Transcriptomic Resources and Marker Validation for Diploid and Polyploid Veronica (Plantaginaceae) from New Zealand and Europe

Authors: Mayland-Quellhorst, Eike, Meudt, Heidi M., and Albach, Dirk C.

Source: Applications in Plant Sciences, 4(10)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600091

Applications in Plant Sciences

Applications in Plant Sciences 2016 4(10): 1600091

Genomic Resources Article

Transcriptomic resources and marker validation for diploid and polyploid *Veronica* (Plantaginaceae) from New Zealand and Europe[1]

Eike Mayland-Quellhorst[2,4], Heidi M. Meudt[2,3], and Dirk C. Albach[2]

[2]Carl-von-Ossietzky Universität Oldenburg, Carl-von-Ossietzky Straße 9–11, Oldenburg 26111, Germany; and [3]Museum of New Zealand Te Papa Tongarewa, Cable Street, P.O. Box 467, Wellington 6140, New Zealand

- *Premise of the study:* Polyploidy may generate novel variation, leading to adaptation and species diversification. An excellent natural system to study polyploid evolution in a comparative framework is *Veronica* (Plantaginaceae), which comprises several parallel, recently evolved polyploid series.
- *Methods:* Over 105 million Illumina paired-end sequence reads were generated from cDNA libraries of leaf tissue from eight individuals representing three European and four New Zealand species. Forty-eight simple sequence repeat (SSR) and 48 low-copy nuclear (LCN) markers were developed and validated with Fluidigm microfluidic PCR and Illumina MiSeq amplicon sequencing on 48 different individuals each.
- *Results:* Individual Trinity assemblies were similar regarding annotated transcripts (13,009–14,271), mean contig length (635–742 bp), N50 value (916–1133 bp), E90N50 value (1099–1308 bp), contigs with positive BLAST hits (42–63%), and gene ontology terms. Analyses of 29,738 single-nucleotide polymorphisms (8746 phylogenetically informative) mined from these transcriptomes plus two outgroups (*Picrorhiza kurrooa* and *Plantago ovata*) showed moderate to high bootstrap support for all branches and reticulation among sampled European *Veronica*.
- *Discussion:* The transcriptome sequences themselves, as well as the validated SSR (40/48) and LCN (11/48) markers derived from them, show inter- and intraspecific genetic variation. These resources will be invaluable for future population genetic, phylogenetic, and functional genetic investigations in polyploid *Veronica*.

Key words: low-copy nuclear (LCN) markers; polyploidy; simple sequence repeat (SSR) markers; single-nucleotide polymorphisms (SNPs); transcriptome; *Veronica.*

Polyploidy (whole genome duplication) is a very important process that has shaped flowering plant evolutionary history (Soltis et al., 2009). Much progress in the study of polyploid evolution has been made in the past two decades regarding both ancient paleopolyploidization (Doyle et al., 2008; Soltis et al., 2009) as well as very recent neopolyploidization (Buggs et al., 2009; Abbott et al., 2013). An important research gap (Soltis et al., 2009) is understanding polyploids of intermediate age that have diploid ancestors in the same genus, so-called mesopolyploids, which are characterized by diploid-like reproduction but whose parental subgenomes are still discernible (Mandáková et al., 2010).

Several mesoallopolyploid crop systems (e.g., cotton, soybean, tobacco, wheat) are becoming well understood and have excellent genetic resources; however, understanding natural systems is also important. Specifically, studying natural mesopolyploid species radiations may be key to understanding the importance of polyploidy in angiosperm diversification (Soltis et al., 2009). Recent plant species radiations are a significant contributor to generating plant biodiversity, and evidence suggests that polyploidy has played an important role in these radiations (Mayrose et al., 2011). Many fundamental and biologically interesting questions regarding polyploidy and diversification in plants are yet to be investigated in such systems (Doyle et al., 2008; Soltis et al., 2009; Mayrose et al., 2011).

The large, nearly cosmopolitan genus *Veronica* L. (Plantaginaceae) comprises approximately 450 species of annual and perennial herbs, shrubs, and small trees with centers of diversity in both Eurasia and New Zealand. The genus is an excellent example of a natural mesopolyploid (~20 million years old) system comprising multiple lineages, including several recent species radiations, in which polyploidy and hybridization have accompanied diversification (Albach et al., 2008; Meudt et al., 2015). Northern Hemisphere *Veronica* species are diploids or polyploids with chromosome numbers ranging from $2n = 14–80$ and base numbers of $x = 6–9$ and 17 (Albach et al., 2008). By contrast, Southern Hemisphere species—which evolved as a single lineage from Northern Hemisphere ancestors ~10 million years ago (Wagstaff et al., 2002; Albach and Meudt, 2010; Meudt et al., 2015)—all have high chromosome numbers ($2n = 40–124$) with base chromosome numbers of $x = 20$ or 21 (Albach et al., 2008).

[1]Manuscript received 28 July 2016; revision accepted 2 September 2016.

The authors thank the Albach and Zotz AGs at Carl-von-Ossietzky Universität Oldenburg for their support; W. Lobin (Bonn), J. Seguí Colomar (Mallorca), P. J. Garnock-Jones (Wellington), and A. C. Clarke (Cambridge) for assistance collecting material; the Alexander von Humboldt Foundation for an Experienced Researcher Fellowship (to H.M.M.); New Zealand Core funding for Crown Research Institutes from the Ministry of Business, Innovation and Employment's Science and Innovation Group; and two anonymous reviewers for their constructive comments.

[4]Author for correspondence: eike.mayland.quellhorst@uni-oldenburg.de

doi:10.3732/apps.1600091

Several studies focusing on *Veronica* in both hemispheres have used standard DNA sequencing and amplified fragment length polymorphism (AFLP) fingerprinting techniques to elucidate patterns of relationship from phylogeography (Meudt and Bayly, 2008) to phylogeny of the genus as a whole (Wagstaff et al., 2002; Albach and Meudt, 2010; Meudt et al., 2015) or of particular polyploid complexes (e.g., Albach, 2007), and used these to infer the evolution of chromosome number, genome size, breeding systems, and habit (Albach and Greilhuber, 2004; Albach et al., 2008; Meudt et al., 2015). Nevertheless, a lack of variable genetic markers using standard DNA sequencing and genotyping techniques, and a lack of appropriate phylogenetic analysis methodologies that can incorporate reticulate evolution and allopolyploids, have hampered further progress in studies of *Veronica* and polyploid evolution at the population, species, and generic levels.

It has been known for some time that low-copy nuclear (LCN) markers can be extremely useful for phylogenetic reconstruction at the genus (interspecific) level, including for elucidating the evolutionary history of polyploids, for which standard uniparental DNA sequencing markers from chloroplast DNA or the internal transcribed spacer (ITS) region are not informative (e.g., Sang, 2002). Apart from LCN markers, microsatellites or simple sequence repeat (SSR) markers are useful for closely related species when traditionally genotyped and analyzed for studies at the infrageneric level, but SSRs and their flanking regions may also be useful as phylogenetic markers when high-throughput sequenced (Chatrou et al., 2009; Germain-Aubrey et al., 2016).This, however, requires new bioinformatic tools such as the workflows MarkerMiner (Chamala et al., 2015) and QDD (Meglécz et al., 2014) for the development of LCN and SSR markers, respectively, using genomic and transcriptomic resources.

High-throughput de novo transcriptome sequencing, or RNA-Seq, has proven to be an excellent source of genetic data for gene characterization and marker development in studies of natural systems with little or no additional genetic resources available (Strickler et al., 2012; Alvarez et al., 2015), as is the case for *Veronica*. The benefits of RNA-Seq are simultaneous characterization of genes and gene expression, reduced representation for large, complex genomes, and the generation of large amounts of sequence data without a reference genome. RNA-Seq also presents its challenges, particularly assembly without a reference genome, and assembly of polyploid genomes. Polyploid transcriptome assembly is an active area of research. A major issue is the differentiation of homoeologs from orthologs. Some studies have tested different pipelines, such as combining multiple *k*-mer assemblies in polyploid wheat (Krasileva et al., 2013), or combining assemblies from different assemblers and then using a second step to cluster redundant contigs in polyploid tobacco (Nakasugi et al., 2014). However, there are few examples to date of comparisons in natural, noncrop systems with few prior genomic resources. To date, there are no clear answers regarding which assembler, combination of assemblers, or assembly pipeline is best for polyploids and their diploid progenitors or close relatives.

The aim of our study was, therefore, twofold. First, we aimed to generate transcriptomic data for *Veronica*; second, we aimed to use these transcriptomic resources to develop and validate phylogenetically informative sequencing markers. Specifically, in this paper, we generate the first transcriptome resources for the genus *Veronica*, using short-read Illumina HiSeq 2000 (Illumina, San Diego, California, USA) sequencing of eight individuals representing seven species and five different ploidy levels. We then assemble, identify, and broadly characterize and compare a large number of expressed sequences. Single-nucleotide polymorphisms (SNPs) are mined from transcriptomes of these eight individuals plus those of two additional Plantaginaceae outgroups (*Plantago ovata* Forssk. and *Picrorhiza kurrooa* Royle ex Benth., available from public databases) and compared using phylogenetic and network analyses. Secondly, we used the transcriptomic data to discover, design, and develop two types of genetic markers (i.e., LCN and SSR markers). To test the success of our approach, we then used microfluidic PCR and Illumina MiSeq to validate 48 loci in 48 individuals for both LCN and SSR markers. We provide examples of sequence alignments and downstream phylogenetic analyses for representative loci showing their potential phylogenetic utility in *Veronica* when resequenced with high-throughput sequencing. The resource and marker development of the current study provide new, variable markers for future evolutionary studies of the genus. Furthermore, a parallel study currently underway will further examine assembly methods and analyze the transcriptomes themselves to quantify and compare underlying interspecific gene divergence and investigate the timing and mode of polyploidy in the sampled *Veronica* polyploids and their close relatives (Meudt et al., unpublished data). The current study is thus a critical first step toward ultimately understanding the role of polyploidy in generating novel genetic and morphological variation that leads to adaptation and species diversification (Doyle et al., 2008; Soltis and Soltis, 2009).

MATERIALS AND METHODS

***RNA extraction, cDNA library prep, and Illumina sequencing*—**We sampled leaf tissue from seven greenhouse-grown individuals and from one field-collected individual representing seven species of two polyploid complexes in *Veronica* from New Zealand and Eurasia with three ploidy levels each (Appendix 1). The field-collected material was stored at −80°C on RNA*later* (Life Technologies, Carlsbad, California, USA). Because we wanted to take a broad approach to analyze polyploidy in *Veronica* and develop markers for the entire genus, we sampled multiple species in two divergent lineages, rather than multiple individuals per species. Cultivated plants were grown in the same greenhouse in Oldenburg, Germany. Leaf material was harvested and placed directly into tubes with liquid nitrogen, stored at −80°C until extraction, and ground to a powder with a prechilled mortar and pestle while adding liquid nitrogen. RNA was extracted using the RNeasy kit (QIAGEN GmbH, Hilden, Germany) following manufacturer's instructions using 500 µL RLC buffer with 4% PVP and 1% β-mercaptoethanol. A DNase I digest and RNase inhibitor reaction was performed using 0.5 µL (20 units) RNase inhibitor, 6.0 µL 10× DNase I buffer, and 1.0 µL DNase I to the resulting 60 µL RNA extract and incubated at 37°C for 15 min. Then, 2.6 µL EDTA (0.2 M, pH = 8; final conc. 8 mM) was added, incubated for 10 min at 75°C, and the RNA was reprecipitated by adding 1 : 10 3 M sodium acetate, 2.5 volume 100% ethanol, incubating on ice for 20 min, centrifuging at full speed for 5 min, washing with 100 µL 75% ethanol, centrifuging at full speed, air-drying the resultant pellet for 10–15 min, redissolving in 25 µL RNase-free water, and storing at −80°C. Small aliquots of raw RNA extract and the reprecipitated RNA extract were run on the Tecan Infinite Pro F200 (Tecan, Crailsheim, Germany) and Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) to measure RNA quality and quantity. RNA from eight individuals with RNA Integrity Number (RIN) of 6.8 or greater, 260 : 280 ratio between 1.9–2.1, and at least 50 ng/µL (Appendix 1) were sent to BGI (BGI-Hong Kong Co. Ltd, Hong Kong, China) for Illumina TruSeq cDNA library preparation on normalized RNA and high-throughput Illumina HiSeq 2000 100-bp paired-end de novo transcriptome sequencing. The transcriptomic data generated here are publicly available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive for submission SPR074674 and the Trinity assemblies in the NCBI Transcriptome Shotgun Assembly Sequence Archive (Table 1; http://www.ncbi.nlm.nih.gov/sra/SRP074674).

TABLE 1. Information about Illumina sequencing reads and Trinity assemblies for the eight individuals of *Veronica* sampled.

Species (Ploidy)	Geography	SRA accession/ TSA accession[a]	No. of raw (clean) reads	No. of contigs (Trinity)	N50 value (mean/median contig length)	No. (%) annotated contigs with positive BLAT hits	No. (%) GO terms assigned[b]	No. of LCN markers[c]
Veronica catarractae (6*x*)	New Zealand	SAMN04961631/ GEVT00000000	23,711,074	66,671	1078 (732/493)	37,287 (56)	13,940 (21)	580
V. hectorii subsp. *coarctata* (6*x*)	New Zealand	SAMN04961628/ GEVQ00000000	24,385,098	64,950	1097 (742/511)	36,839 (57)	14,068 (22)	573
V. planopetiolata (12*x*)	New Zealand	SAMN04961630/ GEVS00000000	25,055,264	73,820	1020 (698/476)	39,915 (54)	14,197 (19)	625
V. ochracea (18*x*)	New Zealand	SAMN04961629/ GEVR00000000	24,050,110	61,752	1065 (722/494)	37,071 (60)	14,211 (23)	606
V. panormitana (2*x*)	Europe	SAMN04961624/ GEVN00000000	24,429,090	41,451	1118 (741/482)	25,047 (60)	13,714 (33)	571
V. trichadena (2*x*)	Europe	SAMN04961625/ GEVU00000000	24,269,936	58,998	916 (635/403)	24,583 (42)	13,009 (22)	460
V. cymbalaria (4*x*)	Europe	SAMN04961626/ GEVO00000000	23,406,760	46,573	1133 (767/539)	29,458 (63)	13,564 (29)	506
V. cymbalaria (6*x*)	Europe	SAMN04961627/ GEVP00000000	24,845,504	73,889	992 (671/431)	36,589 (50)	14,271 (19)	634

Note: GO = gene ontology; LCN = low-copy nuclear; SRA = Sequence Read Archive; TSA = Transcriptome Shotgun Assembly.
[a]Sequence Read Archive (SRA) accession numbers for SRA submission SPR074674 (http://www.ncbi.nlm.nih.gov/sra/SRP074674).
[b]BLAT search with MapMan categories.
[c]Number of LCN markers detected in MarkerMiner (Chamala et al., 2015), contigs longer 600 bp.

