Transcriptome Sequencing and Simple Sequence Repeat Marker Development for Three Macaronesian Endemic Plant Species

Authors: White, Oliver W., Doo, Bethany, Carine, Mark A., and Chapman, Mark A.

Source: Applications in Plant Sciences, 4(8)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600050

Applications in Plant Sciences 2016 4(8): 1600050

Genomic Resources Note

Transcriptome sequencing and simple sequence repeat marker development for three Macaronesian endemic plant species[1]

Oliver W. White[2,3], Bethany Doo[3], Mark A. Carine[2], and Mark A. Chapman[3,4]

[2]Plants Division, Natural History Museum, London SW7 5BD, United Kingdom; and [3]Centre for Biological Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

- *Premise of the study:* Oceanic islands offer unparalleled opportunities to investigate evolutionary processes such as adaptation and speciation. However, few genomic resources are available for oceanic island endemics. In this study, we publish transcriptome sequences from three Macaronesian endemic plant species (*Argyranthemum broussonetii* [Asteraceae], *Descurainia bourgaeana* [Brassicaceae], and *Echium wildpretii* [Boraginaceae]) that are representative of lineages that have radiated in the region. In addition, the utility of transcriptome data for marker development is demonstrated.
- *Methods and Results:* Transcriptomes from the three plant species were sequenced, assembled, and annotated. Between 1972 and 2282 simple sequence repeats (SSRs) were identified for each taxon. Primers were designed and tested for 30 of the candidate SSRs identified in *Argyranthemum*, of which 12 amplified well across three species and eight were polymorphic.
- *Conclusions:* We demonstrate here that a single transcriptome sequence is sufficient to identify hundreds of polymorphic SSR markers. The SSRs are applicable to a wide range of questions relating to the evolution of island lineages.

Key words: *Argyranthemum broussonetii*; *Descurainia bourgaeana*; *Echium wildpretii*; Macaronesia; marker development; simple sequence repeats (SSRs); transcriptomics.

The availability of next-generation sequencing (NGS) technology for nonmodel organisms in prime ecological scenarios has revolutionized evolutionary biology (Egan et al., 2012). An exciting prospect of NGS is the potential to improve our understanding of the genetic basis of processes such as adaptation and speciation (Stapley et al., 2010; Kelley et al., 2012; Chapman et al., 2013; Sousa and Hey, 2013; Rius et al., 2015; Twyford et al., 2015). Volcanic oceanic islands have long served as model systems for the study of such evolutionary processes (Emerson, 2002); however, the capabilities of NGS for oceanic island endemics are only starting to be realized (Kueffer et al., 2014). For example, NGS approaches have been employed to investigate the radiation of Darwin's finches from the Galápagos Islands (Lamichhaney et al., 2015) and to untangle the complex phylogenetic relationships of *Tolpis* Adans., a genus of flowering plants from Macaronesia (Mort et al., 2015). NGS has also been successfully applied to other "island-like" scenarios such as the radiation of cichlid fishes in African and Neotropical lakes (Fan et al., 2012).

While NGS is becoming more affordable, the cost of obtaining genome-level sequences from multiple individuals or population sampling is still high. However, a large genetic resource from just a single or a few individuals (e.g., a transcriptome sequence or an expressed sequence tag [EST] library) offers the ability to produce highly cost-effective PCR-based molecular markers that can be amplified in many individuals at a fraction of the cost (Ellis and Burke, 2007). To generate further interest in this area and to develop a novel genetic resource, we have sequenced and assembled transcriptome sequences for three plant species (*Argyranthemum broussonetii* (Pers.) Humphries [Asteraceae], *Descurainia bourgaeana* (E. Fourn.) Webb ex O. E. Schulz [Brassicaceae], and *Echium wildpretii* H. Pearson ex Hook. f. [Boraginaceae]) that belong to three endemic radiations of Macaronesia (Fig. 1).

Isolated oceanic archipelagos are botanically diverse and rich in endemic species, making them ideal systems to investigate the origin and evolution of plant diversity (Losos and Ricklefs, 2009; Kueffer et al., 2014). These taxa have been selected because they belong to genera that offer exceptional "natural laboratories" in the Macaronesian archipelagos with which to investigate a range of evolutionary phenomena.

Argyranthemum Webb is the largest endemic genus found in Macaronesia with at least 23 species (Fig. 2A) (Humphries, 1976, 1979; Francisco-Ortega et al., 1996), including a rare putative example of homoploid hybrid speciation (Brochmann et al., 2000; Fjellheim et al., 2009). There are seven species of *Descurainia* Webb & Berthel. endemic to the Canary Islands where they exhibit multiple independent adaptations to high-altitude habitats

[1]Manuscript received 22 April 2016; revision accepted 5 July 2016.

