Hans Merensky Avocado Genomics Project

at the University of Pretoria

Compiled by

Prof. Noëlani Van den Berg

and Dr. Sarah Mwangi

in collaboration, with

members of the AvoGenome Consortium

February 2017

Avogenome Consortium

In October 2016, the following avocado researchers initiated a SKYPE working group to discuss important advances in avocado genomics research. As a result of these positive discussions, the Avogenome sequencing project was established. Members of this group will all share data that are generated under the Fort Lauderdale agreement.

- Dr. David Kuhn, USDA, Florida
- Dr. Patricia Manosalva, UCR, California
- Prof. Noëlani van den Berg, UP, South Africa
- Dr. Sarah Mwangi, UP, South Africa
- Dr. Antonio Javier Matas Arroyo, Departamento de Biología Vegetal, University of Malaga, Spain
- Prof. Aureliano Bombarley Gomez, Virginia Tech Horticulture, USA
- Dr. Alan Chambers, University of Florida, USA

Scope of the project

After discussing needs and financial considerations the group decided to join forces to sequence the avocado genome. Four PIs from the group (Dr. David Kuhn, Dr. Patricia Manosalva, Prof. Aureliano Bombarely Gomez and Prof. Noëlani van den Berg) will form the core of this sequencing initiative. Dr. Sarah Mwangi, as our bioinformatician, will extensively be involved in the annotation of the genome alongside Prof. Bombarely. Based on Dr. Mwangi's assessment of the avocado genome data that we have from the Mexicans (Herrera et al.) the group feels that the draft assembly is not of sufficient quality and we therefore propose to sequence a homozygous avocado, followed by the re-sequencing of two rootstocks, one from the UCR breeding programme and one from WTS.

We propose the re-sequencing of the genome of **R0.06**, a *Phytophthora*-tolerant avocado rootstock selected by Westfalia Technological Services. The rootstock was selected as target because of its commercial importance to Westfalia Fruit and due to its performance in the field. We propose a hybrid whole genome sequencing approach using Illumina and PacBio. The sequence data will be assembled and annotated by Prof. Aureliano Bombarely Gomez and Dr. Sarah Mwangi.

To facilitate the process of developing molecular markers for avocado we suggest that an avocado half-sibling breeding population be established on the Westfalia Fruit Estate, Tzaneen as well as in at least one other location in South Africa. We also suggest that Dr. Kuhn be involved in this process, as he has done a similar population in Florida for avocado fruit cultivars. The USDA-Miami has both the California and Florida populations.

The second objective is to genotype avocado rootstocks and to determine parentage of valuable material. A selection of 91 rootstocks from the WTS genepool block has been collected and will be analysed with the use of the Illumina SNP chip designed by Dr. Kuhn, USDA, Florida. A MSc student, Juanita Hanneman, has joined the group and will be supervised by Prof. Noëlani van den Berg, Dr. Sarah Mwangi, Dr. David Kuhn and Prof. Zander Myburg.

Finally, we would like to use the genome data of both avocado and *Phytophthora cinnamomi* for in depth bioinformatic analyses to aid in the identification of avocado defence targets and pathogen effectors. Three *P. cinnamomi* genomes are now available through the Joint Genomes Institute and we feel that it is not necessary to sequence another genome. The current amount of genome data is sufficient for our purposes.

Background

Avocado is an economically important fruit crop worldwide and belongs to the Magnoliidae clade, a basal lineage of the flowering plants. It is a member of the Lauraceae family that contains 2500-3000 species, mosly trees (Chanderbali *et al.*, 2008). *Persea americana* comprises three botanical races, Mexican, West-Indian and Guatemalan. The centre of genetic diversity and domestication is in Mesoamerica and this diversity can be exploited to identify genes involved in important traits.

Avocado has a long vegetative period of about 6 to 8 years, and the difficulties of doing directed crosses complicates avocado breeding. Knowledge of the avocado genome and related genetic diversity will provide important information and a molecular toolbox to facilitate the establishment of more effective breeding programs. Opportunities also arise to study the genetics underpinning complex traits such as disease tolerance and tolerance to abiotic stresses.

Commercial avocado production is based on grafting fruit cultivars onto rootstocks, usually of Mexican or Guatemalan backgrounds or hybrids. Several promising rootstocks have been selected over the years by researchers at WTS but the parentage of these superior rootstocks remains unresolved, with only the identity of the mother being known. Therefore, directed crossing cannot be incorporated into the current selection process.

In avocado, researchers have started exploring the viability and use of molecular marker techniques for crop improvement. In the early 1990's DNA fingerprinting was performed on avocado using mini- and microsatellite probes (Lavi *et al.*, 1991) and in 1997 a genetic linkage map for avocado was constructed (Sharon *et al.*, 1997).

