Bioclimatic discordance: Combining molecular and environmental data to identify floristic refugia and corridors

Maurizio Rossetto[1], Linda Beaumont[2], Sourav Das[2], Sam Yap[1]

[1]National Herbarium of NSW, Royal Botanic Gardens and Domain Trust, Sydney NSW 2000

[2]Biological Sciences, Macquarie University NSW 2109

For further correspondence contact: maurizio.rossetto@rbgsyd.nsw.gov.au

Final report prepared for the NSW Office of Environment and Heritage,

April 2018

This project received funding from the Biodiversity Node of the NSW Adaptation Research Hub

Executive Summary

This study develops a new comparative approach that brings together temporal changes in environmental suitability, with genome-wide signatures of landscape-level dynamics to identify historical refugia. Our results suggest that rainforest vegetation did not respond to past climatic fluctuation as a uniform assemblage, but that current distributional patterns are the result of distinct histories. This information can help us better predict community-level resilience to future change, and the approach developed is widely applicable to other communities and/or landscapes.

Introduction

The analysis of molecular data can identify areas that have served as population refugia (generally identified by the accumulation of high levels of diversity) and differentiate them from expansion areas (generally harbouring lower intraspecific diversity; Rossetto et al. 2015). Habitat Suitability Models (HSM) can explore temporal changes in availability of suitable habitat, based on historical, current and future datasets. These models assess relationships between occurrence patterns and environmental characteristics, to define suitable conditions for the species (Franklin, 2010). These relationships can then be used to assess the distribution of suitable habitat under past, present or future climate scenarios, assuming a constant biotic response or niche stability (Franklin, 2010). HSMs are a key tool used to assess paleo-distributions of species (e.g. Gavin et al. 2014; Varella et al. 2011).

The combination of suitable molecular data (whole genome scan), with historical (e.g. LGM, HMax) and current climatic datasets can identify centres of diversity as well as areas of bioclimatic stability. This approach represents an advancement in identifying landscape-level regions that have acted as refugia (as well as those that act as barriers) during previous periods of climate change, and in projecting the response of vegetation to future climate change.

Consequently, this project combines data from multiple, functionally selected species (long-lived rainforest trees with similar dispersal potential) to identify the landscape features and geographic areas that have operated as refugia during temporal climatic fluctuations and, consequently provide much improved tools for predicting future vegetation-level responses and priority areas.

Methods

Model Development: Habitat Suitability Models (HSM) were developed for 30 rainforest tree species using MaxEnt (Phillips et al., 2006), with occurrence data from each species downloaded from the Atlas of Living Australia. The models were optimised by producing a unique set of environmental variables (climate and topographic indices) that best describes the distribution for each species. While we acknowledge the importance that soil and disturbance history can play in influence species' distributions, this information is unavailable for past time periods. HSMs were then projected onto climate scenarios for multiple time periods: Last Glacial Maximum (LGM, ~22,000 ybp), mid-Holocene (MH, ~ 6000 ypb), and the current period. For each species, paleo-refugia (areas that remained stable from the LGM to the MH and the current period) were identified across Australia. These paleo refugia were used for the molecular validation step, and centres of paleo refugia, i.e. regions projected to have been climatically stable for multiple species, were identified by stacking HSMs from all 30 species.

A map identifying paleo refugia from stacked HSMs and regions defined as "core" rainforest regions in NSW was produced (based on the Australian National Vegetation Information System; Australian Government Department of the Environment & Water Resources, 2006). Four regions were investigated in greater detail as a proof of concept for the genetic validation study: Nightcap-Border Ranges (N), Dorrigo (D), Barrington Tops (BT), and Royal National Park (RNP).