Quality control, preprocessing of reads, assembly, and Blast2GO analyses— The following analyses were carried out on each of the eight individuals separately. Demultiplexed Illumina sequencing results were retrieved in FASTQ format via FTP from BGI. Between 12.8 and 13.5 million paired-end reads were generated per individual in both the forward and reverse directions (Table 1), from which single reads, adapters, and reads with a quality score (QC) cutoff of less than 20 had already been removed. After testing the effect of different QC cutoffs on the resulting sequence reads and assemblies of *V. trichadena* Jord. & Fourr., we used QC = 40 in the bash script TrimClip.sh (De Wit et al., 2012) to remove reads QC < 40. Reads were screened for contaminant sequences from *H. sapiens*, *E. coli*, mtDNA, and cpDNA using mirabait (MIRALIB version 4.0; Chevreux et al., 1999) with default settings, the respective databases downloaded from NCBI, and then removed. We used QualityStats.sh (De Wit et al., 2012) and the Galaxy web interface (Afgan et al., 2016) to summarize quality score and nucleotide distribution data for the forward and reverse reads, CollapseDuplicateCount.sh (De Wit et al., 2012) to calculate the fraction of duplicate reads and singletons, PECombiner.sh (De Wit et al., 2012) to remove orphan reads and put remaining reads in the same order in forward and reverse files, and the Velvet helper script shuffleSequences_fastQ.pl to put those two files together in one interleaved file (necessary for Velvet/Oases assembly). The resulting clean sequence reads were assembled de novo using several different assemblers including Trinity, trans-ABySS, SOAPdenovo-Trans, and Velvet/Oases. Relative to the other assemblers, Trinity produced more hits with >80% similarity to contigs >600 bp against *Arabidopsis thaliana* (L.) Heynh. (data not shown; comparisons done using MarkerMiner 1.0 [Chamala et al., 2015]). Therefore, we chose the de novo assemblies produced using Trinity version r20140717 (Grabherr et al., 2011, compiled for 64-bit Ubuntu) using default settings on the resulting clean sequence reads. For the purposes of marker development, a highly accurate discrimination of homoeologs in polyploids is not necessary at the transcriptome assembly stage, as the discrimination is done in the second resequencing step. Additional comparisons of different assemblers and assembly pipelines, particularly regarding polyploid transcriptomes, were outside the scope of the current study and will be addressed in a subsequent study (Meudt et al., unpublished data). Trinity assemblies of all four New Zealand, all four European, and all eight *Veronica* individuals were also made. Table 1 shows information about the sequence reads and statistics from the eight different individual Trinity assemblies. Functional annotation of contigs from the different assemblies was conducted using BLAT (Kent, 2002) with default settings against the TAIR database (version 10 represented gene model from 2011-01-03; Lamesch et al., 2012) and MapMan hierarchical categories (Ath_AGI_LOCUS_TAIR10_Aug2012; http://mapman.gabipd.org/web/guest/mapmanstore). Mean contig length ranged from 635–742 bp, N50 value from 916–1133 bp, E90N50 value from 1099–1307 bp (which is computed with the contig_ExN50_statistic.pl script of the Trinity package and represents the N50 of 90% of the expressed transcripts), and number (and percentage) of contigs with positive BLAST hits from 24,583–39,915 (42–63%). To demonstrate the quality and utility of the transcriptomic resources developed here, we compared the transcriptome sequences of our eight sampled individuals relative to each other and to two outgroups, *Picrorhiza kurrooa* (http://scbb.ihbt.res.in/Picro_information/; SRR392742; Gahlan et al., 2012) and *Plantago ovata* (SRR629688; Kotwal et al., 2016). To do this, we mined the data from these 10 individuals for SNPs using Site Identification from Short Read Sequences (SISRS) version 1.0 (Schwartz et al., 2015; https://github.com/rachelss/SISRS/releases). SISRS identifies SNPs for phylogenetic studies directly from raw high-throughput sequences without a reference genome and without a priori knowledge of potentially informative loci. Briefly, SISRS first assembles raw sequence reads into a "composite genome" using Velvet, maps the raw reads and individual contigs against this composite genome with Bowtie 2, and then calls SNPs with a Python script (Schwartz et al., 2015). SNP discovery was performed using SISRS on four different data sets: (1) all eight *Veronica* individuals combined plus *P. kurrooa* and *P. ovata* as outgroups, (2) all eight *Veronica* individuals only, (3) the four New Zealand individuals only, and (4) the four European individuals only. The SNP data were converted to NEXUS format and analyzed using NeighborNet networks (SplitsTree version 4.14.2; Huson and Bryant, 2006). In addition, GARLI version 2.01.1067 (Zwickl, 2006) was used for phylogenetic tree reconstruction under maximum likelihood. We first performed a GARLI run with 10 replicates to estimate the model parameters for the model of evolution estimated with jModelTest version 2.1.5 (012010F; Darriba et al., 2012) [setting ratematrix = a b c a b a statefrequencies = estimate]; six of the 10 replicates had the same best lnL score. These estimated model parameters were then fixed for a bootstrap analysis, which was performed with 1000 replicates [parametervaluestring = M1 r 1.00000 7.30163 1.61422 1.00000 7.30163 1.00000 e 0.27231 0.22561 0.22541

0.27667]. The resulting tree was compared to previously published phylogenetic estimates.

***Marker development*—**Two different types of markers were developed from the *Veronica* transcriptome resources generated here, LCN and SSR markers.

Low-copy nuclear markers—MarkerMiner was used with default settings to identify LCN markers from a curated set of conserved ortholog set (COS) loci (De Smet et al., 2013). MarkerMiner was developed and tested using transcriptome assemblies from 77 Lamiales species (including six from Plantaginaceae; Chamala et al., 2015), and uses a reciprocal BLAST of all transcriptomes with one another and to the reference *A. thaliana* genome. *Arabidopsis thaliana* (Brassicales) is the phylogenetically closest reference available in MarkerMiner to *Veronica* (Lamiales). Of the 1228 loci returned, 73 were classified as being "strictly" and 1155 as "mostly" single copy. MAFFT alignments of the 330 loci found in six or more individuals, of which 15 were "strictly" and 314 "mostly" single copy, were used to develop primers in Geneious (version 8.7) with Primer3 (Untergasser et al., 2012), aiming for a melting temperature of 60°C. Loci were checked manually for large introns in Geneious by comparing the MarkerMiner alignment to *A. thaliana.* We chose 13 "strictly" single-copy loci with a successful primer search and 35 additional "mostly" single-copy loci with successful primer searches such that all five *A. thaliana* chromosomes were equally represented in this marker set. These 48 loci were validated using Fluidigm microfluidic PCR and Illumina MiSeq amplicon sequencing of 48 individuals representing 46 *Veronica* species (19 from the Southern Hemisphere) and all subgeneric lineages (Appendix 2). The combination of this technique with Illumina MiSeq amplicon sequencing of 300-bp paired-end reads has proven useful and highly efficient in recent studies for development of novel and effective nuclear sequencing markers and improving understanding of phylogenetic relationships in nonmodel genera (Gostel et al., 2015; Uribe-Convers et al., 2016). This method enables the amplification of 48 samples and 48 primer pairs in 4-μL reaction volumes, such that the total volume of these reactions equals, e.g., 10 standard 25-μL reaction volumes. Each reaction contained 2 ng DNA, 200 nM of each primer, 0.1 μL 5 U/μL VELOCITY DNA polymerase (Bioline, Luckenwalde, Germany), 1× buffer, 0.1 μL 10 mM dNTPs, 0.25 μL 1 M DMSO, and 0.5 μL 5 M betaine. The samples were initially denatured for 2 min at 98°C; followed by 45 cycles of denaturation for 15 s at 98°C, annealing for 30 s at 55°C, elongation for 30 s at 72°C; and finalized with 5-min elongation at 72°C. Preliminary testing showed that more cycles were necessary due to some low-quality DNA samples. Barcoding and Illumina sequencing was done by LGC Genomics (Berlin, Germany) with Illumina MiSeq v3 chemistry. For each LCN locus, resulting sequences were trimmed with BBMap tools (https://sourceforge.net/projects/bbmap/), de novo assembled with CAP3 (99% identity; Huang and Madan, 1999), aligned to the respective locus sequence of the transcript with MAFFT (setting E-INSI; Katoh and Standley, 2013), and examined in Geneious for sequence length, similarity to original transcript, *A. thaliana* gene, and number of individuals successfully sequenced. In addition, the alignment was exported to GARLI, in which numbers of sequences, SNPs, and parsimony informative characters (PICs) were calculated. For one randomly chosen example LCN marker, a phylogeny was reconstructed using the same settings as described above for SNPs.

Simple sequence repeats—Numerous SSRs were identified from Trinity assembly of the New Zealand individuals only using QDD version 3.1 (Meglécz et al., 2014; Table 1). Settings for the search were a length of 250–350 bp of the locus and primer melting temperatures of 59–61°C. After filtering for quality (taking QDD categories A and B), repeats (removing dinucleotides for example), and length of predicted PCR product, 48 loci were chosen from the 1124 potential SSRs with primer sites found by QDD. These were validated using Fluidigm microfluidic PCR and Illumina MiSeq amplicon sequencing (see above) of 48 individuals representing 20 Australasian species and one interspecific hybrid (Appendix 3). For each SSR marker, which included the SSR repeat area and flanking regions, resulting sequences were analyzed in the same way as the LCN data (see above) and examined in Geneious and GARLI regarding SSR motif, sequence length, number of individuals successfully sequenced, number of alleles sequenced, and pairwise genetic distance. In addition, for one randomly chosen example SSR locus, the alignment was exported to GARLI and a phylogeny was reconstructed using the same settings as described above for SNPs.

RESULTS

***Transcriptomes*—**Functional annotation of individual assemblies was similar for each of the eight individuals, with gene ontology (GO) terms assigned to 13,009–14,271 contigs (19–33%; Table 1). There was large overlap of annotated contigs of the different assemblies whether looking at assemblies of individuals of New Zealand species only (26,524 or 89.4% shared annotated contigs; Fig. 1A), European species only (25,456 or 87.8%; Fig. 1B), or all New Zealand vs. all European species (29,839 or 94.3%; Fig. 1C). On the other hand, individual species had 114–453 (0.4–1.6%) unique annotated contigs relative to other species from the same geographical area, and

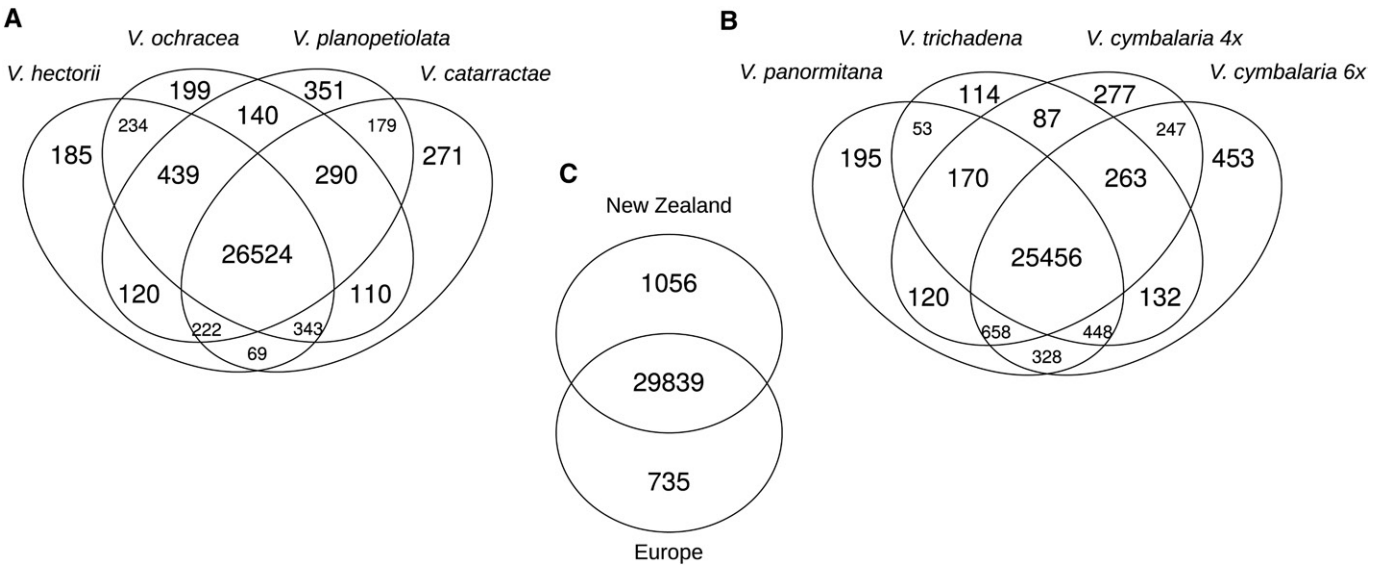

Fig. 1. Venn diagrams showing the number of annotated contigs from the *Veronica* Trinity assemblies. (A) Four New Zealand individuals. (B) Four European individuals. (C) All New Zealand vs. all European individuals.

the numbers for New Zealand and European species were comparatively very similar (compare Fig. 1A and 1B). Within the New Zealand species, *V. hectorii* Hook. f. and *V. ochracea* (Ashwin) Garn.-Jones shared the most unique annotated contigs (234 or 0.8%) relative to the other five species pairs, whereas *V. catarractae* G. Forst. and *V. ochracea* shared the fewest (110 or 0.4%; Fig. 1A). Within the European species, the species pair with the most unique shared annotated contigs was *V. panormitana* Tineo ex Guss. (2*x*) and *V. cymbalaria* Bodard (6*x*) (238 or 1.1%), whereas the two diploids *V. panormitana* and *V. trichadena* shared the fewest (53 or 0.2%) (Fig. 1B).