The authors thank Alfredo Reyes-Betancort for his guidance throughout fieldwork in the Canary Islands and Bonn Botanic Gardens (Bonn, Germany), and the Millennium Seed Bank Partnership (Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom) for the kind donation of seed material. We also thank Alex Harkess (University of Georgia, Athens, Georgia, USA) for advice on the NGS library prep, and Alison Baylay and Tom Bibby (National Oceanography Centre, Southampton, United Kingdom) for running the MiSeq. O.W.W. is funded by a Natural History Museum–University of Southampton studentship. Three reviewers provided insightful comments on an earlier version of the manuscript.

[4]Author for correspondence: m.chapman@soton.ac.uk

doi:10.3732/apps.1600050

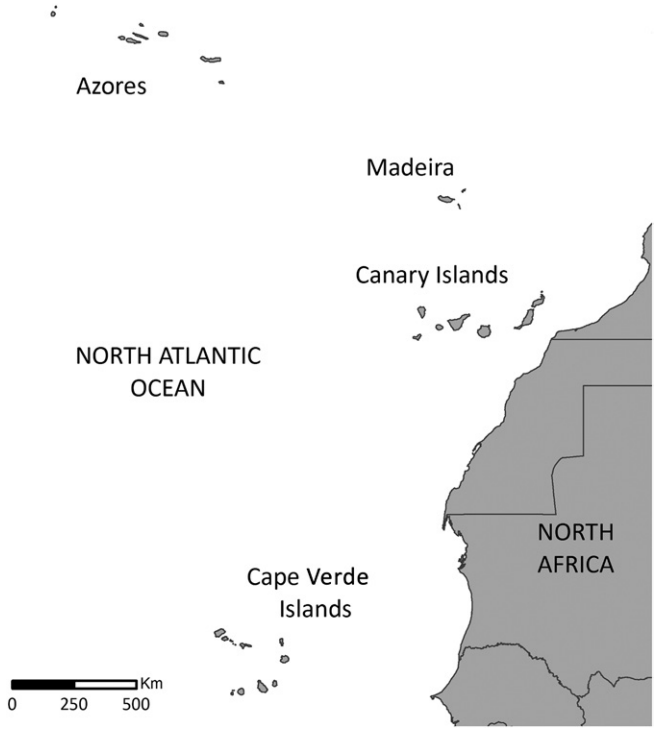

Fig. 1. The Macaronesian archipelagos in the North Atlantic Ocean.

(Fig. 2B) (Goodson et al., 2006). A total of 27 *Echium* L. species are endemic to Macaronesia where they occur in a wide range of habitats and exhibit conspicuous differences in morphology, with annual herbs, candelabra shrubs, and monocarpic rosettes all represented (Fig. 2C) (Böhle et al., 1996).

Transcriptome data for Macaronesian endemic taxa could be used in a number of ways to assist evolutionary studies. Phylogenetic analyses of Macaronesian endemic lineages to date have often lacked sufficient resolution to interpret patterns of evolution due to a lack of genetic variation (Mort et al., 2015), likely a result of the recent and rapid nature of island radiations (Francisco-Ortega et al., 1997) and/or reliance upon commonly used universal molecular markers such as the internal transcribed spacer (ITS) (Sun et al., 1994) or noncoding chloroplast regions (Shaw et al., 2007). Comparative analysis of transcriptome sequences can be used to identify universal markers that can be widely amplified across taxa, but still exhibit variation within taxa, facilitating phylogenetic reconstruction of poorly resolved groups (Chapman et al., 2007; Chamala et al., 2015). Annotated transcriptomes are also useful for the identification of specific genes of interest.

Transcriptome sequences can also be mined for microsatellites or simple sequence repeats (SSRs). SSRs are advantageous over other PCR-based markers because they are codominant, often highly polymorphic, and are transferable to closely related species (Ellis and Burke, 2007). SSRs can be used to investigate genetic diversity, assess population structure and gene flow, and inform conservation strategies (Ellis and Burke, 2007). The development of SSRs using traditional methods is costly and time consuming, but more recent approaches have involved the development of SSRs from EST databases (Ellis and Burke, 2007) and transcriptome sequences (Wang et al., 2013; Chapman, 2015).

For this study, we sequenced and annotated transcriptomes for three Macaronesian endemic plant species and focused on the potential application of transcriptome resources for the identification of SSR loci. As proof of concept, we designed and trialed primers for 30 SSR loci in *Argyranthemum*. All of the resultant data, including transcriptome assemblies, BLAST search results, and SSR loci, are available from the corresponding author (upon request) as a resource for future genetic studies in these Macaronesian endemic lineages.

METHODS AND RESULTS

Cypselae (single-seeded fruit) of *A. broussonetii* were collected from Barranco de Valle Crispín in the Anaga peninsula of Tenerife during June 2015 (collected under a permit from the Cabildo de Tenerife, no. 18297). Seeds of *E. wildpretii* and *D. bourgaeana* were sourced from Bonn Botanic Gardens (Bonn, Germany) and the Millennium Seed Bank Partnership (Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom), respectively. These were soaked overnight in 0.5 mg mL^{-1} gibberellic acid, then rinsed and placed on damp filter paper in Petri dishes at 22°C until germination. Germinated seeds were transferred to a 2 : 1 mixture of Levington's F2+S and vermiculite in a greenhouse with 16-h days, supplemented with artificial light.

