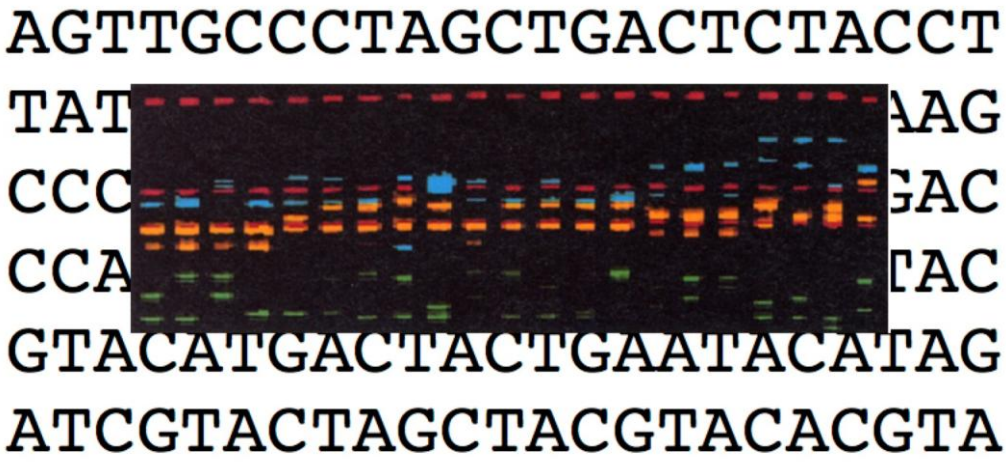
GRAPEVINE TYPE REFERENCE PILOT STUDY



FINAL REPORT to
GRAPE AND WINE RESEARCH & DEVELOPMENT CORPORATION

Project Number: **CSP 1102**

Principal Investigator: **MARK R. THOMAS**

Research Organisation: **CSIRO PLANT INDUSTRY**

Date: **July 2013**

Project title: Grapevine Type Reference Pilot Study

Project Number: CSP 1102

Author Details: Dr Mark R. Thomas

Institution: CSIRO Plant Industry
Postal correspondence: PO Box 350, Glen Osmond, SA 5064
Phone: 08 83038624
Email: mark.r.thomas@csiro.au

Publisher: CSIRO

Copyright statement:
© 2013 CSIRO To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

Disclaimer:
CSIRO advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, CSIRO (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

Abstract 1
Executive Summary 2
Acknowledgments 4
1 Background 5
1.1 Industry Context 6
2 Project Aims 6
3 Methods 6
3.1 DNA samples from international and national collections 7
3.2 DNA profiling of selected varieties 7
3.3 Database development and data analysis 7
4 Results & Discussion 7
4.1 DNA samples provided by four countries 8
4.2 DNA profiling 8
4.3 Variety identification using a 48 SNP set 9
4.4 New SNP markers for improved grapevine identification 10
4.5 Sources of errors in identification 10
4.6 Web accessible SNP database for grapevine identification 11
4.7 DNA 'gold standards' versus plant reference standards 11
5 Outcome & Conclusion 11
6 Recommendations 14
6.1 Future Research Directions 14
6.2 Industry Context and Adoption Scenarios 15
Appendix 1 Communication 16
Appendix 2 Intellectual Property 17
Appendix 3 References 18
Appendix 4 Staff 20

Abstract

This pilot project successfully trialed an international collaborative arrangement to validate the identification of a select group of grapevine varieties (scions and rootstocks) held in Australian germplasm collections. A new marker type, SNPs (Single Nucleotide Polymorphisms), was evaluated and was shown to be a suitable marker type that could be used for variety identification. Accurate variety identification is essential for compliance with the Label Integrity Program (LIP) and best practice would involve a mechanism for traceability from the wine back to the vineyard to the nursery and mother vine or imported vine. A DNA based method would fulfill this goal and this project reports the development of a system based on SNP markers that with further development could be used by researchers and industry for variety identification and traceability.

Executive Summary

The naming of grapevine varieties and identification is complex due to a legacy of many centuries of cultivation in many countries and significant movement of material between regions within countries and between countries over a very long period of time. This has caused a large number of varieties to be misnamed and to have multiple names or synonyms. Until recently the lack of a coordinated multi-national approach to naming and identification has contributed to the legacy. The development twenty years ago by CSIRO of an objective method of grapevine identification based on DNA markers known as microsatellites or SSRs has enabled researchers throughout the world to characterise national collections to identify variety identification errors and synonyms and the genetic diversity within collections. Commercial and semi-commercial services are offered to industry throughout the world with one service in Australia offered by the AGRF.