Molecular Validation: Genetic data from 14 rainforest species across New South Wales were used for validating the model-based rainforest paleo refugia. These are: *Callicoma serratifolia, Diploglottis australis, Doryphora sassafras, Elaeocarpus reticulatus, Neolitsea dealbata, Orites excelsus, Pittosporum multiflorum, Polyscias murrayi, Sloanea australis, Stenocarpus salignus, Toona australis, Tristaniopsis collina, Tristaniopsis laurina, Wilkiea huegeliana.*

Leaf material was collected across the NSW distribution of these 14 species using a uniform approach, and was genotyped through Diversity Arrays Technology (DArTseq), a microarray hybridisation-based technique that simultaneously assays thousands of SNP (Single Nucleotide Polymorphism) markers across the genome using Next Generation Sequencing (NGS). The output from the DArTseq genotyping platform undergoes quality filtering through our standardised DArTseq analytical pipeline, *RRtools* within R version 3.1.2 (R Development Core Team 2014). This tool developed specifically in-house, handles the multi-species genetic data by assessing whether it follows Hardy-Weinberg equilibrium, and filters missing data, reproducibility and clonality.

For each species, genetic diversity was calculated for each population and region using allelic richness (i.e., mean number of alleles within population), expected heterozygosity and Shannon's diversity index using the R package, *poppr* (Kamvar et al. 2014). These results were combined into a large dataset that was then analysed to compare genetic diversity within and between regions (i.e. to identify refugium vs recolonised area), and to identify differences between groups of species. The observed diversity patterns for each region were collated with the stacked HSMs to cross validate the results.

Results & Recommendations

The congruence of areas projected to have suitable habitat over consecutive time periods can be used to identify future climate refugia. Paleo refugia (areas that have remained suitable from the Last Glacial Maximum (LGM) and mid-Holocene to the present day) were identified by stacking Habitat Suitability Models (HSMs) from 30 rainforest species across Australia (Fig. 1). It is generally assumed that paleo refugia have been stable for a considerable length of time, and consequently may harbour higher genetic diversity than areas that were recolonized, or areas that have undergone climate-induced bottlenecks.

To test the validity of paleo-refugia derived from stacked HSMs (with a focus on New South Wales), we analysed representative genetic data from 14 species included within the stacked paleo models (Fig. 1).

Based on our previous published work (Rossetto *et al*. 2015) we expected that refugial areas such as Nightcap-Border Ranges would display higher genetic diversity than recolonised areas such as Washpool. We also expected congruence between the areas of high genetic diversity and the areas identified as paleo-refugia by the stacked HSM models.

Figure 3 shows a summary of genetic results using average heterozygosity across all species as a measure of local diversity for four representative rainforest regions. The results can be compared to the modelled expectation of refugia (green) vs. recolonised areas (purple) in Figure 2, to show that the two datasets are in general agreement. For example, Nightcap-Border Ranges and Dorrigo have been identified as areas of climatic environmental stability for many (but not all of course) species, as well as areas of high genetic diversity. Conversely, Royal National Park and, surprisingly, Barrington Tops were identified by both datasets as mostly recolonised areas (Table 1 and Fig. 3).

However, some additional patterns also emerged from single species genomic data (Table 1, Fig. 4). For instance, while broad patterns do confirm the southern areas as mostly recolonised, a set of species appear to have high levels of genomic diversity at higher latitudes (especially within RNP; Table 1). The source of disagreement between cumulative patterns and species-level datasets originates from the fact that not all taxa have the same historical dynamics, and their specific patterns might not clearly emerge in cumulative datasets. Differences in temporal dynamics are influenced by a range of factors, such habitat preference (e.g. cool adapted species appear to be more diverse in southern areas; Table 1), different biogeographic histories (e.g. Indo Malesian lineages with a recent history of north to south migration are more diverse in northern areas; Table 1), different dispersal mechanisms, etc.

To illustrate these differences in greater detail, Figure 5 shows the different dynamics between a cool adapted species, *Doryphora sassafras*, and a recent Indo-Malesian migrant, *Toona australis*. Differential dynamics, greater diversity in the south for the cool adapted species and in the north for the recent migrant, correspond to the patterns obtained in species-specific HSMs (greater stability in the south for *Doryphora* and in the north for *Toona*; Fig. 6).

Thus, genetic data from species selected according to distributional patterns, functional diversity and biogeographic history, can be used as a feedback mechanism to inform the stacking strategy used in the development of paleo HSMs. We are planning to further confirm these findings by investigating fine-scale directional and temporal dynamics across selected species through coalescent models. From there we will develop new stacked HSMs based on expanded functionally- and biogeographically-grouped species.