RNA extraction was carried out from true leaves of seedlings using a QIAGEN RNeasy Kit (QIAGEN, Crawley, United Kingdom) following the manufacturer's protocol, with on-column DNase digestion (RNase-free DNase, QIAGEN). One microgram of RNA was prepped for sequencing using the KAPA Stranded RNA-Seq Library Preparation Kit (Kapa Biosystems, Wilmington, Massachusetts, USA) with unique adapters for each sample to allow de-convolution of reads. Library amplification was carried out with seven cycles of PCR. The three samples were combined and sequenced on ~3/4 lane of an Illumina MiSeq (Illumina, San Diego, California, USA) with 300-bp paired-end reads at the National Oceanography Centre, University of Southampton, United Kingdom.

Between 3.5 and 3.9 million paired-end 100-bp reads were generated for each of the samples (Table 1). Reads have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID PRJNA324223), and the FASTA-formatted transcriptome sequences are available from the authors upon request.

The following steps were carried out on the raw transcriptome sequences for each of the Canary Island endemics separately. Poor quality sequence data and adapter sequences were trimmed with Trimmomatic (Bolger et al., 2014). Parameters included Illumina clip with seed mismatches 2, palindrome clip threshold 30, simple clip threshold 10, minimum adapter length 8, keep both reads equals TRUE, leading quality and trailing quality 5, sliding window trimming with a window size 4, required quality 15, and minimum read length of 36. The resulting reads were used to create a de novo transcriptome assembly with Trinity (version 2.0.6; Grabherr et al., 2011). Libraries were normalized to a k-mer coverage of 30 so as to reduce computation time, and then assembled using Trinity with the settings –min_kmer_cov 2 to increase the stringency for reads being assembled together, and –max_diffs_same_path 4, –max_internal_gap_same_path 15, which allowed for more divergent reads (up to four nucleotide differences and up to a 15-bp gap) to be assembled into the same transcript. This takes into account the likely heterozygosity of the species.

Although the number of raw and normalized reads from each of these species was relatively similar, de novo assembly of these sequences resulted in transcriptomes with different assembly characteristics. *Argyranthemum broussonetii* had the largest number of transcripts (80,620 genes and 94,522 transcripts) and *D. bourgaeana* the fewest (44,287 genes and 54,221 transcripts). *Echium wildpretii* was intermediate, comprising 58,526 genes and 69,509 transcripts (Table 1, Fig. 3). Despite the *A. broussonetii* transcriptome comprising more transcripts, transcript length was generally shorter, whereas the transcriptome of *D. bourgaeana* had longer transcripts, with *E. wildpretii* intermediate (Table 1, Fig. 3). The larger number of contigs in *Argyranthemum* could indicate the transcripts are more fragmented, but this could also reflect the palaeopolyploidy of this group of plants.

BLAST (Altschul et al., 1990) was used to compare each of the assembled transcriptomes with annotated coding sequences of *Arabidopsis thaliana* (L.) Heynh. and *Solanum lycopersicum* L. using a local BLASTX search and peptide sequences downloaded from The *Arabidopsis* Information Resource (https://www.arabidopsis.org/) and the Tomato Genome Sequencing Project (ftp://ftp.solgenomics.net/tomato_genome). BLAST parameters included an expectation

