New insights from population genomics into the invasive *Lantana camara* L species complex

Patricia Lu-Irving[1], Francisco Encinas-Viso[2], Jason Callander[3], Michael Day[3], Johannes Le Roux[4]
[1]Research Centre for Ecosystem Resilience, Australian Institute of Botanical Science, Royal Botanic Garden Sydney, NSW, Australia
[2]Centre of National Australian Biodiversity Research, CSIRO, Canberra, ACT, Australia
[3]Department of Agriculture and Fisheries, GPO 267, Brisbane, QLD, Australia
[4]School of Natural Sciences, Macquarie University, Sydney, NSW, Australia

Summary Extensive morphological diversity in the invasive *L. camara* species complex has resisted clear taxonomic characterisation, yet molecular studies to date have revealed minimal genetic structure. Analysis of thousands of genome-wide markers successfully detects differentiation among sub-lineages within the complex, revealing that two of the morphological variants in Australia comprise multiple, distinct lineages. The common pink-flowering morphotype appears to be monophyletic, whereas the pink-edged red flowering morphotype does not. Pending further investigation and expanded sampling, these findings hold promise for improving weed management through a deeper understanding of the systematics of the *L. camara* complex (e.g., by enabling selection of biological control agents that are best matched with target host populations).

Keywords biological control, biosecurity, invasive species, weeds

INTRODUCTION

Lantana camara is weed of national significance in Australia and has been a target for biological control for over 100 years. However, the diversity of forms that make up this species complex has limited the success of biological control programmes, with natural enemies performing better on some varieties than others (Day et al. 2003a). Resolving the evolutionary diversity within the *L. camara* species complex is expected to lead to improved management of these globally significant weeds (e.g., through better biological control agent-host matching). However, natural species boundaries within the complex have been obscured by anthropogenic hybrid introductions, and subsequent spontaneous hybridisation among native and naturalised plants. Studies to date of both morphological and molecular variation in *L. camara* have yet to reveal clear and consistent patterns which can be used to identify invasive populations across the complex's entire geographic distribution.

Within Australia, *L. camara* is thought to consist of multiple taxa (Smith & Smith, 1982) and varietal groups (Day et al. 2003b): suggesting that the complex comprises distinct sub-lineages that are divergent enough to warrant recognition. However, molecular studies to date of up to ~200 genetic markers (Scott et al. 1997, Watts 2009) and Sanger sequencing of 16 nuclear and chloroplast loci (Watts 2009) have found insufficient evidence to support this view.

Here, we report preliminary results from analysis of genome-wide markers in the *L. camara* complex (hereafter referred to by the common name, "lantana"). We used a Genotyping-By-Sequencing (GBS) approach to discover thousands of bi-allelic Single Nucleotide Polymorphisms (SNPs) in 101 Australian and extra-Australian lantana samples. We analysed these data to investigate whether there is detectable genetic structure in Australian lantana, and, if so, the extent to which it aligns with flower colour (the morphological character most used to define subgroups within lantana). Day et al. (2003b) described five broad varietal groups based on flower colour: common pink, pink-edged red, red, orange, and white. Since these concepts have yet to receive formal taxonomic recognition, we use the term "morphotype" to refer to them.

MATERIALS AND METHODS

Sampling Healthy leaf tissue was collected in the field and preserved by desiccation. Most samples analysed here were collected from the invaded range in eastern Australia (87 samples across 23 sites from latitude 37.0ºS to 17.5ºS); 6 plants were sampled per morphotype per site at 10 sites (9 sites with 1 morphotype, 1 site with 2); 1 plant was sampled per site at 12 sites. The remainder of the samples originated from northern and western Australia (3), Hawaii (2), South Africa (2), and the Americas where the genus *Lantana* is native (North America: 5; Caribbean: 1; South America: 1).

Sampling focused on the common pink-flowering morphotype (52 samples over 15 sites), and the pink-edged red flowering morphotype (16 samples over 6 sites).