This pilot study shows that while modelling the response to past climate change for single species cannot be used to generalise across whole communities, stacked (multi-species) models that consider functional and historical factors can be highly informative. Species-wide genome scans of representative species provide the critical landscape dynamic interpretation needed for group selection.

While this pilot study was conducted on rainforest vegetation, the same approach is applicable to any vegetation type. With the recent advances in the automation of genomic research obtaining the necessary genetic validation information is relatively simple. The ability of validating paleo refugia provides increased confidence in the use of a similar HSM-derived approach for identifying putative future refugia (areas that a likely to remain sufficiently climatically stable even within the parameters of predicted future climate).

References

Franklin J (2010) *Mapping species distributions: spatial inference and predictio*n. Cambridge University Press.

Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Rodríguez-Sánchez F, Dobrowski SZ, Hampe A, Hu FS, Ashcroft MB, Bartlein PJ (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist* **204**, 37-54.

Kamvar ZN, Tabima JF, Grunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Phillips S, Anderson R, Schapire R (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

R Development Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Rossetto M, McPherson H, Siow J, Kooyman R, van der Merwe M, Wilson PD (2015) Where did all the trees come from? A novel multidisciplinary approach reveals the impacts of biogeographic history and functional diversity on rain forest assembly. *Journal of Biogeography* **42**: 2172-2186.

Varela S, Lobo LJ, Hortal J (2011) Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. *Palaeography Palaeoclimatology Palaeoecology*. **310**, 451-463.

