Government of South Australia
Primary Industries and Regions SA

Adelaide Hills
Vine Improvement Inc

Wine
Australia
for
Australian
Wine

The Australian Wine
Research Institute

Clonal mapping of Pinot Noir

FINAL REPORT

Principal Investigator: Anthony Borneman

Research Organisation: The Australian Wine Research Institute

Date: February 2022

Disclaimer/copyright statement:

This document has been prepared by The Australian Wine Research Institute ("the AWRI") as part of fulfilment of obligations towards the Wine Australia Project Agreement AWR 2001 SAWIDS Clonal mapping of Pinot Noir (the Project Agreement) and is intended to be used solely for that purpose and unless expressly provided otherwise does not constitute professional, expert or other advice.

The information contained within this document ("Information") is based upon sources, experimentation and methodology which at the time of preparing this document the AWRI believed to be reasonably reliable and the AWRI takes no responsibility for ensuring the accuracy of the Information subsequent to this date. No representation, warranty or undertaking is given or made by the AWRI as to the accuracy or reliability of any opinions, conclusions, recommendations or other information contained herein except as expressly provided within this document. No person should act or fail to act on the basis of the Information alone without prior assessment and verification of the accuracy of the Information.

To the extent permitted by law and except as expressly provided to the contrary in this document all warranties whether express, implied, statutory or otherwise, relating in any way to the Information are expressly excluded and the AWRI, its officer, employees and contractors shall not be liable (whether in contract, tort, under any statute or otherwise) for loss or damage of any kind (including direct, indirect and consequential loss and damage of business revenue, loss or profits, failure to realise expected profits or savings or other commercial or economic loss of any kind), however arising out of or in any way related to the Information, or the act, failure, omission or delay in the completion or delivery of the Information.

The AWRI acknowledges and agrees that the Information was commissioned by Adelaide Hills Vine Improvement Inc. with support from the AWRI, Wine Australia and Primary Industries and Regions South Australia (PIRSA) under the terms of the Project Agreement and agrees to the provision of copyright permissions as required by this Agreement.

The Information must not be used in a misleading, deceptive, defamatory or inaccurate manner or in any way that may otherwise be prejudicial to the AWRI.

Abstract

Pinot Noir is one of the highest value grape varieties in Australia (up to $5,000/t), with sales of grapes contributing $5.7 million per annum to the South Australian wine industry and representing 60% of the value directly generated from Adelaide Hills red grapes. Of the 672 ha of Pinot Noir grown in the region (42% of South Australia's Pinot Noir), more than 35% was burnt in the December 2019 Cudlee Creek fire. The fire resulted in the loss of more than 2,000 tonnes of grape production and significant losses in revenue for around 60 Adelaide Hills grape and fine wine producers.

Prior to the fires, the Adelaide Hills region had the majority of Pinot Noir source blocks in South Australia. The bushfire destroyed 6 out of 28 source blocks, significantly reducing the supply of high-demand clonal planting material to industry. While potential vineyards for new source blocks were identified, there was a lack of evidence on the provenance of the clonal material, in particular for the sought-after Pinot Noir clone D4V2. Definitive clonal identification of alternative sources of planting material was required.

To address these challenges, a broad range of Pinot Noir germplasm was sequenced, allowing generation of a foundational sequence dataset of Pinot Noir clones used for grape and wine production in the Adelaide Hills, including key clones that were lost due to bushfire damage. The project is a world-first application of state-of-the-art whole genome sequencing for the assessment of genetic diversity in Pinot Noir grapevines.

The results enabled identification of clonal genetic markers for individual Pinot Noir clones and provided the means to identify potential sources of high-confidence germplasm for the re-establishment of source blocks within the Adelaide Hills. This will in turn enhance confidence in replanting investments, facilitate vineyard recovery and improve income for wine producers in the region.

Contracted deliverables

All deliverables have been achieved.