In the study by Lavi *et al.* (1991) it was demonstrated that DNA fingerprinting (DFP) on using mini- and microsatellite probes resulted in each cultivar being identified by its DFP pattern and each race also being characterised by a specific DFP pattern. The study suggests that the DNA fingerprints will be useful for breeding for economically important traits, allowing the rapid screening of seedlings in avocado breeding programmes.

The genetic linkage map constructed by Sharon *et al.* (1997) for avocado was aimed at the use of the linkage map to aid in QTL (quantitative trait loci) identification for the breeding of crops with desired traits. Fifty polymorphic simple sequence repeat (SSR) loci, 17 random amplified polymorphic DNA (RAPD) and 23 mini-satellite DNA Fingerprint (DFP) bands were used to construct the avocado genetic map. From the 90 SSR, RAPD and DFP markers 56 markers remained unlinked. It was concluded that for avocado and other sub-tropical fruit trees SSR markers would be the most desirable marker system for the purpose of linkage map development since they were polymorphic, abundant and locus-specific and overcame the shortcomings of RAPD markers that were not that informative and DFP markers that are not locus-specific (Sharon *et al.*, 1997). Another study involved in the development of microsatellite markers for avocado suggested that the use of a single cultivar for the development of a microsatellite library is too narrow for broad spectrum marker development due to the genetic diversity present in avocado (Ashworth *et al.*, 2004)

However, all the current molecular markers and data are representative of fruit and flowering characteristics, with very little, if any, work on the molecular aspects of avocado rootstocks. The absence of a rootstock breeding population where phenotypic data is collected to combine with the genotypic data remains a constraint.

In the last 8 years genomic and transcriptomic data has been generated in our research programme and by researchers in Mexico. With the completion of the Mexican avocado genome and the Dusa transcriptome many research hypotheses can now be investigated. Our relationship with the Mexican group has secured access to the genome data that is otherwise not publically available. In our opinion the assembly is not of sufficient quality to re-assemble a new genome such as R0.06. The current genome has been assembled into 3,000 scaffolds indicating that the assembly is not complete. We can however still utilize this data for gene mining.

As with any other food and fruit crop that is to be produced sustainably in the future it is essential that scientific research of avocado remain at the forefront of research trends around the world. We propose in this document a strategy that will contribute to the generation of important genomic data of the WTS rootstock R0.06 as well as to develop molecular markers that will aid in resolving parentage in avocado and to genotype important avocado germplasm.

Collaborators:

Members of the AvoGenome consortium

Dr. Zelda van Rooyen, WTS

Prof. Zander Myburg, Dept. of Genetics, UP

Workpackage 1: Sequencing a pure race avocado genome, followed with re-sequencing of two rootstocks.

Time frame: January 2017 – December 2018

The avocado genome sequencing/re-sequencing project will provide significant advances in research methodologies to aid in tree breeding and selection. Fundamental questions such as evolutionary biology and relationships between gene expression, molecular and biochemical pathways and physiological processes can be investigated by utilizing “omics” data. This will allow us to understand complex responses in avocado to biotic and abiotic stress.

A pure race avocado (the homozygocity was determined using 5050 SNP markers) will be sequenced via hybrid approaches (60X PacBio Sequel and corrected with 100X Illumina) for a high quality *de novo* assembly. The R0.06 and a UCR-PRR tolerant genome will be re-sequenced on the Illumina Hi-Seq platform. Two pair-end libraries with different insert sizes will be sequenced with the expectation of obtaining >30x coverage of the 928 Mb avocado genome. The data will be assembled against the high quality pure race genome using standard bioinformatics tools. Gene prediction algorithms in conjunction with expression data will be used to annotate the genome.

Phase 1. Collect and prepare plant material for sequencing

Phase 2. Sequencing, *de novo* assembly and annotation of the pure race avocado genome sequence

Phase 3. Resequencing and assembly of R0.06 and a UCR rootstock using the genome assembled in phase 2.

Work package 2: Establishing a half-sib population for avocado with R0.06 as the parent and collecting phenotypic data of this population.

Time frame: January 2018 – Ongoing

The objective is to develop a structured half-sib population using R0.06 as the mother to assess important traits. Phenotypic data (traits selected by WTS) will be collected over several seasons. This data is important to link with the genotypic data. This population should be duplicated in at least one other location in South Africa. The trial will be designed with inputs from Prof. Myburg and Dr. David Kuhn and will be maintained by WTS.

Phase 1. Identify a suitable ungrafted R0.06 tree and a suitable piece of land to establish the half-sib population trial. (To be discussed with WTS)

Phase 2. Harvest and germinate at least 200 R0.06 fruit in the WTS nursery.

