Assessing diversity and clonal variation of Australia's grapevine germplasm: Curating the FALCON-Unzip Chardonnay *de novo* genome assembly

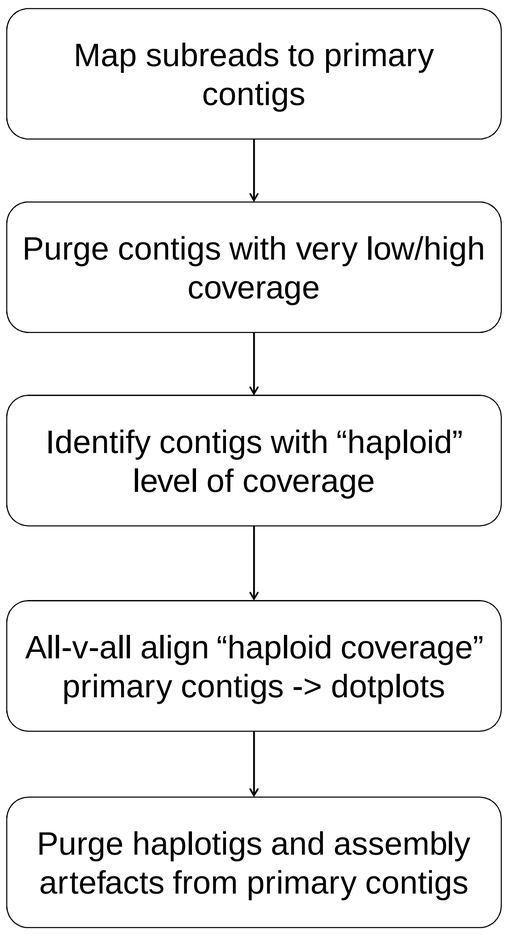
Michael J. Roach[1], Simon A. Schmidt[1] and Anthony R. Borneman[1]

[1]The Australian Wine Research Institute, PO Box 197, Glen Osmond (Adelaide) SA 5064, Australia

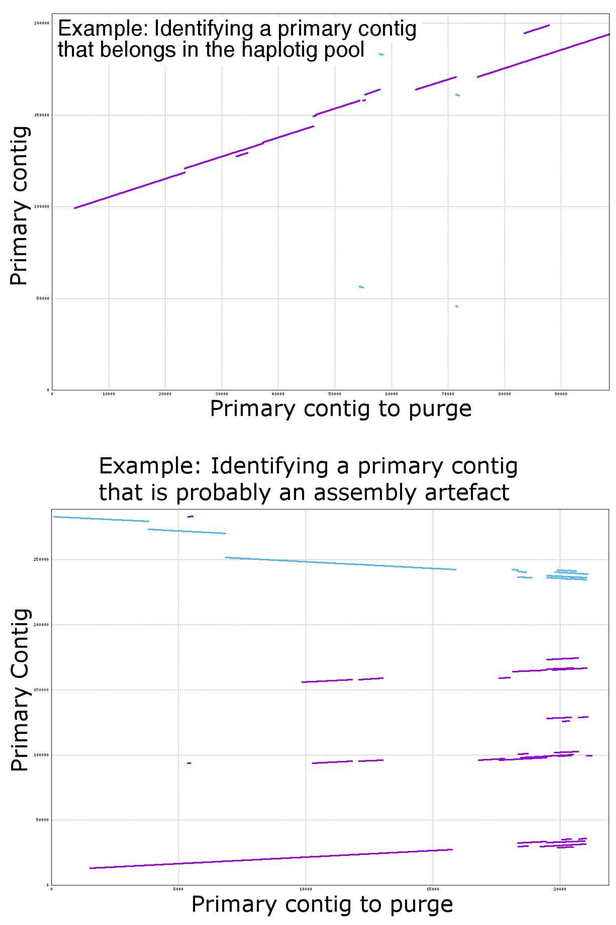
Background: Until recently only two genome assemblies were publically available for grapevine—both *Vitis vinifera L. Cv.* Pinot Noir (PN). The best available PN genome assembly (Jaillon et al. 2007) is not representative of the genome complexity that is typical of wine-grape cultivars in the field and it is highly fragmented. To assess the genetic complexities of Chardonnay grapevine, assembly of a new *de novo* reference genome was needed. Here we describe a draft assembly using PacBio SMRT sequencing data and PacBio's new phased diploid genome assembler FALCON-Unzip (Chin et al. 2016).

I'm unzipped, now what do I do? 1A) Genome curation pipeline overview **1B)** Identifying haplotigs and assembly graph artefacts with dotplots **1C)** Assembly stats, comparison and improvement **1D)** Validation with BUSCO (Simão et al. 2015)

1A)



1B)



1C)

	Pinot Noir (2007)	Chardonnay genome assembly raw	Chardonnay genome assembly curated
Size (Mb):	486.197	590.740	494.390
Contigs:	14 634	1 803	978
N50 (bp):	102 851	743 564	922 070
Largest Contig (bp):	653 287	6 339 470	6 347 624

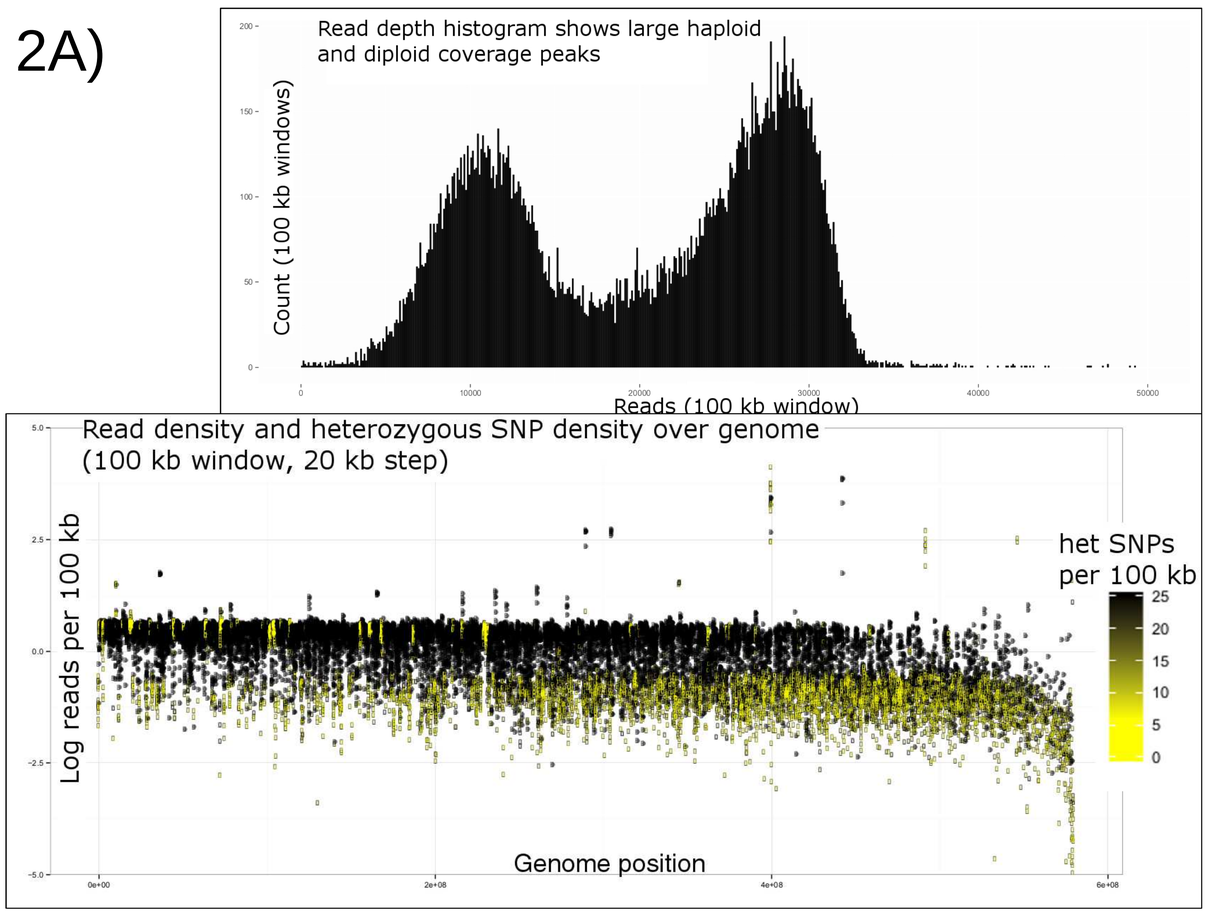
1D)

