Hort Innovation

Final Report

Establishing an open-source platform for unravelling the genetics of Macadamia: integration of linkage and genome maps

Project leader:

Dr Catherine Nock

Delivery partner:

Southern Cross University

Project code:

MC15008

Project:

Establishing an open-source platform for unravelling the genetics of Macadamia: integration of linkage and genome maps MC15008

Disclaimer:

Funding statement:

This project has been funded by Hort Innovation with co-investment from Australian Macadamia Society and contributions from the Australian Government. Hort Innovation is the grower-owned, not-for-profit research and development corporation for Australian horticulture.

Publishing details:

ISBN 978 0 7341 4611 3

Published and distributed by: Hort Innovation

Level 7
141 Walker Street
North Sydney NSW 2060

Telephone: (02) 8295 2300

www.horticulture.com.au

Content

Summary

Macadamia is the first Australian native plant to become a major international food crop and is an important contributor to Australian horticultural industry. Genomic resources are needed to understand the genetic basis of important traits and to accelerate selective breeding of new improved varieties for the macadamia orchards of the future. Until recently however, sequence data for macadamia were limited compared to other crops.

The genome of *Macadamia integrifolia* cultivar HAES 741 is being sequenced and assembled at Southern Cross University. The objective of the project MC15008 *'Establishing an open-source platform for unravelling the genetics of Macadamia: integration of linkage and genome maps'* was to develop a genetic linkage map for macadamia to facilitate genome assembly and provide a platform to identify and locate the genes controlling important traits such as pest and disease resistance, quality and yield. This project is particularly timely given the reliance of AS17000 *National Tree Genomics* program on genomic resources for macadamia. It also provides valuable resource for MC14000 *Macadamia second generation breeding and conservation* and MC19000 *National macadamia breeding and evaluation program*. Early in this project, the first draft macadamia genome assembly was released (Nock et al. 2016). In 2017, an improved v2 assembly was developed using additional Illumina short-read and PacBio long-read sequence data (4,098 scaffolds, 413 kb N50).

Mapping populations with HAES 741 parentage were developed specifically for this project and DNA paternity analysis was conducted to confirm the parentage of progeny. This identified self-fertilisation and pollen contamination in populations developed using controlled crosses. In addition, relatively high levels of self-fertilisation were detected in some cultivars including HAES 741, 660, 791 and 842 (Langdon et al. 2019). This was unexpected given that the open-pollinated seed parents in this study were located in a regional variety trial plot. Further research is needed to determine the contribution of self-pollination to yield in macadamia orchards. There is also a timely opportunity to work with the industry to undertake screening of cultivar distribution at landscape scale to understand how 'incompatibility groups' may affect yield.

A series of nine genetic linkage maps was constructed for cultivars HAES 741, HV A4 and HV A268. Each map comprised 14 linkage groups, consistent with the haploid chromosome number for macadamia. The use of multiple populations and maps increased marker density overall and successfully anchored 70% of the v2 genome assembly into 14 pseudo-chromosomes (Langdon et al. 2020, in press). Following annotation of the anchored assembly, 34,274 genes protein-coding genes were predicted. The genome assemblies, linkage maps, gene predictions and database established through this project provide a platform for macadamia genetics. These resources are expected to underpin future breeding and support the long-term conservation of natural populations.

Keywords

macadamia; genome; genetic markers; DNA paternity testing; linkage map; database; breeding; conservation

Introduction

The application of genomic data is transforming selective breeding & agricultural productivity. Genomic technologies have rapidly improved over the past decade and the genomes of most important crop species have been sequenced (Varshney et al. 2014, Chen et al. 2018). Macadamia is the only Australian native plant to become a major international food crop and is an important contributor to Australian horticultural industry. Genomic resources are needed to understand the genetic basis of important traits and to accelerate selective breeding of new improved varieties.

The first macadamia genomic data were collected at Southern Cross University (SCU) and a collaboration with the University of Queensland (UQ) was established to sequence the genome of one of the most widely grown cultivars in Australia, *Macadamia integrifolia*, HAES 741. The project MC15008 *Unravelling the genetics of macadamia: integration of linkage and genome maps* was established to:

- Develop markers to enable the evaluation of industry, breeding and wild germplasm for varietal identification, parentage and genetic diversity, and
- Create a high-density linkage map for genome assembly and to facilitate gene discovery, and marker-assisted selection in macadamia breeding.
- Provide a genomic platform for improved understanding of the genetics behind key traits for conservation and crop development including local adaptation, pest and disease resistance, and productivity.

The project is aligned with priorities outlined in the Macadamia Strategic Investment Plan 2017-2021. In particular, it contributes to Outcome 2, Improved production systems covering plant breeding, intensive orchards and novel technologies and Outcome 3, Improved capacity to lead and support current and future industry needs. It links to other Hort Innovation funded projects including MC14000 *Macadamia Second Generation Breeding and Conservation*, MC15007 *Still wild about macadamias – conserving a national icon*, and AS17000 *National Tree Genomics Program*.

Methodology

Genome Assembly

The first draft genome assembly for macadamia, constructed from Illumina short-read and mate pair data was released during the first year of this project https://www.ncbi.nlm.nih.gov/assembly/GCA_900087525.1. Following annotation, ~35,000 protein-coding genes were predicted. Of these over 90% were expressed in at least one of the leaf, shoot or flower tissues examined (Nock et al. 2016, https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3272-3). During the second year of the project, an improved assembly (v2) with increased sequence scaffold lengths (4,416 scaffolds, 414 kb N50) was constructed using additional Illumina short-read and PacBio long-read sequence data. https://www.ncbi.nlm.nih.gov/assembly/GCA_900631585.1

Mapping Populations

Bi-parental, self and open-pollinated populations were developed specifically for this project by the Australian macadamia breeding program team and Southern Cross University. The parentage of each population included HAES 741, the same variety used for genome assembly (Figure 1). *M. integrifolia* - *M. tetraphylla* hybrid cultivars from Hidden Valley Plantations HV A268 and HV A4 were also selected as parents because they differ in a range of economically important traits.

