Comparative genomics for three Sweet Cherry varieties

[1]Javier Cáceres-Molina, [1]Karin Rothkegel, [1]Tomás Carrasco-Valenzuela, [1,4]Claudio Urra, [1,3]Claudio Meneses, [2]Humberto Prieto and [1]Andréa Miyasaka Almeida

[1]Centro de Biotecnología Vegetal, Universidad Andrés Bello. Santiago, Chile. [2]Instituto de Investigaciones Agropecuarias, INIA La Platina, Santiago, Chile. [3]FONDAP Center for Genome Regulation, Santiago, Chile. [4]UC DAVIS-Chile, Life Science Innovation Center, Santiago, Chile.

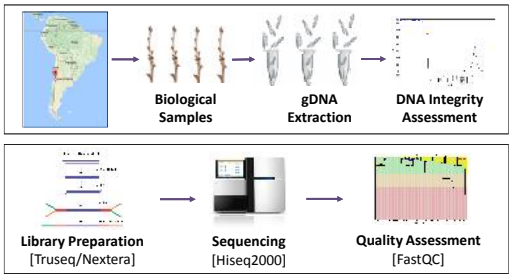
amiyasaka@unab.cl

Abstract

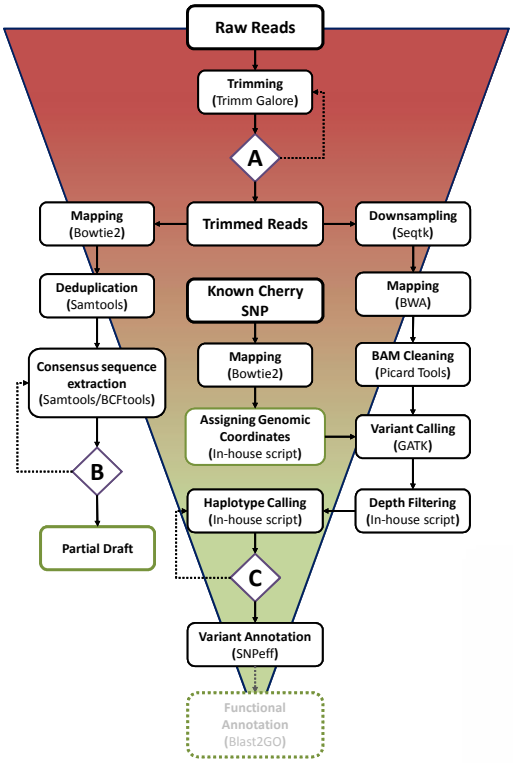
The Rosaceae is the third largest world plant family comprising approximately 3,400 species, many of them edible seasonal fruits which are prized for their unique flavors, colors and nutritional properties. Sweet cherry (Prunus avium) to this family and is an important fruit crop with a world production over 2 tons annually. It is a diploid species (2n=2x=16) with an estimated genome size of 338 Mbp as reported in the Genome Database for Rosaceae (GDR). Sweet cherry, similar to other species of the Rosaceae family, exhibits stylar gametophytic self-incompatibility and therefore has a highly heterozygous genomic background. Despite its commercial importance, there's a lack of genomic information in order to facilitate the identification of features associated to agronomical important traits. In order to overcome this limitation, our proposal was to take advantage of recently published genome of 'Satonishiki' Japanese sweet cherry variety to perform *in silico* comparative genomics using a whole genome sequencing strategy to search for meaningful structural differences between three sweet cherry varieties 'Karina', 'Kordia' and 'RoyalDawn', contrasting for diverse traits such as chill requirement.

METHODS

1. Sample preparation and sequencing:



2. Variant discovery pipeline:



RESULTS

Table 1. Libraries stats for the Hiseq libraries of Karina, Kordia and Royal varieties, summarizing the post-trimmed metrics for each one.

Variety	Chill Requirement	Type	R	Reads	Lenght (bp)	%GC
Satonishiki	High	-	-	-	-	-
Karina	Mid-High	Paired-end	R1	169,871,230	101	38
			R2	169,871,230	101	38
Kordia	High	Paired-end	R1	103,938,325	20-116	39
			R2	103,938,325	20-116	38
Royal Dawn	Low	Paired-end	R1	133,418,415	20-116	39
			R2	133,418,415	20-116	38

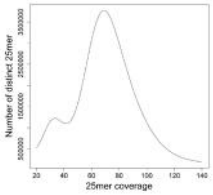


Figure 1. Learning from the data. Distribution plot of the 25mer species count of Karina libraries.