	Chardonnay genome assembly raw	Chardonnay genome assembly curated
Complete BUSCOs:	892	892
Fragmented BUSCOs:	20	23
Missing BUSCOs:	44	41
Predicted completeness:	~ 95.4 %	~ 95.8 %

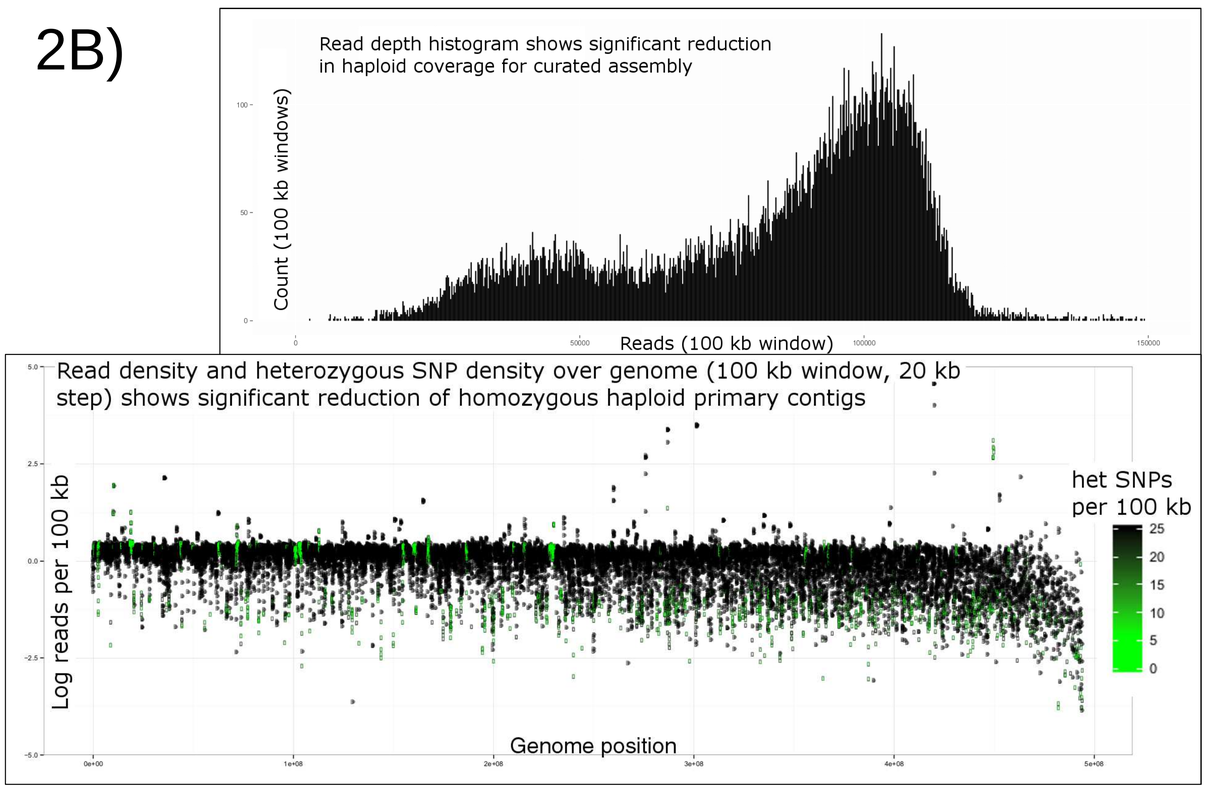
Read coverage and SNP density used to validate curation.

Paired-end illumina reads mapped to primary contigs and heterozygous SNPs called. **2A)** raw assembly **2B)** curated assembly

2A)



2B)



Conclusions:

- Long read sequencing enables highly contiguous phased genome assemblies
- FALCON-Unzip assemblies can be readily curated (even for problematic genomes)
- Phasing aids in genome curation

References:

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J.R., Cantu, D., Rank, D.R. and Schatz, M.C. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. Nature methods.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quetier, F., Wincker, P. and Public, F.-I. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–U5.

Simão, Felipe A. and Waterhouse, Robert M. and Ioannidis, Panagiotis and Kriventseva, Evgenia V. and Zdobnov, Evgeny M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics.

Acknowledgments: Special thanks to Jason Chin and PacBio for providing early access to FALCON-Unzip. We thank VLSCI and QRIScloud for providing the compute and support during assembly. Bioplatforms Australia are gratefully acknowledged for their generous funding and support.

This project 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.