Figure 1. Macadamia mapping populations used for genetic linkage mapping, left. PhD student Kirsty Langdon with progeny from the HAES 741 open-pollinated population at Southern Cross University, right.

Paternity Analysis

In order to confirm the parentage of the progeny used in this project, genetic markers were developed and optimised for macadamia. DNA paternity testing using 11 microsatellite markers was conducted on open-pollinated seedlings from cultivar HAES 741, and mature kernel from eight different cultivars, seven of which have been used in previous self and cross pollination studies. The study was based in a varietal trial plot containing 40 different cultivars. Paternity was assigned to 92% of seedlings and 87% of kernel assessed. For four cultivars, selfing rates of 20-40% were detected. Estimates of relatedness between cultivars indicate a first-degree relationship between the two most widely grown cultivars in Australia, HAES 741 and HAES 344, at the level of full-sib or parent-offspring.

Insight into the breeding system of macadamia is likely to have major implications for landscape and orchard-scale management. These results point to significant differences in the capacity for self-fertilisation between cultivars and provide new molecular evidence of pollen flow under orchard conditions (Langdon et al. 2019, https://link.springer.com/article/10.1007/s11295-019-1336-7). This research forms a chapter of Kirsty Langdon's PhD thesis and was published in the journal *Tree Genetics and Genomes*. An article summarising the results was submitted to the Australian Macadamia Society and will be published in the Autumn 2020 edition of AMS New Bulletin (Appendix 1).

Map construction

Leaf material was collected from over 400 seedlings and high-quality DNA (>50 ng/µl) was extracted following methods described in Langdon et al. (2019). Genetic marker data were generated by Diversity Array Technologies Pty. Ltd. using DArTseq methodology previously developed and described for macadamia (Kilian et al. 2012, Alam et al. 2018). Unfiltered data from 6,071 co-dominant SNP markers and 10,071 dominant markers were generated for each of the progeny and parental cultivars from the four populations. Following quality control and filtering 901 (741 self), 3,805 (741 x A4), 3,533 (741 x A268), and 2,650 (A268 x 741) informative markers were retained for genetic linkage mapping.

Seven genetic linkage maps were developed in JoinMap v5.0. These included HAES 741 self, maternal and paternal maps, HV A268 maternal and paternal, and HV A4 paternal maps. In addition, integrated maps with higher marker density for HAES 741 and HV A268 were developed. The amalgamation of information from multiple populations and linkage maps was successful in increasing marker density and detecting variation in recombination frequency between parental genotypes. Compared with a single-species bi-parental population, the use of four populations involving a common parent and two additional parents of diverse origin both increased the number of recombination events and informative markers.

Genome anchoring

Genome anchoring was performed using ALLMAPS. This is a method that configures sequence scaffolds to maximise the collinearity of markers located on physical and genetic linkage maps (Tang et al. 2015). Multiple linkage maps can be included in a single run and scaffolds are ordered and orientated to generate sequences that are concordant with the input linkage maps. Figure 5 illustrates marker density across the 14 macadamia chromosomes relative the integrated HAES 741 and HV A268 integrated maps. The generation of a series of linkage maps from multiple mapping populations provided suitable marker density to anchor 70% of the v2 genome assembly to produce a chromosome-scale assembly. This research forms a chapter of Kirsty Langdon's PhD thesis and has been accepted for publication in *Scientific Reports* (Langdon et al. 2020, in press). The paper and associated data https://epubs.scu.edu.au/data_collections/48/ will be available open-source on publication.

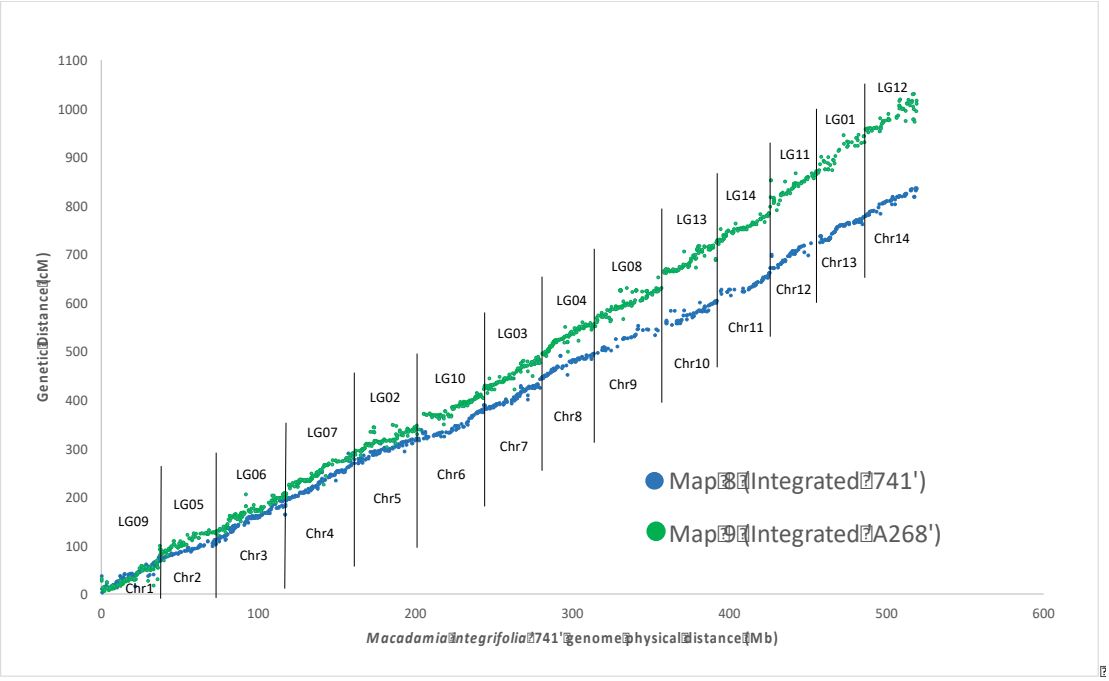

Figure 5. Marker positions in the integrated HAES 741 and HV A268 integrated linkage maps relative to the *M. integrifolia* HAES 741 genome. The X axis is physical distance in chromosome order. The Y axis is genetic distance in centimorgans (cM). Vertical lines indicate the boundaries between linkage groups (LG). and chromosomes (Chr).