Year 1	Output	Activities	Target date
A	Sample set of up to 96 green vine tissue samples supplied by Adelaide Hills Vine Improvement for DNA extraction and sequencing	1. Collect leaf samples from the Pinot Noir source blocks at Adelaide Hills Vine Improvement thought to be clone D4V2, the original import of PN D4V2 in the closed CSIRO collection. and 11 other major clones of Pinot Noir in the Monash and other germplasm collections and source blocks being used for cutting production for the propagation sector. Pinot Noir clones to include multiple samples for D4V2, D2V6, D5V12, G8V3, G8V7, MV6, 114, 115, 386, 521, 667, 777, and potentially others. 2. AHVII to conduct ampelographic inspection to verify Pinot Noir varietal status and carry out virus testing according to the recently reviewed Part 1 of the Vine Industry Nursery Accreditation Scheme on vine health status	October 2020
B	Draft phylogenetic tree of candidate D4V2 Pinot Noir clones from Adelaide Hills Source Blocks and vines from other sources	1. Extract DNA from vine leaf samples provided by Adelaide Hills Vine Improvement, and prepare for sequencing 2. Carry out whole genome sequencing of up to 96 Pinot Noir DNA samples using Australian service provider 3. Map clonal DNA reads against published Pinot Noir reference genome 4. Carry out bioinformatics analysis and prepare phylogenetic tree 5. Assess genetic relationships and identify outliers 6. Adelaide Hills Vine Improvement committee and AWRI researchers meet to assess genetic relationships and clonal identity of potential Pinot Noir D4V2 clones in source blocks as well as other Pinot Noir clones.	April 2021

Year 2	Output	Activities	Target date
A	Additional set of green vine tissue samples supplied by Adelaide Hills Vine Improvement as reference material for second round of DNA typing	1. If necessary, source DNA samples from overseas to assist with validation of clone identification. 2. Collect leaf samples from equivalent Pinot Noir clones from Monash and other Australian germplasm sources. 3. Verify and document identity and vine health status.	September 2020
B	Verified phylogenetic tree of Pinot Noir clones and confirmed sources of clean D4V2 planting material	1. Extract DNA from additional samples provided by Adelaide Hills Vine Improvement, and prepare for sequencing. 2. Carry out whole genome sequencing of additional Pinot Noir DNA samples using Australian service provider 3. Map clonal DNA reads against published Pinot Noir reference genome 4. Finalise bioinformatic analysis and complete revised phylogenetic tree 5. Deposit Pinot Noir clonal genomic DNA sequences into public sequence database (e.g. GenBank) 6. Identify germplasm which may be suitable for replacing D4V2 Pinot Noir source blocks burnt in the Cudlee Creek bushfire of December 2019	December 2021
C	Final report	1. Develop communication plan for recommendations to industry stakeholders, Australian vine improvement groups and growers. 2. Identify needs and opportunities for commercialisation of a DNA testing service for Pinot Noir and other varieties. 3. Prepare final report with input from Adelaide Hills Vine Improvement and submit to Wine Australia.	February 2022

Methods

Sample collection

Leaf samples (~5 g) were obtained for selected Pinot Noir clones as indicated in Appendix 1, snap frozen in liquid nitrogen and stored at -80°C. For DNA isolation, approximately 1 g of leaf tissue was ground using a mortar and pestle under liquid nitrogen, with DNA isolated from a 100 mg aliquot of ground tissue using a DNAeasy Plant mini kit (Qiagen). Remaining aliquots were stored at -80°C.

DNA sequencing

DNA sequencing was performed at the Ramaciotti Centre for Genomics. Sequencing libraries were produced using the Illumina DNA kit and sequenced on a Novaseq6000 using 2x150bp chemistry.

Marker identification and phylogenomics

Raw short reads were quality controlled using fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and ngsReports (Ward et al. 2020). Short read data was then used as input into a custom variant calling pipeline, which uses k-mer information to select true variants, for variant calling (Roach et al. 2018). In brief, reads were mapped to the Pinot Noir reference genome (Velasco et al. 2007) using BWA mem (Li, 2013). Mapped reads that aligned within repetitive regions were then masked using SAMtools (Li et al. 2009), informed by a repeat database constructed using repeatmodeler2 (Flynn et al. 2020). Variants were then called using SAMtools and rescored to identify a base set of marker variants. A k-mer database, constructed using Jellyfish (Marçais and Kingsford 2011), was then deployed to purge false positive variants from the base set. K-mers were further used to confirm true positive genotypes through k-mer overlap across both the variant and reference allele.