Figures

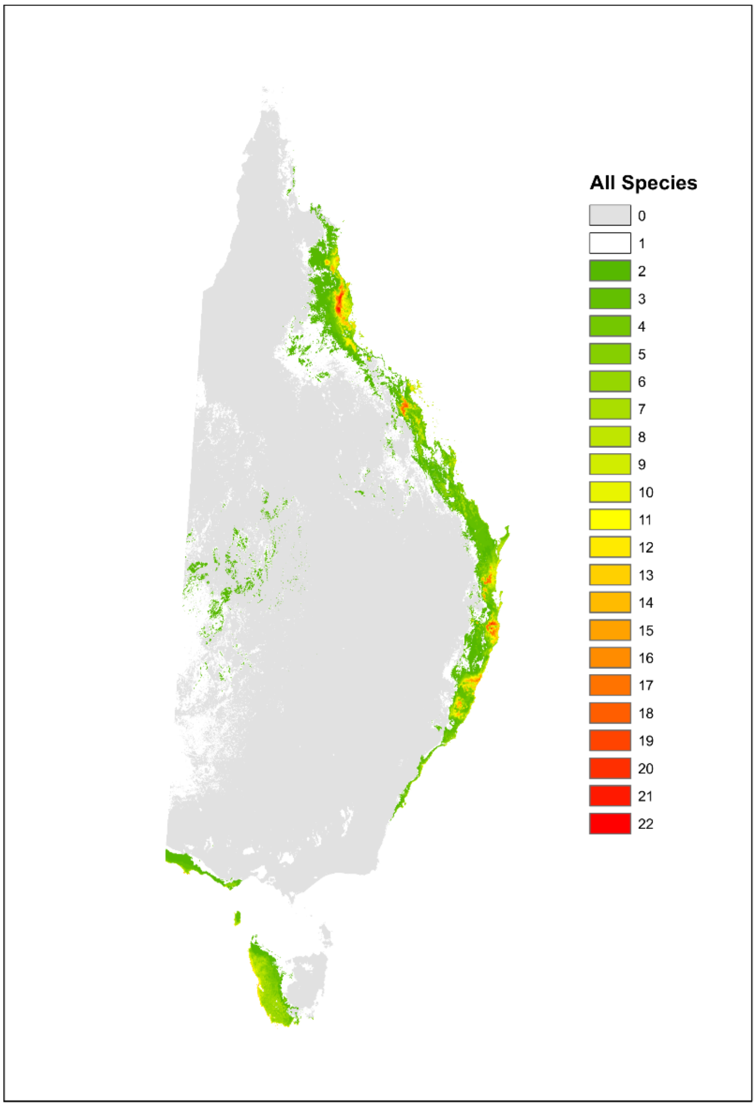


Fig. 1: Centres of paleo refugia for 30 tropical and subtropical Australian rainforest plant species, based on habitat suitability maps generated with Maxent. Paleo refugia are defined as the intersection of areas classified as suitable over consecutive time periods: Last Glacial Maximum (LGM), mid-Holocene and the current period. Colours identify the number of species that persist at each location through time (i.e. red identifies regions that were projected to be refugia for more species while green are regions projected to be refugia for fewer species). (*Source: Sourav Das, MSc. Thesis*)

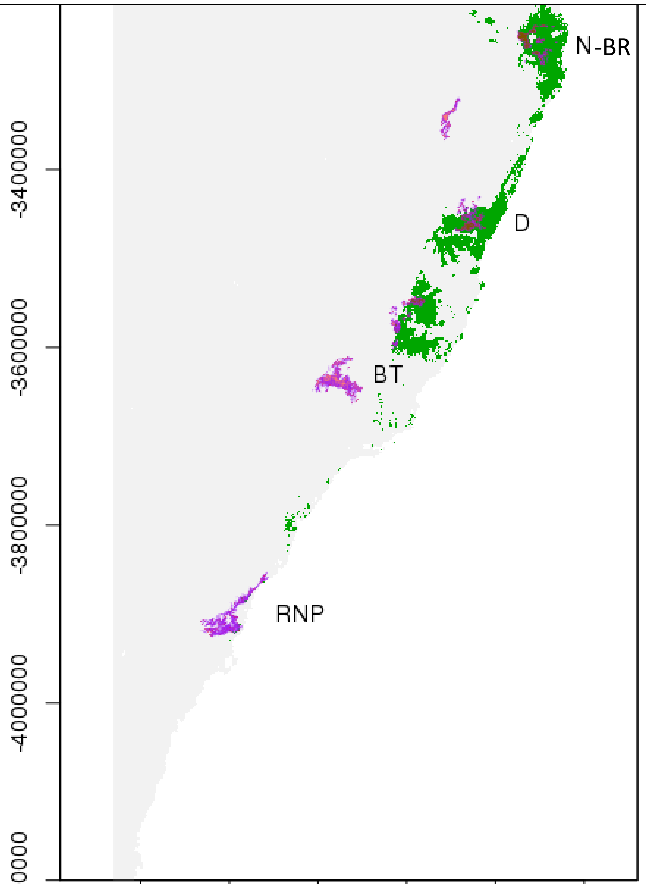


Fig. 2: Target rainforest areas for this project (NSW). Coloured highlights identify targeted rainforest vegetated areas, the green subset identifies stable refugia for 15 or more species (based on Fig. 1). (BT: Barrington Tops; D: Dorrigo; N-BR: Nightcap Border Ranges; RNP: Royal NP). (*Source: Sam Yap, Ph.D Thesis*)