Chromosome-scale assembly for *Macadamia integrifolia*, HAES 741

Following annotation of the anchored assembly, 34,274 genes were predicted. In comparison to the earlier assemblies, the genes were more complete and this assembly contains 91% of the expected gene content.

The genomic resources developed through this project have been deposited in public databases will be open-source and available for the benefit of the macadamia industry. The v1.1 and v2 assemblies have been released [NCBI BioProject PRJEB13765, https://www.ncbi.nlm.nih.gov/bioproject/339709]. For example, the v2 assembly https://www.ncbi.nlm.nih.gov/assembly/GCA_900631585.1 was used by Katie O'Connor during her PhD research at University of Queensland as part of the Australian breeding program to estimate linkage disequilibrium in a macadamia breeding population and for genome wide association studies to identify regions of the genome associated with nut weight, kernel recovery and other important traits (O'Connor et al. 2019, O'Connor 2019, O'Connor et al 2020). This Whole Genome Shotgun project has been deposited at NCBI GenBank [WGS accession JAAEEG000000000, BioProject PRJNA593881] and will be released pending publication of a data

note (Nock et al., https://epubs.scu.edu.au/data_collections/49/). These genomic resources are expected to underpin future breeding and support the long-term conservation of natural populations.

Database development

CropStore$_{DB}$ is a curated open-source database for the management of population, genetic map, QTL and trait measurement data (Figure 1).
During this project, we have:

- collated data for plants, populations, maps and sequence-tagged loci.
- established a series of nomenclature data standards for these entities to facilitate current and future data curation, based on discussion amongst the project participants, the broader Australian consortium including colleagues at QAAFI.
- entered macadamia data for release into the public domain under the CC BY 4.0 Creative Commons license via http://www.cropstoredb.org/cs_macadamia.html. This includes:
    - a summary of data records for various entities
    - a web-browser accessible graphic user interface to query and navigate available datasets
    - a Windows PC executable app is available upon request.

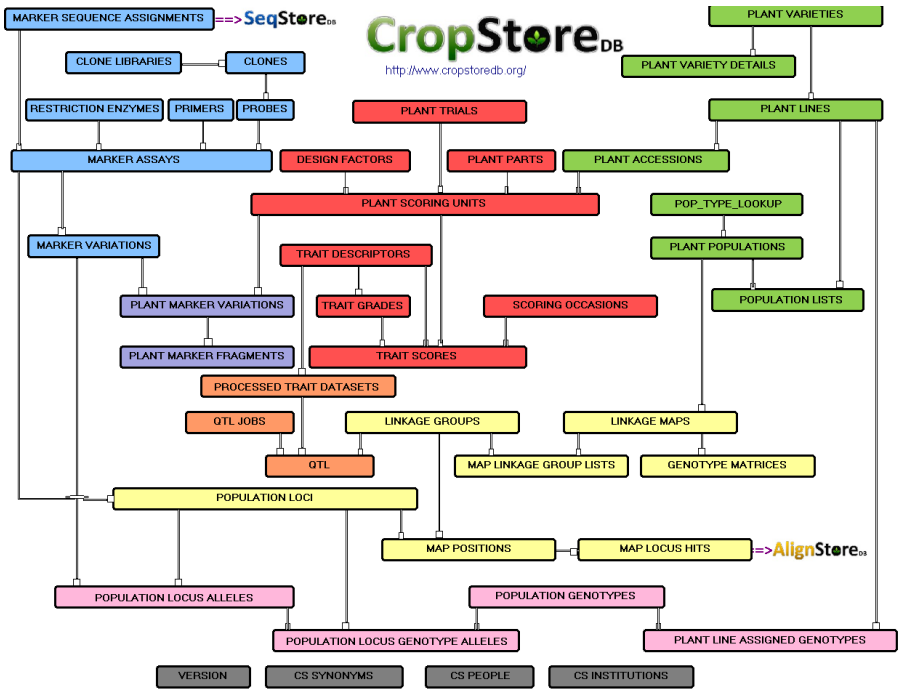

Figure 1. Schema for the CropStoreDB database, v 9.03

The data curated and released from this project represent a stable open-source platform for long-term storage and access to genetic and geolocation data for macadamia germplasm. It has considerable strategic value, and with adequate future investment has capacity to underpin global coordinated efforts to harmonise macadamia genetics, genomics and pre-breeding R&D. However, this requires further investment in order to maximise the added value to the macadamia genepool. Ongoing discussion with the Macadamia Conservation Trust and Australian macadamia breeding program team throughout the course of this project is expected to enable long-term secure storage and access to genetic and geolocation data for macadamia accessions and populations.

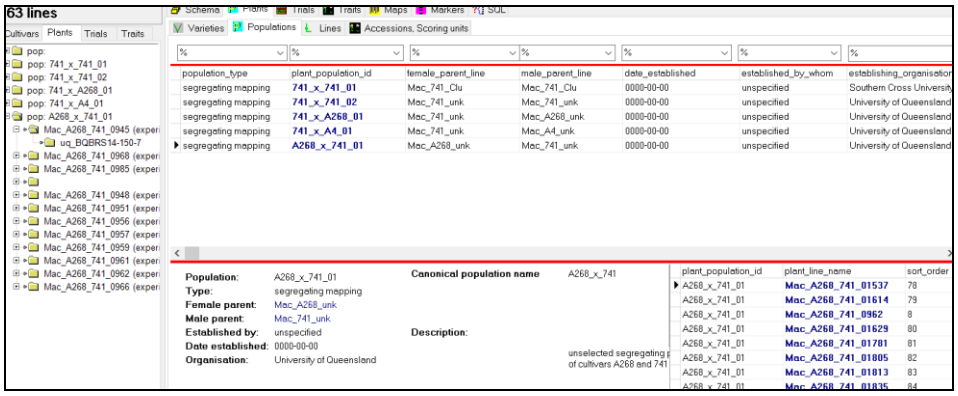

Figure 2. CropStoreDB interface with population, plant line, plant accession and drill-down data