Clonal marker variants were then used to reconstruct phylogenetic trees using two methods:
i) a maximum likelihood phylogeny that was constructed using a concatenated alignment of all variants with RAxML (Kozlov et al. 2019)
ii) a cladogram that was constructed using SVDquartets (Chifman and Kubatko 2014) to account for evolutionary rate variance between unlinked loci.

Node support for each tree was determined from 1,000 bootstrap replicates and topologies were then manually compared to identify sample clade assignment.

Results

To obtain a comprehensive set of samples that reflected the Pinot Noir clones of interest to Adelaide Hills Vine Improvement (AHVII), leaf samples were obtained for a degenerate set of 197 grapevines, which represented 30 different Pinot Noir clones from across 21 different sites. Within this set were six independent representatives of D4V2 and multiple representatives of most other clones.

From this initial set, a subset of 96 vines was selected that captured the major clonal groups (Table 1). Genomic DNA was extracted from these 96 leaf samples, which was subjected to whole genome sequencing using the Illumina short-read platform at the Ramaciotti Centre for Genomics.

Following sequencing, raw data from all 96 samples was then processed using the AWRI clonal-marker identification pipeline, with mapping performed against the existing Pinot Noir PN0024 reference genome. All samples were shown to have genetic backgrounds consistent with the targeted cultivar, Pinot Noir.

A total of 3,208 marker single nucleotide polymorphisms (SNPs) were characterised across the 96 samples, which were then used to build a genetic phylogeny of the Pinot Noir clones. Both coalescent and maximum-likelihood phylogenies were produced, with both analysis methods producing congruent results (Figure 1).

Genetic relationships were shown to be robust across multiple samples of the same clone. DNA typing data also supported known historic ancestry and/or geographic relationships, including:

- G8V7 (FPS 15) and G8V3 (FPS 13) both have origins in the Martini vineyard (1966) (https://fps.ucdavis.edu/fgrdetails.cfm?varietyid=1184)
- H7V15 (FPS 22) and D5V12 (FPS 19) (collected from same vine 1974) (https://fps.ucdavis.edu/fgrdetails.cfm?varietyid=1184)
- UCD05 (FPS 05) and D4V2 (FPS 04) derived from same stock via heat treatment (1965) (https://fps.ucdavis.edu/grapebook/winebook.cfm?chap=PinotNoir)
- Mariafeld, D2V5 (FPS 01A) and D2V6 (FPS 02A) having origins in Wädenswil, Switzerland (https://fps.ucdavis.edu/grapebook/winebook.cfm?chap=PinotNoir)
- MV6 and Mount Pleasant sourced from Mount Pleasant, Hunter Valley, New South Wales (Nichols)

