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

Michael J. Roach¹, Simon A. Schmidt¹ and Anthony R. Borneman¹

¹The 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) Map subreads to primary contigs

1B) Example: Identifying a primary contig that belongs in the haplotig pool

1C) Identifying contigs with “haploid” level of coverage

1D) All-v-all align “haploid coverage” primary contigs -> dotplots

1A) 1B) Complete BUSCOs: Fragmented BUSCOs: Missing BUSCOs: Predicted completeness:

2A) Read coverage and SNP density used to validate curation.

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

2A) Read depth histogram shows large haploid and diploid coverage peaks

2B) Read depth histogram shows significant reduction in haploid coverage for curated assembly

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:

