Conservation genetics of wild populations and botanic garden collections of Australian cycads

James A. R. Clugston and Nathalie S. Nagalingum

This project focuses on three species of cycad endemic to the Northern Territory of Australia: *Cycas armstrongii* Miq., *C. calcicola* Maconochie, *C. maconochiei* Chirgwin & K.D.Hill, and a single hybrid population of *C. armstrongii* x *maconochei*. Our aims are to use genetic data to assess the conservation status of these *Cycas* in order to allow fine turning of conservation management plans, and assess the putative hybrid status of *C. armstrongii* x *maconochei*. In addition, cultivated material obtained from ex-situ botanic garden collections will be directly compared to the genetic diversity of wild populations. This will enable us to determine how well the genetic diversity of the wild populations are represented in ex-situ conservation collections. Ultimately, this will improve the representation of the wild genetic diversity in ex-situ conservation collections.

This study adopts next generation sequencing (NGS) for conservation, and is the first to apply this new technology to cycads. Specifically, we are using an NGS method called restriction site associated DNA sequencing (RADseq) for generating new data on the genetic diversity of wild populations of *Cycas* in Australia. The results will provide vital insights into genetic variability, inbreeding, outbreeding, and the genetic differentiation within and between populations of *Cycas*.

Taxon sampling

Field collected samples. Silica-dried leaflets and herbarium voucher specimens were collected from wild *C. armstrongii*, *C. calcicola*, *C. maconochei* populations, and a *C. armstrongii* x *maconochei* hybrid population on a fieldtrip (May 2015) to the Northern Territory in Australia (Figure 1, Table 1).

The original proposal included *C. pruinosa*. However, the populations that occur in NT were not possible to reach by car. Due to the inaccessibility of the populations we selected another species, *C. calcicola. Cycas calcicola* is considered to be abundant around Litchfield National Park and the Katherine region, occurring as two disjunct and possibly independently breeding populations. Many of these populations are also easily accessible by car.

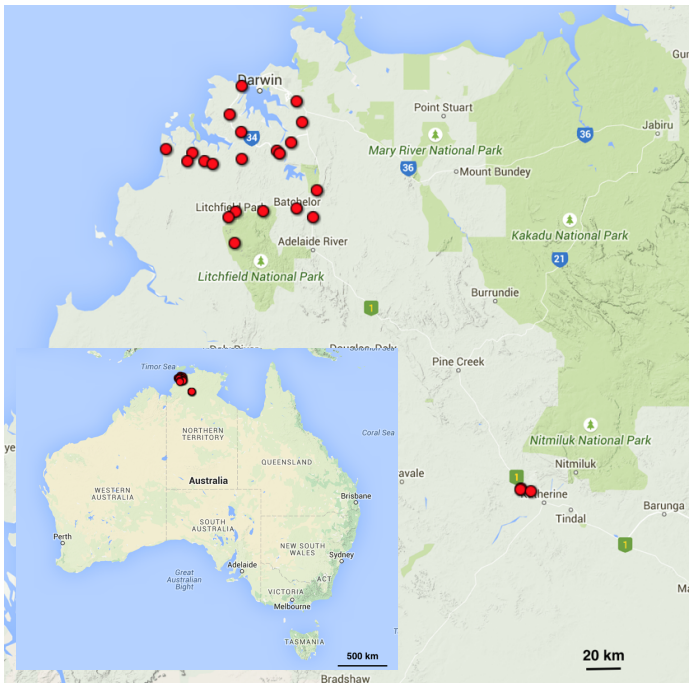

Figure 1 Location of sampling sites in the Northern Territory in relation to Australia. Red markers concentrated around the Darwin and Katherine region indicate location of all population level sampling for *C. armstrongii*, *C. calcicola*, *C. maconochei* populations, and *C. armstrongii* x *maconochei* hybrid population.

Table 1 Total number of populations and samples of *C. armstrongii*, *C. calcicola*, *C. maconochei*, and *C. armstrongii* x *maconochei* hybrid.

Species	Populations	Samples
C. calcicola	6	60
C. armstrongii	9	90
C. maconochei	8	77
C. armstrongii x maconochiei	1	10

Cycas armstrongii. The collections for *C. armstrongii* consist of a total of nine populations (Table 1, Figure 2a). Samples were obtained from the Darwin region and west of Darwin on Cox Peninsula Road, where the species was found to be abundant comprising many large populations. Further populations were also sampled close to Lake Bennett and around the Litchfield National Park area (south of Darwin; Figure 2a).

Cycas maconochei. A total of eight populations of *C. maconochei* were sampled (Table 1, Figure 2b). Three populations were sampled along Cox Peninsula towards Wagait Beach where *C. maconochei* was found to be abundant, and five populations were sampled along Fog Bay Road towards Dundee Beach.

C. armstrongii x maconochei. A single hybrid population of *C. armstrongii* x *maconochei* was sampled (Table 1, Figure 2c). It is known only as a single population along Fog Bay Road between the range of *C. armstrongii* and *C. maconochei*.