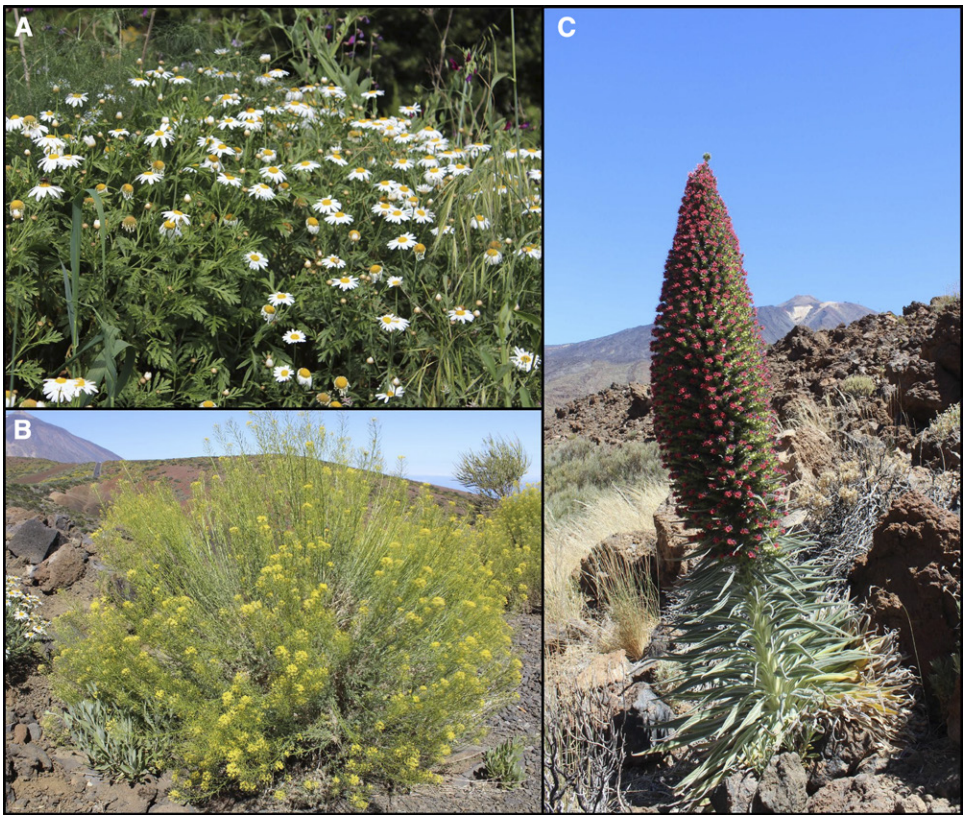

Fig. 2. Macaronesian endemics *Argyranthemum* sp. (A), *Descurainia bourgaeana* (B), and *Echium wildpretii* (C). Photos taken by O. White.

value (*E*-value) of 1.0×10^{-20} and alignment length greater than or equal to 50. One hit per transcript is reported (Appendix S1).

A substantial proportion of the contigs in our transcriptomes matched with an annotated coding sequence from either *A. thaliana* or *S. lycopersicum* (Table 1). This is especially true of the comparison between our *Descurainia* transcriptome and *A. thaliana*. Of the 54,221 transcripts in the assembled *Descurainia* transcriptome, 43,329 (80%) matched an annotated sequence from *A. thaliana*. In addition, 16,645 *A. thaliana* genes (50% of the estimated number of genes) were recovered in the *Descurainia* transcriptome. The proportions of transcripts that matched an annotated coding sequence in the other BLAST search combinations were more moderate, ranging from 39–63% of transcripts matching with an annotated sequence and 36–38% of genes being recovered from either *A. thaliana* or *S. lycopersicum* (Table 1).

SSRs were identified in each of the assembled transcriptomes (Appendix S1) using MISA (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/misa.html). Minimum repeat numbers for di-, tri-, and tetranucleotide repeat markers were eight, six, and four.

A large number of microsatellites were identified in each of the transcriptomes (Appendix S1): 2282, 1972, and 1284 in *A. broussonetii*, *D. bourgaeana*, and *E. wildpretii*, respectively. Several of these were within the first or last 50 bp of the transcript, were in compound formation, or there was more than one SSR per transcript. After removal of these, the number of SSRs was 1288, 1219, and 737 for *A. broussonetii*, *D. bourgaeana*, and *E. wildpretii*, respectively. The number of trinucleotide repeat markers was notably higher than di- or tetranucleotide markers, typical of SSRs identified in coding regions from EST libraries/transcriptomes.

Primers were designed for 30 SSRs identified in *A. broussonetii* using Primer3 (Untergasser et al., 2012). Contigs were avoided if (1) the SSR was in the first or last 50 bp (ensuring there was sufficient space for primer design), (2) the SSR was in a compound formation, and (3) there was more than one SSR present in the locus. Longer SSRs were preferentially chosen because a number of studies have suggested that longer SSRs are more likely to be polymorphic (Burstin et al., 2001; Mun et al., 2006; Wang et al., 2012). The 30 loci (Appendix S2) were screened across 10 DNA samples from four members of the genus *Argyranthemum*, including *A. frutescens* (L.) Sch. Bip. subsp. *frutescens*, *A. frutescens* subsp. *succulentum* Humphries, *A. broussonetii* subsp. *broussonetii*, and *A. tenerifae* Humphries. These taxa were selected to test how broadly our SSR loci can be used effectively, from closely related taxa such as the *A. frutescens* subspecies to more distantly related species such as *A. broussonetii* and *A. tenerifae*.

DNA was extracted from silica gel–dried leaf material of *Argyranthemum* using a modified cetyltrimethylammonium bromide (CTAB)–based method (Doyle and Doyle, 1987). For PCR, the forward primers were designed with the M13(–29) sequence (CACGACGTTGTAAAACGAC) appended to the 5′ end such that a third fluorescently labeled M13(–29) primer (either TET or FAM) could be incorporated in the PCR (Schuelke, 2000). Each PCR contained 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 0.01% Tween 20, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.04 µM forward primer, 0.2 µM reverse primer, 0.2 µM fluorescent primer, 1 unit of *Taq* DNA polymerase, 15 ng DNA, and was made up to 15 µL with water. PCR conditions consisted of an initial denaturation (94°C for 3 min); 10 touchdown cycles of 94°C for 30 s, 65°C for 30 s (decreasing by 1°C per cycle), and 72°C for 60 s; 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s; and a final elongation (72°C for 7 min). Amplification success was assessed by running PCR products on 1% agarose gels stained with GelRed (Biotium, Hayward, California, USA).

