The DNA of innovation

By Anthony Borneman, Markus Herderich and Dan Johnson
The Australian Wine Research Institute, PO Box 197, Glen Osmond, SA 5064, Australia.
Email: anthony.borneman@awri.com.au

Mangaging director Dan Johnson

Winemaking is arguably the world's oldest biotechnology, with a history dating back more than 7000 years. The past century has seen unprecedented advances in wine research with the advent of genomics and DNA sequencing. These technologies stand to transform our understanding of wine yeast, boost efforts to combat pests, diseases and contaminants and provide greater insight into wine character, regionality and terroir. This AWRI Report explains how and why DNA sequencing technologies are relevant to grape and wine producers and outlines their potential for the future.

THE SCIENCE OF DNA

DNA is the blueprint of life. It resides within every living organism, from bacteria to complex plants and animals. It represents an encrypted repository of the thousands of individual instructions that are required for cells to grow and respond to their environment.

Genomics is the science behind decoding and understanding DNA. Until recently, this discipline was limited by the enormous time and cost required to obtain the genomic information of even the simplest life forms, such as bacteria or yeast, and translate it into a useable form.

To understand the complexity of this work, a brief overview of DNA structure is required. Rather than being a single, homogenous chemical, DNA (or deoxyribonucleic acid) is comprised of chains of four slightly different sub-units or bases, represented by the letters A, C, G and T. These bases are joined together to form very long polymers (Figure 1).

Instructions for the growth and development of an organism are encoded by the DNA bases and their precise order along the DNA strand. In a typical human genome, there are at least 20,000 individual instructions (genes) spread across three billion DNA letters – equivalent to roughly 100,000 A4 pages of 12-point text. The science of genomics unravels this information. It decodes the instructions into a form that we can understand and, therefore, use.

Next-generation DNA sequencing is revolutionising this work and biological science as a whole. DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It utilises a variety of technologies to determine the order of the four bases—adenine, guanine, cytosine, and thymine (A, G, C, T)—in a strand of DNA. Lower costs and processing times allow unprecedented access to genomic data; links between DNA and particular traits or characteristics can now be explored, even in complex plants and animals. Cutting-edge technologies are taking this information and using it in practical applications such as personalised medicine or genome-assisted breeding of crops or animals.

Applying next-generation DNA sequencing in the wine industry provides an unprecedented ability to accurately understand, select and track biological factors in the winemaking process. Importantly, unlike genetic modification, genomics does not seek to alter the genetic make-up of an organism. As such, genomics can be utilised

AT A GLANCE

- Next-generation DNA sequencing systems are delivering data faster and cheaper than ever before
- Genetic information from vines, yeast and bacteria stands to revolutionise grape and wine production, including unravelling the constitution of terroir, providing new tools to combat vineyard pathogens and understanding the composition of wild ferments
- The genomes of some strains of wine yeast have already been decoded, offering new openings for research into fermentation performance and flavour
- Consumers will benefit through improved products and an understanding of taste preferences.

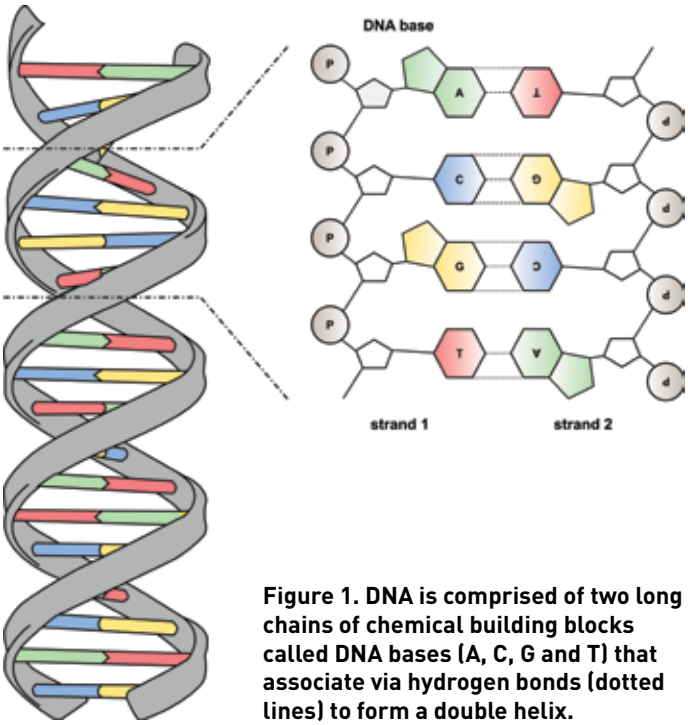

Figure 1. DNA is comprised of two long chains of chemical building blocks called DNA bases (A, C, G and T) that associate via hydrogen bonds (dotted lines) to form a double helix.