Unfortunately the major collections in Australia, held by CSIRO and SARDI, that represent up to 80% of unique material held in Australia (McMichael et al., 2013) have not been completely characterised by DNA profiling. An article by Walker and Thomas in 2009 reported on the importance of Australian germplasm collections to both researchers and industry and noted that there was a degree of urgency for the Australian industry to address strategic issues relating to collections in Australia given the impending closure of the CSIRO Merbein site. Subsequently the CSIRO collection was moved to a new Sunraysia site when Merbein was closed. Both the CSIRO and SARDI collections are currently closed to industry.

GWRDC has funded 3 projects that are interrelated and finish in 2013; this one CSP 1102, GWR 1112 "Review of Grapevine Germplasm Collections In Australia" and GWR 1002 "Australian Vine Quality Standards Scheme - Vine Assurance".

The objective of this pilot project was to develop an international collaborative arrangement between Australia, USA, Spain and Germany to evaluate mechanisms and methods for DNA profiling grapevine cultivars and through this validate a list of grapevine varieties (scions and rootstocks) to determine the feasibility of this approach for adoption by the Australian industry to assist with compliance with the Label Integrity Program (LIP), voluntary compliance with the Australian Standard for Grapevine Propagation Material and for DNA profiling the two main grapevine variety collections in Australia held by CSIRO and SARDI.

The objective was achieved and a DNA profiling system based on SNP DNA markers was developed in addition to a database that could be further developed for use by industry and researchers to determine variety identity. The advantages of SNP markers over SSR markers are described in the report.

The varieties DNA profiled in this study represent only a fraction of the varieties held in the two main Australian collections. It is recommended that the Australian industry consider funding a new project to characterise and validate by DNA profiling the identity of unique varieties in the main collections. Suggestions on how this might be achieved are outlined in the Recommendations section. If the Australian industry accepts the above recommendation then an essential component of the project should include a means to provide the industry with at least one Australian commercial provider of a

DNA profiling service for grapevine identification. It is also recommended that an advisory reference group composed of representatives of peak industry bodies and key scientists be associated with the project to ensure it meets the needs of the Australian industry and has the necessary global linkages with organisations (e.g. International Organisation of Vine and Wine; OIV) involved in clarifying and assigning variety names and synonyms.

The research results and outcomes described in this project report would not have been possible without the valuable contributions of the people listed in the Acknowledgement section.

Acknowledgments

The author of this report would like to thank Peter Hayes and GWRDC staff for assistance in developing the project and during the course of the project. CSIRO staff; Rob Walker, Peter Clingeleffer, Ying Zhu, Don Mackenzie and Hilary Davis are thanked for valuable contributions to the project. Mike McCarthy is thanked for access to the SARDI germplasm collection. Sean Myles is thanked for providing unpublished DNA sequence information surrounding grapevine SNPs. Greatly appreciated was the advice and guidance from members of the Industry Reference Group associated with the project.

This research would not have been possible without the valuable contributions of the international collaborators; Reinhard Töpfer and Erika Maul (Julius Kühn-Institut, Germany), Javier Ibáñez and José M. Martinez-Zapater (Instituto de Ciencias de la Vid y del Vino, Spain), Gan-Yuan Zhong, ChihCheng Thomas Chao, Heidi R. Schwaninger and Anne Koehmstedt (USDA-ARS, USA).

1 Background

DNA profiling (typing) has become the method of choice for identifying grapevine cultivars and determining instances of synonyms (different names for the same variety) and homonyms (different varieties identified under the same name). The large body of scientific literature indicates that the extent of these naming errors can be very high especially in Europe that is the origin of most of the varieties imported into Australia. Naming errors have also been identified in Australia. Twenty years ago CSIRO published a DNA method for grapevine identification based on DNA markers known as SSRs or microsatellites (Thomas and Scott, 1993; Thomas et al., 1994). This methodology has been adopted worldwide (over 600 scientific papers have referenced these papers) and is currently used to identify grapevine cultivars and has been used in other countries to characterize variety collections.

Recent advances in DNA technology have seen a new DNA marker type know as SNPs (Single Nucleotide Polymorphisms) emerge as a potentially suitable DNA marker for grapevine identification as it has a number of advantages over microsatellite markers including the ease of transferability and comparison between laboratories (Cabezas et al., 2011).

Research laboratories use DNA profiling for grapevine cultivar identification, genetic relatedness studies, linking traits to DNA markers and genes and for investigating genetic diversity. Throughout the world a small number of commercial and semi- commercial laboratories (often associated with a research laboratory) offer a grapevine DNA profiling service based on SSR markers and proprietary databases. These commercial databases are not cross-referenced. In Australia, currently the AGRF (Australian Genome Research Facility) provides a SSR based grapevine DNA profiling service to industry using a database developed by CSIRO. AGRF has corporate accreditation to ISO 17025.