Primers that produced an apparently single product between 100 and 400 bp were selected for further analysis. PCR products were diluted 1 : 30 and combined in such a way that multiple loci (differing in size and/or fluorescent label) could be resolved in a single lane on an ABI3730xl (Applied Biosystems, Carlsbad, California, USA) at the Department of Zoology, University of Oxford, United Kingdom. Alleles were scored from the raw traces using GeneMarker 2.6.7 (SoftGenetics, State College, Pennsylvania, USA), and we performed a principal coordinates analysis (PCoA) of the 10 *Argyranthemum* samples based on polymorphic loci using GenAlEx (Fig. 4) (Peakall and Smouse, 2012).

All 30 SSR primer pairs tested in *Argyranthemum* produced a PCR product; however, only 12 primer pairs produced an apparently single PCR product between 100 and 400 bp. Of these, eight were polymorphic in the 10 DNA samples, one was monomorphic, and three produced multiple nonspecific PCR products. Although our sample number is small, PCoA of SSR loci shows that these markers are able to differentiate between *A. frutescens* and *A. broussonetii* as well as differentiate the subspecies of *A. frutescens* (Fig. 4). The position of *A. tenerifae* is not clear, as it appears to overlap with *A. frutescens* subsp. *frutescens*. With

TABLE 1. Summary statistics for the de novo assembled transcriptomes, BLASTX searches, and simple sequence repeat identification.

Statistic	*Argyranthemum broussonetii*	*Descurainia bourgaeana*	*Echium wildpretii*
Raw reads			
No. of raw reads	3,950,166	3,958,226	3,550,703
No. of trimmed reads	3,842,373	3,774,119	3,375,554
No. of normalized reads	1,471,858	1,234,637	1,228,003
Normalized reads, %	38	33	36
Assembly statistics			
No. of genes	80,620	44,287	58,526
No. of transcripts	94,522	54,221	69,509
N50 all transcripts, bp	921	1233	1041
Median length all transcripts, bp	411	549	446
Average length all transcripts, bp	654	833	707
Total no. of assembled bases	61,826,463	45,213,744	49,160,261
BLASTX analyses[a]			
A. thaliana BLAST hits	37,265	43,329	37,060
A. thaliana BLAST hits, %	39	80	53
A. thaliana genes recovered	12,235	16,645	12,072
A. thaliana genes recovered, %[b]	36	50	36
S. lycopersicum BLAST hits	39,521	34,042	39,758
S. lycopersicum BLAST hits, %	42	63	57
S. lycopersicum genes recovered	13,139	12,534	13,101
S. lycopersicum genes recovered, %[b]	38	36	38
MISA statistics			
Total no. of identified SSRs	2282	1972	1284
Total no. of SSR-containing transcripts	2232	1919	1251
No. of transcripts with >1 SSR	50	51	31
No. of SSRs in compound formation	68	34	31
No. of dinucleotide SSRs	478	604	230
No. of trinucleotide SSRs	1097	1098	799
No. of tetranucleotide SSRs	589	183	191
No. of SSR loci suitable for primer design[c]	1288	1219	737

[a] BLAST parameters included an expectation value (*E*-value) of 1.0×10^{-20} or less.
[b] Proportion of the *A. thaliana* (33,602) or *S. lycopersicum* (34,727) annotated coding sequences recovered from the target species.
[c] SSR loci were excluded if they were located in the first or last 50 bp of the contig, if SSRs were in compound formation, or if there were multiple SSR loci per contig.

the small sample sizes of all the taxa it is hard to explain this; however, this could indicate these species are more closely related than previously discussed, or that some hybridization between these taxa has been uncovered.

CONCLUSIONS

This generation of transcriptome sequences from Macaronesian endemic species of flowering plants are the first resources of their kind for these archipelagos as far as we are aware. The de novo assembled transcriptomes have recovered a considerable portion of the expressed genes as indicated by our BLAST comparisons with *A. thaliana* and *S. lycopersicum* (Table 1) and 1200 to 2200 SSRs. The results of the microsatellite screen in *Argyranthemum* revealed that eight out of 30 markers (27%) were polymorphic and easy to score. Assuming this is representative of the entire transcriptome and across island lineages, we might expect that we are able to amplify and score ~350, ~330, and ~200 polymorphic SSRs in *Argyranthemum*, *Descurainia*, and *Echium*, respectively.

