American Journal of Botany

American Journal of Botany: e369–e371. 2011.

*AJB* Primer Notes & Protocols in the Plant Sciences

# Rapid isolation and cross-amplification of microsatellite markers in *Plectritis congesta* (Valerianaceae) with 454 sequencing[1]

Jamie R. McEwen[2,3], Jana C. Vamosi[2], and Sean M. Rogers[2]

[2]Department of Biological Sciences, University of Calgary, 2500 University Dr. NW, Calgary, Alberta T2N 1N4, Canada

- *Premise of the study:* Microsatellite markers were isolated and characterized in *Plectritis congesta* for studying the evolution of this highly variable species.
- *Methods and Results:* We used 454 sequencing of DNA enriched for microsatellite repeats to develop microsatellite markers. This produced 262 079 reads with an average length of 324 bp, representing approximately 800 microsatellite regions from which 48 primers were tested. Eleven markers reliably amplified without optimization. These primer pairs showed a high degree of heterozygosity and allelic diversity. Unexpectedly, half of the markers contained multiple peaks, with up to four alleles per individual, which suggests that either polyploidy or isolated gene duplication has occurred within this clade. These primers successfully cross-amplified in *P. macrocera*, indicating the utility of these markers for the genus.
- *Conclusions:* With variation in mating system and habitat, a mix of duplicated and nonduplicated markers, and high genetic variance, *Plectritis* is an ideal candidate model genus for studying the ecological and evolutionary consequences of gene duplication.

**Key words:** 454 sequencing; diploidization; microsatellite; *Plectritis*; polyploidy; Valerianceae.

*Plectritis* (Lindl.) DC. is considered a challenging genus from a taxonomic and biogeographical perspective (Bell and Donoghue, 2005). *Plectritis* (now considered to be nested within the genus *Valeriana* L.) is a small clade of three to five species, with a surprisingly wide latitudinal extent, ranging from western North America to western South America (Bell and Donoghue, 2005). *Plectritis congesta* (Lindl.) DC. in particular is largely restricted to coastal bluffs and exhibits an extremely variable mixed mating system, potentially due to its restriction to small isolated island populations. Because of its unique geographic distribution, *Plectritis* is a model system to investigate the evolutionary dynamics associated with fragmentation, migration, and variable population sizes. Answering these questions is currently limited by a complete lack of DNA markers in this genus, or in the Valerianaceae.

High-throughput sequencing has revolutionized marker development, improving the reliability, cost, time requirements, and ease of obtaining sequencing information for marker development in organisms without existing genomic resources (Csencsics et al., 2010). Although the development of molecular markers has become easier and more affordable with high-throughput sequencing, cloning methods remain common, and the technology has only begun to be applied to a broad range of species (Csencsics et al., 2010). In this paper, we report the use of high-throughput sequencing combined with a bioinformatics workflow to generate a large list of independent microsatellite markers in a species without any prior genetic information or relation to a model species.

## METHODS AND RESULTS

DNA was extracted from a single *P. congesta* individual collected as seed from Vancouver Island, British Columbia (48.535°N, 123.397°W), using a standard cetyltrimethylammonium bromide (CTAB) extraction protocol. DNA for primer testing was obtained from individuals from the Gulf Islands of British Columba: Gabriola Island (49.154°N, 123.831°W), Galiano Island (48.905°N, 123.413°W), and Salt Spring Island (48.816°N, 123.473°W). DNA quality and quantity were determined with agarose gel electrophoresis and a Beckman Coulter DTX 880 Multimode Detector (Beckman Coulter, Brea, California, USA) spectrophotometer. Restriction-ligation and microsatellite enrichment were performed using methods developed by Glenn and Schable (2005). Briefly, DNA was digested with the restriction enzymes *Rsa*I and *Xmn*I (New England Biosystems, Ipswitch, Massachusetts, USA), followed by a ligation of an SNX linker. These linker-ligated fragments were then hybridized to a repetitive DNA biotinylated oligonucleotide mixture ($(AG)_{12}$, $(TG)_{12}$, $(AAC)_6$, $(AAG)_6$, $(AAT)_8$, $(ACT)_{12}$, $(ATC)_8$), and the fragments containing complementary repeats were purified by Dynabead enrichment (Dynal, Applied Biosystems, Carlsbad, California). The enriched microsatellite-containing fragments were recovered by PCR and then sent for sequencing at the Plant Biotechnology Institute (PBI) in Saskatoon, Saskatchewan, Canada.

The microsatellite-enriched DNA fragments were MID-tagged and sequenced on a one-quarter plate in parallel with four other species using a Roche 454 sequencing platform (Roche, Basel, Switzerland). The raw data were downloaded and converted to FASTA format using sff_extract (Blanca and Chevreux, 2010). The first 50 nucleotides were trimmed, and the remaining sequences containing the linkers were removed using the FASTX Toolkit (Gorden, 2010). This prevented the design of primers that included the linkers

[1]Manuscript received 4 April 2011; revision accepted 17 July 2011.

The authors thank N. Kane for sharing his knowledge of bioinformatics and generously providing Perl scripts for sequence processing and L. Adderley for collecting the plant material. This study was funded by National Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants to S.M.R. and J.C.V. and an NSERC-CANPOLIN grant to J.C.V. (This represents publication #21 of NSERC-CANPOLIN.)