Two European grapevine databases have been developed based on international collaborative research efforts in an attempt to cross-reference varieties – the European *Vitis* Database (EVD) and the *Vitis* International Variety Catalogue (VIVC) (Lacombe et al., 2011; Maul et al., 2012). Both databases can be accessed by a web interface (http://www.eu-vitis.de/ http://www.vivc.de/) and are managed by the Julius Kühn-Institut (JKI) research organization based in Germany. Both databases were originally based on ampelographic descriptors with recent research investigating the inclusion of SSR data. Personal correspondence from Erika Maul at JKI provided the following information. Development of the VIVC was initiated in 1983 by Prof. Alleweldt when no international agreement existed on registration standards for collected material and the database now includes both ampelographic descriptors, synonyms, genetic information and some SSR marker data for the varieties. The distinguishing feature of the European *Vitis* Database is the inclusion of clonal information and the ability to trace back to the plant that has been described in the database. Both databases are incomplete and are “works in progress” but represent research efforts to cross-reference varieties in collections especially between European countries participating in the research. For the time being the intention is to keep both databases separate. Research organizations in other countries have separate initiatives and databases for national grapevine collections (e.g. National Grape Registry in the USA http://ngr.ucdavis.edu, ENSAM-INRA in France http://bioweb.ensam.inra.fr/collections_vigne/Variete.php#). The quantity, quality and up-to-dateness

of the data in all research databases is dependent on research funds. Recent studies in USA, France and Italy have used DNA profiling results from germplasm collections to investigate genetic diversity, genetic structure, parentage and grapevine domestication events (Cipriani et al., 2010; Myles et al., 2010; Laucou et al., 2011; Bacilieri et al., 2013; Lacombe et al., 2013).

1.1 Industry Context

Australia has two main germplasm (variety) collections, the largest one held by CSIRO and the second largest one held by SARDI. A recent article describes the history of grapevine germplasm collections in Australia and the benefits to both industry and research scientists (Walker and Thomas, 2009). Both collections are currently closed to industry. The 2013 GWRDC Final Report (GWR 1112) "Review of Grapevine Germplasm Collections In Australia" details the current state of both research and industry collections. DNA profiling the major Australian collection(s) would benefit both scientists and industry and may lead towards a pathway for an Australian National Collection.

In Australia, compliance with the Label Integrity Program (LIP) is required for wine labeling and if a variety name is on the label an audit trail must be kept. A DNA based method for variety identification would considerably reduce the risk to members of the Australian industry and Australia's international reputation by minimizing the chance of accidental mislabeling events. Development of a scheme that can trace grapevine genotypes from vine importation to nursery to vineyard to the winery for label integrity would be of benefit to the industry. The GWRDC funded a project (GWR 1002) "Australian Vine Quality Standards Scheme - Vine Assurance" to provide industry with a voluntary Australian Standard for Grapevine Propagation Material. The Australian Standard for Grapevine Propagation Material, AS 5588:2013 was published in September 2013 (www.saiglobal.com/shop). An important component of this scheme will be DNA profiles that accurately identify grapevine cultivars.

2 Project Aims

The objective of this pilot project was to develop an international collaborative arrangement between Australia, USA, Spain and Germany to evaluate mechanisms and methods for DNA profiling grapevine cultivars and through this validate the DNA type reference genotype for specific grapevine varieties and determine the feasibility of this approach for wider adoption within the Australian industry to assist with compliance with the Label Integrity Program, voluntary compliance with the Australian Standard for Grapevine Propagation Material and for DNA profiling the two main grapevine variety collections in Australia, held by CSIRO and SARDI.

3 Methods

A summary is provided here. The details of the experimental procedures and results for this project and will be publically released once publication in the peer reviewed literature is achieved. Interested persons should contact GWRDC for more information.

3.1 DNA samples from international and national collections

In Australia, tissue from selected varieties was collected from both the CSIRO and SARDI germplasm collections. DNA was extracted using standard procedures with commercial kits from both leaf and cambium tissue of wooded canes or by an automated DNA extraction platform at the AGRF.

DNA samples were sent to Australia from the three international collaborators: Julius Kühn-Institut in Germany (Reinhard Töpfer and Erika Maul), Instituto de Ciencias de la Vid y del Vino in Spain (Javier Ibáñez and José M. Martinez-Zapater) and USDA- ARS in the USA (Gan-Yuan Zhong, ChihCheng Thomas Chao, Heidi R. Schwaninger and Anne Koehmstedt).