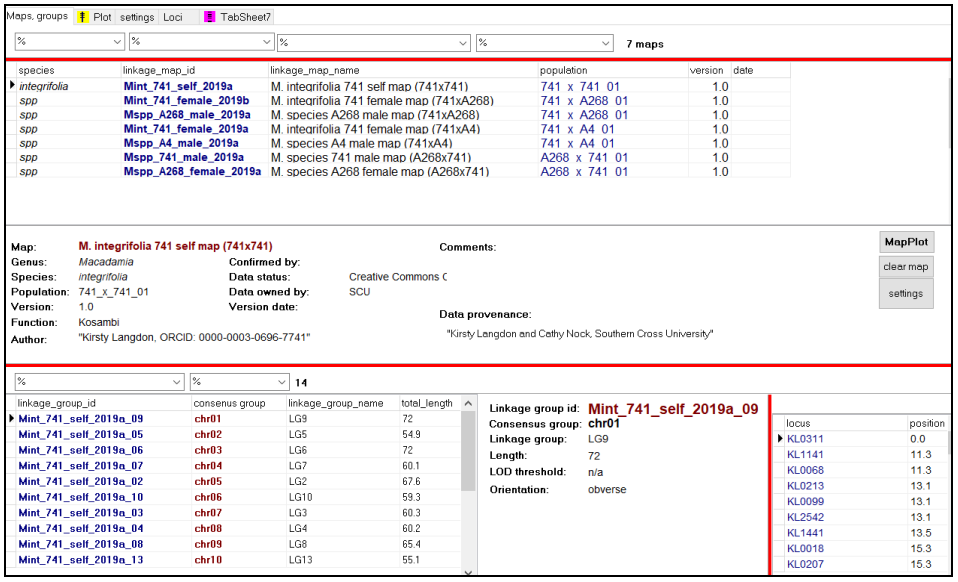

Figure 3. CropStoreDB interface with seven genetic linkage maps loaded

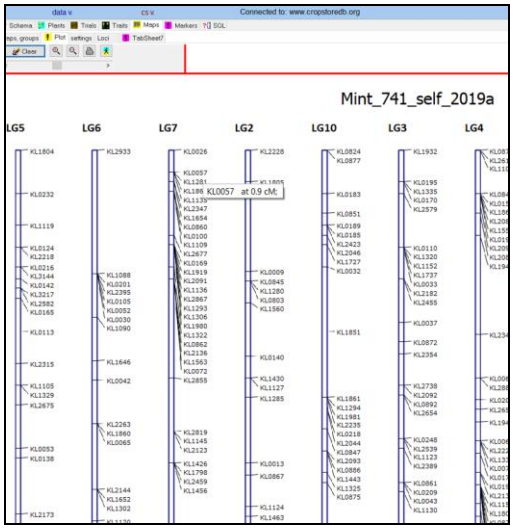

Figure 4. CropStoreDB interface with graphic display of linkage groups and active links from each marker locus

Outputs

High-density genetic linkage maps for the 14 chromosomes of macadamia

One of the main expected outputs of this project was a genetic linkage map for macadamia. During her PhD research, Kirsty Langdon constructed seven high-density genetic linkage maps and integrated maps for cultivars HAES 741 and HV A268. This research forms a chapter of Kirsty's PhD thesis 'Maximising recombination across macadamia populations to generate linkage maps for genome anchoring'. It has been accepted for publication in *Scientific Reports* (Langdon et al. 2020, in press). The associated data have been deposited and will be available on publication https://epubs.scu.edu.au/data_collections/48/.

An open-source platform to identify markers and genes linked to important crop traits

- Release of the first genome assemblies for macadamia
  v1.1
  - Nock et al. 2016, https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3272-3
  - Genome assembly in open-source NCBI database https://www.ncbi.nlm.nih.gov/bioproject/339709
  - Associated computational data is available at https://epubs.scu.edu.au/data_collections/46/

  v2
  - Genome assembly in open-source NCBI database https://www.ncbi.nlm.nih.gov/bioproject/339709
  - O'Connor et al. 2019, https://link.springer.com/article/10.1007/s11295-019-1331-z
  - O'Connor et al 2019, https://bmcgenomics.biomedcentral.com/track/pdf/10.1186/s12864-020-6575-3

- Development of a chromosome-scale genome assembly for macadamia
  v3
  - Genome assembly deposited in NCBI Genbank database [BioProject PRJNA593881].
  - Langdon et al. 2020 in press, *Scientific Reports*
  - Nock et al. (in preparation, https://epubs.scu.edu.au/data_collections/49/)

An integrated database for macadamia genetic and provenance data

During this project, we have developed an open-source database framework in CropStore$_{DB}$ for the long-term storage and management of population, genetic map, QTL and trait measurement data for macadamia. Data relating to the plants, populations, maps used in this project have been deposited in CropStore$_{DB}$ http://www.cropstoredb.org/cs_macadamia.html.

PhD Thesis

Kirsty Langdon commenced her PhD research at Southern Cross University under a scholarship provided through this project on 22 March 2016. Her candidature was confirmed on 15 March 2017 and in-candidature review was completed on 3 October 2018. A draft thesis entitled 'Unravelling the genetics of macadamia: integration of linkage and genome maps' was submitted to PhD supervisors Dr Catherine Nock and Professor Graham King on 6 December 2019. The thesis includes a chapter 'DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia' that was accepted for publication in *Tree Genetics and Genomes* (Langdon et al. 2019, https://link.springer.com/article/10.1007/s11295-019-1336-7). Another chapter 'Maximising recombination across macadamia populations to generate linkage maps for genome anchoring' has been accepted for publication in *Scientific Reports* (Langdon et al. 2020, in press). A third chapter 'Exploring the complexities of linkage mapping' will be submitted for publication following examination of the thesis. The thesis is being finalized and will be submitted for examination in April 2020, within the 4-year period of PhD candidature.