GO term results were also very similar; of 35 GO categories, the number of unique transcripts were largely overlapping for all species pairs, as is shown for the most divergent species pair of the eight transcriptomes sequenced (i.e., New Zealand *V. hectorii* vs. European *V. panormitana*; Fig. 2). The GO categories with the largest numbers of unique transcripts (ca. 500–3000) for these *Veronica* leaf transcriptomes were (from highest to lowest) "not assigned," "protein," "RNA," "signaling," "transport," "misc," "cell," and "DNA" (Fig. 2A).

SNP discovery using SISRS resulted in the following number of SNPs and potential PICs: 10-individual data set including outgroups (29,738 SNPs, 8746 PICs), eight *Veronica* individuals only (45,751 SNPs, 40,217 PICs), four New Zealand individuals only (41,167 SNPs, 2302 PICs), and four European individuals only (65,278 SNPs, 1735 PICs). When the 10-individual data set was analyzed using SplitsTree (Fig. 3A–C), the NeighborNet network clearly showed a main split between all *Veronica* transcriptomes vs. the two outgroups (Fig. 3A). Although some reticulation was present among the New Zealand species (Fig. 3C), reticulation is more pronounced among the European individuals (Fig. 3B), which comprise two allopolyploids and their putative diploid parental species. The phylogenetic analysis of the same data set contained moderate to high support for all branches in the phylogeny (Fig. 3D). Among the New Zealand individuals, *V. hectorii* (6*x*) and *V. ochracea* (18*x*) are very closely related to each other; *V. ochracea* may be an allopolyploid of *V. hectorii* and another unsampled species (Wagstaff and Wardle, 1999). Within the European lineage, *V. cymbalaria* (4*x*) is positioned between both diploid parental species as expected (Albach, 2007; Fig. 3B, 3D), and we suspect *V. cymbalaria* (6*x*) to be a backcross allopolyploid of *V. cymbalaria* (4*x*) × *V. panormitana* (2*x*) based on the larger similarity with that species compared with *V. trichadena* (Fig. 1; 328 vs. 132 unique annotated contigs).

Marker development

Low-copy nuclear markers—A range of 3–44 (average: 23.4, median: 22) of 48 individuals were successfully sequenced for each of the 48 loci, with 22 of 48 (46%) loci successfully amplifying in at least 24 (>50%) individuals (Appendix 4). For each individual, 4–40 loci were successfully amplified (average: 23.4, median: 21.5), and again less than half of the individuals (22/48, 46%) had successful amplification of at least 25 (>50%) loci (data not shown). Only one-quarter (11/48) of the loci aligned well with the corresponding transcript; these loci had mean lengths of 327–480 bp, contained large numbers of SNPs and PICs, and BLASTed to known *A. thaliana* genes (Appendix 4).

Figure 4 shows an alignment and GARLI tree of sequences from 22 of the 42 individuals successfully sequenced for one randomly chosen example locus, LCN-04 (two outgroups plus 10 European and 10 New Zealand *Veronica* individuals/species). Nearly twice as many different sequences were generated for the 10 New Zealand individuals shown here (27 sequences; 6*x* or 18*x*, "*V. townsonii* E6 18" to "*V. melanocaulon* D5 11" in the tree) relative to the 10 European individuals (14 sequences; 2*x* or 4*x*, "*V. missurica* E2 13" to "*V. chamaedrys* C3 10"). Of the 10 New Zealand individuals, five have only one sequence and are all in the same clade (*V. albiflora* (Pennell) Albach, *V. cupressoides* Hook. f., *V. densifolia* F. Muell., *V. lavaudiana* Raoul, and *V. senex* (Garn.-Jones) Garn.-Jones), whereas the other five have 2–8 sequences that fall into the first clade or one

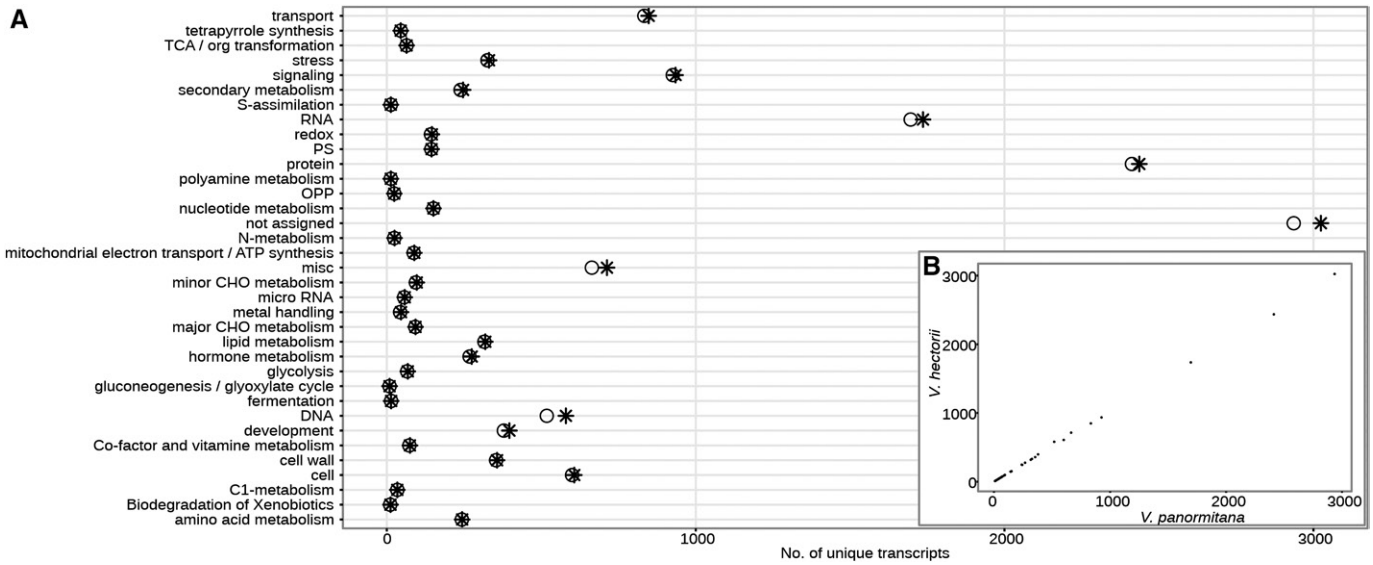

Fig. 2. (A) Number of unique transcripts (x axis) for each of 35 hierarchical gene ontology (GO) categories (y axis) for the Trinity assemblies of leaf transcriptome data from one individual each of *Veronica panormitana* (European diploid, *) and *V. hectorii* (New Zealand hexaploid, ○). (B) Comparison of number of genes for *V. panormitana* vs. *V. hectorii*. Results from the other six individuals were very similar (data not shown).

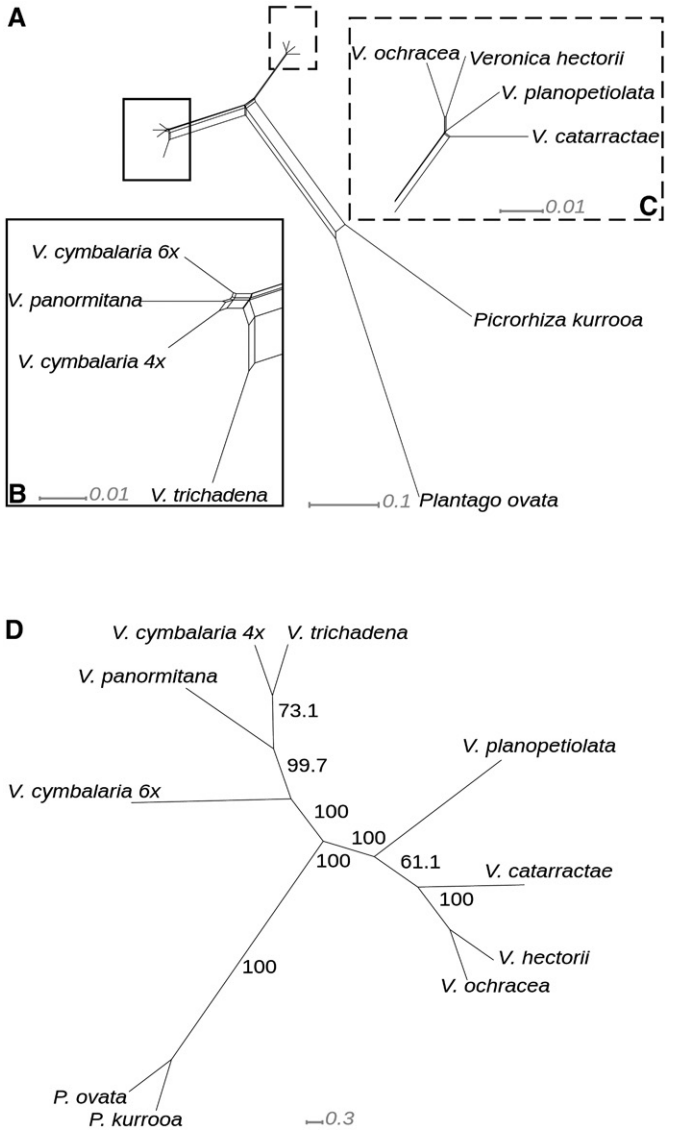

Fig. 3. Network and phylogenetic analyses of SNPs mined from leaf transcriptome data using SISRS for eight individuals of *Veronica* and two outgroups. (A) SplitsTree NeighborNet network. (B) Detail of network showing relationships of the four New Zealand *Veronica* individuals. (C) Detail of network showing relationships of the four European *Veronica* individuals. (D) GARLI phylogenetic tree with bootstrap values from 1000 replicates.

of two other clades (Fig. 4). As another example highlighting the low-copy nature of the loci that were sequenced, locus LCN-38 has two orthologous copies, which is expected due to the categorization of the *A. thaliana* gene AT3G59380 as "mostly" single copy (data not shown; comparisons done using MarkerMiner 1.0 [Chamala et al., 2015]). Additional phylogenetic analyses of the other LCN loci are outside the scope of this study and will be performed elsewhere (Meudt et al., unpublished data).

Simple sequence repeats—Overall, 3–47 (mean: 37.7, median: 44.5) of 48 individuals were successfully sequenced for each of the 48 SSR loci, including 40 of 48 loci that were successfully sequenced for at least 26 (>50%) individuals (Appendix 3). For each individual, 0–43 loci were successfully sequenced (average: 37.7, median: 40), with all but two individuals with at least 29 (>60%) loci successfully sequenced (individuals *V. catarractae* B1 and *V. colostylis* H3 failed for all 48 and 40 loci, respectively). In general, sequences ranged from 98–851 bp in length (average: 324) and contained one or more length- and/or sequence-variable SSR motifs as well as flanking SNPs and indels within and among individuals (e.g., Fig. 5). Number of sequenced alleles (which are supported by at least 10 raw sequencing reads) per individual ranged from 1–39 (mean: 4.32, median: 3.0, $n = 47$), with the lower polyploids having fewer alleles than the higher polyploids (6*x*, mean: 3.96, $n = 37$; 12*x*, mean: 5.25, $n = 5$; 18*x*, mean: 6.09, $n = 5$).

As the focus of SSRs is often population genetics, we analyzed two subsets of the larger SSR data set in more detail, i.e., eight individuals of *V. chionohebe* Garn.-Jones (4), *V. trifida* Petrie (2), and their interspecific hybrid (2) (all $2n = 42$) (Appendix 5), and six individuals of *V. thomsonii* Cheeseman ($2n = 42$), respectively (Appendix 6). For all loci in the two subsets, sequences were on average of 317–327 bp, with 1–26 alleles (mean: 4.0–4.3), 54–80 SNPs, and 41.7–52.5 PICs (see "Totals" rows in Appendix 5 and 6). Figure 5 shows an alignment of 54 different SSR sequences from one locus (SSR-08) of the eight-individual *V. chionohebe/V. trifida* subset. In locus SSR-08, the sequences ranged from 311–387 bp (average: 357 bp). The sampled individuals had on average 6.8 alleles, and individuals of *V. chionohebe* had half as many unique alleles (3–6 each) as individuals of *V. trifida* and the interspecific hybrid (8–10). The sequences of locus SSR-08 were highly variable (note the many colored bars in the alignment in Fig. 5), with 126 SNPs and 109 PICs, and 0–0.14 pairwise genetic distances (mean and median: 0.08) (Appendix 5). In the phylogenetic tree, there is support for some taxonomic clustering of sequences of *V. chionohebe* and *V. trifida*, respectively, with hybrid sequences in highly supported clades with *V. chionohebe* or *V. trifida* in three vs. four cases, respectively (see tree in Fig. 5). Additional analyses of the other SSR loci are outside the scope of this study and will be performed elsewhere (Meudt et al., unpublished data).

DISCUSSION

The development of transcriptomic and genomic resources and variable genetic markers in so-called natural "mesopolyploid" species radiations is key to addressing fundamental questions about polyploidy and diversification. For polyploids, functional genomic resources in particular are important to facilitate the study of gene evolution. *Veronica* is an example of a natural mesopolyploid species radiation that to date has lacked such genomic and genetic resources, and this has hindered progress in studying polyploid evolution at the population, species, and generic levels. The transcriptomic and genetic resources developed here will make further detailed studies regarding the role of polyploidy in adaptation and species diversification in *Veronica* possible.