A list of 80 scion and rootstock varieties was provided to the international collaborators. The list consisted of common scion varieties, emerging or alternative scion varieties and rootstock varieties to assess both the ability of the international collaborators to provide the material and to determine the ability of SNP markers to discriminate a diverse range of genetic material.

3.2 DNA profiling of selected varieties

For SNP DNA profiling the MassARRAY iPLEX assay with matrix assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) detection (Sequenom, http://www.sequenom.com/) was used as it had proved to be a useful technology for previous grapevine DNA profiling research (Chaib et al., 2010). The SSR analysis was essentially as previously described (Thomas et al., 1994). Both the SSR and SNP genotyping was performed by AGRF.

3.3 Database development and data analysis

A FileMaker database was developed for research purposes to hold DNA profiles and associated information for analysis. Comparison of variety DNA profiles extracted from the database was done using custom R scripts and the program clustalw 2.1 (Larkin et al., 2007) was used to generate genetic relatedness trees.

A MySQL database with a PHP implemented web front end was developed to hold selected validated variety SNP DNA profiles for web based input queries. This database was developed to determine the feasibility of providing a web accessible database for both researchers and commercial providers to query SNP DNA profiles against validated variety DNA profiles.

4 Results & Discussion

A summary is provided here. The details of the results for this project and the experimental procedures will be publically released once publication in the peer reviewed literature is achieved. Interested persons should contact GWRDC for more information.

4.1 DNA samples provided by four countries

DNA samples of the 80 varieties were requested from USA, Germany and Spain. It was requested that DNA samples be provided only from varieties in national germplasm collections. The DNA samples would then be DNA profiled to compare the same varieties from all countries including Australia. If the same DNA profile for a variety was obtained from a minimum of 3 country collections a high level of confidence in the correct identity of the variety would exist and the DNA profile could be classified as a "type reference genotype". The type reference genotype would then represent the 'gold standard' against which other plants of the same variety could then be compared against in the future.

After DNA samples were received from the 3 countries it became apparent that not all countries had all the 80 varieties in the national collections or were unable to send DNA samples due to a number of reasons. Variety DNA samples provided were; Germany 61, USA 67, Spain 35. The number of samples per variety ranged from 1 (Australia only) to all 4 countries (Table 1).

Table1. Summary of DNA samples provided per variety.

DNA samples obtained per variety	Variety number
1	2
2	20
3	31
4	27

From Table 1, a maximum of 58 varieties could theoretically meet the criteria of at least 3 matching DNA profiles for a high level of confidence in the correct variety identification. A reasonable level of confidence in correct variety identification would be met for 78 varieties if only two DNA profiles matched. For 2 varieties no DNA comparison was possible except to compare the DNA profile to the other 78 varieties to confirm that the DNA profile was unique and did not match the other variety DNA profiles.

4.2 DNA profiling

The DNA marker type and the technology used for plant and animal identification is subject to change over time as new DNA marker types replace existing ones and cheaper more efficient technologies replace older technologies. The current rapid pace in genomic technology developments will see continued change in both DNA marker types and the technologies available for identification applications.

The two main DNA marker types used for grapevine identification today are SSR and SNP markers. SSR markers have the advantage that they are multi-allelic thereby increasing the discriminating power of a single marker when applied to a diploid crop such as grapevine with some grapevine SSR markers having more than 10 alleles. SNP markers are bi-allelic for diploid organisms and are less informative with only two alleles possible. The other major difference between SSR and SNP markers is that SSR allele differences are determined by DNA length differences while SNP marker differences are based on DNA sequence differences.

Initially there were a wide variety of different technologies used to measure SSR allele length differences but the different technology platforms produced different allele lengths for the same allele making it difficult to compare results between laboratories. The technology described in 1994 for SSR grapevine identification (Thomas et al., 1994) is now the dominant technology used today for grapevine identification. Using this technology, reproducibility of allele calling is very high over time within a laboratory and this is one of the reasons it is used by scientists and commercial service providers. Reproducibility between laboratories remains an issue even when the same technology platform is used and Table 2 shows the allele differences recorded for the cultivar Viognier when SSR DNA profiles across 6 loci are compared between profiles in the AGRF database and the VIVC database.

Table 2. Comparison of SSR alleles at 6 loci for the variety Viognier reported by the *Vitis* International Variety Catalogue (VIVC) in Europe and AGRF in Australia.