Previous studies of rapidly radiating lineages, such as those represented by the three species we focused on in this study, have been hampered by a lack of genetic variation between taxa using sequence-based "universal" markers (Böhle et al., 1996; Francisco-Ortega et al., 1997; Goodson et al., 2006). The markers we have identified will allow much more fine-scale resolution of evolutionary processes in these lineages. It is encouraging to find that eight SSR loci were able to differentiate *A. broussonetii* and both subspecies of *A. frutescens* (Fig. 4). The lack of genetic differentiation between *A. frutescens* subsp. *frutescens* and *A. tenerifae* may be due to our low sample size, but this is an area of study that warrants further study. Indeed, the sister relationships of this species are not clear on the basis of the chloroplast restriction site phylogeny by Francisco-Ortega et al. (1996).

Other recent studies have also used transcriptome sequences to identify SSR loci, with similar success rates. Wang et al. (2013) used a transcriptome from *Chrysanthemum nankingense* Hand.-Mazz. to identify 2813 putative SSR loci, with approximately 20% of the 100 tested showing polymorphism. Chapman (2015) also used transcriptome sequences to identify between 1139 and 2567 SSR loci for each of four underutilized legumes. In a follow-up study, 36 primer pairs were designed for one of these species, of which six (17%) were found to be polymorphic in a small number of accessions (Robotham and Chapman, 2016).

For our study, we only attempted to amplify and genotype a small number of SSR loci, and we did not try to optimize the PCR for markers that amplified nonspecific products. In addition, the settings and programs used in the SSR discovery are likely to resolve more or less markers depending on how strictly one sets the parameters. Nevertheless, the identification of SSR loci using transcriptome sequences is more cost effective, less time consuming, and methodically straightforward compared to earlier methods (Ellis and Burke, 2007).

Previous studies have used NGS technology to identify SSRs from oceanic island endemic plants. Takayama et al. (2013, 2015) have developed and used 10 SSR markers for *Robinsonia* DC. (Asteraceae), endemic to the Juan Fernández Archipelago, to

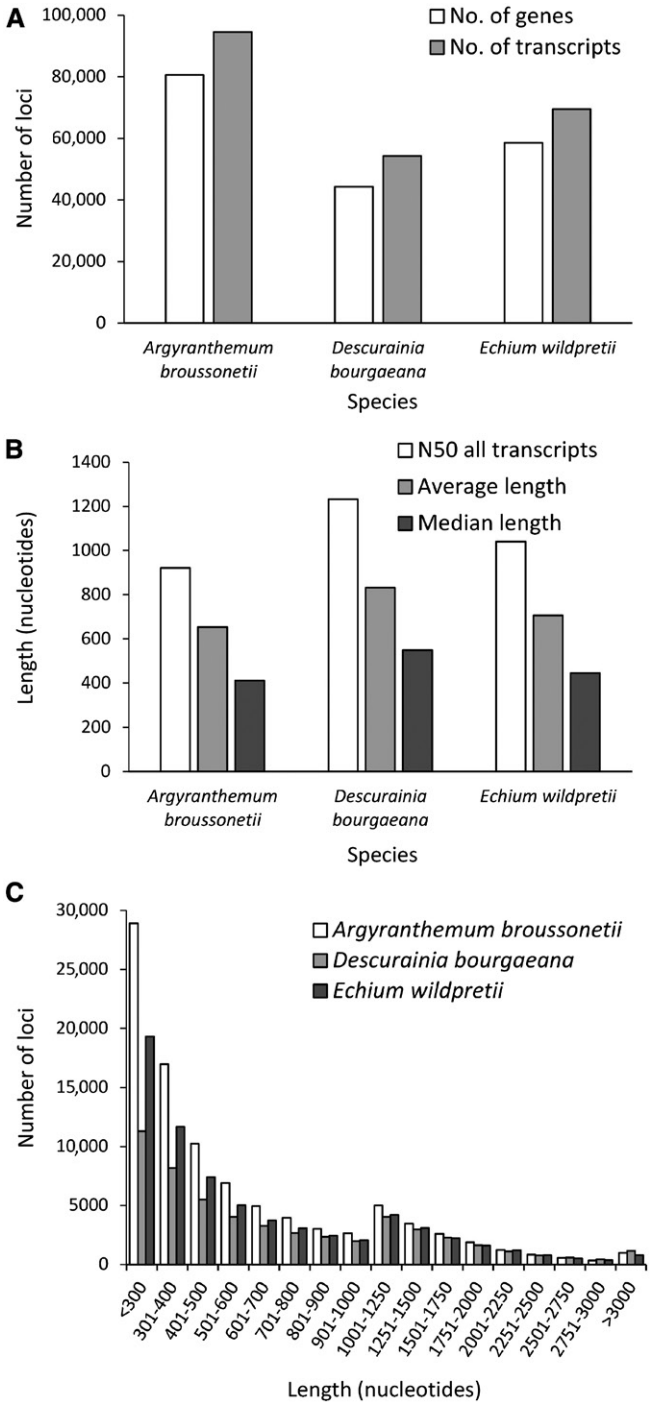

Fig. 3. Summary statistics for the three de novo transcriptome assemblies. (A) Number of genes and transcripts assembled for each species. (B) N50, mean, and median transcript length. (C) Transcript lengths for the three transcriptomes (note the change in bin size along the x axis).

assess population structure and evolutionary relationships. The SSRs identified in the current study can be harnessed in a similar way for a range of evolutionary studies, ranging from homoploid hybrid speciation in *Argyranthemum*, adaptation to high altitude in *Descurainia*, and the evolution of monocarpy in *Echium*.