Genotyping Tissue samples were submitted to Diversity Arrays Pty Ltd. (DArT) for DNA extraction, sequencing, and SNP calling using the

DArTseq GBS pipeline; this approach is described in detail by Rossetto et al. (2019).

Population genomic analysis All data processing and analyses were conducted in the R environment (R Core Team 2020). The SNP loci returned by DArT were filtered by reproducibility, representation among samples, and independence; SNPs passed filter if they were >96% reproducible, missing from =<20% of samples, and not co-located on sequencing tags.

The filtered data were analysed using Splitstree v4.14.6 (Huson & Bryant 2006) to infer a splits graph of evolutionary relationships among samples. This approach enabled representation of a non-treelike (i.e. potentially reticulate) evolutionary history (Huson & Bryant 2006), which is to be expected in lantana, given its history of extensive hybridisation.

For sites/morphotypes where six samples were collected, pairwise population F_{ST} values were calculated using the relative beta estimator implemented in the "SNPrelate" v1.20.1 package (Zheng et al. 2012). T providing a measure of genetic differentiation among populations.

RESULTS

The raw SNP matrix consisted of 101 samples and 60,634 loci, 42,958 (71%) of which were found in less than 80% of samples (i.e., were missing from 20% of samples or more). After filtering, 10,847 SNPs remained in the final data set.

Multiple, distinct lineages were resolved in a phylogenetic network (Fig. 1). The most clearly defined lineage consisted entirely of eastern Australian common pink-flowering lantana (the morphotype with broadest geographic distribution; "pink " hereafter).

Genetic differentiation (pairwise F_{ST}) between populations of pink lantana was close to zero, even over large geographic distances (>1,000 km). In comparison, F_{ST} between populations of the pink morphotype and pink-edged red morphotype ("PER" hereafter) were very high, even in areas where they occurred in close proximity or in sympatry (Table 1).

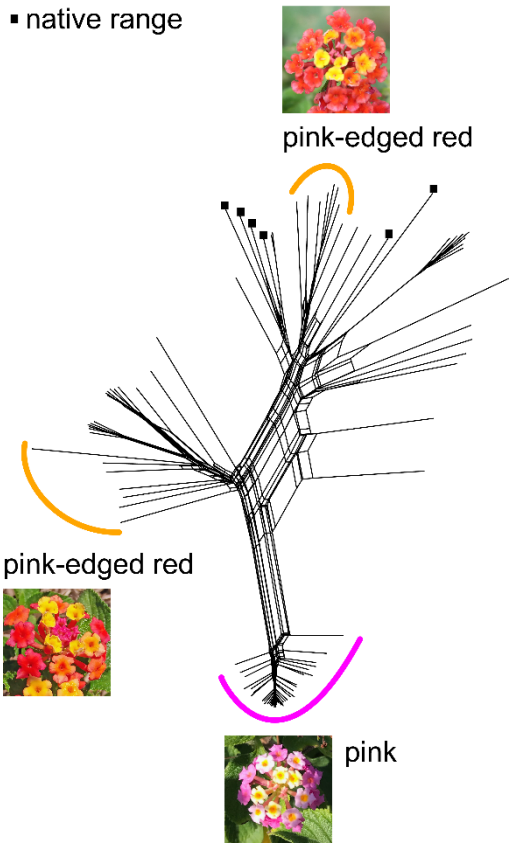

Figure 1 (opposite). Phylogenetic network of 101 lantana samples inferred from 10,847 SNPs. Clusters of eastern Australian pink-flowering and pink-edged red flowering lantana samples are indicated, as are samples from the native range.

Table 1. Pairwise F_{ST} and geographic distance (km) among selected populations of pink-flowering and pink-edged red flowering morphotypes of *Lantana camara*. Boonah was the one site at which two morphotypes were sampled in sympatry; of the other sites with 6 samples per morphotype, Mt. Fisher was most geographically distant from Boonah.

