Selection of DNA barcoding regions for identification and genetic analysis of two *Echium* invaders in Australia: *E. plantagineum* and *E. vulgare*

Xiaocheng Zhu[1], Lucie Meyer[2], David Gopurenko[1,3], Paul A. Weston[1], Geoff M. Gurr[4], Ragan M. Callaway[5], Brendan J. Lepschi[6] and Leslie A. Weston[1]

[1] Graham Centre for Agricultural Innovation (NSW Department of Primary Industries and Charles Sturt University), Locked Bag 588, Wagga Wagga, NSW 2678, Australia

[2] AgroSup Dijon, 26 boulevard Petitjean, BP 87999, 21079 – Dijon Cedex, France

[3] NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, PMB, Wagga Wagga, NSW 2650, Australia

[4] Graham Centre for Agricultural Innovation (NSW Department of Primary Industries and Charles Sturt University), PO Box 883, Orange, NSW 2800, Australia

[5] Division of Biological Sciences, University of Montana, Missoula, Montana, MT 59812, USA

[6] Australian National Herbarium, Centre for Australian National Biodiversity Research, GPO Box 1600, Canberra, ACT 2601, Australia

(xzhu@csu.edu.au)

Summary *Echium plantagineum* L. (Paterson's curse) is an exotic noxious weed in Australia which infests over 30 million ha and costs the Australian meat and wool industries over $125 million annually. The species is morphologically similar to *E. vulgare* L. (viper's bugloss), a closely related exotic weed also in Australia. This study aims to find useful DNA barcoding regions for distinguishing these two species and for investigating their population genetic diversity. Levels of polymorphism at six DNA barcoding regions were evaluated in *E. vulgare* (N = 15) and *E. plantagineum* (N = 25) sampled in Australia. PCR amplifications were successful in both species at one nuclear region (ITS) and three chloroplast regions (*trnH-psbA* spacer, *trnL* intron and *trnL-trnF* spacer). Minimum pair-wise sequence differences between the two species were 6.50%, 3.78%, 0.84% and 1.01% at ITS, *trnH-psbA*, *trnL* intron and *trnL-trnF*, respectively. The plastid gene regions were also useful for investigating intra-specific genetic diversity. Six, three and five haplotypes were detected among the 25 *E. plantagineum* samples at *trnH-psbA*, *trnL* intron and *trnL-trnF*, with maximum intra-specific pair-wise sequences differences of 2.32%, 0.42% and 0.76%, respectively. In addition, two haplotypes were detected in *E. vulgare* at each of the three cpDNA regions. This study resulted in development of a genetic method to reliably distinguish *E. vulgare* from *E. plantagineum* in Australia and for further investigation and assessment of genetic diversity within these two species, which is critical for identification and management of *E. plantagineum*.

Keywords Paterson's curse, salvation Jane, viper's bugloss, noxious weed, nuclear genes and chloroplast genes.

INTRODUCTION

Echium plantagineum and *E. vulgare* are noxious exotic weed species in Australia. They were first introduced to Australia in the early 1800s. The two species are morphologically similar and share many life history traits for reproduction and gamete dispersal. Both species are drought tolerant, but *E. plantagineum* occurs in all states of Australia (Figure 1) and cost over $125 million annually, while *E. vulgare* is restricted to south-eastern Australian states of South Australia, New South Wales, Victoria and Tasmania (Figure 1). Reliable identification of these two species at a seedling growth stage is necessary to ensure efficacious weed management. As herbicide control of *E. plantagineum* is more effective at the rosette stage.

Current information on genetic diversity of *E. plantagineum* in Australia is very limited. Burdon and Brown (1986) compared four Australian and two European populations using isozyme markers and highlighted the similar level of genetic diversity between the Australian and European *E. plantagineum*. However, isozyme polymorphisms detected within specimens can vary among tissues, growth stages and environments. Furthermore, isozyme analyses can be confounded by preservation methods used during specimen collection and preparation. Therefore, more reproducible markers for targeted sequencing of key gene regions are required to reliably evaluate the genetic diversity within this species.

DNA barcoding is a sequence based species identification method which can assist alpha-taxonomic discovery and diagnostics (Hebert *et al*. 2003). A combination of several plastid DNA barcode regions is typically recommended for plant species identifications, including *rbcL* + *matK* (CBOL Plant Working Group 2009) and *trnH-psbA* + *rbcL* (Kress and

Erickson 2007). In addition, nuclear genes such as internal transcribed spacer (ITS) regions of the nuclear ribosomal cistron (18S-5.8S-26S) can provide species identifications independent of that detected from the chloroplast genome.