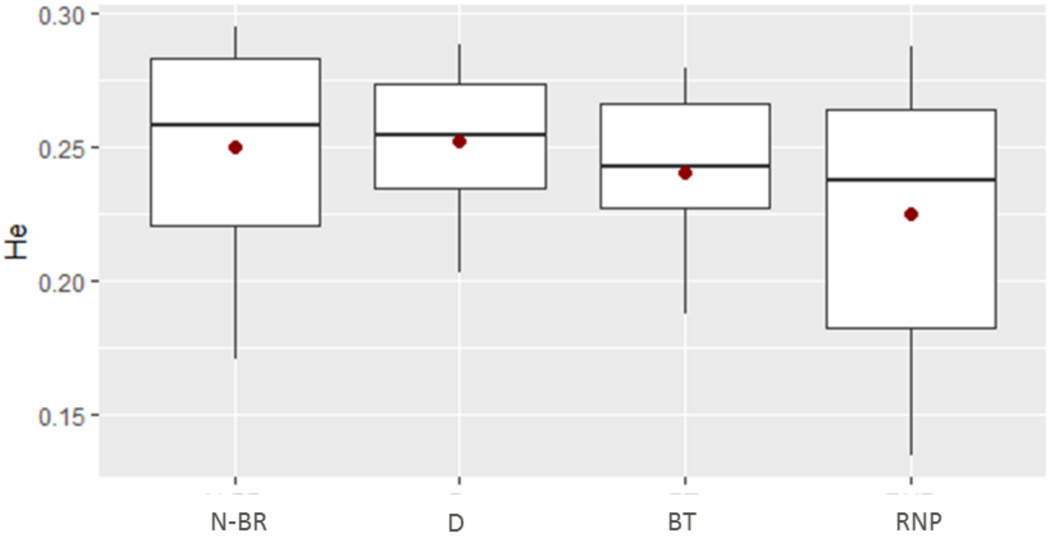
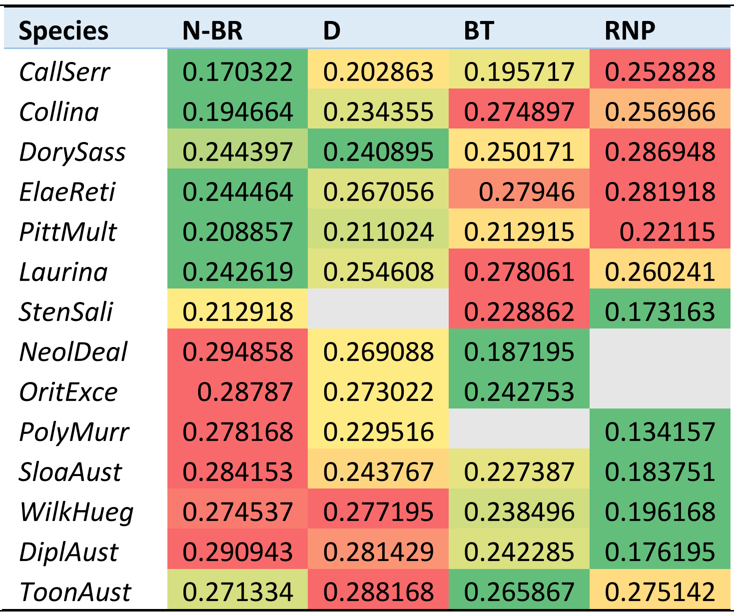


Fig. 3: Differences in expected heterozygosity across four sites. Red dot indicates the average expected heterozygosity across all species at each site. (*Source: Sam Yap, Ph.D Thesis).*



Species	N-BR	D	BT	RNP
CallSerr	0.170322	0.202863	0.195717	0.252828
Collina	0.194664	0.234355	0.274897	0.256966
DorySass	0.244397	0.240895	0.250171	0.286948
ElaeReti	0.244464	0.267056	0.27946	0.281918
PittMult	0.208857	0.211024	0.212915	0.22115
Laurina	0.242619	0.254608	0.278061	0.260241
StenSali	0.212918		0.228862	0.173163
NeolDeal	0.294858	0.269088	0.187195	
OritExce	0.28787	0.273022	0.242753	
PolyMurr	0.278168	0.229516		0.134157
SloaAust	0.284153	0.243767	0.227387	0.183751
WilkHueg	0.274537	0.277195	0.238496	0.196168
DiplAust	0.290943	0.281429	0.242285	0.176195
ToonAust	0.271334	0.288168	0.265867	0.275142

Table 1: Four sites are ranked based on species diversity measured as heterozygosity (BT: Barrington Tops; D: Dorrigo; N-BR: Nightcap Border Ranges; RNP: Royal NP). (*Source: Sam Yap, Ph.D Thesis*)