Phase 3. Collect phenotypic data (PRR tolerance, leaf morphology, stem diameter, presence of lenticels on stem etc) of the half-sib plants

Phase 4. Establish the population in the field and continue to collect phenotypic data.

Work package 3: SNP Genotyping of avocado rootstocks.

Time frame: January 2017– December 2020

The development of SNP markers has now become rather cost-effective due to Next Generation Sequencing. SNPs are a high-throughput method to genotype thousands of trees and have successfully been used in several tree crops like eucalyptus and also avocado. SNP markers have been previously developed for fruit cultivars of avocado and are used to genotype these trees. We will use the current SNP chip developed for fruit cultivars (USDA, Florida) to genotype the avocado rootstocks.

Phase 1. Collect material from important avocado germplasm (as determined by WTS).

Phase 2. DNA extraction and quality control

Phase 3. SNP analysis using the SNP chip designed for avocado fruit cultivars (this will be outsourced to Dr. David Kuhn, USDA, Florida).

Work package 4: Bioinformatic analyses of the avocado and *Phytophthora cinnamomi* genomes and transcriptomes to aid in the identification of defence mechanisms and pathogen virulence and pathogenicity factors.

Time frame: January 2017 – December 2020

The interaction between host and parasite has profound consequences for biology, thus, understanding phenotypic and genotypic processes controlling these interactions is crucial. High Throughput Sequencing (HTS) technologies have provided a cost effective alternative by which host–parasite interacting mechanisms can be unravelled. The availability of genomic data for both the avocado and *Phytophthora cinnamomi* genomes will enable analysis of host–parasite interactions at a genomic and transcriptomic level.

Phase 1. Comparison of the Dusa transcriptome and Mexican genome to determine homology.

Phase 2. RNA seq analysis of a tolerant avocado infected with *P. cinnamomi* over time-course. (This data has been generated as part of a MSc project in the Avocado Research Group)

New projects will arise as we advance in the proposed projects.

Conclusion

The project that we propose has been designed with the inputs from several avocado researchers as well as Prof. Myburg who is the PI of the Forest Molecular Genetics group at UP. We aim to address the unresolved questions pertaining to avocado parentage and to further generate genome sequencing for the superior rootstock R0.06. Through this initiative we also provide opportunities for student training and capacity building in genomics and bioinformatics.

References

Ashworth, V.E.T.M., Kobayashi, M.C., De La Cruz M. & Clegg M.T. (2004). Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. *Scientia Horticulturae* 101, 255-267.

Borrone, J., Brown, J., Tondo, C., Mauro-Herrera, M., Kuhn, D., Violi, H., Sautter, R. & Schnell, R. (2009). An EST-SSR based linkage map for *Persea americana* Mill. (avocado). *Tree Genetics and Genomes*. 10.1007/s11295-009-0208-y.

Clegg, M.T., Ashworth, V. & Chen, H. (2007). Assessing the genetic determination of valuable avocado traits using microsatellite (SSR) markers and quantitative trait loci (QTL) analysis. Production Research Report, California avocado commission.

Clegg, M.T., Ashworth, V., Chen, H. & Xu, S. (2006). Assessing the genetic determination of valuable avocado traits using microsatellite (SSR) markers and quantitative trait loci (QTL) analysis. Proceedings of the California Avocado Research Symposium. Pp. 57-63.

Lavi, U., Hillel, J., Vainstein, A., Lahav, E. & Sharon, D. (1991). Application of DNA fingerprints for identification and genetic analysis of avocado. *Journal of the American Society for Horticultural Science* 116, 1078-1081.

Sharon, D., Cregan, P.B., Mhameed, S., Kusharska, M., Hillel, J., Lahav, E. & Lavi, U. (1997). An integrated genetic linkage map of avocado. *Theoretical and Applied Genetics* 95, 911-921.

Progress Report – May 2017

Avocado Sequencing Project

1. The development of SNP markers for the diversity analysis, identification and parentage reconstruction in avocado cultivars – Juanita Hanneman
    - The leaves of 91 avocado rootstocks were collected and submitted to the USDA in Florida for DNA extraction and genotyping
    - SNP genotyping has been completed and we have received the data
    - Data is being analyzed under the supervision of Dr. Sarah Mwangi
2. Sequencing and assembly of a “homozygous” avocado in collaboration with the AvoGenome Consortium – Dr. Sarah Mwangi
    - Dr. Manosalva has identified the avocado material that will be sequenced.
    - DNA will be extracted by Dr. David Kuhn at the USDA in Florida.
    - We are awaiting the funding from all parties involved to be released and then the sequencing will commence