	MT FISHER (PINK)	BOONAH (PINK)	BOONAH (PER)
MT FISHER (PINK)	0 KM	1,369 KM	1,369 KM
	F_{ST}= 0	F_{ST}= -0.03	F_{ST}= 0.56
BOONAH (PINK)	-	0 KM	0.32 KM
	-	F_{ST}= 0	F_{ST}= 0.51

DISCUSSION

The unprecedented quantity of data yielded by our GBS approach enabled detection of patterns in genomic variation which was not possible with prior molecular approaches. Our analysis of genome-wide markers provided clear evidence of distinct lineages within Australian lantana (i.e., monophyletic groups which are genetically divergent from other such groups). That 71% of SNPs discovered by DArTseq were not found across all samples is consistent with this interpretation (i.e. a large proportion of loci were unique to one lineage or a subset of lineages).

The lineages revealed show strong correspondence with described lantana morphotypes (*sensu* Day et al. 2003b). This is in contrast with prior molecular studies reporting limited differentiation among morphotypes (Scott et al. 1997; Watts 2009), which we attribute to the methodological advances our study was able to apply. We report here on our findings specifically concerning the two most widespread morphotypes in Australia, for which we had greatest sampling effort: pink (52 samples) and PER (16 samples); further work is ongoing which will report in greater detail on a more comprehensive sample including other morphotypes as well as extra-Australian populations.

Samples of pink lantana were phylogenetically distinct from other samples (Fig. 1), and we found no genetic differentiation between populations of this morphotype (even between populations >1,000 km apart; Table 1). However, populations of pink and PER morphotypes were strongly differentiated, even when in sympatry (Table 1). We conclude that the common pink-flowering morphotype corresponds with a distinct lineage, a meaningful biological and evolutionary unit. On the other hand, the PER morphotype does not meaningfully identify a lineage, with at least two distinct and divergent groups of plants appearing to share this phenotype (Fig. 1).

Biological control implications Since the pink lantana morphotype represents a uniform host genetic background, agents which prefer it (e.g. *Prospodium tuberculatum* (Spegazzini) Arthur) would be expected to affect different populations of this morphotype consistently. However, agents preferring the PER morphotype (e.g. *Teleonemia scrupulosa* Stål and *Aceria lantanae* Cook) might not be expected to express this preference consistently among host populations, because this morphotype encompasses multiple and genetically-distinct lineages of lantana. Further study is required to test these predictions and report in depth on the applications of the findings presented here.

ACKNOWLEDGMENTS

We thank the colleagues and volunteers who contributed samples to this project, in particular the 18 individual members of Landcare groups including Upper Clarence Combined Landcare and Landcare Illawarra. This project was supported by the Department of Agriculture, Water, and the Environment (activity ID 4-FY9KQZ3).

REFERENCES

Day, M.D., Wiley, C.J., Playford, J., Zalucki, M.P., 2003a. Lantana current management status and future prospects. ACIAR Canberra. Monograph 102.

Day, M.D., Broughton, S., Hannan-Jones, M.A., 2003b. Current distribution and status of Lantana camara and its biological control agents in Australia, with recommendations for further biocontrol introductions into other countries. Biocontrol News and Information 24(3): 63N – 76N.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23(2): 254 – 267.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rossetto, M., Bragg, J., Kilian, A., McPherson, H., van der Merwe, M., Wilson, P.D., 2019. Restore and Renew: a genomics-era framework for species provenance delimitation. Restoration Ecology 27(3): 538 – 548.

Scott, L.J., Graham, G.C., Hannan-Jones, M., Yeates, D.K., 1997. DNA Profiling resolves the limited importance of flower colour in defining varieties of Lantana camara. Electrophoresis 18: 1560-1563.

Watts, R., 2009. Testing the hybrid swarm model for weedy lantana. Unpublished report.

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie C., Weir, B., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data." Bioinformatics 28(24): 3326-3328.