In the current study, we sequenced and assembled leaf transcriptomes from eight individuals representing seven species of *Veronica* from polyploid species radiations in Europe and New Zealand. There was high overlap of annotated contigs (Fig. 1) and GO terms (Fig. 2) among the eight individuals, as well as good phylogenetic resolution in the network and phylogenetic analyses of SNPs generated using SISRS (Fig. 3). An outstanding challenge with de novo transcriptome assemblies of polyploids

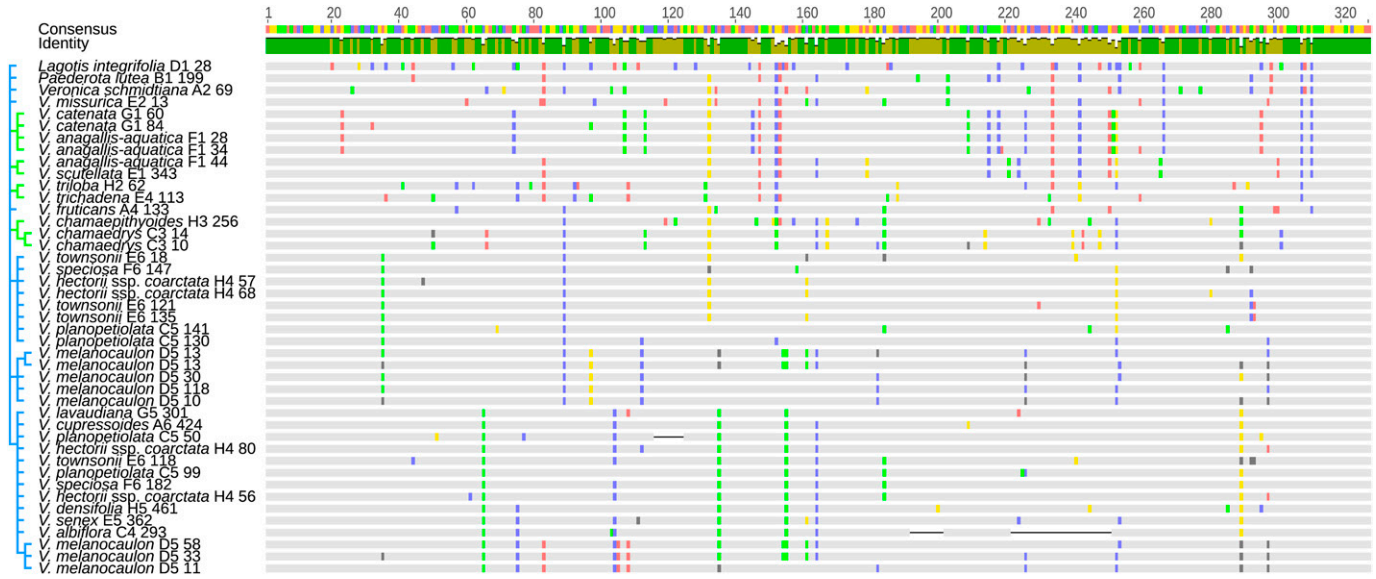

Fig. 4. MAFFT alignment and GARLI phylogenetic tree (visualized in Geneious) for 22 of the 42 individuals (two outgroups plus 10 European and 10 New Zealand *Veronica*) for which sufficient sequence reads of the correct locus were successfully generated from sequences of LCN locus LCN-04 mined using MarkerMiner from Trinity assemblies of leaf transcriptome data. The consensus and identity sequences are shown at the top. Base pairs that are identical to the consensus are shown in gray, whereas SNPs are shown as colors (red = A, blue = C, green = T, yellow = G, black = N). For each sequence in the alignment, species names are followed by sequencing plate location (e.g., D1) and number of sequence reads supporting that allele (range: 10–424). Green branches in the GARLI tree to the left of the individual names have >80% bootstrap support (see Fig. 3 for GARLI settings). Voucher information is shown in Appendix 2.

is differentiating homoeologs from orthologs; however, this was not an issue for developing markers in polyploid *Veronica* from our transcriptome assemblies, as phylogenetic relationships (Fig. 3) are consistent with hypothesized relationships and previous phylogenetic results. Such results demonstrate the utility of these transcriptomic resources for phylogenetic studies, functional analyses across the genus using reverse transcription PCR, or for further comparative transcriptomic analyses of the sampled natural allopolyploids and their diploid parental species in the two main centers of diversity for *Veronica* (i.e., Europe and New Zealand). The large number of transcripts unique to hexaploid *V. cymbalaria* (453) relative to other individuals representing species from which it likely derived (*V. trichadena*: 114 and *V. panormitana*: 195) is surprising and opens the door to studies of differential expression and functional differentiation of genes in polyploids. Common garden experiments are also planned, which will allow comparison of other individuals with the eight sequenced here.

Furthermore, the SSR and LCN genetic markers developed here from the transcriptomes, and validated using microfluidic PCR and high-throughput sequencing, are highly variable and will be extremely useful in future phylogenetic studies of *Veronica* as a whole, as well as studies at the interface of inter- and intraspecific levels of New Zealand *Veronica* (e.g., phylogenetic, phylogeographic, and population genetic studies). From 330 mostly or strictly "low copy" loci common to 6–8 of the sequenced transcriptomes, we developed and sequenced 48 LCN markers in 48 individuals representing all subgeneric lineages in *Veronica*. Of the 22 LCN markers that were successfully sequenced for >50% individuals, 11 aligned well with the corresponding transcripts, were on average 394 bp long, contained large numbers of SNPs and PICs, and BLASTed to known *A. thaliana* genes. These 11 LCN markers are excellent candidates for reconstructing a better-resolved phylogeny of *Veronica*. In addition, of the 1124 SSRs identified in the four New Zealand *Veronica* individuals, we validated 48 in 48 Southern Hemisphere *Veronica* individuals, 40 of which were successfully sequenced for >50% of individuals. Sequenced SSRs and their flanking regions were on average 324 bp long, contained numerous SNPs and PICs, and had mean pairwise genetic distances of 0.01–0.18. The variation seen, particularly in the flanking regions of the sequenced SSRs, is equal to or much greater than that from previous studies using standard DNA sequencing and genotyping markers (e.g., Wagstaff et al., 2002; Meudt and Bayly, 2008). These 40 SSRs have great potential as highly variable sequencing markers (as opposed to being genotyped) at the interface of intra- and interspecific levels regarding questions of population genetics, species limits, and relationships of closely related species in New Zealand *Veronica.* Additionally, challenges presented by genotyping SSRs in polyploids, such as determining allele dosage and unambiguously identifying alleles (Pfeiffer et al., 2011), are overcome by sequencing the SSRs and their flanking regions, which we would recommend for future studies. For both the LCN and the SSR markers, future sequencing projects could be conducted either using traditional methods (PCR, cloning, sequencing) or using high-throughput sequencing. Furthermore, as biparental nuclear markers, the LCN markers and SSRs will be highly effective in elucidating complex relationships in polyploid *Veronica.*

In addition to the potential advances for *Veronica*, our methodological approach may also be useful for other natural polyploid groups that lack genomic or genetic resources. Natural species that are not associated with economically important crop or other "model" species often lack genomic resources and are very limited regarding the availability of variable genetic markers. Furthermore, developing and establishing such markers using

Fig. 5. MAFFT alignment and GARLI phylogenetic tree (visualized in Geneious) of 54 sequences for a subset of eight New Zealand *Veronica* individuals of *V. chionohebe*, *V. trifida*, and their interspecific hybrid from two South Island locations from sequences of SSR locus SSR-08 mined using QDD from Trinity assemblies of leaf transcriptome data. Consensus and identity sequences are shown at the top. Base pairs that are identical to the consensus are shown in gray, whereas SNPs are shown as colors (red = A, blue = C, green = T, yellow = G, black = N). Each of the eight individuals has a unique color: three individuals of *V. chionohebe* (orange, red, and brown), two of *V. trifida* (blue, pink), and two of their hybrid (light and dark green). For each sequence in the alignment, species names are followed by location (Garvie Mountains or Pisa Range), sequencing plate location (A5, B5, C4, D4, E4, F4, G4, or H4), and number of sequence reads supporting that allele (range: 12–187). Green branches in the GARLI tree to the left of the individual names have >80% bootstrap support (see Fig. 3 for GARLI settings). Voucher information is shown in Appendix 3.

traditional methods (e.g., López-González et al., 2015) can be tedious and time-consuming, with more effort required for fewer microsatellites developed. (As an aside, we found eight of the 12 reported SSR loci from López-González et al. [2015] in our transcript sequences, none of which met the quality criteria of our QDD pipeline for our New Zealand–focused sampled species.) There are nearly 4000 plant transcriptomes in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) and 1000 Plants (1KP) project (www.onekp.com; Matasci et al., 2014) online resources (Hodel et al., 2016). The 1KP transcriptomes have recently been used to develop SSRs for over 1000 plant species (Hodel et al., 2016), whereas Chapman (2015) published a method for the development and validation of 10 COS LCN loci for legume crop species from transcriptomes. By combining Chapman et al.'s (2015) standard wet laboratory approach with a scalable, high-throughput microfluidic PCR strategy (Gostel et al., 2015; Uribe-Convers et al., 2016), we here show that screening of 48 SSR or LCN loci is possible in one microfluidic PCR. In fact, this approach could be scaled up from 48 loci to 480 loci, although the latter might have drawbacks, as here 10 loci are amplified in multiplexed reactions, respectively, and results of these de novo marker sequences can contain PCR chimeras. Nevertheless, combining Fluidigm microfluidic PCR and MiSeq amplicon sequencing of LCN and SSR markers, which were designed in MarkerMiner and QDD from transcriptomic data, is a relatively straightforward high-throughput marker validation method as well as an analysis pipeline that can be used on other natural (and polyploid) systems.

LITERATURE CITED

Abbott, R. J., D. C. Albach, S. W. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* 44: W3–W10.

Albach, D. C. 2007. Amplified fragment length polymorphisms and sequence data in the phylogenetic analysis of polyploids: Multiple origins of *Veronica cymbalaria* (Plantaginaceae). *New Phytologist* 176: 481–498.

Albach, D. C., and J. Greilhuber. 2004. Genome size variation and evolution in *Veronica*. *Annals of Botany* 94: 897–911.

Albach, D. C., M. M. Martínez-Ortega, L. Delgado, H. Weiss-Schneeweiss, F. Özgökce, and M. A. Fischer. 2008. Chromosome numbers in Veroniceae (Plantaginaceae): Review and several new counts. *Annals of the Missouri Botanical Garden* 95: 543–566.

Albach, D. C., and H. M. Meudt. 2010. Phylogeny of *Veronica* in the Southern and Northern Hemispheres based on plastid, nuclear ribosomal and nuclear low-copy DNA. *Molecular Phylogenetics and Evolution* 54: 457–471.

Alvarez, M., A. W. Schrey, and C. L. Richards. 2015. Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Molecular Ecology* 24: 710–725.

Buggs, R. J. A., A. N. Doust, J. A. Tate, J. Koh, K. Soltis, F. A. Feltus, A. H. Paterson, et al. 2009. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): Comparison of natural and synthetic allotetraploids. *Heredity* 103: 73–81.

Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 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.

Chatrou, L. W., M. P. Escribano, M. A. Viruel, J. W. Maas, J. E. Richardson, and J. I. Hormaza. 2009. Flanking regions of monomorphic microsatellite loci provide a new source of data for plant species-level phylogenetics. *Molecular Phylogenetics and Evolution* 53: 726–733.

Chevreux, B., T. Wetter, and S. Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information. *In* Computer Science and Biology: Proceedings of the German Conference on Bioinformatics in Hannover, Germany, October 4–6, 1999, 45–56.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.

De Smet, R., K. L. Adams, K. Vandepoele, M. C. E. Van Montagu, S. Maere, and Y. Van de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences USA* 110: 2898–2903.

De Wit, P., M. H. Pespeni, J. T. Ladner, D. J. Barshis, F. Seneca, H. Jaris, N. Overgaard Therkildsen, et al. 2012. The simple fool's guide to population genomics via RNA-Seq: An introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12: 1058–1067.

Doyle, J. J., L. E. Flagel, A. H. Paterson, R. A. Rapp, D. E. Soltis, P. S. Soltis, and J. F. Wendel. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* 42: 443–461.

Gahlan, P., H. R. Singh, R. Shankar, N. Sharma, A. Kumari, V. Chawla, P. S. Ahuja, and S. Kumar. 2012. *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* 13: 126.

Germain-Aubrey, C. C., C. Nelson, D. E. Soltis, P. S. Soltis, and M. A. Gitzendanner. 2016. Are microsatellite fragment lengths useful for population-level studies? The case of *Polygala lewtonii* (Polygalaceae). *Applications in Plant Sciences* 4: 1500115.

Gostel, M. R., K. A. Coy, and A. Weeks. 2015. Microfluidic PCR-based target enrichment: A case study in two rapid radiations of *Commiphora* (Burseraceae) from Madagascar. *Journal of Systematics and Evolution* 53: 411–431.

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.

Hodel, R. G. J., M. A. Gitzendanner, C. C. Germain-Aubrey, X. Liu, A. A. Crowl, M. Sun, J. B. Landis, et al. 2016. A new resource for the development of SSR markers: Millions of loci from a thousand plant transcriptomes. *Applications in Plant Sciences* 4: 1600024.

Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. *Genome Research* 9: 868–877.

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

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kent, W. J. 2002. BLAT—The BLAST-like alignment tool. *Genome Research* 12: 656–664.

Kotwal, S., S. Kaul, P. Sharma, M. Gupta, R. Shankar, M. Jain, and M. K. Dhar. 2016. *De novo* transcriptome analysis of medicinally important *Plantago ovata* using RNA-Seq. *PLoS ONE* 11: e0150273.

Krasileva, K. V., V. Buffalo, P. Bailey, S. Pearce, S. Ayling, F. Tabbita, M. Soria, et al. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology* 14: R66.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, et al. 2012. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.

López-González, N., E. Mayland-Quellhorst, D. Pinto-Carrasco, and M. M. Martínez-Ortega. 2015. Characterization of 12 polymorphic SSR markers in *Veronica* subsect. *Pentasepalae* (Plantaginaceae) and cross-amplification in 10 other subgenera. *Applications in Plant Sciences* 3: 1500059.

Mandáková, T., S. Joly, M. Krzywinski, K. Mummenhoff, and M. A. Lysak. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis. Plant Cell* 22: 2277–2290.

Matasci, N., L. H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257.

Meglécz, E., N. Pech, A. Gilles, V. Dubut, P. Hingamp, A. Trilles, R. Grenier, and J.-F. Martin. 2014. QDD version 3.1: A user-friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources* 14: 1302–1313.