even within organic or biodynamic winemaking practices. The data provided by genomics will allow precision breeding of new grape cultivars or strains of yeast and bacteria that can tailor wine flavours. It can also provide grapegrowers and winemakers with the ability to monitor microbial populations in vineyard soils, wineries or wild fermentations. These data can then be correlated with viticultural practices, soil health or wine quality to determine factors that can lead to increased wine quality or production efficiency.

THE NEXT-GENERATION SEQUENCING REVOLUTION

The field of biological research was revolutionised by the pioneering work of Sanger *et al.* (1977) who provided an 'efficient' means of sequencing DNA. For more than a quarter of a century, 'Sanger' sequencing led the way, until next-generation technologies offered an even more efficient alternative.

Despite nearly 30 years of significant improvements, 'Sanger' sequencing delivered relatively low output combined with high labour and instrument costs. This restricted access and uptake among potential users: the study of entire genomes was limited to large, multinational collaborations and specialised sequencing centres.

All this changed in 2005 with the introduction of massively-parallel pyrosequencing systems. Since then, the development of next-generation DNA sequencing technologies has progressed at an incredible pace with rapid improvements to several competing instruments. Collectively, these instruments are referred to as next-generation sequencers (Table 1).

In the last decade, competition has driven down the expense of DNA sequencing at a tremendous rate, with dramatic increases in output and efficiency. As a result, the raw, per-base

Table 1. Current sequencing technologies (Mardis 2013)

Company	Machine	Machine cost	Output per day (number runs)[a,b]	Effective yeast genomes per day[c]
ABI	AB3730xl	$	0.0016 Gb (24)	0.013
Roche-454	GS XLR70	$$	1.08 Gb (2.4)	10
	GS Junior	$	0.084 Gb (2.4)	0.46
Illumina	Hiseq 2000	$$$	55 Gb (0.09)	90
	Miseq	$	4.5-5.1 Gb (0.6)	25-30
Life Technologies	Proton	$$	60-120 Gb (6-12)	100-200
Pacific Bioscience	PacBio RS II	$$$	2.64 Gb (12)	22

[a] number of runs refers to how many individual samples the machine can process in a 24-hour period
[b] Gb = Gigabase, equivalent to one billion DNA letters
[c] Effective number of genomes takes into account read length as shorter sequencing reads require higher coverage for assembly. AB3730xl and PacBioRS II output is based on 10x coverage, Miseq and Roche-454 sequencers based on 15x coverage and Hiseq and Proton sequencing on 50x coverage.