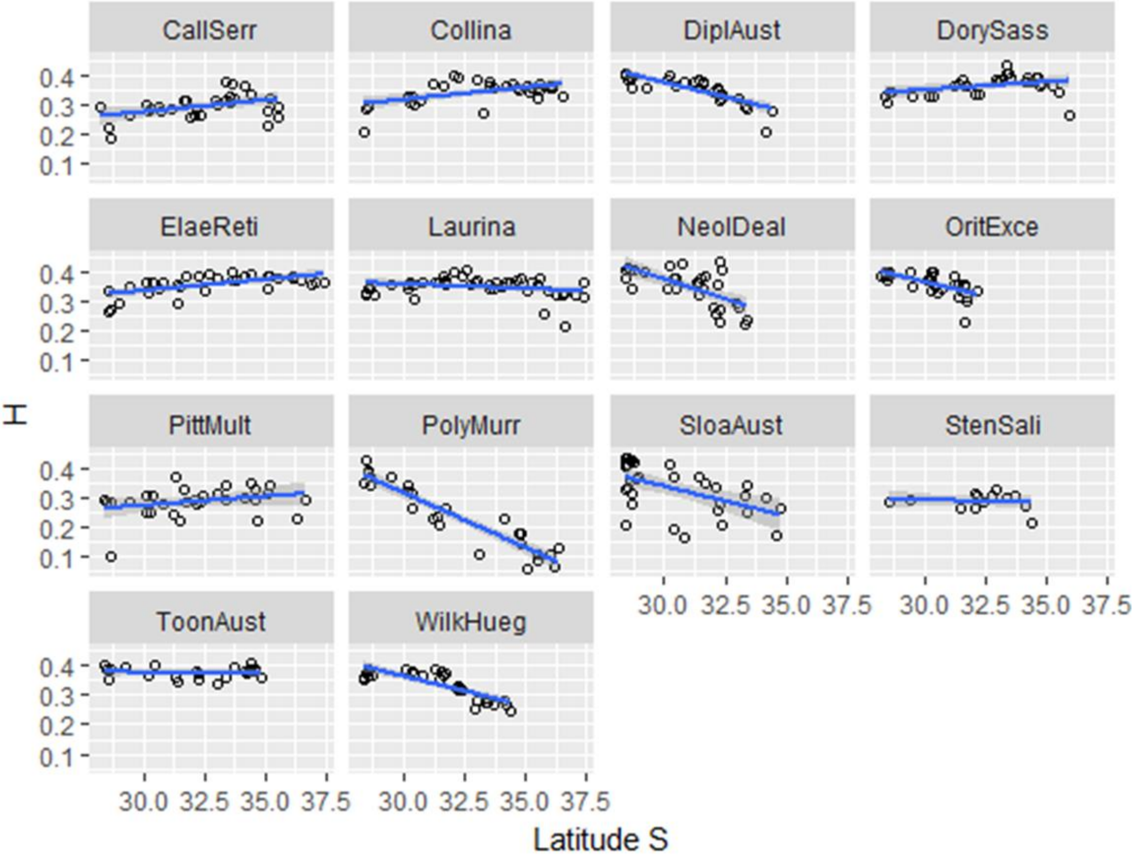


Fig. 4: Relationship between genetic diversity and latitude across 14 species. This suggests that latitudinal directionality of re-expansion patterns differed among species (e.g. north to south vs. south to north). (*Source: Sam Yap, Ph.D Thesis).*