Refereed Scientific Publications

Nine scientific publications related to this project have been published (see Refereed scientific publications). Another is in preparation.

AMS Newsletter Articles

Langdon, K.S. et al. Cracking the macadamia genome. Summer 2016, page 77-78.

Nock C.J. The wild relatives of macadamia. Spring 2017, page 73-74.

Langdon K.S. et al. Paternity testing in the orchard. Who's fathering the nuts? Spring 2018, page 71-72.

Langdon, K.S. and **Nock C.J.** Paternity testing in the orchard. Self-pollination and cross-pollination. Autumn 2020, page 70-73.

Conference Presentations

Nock, C.J., Langdon, K.S., Baten, A., Mauleon, R., Topp, B., Hardner, C.M., O'Connor, K. and **King, G.J.** (2019) Genomics and the macadamia orchard of the future. TropAg 2019, Brisbane. November 11-13.

Furtado, A., **Nock, C.J.**, Coin, L., Murigneux, V., Topp, B., Henry, R.J. (2019) Advances in macadamia genomics. TropAg 2019, Brisbane. November 11-13.

Nock, C.J. (2018) Towards a reference genome for macadamia. 8th International Macadamia Symposium, Lincang, China. October 14-20

Langdon, K.S. (2018) A genetic platform for macadamia. 8th International Macadamia Symposium, Lincang, China. October 14-20

Alam, M., Neal, J., O'Connor, K., **Nock C.**, **Langdon, K**. and Topp, B. (2017) Morphological and molecular characterization of self-pollinated vs naturally out-crossed progeny of macadamia: a case study of 'HAES 741'. International Tropical Agriculture Conference, Brisbane, November 20-22nd.

Lee A., Rost J., Muralidharan , Campbell D., Mehr S., **Nock C.,** Baumert J., Taylor S., (2016) 7S and 11S globulins are likely to be allergens in macadamia nuts. Food Allergy and Anaphylaxis Meeting, Rome, October 13-15th.

Nock, C.J., Baten, A., Langdon, K., Barkla, B., Hardner, C., Topp, B., Furtado, A., Henry, R. and **King, G.J** (2017) The Macadamia Genome Project. International Macadamia Research Symposium, Hawaii, September 12-14.

Peace, C., Hardner, C.M. and **Nock, C.J.** (2017) Genetic origins of macadamia cultivars: what we know so far. International Macadamia Research Symposium, Hawaii, September 12-14.

Hardner C.M., **Nock C.J.,** Batley J., Ahmad Termizi A., Hayashi S., Montenegro J.D. and Edwards D. (2017) Domestication of Macadamia in Hawai'i. International Macadamia Research Symposium, Hawaii, September 12-14.

Alam, M., Hardner, C., **Nock, C.**, O'Connor, K. and Topp, B. (2017) Historical and molecular evidence of genetic identity of HAES741 and HAES660 macadamia cultivars. International Macadamia Research Symposium, Hawaii, September 12-14.

Nock, C.J. (2017) Conserving the wild relatives of macadamia. Rainforests of Subtropical Australia Conference, Robina Qld, March 23-24.

Langdon, K.S., Baten, A., Topp B., Hardner, C.M., King, G.J. and **Nock, C.J.**, (2017) Unravelling the genetics of macadamia: integration of linkage and genome maps. Annual Conference of the Genetic Society of Australasia Conference, Dunedin, New Zealand July 2-6th.

Rost, J., Muralidharan, S., Campbell, D., Merhr, S., **Nock, C.J.** and Lee A.L. (2017) Purification and characterisation of IgE-Binding proteins as putative macadamia nut allergens. FAMS 2017 Food Allergen Management Symposium, Sydney May 21-24th.

Nock C.J., Baten A., Barkla B.J., Furtado A., Henry R.J. and **King G.J.** (2016) Cracking the nut: genome and transcriptome sequencing of *Macadamia integrifolia* (Proteaceae). Society for Molecular Biology and Evolution Conference 2016, Gold Coast July 4-7th.

Outcomes

Genome sequences can aid in species and variety identification, finding the genes and gene combinations underlying valuable traits for crop production and the conservation of wild populations. The genomes of many crop species have been sequenced and are being used to accelerate selective breeding and improve productivity.

This investment has contributed to the development of the first genomic resources for macadamia. MC15008 *'Establishing an open-source platform for unravelling the genetics of Macadamia: integration of linkage and genome maps'* was a voluntary contribution (VC) funded project and its outcomes provide a platform for improved understanding of macadamia genetics.

This platform includes genome assemblies released in 2016 and 2018 that have been important in the identification of candidate genes (Nock et al. 2016), estimating linkage disequilibrium among genetic markers used to characterise macadamia breeding populations (O'Connor et al. 2019) and locating markers associated with important nut characteristics (O'Connor 2019, O'Connor et al. 2020). The assemblies have also been used to identify geminiviruses in the genome (Akinsanmi 2018) and seed storage proteins and putative allergens (Rost et al. 2017).

The genetic linkage maps developed through this project provide information on recombination in macadamia that will be important for selective breeding programs aiming to achieve favourable combinations of genes and traits in new macadamia varieties. The maps were used to order and orient the genome sequence scaffolds to produce a chromosome-scale assembly with DNA sequences representing the 14 chromosomes of macadamia (Langdon et al. 2020, in press). Annotation of this assembly identified 34,274 protein-coding genes representing 91% of the expected gene content of macadamia. A database framework was also established for the management of macadamia population, genetic map, QTL and trait measurement data. Data from plants, populations and maps developed through the project have been entered into the curated, open-source database CropStore$_{DB}$. These genomic resources are expected to support future breeding and the long-term conservation of natural populations.