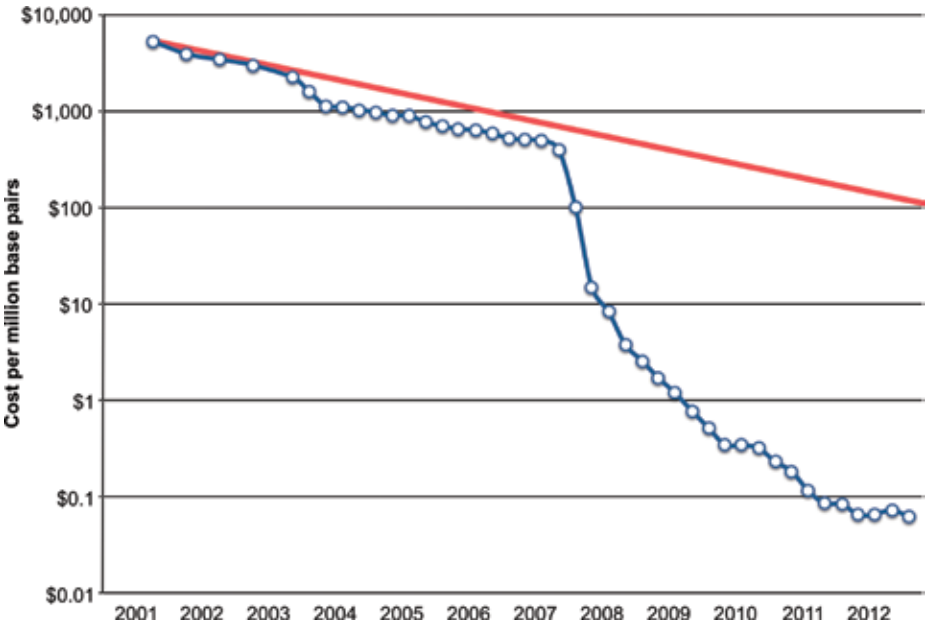

Figure 2. Dramatic drop in the cost of DNA sequencing.

sequencing cost has dropped by a factor of 10 every 18 months (Figure 2, page 53).

Collectively, these advances have made genome sequencing much cheaper and more accessible: today, genomics technologies are used by smaller, individual laboratories at little cost. As a result, genome sequencing is being applied in clinical diagnostics and agricultural research on a scale that was not economically viable even 12 months ago.

Prior to the introduction of next-generation sequencing, the cost of DNA sequencing was reducing at a rate approximating 'Moore's law' (halving in cost every two years; red line). Since its introduction, next generation sequencing has driven the per-base cost of DNA sequencing down by a factor of 10 every 18 months (Wetterstrand 2013)[6,7].

GENOMICS AND THE WINE INDUSTRY

The history of winemaking dates back some 7000 years; it is the past century that has seen the most dramatic advances in the application of cutting-edge scientific research to improve wine production.

Given the roles of grapevines, yeast and bacteria in shaping the composition of finished wines, it is not surprising that genomics is poised to play an ever-increasing role in unlocking their potential as biological inputs (summarised in Figure 3).

Connecting the phenotypic traits and characteristics of grapevines, yeast and bacteria with specific genomic features stands to have the most significant effect on the wine industry. This will allow genetic variations to be identified and links to be made with the production (or reduction) of particular characters.

Microbial strain development

Due to the relatively small sizes of their genomes, both the wine yeast *Saccharomyces cerevisiae* and the malolactic bacterium *Oenococcus oeni* have already been the subject of significant genomic research. Genome sequences are available for more than 80 strains of *S. cerevisiae* (although the vast majority of these are not wine strains) and more than a dozen strains of *O. oeni*. Comparisons among these sequences have already identified genomic differences within species that might be linked to wine-relevant traits such as fermentation robustness or flavour production.

It is now possible to sequence large numbers of yeast and bacteria, even with single sequencing runs on low-cost benchtop sequencing machines such as the Miseq from Illumina (see Table 1). As more strains within a species are sequenced, more comparisons and connections can be made. Once identified, genomic differences can be used as molecular markers to accurately predict the expected phenotype of strains without costly and laborious manual analysis. This dramatically increases the speed at which strains can be selected for commercial application.