Meudt, H. M., and M. J. Bayly. 2008. Phylogeographic patterns in the Australasian genus *Chionohebe* (*Veronica* s.l., Plantaginaceae) based on AFLP and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* 47: 319–338.

Meudt, H. M., B. M. Rojas-Andrés, J. M. Prebble, E. Low, P. J. Garnock-Jones, and D. C. Albach. 2015. Is genome downsizing associated with diversification in polyploid *Veronica* lineages? *Botanical Journal of the Linnean Society* 178: 243–266.

Nakasugi, K., R. Crowhurst, J. Bally, and P. Waterhouse. 2014. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *Nicotiana benthamiana. PLoS ONE* 9: e91776.

Pfeiffer, T., A. M. Roschanski, J. R. Pannell, G. Korbecka, and M. Schnittler. 2011. Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). *Journal of Heredity* 102: 479–488.

Sang, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Schwartz, R. S., K. Harkins, A. C. Stone, and R. A. Cartwright. 2015. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. *BMC Bioinformatics* 16: 193.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

Strickler, S. R., A. Bombarely, and L. A. Mueller. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American Journal of Botany* 99: 257–266.

Thiers, B. 2016 [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. Website http://sweetgum.nybg.org/science/ih/ [accessed 14 September 2016].

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.

Uribe-Convers, S., M. L. Settles, and D. C. Tank. 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: Resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). *PLoS ONE* 11: e0148203.

Wagstaff, S. J., and P. Wardle. 1999. Whipcord hebes—Systematics, distribution, ecology and evolution. *New Zealand Journal of Botany* 37: 17–39.

Wagstaff, S. J., M. J. Bayly, P. J. Garnock-Jones, and D. C. Albach. 2002. Classification, origin and diversification of the New Zealand hebes (Scrophulariaceae). *Annals of the Missouri Botanical Garden* 89: 38–63.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin, Austin, Texas, USA.

Appendix 1. Information about the eight individuals of *Veronica* sampled for RNA-Seq.

Species[a]	GPS coordinates	Chromosome number (Ploidy)[b]	1C-value (pg)[c]	Collection locality and collection no. (Voucher)[d]	RNA 260/280 ratio[e]	RNA conc. (ng/μL)[f]	RNA RIN[f]
Veronica catarractae G. Forst.	NA (cultivated plant)	$2n = 42$ (6x)	1.06	Cult. Botanischer Garten Oldenburg (Germany), ex New Zealand, *Meudt s.n.* (OLD00026)	2.11	1017.00	7.90
V. hectorii Hook. f. subsp. *coarctata* (Cheeseman) Garn.-Jones	NA (cultivated plant)	$2n = 40$ (6x)	1.07	Cult. Botanischer Garten Bonn 1342 (Germany), ex New Zealand, *Meudt s.n.* (OLD00029)	1.94	121.00	7.10
V. planopetiolata G. Simpson & J. S. Thomson	44.52247°S, 168.6736916667°E	$2n = 84$ (12x)	2.45	New Zealand: South Island, Otago, *Meudt HMM339a* (WELT SP091593)	1.93	147.00	7.00
V. ochracea (Ashwin) Garn.-Jones	NA (cultivated plant)	$2n = 124$ (18x)	2.97	Cult. Botanischer Garten Bonn 9509 (Germany), ex New Zealand, *Meudt s.n.* (OLD00071)	2.12	1327.00	6.80
V. panormitana Tineo ex Guss.	36.6672°N, 31.8989°E	$2n = 18$ (2x)	0.36	Turkey: north of Paravallar, *Albach 1114 & S272* (OLD00214)	2.00	53.00	8.00
V. trichadena Jord. & Fourr.	39.678536°N, 2.80062°E	$2n = 18$ (2x)	0.39	Spain: Mallorca, *Meudt HMM346L* (OLD00086)	1.98	302.00	7.50
V. cymbalaria Bodard	36.5325°N, 31.99°E	$2n = 36$ (4x)	0.76	Turkey: Alanya Castle, *Albach 1235* (OLD01171)	2.04	245.00	6.90
V. cymbalaria	37.22778°N, 31.12972°E	$2n = 54$ (6x)	1.38	Turkey: Anatalya, Selgedos, *Albach 1087 & S300* (OLD00481)	2.11	1265.00	7.60

Note: NA = not applicable.

[a]RNA was extracted from leaf material from greenhouse-grown material of all individuals except *V. planopetiolata*, which was from field-collected leaf material stored in RNA*later* (Life Technologies, Carlsbad, California, USA).

[b]Chromosome numbers are from the literature (Albach et al., 2008).

[c]1C-values (Meudt et al., 2015) were assessed for the same individual from which RNA was extracted for this study except for *V. panormitana*, whose 1C-value is based on the average of five other individuals from three different Turkish populations (range 0.35–0.37 pg; Meudt et al., 2015).

[d]Voucher specimens are lodged at herbaria at the Museum of New Zealand Te Papa Tongarewa (WELT) or Carl-von-Ossietzky Universität Oldenburg (OLD).

[e]RNA 260:280 ratio was calculated using the Tecan Infinite Pro F200 (Tecan, Crailsheim, Germany).

[f]RNA concentration and RNA Integrity Number (RIN) were calculated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany); please note that the cDNA construction was made with normalized RNA.

Appendix 2. Information about the 48 individuals of *Veronica* sampled for the LCN marker validation.

Species	Subgenus	Ploidy[a]	Chromosome no.[a]	Country	Voucher (Herbarium and/or Herbarium accession no.)[b]	Location on sequencing plate	No. of LCN markers successfully sequenced (of 48 total)
Lagotis integrifolia (Willd.) Schischk. ex Vikulova	(outgroup)	4	44	Kazakhstan	Tribsch & Essl 10986 (WU)	D1	40
Paederota lutea L. f.	(outgroup)	4	36	Austria	Albach 209 (WU)	B1	32
Veronicastrum stenostachyum (Hemsl.) T. Yamaz.	(outgroup)	4	34	China	Albach 123 (K)	C1	29
Wulfenia carinthiaca Jacq.	(outgroup)	2	18	cult.	Albach 74 (BONN)	A1	36
Veronica anagallis-aquatica L.	*Beccabunga*	?	?	Czech Republic	597087 (BRUENN)	F1	28
V. catenata Pennell	*Beccabunga*	2	18	Czech Republic	597095 (BRUENN)	G1	36
V. gentianoides Vahl	*Beccabunga*	?	?	Georgia	Schneeweiss Geo02/43 (WU)	H1	27
V. arvensis L.	*Chamaedrys*	2	16	Germany	Albach 147 (WU)	B3	19
V. chamaedrys L.	*Chamaedrys*	4	32	Norway	Albach 121 (K)	C3	20
V. crista-galli Steven	*Cochlidiosperma*	2	18	Georgia	Dolmkanov 17.4.1983 (TBS)	G2	20
#V. cymbalaria Bodard	*Cochlidiosperma*	4	36	Turkey	Albach 1235 (OLD01171)	F4	12
#V. cymbalaria	*Cochlidiosperma*	6	54	Turkey	Albach 1087 (OLD00481)	G4	25
V. javanica Blume	*Cochlidiosperma*	2	16		Murata et al. 10050 (BM)	F2	4
#V. panormitana Tineo ex Guss.	*Cochlidiosperma*	2	18	Turkey	Albach 1114 (OLD00214)	D4	21
#V. trichadena Jord. & Fourr.	*Cochlidiosperma*	2	18	Spain	Meudt HMM346L (OLD00086)	E4	18
V. triloba (Opiz) Opiz	*Cochlidiosperma*	2	18	Turkey	Albach 242 (WU)	H2	21
V. brownii Roem. & Schult.	*Labiatoides*	12	72	Australia	NSW 285360	B4	33
V. triphyllos L.	*Pellidosperma*	2	14	Russia	S434, BG Osnabrück, 961; RU, Altei, 1900 m	A3	12
V. cuneifolia D. Don	*Pentasepalae*	2	16	Turkey	Albach 1159 (OLD)	G3	21
V. fuhsii Freyn & Sint.	*Pentasepalae*	?	?	Turkey	Albach 897 (VANF, WU)	F3	32
V. prostrata L.	*Pentasepalae*	2	16	Austria	Albach 860 (MZJG)	E3	21
V. filiformis Sm.	*Pocilla*	2	14	Germany	Albach 144 (WU)	D3	18
V. longifolia L.	*Pseudolysimachium*	4	34	Turkey	Behcet 7435 (OLD)	C2	18
V. longifolia	*Pseudolysimachium*	?	?	UK	Sheahan 48 (K)	D2	18
V. schmidtiana Regel	*Pseudolysimachium*	4	34	Japan	Umezawa 20130 (WU)	A2	19
V. spicata L.	*Pseudolysimachium*	8	68	Austria	Bardy 60 (WU)	B2	11
V. fruticans Jacq.	*Stenocarpon*	2	16	UK	Viv Halcro VH030 (K)	A4	31
V. missurica Raf.	*Synthyris*	4	24	USA	Albach 124 (K)	E2	15
V. chamaepithyoides Lam.	*Triangulicapsula*	4	24	Spain	UA 174 (SALA)	H3	39
V. scutellata L.	*Veronica*	4	36	Austria	Dobes 7026 (WU)	E1	30
V. albiflora (Pennell) Albach	*Pseudoveronica*	6	42	New Guinea	Johns 8965 (K)	C4	36
V. baylyi Garn.-Jones	*Pseudoveronica*	18	116	New Zealand	Garnock-Jones PGJ 2868 (OLD)	C6	22
#V. catarractae G. Forst.	*Pseudoveronica*	6	42	New Zealand	Meudt HMM s.n. (OLD00026)	B5	13
V. colostylis Garn.-Jones	*Pseudoveronica*	6	42	New Zealand	Meudt HMM341C (OLD)	F5	20
V. cupressoides Hook. f.	*Pseudoveronica*	6	42	New Zealand	Garnock-Jones PGJ 2887 (OLD)	A6	26
V. densifolia F. Muell.	*Pseudoveronica*	6	42	New Zealand	Meudt HMM337A (WELT SP091591)	H5	36
#V. hectorii Hook. f. subsp. *coarctata* (Cheeseman) Garn.-Jones	*Pseudoveronica*	6	40	New Zealand	Meudt HMM s.n., cult. Bonn 13428 ex New Zealand (OLD00029)	H4	16
V. hulkeana F. Muell. ex Hook. f.	*Pseudoveronica*	6	42	New Zealand	Garnock-Jones PGJ 2874 (OLD)	H6	11
V. lavaudiana Raoul	*Pseudoveronica*	6	42	New Zealand	Garnock-Jones PGJ 2881 (OLD)	G5	29
V. macrantha Hook. f.	*Pseudoveronica*	6	42	New Zealand	Clarke s.n., cult. K 1969-35034 ex New Zealand (OLD)	B6	16
V. melanocaulon Garn.-Jones	*Pseudoveronica*	6	42	New Zealand	Garnock-Jones PGJ 2883 (OLD)	D5	29
#V. ochracea (Ashwin) Garn.-Jones	*Pseudoveronica*	18	124	New Zealand	Meudt HMM s.n., Bonn 9509 (OLD00071)	A5	26
V. pinguifolia Hook. f.	*Pseudoveronica*	12	80	New Zealand	Meudt HMM s.n. cult. Bonn 265 ex New Zealand (OLD)	D6	17

Appendix 2. Continued.

Species	Subgenus	Ploidy[a]	Chromosome no.[a]	Country	Voucher (Herbarium and/or Herbarium accession no.)[b]	Location on sequencing plate	No. of LCN markers successfully sequenced (of 48 total)
#*V. planopetiolata* G. Simpson & J. S. Thomson	*Pseudoveronica*	12	84	New Zealand	Meudt HMM339a (WELT SP091593)	C5	22
V. senex (Garn.-Jones) Garn.-Jones	*Pseudoveronica*	6	42	New Zealand	Garnock-Jones PGJ 2879 (OLD)	E5	28
V. speciosa R. Cunn. ex A. Cunn.	*Pseudoveronica*	6	40	New Zealand	Garnock-Jones PGJ 2878 (OLD)	F6	33
V. tairawhiti (B. D. Clarkson & Garn.-Jones) Garn.-Jones	*Pseudoveronica*	12	80	New Zealand	Garnock-Jones PGJ 2888 (OLD)	G6	9
V. townsonii Cheeseman	*Pseudoveronica*	6	40	New Zealand	Garnock-Jones PGJ 2901 (WELT SP103482)	E6	26

Note: LCN = low-copy nuclear.

[a] Ploidy and chromosome numbers are from the literature (Albach et al., 2008).

[b] Herbaria acronyms follow Thiers (2016).

#RNA-Seq sample.

Appendix 3. Validation of 48 SSR markers on 48 individuals of 20 species of Southern Hemisphere *Veronica* subg. *Pseudoveronica*.