SSR Marker:	VVS2	VVMD5	VVMD7	VVMD27	VrZAG62	VrZAG79
Viognier (VIVC)	133 : 139	228 : 234	239 : 249	186 : 192	188 : 200	251 : 251
Viognier (AGRF)	130 : 136	222 : 229	237 : 247	183 : 189	186 : 198	248 : 248

The VIVC has attempted to assist users in comparing SSR results by providing standard Cabernet Sauvignon and Chardonnay DNA profiles but the example in Table 2 demonstrates that accurate comparisons between laboratories remains difficult.

SNPs, in theory, are a marker type that is relatively easy to transfer and compare between laboratories as the marker is scored as a single DNA base (either A, T, G or C). Different technologies have been developed to identify SNPs and the technologies are still evolving. The SNPlex™ technology platform used for the development and evaluation of the 48 SNP set for grapevine identification (Cabezas et al., 2011) is now no longer available. The main issue with SNP markers is DNA primer design to allow multiplexing of markers as the more markers that can be multiplexed in one assay the cheaper the cost per marker. Even when primer design is optimized for multiplexing markers issues can still arise. For example, the 48 SNP markers reported by Cabezas et al. (2011) worked in separate PCR assays but when multiplexed into a single 48 multiplex reaction one marker consistently failed effectively giving DNA profiles for only 47 markers.

The advantages of SNP markers over SSR markers, relatively easy to assay large numbers of markers and score genotypes by computer with little or no human involvement at an attractive cost, suggests that SNP markers will eventually replace SSR markers.

4.3 Variety identification using a 48 SNP set

SNP marker primers were designed for the Sequenom MassARRAY iPLEX assay based on DNA sequence surrounding the 48 SNPs reported by Cabezas et al. (2011). DNA from the four countries was then analyzed to determine if DNA profiles matched or disagreed.

For *Vitis vinifera* scion DNA all 48 markers could be used for analysis, however, some markers were either unreliable or not informative for non-vinifera and hybrid rootstocks.

SNP analysis of rootstocks showed that fewer polymorphisms were identified between rootstocks with the SNP markers used). This indicated that the 48 SNP set was better suited for vinifera variety identification than non-vinifera.

4.4 New SNP markers for improved grapevine identification

SNP primer designs were developed for new SNP markers using SNP information from unpublished CSIRO SNP markers and marker information provided by Sean Myles (Myles et al., 2011). Potential SNP markers were first screened by computer analysis for polymorphism in non-vinifera species and distribution across all 18 chromosomes of the grapevine genome. A selection of 218 SNP markers meeting the criteria were chosen for primer design and testing using the Sequenom MassARRAY iPLEX assay.

Single locus specificity of SNP markers is important for accurate SNP allele calls and the new markers were tested against 3 progeny populations to determine by Chi- square analysis if the markers showed Mendelian inheritance. The 3 populations chosen had genetically diverse parents to also investigate if the markers were suitable for both vinifera and non-vinifera varieties. 52 markers distributed across 15 of the 18 grape chromosomes were selected from this study.

The SNP data from the 3 progeny analyses were also used to evaluate the discriminating power of SNP profiles for closely related individuals such as parents and siblings. It was found that the 48 SNP set had decreasing discrimination ability when the genome of individuals had more non-vinifera DNA present. A SNP 52 set composed of 52 selected SNP markers from the 218 set had very good discriminating power for both vinifera and non-vinifera genotypes.

The advantage of SNP markers and the MassARRAY iPLEX technology is that it is relatively easy to add new SNP markers to existing SNP sets to expand the number of SNP markers used for grapevine identification. The cost of the assay depends on how many markers can be multiplexed in one reaction. Past unit prices quoted to CSIRO from AGRF indicated $5 per unit (1 sample x 1 multiplex reaction) for SNP analysis if a minimum order of $1000 was placed. If 30 markers were multiplexed this would be $0.17 per marker. The AGRF grapevine identification service available to industry using SSR markers is $120 per sample and $11.50 per marker when DNA extraction costs are removed. It should be noted that these AGRF prices are indicative only and may change, however, it is clear that SNP markers are a cheaper alternative to SSR marker analysis if samples are combined in a batch to give a value above the minimum order cost.

4.5 Sources of errors in identification

There are a number of error sources that can occur during the grapevine DNA identification process. These error types are not unique to grapevine identification and may include:

1. Sampling tissue from the wrong plant.
2. DNA sample mix up during extraction or storage.
3. Sample mislabeling during the DNA profiling process.
4. Errors in allele calling by software.

Mistakes due to tissue sampling and DNA sample mislabeling can be minimized by implementing improved procedures. Sequenom literature reports 99.7% accuracy in SNP allele calling for the MassARRAY iPLEX technology indicating that while errors will occur it occurs at a low frequency. For grapevine DNA it was found that poor quality DNA and/or low DNA concentration could contribute to this error rate and increase the number of failed allele calls.