In order to confirm the parentage of the mapping populations used in this study, genetic markers and methodology for DNA paternity testing were developed and optimised. This provided new information on pollination under orchard conditions and evidence for significant differences in self-fertility between macadamia cultivars. The SSR markers developed at SCU have been published (Nock et al. 2014, Langdon et al. 2019) and are being utilized in other laboratories in Australia (O'Connor et al. 2015, Trueman et al. 2019) and internationally.

The peer-reviewed scientific publications and conference presentations delivered through this project have improved the profile of macadamia as an Australian native food crop and as a model plant for the large, ancient family Proteaceae. Project results have been communicated to the industry at Australian and International macadamia industry conferences. Collaboration and regular meetings between the MC15008, MC13015 *Still wild about macadamias - conserving a national icon* and MC14000 *Macadamia second generation breeding and conservation* project teams have led to joint publications and plans to continue the collaboration through future research.

Monitoring and evaluation

Regular reviews and meetings of the project team and collaborators were untaken to ensure timely completion of milestone requirements. DNA paternity testing was conducted on the open-pollinated, self and bi-parental mapping populations used in this project. This identified self-fertilisation and pollen contamination with between 6% and 43% of progeny from populations developed for this project from controlled crosses not from the expected cross. A 6-month extension to the project was approved to enable reanalysis and reconstruction of genetic linkage maps. The accuracy and utility of the revised maps was improved using the corrected populations.

Recommendations

1. Through this project, markers and methodology were developed for DNA paternity analysis in macadamia. This provided evidence of the importance of paternity testing to confirm the parentage of controlled crosses and to understand pollination under orchard conditions. Further research is needed to determine the contribution of self-pollination to yield in macadamia orchards. There is also a timely opportunity to work with the industry to undertake screening of cultivar distribution in plantations at landscape scale to understand how 'incompatibility groups' may affect yield.

2. The database developed through this project provides an open-source platform for long-term storage and access to genetic and geolocation data for macadamia germplasm. Further investment will be required to maximise its value to the macadamia genepool. Ongoing discussion with the Macadamia Conservation Trust and Australian macadamia breeding program team throughout the course of this project is expected to enable long-term secure storage and access to genetic and geolocation data for macadamia accessions and populations.
3. This project provides the first genome assemblies for macadamia, representing an important resource for future research to understand the genetic basis of important traits involved in adaptation, yield and pest and disease resistance. Further funding and support will be required to finish the *Macadamia integrifolia* HAES 741 reference genome sequence using new sequencing technologies that became available recently and were beyond the scope and budget of this project. There has been discussion with the AS17000 *National Tree Genomics Program* project team and this expected to proceed under a collaborative arrangement between UQ and SCU.

Refereed scientific publications

Journal article

Langdon, K.S., King, G.J., Baten, A., Mauleon, R., Bundock, P.C., Topp, B.L. and **Nock, C.J.** In press. Maximising recombination across macadamia populations to generate linkage maps for genome anchoring. *Scientific Reports*

O'Connor, K., Hayes, B. Hardner, C.**, Nock, C.J**., **Baten, A.,** Alam, M., Henry, R. and Topp, B.L. 2020. Genome-wide association studies for yield component traits in a macadamia breeding population. *BMC Genomics*, 21:199. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-6575-3

Nock, C.J., Hardner, C.M., Montenegro, J.D., Ahmad Termizi, A.A., Hayashi, S., Playford, J., Edwards, D., Batley, J., 2019. Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. *Frontiers in Plant Science* **10**, 334. https://www.frontiersin.org/articles/10.3389/fpls.2019.00334/full

Langdon, K.S., King, G.J. and Nock, C.J., 2019. DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. *Tree Genetics and Genomes* **15**, 29. https://link.springer.com/article/10.1007/s11295-019-1336-7

Alam, M.M., Hardner, C.M., **Nock, C.J.**, O'Connor, K.M. and Topp, B.L., 2019. Historical and molecular evidence of genetic identity of HAES 741 and HAES 660 macadamia cultivars. *Hort Science* **54(4)**, 616-620. https://journals.ashs.org/hortsci/abstract/journals/hortsci/54/4/article-p616.xml?rskey=dhCAYM

O'Connor, K., Killian, A., Hayes, B. Hardner, C., **Nock, C.J., Baten, A**., Alam, M. and Topp, B.L., 2019. Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silico DArT markers. *Tree Genetics and Genomes* **1**5, 24. https://link.springer.com/article/10.1007/s11295-019-1331-z

Nock, C.J., Baten, A., Barkla, B.J., Furtado, A., Henry, R.J. and **King, G.J.**, 2016. Genome and transcriptome sequencing confirms the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* **17(1)**, 937 https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3272-3

Book chapter

Topp, B.L., **Nock, C.J**., Hardner, C.M., Alam, M.M. and O'Connor, K.M., 2019. Macadamia (*Macadamia spp.*) Breeding. Chapter 7 In: Advances in Plant Breeding Strategies, Vol 4 Nuts and Beverage Crops. Springer https://link.springer.com/chapter/10.1007/978-3-030-23112-5_7

Conference proceedings

Rost, J., Muralidharan, S., Campbell, D., Merhr, S., **Nock, C.J.** and Lee A.L., 2016. Discovery of 7S and 11S globulins as putative allergens in macadamia nut by combining allergenomics and serum IgE binding. *Internal Medicine Journal* 46(S4):10.

References

Akinsanmi, F., 2018. Determining the extrent and causes of abnormal vertical growth. MC15011 Hort Innovation Project Final Report. Sydney, p. 143

Alam, M., Neal, J., O'Connor, K., Kilian, A. & Topp, B., 2018. Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PloS one* **13**, e0203465

Chen, F., Dong, W., Zhang, J., Guo, X., Chen, J., Wang, Z., et al., 2018. The sequenced angiosperm genomes and genome databases. *Frontiers in Plant Science*, **9**, 418.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C. and Aschenbrenner-Kilian, M., 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. In: Data production and analysis in population genomics. Humana Press, Totowa, NJ, pp. 67-89.

Langdon, K.S., King, G.J. and Nock, C.J., 2019. DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. *Tree Genetics and Genomes* **15**:29.