Phylogenetic relationships within the genus *Echium* have been studied using DNA sequence information (Böhle *et al.* 1996, Romeiras *et al.* 2011). However, their sample replication at each species was limited to a few specimens sampled from their native ranges. Incomplete sampling can underestimate the level of genetic diversity within a species and lead to erroneous DNA barcode species identifications in cases where species share genetic lineages through ancestry or homoplasy (Liu *et al.* 2012). This study aims to identify useful DNA barcoding regions for distinguishing these two *Echium* species and for investigating their population genetic diversity. Levels of intra- and inter-specific polymorphism at six DNA barcoding regions were evaluated in *E. vulgare* (N = 15) and *E. plantagineum* (N = 25) sampled in Australia. The DNA barcode regions identified will be further used to determine intra- and inter-specific genetic diversity for our ongoing investigation of *Echium* taxonomy.

MATERIALS AND METHODS

Twenty five *E. plantagineum* and 15 *E. vulgare* individuals were collected from 10 and eight geographically distinct sites in Australia, including taxonomically identified reference specimens provided by the Australian National Herbarium for *E. plantagineum* and *E. vulgare* (two and four specimens, respectively). Young leaf tissue was collected individually and placed in 70% ethanol until extraction.

Genomic DNA isolation, PCR, sequencing and alignment procedures were performed according to Gopurenko *et al.* (2013), with adjustment to PCR conditions to accommodate for different $MgCl_2$ concentrations and annealing temperatures (Table 1). One nuclear DNA region and five chloroplast DNA (cpDNA) regions were evaluated using ten primer combinations (Table 1). Aligned sequences were analysed using MEGA6 to calculate intra- and inter-specific distances using the Kimura 2-parameter (K2P) model. Indels and ambiguous positions were excluded from analyses.

RESULTS AND DISCUSSION

PCR and sequencing was 100% successful in both species at four selected gene regions which included ITS, *trnH-psbA* spacer, *trnL* intron and *trnL-trnF* spacer (Table 1), and based on these results, these regions were selected for further study. Although

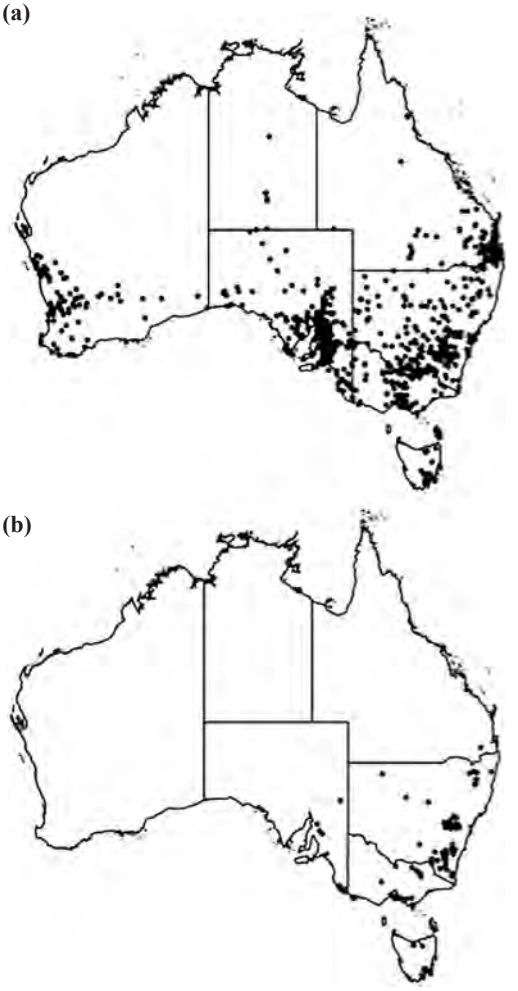

Figure 1. Distribution of *Echium plantagineum* (a) and *E. vulgare* (b). Source: Australia's Virtual Herbarium, 2013.

rbcL+matK are two cpDNA loci often recommended for DNA barcoding of plants (CBOL Plant Working Group 2009), recent studies have questioned their discriminatory power across the broad spectrum of plant species (Zhang *et al.* 2013). We did not include the *rbcL* gene in our research because very limited inter-specific divergence (three bases differences out of 1196 bp) was found between *E. plantagineum* and *E. vulgare rbcL* sequence accessions in GenBank. The *matK* gene was excluded because of PCR failure with *E. vulgare* when using the two *matK* primer pairs. Primer ucp_a and ucp_b were reported to amplify *trnT-trnL* in *E. vulgare* (Böhle *et al.* 1996), but failed in our study (Table 1).