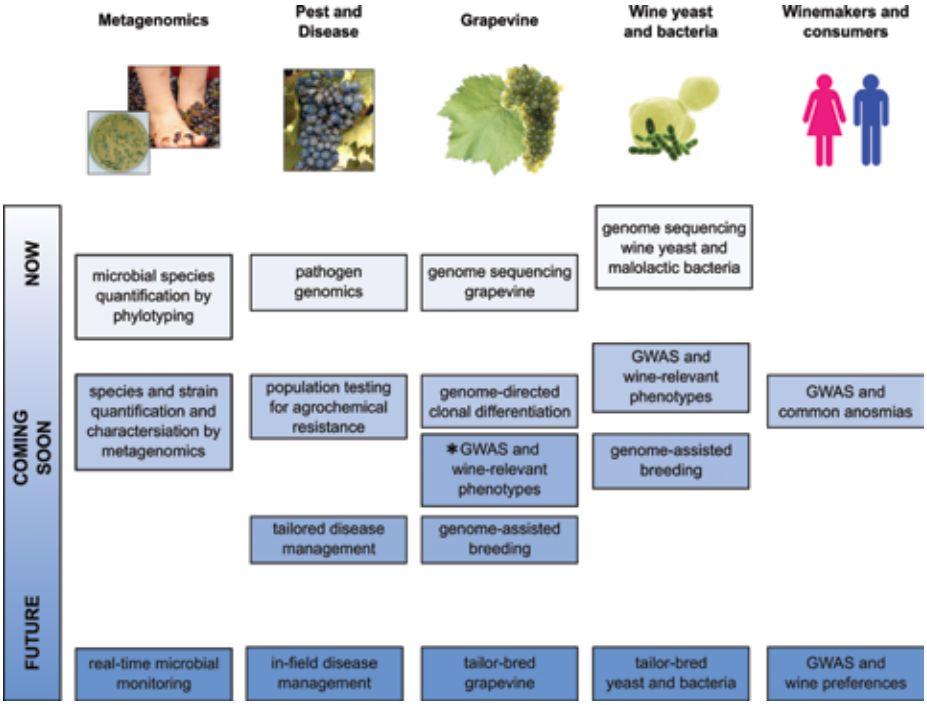

Figure 3. Summary of applications of next-generation sequencing in the wine industry. *GWAS=Genome-wide association studies

Genomics in the study of pests, diseases and contaminants

Genome sequences are currently available for a small number of vineyard pests and diseases including Botrytis rot, Pierce's disease and winery contaminants such as *Brettanomyces* (see Amselem *et al.* 2011, Simpson *et al.* 2000, and Curtin *et al.* 2012 respectively).

Comprehensive genome information for important pathogens such as powdery and downy mildew, phylloxera, and problem strains of already sequenced pathogens and spoilage microorganisms are still lacking, however. Fortunately, the advances provided by next-generation sequencing should see these gaps filled in the near future.

Agrochemical resistance is also a priority. In regions with wet summers, constant pressure from pathogens can be exacerbated by resistance to fungicides. Next-generation sequencing offers the ability to track known markers for agrochemical resistance accurately (Wicks and Wilson 2012). These data will be invaluable for grape producers as they plan strategically to use the right combinations of agrochemicals to achieve the best results within a single season and manage agrochemical resistance in the longer term.

Genomics in the study of diversity, regionality and terroir

Metagenomics involves the sequencing of DNA isolated from environmental samples (e.g. water, soil, air, faeces) that are composed of complex mixtures of microorganisms. The 'Human Microbiome Project' and the 'Earth Microbiome Project' have highlighted the importance of metagenomics: both projects have sought to determine the microbial composition of thousands of samples taken from the human body and from natural environments, respectively. Similar to single species studies, the aim of metagenomics is to link specific microbial genomes (or metabolic pathways) in heterogeneous samples with specific traits.

Due to the complexity of many microbial communities, true metagenomic sequencing cannot be achieved, however, even with the most cutting-edge of current next-generation technologies. Instead, a scaled-back form of metagenomics (often termed phylotyping) can be applied – this uses a small portion of DNA as a 'genomic barcode'.

For the wine industry, metagenomics is already proving to be a huge help in the study of vineyard microbiota and wild fermentations. Comparisons can be made between microbial populations from conventionally farmed, organic and

biodynamic vineyards; researchers can also compare similarly managed vineyards in different geographical locations. This will provide scientific data regarding the effects of geography combined with different vineyard practices on soil composition and greater insight into possible relationships between soil microbiota, sustainable vineyard management and regional terroir. It will also offer a way to systematically and rigorously assess the effects of viticultural interventions on soil microorganisms after base-line measurements have been taken.

Monitoring the composition of wild ferments may be one of the most useful applications of metagenomics. Wild ferments are typically characterised by a progression of diverse microbial species that, due to a combination of factors, generally conclude with *S. cerevisiae* strains being the dominant yeast at work. However, it is the varied metabolic contribution of the non-*Saccharomyces* yeasts at the beginning of wild fermentations that are thought to provide the complex characteristics that make wild ferment wines preferable to many of their inoculated counterparts.

Previous analytical methods were too labour-intensive to identify wild ferment species efficiently in complex microbial mixtures. Applying metagenomic tools such as phylotyping will provide data that can be used to correlate the incidence of wild ferment species in individual fermentations with final wine composition, or to judge the effect of geography or viticultural and winemaking interventions (harvest method, temperature, SO_2) on wine microbiota. Research has already shown that the technology is useful for tracking bacterial and fungal composition during wine production (Bokulich and Mills 2013, Bokulich *et al.* 2012).