Langdon, K.S., King, G.J., Baten, A., Mauleon, R., Bundock, P.C., Topp, B.L. and **Nock, C.J.** in press. Maximising recombination across macadamia populations to generate linkage maps for genome anchoring. *Scientific Reports*

Nock, C.J., Baten, A., Barkla, B.J., Furtado, A., Henry, R.J. and King, G.J., 2016. Genome and transcriptome sequencing confirms the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* **17(1)**, 937

O'Connor, K. 2019. Selection strategies to improve yield in macadamia using component traits and genomics. PhD thesis, University of Queensland.

O'Connor, K., Killian, A., Hayes, B. Hardner, C., Nock, C.J., Baten, A., Alam, M. and Topp, B.L., 2019. Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDArT markers. *Tree Genetics and Genomes* **15**, 24

O'Connor, K., Hayes, B. Hardner, C.**, Nock, C.J**., Baten, A., Alam, M., Henry, R. and Topp, B.L. 2020. Genome-wide association studies for yield component traits in a macadamia breeding population. *BMC Genomics* 21:199

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J.C., Schnable, P.S., Lyons, E. and Lu, J., 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biology* **16**, 3.

Trueman, S., Kämper, W., Richards, T. Hossein Bai, S., Ogbourne, S., Nicols, J. and Wallace,. H. 2019. The results are in: most nuts come from cross-pollination. Australian Macadamia Society News Bulletin, Autumn edition, pp 72-73.

Varshney, R. K., Terauchi, R., & McCouch, S. R., 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biology*, **12(6)**, e1001883.

Intellectual property, commercialisation and confidentiality

No project IP, project outputs, commercialisation or confidentiality issues to report

Acknowledgements

The MC15008 project team included Catherine Nock, Abdul Baten, Graham King and PhD student Kirsty Langdon. We would like to thank David Knappick and the Knappick Foundation Ltd. for the generous contribution to the Macadamia Conservation Trust that made this research possible. We also thank our project collaborators Bruce Topp and Craig Hardner from University of Queensland for their contribution to this project, and Hort Innovation project managers Brenda Kranz and Vino Rajandran for their support and advice throughout.

The project would not have been possible without the support and involvement of:

- Macadamia Conservation Trust: Ian McConachie, Denise Bond, Brice Kaddatz, Paul O'Hare
- Australian Macadamia Society: Jolyon Burnett
- Southern Cross University: Graham King, Kirsty Langdon, Abdul Baten, Ramil Mauleon, Asuka Kawamata, Peter Bundock, Alicia Hidden, Carolyn Raymond, Eslam Ibrahim, Lee Kearney, Paul Kelly
- University of Queensland: Bruce Topp, Craig Hardner, Katie O'Connor, Mobashwer Alam, Agnelo Furtado and Robert Henry
- Healthy Land and Water, Queensland: Liz Gould
- Staff at QDAF Maroochy Horticultural Research Station and Bundaberg Research Station, and NSW DPI Centre for Tropical Horticulture
- Gray Plantations: Kim Wilson, Alex Yong

Appendices

Appendix 1. Langdon, K.S. and Nock, C.J. 2019. DNA paternity testing in the orchard: self- and cross-pollination. Australian Macadamia Society News Bulletin, Autumn 2020 pp.70-72.

Confidential Appendices (provided to Hort Innovation separately)

Appendix 2. Langdon, K.S., King, G.J. and Nock, C.J., 2019. DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. *Tree Genetics and Genomes* **15**:29.

Appendix 3. Langdon, K.S., King, G.J., Baten, A., Mauleon, R., Bundock, P.C., Topp, B.L. and Nock, C.J., In press. Maximising recombination across macadamia populations to generate linkage maps for genome anchoring. *Scientific Reports*

Hort Innovation | MACADAMIA FUND

Asuka Kawamata, Kirsty Langdon and Cathy Nock at work in the genetics laboratory at Southern Cross University.

DNA paternity testing in the orchard: self-pollination and cross-pollination

Kirsty Langdon and Cathy Nock, Southern Cross Plant Science, Southern Cross University

Project snapshot

Macadamia is preferentially outcrossing but it has long been presumed that some varieties may be self-fertile. As part of her PhD research, Kirsty Langdon paternity tested seedlings and nuts from a Regional Variety Trial plot. While most were cross-pollinated, selfing rates of 20–40% were found for some varieties (741, 660, 842, 791). The results indicate that there are significant differences between varieties in the capacity for self-fertilisation.

This research was undertaken in the Hort Innovation VC-funded project (MC15008): *Establishing an open-source platform for unravelling the genetics of macadamia: integration of linkage and genome maps.*

Paternity testing uses DNA and associated genetic markers to identify the father in situations where the mother is known. In both humans and agriculture, this often involves genetic markers called microsatellites. Dr Cathy Nock at Southern Cross University has developed macadamia-specific microsatellite markers and a marker profile database of reference genotypes for the most widely grown macadamia varieties. This toolkit is being used for DNA authentication of recently released varieties and for DNA paternity testing to find the pollen parent of nuts, and has great potential for understanding pollen movement in and between orchards.

The aim of Kirsty's PhD project, supervised by Dr Nock and Professor Graham King, was to develop genetic linkage maps for genome anchoring as part of the Australian effort to sequence the macadamia genome. To do this, a range of mapping populations were developed. One of these was an open-pollinated population from the same 741 tree that had been used for genome sequencing.