Cycas calcicola. The *C. calcicola* collections consist of six populations: three populations in Litchfield National Park and three from the Katherine region (Table 1, Figure 2d). The collected populations represent the northernmost and southernmost limits of the species' range, which will provide insights into the geographic differentiation of the species.

Cultivated samples. Fresh silica-dried leaflet material for *C. armstrongii*, *C. calcicola* and *C. maconochei* were obtained from the ex-situ conservation collections of Darwin Botanic Garden, University of California Berkley Botanic Garden, Montgomery Botanical Center and the Royal Botanic Garden Sydney. This material will be compared directly to wild collected populations.

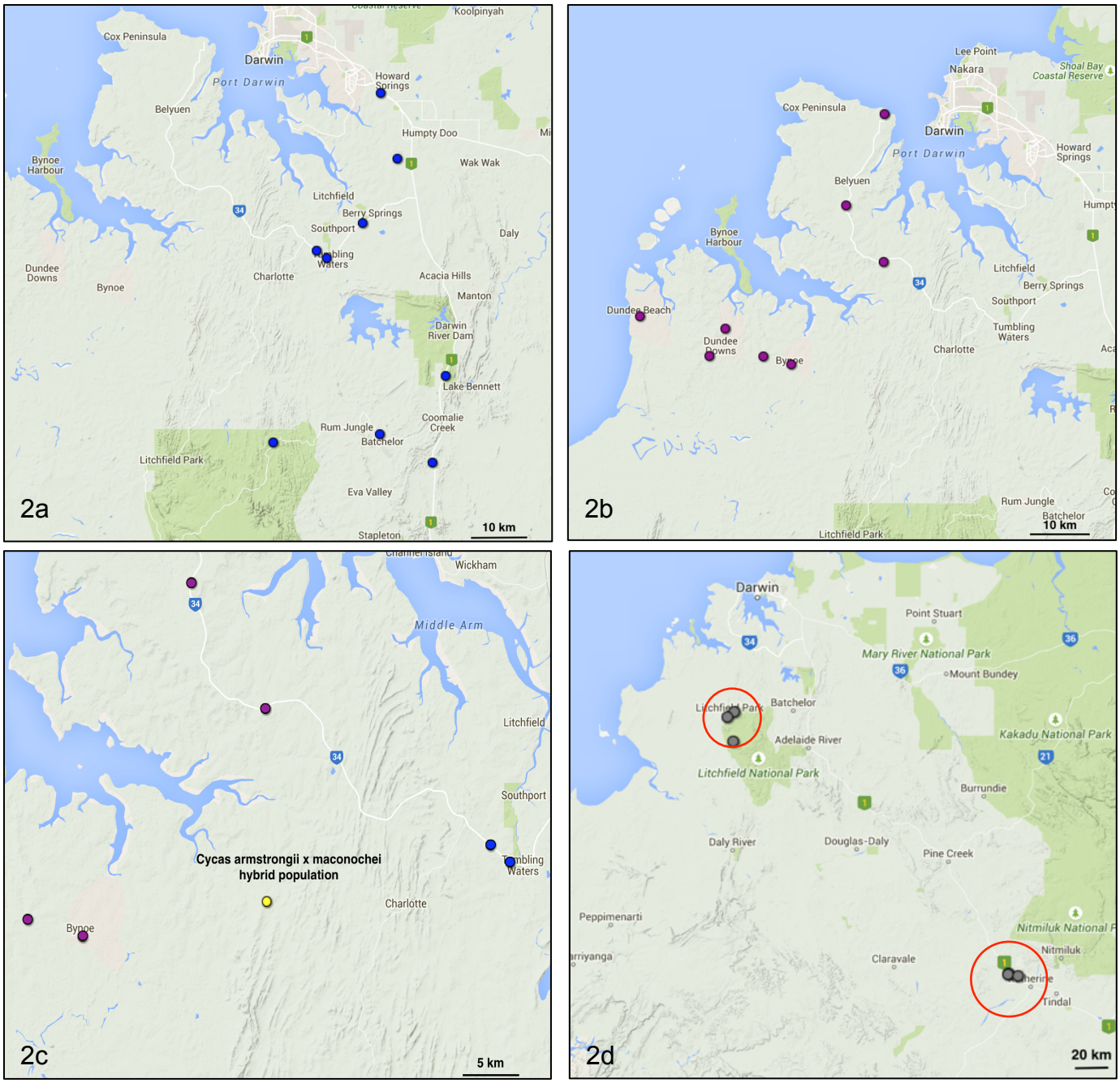

Figure 2 Location of populations sampled in the Northern Territory, Australia. 2a *Cycas armstrongii* populations sampled around Darwin, the Cox Peninsula and Litchfield National Park region. 2b *Cycas maconochei* populations sampled along the Cox Peninsula and Fog Bay Road. 2c *Cycas armstrongii* x *maconochei* hybrid population (yellow) sampled along Fog Bay Road between the range of *C. armstrongii* (blue) and *C. maconochei* (purple). 2d *Cycas calcicola* populations sampled in Litchfield National Park and around the Katherine region, the two disjunct groups of populations are circled in red.