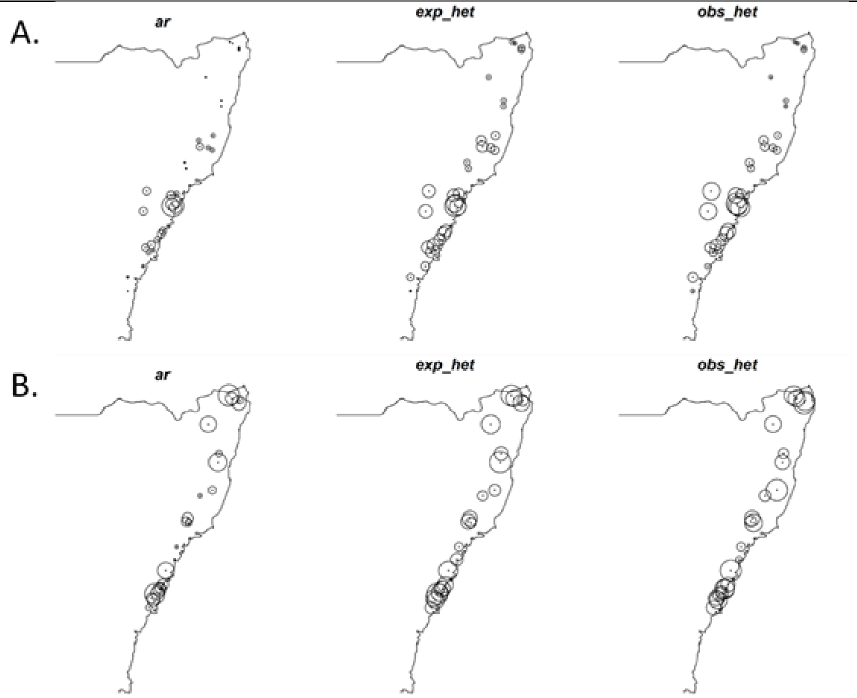


Fig. 5: Example of contrasting historical dynamics between A. *Doryphora sassafras* (Gondwanan heritage, south to north expansion); B. *Toona australis* (Indo-Malesian origins, very recent expansion from northern Australia). (*Source: Sam Yap, Ph.D Thesis*)

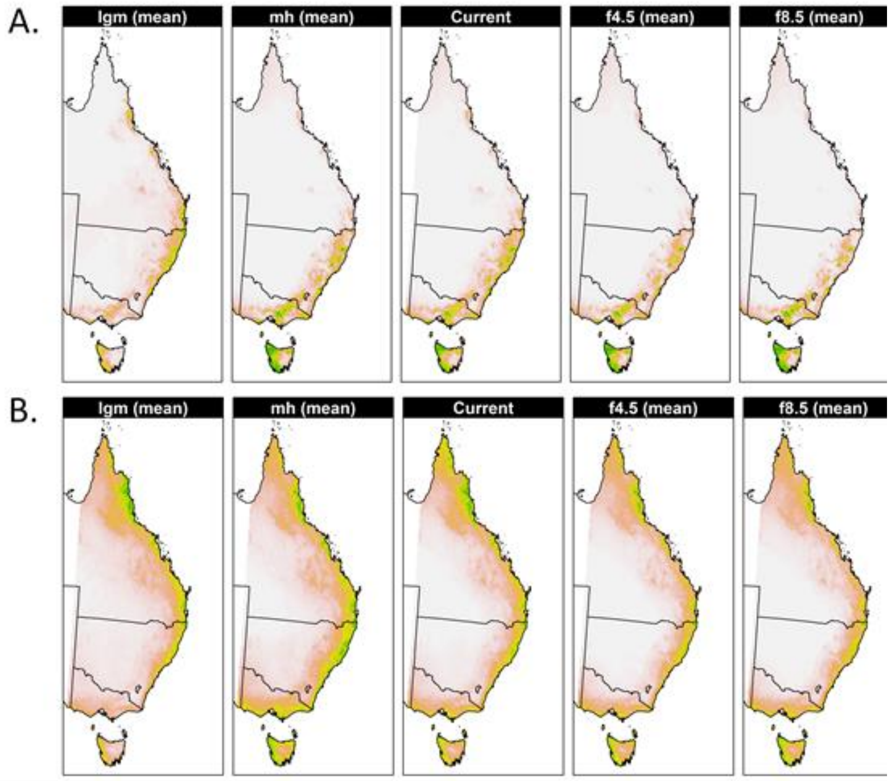


Fig. 6: Change in habitat suitability for A. *Doryphora sassafras*; B. *Toona australis*. Maps represent Last Glacial Maximum (LGM), mid-Holocene (MH), the current period (Current), and future period (2070) under two representative concentration pathways, RCP 4.5 (F4.5) and RCP 8.5 (F8.5). Modelling suggests a contraction in habitat availability for *Doryphora* and expansion for *Toona* (consistent with genetic data). (*Source: Sourav Das, MSc. Thesis*).