In April, over two consecutive years, over 120 seed were collected from the 741 tree in a regional varietal plot in northern NSW and germinated. Before the seedlings could be used for mapping, DNA paternity analysis was undertaken to determine their pollen parent

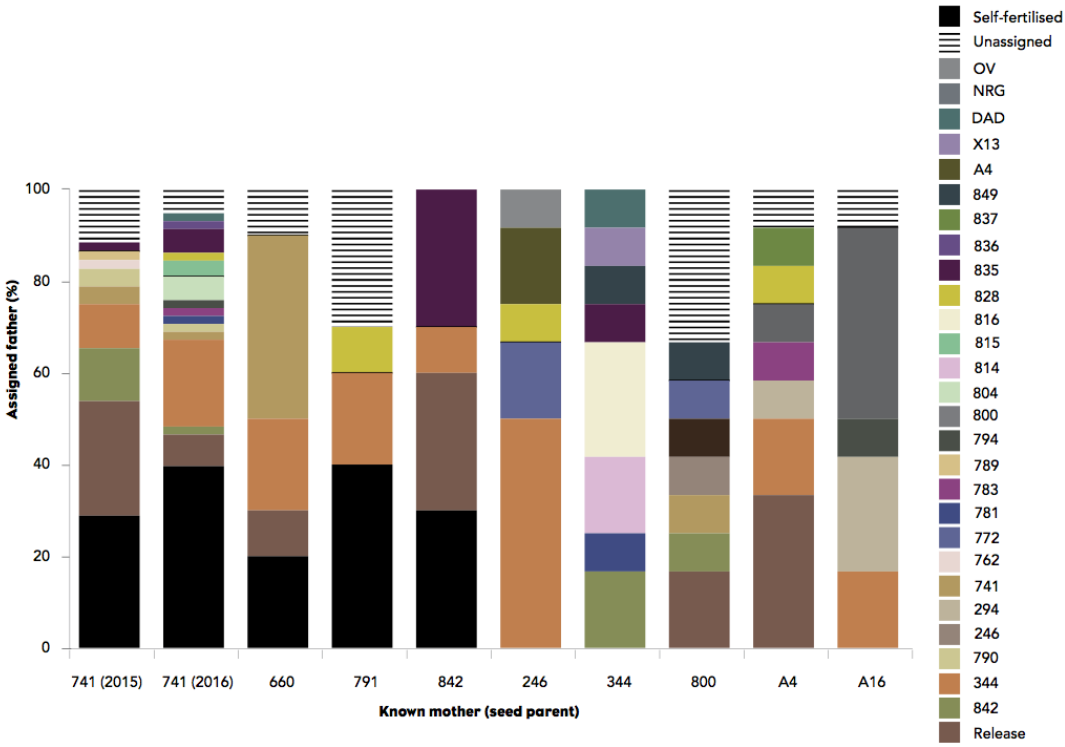

Figure 1. Paternity assignment of the progeny from nine varieties in the multi-cultivar trial plot. Self-fertilised progeny are shown in black. (OV is Own Venture, DAD is Daddow, NRG is Norm Greber)

(father). Most of the nuts were cross-pollinated, with the pollen coming from 17 different varieties (Figure 1). Still, it was surprising to find that so many were self-pollinated because the mother 741 tree was located in a randomised trial plot of 40 varieties. In 2015, 15 seedlings (29%) were selfs. In 2016, 23 (40%) were self-fertilised.

Given the unexpectedly high levels of selfing for 741, further investigation of other varieties in the plot was undertaken. In 2017, nuts were collected from eight additional varieties: 246, 344, 660, 791, 800, 842, A4 and A16. This time, Kirsty optimised a method to extract DNA directly from the kernel rather than waiting for the seed to germinate. There were no selfs in the nuts from 246, 344, 800, A4 or A16. It is possible that this could be due to the relatively small sample size, because selfing (1–20%) was reported previously for A16 in the Hort Innovation project (MC98027) *Maximising the benefits from cross-pollination in macadamia orchards.*

We found that between 20% and 40% of the nuts from 660, 791 and 842 were self-fertilised (Figure 1). Variation in the levels of self-fertilisation between the nine cultivars tested was significant. Recently, Stephen Trueman's team carried out paternity testing in commercial orchards north of Bundaberg using the same genetic markers (*AMS Bulletin,* Autumn 2019). They tested nuts from three varieties: A4, 816 and Daddow. As with our study, they found that most nuts were cross-pollinated. There were very few self-pollinated nuts (< 8%) even in the middle of large single-block plantings. Together, the DNA paternity studies on macadamia suggest that many of the varieties tested so far exhibit low selfing rates. Other varieties, including 741, 660, 791 and 842 seem to be self-fertile with selfing rates of up to 40%, even in a randomised trial plot.

As well as comparing rates of self- and cross-fertilisation, we were also able to determine the pollen parent of 93% of seedlings and 87% of the nuts tested. This was possible because we genotyped every variety in the plot. As shown in Figure 1, there were large differences in the range of pollen parents, and this provides some insight into cross-compatability. Interestingly, the predominant pollen parents of cross-pollinated nuts were often the closest neighbours, suggesting that pollination may be occurring primarily over short distances.

Access to the broad range of varietal DNA genotypes in the SCU database also allowed estimation of relatedness between varieties. This confirmed the reported parentage of some varieties such as Hidden Valley Plantation's A4 and A16. One of the most

interesting findings was the close relationship between two of the most widely grown cultivars in Australia (344 and 741) at the level of parent-offspring or full siblings.

Future directions

This was a small pilot study. Ongoing research is needed to confirm these findings on a larger scale, and to explore the relative contribution of self- and cross-pollination to yield in commercial orchards. Our research demonstrates the power of DNA paternity testing to provide detailed information on pollination under natural orchard conditions. Improved understanding of cross-compatibility, self-fertility and pollination distance will be important to optimise design of individual Australian orchards, as well as for planning at landscape scale.

New paper on DNA paternity testing in macadamia

A scientific paper by Kirsty Langdon, Graham King and Cathy Nock with more details on some of the findings of this research was peer-reviewed and published in *Tree Genetics and Genomes* 15:29, https://doi.org/10.1007/s11295-019-1336-7.

Acknowledgments
This research, and Kirsty Langdon's PhD scholarship, was funded by Hort Innovation (MC15008) with voluntary contributions from the Knappick Foundation, Macadamia Conservation Trust, Southern Cross University and the Australian Government. We thank Professor Graham King for his invaluable role in this research project, Kim Wilson and Alex Yong for assistance and access to the orchard, Asuka Kawamata for her expertise and support in the lab, and Alicia Hidden for horticultural support.