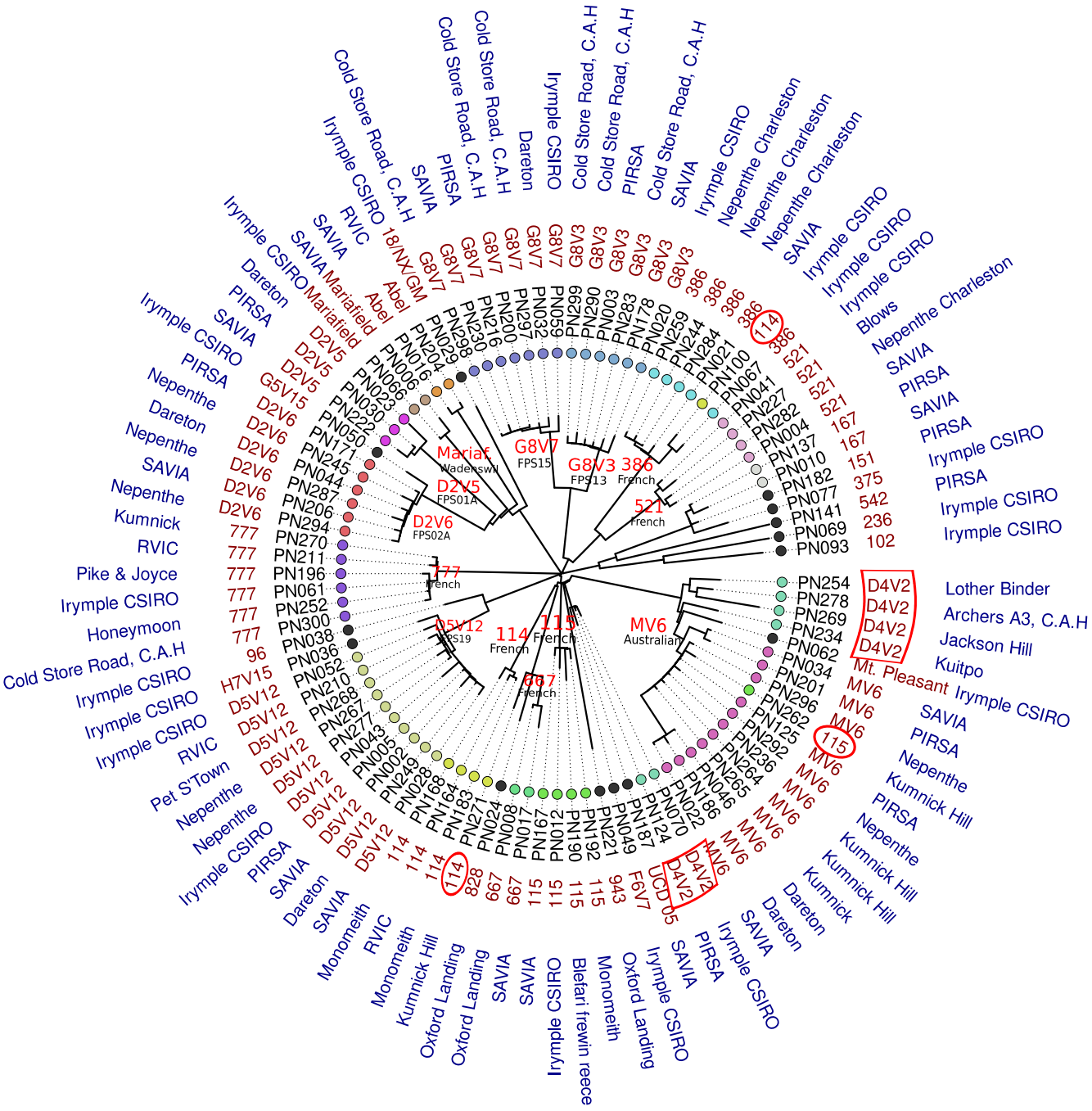


Figure 1. Pinot Noir clonal phylogeny, based on whole genome sequencing of 96 vines. Nodes are shaded by their original clonal designation, with clonal 'clades' labelled with red text. Notes on clonal provenance are also included. Clonal samples displaying incongruent results are highlighted by red boxes/circles.

Seven samples among clones 114, 115 and D4V2 (marked in red) were determined to be outliers compared to the consensus genetic profile associated with the expected clonal ID. This included all samples of D4V2 sourced from the Adelaide Hills, which displayed a genetic profile distinct from samples of D4V2 from CSIRO and PIRSA and were instead positioned in a broad clade with samples originally sourced from the Hunter Valley (HV clade), including MV6 and Mount Pleasant (Figure 1).

As the observed outlier results may reflect sampling and/or processing errors, the seven first-round samples were re-sequenced, in addition to fourteen independent samples (different vines) of the outlying clones from across the affected locations.

Data from all 21 second-round samples was processed as for the first-round samples, with a revised phylogeny produced (Figure 2). The results of the confirmation samples supported the first round of sequencing with respect to D4V2 source blocks from within the Adelaide Hills. Despite the addition of the confirmation samples, all D4V2 germplasm sourced from the Adelaide Hills was shown to reside within the 'HV' clade. This clade is separate from that which contains the examples of D4V2 sourced from the PIRSA and CSIRO germplasm collections.

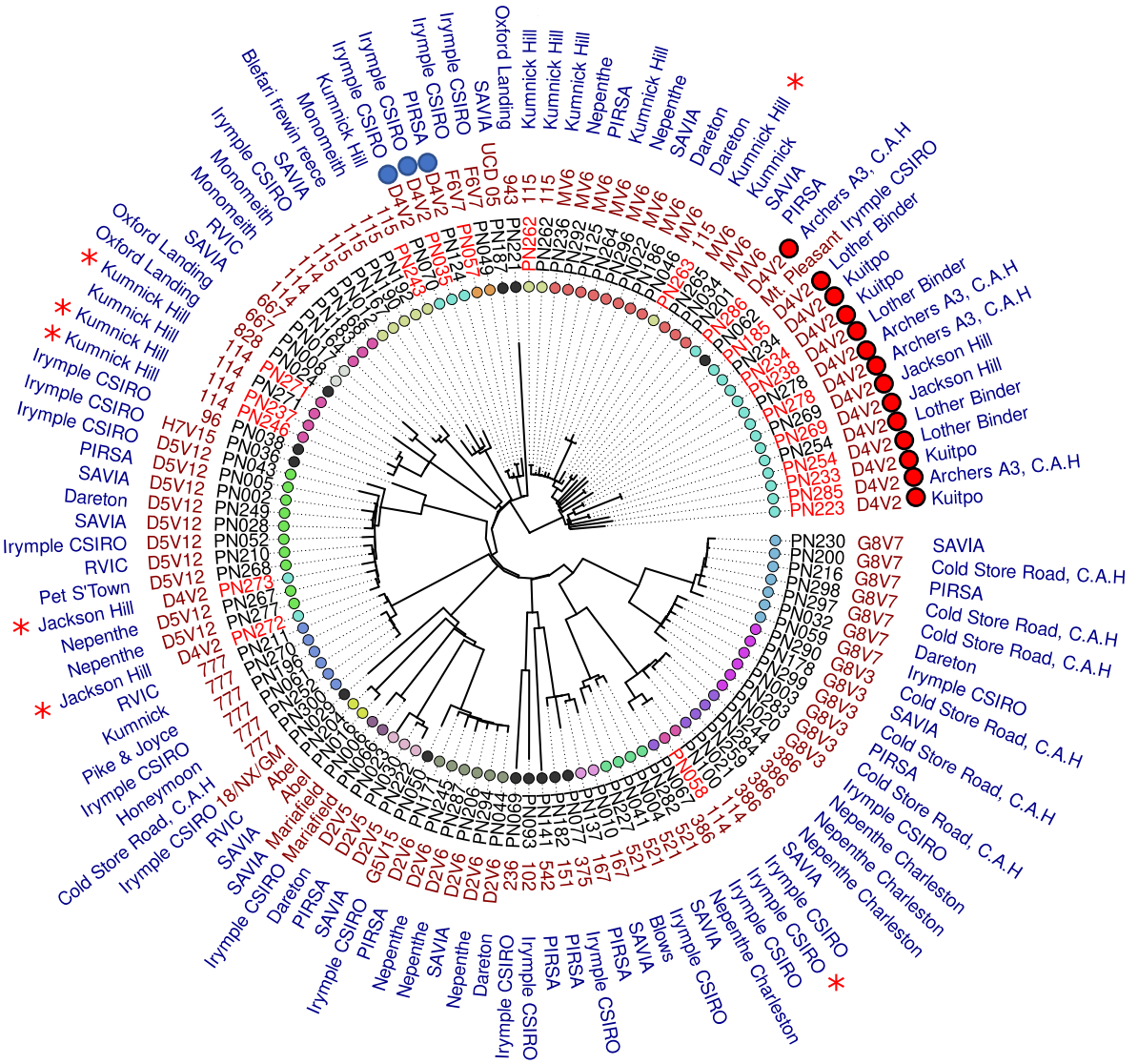


Figure 2. Compiled clonal phylogenetic results for Pinot Noir sequencing rounds 1 and 2. Round 2 samples are indicated in red text (with asterisks denoting independent sources of material relative to Round 1). Node tips are shaded according to existing clone annotation. The Adelaide Hills and CSIRO/PIRSA clades of D4V2 are indicated by the red and blue dots, respectively.

Furthermore, this material displayed levels of inter-sample genetic variation which is consistent with highly related, but not clonal, origins. The genetic distance between the Adelaide Hills D4V2 samples would be consistent with a 'mass-selected' cohort of related cuttings with origins in the Hunter Valley and genetically similar to MV6 and Mount Pleasant.

