synonymous SNV | 753C>T | S251S | 000650F | 166557 | - | Putative B3 domain-containing protein At2g27410 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | nonsynonymous SNV | 539G>A | R180Q | 000840F | 33936 | 0.465 | Mitogen-activated protein kinase kinase kinase 9 |
| ***277*** | nonsynonymous SNV | 902G>A | R301H | 000000F | 4916333 | -4.948 | Protein MAM3 |
| | stopgain | 282T>G | Y94X | 000006F | 3175203 | - | |
| | synonymous SNV | 165G>A | L55L | 000007F | 2311151 | - | RNA-directed DNA polymerase homolog |
| | stopgain | 340C>T | Q114X | 000010F | 1213916 | - | |
| | nonsynonymous SNV | 386G>T | R129L | 000011F | 78945 | -3.141 | Acyltransferase-like protein At1g54570, chloroplastic |
| | nonsynonymous SNV | 1022C>T | P341L | 000013F | 1639155 | -0.033 | Microtubule-associated serine/threonine-protein kinase 4 |
| | nonsynonymous SNV | 188C>T | P63L | 000014F | 523106 | -7.14 | Receptor protein-tyrosine kinase CEPR1 |
| | nonsynonymous SNV | 1043C>T | S348L | 000024F | 683670 | -0.605 | Putative E3 ubiquitin-protein ligase RF298 |
| | nonsynonymous SNV | 614C>T | P205L | 000042F | 977870 | -9.078 | Acidic endochitinase |
| | nonsynonymous SNV | 97C>A | L33I | 000049F | 1224051 | -0.372 | |
| | synonymous SNV | 465G>A | P155P | 000057F | 1469242 | - | |
| | nonsynonymous SNV | 114A>T | R38S | 000070F | 1519423 | -2.535 | Cellulose synthase A catalytic subunit 3 [UDP-forming] |
| | nonsynonymous SNV | 2305T>C | F769L | 000118F | 436902 | -5.699 | E3 ubiquitin-protein ligase TRIP12 |
| | nonsynonymous SNV | 1397A>G | Y466C | 000143F | 242181 | -0.35 | Aspartic proteinase nepenthesin-2 |
| | nonsynonymous SNV | 1159A>G | R387G | 000153F | 876372 | -6.942 | Pentatricopeptide repeat-containing protein At2g01510, mitochondrial |
| | nonsynonymous SNV | 485G>A | G162E | 000412F | 320154 | -1.833 | |
| | synonymous SNV | 945C>T | G315G | 000440F | 60685 | - | |
| | nonsynonymous SNV | 16G>A | E6K | 000702F_002 | 176763 | -0.397 | Transposon Ty1-DR1 Gag-Pol polyprotein |
| | | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ***76*** | synonymous SNV | 318T>A | G106G | 000006F | 1970181 | - | DUF21 domain-containing protein At5g52790 |
| | synonymous SNV | 672A>G | E224E | 000007F | 3124188 | - | |
| | nonsynonymous SNV | 2029T>C | F677L | 000020F | 476326 | -5.316 | Primary amine oxidase 2 |
| | nonsynonymous SNV | 499G>A | V167M | 000022F | 2096116 | 0.03 | DnaJ homolog subfamily A member 4 |
| | synonymous SNV | 1173C>T | S391S | 000283F | 511688 | - | Polygalacturonase ADPG1 |
| | nonsynonymous SNV | 3383G>A | R1128Q | 000316F | 247779 | 0.018 | Disease resistance protein TAO1 |
| | nonsynonymous SNV | 432C>G | F144L | 000337F | 397243 | 0.209 | Alanine--tRNA ligase |
| | synonymous SNV | 534T>C | P178P | 000373F | 20837 | - | Transposon Tf2-8 polyprotein |
| ***96*** | nonsynonymous SNV | 619G>A | A207T | 000029F | 803809 | -3.848 | Alkaline/neutral invertase A, mitochondrial |
| | nonsynonymous SNV | 873G>A | M291I | 000044F | 454296 | -1.16 | Protein NDH-DEPENDENT CYCLIC ELECTRON FLOW 5 |
| ***1066*** | nonsynonymous SNV | 1118C>T | P373L | 000000F | 5864702 | -9.739 | Cytochrome P450 81F1 |
| | nonsynonymous SNV | 1530G>T | E510D | 000008F | 1267312 | -2.761 | |
| | nonsynonymous SNV | 190A>T | S64C | 000008F | 2911091 | -5 | EG45-like domain containing protein |
| | stopgain | 511C>T | Q171X | 000046F | 750955 | - | |
| | nonsynonymous SNV | 407C>T | P136L | 000048F | 981856 | 3.908 | Polyphenol oxidase F, chloroplastic |
| | stopgain | 34C>T | R12X | 000066F | 1279485 | - | |
| | nonsynonymous SNV | 2525C>A | S842Y | 000084F | 759222 | -0.389 | DUF724 domain-containing protein 6 |
| | nonsynonymous SNV | 2596G>C | A866P | 000146F | 776213 | -0.52 | Kinesin-like protein KIN-7L, chloroplastic |
| | synonymous SNV | 930T>C | A310A | 000151F | 87486 | - | |
| | frameshift deletion | 959delG | R320fs | 000154F | 867001 | - | |