While moving from phylotyping to full metagenomic sequencing will lead to higher costs, it will allow the contributions of individual strains of wild ferment species to be tracked effectively. These extra data will be invaluable when the presence of specific strains of yeast or bacteria result in unexpected (either desirable or undesirable) oenological outcomes or when winemakers wish to know if commercial microbial strains interfere with their 'wild' fermentations. Alternatively, it might become apparent from sequencing data and metagenomics analysis that specific wineries or geographical locations harbour unique strains of wine yeast and bacteria that contribute to their distinctive terroir.

The future

Development of genomic sequencing technologies continues to progress at an astonishing rate. New approaches, such as nanopore-based techniques (e.g. www.oxfordnanopore.com), promise to make sequencing even cheaper, faster and, perhaps, accessible in the field via a simple USB stick attached to a laptop computer. The application of such simplified sequencing technologies will enable close-to-real-time data to be gathered on pathogen loads and likely levels of agrochemical resistance in a vineyard, providing ways to tailor viticultural intervention. Likewise, the ability to analyse the composition of wild ferments in real-time will enable winemakers to intervene in individual fermentations that display sub-optimal mixtures of microflora or that contain unwanted microbial contamination. This has the potential to prevent write-offs from 'failed' wild ferments.

Advances in human personalised genomics might also offer new opportunities to grape and wine producers. As more links between genetic differences and our ability to smell and taste are made, the application of human personalised genomics will no longer be limited to health outcomes; scientists could also reach a better understanding of why some people perceive flavour and aroma differently. There are already recognised genetic variants associated with the perception of bitterness or in perceiving a 'soapy' taste in coriander. As additional genetic associations are made, it might be that taste preferences can be predicted at birth from a standard genome sequence analysis. This offers benefits for consumers as well as grape and wine producers. At the very least, future winemakers may be forewarned if they have any genetic

predispositions that leave them anosmic to (unable to perceive) specific aroma compounds and, therefore, affect the quality of the wine they produce (AWRI Annual Report 2011).

CONCLUSION

Winemaking is a well-established microbiological process; yet, the rapid development of next generation sequencing is poised to revolutionise all aspects of this ancient biotechnology. Embracing new genomic sequencing technologies will provide researchers and producers with insights into the biological factors and inputs (grapevine, yeast, bacterial and human) that influence winemaking. It is beyond doubt that the increasing simplicity, uptake and accessibility of these new technologies will improve decision making, competitiveness, innovation and economic sustainability among Australian grape and wine producers. Innovation in this emerging area also stands to benefit consumers; communicating the advantages of genomics technologies effectively to the wider public is, therefore, also a priority.

ACKNOWLEDGEMENTS

This work is funded by Australian grapegrowers and winemakers through their investment body, the Grape and Wine Research and Development Corporation. The Australian Wine Research Institute is a member of the Wine Innovation Cluster. The authors thank Sharon Mascall-Dare and Rae Blair for their editorial assistance.

REFERENCES

Amselem, J.; Cuomo, C.A.; van Kan, J.A.L.; Viaud, M.; Benito, E.P.; Couloux, A.; Coutinho, P.M.; de Vries, R.P.; Dyer, P.S.; Fillinger, S. *et al.* (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. Plos Genet. 7, e1002230.

AWRI Annual Report (2011) Taint Investigations 47-48.

Bokulich, N.A. and Mills, D.A. (2013) Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl. Environ. Microbiol. 79:2519–2526.

Bokulich, N.A.; Joseph, C.M.L.; Allen, G.; Benson, A.K. and Mills, D.A. (2012) Next-generation sequencing reveals significant bacterial diversity of botrytised wine. Plos One. 7,e36357.

Curtin, C.D.; Borneman, A.R.; Chambers, P.J. and Pretorius, I.S. (2012) De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast Dekkera bruxellensis AWRI1499. Plos One. 7, e33840.

Mardis, E.R. (2013) Next-Generation Sequencing Platforms. Annu. Rev. Anal. Chem. 6:287-303.

Simpson, A.J.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.; Araya, J.E.; Baia, G.S.; Baptista, C.S. *et al.* (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The Xylella fastidiosa Consortium of the Organisation for Nucleotide Sequencing and Analysis. Nature 406:151–159.

Wetterstrand, K.A. (2013) DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcosts. Accessed 7 June 2013.

Wicks, T. and Wilson, D. (2012) More strobilurin-resistant powdery mildew detected in vineyards. Grapegrower and Winemaker 580:24.