Table 1. Primer pairs used for amplification in *Echium plantagineum* and *E. vulgare*.

Target region	Forward primer	Reverse primer	Size of amplicons (bp)		T_A (°C)	$MgCl_2$ (mM)
			E. plantagineum	*E. vulgare*		
ITS	ITSL	ITSR	+	+	54	3
	ITS5A	ITS4	−	−	45−65*	2.5
	ITS5A	ITSR	740	741	60	3
	ITSL	ITS4	+	+	54	3
matK	390F	1326R	−	−	45−65*	3
	3Fx_KIM	1Rx_KIM	+	−	45−65*	3
trnH-psbA	trnH	psbA	323	240	62	2.2
trnT-trnL spacer	ucp_a	ucp_b	+	−	45−65*	3
trnL intron	ucp_c	ucp_d	549–553	556	58	2.2
trnL-trnF spacer	ucp_e	ucp_f	441	441	64	2.2

+ refers to positive amplification but not sequenced;
− refers to negative amplification;
T_A: annealing temperature;
* ranges of annealing temperature were evaluated if no amplification was detected.

Forty sequences were obtained from each of the four regions (Table 2). Alignments were truncated to 636, 273, 494 and 399 bp sequences for ITS, *trnH-psbA*, *trnL* intron and *trnL-trnF* spacer, respectively. There were 41 parsimony informative sites, three ambiguous positions and a single base pair indel in the ITS alignment. For the cpDNA regions, 14, eight and seven parsimony informative sites were found in *trnH-psbA*, *trnL* and *trnL-trnF*, respectively. There were eight indels in the *trnH-psbA* alignment, and these ranged from 1–31 bp, eight indels of 1–5 bp in length in the *trnL* alignment and no indels in the *trnL-trnF* alignment. A four-base inversion region was observed among specimens in the *trnH-psbA* alignment. In addition, there were two continuous mononucleotide repeats (A and T, both ranging from 9–12 bp) from position 276 to 300 in the *trnL* alignment.

Two haplotypes were found in ITS regions for each species at ambiguous positions. Five, six and three haplotypes were found in *trnH-psbA*, *trnL* and *trnL-trnF* in *E. plantagineum* with maximum intra-specific distances of 2.32%, 0.42% and 0.76%, respectively. For *E. vulgare*, two haplotypes were found in each of the three cpDNA regions. The maximum intra-specific distance of *E. vulgare* in *trnH-psbA*, *trnL* and *trnL-trnF* was 1.84%, 0.63% and 0.25%, respectively.

Minimum K2P sequence differences between *E. plantagineum* and *E. vulgare* were 6.50%, 3.78%, 0.84% and 1.01% for ITS, *trnH-psbA*, *trnL* and *trnL-trnF*, respectively. Minimum inter-specific distances

Table 2. Evaluation of four DNA regions in *Echium plantagineum* (EP) and *E. vulgare* (EV).

	ITS	*trnH-psbA*	*trnL*	*trnL-trnF*
Aligned length (bp)	636	273	494	399
No. haplotype in EP	2	5	6	3
No. haplotype in EV	2	2	2	2
No. indel between species	1	7	3	0
No. indel within EP	0	1	5	0
No. indel within EV	0	0	5	0
No. informative sites	41	14	8	7
Minimum inter-specific distance %	6.50	3.78	0.84	1.01
Maximum intra-specific distance in EP %	0.16	2.32	0.42	0.76
Maximum intra-specific distance in EV %	0.31	1.84	0.63	0.25

for all regions were higher than maximum intra-specific distances (Table 2). In addition, no overlap between intra- and inter-specific genetic distances was found in any of the four regions evaluated, suggesting the presence of DNA barcode gaps (Figure 2), useful for the separation and identification of Australian *E. plantagineum* and *E. vulgare* at all four genes evaluated.

Our results also indicate the four sequence regions (ITS, *trnH-psbA*, *trnL* and *trnL-trnF*) have high levels of intra-specific difference and may be useful regions for investigation of genetic diversity within populations of Australian *E. plantagineum* and *E. vulgare*.

In conclusion, our studies have highlighted one nuclear and three cpDNA barcoding regions that are reliable for distinguishing *E. plantagineum* and *E. vulgare*, as well as genetic diversity within each species. Future studies will focus on the phylogenetic relationship of *E. plantagineum* and other *Echium* species in Australia and also in their native range. Such research will further our understanding of the invasiveness of *E. plantagineum*, help to pin point respective sites of origin of related *Echium* species in the Iberian Peninsula, and assist in potential future selection of biocontrol agents for successful management in Australia.