Species name	Section and informal group	Ploidy[a]	Chromosome no.[a]	Country	Voucher and collection locality (Herbarium and/or Herbarium accession no.)[b]	Location on sequencing plate	No. SSR loci successfully sequenced (of 48 total)
Veronica calycina R. Br.	sect. *Labiatoides*	6	36	Australia	RGC 19644, near Lithgow, NSW (NSW, OLD)	C3	40
V. derwentiana Andrews subsp. *subglauca* (B. G. Briggs & Ehrend.) B. G. Briggs	sect. *Labiatoides*	6	40	Australia	RGC 19649, near Lithgow, NSW (NSW, OLD)	D3	37
V. chionohebe Garn.-Jones	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1823, Pisa Range (WELT SP084028/A)	E4	42
V. chionohebe	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1824, Pisa Range (WELT SP084029)	F4	38
V. chionohebe	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1844, Garvie Mountains (WELT SP084043)	C4	40
V. chionohebe	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1845, Garvie Mountains (WELT SP084044)	D4	40
V. chionohebe × *V. trifida* Petrie	sect. *Hebe*, snow hebe × speedwell hebe hybrid	6	42	New Zealand	MJB 1848, Garvie Mountains (WELT SP084059)	G4	38
V. chionohebe × *V. trifida*	sect. *Hebe*, snow hebe × speedwell hebe hybrid	6	42	New Zealand	MJB 1849, Garvie Mountains (WELT SP084060/A)	H4	39
V. ciliolata (Hook. f.) Garn.-Jones subsp. *ciliolata*	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1696, Mt. Brewster (WELT SP083925)	D6	40
V. ciliolata subsp. *ciliolata*	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1813, Mt. Cook (WELT SP084020)	C6	43
V. ciliolata subsp. *fiordensis* (Ashwin) Meudt	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1673, Mt. Burns (WELT SP083910)	A6	40
V. ciliolata subsp. *fiordensis*	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1837, Livingstone Range (WELT SP084037)	B6	42
V. densifolia F. Muell.	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1805, Hunter Hills (WELT SP084053)	H6	41
V. densifolia	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1858, Garvie Mountains (WELT SP084058)	G6	38
V. pulvinaris (Hook. f.) Cheeseman	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1728, Temple Basin (WELT SP083950)	E6	42
V. pulvinaris	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1761, Mt. Arthur (WELT SP083968)	F6	37
V. thomsonii Cheeseman	sect. *Hebe*, snow hebe	6	42	New Zealand	HMM 259, Mt. St. Bathans (WELT SP085925)	F5	41
V. thomsonii	sect. *Hebe*, snow hebe	6	42	New Zealand	HMM 261, Mt. St. Bathans (WELT SP085937)	H5	40
V. thomsonii	sect. *Hebe*, snow hebe	6	42	New Zealand	HMM 265, Mt. St. Bathans (WELT SP085931)	G5	40
V. thomsonii	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1851, Garvie Mountains (WELT SP084047/A)	C5	43
V. thomsonii	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1852, Garvie Mountains (WELT SP084048)	D5	39
V. thomsonii	sect. *Hebe*, snow hebe	6	42	New Zealand	MJB 1853, Garvie Mountains (WELT SP084049)	E5	42
V. trifida Petrie	sect. *Hebe*, speedwell hebe	6	42	New Zealand	MJB 1841, Garvie Mountains (WELT SP084041)	A5	37
V. trifida	sect. *Hebe*, speedwell hebe	6	42	New Zealand	MJB 1842, Garvie Mountains (WELT SP084041)	B5	41
V. brachysiphon (Summerh.) Bean	sect. *Hebe*, hebe	18	120	New Zealand	PGJ 2902, cult. Otari (WELT SP103452)	G2	40
V. brachysiphon (as *Hebe vernicosa* in Kew Gardens)	sect. *Hebe*, hebe	18	120	New Zealand	HMM s.n., cult. Kew Gardens 1997-5679 (OLD)	H2	35
V. catarractae G. Forst.	sect. *Hebe*, speedwell hebe	6	42	New Zealand	PGJ 2875, cult. Wellington (OLD)	B1	0

Appendix 3. Continued.

Species name	Section and informal group	Ploidy[a]	Chromosome no.[a]	Country	Voucher and collection locality (Herbarium and/or Herbarium accession no.)[b]	Location on sequencing plate	No. SSR loci successfully sequenced (of 48 total)
#*V. catarractae* (purchased as *Parahebe* 'Snow')	sect. *Hebe*, speedwell hebe	6	42	New Zealand	HMM s.n., cult. Botanischer Garten Oldenburg (OLD00026)	A1	41
V. colostylis Garn.-Jones	sect. *Hebe*, speedwell hebe	6	42	New Zealand	HMM338a, Arrowtown (WELT SP091592)	H3	8
V. colostylis	sect. *Hebe*, speedwell hebe	6	42	New Zealand	HMM341c, Moke Creek (WELT SP091595)	G3	30
V. hectorii Hook. f.	sect. *Hebe*, hebe	6	40	New Zealand	PGJ 2910, cult. Otari (WELT SP103460)	D1	38
#*V. hectorii* subsp. *coarctata* (Cheeseman) Garn.-Jones	sect. *Hebe*, hebe	6	40	New Zealand	HMM s.n., Bonn 13428 (OLD00029)	C1	41
V. hulkeana F. Muell. ex Hook. f. subsp. *evestita* (Garn.-Jones) Garn.-Jones 'Lena'	sect. *Hebe*, sun hebe	6	42	New Zealand	PGJ 2874, cult. Wellington (OLD)	A4	32
V. lavaudiana Raoul	sect. *Hebe*, sun hebe	6	42	New Zealand	PGJ 2881, cult. Wellington (OLD)	B4	41
V. macrantha Hook. f.	sect. *Hebe*, unresolved, early branching	6	42	New Zealand	HMM s.n., cult. Kew Gardens 1969-35034 (OLD)	D2	36
V. macrantha	sect. *Hebe*, unresolved, early branching	6	42	New Zealand	PGJ 2924, cult. Otari (WELT SP103475)	C2	41
#*V. ochracea* (Ashwin) Garn.-Jones	sect. *Hebe*, hebe	18	124	New Zealand	HMM s.n., Bonn 9509 (OLD00071)	E1	42
V. ochracea	sect. *Hebe*, hebe	18	124	New Zealand	PGJ 2911, cult. Otari (WELT SP103461)	F1	36
V. ochracea 'James Stirling'	sect. *Hebe*, hebe	18	124	New Zealand	HMM s.n., cult. Kew Gardens 1992-1403 (OLD)	G1	39
V. odora Hook. f. (as *Hebe vernicosa* in Botanischer Garten Bonn)	sect. *Hebe*, hebe	12	84	New Zealand	HMM s.n., cult. Bonn 17475 (OLD)	A3	29
V. odora 'New Zealand Gold'	sect. *Hebe*, hebe	12	84	New Zealand	HMM s.n., cult. Kew Gardens 1989-2000 (OLD)	B3	40
#*V. planopetiolata* G. Simpson & J. S. Thomson	sect. *Hebe*, speedwell hebe	12	84	New Zealand	HMM339a, Shotover Saddle (WELT SP091593)	H1	42
V. planopetiolata	sect. *Hebe*, speedwell hebe	12	84	New Zealand	HMM339b, Shotover Saddle (WELT SP091593)	A2	36
V. planopetiolata	sect. *Hebe*, speedwell hebe	12	84	New Zealand	HMM339c, Shotover Saddle (WELT SP091593)	B2	43
V. salicornioides Hook. f.	sect. *Hebe*, hebe	6	42	New Zealand	HMM s.n., cult. Kew Gardens 1989-2004 (OLD)	F2	38
V. salicornioides	sect. *Hebe*, hebe	6	42	New Zealand	PGJ 2923, cult. Otari (WELT SP103474)	E2	42
V. vernicosa Hook. f.	sect. *Hebe*, hebe	6	42	New Zealand	PGJ 2925, cult. Otari (WELT SP103476)	E3	41
V. vernicosa	sect. *Hebe*, hebe	6	42	New Zealand	PGJ 2926, cult. Otari (WELT SP103477)	F3	39

[a] Ploidy and chromosome numbers are from the literature (Albach et al., 2008).

[b] Herbaria acronyms follow Thiers (2016). Voucher specimens are lodged at herbaria at the Museum of New Zealand Te Papa Tongarewa (WELT), Carl-von-Ossietzky Universität Oldenburg (OLD), or National Herbarium of New South Wales (NSW). Collection initials: MJB = Michael J. Bayly, HMM = Heidi M. Meudt, PGJ = Phil Garnock-Jones, RGC = R. G. Coveny.

#RNA-Seq sample.

Appendix 4. Validation of 48 LCN markers on 48 individuals of 46 species of *Veronica*, representing all subgeneric lineages in the genus.

Locus	Primer sequences (5′–3′)	Sequence same as original transcript?	No. of individuals successfully sequenced	No. of different sequences in GARLI alignment	Length (bp, range)	Length (bp, mean)	No. SNPs	No. PICs	*A. thaliana* gene
LCN-03	F: AGCAGTGCCTCTAGTCTGTTT R: CCGCTAATGGCACCTGAATTG	complete	18	37	158–866	480	607	345	AT3G07080: EamA-like transporter family
LCN-04	F: AGGTTTATACATTTGCGGCG R: TTCCCGCACCCTCCAAAC	complete	42	43	288–331	327	117	77	AT3G07720: Galactose oxidase/ kelch repeat superfamily protein
LCN-08	F: CCCTCCAGAGAAGAGCTTAACG R: GCCCTTTGCCTCCTCCATATAG	complete	27	44	311–791	374	427	131	AT4G17100: UNKNOWN
LCN-10	F: GCAAAGACCAGTTCAAACTTTGAG R: AGAGGCTTGCTGACCTTCAAC	complete	35	68	234–820	440	523	303	AT4G33460: ABC transporter family protein
LCN-13	F: TCTAACTGGTTGTCATCCGCT R: CCAAGGATCCAAGAGCCCATT	partial	18	19	310–912	422	354	112	AT5G65760: Serine carboxypeptidase S28 family protein
LCN-20	F: GGCATACGTGAAGACCTGGG R: AGCAACAATGGCACCACTTG	partial	44	85	306–681	384	280	176	AT1G57770: FAD/NAD(P)-binding oxidoreductase family protein
LCN-25	F: AGGAGTGATTCGAGCAGTGC R: ACTTGTTCCCCCAATCCACC	partial	43	89	310–726	396	439	287	AT2G05830: NagB/RpiA/CoA transferase-like superfamily protein
LCN-38	F: AAGACCCTTGGAGGATGGGA R: TAGTGCTCTTTCGCCACTCC	complete	42	104	310–762	361	436	299	AT3G59380: farnesyltransferase A
LCN-43	F: TATGACTGCTGCTGGTCTTGG R: AGACCACGTTCTAATTCGCCA	complete	30	50	139–750	396	329	185	AT4G35850: Pentatricopeptide repeat (PPR) superfamily protein
LCN-46	F: TGCAACTCCTTTTTGGGGGT R: ACTTCATCATGGGGGCAGTG	complete	44	85	307–707	373	314	224	AT5G13800: pheophytinase
LCN-48	F: AAGGTAACGCCGCCAAGTAT R: TGCGCAGTTTATGGGTACGA	complete	34	42	147–734	379	445	280	AT5G14520: pescadillo-related
LCN-01	F: AGCATCGCTTGGACAGGTTTA R: ATTCCCCCATCATGCCGAAAT	no	17						AT1G71810: protein kinase superfamily protein
LCN-02	F: TGGGAGCAGCGCCTTAATTC R: CCACAACATCCCTCTTCAGCT	no	22						AT2G25950: protein of unknown function (DUF1000)
LCN-05	F: TTGCCGCCTCCTGATCATATC R: AGAACTGCAACATCTCTGGCA	partial	12						AT3G20790: NAD(P)-binding Rossmann-fold superfamily protein
LCN-06	F: GTGAGCAGGTTTTTCGAGTGG R: AAGCTTCTGCACTCCCTTTGA	no	33						AT4G09730: RH39
LCN-07	F: GGAGATCAATCGCTTTTGGAGTC R: TGGCATATTGTTCAACTCCATCG	no	0						AT4G09750: NAD(P)-binding Rossmann-fold superfamily protein
LCN-09	F: AAAGCTGGTGAACTTGCAGTG R: GGCAGCCCATAAGCAATGTTC	no	16						AT4G25450: nonintrinsic ABC protein 8
LCN-11	F: GTGCATTTGCCATGGAATCCC R: TACGTCCACGACCGTTATTCC	no	3						AT4G37040: methionine aminopeptidase 1D
LCN-12	F: GGAATGGTGGTAGGATTGGGG R: CCTCCAAACCTCAGCATCTCC	no	20						AT5G44520: NagB/RpiA/CoA transferase-like superfamily protein

Appendix 4. Continued.

Locus	Primer sequences (5′–3′)	Sequence same as original transcript?	No. of individuals successfully sequenced	No. of different sequences in GARLI alignment	Length (bp, range)	Length (bp, mean)	No. SNPs	No. PICs	*A. thaliana* gene
LCN-14	F: CGGATCGTTACATTGCTAGCTG R: GCACCTGACAAGCAAACTGTAG	no	13						AT1G04420: NAD(P)-linked oxidoreductase superfamily protein
LCN-15	F: CGGTGGGTGGAAGCATTTTG R: TCCAACAGAAGTGGACCAGC	partial	28						AT1G16180: Serine-domain containing serine and sphingolipid biosynthesis protein
LCN-16	F: ACTCCTTTCCCGCATTCCTG R: CCTCACCATCTCGAAGCTGG	no	30						AT1G19600: pfkB-like carbohydrate kinase family protein
LCN-17	F: AGACTCTACCCACAGCCTCC R: TGGGGATGATAGGGGGCC	no	11						AT1G31800: cytochrome P450, family 97, subfamily A, polypeptide 3
LCN-18	F: AGTTTGGTGGTGGGCATAGG R: GAAGATCAGGCTCGGGGAAG	partial	24						AT1G48520: GLU-ADT subunit B
LCN-19	F: CTGTTGCGCTTGGGTCATG R: TTGAGCTCCACCAAGACCAC	no	31						AT1G53280: Class I glutamine amidotransferase-like superfamily protein
LCN-21	F: TGGTGTCATTGGAGCTGGTC R: TGCCATTCCTTCTCGAGTCC	no	9						AT1G68010: hydroxypyruvate reductase
LCN-22	F: TGGGTGAAGGGTCTTTTGGTG R: CCAACTCTCAAATCAGTAGCTGC	no	21						AT1G68830: STT7 homolog STN7
LCN-23	F: AAGCATGTGGGAGAAGAGGC R: CAAGCACCAATCGCTCTGAC	no	24						AT1G71240: Plant protein of unknown function (DUF639)
LCN-24	F: GGAACTCCTATGCCTCAGGTTG R: TCTTCATTAGTTGTCCCCACACC	no	13						AT1G75210: HAD-superfamily hydrolase, subfamily IG, 5′-nucleotidase
LCN-26	F: GATAACTGGAGCGACGGGATT R: GCTAGAGCACCACCCTCTTTT	no	32						AT2G21280: NAD(P)-binding Rossmann-fold superfamily protein
LCN-27	F: TGGGATGCAGTATCATTGGCA R: CAGCTGTAGGTTGTGACTGGT	partial	18						AT2G23390: UNKNOWN
LCN-28	F: TGCCTCCACCAGTCAAGATG R: CCATCCTCCCCAAGCATCAA	no	16						AT2G27680: NAD(P)-linked oxidoreductase superfamily protein
LCN-29	F: GCTAGAGCCCCAAAGAGCAA R: TCCTCCACATATGCAACCGG	partial	11						AT2G30390: ferrochelatase 2
LCN-30	F: ATGGAAAGGAGTGGGAGCTG R: TTGGCTGGACTGACCCATTC	no	22						AT2G44760: Domain of unknown function (DUF3598)
LCN-31	F: TCAACTTTGCAGCATTGGAGC R: CAACAGCGGCAATGTCAAAGA	partial	19						AT3G06510: Glycosyl hydrolase superfamily protein
LCN-32	F: AAAATGGGTGCTGCTGTTGG R: ACAAGGCCATACCCATGCAT	no	39						AT3G17810: pyrimidine 1
LCN-33	F: TGCACGATCACCTCCTTGTC R: AGAATGGTTCCGGAGCTGTG	no	28						AT3G17940: Galactose mutarotase-like superfamily protein
LCN-34	F: CACAGAAAGGCAGAATCAGGC R: TGATCCAATCAGAGGTGCGT	partial	12						AT3G23620: Ribosomal RNA processing Brix domain protein

Appendix 4. Continued.