4.6 Web accessible SNP database for grapevine identification

Implementation of a SNP grapevine identification system requires a database holding DNA profiles of varieties that can be used to query new DNA profiles against. A MySQL database was developed containing the 266 SNP markers developed in this project and associated SNP DNA profiles for varieties studied in this project.

The database and web front end is a beta version and this type of database containing reference genotypes will be essential for any research or commercial service requiring a SNP database to query samples against to determine grapevine identity.

4.7 DNA 'gold standards' versus plant reference standards

A "type reference" can refer to the plant from which DNA is extracted, the DNA sample or the DNA profile of that plant. The main issue with using an individual plant of known location in a germplasm collection as a "type reference" is that there is a risk that the plant may not be available at some future date due to plant death or closure, relocation or change of ownership of the collection. The risk associated with using a DNA profile as a "type reference" is that new technology and marker types may make the collected DNA profiles obsolete. This risk is high at the present time as the genomics field is undergoing rapid change as new technologies replace older technologies over relatively short time frames.

Stored DNA samples that have been verified by a DNA profiling method as a reference for a particular variety may provide the solution. Once verified, these DNA samples would be deemed 'Gold Standards' and stored by an assigned institution for future reference. Aliquots of the DNA 'Gold Standard' could be shared between international laboratories if an international initiative was developed to cross-reference international grapevine collections. The stored DNA 'Gold Standards' could also be used in the future for the development of variety DNA profiles using new technologies and marker types. This would also alleviate the need and cost to re- extract DNA from plants at a future date.

A DNA 'Gold Standard' reference collection for long-term storage would normally be stored in a freezer. To investigate the option of storing a DNA reference collection at room temperature and transporting DNA at room temperature between international laboratories a small study was conducted within this project. Grapevine DNA was stored at room temperature either as a liquid or dry using the commercial reagent. After 39 weeks storage at room temperature the samples were DNA profiled using the SNP 48 set. The DNA profiles matched the DNA profiles from the frozen DNA control sample confirming that room temperature storage and transport of a DNA 'Gold Standard' Reference collection might be a viable future option.

5 Outcome & Conclusion

The objective of this pilot study was to evaluate mechanisms and methods for validating the names of grapevine varieties held in Australia. This was achieved but the study also identified a number of issues that require further consideration for any future study.

An international collaborative framework was established for the testing of DNA samples of the same variety from 4 countries; Australia, USA, Germany and Spain. The collaboration was successful and may represent a framework to expand upon in any future study seeking to validate grapevine varieties held in Australian germplasm collections. Scientists from both Portugal and Italy with access to national germplasm collections have approached the author and expressed interest in being part of any future initiative. Any future study would need to ensure that all collaborative partners develop appropriate quality control procedures in sample collection and distribution. This would be essential if a decision was made to produce DNA 'Gold Standards' for each variety. A common problem with international collaborations is funds available to each partner to participate in an international project. If GWRDC decided to fund a future project to DNA profile Australian varieties the involvement of an international collaborator or collaborators would still be required for cross referencing and validation.

A number of countries in Europe have previously received EU funding to harmonise European grapevine collections and identify mistakes and synonyms but personal communications suggest the funding is coming to an end. The focus for individual countries and the current European collaborative initiative is to investigate genetic diversity in grapevine germplasm collections to identify and preserve unique germplasm and investigate genetic structure. The two European grapevine databases (European Vitis Database and Vitis International Variety Catalogue) were developed based on international collaborative research efforts to cross-reference varieties and mostly contain phenotypic information and location of the plants referenced. The available DNA profiling information is currently limited to 6 SSR markers for only some varieties. It is known that this number of SSR markers is too few to accurately identify grapevine varieties with 10 the standard number and sometimes 20 markers required (Laucou et al., 2011; Lacombe et al., 2013). As shown in the Results section SSR marker values also do not easily transfer between laboratories.

This project investigated a new DNA marker type known as Single Nucleotide Polymorphisms (SNPs) for DNA identification of grapevine as it has a number of advantages over SSR markers. The marker type proved useful for grapevine identification with a SNP 48 set effective at identifying *V. vinifera* varieties while a 52 SNP set was better at distinguishing rootstock and non-vinifera varieties. The technology used for SNP analysis was a MassARRAY iPLEX assay with MALDI-TOF MS detection.

The Sequenom MassARRAY iPLEX MALDI-TOF MS system is a high throughput system that would be suitable for DNA profiling in a cost effective way the two major Australian germplasm collections (CSIRO and SARDI) that contain approximately 50% of germplasm and accessions, and up to 80% of unique material (McMichael et al., 2013) or even all the unique material in Australia if available. Use of this technology by the Australian industry after a SNP reference database was developed would require a service provider that could offer a service to industry at an acceptable cost. For this type of technology it is suggested that the industry consider combining sample submissions to minimize the cost per sample. The alternative is to use a low throughput technology for industry provided samples.