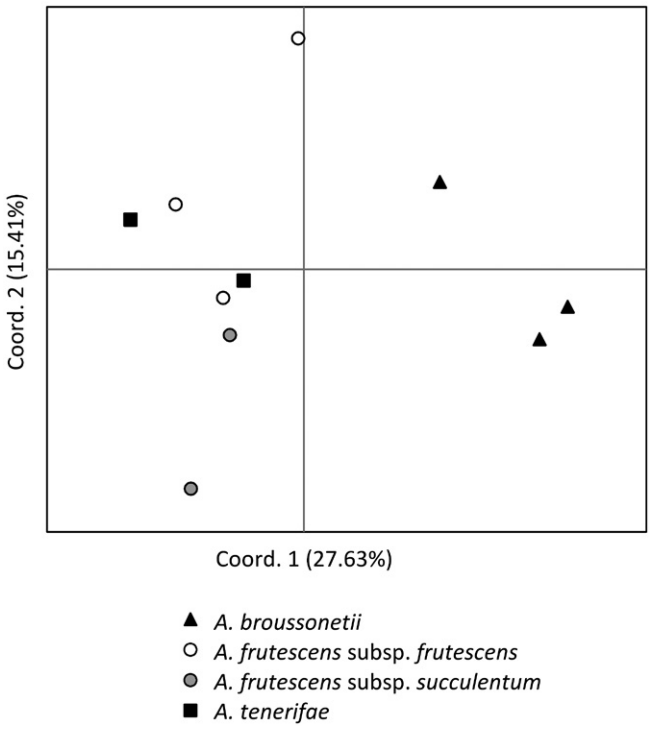

Fig. 4. Principal coordinates analysis (PCoA) of 10 samples of *Argyranthemum* based on genotypic information from eight SSR loci.

Although we have focused on SSR identification, there are further potential applications for the transcriptomes generated in future studies of *Argyranthemum*, *Descurainia*, or *Echium*, such as the development of universal markers that can be widely amplified across taxa (Wu et al., 2006; Chapman et al., 2007; Chapman, 2015) and the use of transcriptomes in studies of adaptation and speciation to link divergent genes to genes of known function.

Transcriptome sequences are a valuable source of genomic information that can be readily acquired from nonmodel organisms in prime ecological scenarios. The availability of free de novo transcriptome assembly software means there is no need for a reference genome to map sequences against and no subsequent costs. As such, this NGS approach can be applied to any nonmodel organism and downstream processing does not require any extra expenditure. Simultaneously providing gene expression and sequence polymorphism for thousands of genes, transcriptomics is one branch of NGS technology that holds exciting potential to inform evolutionary biologists about the genetic changes underlying processes such as adaptation and speciation (Sousa and Hey, 2013; Rius et al., 2015).

LITERATURE CITED

Altschul, S. F., G. Warren, W. Miller, W. E. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Böhle, U. R., H. H. Hilger, and W. F. Martin. 1996. Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proceedings of the National Academy of Sciences, USA* 93: 11740–11745.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30: 2114–2120.

Brochmann, C., L. Borgen, and O. E. Stabbetorp. 2000. Multiple diploid hybrid speciation of the Canary Island endemic *Argyranthemum sundingii* (Asteraceae). *Plant Systematics and Evolution* 220: 77–92.

Burstin, J., G. Deniot, J. Potier, C. Weinachter, G. Aubert, and A. Baranger. 2001. Microsatellite polymorphism in *Pisum sativum*. *Plant Breeding* 120: 311–317.

Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, D. E. Soltis, and P. S. Soltis. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Chapman, M. A. 2015. Transcriptome sequencing and marker development for four underutilized legumes. *Applications in Plant Sciences* 3: 1400111.

Chapman, M. A., J. Chang, D. Weisman, R. V. Kesseli, and J. M. Burke. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theoretical and Applied Genetics* 115: 747–755.

Chapman, M. A., S. J. Hiscock, and D. A. Filatov. 2013. Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution* 30: 2553–2567.

Doyle, J., and J. L. Doyle. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochemical Bulletin* 19: 11–15.

Egan, A. N., J. Schlueter, and D. M. Spooner. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99: 175–185.

Ellis, J. R., and J. M. Burke. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity* 99: 125–132.

Emerson, B. C. 2002. Evolution on oceanic islands: Molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology* 11: 951–966.