Locus	Primer sequences (5′–3′)	Sequence same as original transcript?	No. of individuals successfully sequenced	No. of different sequences in GARLI alignment	Length (bp, range)	Length (bp, mean)	No. SNPs	No. PICs	*A. thaliana* gene
LCN-35	F: AAATCGCTCACCGGTGTTTG R: TTGCAGTTGGGAAGTTCCAAAA	partial	9						AT3G52190: phosphate transporter traffic facilitator1
LCN-36	F: GATCCGGGTCAAATCCACCA R: AACGGCAATGACAATGGCAC	partial	31						AT3G56460: GroES-like zinc-binding alcohol dehydrogenase family protein
LCN-37	F: CAAGGAGCTTGGTAGGAGGC R: GAGACAGAAGAAGCGGGACC	no	7						AT3G56940: dicarboxylated iron protein, putative (Crd1)
LCN-39	F: CCGGTGATCTTGTTCGCATG R: AATTGGAGCGCTCGACTCTT	no	36						AT3G62910: Peptide chain release factor 1
LCN-40	F: TGGGAAACTCGGAATGGGTG R: CGGAATGCTGCTTGATGTGT	no	43						AT4G02790: GTP-binding family protein
LCN-41	F: AGGTGGGCTGAATGGAATGG R: CCTCCAATTGTCCCCACTGG	no	13						AT4G09020: isoamylase 3
LCN-42	F: AAGTGGTTGCCGTGCCAT R: GCCTCTGGTCGTATGTAGCC	partial	9						AT4G21470: riboflavin kinase/FMN hydrolase
LCN-44	F: ACAAAGGATGAGATCGAACGGT R: TGCCCAAGAAAGTGCTGAAAC	partial	14						AT5G06260: TLD-domain containing nucleolar protein
LCN-45	F: GGCAGACTTGGTCATGGACA R: CCCCAGCCCCATGTGTAAAT	no	36						AT5G08710: Regulator of chromosome condensation (RCC1) family protein
LCN-47	F: TTCTGCAGCAGCTCAAAGGA R: AAATCTCTGGCGCTCTCGTC	no	22						AT5G14250: Proteasome component (PCI) domain protein

Note: LCN = low-copy nuclear; PIC = parsimony informative character; SNP = single-nucleotide polymorphism.

Applications in Plant Sciences 2016 4(10): 1600091
doi:10.3732/apps.1600091

APPENDIX 5. Validation of 48 SSRs on a subset of the 48 New Zealand and Australian individuals of *Veronica* sequenced. Shown are eight individuals of the *Veronica chionohebe/V. trifida* subset (A5, B5, C4, D4, E4, F4, G4, and H4; see Appendix 3).

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles											Pairwise genetic distance				No.			Notes
			Range	Mean		A5	B5	C4	D4	E4	F4	G4	H4	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-01	F: TGGAACAGCCATTGCATCAAA R: TCGTCGACTTACCAGTTCCAG	ACA (ATG)	310–692	353	8	3	2	2	3	3	3	3	4	2	4	2.9	0	0.03	0.01	0.01	23	14	2	two large introns, motif in central exon, sequences partially not covering complete locus
SSR-02	F: GATTGTTTCAGCCAAGAGATTCTCA R: CTTGTTCCGACGCAGACCAT	GAT	208–476	328	7	0	2	2	2	1	1	1	3	1	3	1.5							?	incomplete; several genes amplified by primers; same locus as SSR-42
SSR-03	F: TTGAGACGCAAGATTTCTGCAA R: CCCTCACGCGCTCTATCATT	ACT																					at least 1	several genes amplified by primers
SSR-04	F: TTGTTCAACCAGTCGGACGT R: CCGCTTCGAGGACTTGCTAG	GAT	127–239	216	8	8	3	4	5	4	4	5	5	3	8	4.8	0	0.06	0.02	0.01	23	15	0	
SSR-05	F: GTCGAAATCGGATTTACTAGCTAAGT R: AGTCGGGAAAGAGATTGGGC	CATA	293–326	298	8	4	4	1	1	1	4	5	6	1	6	3.3	0	0.08	0.04	0.02	28	27	0	
SSR-06	F: AATAAACTGACGACAGCGCG R: ACTGTGAGTCTGCCTTACGC	TGA			3																		?	
SSR-07	F: AGCAGTGAGAGCCAACATCC R: CGAAACGCCCTCTTACACGA	TAC	358–424	386	8	7	10	10	9	9	9	11	10	7	11	9.4	0	0.33	0.18	0.14	190	174	1	three orthologues?
SSR-08	F: CCATCAAACCCTTCCAAGCTG R: TGGCCTCTTACTTCCTACGTG	GAT	311–387	357	8	9	10	4	3	6	6	8	8	3	10	6.8	0	0.14	0.08	0.08	126	109		
SSR-09	F: TGGTCACTCTTTCGTGTTGGA R: CCATAAATTTGTGCTGCCTCCA		310–401	325		2	1	3	1	2	0	0	0	1	0	0							3	at least three introns, sequence not covering complete locus
SSR-10	F: CGTAAATTGGATCAGGTCGCC R: CGTAGCTAGTTTGTCATTGGATGG	AGT	266–280	272	8	3	4	1	1	4	3	4	4	1	4	3	0	0.04	0.02	0.02	20	14	1	
SSR-11	F: AAACGACGTCGGACTGAGAC R: GGGATAACATTGCTCACTCACC	ACGA (TTG, ATT, AG)	264–293	283	8	2	8	9	4	5	5	4	4	2	9	5.1	0	0.08	0.04	0.04	42	31	0	two orthologues?
SSR-12	F: TTGCAGTCGGCTTTAAAGATCC R: ATACCAGCCATATCAGAGCGC	AATC															0	0.07	0.02	0.02			0	two orthologues?
SSR-13	F: TCCTTCCTACTTGCCAAACTCT R: TCACGCACAGAGGACTGAAC																							
SSR-14	F: TGTTGACTCAATCCGTCTCCG R: TCTGCTTTGCTACCTGTCTTCT	TTAA	290–306	293	8	2	1	1	1	4	4	2	2	1	4	2.1	0	0.02	0.01	0.01	9	6	0	one unambiguous locus
SSR-15	F: GGCAGAAGAAACGGTTGCAG R: GACCTTTATGCCGTCTGCCT	GAT	310–355			2	2	0	0	2	0	0	0	1	0	0							2	sequences not covering complete locus and do not overlap
SSR-16	F: GAGACAACTGCTGCACTTGC R: TTAGTCCACCAGTGTCCACG	ATC	301–620	377	7	3	4	2	3	4	1	0	3	1	4	2.5	0	0.51	0.06	0.02	91	43	1	sequences nearly not overlapping
SSR-17	F: AACTTGCTCGTCTCCACCAG R: CCGATGGATTCAGAAACCAACAA	GAT	203–257	238	8	1	1	7	2	5	6	5	6	1	7	4.1	0	0.09	0.04	0.03	39	27	0	two orthologues?
SSR-18	F: TCTGTGCTACAACTAGTACAAGGAG R: GGATGGATCCCTTTCTTGAAATAAGG																							not reference transcript sequenced
SSR-19	F: TGGCAACATGCAACTGTGTT R: ACGAGAATACCATACTTCATGTTCG	TATC (ATA, TAC)	268–303	282	8	6	6	2	1	4	1	4	3	1	6	3.4	0	0.09	0.04	0.03	35	33	0	two orthologues?
SSR-20	F: CATTCGTATTACTGTAAATGGTTTGCC R: GCAAACAGCACAAATATTTCACCA	GTTA (ACA, GTGA)	186–253	228	8	2	2	8	3	4	8	8	5	2	8	5	0	0.1	0.04	0.05	33	31	0	two orthologues?

Appendix 5. Continued.

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles											Pairwise genetic distance				No.			Notes
			Range	Mean		A5	B5	C4	D4	E4	F4	G4	H4	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-21	F: ATGGATGAAGGGCCAGTTAAGG R: CCGCCAACTCCTCATCTAATTCA	GAT	238–256	246	8	5	3	4	5	1	6	5	6	1	6	4.4	0	0.12	0.05	0.02	38	32	0	two or three orthologues?
SSR-22	F: AGGGTCGTTATGGAAACCGG R: GACATCACCAGTCATCCGCA	GAT	286–348	332	8	1	1	2	2	5	1	4	4	1	5	2.5	0	0.36	0.11	0.01	111	99	1	two orthologues?
SSR-23	F: CACAACCAAAGTAGCAGCACT R: TGTGAGTTCGCGTAAAGGGA																							three orthologues sequenced? sequence different to transcript
SSR-24	F: GATGCCATTGTTGGATGAATTTCG R: AGCTGCAACTCCTCCTTCAA																							sequences different to transcript
SSR-25	F: GGTGGTAAAGGCACCGTTAGA R: CGACGAGCTCAGGTACGTC				0																			not amplified/ sequenced
SSR-26	F: GTGCGCGAACAAGTTTGGTT R: TCACTAATCCACCTGATCCGTC	ATC	288–315	304	8	3	3	3	2	3	4	3	4	2	4	3.1	0	0.19	0.12	0.14	88	80	0	three orthologues?
SSR-27	F: CGGAGAGGTGCAATATACAAATGT R: GGACAACGCATTAGGAAGTGG	ACT	259–262	260	8	2	1	2	1	2	4	3	3	1	4	2.3	0	0.02	0.01	0.01	8	4	0	one unambiguous locus
SSR-28	F: GCGAAATGCAACATTCCACTG R: GGAGACACGGAACCTGAACA	ACT	266–284	276	8	3	4	5	3	4	7	5	5	3	7	4.5	0	0.05	0.02	0.02	25	17	0	two orthologues?
SSR-29	F: GACACCAAACTTGTCTTCAACGT R: AAAGAGGTTGTGAATTCACTAGAAGTT	ACT	300–355	342	8	6	13	2	2	6	2	7	10	2	13	6	0	0.14	0.03	0.01	65	48	1	two orthologues?
SSR-30	F: TTCTTGCTCTTGTGTTGGTTCC R: TTCACTTCAAACCTTTGTCACTACC	TGA	278–290	283	8	2	1	2	2	1	8	2	3	1	8	2.6	0	0.04	0.01	0.01	16	10	0	three orthologues?
SSR-31	F: CGATGACGATGAGGACGACG R: CATTTGATGCACCTCCATGCT				2																			sequences different to transcript
SSR-32	F: GTGCCTAGATATCACCAAGATAGAAGA R: GACCAGAAGATCAGACTCAGCA	GAT	158–248	235	8	1	2	4	4	7	7	3	4	1	7	4	0	0.07	0.03	0.03	17	21	0	two orthologues?
SSR-33	F: GCTGCACCTGGGATTCAAAG R: ACTGTGAGTCTGCCTTACGC				5																			three orthologues sequenced? sequence different to transcript
SSR-34	F: ATTGCTCAACATGTTTGCCTCT R: TGTCACAGTTTGGCGATATTGG				3																			sequences different to transcript
SSR-35	F: TCGTCATCGCTGAAACCATCA R: ACACTTGATCTGCTTGTTGCC	ATC (CAA)	317–501	481	8	5	6	4	5	6	3	1	8	1	8	4.8	0	0.34	0.06	0.05	141	59	1	three orthologues?
SSR-36	F: AAACCCAATTCAAAGCAATGACAC R: ACCCTCATTTCTCCAAACCAACT	TCA	238–253	245	8	3	2	2	1	2	2	3	3	1	3	2.3	0	0.07	0.02	0	17	13	0	two orthologues?
SSR-37	F: AGTTGACGCCTTGTTTGGTTC R: CACGCAAACACCACATTCCC	GAT	112–280	272	8	13	16	9	13	11	18	19	20	9	20	14.9	0	0.09	0.03	0.03	45	30	0	two or three orthologues?
SSR-38	F: CCCTAAAGTTCAAGCATCTATACCAG R: TGCTGCAGCTTCAAATGTTTCA		310–569	521	8	7	6	6	4	6	6	6	7	4	7	6	0	0.27	0.09	0.1	174	154	2	three orthologues?
SSR-39	F: ACTTGCTGCAACTTGCTAAACA R: TGGATGACAATGAAAGAGAAAGAAGAC	TCA	450–480	457	8	4	3	4	6	4	7	5	6	3	7	4.9	0	0.05	0.02	0.03	55	35	2	two orthologues?
SSR-40	F: GCGTGGCTTGATGAACTTGG R: ATGCTAGTTGAAGCCGTGCA				1																			sequences different to transcript
SSR-41	F: GTAAGACAAGTAGATTTGGTTCACTCT R: GCGGTGTCTCCTTTGTTATGTT		375–380	379	5	1	1	1	1	1	0	0	0	1	1	0.6	0	0.05	0.03	0.04	23	6	1	
SSR-42	F: ACGTAACTCAAATAACGATGCAAGT R: AGCTCATTTCCCAGTCATTTAGC	GAT			7																			three orthologues sequenced? sequence different to transcript

Appendix 5. Continued.