| | nonsynonymous SNV | 2024T>C | L675P | 000160F | 260906 | -6.767 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | nonsynonymous SNV | 733G>A | D245N | 000184F | 146990 | -1.533 | Retrovirus-related Pol polyprotein from transposon RE1 |
| | frameshift deletion | 61delG | G21fs | 000227F | 555707 | - | Rab escort protein 1 |
| | nonsynonymous SNV | 2650G>A | V884M | 000325F | 57379 | -0.2 | Retrovirus-related Pol polyprotein from transposon RE2 |
| | nonsynonymous SNV | 154G>A | A52T | 000360F | 365282 | -3.85 | |
| | stopgain | 865C>T | Q289X | 000379F | 175644 | - | |
| ***Waite Star*** | synonymous SNV | 2232G>A | L744L | 000061F | 374981 | - | Transposon Tf2-4 polyprotein |
| | frameshift deletion | 1375delA | T459fs | 000069F | 338355 | - | Dual specificity protein phosphatase 1-A |
| | synonymous SNV | 450C>T | P150P | 000325F | 347467 | - | Protein DETOXIFICATION 16 |
| | nonsynonymous SNV | 184T>C | F62L | 000954F | 107534 | -5.864 | Serine carboxypeptidase-like 27 |
| ***118*** | nonsynonymous SNV | 386G>A | R129H | 000027F | 905715 | 1.123 | Scarecrow-like protein 33 |
| | nonsynonymous SNV | 439G>A | V147I | 000087F_902 | 740336 | -0.027 | LIM domain-containing protein PLIM2a |
| ***809*** | nonsynonymous SNV | 814T>C | S272P | 000004F | 3194688 | -3.37 | 1-deoxy-D-xylulose-5-phosphate synthase, chloroplastic |
| | nonsynonymous SNV | 409C>A | L137I | 000019F | 2052594 | -1.431 | Uncharacterized protein At5g43822 |
| | nonsynonymous SNV | 709G>A | G237S | 000020F | 2217262 | -5.534 | Putative glycerol-3-phosphate transporter 5 |
| | nonsynonymous SNV | 313G>A | D105N | 000030F | 136083 | -1.132 | |
| | synonymous SNV | 315G>A | G105G | 000037F | 1971138 | - | |
| | stopgain | 501delT | Y167X | 000081F | 474919 | - | |
| | synonymous SNV | 708C>T | L236L | 000097F | 563782 | - | |
| | nonsynonymous SNV | 8T>A | L3H | 000140F | 171713 | 0.671 | Putative disease resistance RPP13-like protein 1 |
| | nonsynonymous SNV | 1709C>T | A570V | 000230F | 388095 | -3.649 | Xanthine dehydrogenase/oxidase |
| | nonsynonymous SNV | 589T>C | W197R | 000310F | 411561 | -13.858 | P-loop NTPase domain-containing protein LPA1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | frameshift deletion | 931_943del | R311fs | 000321F | 444838 | - | Peregrin |
| | nonsynonymous SNV | 274G>A | E92K | 000339F | 467880 | -0.033 | |
| | nonsynonymous SNV | 3797A>C | K1266T | 000391F | 107563 | -1.153 | LINE-1 reverse transcriptase homolog |
| ***548*** | stopgain | 399G>A | W133X | 000001F_011 | 2409320 | - | |
| | nonsynonymous SNV | 1016C>T | P339L | 000012F | 3164663 | -8.894 | Lipase-like PAD4 |
| | nonsynonymous SNV | 3266C>T | P1089L | 000042F | 470206 | -9.199 | Probable global transcription activator SNF2L2 |
| | nonsynonymous SNV | 1913C>T | S638F | 000062F | 376977 | -3.745 | Probable disease resistance protein At1g61180 |
| | nonsynonymous SNV | 1905T>G | C635W | 000065F | 286229 | -8.591 | Putative K(+)-stimulated pyrophosphate-energized sodium pump |
| | frameshift deletion | 798delG | E266fs | 000074F | 630734 | - | |
| | nonsynonymous SNV | 593G>A | R198H | 000081F | 800974 | -4.832 | Proline--tRNA ligase, chloroplastic/mitochondrial |
| | nonsynonymous SNV | 1131C>A | N377K | 000088F | 457928 | -2.218 | Gibberellin 20 oxidase 2 |
| | synonymous SNV | 510C>T | V170V | 000195F | 104054 | - | 3beta-hydroxysteroid-dehydrogenase/decarboxylase isoform 1 |
| | nonsynonymous SNV | 2534C>T | A845V | 000255F | 198105 | -2.84 | Coatomer subunit beta-2 |
| | synonymous SNV | 552G>A | P184P | 000453F | 70437 | - | Protein NRT1/ PTR FAMILY 5.7 |
| | frameshift deletion | 2522_2523del | L841fs | 000486F | 69152 | - | DNA mismatch repair protein msh3 |
| ***548, 76*** | nonsynonymous SNV | 799G>A | A267T | 000046F | 127922 | -0.367 | |
| | nonsynonymous SNV | 1668A>T | L556F | 000060F | 404031 | -0.636 | Chaperonin 60 subunit beta 2, chloroplastic |
| | nonsynonymous SNV | 701A>T | K234I | 000181F | 786145 | -2.089 | Pentatricopeptide repeat-containing protein At1g18900 |