ACKNOWLEDGMENTS

We acknowledge the Australian Research Council Discovery Program for project support and Dominik Skoneczny, Shane Hildebrand and Brigette Ryan for sampling assistance.

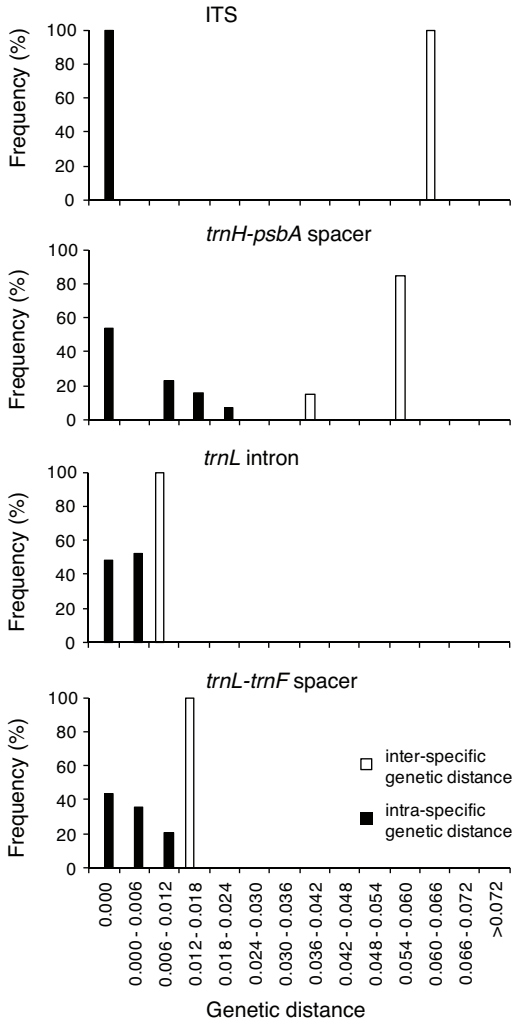

Figure 2. Frequency distribution of intra- and inter-specific genetic distances of that ITS, *trnH-psbA*, *trnL* and *trnL-trnF* regions showing the presence of DNA barcode gap.

REFERENCES

Australia's Virtual Herbarium (2014). Australia's Virtual Herbarium. http://avh.ala.org.au/ occurrences/search?taxa=Echium+plantagineum#tab_mapView (viewed 10 May 2014).

Böhle, U.R., Hilger, H.H. and Martin, W.F. (1996). Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proceedings of the National Academy of Sciences of the United States of America* 93, 11740-5.

Burdon, J.J. and Brown, A.H.D. (1986). Population genetics of *Echium plantagineum* L.-target weed for biological control. *Australian Journal of Biological Sciences* 39, 369-78.

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106, 12794-7.

Gopurenko, D., Fletcher, M.J., Löcker, H. and Mitchell A. (2013). Morphological and DNA barcode species identifications of leafhoppers, planthoppers and treehoppers (Hemiptera: Auchenorrhyncha) at Barrow Island. *In* 'Records of the Western Australian Museum, supplement 83: the terrestrial invertebrate fauna of Barrow Island', eds N.R. Gunawardene, J.D. Majer, C.K. Taylor *et al.*, pp. 253-85. (Western Australian Museum Publications).

Hebert, P.D.N., Cywinska, A., Ball, S.L. and DeWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences* 270, 313-21.

Kress, W.J. and Erickson, D.L. (2007). A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *Plos One* 2.

Liu, J., Provan, J., Gao, L.M. and Li, D.Z. (2012). Sampling strategy and potential utility of indels for DNA barcoding of closely related plant species: a case study in taxus. *International Journal of Molecular Sciences* 13, 8740-51.

Romeiras, M.M., Paulo, O.S., Duarte, M.C. Pina-Martins, F., Cotrim, M.H., Carine, M.A. and Pais, M.S. (2011). Origin and diversification of the genus *Echium* (Boraginaceae) in the Cape Verde archipelago. *Taxon* 60, 1375-85.

Zhang, W., Fan, X.H., Zhu, S.F., Zhao, H. and Fu, L.Z. (2013). Species-specific identification from incomplete sampling: applying DNA barcodes to monitoring invasive *Solanum* plants. *Plos One* 8.