Table 2. Mapping efficiency and coverage statistics for the mapping step and variant filtration.

		Allele Read Depth			
Variety	Mapping (%)	Min	Max	Mean	SD
Karina	90.5	14	3,816	51.4	51.2
Kordia	87.9	14	2,867	42.7	39.4
Royal Dawn	86.6	14	2,840	43.8	40.8

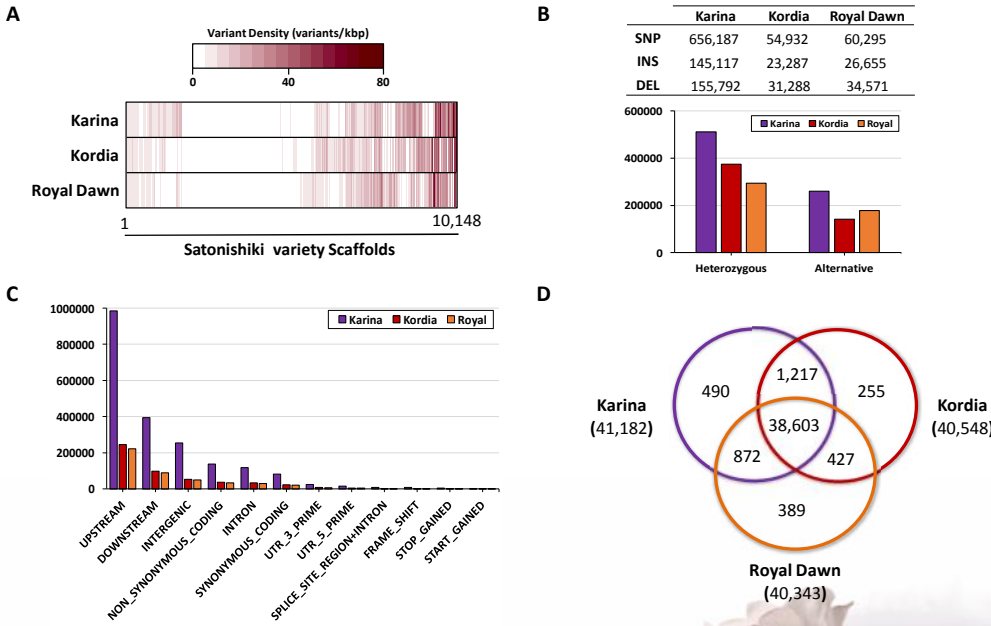


Figure 2. Variant discovery results. **A.** Variant density amongst the Satonishiki scaffolds **B.** Number of variants found of each type and genotype distribution. **C.** Variant effect distribution by ontology. **D.** Venn diagram showing the transcripts affected with variants.

CONCLUSIONS

The present work shows the results of the standardization of a pipeline for variant discovery in sweet cherry (*Prunus avium*), taking advantage of the recently published 'Satonishiki' variety genomic sequences, open source tools and in house developed computational programs. In the process, we mapped into the 'Satonishiki' scaffolds a population of SNP previously referenced to peach (*Prunus persica*) genome. Taking into account the mapping yield, one can think that the genetic background of the 'Karina' variety is the most similar to the Japanese variety, in the other hand the number of variants discovered in 'Karina' tells us the opposite. Despite the down sampling of the reads, 'Kordia' and 'Royal Down' varieties showed less allele depth than 'Karina', supporting the notion that the later is the most similar to the reference. The number of genes affected with variants otherwise tells us that Karina has the higher variability, suggesting that there's an uneven representation of several regions of the 'Kordia' and 'Royal Dawn' genomes. In order to overcome this, a higher depth sequencing is required. Also it is mandatory to refine the gene annotation in which a RNA-seq experiment could help to guide the generation of a new annotations of the genome.

AKNOWLEDGEMENTS

Financing support: projects FONDEF G09I1008, Consorcio Biofrutales-CORFO 13CTI21520-SP05.