The Australian industry would benefit from having at least one Australian service provider offer a grapevine typing service as there are risks associated with relying on overseas providers. AGRF has corporate accreditation to ISO 17025 and currently provides the industry with a grapevine SSR identification service.

A SNP identification service offered to industry or for use in an international research collaboration will require access to a SNP reference database. A beta version of a web accessible database was

developed in this project as a proof of concept.

It is suggested that the Australian industry consider the development of a DNA ‘Gold Standard’ reference collection to be held in Australia to future proof against changes in plant access and DNA marker and technology changes.

6 Recommendations

This pilot project successfully trialed an international collaborative arrangement to validate the identification of a select group of grapevine varieties (scions and rootstocks) held in Australian germplasm collections. A new marker type, SNPs (Single Nucleotide Polymorphisms), was used as it offered a number of advantages over microsatellite (SSR) markers. A beta version of a grapevine variety SNP database was developed that could be used by both researchers and commercial service providers to query unknown samples against.

6.1 Future Research Directions

The grapevine germplasm collections in Australia have not been completely DNA profiled against international agreed type reference standards to validate the identity of the varieties held in the collections. Research consensus in relation to internationally agreed standards for grapevine varieties is under ongoing discussion. Initiatives in many countries to validate national grapevine collections have been based on SSR marker technology that was originally developed for grapevine by CSIRO (Thomas and Scott, 1993; Thomas et al., 1994) 20 years ago. A similar initiative in Australia has not occurred apart from an early study to provide a SSR database of a set number of varieties to a commercial service provider to allow industry to submit samples for testing.

The 2013 GWRDC Final Report (GWR 1112) “Review of Grapevine Germplasm Collections In Australia” details the current state of both research and industry collections. The two main collections, the largest one held by CSIRO and the second largest one held by SARDI are currently closed to industry but hold up to 80% of unique material (McMichael et al., 2013). DNA profiling the major Australian collection(s) would benefit both scientists and industry and may lead towards a pathway for an Australian National Collection. Scientists would benefit from the knowledge obtained by DNA profiling the Australian gemplasm collections in projects investigating grapevine improvement (rootstocks and scions), biosecurity, genetic diversity and adaptation to climate change.

It is recommended that the Australian industry consider funding a project to DNA profile the major Australian collections. It is further recommended that the SNP marker type be used for DNA profiling and a high throughout SNP technology be used to minimize costs. To avoid potential future issues with access to reference plants and changes in technology and DNA marker types it is recommended that a DNA ‘Gold Standard’ reference collection be held in Australia. Involvement of a least one international research group in a new project is essential to validate Australian varieties against. The ideal situation would be to involve multiple international groups so that Italian varieties are validated against the Italian collection, Spanish varieties against the Spanish collection, etc. However the willingness of international collaborators and budget constraints will limit what is possible.

6.2 Industry Context and Adoption Scenarios

In Australia, compliance with the Label Integrity Program is required for wine labeling and if a variety name is on the label an audit trail must be kept. A DNA based method for variety identification would considerably reduce the risk to members of the Australian industry and Australia's international reputation by minimizing the chance of accidental mislabeling events. The Australian Standard for Grapevine Propagation Material, AS 5588:2013 provides industry with a voluntary Australian Standard. An important component of this scheme will be DNA profiles that accurately identify grapevine varieties and application of this to trace a variety from vine importation to nursery to vineyard to the winery for label integrity.

The industry benefits of a DNA identification method for grapevine extend beyond compliance with the Label Integrity Program and distribution of material through the supply chain. Germplasm diversity is essential for research projects that seek to provide industry with improved planting material that reduce costs to industry, improve wine quality, are better adapted to the changing Australian climate and are resistant to endemic pests and diseases and exotics that represent a biosecurity threat. DNA profiling the Australian research collections would determine the extent of the genetic diversity in Australia and what new plant material needs to be imported into Australia to extend the existing genetic diversity to fill the missing gaps where important traits are missing.

If the Australian industry accepts the above recommendation that a project to DNA profile the major Australian collections be funded then an essential component of that project should include a means to provide the industry with at least one Australian commercial provider of a DNA profiling service for grapevine identification. It is also recommended that an advisory reference group composed of representatives of peak industry bodies and key scientists be formed to provide guidance to the project to ensure it meets the needs of the Australian industry and has the necessary global linkages with organisations (e.g. International Organisation of Vine and Wine (OIV)) involved in clarifying and assigning variety names and synonyms.