D4V2 germplasm from CSIRO and PIRSA therefore represent the likely 'correct' source of this clone in Australia. Supporting this observation is the tight relationship between the CSIRO and PIRSA clones of D4V2 and the SAVIA clone UCD05 (FPS05). UCD05 represents a heat-treated sibling of D4V2 (FPS04), which is fully supported by the clonal phylogenetic results. Accordingly, UCD05 may represent a secondary option as a source of clonal material that is highly similar to D4V2 at the genetic level.

Conclusions

Clonal-marker identification allowed for high-resolution genetic typing of a broad subset of Pinot Noir clones from around Australia. It was possible to demonstrate robust genetic relationships between clones from disparate sources and those that shared known historic associations or shared ancestry.

In most cases, samples grouped as expected and with other grapevines of the same original clonal designation. There were seven samples in which a discrepancy existed between the supplied clonal name and the genetic clade in which the sample was placed. These included two samples of clone 114, a single sample of clone 115 and all samples of clone D4V2 sourced from AHVII (Kuitpo, Lother Binder, Jackson Hill, Archers A3).

In the case of D4V2, the majority of the samples of D4V2 from AHVII do not cluster with samples of D4V2 from the CSIRO or PIRSA collections. Instead, these clones show their highest genetic similarity with samples of clone MV6 (11 in total from 5 different organisations) and Mount Pleasant (CSIRO, Irymple). However, they do not represent examples of either of these clones. The level of genetic diversity observed between the AHVII D4V2 samples is larger than would be expected for a single clonal group, such that they potentially represent a mass-selected cohort of vines with origins in the Hunter Valley.

MV6 is considered a 'premium' clone of Pinot Noir. The desirable characteristics of the AHVII 'D4V2' material may therefore reflect this shared genetic ancestry. Individual samples from this genetically mixed group of material have the potential to be specifically selected and propagated as new clonal material under different clonal designations.

In the light of these results, it is recommended that the two potential Pinot Noir source blocks not be used to provide cuttings for propagation of the D4V2 clone. In addition, plantings of the clone elsewhere in the Adelaide Hills should not be used to re-establish AHVII source blocks of D4V2 lost to the Cudlee Creek bushfires.

The sources of the 'D4V2' variant should be traced back and the Adelaide Hills clone(s) renamed. Future work could assess the 'D4V2' variant alongside the true and traceable D4V2 and MV6. Investigation of characters such as growth habit, bunch and berry structure and maturity data of both grapes and wine will provide an indication of its attributes as a desirable clone of Pinot Noir and how it differs from D4V2 and MV6.

References

Chifman, J. Kubatko, L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics.* 30(23): 3317-3324.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C. Smit, A.F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Nat. Acad. Sci.* 117(17): 9451-9457.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. Stamatakis, A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 35(21): 4453-4455.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16): 2078-2079.

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

Marçais, G., Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 27(6): 764-770.

Roach, M.J., Johnson, D.L., Bohlmann, J., van Vuuren, H.J.J., Jones, S.M., Pretorius, I. S., Schmidt, S. A., Borneman, A.R. 2018. Population sequencing reveals clonal diversity and ancestral inbreeding in the grapevine cultivar Chardonnay. *PloS Genetics* 14(11): e1007807.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D., Pindo, M., FitzGerald, L.M., Vezzulli, S., Reid, J., Malacarne, G. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PloS One*. 2(12): p.e1326.

Ward, C.M., To, T.H., Pederson, S.M. 2020. ngsReports: a Bioconductor package for managing FastQC reports and other NGS related log files. *Bioinformatics.* 36(8): 2587-2588.

Acknowledgements

Adelaide Hills Vine Improvement Inc (AHVII)
Adelaide Hills Vine Improvement Source Block owners
Australian Vine Improvement Association (AVIA)
Bioplatforms Australia
Commonwealth Scientific and Industrial Research Organisation (CSIRO)
Department of Primary Industries and Regions (PIRSA)
K1 Wines
S. Smith and Sons
Wine Australia