Molecular techniques and current progress

In the original grant we proposed shotgun sequencing to identify microsatellites. However, newer and more comprehensive genomic approaches have subsequently become feasible, namely RADseq (Restriction site Associated DNA sequencing). RADseq can identify potentially thousands of polymorphisms throughout the genome (Peterson et al. 2012; Toonen et al. 2013), e.g., a study of *Helianthus annuus* L. recovered 5,019 informative SNP markers (Talukder et al. 2014). The data that we obtain from RADseq will consist of hundreds of thousands of small stretches of the genome for population genotyping. The application of RADseq has been successfully used across a range of organisms, and in non-model angiosperms, to understand ancestral hybridisation,

adaptation, the emergence of natural species barriers, and genomic differentiation in and among populations (Eaton & Ree 2013; Roda et al. 2013).

We have currently tried, tested, and refined the ezRAD protocol v3 (Toonen et al. 2013) to better suit *Cycas* and yield optimum results. The ezRAD protocol has a number of steps all involving the use of Illumina TruSeq Nano DNA library preparation kit. The protocol is as follows (1) genomic DNA is digested into fragments using a double digest reaction, i.e., with two restriction enzymes (EcoR1-HF and Mse1). (2) Fragments are then "end repaired" to convert any overhang. (3) Fragment size selection is carried out, selecting fragments between 200-600 base pairs in length (discarding fragments above and below this range). (4) Fragmented ends are adenylated (to prevent fragments from sticking to each other). (5) Dual index adapters are ligated to fragments. (6) PCR is carried out to amplify the fragments. Once the fragments have been amplified all samples are pooled (=multiplexed) into a single tube, and are sequenced using an Illumina NextSeq 500. The ezRAD protocol offers some advantages over other RADseq protocols due to better horizontal read coverage and less amplification biases through use of the illumina TruSeq Nano library preparation kit (Andrews et al. 2016).

The first batch of 95 libraries (comprising 86 samples of *C. calcicola* and 9 cycad genera in Zamiaceae) has been prepared as described above, and has been through numerous quality controls. These libraries have already been sequenced on the Illumina NextSeq 500 at the Ramaciotti Center for Genomics at the University of New South Wales. The data have been returned to us (67GB), and they are undergoing quality control and assessment. Two additional libraries (95 samples each) for *C. armstrongii*, *C. maconochei, and C. armstrongii* x *maconochei* are now in preparation.

Expected data analysis

Using the dual index markers sensu Peterson (2012), the pooled/multiplexed samples will be de-multiplexed in order for each sample to be identified individually. De novo assembly of genomic fragments and single nucleotide polymorphism (SNP) discovery will be performed using CLC Genomics Workbench and Geneious.

A range of analyses will be performed. Firstly, using the program Structure, we will conduct a population structure analysis to detect underlying genetic variation at population level, and to identify sets of individual genotypes in each population (Hubisz et al. 2009). Structure will also be used to test the genetic ancestry of the populations. Secondly, using Genepop, we will to test if populations deviate from the Hardy Weinberg equilibrium and test for genetic differences between populations using a population differentiation test (Rousset 2008). Lastly, the software package IBD: isolation by distance will be used to test for the genetic connectivity of the populations and to run a geographic distance test to evaluate the relationships between geographic distance and genetic divergence. The program pyRAD will be used to understand various population genetic parameters including Fst and Fis values (Eaton 2014). pyRAD will again be used to assess the genetic variation in plants from *ex-situ* conservation collections, and link genotypes to their wild counterparts. The results from this final analysis will determine if botanic garden collections represent the genetic diversity of the wild populations, and design a conservation management plan.

Student recruitment

The project has successfully recruited James Clugston who has already completed an MSc and has a wide knowledge of cycads. James enrolled in a Ph.D. in September 2015 at the University of Edinburgh, but is based at the Royal Botanic Garden Sydney in Australia for 2 years. Together with supervisor Nathalie Nagalingum, James travelled to the Northern Territory to collect the plants for this project. At the Royal Botanic Garden Sydney, James has been developing the RADseq techniques for the project with good success.

References

Andrews, K.R. et al., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2), pp.81–92.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), pp.1844–1849.

Eaton, D.A.R. & Ree, R.H., 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, 62(5), pp.689–706.

Hubisz, M.J. et al., 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*.

Peterson, B.K. et al., 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5), pp.1–11.

Roda, F. et al., 2013. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology*, 22(11), pp.2941–2952.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), pp.103–106.

Talukder, Z.I. et al., 2014. A High-Density SNP Map of Sunflower Derived from RAD-Sequencing Facilitating Fine-Mapping of the Rust Resistance Gene R12. *PLoS ONE*, 9(7), pp.1–14.

Toonen, R.J. et al., 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. *Peerj*, 1(14), pp.e203–e203.