Appendix 1 Communication

During the course of this project communications occurred with members of the Industry Reference Group associated with the project.

Appendix 2 Intellectual Property

1. SNP primer designs for SNP analysis using a Sequenom MALDI-TOF MS detection system.

2. SNP DNA profiles of selected varieties.

3. A MySQL database with a web front end. The database contains SNP DNA profiles of selected validated varieties. The database is expandable to accommodate SNP DNA profiles of additional varieties and new markers. The web front end allows users with access to query the database with SNP DNA profiles of unknown samples to determine variety DNA profile matches.

Appendix 3 References

Bacilieri R, Lacombe T, Le Cunff L, Di Vecchi-Staraz M, Laucou V, Genna B, Peros JP, This P, Boursiquot JM (2013) Genetic structure in cultivated grapevines is linked to geography and human selection. BMC Plant Biology **13:** 25

Cabezas JA, Ibanez J, Lijavetzky D, Velez D, Bravo G, Rodriguez V, Carreno I, Jermakow AM, Carreno J, Ruiz-Garcia L, Thomas MR, Martinez-Zapater JM (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biol **11:** 153

Chaib J, Torregrosa L, Mackenzie D, Corena P, Bouquet A, Thomas MR (2010) The grape microvine - a model system for rapid forward and reverse genetics of grapevines. Plant Journal **62:** 1083-1092

Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A, Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theor Appl Genet **121:** 1569-1585

Lacombe T, Audeguin L, Boselli M, Bucchetti B, Cabello F, Chatelet P, Crespan M, D'Onofrio C, Dias JE, Ercisli S, Gardiman M, Grando MS, Imazio S, Jandurova O, Jung A, Kiss E, Kozma P, Maul E, Maghradze D, Martinez MC, Munoz G, Patkova JK, Pejic I, Peterlunger E, Pitsoli D, Preiner D, Raimondi S, Regner F, Savin G, Savvides S, Schneider A, Spring JL, Szoke A, Veres A, Boursiquot JM, Bacilieri R, This P (2011) Grapevine European Catalogue: Towards a Comprehensive List. Vitis **50:** 65-68

Lacombe T, Boursiquot JM, Laucou V, Di Vecchi-Staraz M, Peros JP, This P (2013) Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). Theor Appl Genet **126:** 401-414

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948

Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Vares D, Peros JP, Boursiquot JM, This P (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet **122:** 1233-1245

Maul E, Sudharma KN, Kecke S, Marx G, Muller C, Audeguin L, Boselli M, Boursiquot JM, Bucchetti B, Cabello F, Carraro R, Crespan M, de Andres MT, Dias JE, Ekhvaia J, Gaforio L, Gardiman M, Grando S, Gyropoulos D, Jandurova O, Kiss E, Kontic J, Kozma P, Lacombe T, Laucou V, Legrand D, Maghradze D, Marinoni D, Maletic E, Moreira F, Munoz-Organero G, Nakhutsrishvili G, Pejic I, Peterlunger E, Pitsoli D, Pospisilova D, Preiner D, Raimondi S, Regner F, Savin G, Savvides S, Schneider A, Sereno C, Simon S, Staraz M, Zulini L, Bacilieri R, This P (2012) The European Vitis Database (http://www.eu-vitis.de) - a technical innovation through an online uploading and interactive modification system. Vitis **51:** 79-85

McMichael P, Hamilton R, Tassie E (2013) GWR 1112 Review of Grapevine Germplasm Collections in Australia. GWRDC Final Report, p 56

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES (2011) Genetic structure and domestication history of the grape. Proc Natl Acad Sci USA **108:** 3530-3535

Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D (2010) Rapid Genomic Characterization of the Genus *Vitis*. PLoS ONE **5:** e8219

Thomas MR, Cain P, Scott NS (1994) DNA Typing of Grapevines - a Universal Methodology and Database for Describing Cultivars and Evaluating Genetic Relatedness. Plant Molecular Biology **25:** 939-949

Thomas MR, Scott NS (1993) Microsatellite Repeats in Grapevine Reveal DNA Polymorphisms When Analyzed as Sequence-Tagged Sites (STSs). Theoretical and Applied Genetics **86:** 985-990

Walker R, Thomas MR (2009) Evolution of grapevine variety collections in Australia. Grapegrower & Winemaker 2009 Annual Technical Issue. **2009:** 78-80

Appendix 4 Staff

CSIRO Plant Industry: Dr Mark Thomas, Dr Rob Walker, Mr Don Mackenzie and Mrs Hilary Davis.