Table 1. Pinot Noir samples sequenced in this study

ID	Clone	Source
PN002	D5V12	SAVIA
PN003	G8V3	PIRSA
PN004	521	SAVIA
PN005	D5V12	PIRSA
PN006	Mariafeld	SAVIA
PN008	667	Oxford Landing
PN010	167	SAVIA
PN012	115	Irymple (CSIRO)
PN016	Abel	SAVIA
PN017	667	SAVIA
PN020	G8V3	Irymple (CSIRO)
PN021	386	SAVIA
PN022	MV6	SAVIA
PN023	D2V5	Dareton
PN024	828	Oxford Landing
PN028	D5V12	SAVIA
PN029	18/NX/GM	Irymple (CSIRO)
PN030	D2V5	PIRSA
PN032	G8V7	Dareton
PN034	MV6	SAVIA
PN035	D4V2	Irymple (CSIRO)
PN036	H7V15	Irymple (CSIRO)
PN038	96	Irymple (CSIRO)
PN041	521	Irymple (CSIRO)
PN043	D5V12	Irymple (CSIRO)
PN044	D2V6	Dareton
PN046	MV6	Dareton
PN049	F6V7	Irymple (CSIRO)
PN050	G5V15	Irymple (CSIRO)
PN052	D5V12	Irymple (CSIRO)
PN057	F6V7	Irymple (CSIRO)
PN058	114	Irymple (CSIRO)
PN059	G8V7	Irymple (CSIRO)
PN061	777	Irymple (CSIRO)
PN062	Mt. Pleasant	Irymple (CSIRO)
PN066	Mariafeld	Irymple (CSIRO)
PN067	386	Irymple (CSIRO)
PN069	236	Irymple (CSIRO)
PN070	D4V2	Irymple (CSIRO)
PN077	375	Irymple (CSIRO)
PN093	102	Irymple (CSIRO)
PN100	114	Irymple (CSIRO)
PN124	D4V2	PIRSA
PN125	MV6	PIRSA
PN137	167	PIRSA
PN141	542	PIRSA
PN164	114	RVIC
PN167	115	SAVIA
PN168	114	Monomeith
PN171	D2V6	PIRSA
PN178	G8V3	SAVIA
PN182	151	PIRSA
PN183	114	Monomeith
PN185	D4V2	Lother Binder

PN186	MV6	Dareton
PN187	UCD 05	SAVIA
PN190	115	Blefari frewin reece
PN192	115	Monomeith
PN196	777	Pike & Joyce
PN200	G8V7	Cold Store Road, C.A.H
PN201	MV6	PIRSA
PN204	Abel	RVIC
PN206	D2V6	SAVIA
PN210	D5V12	RVIC
PN211	777	RVIC
PN216	G8V7	PIRSA
PN221	943	Oxford Landing
PN222	D2V5	SAVIA
PN223	D4V2	Kuitpo
PN227	521	Blows
PN230	G8V7	SAVIA
PN233	D4V2	Kuitpo
PN234	D4V2	Kuitpo
PN236	MV6	Kumnick Hill
PN237	114	Kumnick Hill
PN238	D4V2	Lother Binder
PN243	115	Kumnick Hill
PN244	386	Nepenthe Charleston
PN245	D2V6	Nepenthe
PN246	114	Kumnick Hill
PN249	D5V12	Dareton
PN252	777	Honeymoon
PN254	D4V2	Lother Binder
PN259	386	Nepenthe Charleston
PN262	115	Kumnick Hill
PN263	115	Kumnick Hill
PN264	MV6	Kumnick Hill
PN265	MV6	Kumnick Hill
PN267	D5V12	Nepenthe
PN268	D5V12	Pet S'Town
PN269	D4V2	Jackson Hill
PN270	777	Kumnick
PN271	114	Kumnick Hill
PN272	D4V2	Jackson Hill
PN273	D4V2	Jackson Hill
PN277	D5V12	Nepenthe
PN278	D4V2	Archers A3, C.A.H
PN282	521	Nepenthe Charleston
PN283	G8V3	Cold Store Road, C.A.H
PN284	386	Nepenthe Charleston
PN285	D4V2	Archers A3, C.A.H
PN286	D4V2	Archers A3, C.A.H
PN287	D2V6	Nepenthe
PN290	G8V3	Cold Store Road, C.A.H
PN292	MV6	Nepenthe
PN294	D2V6	Nepenthe
PN296	MV6	Nepenthe
PN297	G8V7	Cold Store Road, C.A.H
PN298	G8V7	Cold Store Road, C.A.H
PN299	G8V3	Cold Store Road, C.A.H
PN300	777	Cold Store Road, C.A.H