Fan, S., K. R. Elmer, and A. Meyer. 2012. Genomics of adaptation and speciation in cichlid fishes: Recent advances and analyses in African and Neotropical lineages. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367: 385–394.

Fjellheim, S., M. H. Jørgensen, M. Kjos, and L. Borgen. 2009. A molecular study of hybridization and homoploid hybrid speciation in *Argyranthemum* (Asteraceae) on Tenerife, the Canary Islands. *Botanical Journal of the Linnean Society* 159: 19–31.

Francisco-Ortega, J., R. K. Jansen, and A. Santos-Guerra. 1996. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. *Proceedings of the National Academy of Sciences, USA* 93: 4085–4090.

Francisco-Ortega, J., A. Santos-Guerra, A. Hines, and R. K. Jansen. 1997. Molecular evidence for a Mediterranean orgin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany* 84: 1595–1613.

Goodson, B. E., A. Santos-Guerra, and R. K. Jansen. 2006. Molecular systematics of *Descurainia* (Brassicaceae) in the Canary Islands: Biogeographic and taxonomic implications. *Taxon* 55: 671–682.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Humphries, C. J. 1976. A revision of the Macaronesian genus *Argyranthemum* Webb ex Schultz Bip. (Compositae-Anthemideae). *Bulletin of the British Museum (Natural History)* 5: 145–240.

Humphries, C. J. 1979. Endemism and evolution in the Macaronesia. *In* Plants and islands, pp. 171–199. Academic Press, New York, New York, USA.

Kelley, J. L., C. N. Passow, M. Plath, L. Arias Rodriguez, M.-C. Yee, and M. Tobler. 2012. Genomic resources for a model in adaptation and speciation research: Characterization of the *Poecilia mexicana* transcriptome. *BMC Genomics* 13: 652.

Kueffer, C., D. R. Drake, and J. M. Fernández-Palacios. 2014. Island biology: Looking towards the future. *Biology Letters* 10: 20140719.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518: 371–375.

Losos, J. B., and R. E. Ricklefs. 2009. Adaptation and diversification on islands. *Nature* 457: 830–836.

Mort, M. E., D. J. Crawford, J. K. Kelly, A. Santos-Guerra, M. Menezes de Sequeira, M. Moura, and J. Caujape-Castells. 2015. Multiplexed-shotgun-genotyping data resolve phylogeny within a very recently derived insular lineage. *American Journal of Botany* 102: 634–641.

Mun, J. H., D. J. Kim, H. K. Choi, J. Gish, F. Debellé, J. Mudge, R. Denny, et al. 2006. Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* 172: 2541–2555.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics (Oxford, England)* 28: 2537–2539.

Rius, M., S. Bourne, H. G. Hornsby, and M. A. Chapman. 2015. Applications of next-generation sequencing to the study of biological invasions. *Current Zoology* 61: 488–504.

Robotham, O., and M. Chapman. 2016. Erratum to: Population genetic analysis of hyacinth bean (*Lablab purpureus* (L.) Sweet, Leguminosae) indicates an East African origin and variation in drought tolerance. *Genetic Resources and Crop Evolution* doi:10.1007/s10722-015-0356-x.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany* 94: 275–288.

Sousa, V., and J. Hey. 2013. Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews. Genetics* 14: 404–414.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, et al. 2010. Adaptation genomics: The next generation. *Trends in Ecology & Evolution* 25: 705–712.

Sun, Y., D. Z. Skinner, G. H. Liang, and S. H. Hulbert. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* 89: 26–32.

Takayama, K., P. López Sepúlveda, G. Kohl, J. Novak, and T. F. Stuessy. 2013. Development of microsatellite markers in *Robinsonia* (Asteraceae) an endemic genus of the Juan Fernández Archipelago, Chile. *Conservation Genetics Resources* 5: 63–67.

Takayama, K., P. López-Sepúlveda, J. Greimler, D. J. Crawford, P. Peñailillo, M. Baeza, E. Ruiz, et al. 2015. Relationships and genetic consequences of contrasting modes of speciation among endemic species of *Robinsonia* (Asteraceae, Senecioneae) of the Juan Fernández Archipelago, Chile, based on AFLPs and SSRs. *New Phytologist* 205: 415–428.

Twyford, A. D., M. A. Streisfeld, D. B. Lowry, and J. Friedman. 2015. Genomic studies on the nature of species: Adaptation and speciation in *Mimulus*. *Molecular Ecology* 24: 2601–2609.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3—New capabilities and interfaces. *Nucleic Acids Research* 40: e115.

Wang, H., R. V. Penmetsa, M. Yuan, L. Gong, Y. Zhao, B. Guo, A. D. Farmer, et al. 2012. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biology* 12: 10.

Wang, H., J. Jiang, S. Chen, X. Qi, H. Peng, P. Li, A. Song, et al. 2013. Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. *PLoS ONE* 8: 1–10.

Wu, F., L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. *Genetics* 174: 1407–1420.