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles											Pairwise genetic distance				No.			Notes
			Range	Mean		A5	B5	C4	D4	E4	F4	G4	H4	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-43	F: ACCATCAAACCCTTCCAAGCT R: TTTGGGATTGGCGCCTCTAC	ATG	192–416	382	8	1	1	1	1	2	3	2	2	1	3	1.6							1	two orthologues? second sequence different to transcript
SSR-44	F: GTTATAAGCATCACCAGCGTGG R: AGGTAGGAGCATGCTCGTTG	ATC (TCG, CACC)	283–310	297	8	2	2	4	2	4	5	3	3	2	5	3.1	0	0.06	0.03	0.03	31	21	0	two or three orthologues?
SSR-45	F: GTTGGTGTTGAAGATGGACATGA R: ACAATTGTTCCATCAGGTTGTGAA		147–632	316	8																			two orthologues? second sequence different to transcript
SSR-46	F: TCGCTGTAATGCCAAGAGCC R: GCGTTGGTCCAAGAAAGCAA				3																			sequences different to transcript
SSR-47	F: CAGGACCAGATGGCTGACAA R: ACCACTTGTCATTAAACAAACCCT	TGAGAT (GGAATT, TGT)	264–288	272	8	1	2	2	5	4	14	10	13	1	14	6.4	0	0.04	0.01	0.02	19	15	0	one unambiguous locus?
SSR-48	F: CTCTTCACTTCATGAAATGTATCGAGA R: CAATCTCTTGCCGCTTTATATCAGA				0																			failed
Totals			112–692	316.8	6.7	3.7	4.1	3.5	3.1	4.1	4.8	4.4	5.1	1	20	4	0	0.51	0.04	0.04	54.7	41.7	0–3	

Note: PICs = parsimony informative characters; SNPs = single-nucleotide polymorphisms; SSRs = simple sequence repeats.

APPENDIX 6. Validation of 48 SSRs on a subset of the 48 New Zealand and Australian individuals of *Veronica* sequenced. Shown are six individuals of the *V. thomsonii* subset (C5, D5, E5, F5, G5, and H5; see Appendix 3).

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles									Pairwise genetic distance				No.			Notes
			Range	Mean		C5	D5	E5	F5	G5	H5	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-01	F: TGGAACAGCCATTGCATCAAA R: TCGTCGACTTACCAGTTCCAG	ACA (ATG)	310–694	342	6	4	4	4	3	3	2	2	4	3.3	0	0.06	0.01	0.01	29	12	2	two large introns, motif in central exon, sequences partially not covering complete locus
SSR-02	F: GATTGTTTCAGCCAAGAGATTCTCA R: CTTGTTCCGACGCAGACCAT	GAT			5																?	incomplete; several genes amplified by primers; same locus as SSR-42
SSR-03	F: TTGAGACGCAAGATTTCTGCAA R: CCCTCACGCGCTCTATCATT	ACT			5																at least 1	several genes amplified by primers
SSR-04	F: TTGTTCAACCAGTCGGACGT R: CCGCTTCGAGGACTTGCTAG	GAT	127–242	214	6	2	3	9	6	6	2	2	9	4.7	0	0.05	0.01	0.01	19	9	0	
SSR-05	F: GTCGAAATCGGATTTACTAGCTAAGT R: AGTCGGGAAAGAGATTGGGC	CATA	293–314	298	6	6	6	3	1	4	2	1	6	3.7	0	0.08	0.04	0.05	39	28	0	
SSR-06	F: AATAAACTGACGACAGCGCG R: ACTGTGAGTCTGCCTTACGC	TGA			5								0								?	not transcript sequence amplified
SSR-07	F: AGCAGTGAGAGCCAACATCC R: CGAAACGCCCTCTTACACGA	TAC	388–433	386	6	9	12	10	9	10	10	9	12	10	0	0.35	0.18	0.14	179	163	1	three orthologues?
SSR-08	F: CCATCAAACCCTTCCAAGCTG R: TGGCCTCTTACTTCCTACGTG	GAT	342–418	357	6	8	6	8	4	4	6	4	8	6	0	0.17	0.08	0.08	146	87	0	
SSR-09	F: TGGTCACTCTTTCGTGTTGGA R: CCATAAATTTGTGCTGCCTCCA		310–798	393	5	1	1	1	2	1	0	1	2	1							2?	sequences not covering locus
SSR-10	F: CGTAAATTGGATCAGGTCGCC R: CGTAGCTAGTTTGTCATTGGATGG	AGT	145–275	258	6	2	4	2	1	4	1	1	4	2.3	0	0.36	0.09	0.03	105	24	1?	two misamplifications
SSR-11	F: AAACGACGTCGGACTGAGAC R: GGGATAACATTGCTCACTCACC	ACGA (TTG, ATT, AG)	265–294	281	6	7	7	12	5	11	5	5	12	7.8	0	0.08	0.03	0.03	36	33	0	two orthologues?
SSR-12	F: TTGCAGTCGGCTTTAAAGATCC R: ATACCAGCCATATCAGAGCGC	AATC	199–263	229	6	5	8	9	6	5	6	5	9	6.5	0	0.07	0.02	0.02	39	21	0	two orthologues?
SSR-13	F: TCCTTCCTACTTGCCAAACTCT R: TCACGCACAGAGGACTGAAC				5																?	sequences not covering locus
SSR-14	F: TGTTGACTCAATCCGTCTCCG R: TCTGCTTTGCTACCTGTCTTCT	TTAA	286–311	293	6	2	2	1	2	1	4	1	4	2	0	0.28	0.08	0.01	84	74	0	two orthologues? maybe two misamplifications
SSR-15	F: GGCAGAAGAAACGGTTGCAG R: GACCTTTATGCCGTCTGCCT	GAT	312–1051	681	2	0	0	0	1	0	1	1	1	0.3							2	one misamplification
SSR-16	F: GAGACAACTGCTGCACTTGC R: TTAGTCCACCAGTGTCCACG	ATC	310–627	466	6	1	3	4	2	2	3	1	4	2.5	0	0.08	0.1	0.02	75	9	1	sequences not covering locus
SSR-17	F: AACTTGCTCGTCTCCACCAG R: CCGATGGATTCAGAAACCAACAA	GAT	222–252	239	6	9	6	6	4	7	10	4	10	7	0	0.51	0.09	0.06	173	52	0	two orthologues? three misamplifications

Appendix 6. Continued.

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles									Pairwise genetic distance				No.			Notes
			Range	Mean		C5	D5	E5	F5	G5	H5	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-18	F: TCTGTGCTACAACTAGTACAAGGAG R: GGATGGATCCCTTTCTTGAAATAAGG				4																?	sequences not covering locus
SSR-19	F: TGGCAACATGCAACTGTGTT R: ACGAGAATACCATACTTCATGTTCG	TATC (ATA, TAC)	260–295	274	6	4	3	5	4	4	3	3	5	3.8	0	0.08	0.02	0.01	31	24	0	two orthologues?
SSR-20	F: CATTCGTATTACTGTAAATGGTTTGCC R: GCAAACAGCACAAATATTTCACCA	GTTA (ACA, GTGA)	190–253	232	6	5	4	10	5	4	6	4	10	5.7	0	0.09	0.04	0.04	26	24	0	two orthologues?
SSR-21	F: ATGGATGAAGGGCCAGTTAAGG R: CCGCCAACTCCTCATCTAATTCA	GAT	238–256	245	6	5	4	5	6	4	5	4	6	4.8	0	0.12	0.05	0.02	39	32	0	two orthologues?
SSR-22	F: AGGGTCGTTATGGAAACCGG R: GACATCACCAGTCATCCGCA	GAT	285–345	333	6	3	6	6	1	3	2	1	6	3.5	0	0.41	0.15	0.01	133	126	1	two orthologues?
SSR-23	F: CACAACCAAAGTAGCAGCACT R: TGTGAGTTCGCGTAAAGGGA				6																?	sequences not covering locus, two loci?
SSR-24	F: GATGCCATTGTTGGATGAATTTCG R: AGCTGCAACTCCTCCTTCAA				5																?	sequences not covering locus
SSR-25	F: GGTGGTAAAGGCACCGTTAGA R: CGACGAGCTCAGGTACGTC				0																	no sequences
SSR-26	F: GTGCGCGAACAAGTTTGGTT R: TCACTAATCCACCTGATCCGTC	ATC	302–306	302	6	3	4	4	2	2	2	2	4	2.8	0	0.2	0.12	0.17	95	81	0	three orthologues?
SSR-27	F: CGGAGAGGTGCAATATACAAATGT R: GGACAACGCATTAGGAAGTGG	ACT	259–263	259	6	2	2	3	1	1	1	1	3	1.7	0	0.04	0.01	0.01	14	4	0	one unambiguous locus
SSR-28	F: GCGAAATGCAACATTCCACTG R: GGAGACACGGAACCTGAACA	ACT	257–290	273	6	4	3	6	4	3	4	3	6	4	0	0.05	0.02	0.03	25	18	0	two orthologues?
SSR-29	F: GACACCAAACTTGTCTTCAACGT R: AAAGAGGTTGTGAATTCACTAGAAGTT	ACT	300–351	329	6	7	4	5	7	4	2	2	7	4.8	0	0.13	0.06	0.06	61	58	1	three orthologues?
SSR-30	F: TTCTTGCTCTTGTGTTGGTTCC R: TTCACTTCAAACCTTTGTCACTACC	TGA	275–285	279	6	1	2	2	3	5	2	1	5	2.5	0	0.04	0.01	0.01	13	5	0	one unambiguous locus
SSR-31	F: CGATGACGATGAGGACGACG R: CATTTGATGCACCTCCATGCT				0																	no sequences
SSR-32	F: GTGCCTAGATATCACCAAGATAGAAGA R: GACCAGAAGATCAGACTCAGCA	GAT	236–251	241	6	3	6	9	5	6	3	3	9	5.3	0	0.07	0.03	0.03	26	23	0	two or three orthologues?
SSR-33	F: GCTGCACCTGGGATTCAAAG R: ACTGTGAGTCTGCCTTACGC				5																0	different orthologous locus sequenced?
SSR-34	F: ATTGCTCAACATGTTTGCCTCT R: TGTCACAGTTTGGCGATATTGG				4																	sequences different to transcript
SSR-35	F: TCGTCATCGCTGAAACCATCA R: ACACTTGATCTGCTTGTTGCC	ATC (CAA)	463–501	484	5	7	4	7	6	6	0	1	7	5	0	0.47	0.16	0.06	192	141	1	two orthologues?
SSR-36	F: AAACCCAATTCAAAGCAATGACAC R: ACCCTCATTTCTCCAAACCAACT	TCA	240–253	245	6	2	2	4	4	2	4	2	4	3	0	0.02	0.01	0.01	8	6	0	one unambiguous locus
SSR-37	F: AGTTGACGCCTTGTTTGGTTC R: CACGCAAACACCACATTCCC	GAT	184–280	275	6	10	26	16	14	10	2	2	26	13	0	0.08	0.03	0.03	32	29	0	two or three orthologues?
SSR-38	F: CCCTAAAGTTCAAGCATCTATACCAG R: TGCTGCAGCTTCAAATGTTTCA		310–568	536	6	10	8	6	8	6	8	6	10	7.7	0	0.23	0.08	0.06	171	98	2	two or three orthologues?
SSR-39	F: ACTTGCTGCAACTTGCTAAACA R: TGGATGACAATGAAAGAGAAAGAAGAC	TCA	406–480	456	6	6	4	3	5	4	5	3	6	4.5	0	0.06	0.03	0.03	52	36	2	two orthologues?
SSR-40	F: GCGTGGCTTGATGAACTTGG R: ATGCTAGTTGAAGCCGTGCA				3																	sequences different to transcript, but consistent

Appendix 6. Continued.

SSR locus	Primer sequences (5′–3′)	SSR motif, Main (additional)	Sequence length (bp)		No. of individuals successfully sequenced	No. of alleles									Pairwise genetic distance				No.			Notes
			Range	Mean		C5	D5	E5	F5	G5	H5	Min	Max	Mean	Min	Max	Mean	Median	SNPs	PICs	Introns	
SSR-41	F: GTAAGACAAGTAGATTTGGTTCACTCT R: GCGGTGTCTCCTTTGTTATGTT		376–380	379	5	2	0	1	1	1	1	1	2	1	0	0.06	0.02	0.04	24	4	1	
SSR-42	F: ACGTAACTCAAATAACGATGCAAGT R: AGCTCATTTCCCAGTCATTTAGC	GAT	125–362	313	6	3	2	3	2	1	2	1	3	2.2	0	0.39	0.18	0.15	180	68	0	two orthologues?
SSR-43	F: ACCATCAAACCCTTCCAAGCT R: TTTGGGATTGGCGCCTCTAC	ATG	157–416	366	4	2	0	2	0	0	1	1	2	0.8							0	misamplifications
SSR-44	F: GTTATAAGCATCACCAGCGTGG R: AGGTAGGAGCATGCTCGTTG	ATC (TCG, CACC)	283–316	296	6	3	4	4	3	5	2	2	5	3.5	0	0.05	0.03	0.03	21	16	0	two orthologues?
SSR-45	F: GTTGGTGTTGAAGATGGACATGA R: ACAATTGTTCCATCAGGTTGTGAA		148–359	303	6	6	1	10	6	1	1	1	10	4.2	0	0.6	0.43	0.47	405	311	?	three orthologues? second sequence different to transcript
SSR-46	F: TCGCTGTAATGCCAAGAGCC R: GCGTTGGTCCAAGAAAGCAA				3																	sequences different to transcript
SSR-47	F: CAGGACCAGATGGCTGACAA R: ACCACTTGTCATTAAACAAACCCT	TGAGAT (GGAATT, TGT)	264–294	271	6	3	7	4	6	2	2	2	7	4	0	0.03	0.01	0.01	17	10	0	two orthologues?
SSR-48	F: CTCTTCACTTCATGAAATGTATCGAGA R: CAATCTCTTGCCGCTTTATATCAGA				1																?	
Totals			125–1051	327.3	5	4.3	4.6	5.4	4.1	3.9	3.2	1	26	4.3	0	0.6	0.1	0.1	80.3	52.5	0–2	

Note: PICs = parsimony informative characters; SNPs = single-nucleotide polymorphisms; SSRs = simple sequence repeats.