[3]Author for correspondence: jrmcewen@ucalgary.ca

doi:10.3732/ajb.1100160

TABLE 1. Primers for amplifying microsatellite DNA in *Plectritis congesta*. For each locus, primer pairs, annealing temperature, repeat motif, size range, apparent ploidy, and GenBank accession numbers are reported.

| Locus | Primer sequence (5′–3′) | $T_a$ (°C) | Repeat motif | Size range (bp) | Apparent ploidy[a] | GenBank Accession No. |
|---|---|---|---|---|---|---|
| Pcon202 | F: GTTCTTCACTGGTTCGGTCG<br>R: GGAAGAGCCTAATGATGACGG | 55 | (AC)15 | 192–300 | Polyploid | JF714687 |
| Pcon276 | F: GGGCGATGATTGCCTAACAC<br>R: ACCAACGTAGGACAAGGACC | 55 | AC(8), AG(25), AG(13) | 194–300 | Polyploid | JF714688 |
| Pcon315 | F: ATGCTCGTCGGTTTGAACAC<br>R: CAGAATCACATCGGCTTCACC | 55 | ATC(12) | 289–315 | Diploid | JF714689 |
| Pcon349 | F: TCCCAAGCGACACAATCAAC<br>R: CCCTAGGTCTCAGTTGAACAAG | 55 | ATC(10) | 336–357 | Diploid | JF714690 |
| Pcon151 | F: TCAGTTGCCAGCAGAACAAG<br>R: ACATAACCGGGTCCTCTTCG | 55 | ATC(10) | 132–161 | Diploid | JF714691 |
| Pcon391 | F: CAGGACAAAGACGTGGTTGG<br>R: CAAGTCAACCGTCAACTGGG | 55 | ATC(13) | 339–409 | Polyploid | JF714692 |
| Pcon194a | F: AGATTAAGCTGTGGACTTCTGC<br>R: GCCTTGTTACTGGATCCTGTG | 55 | ATC(13) | 158–239 | Diploid | JF714693 |
| Pcon231b | F: GCGCGCCATCTTACTTACAG<br>R: TGGATTGAGGATGAACACAAGC | 55 | ATC(10) | 162–243 | Polyploid | JF714694 |
| Pcon222b | F: AACGCACCTCTGAAATCACC<br>R: CACTCATTGCCGGTCCAAAG | 55 | ATC(11) | 214–226 | Polyploid | JF714695 |
| Pcon228 | F: TCACGTGGTTGGTTGGTTTC<br>R: GATTTCCTTCGGATTGGCGG | 55 | AG(8) | 222–228 | Diploid | JF714696 |
| Pcon312b | F: CAATCTGAACCGCCGCTTAG<br>R: CGGAGCTTGAGTCAGATGCC | 55 | ATC(10) | 309–336 | Polyploid | JF714697 |

Note: $T_a$ = annealing temperature.
[a] Some markers amplified up to four alleles in a single individual and are reported here as polyploid markers.

and purged linker dimers from the data. Sequences less than 200 bp were removed from the resulting data, increasing the average read length of remaining sequences to 412 bp.

Sequences containing di-, tri-, tetra-, penta-, and hexanucleotide repeats were identified with MSATCOMMANDER (Faircloth, 2008), which finds microsatellite DNA and uses Primer3 (Rozen and Skaletsky, 2000) to design primers. We used the default values for guanine-cytosine (GC) content and annealing temperature ($T_a$) (GC content 30–70%, primer optimal $T_a$ of 60°C). These sequences were extracted from the processed FASTA file and assembled with the GS de novo Assembler (Roche) to ensure that designed primers would not amplify the same region. Primers were designed from the resulting assembled contigs using MSATCOMMANDER (Rozen and Skaletsky, 2000; Faircloth, 2008) selecting for di-, tri-, tetra-, penta-, and hexanucleotide repeats with a product size of 150–600 bp. Forty-eight primers from 150 to 500 bp containing di-, tri-, tetra-, and pentanucleotide repeats were selected for primer testing.

The selected primers were first tested for amplification success. PCR was performed with 0.5 U of TopTaq *Taq* polymerase (QIAGEN, Valencia, California, USA) and 1× TopTaq PCR buffer (QIAGEN), using 1.5 mM $MgCl_2$, 1 mM dNTP, 1 μM primers, and 50 ng of DNA in a total reaction volume of 10 μL. Thermocycling was conducted on BioRad C1000 and S1000 thermocylers (BioRad, Hercules, California) according to the following cycle: 94°C for 4 min, 30 cycles of 94°C for 30 s, 55°C for 45 s, 72°C for 2 min, followed by a 72°C extension period for 5 min. Primers successfully amplifying polymorphic markers were resynthesized and 5′-labeled with 6-FAM dye (Applied Biosystems), and PCR was redone on 20 different *P. congesta* individuals and two *P. macrocera* Torr. & A. Gray individuals from three populations from the Gulf Islands of British Columbia, Canada, to characterize microsatellite markers and test for cross-amplification. Labeled PCR product was run on an Applied Biosystems 3730 DNA Analyzer using the LIZ size standard (Applied Biosystems), and allele range was assessed with GeneMapper version 3.2 (Applied Biosystems).

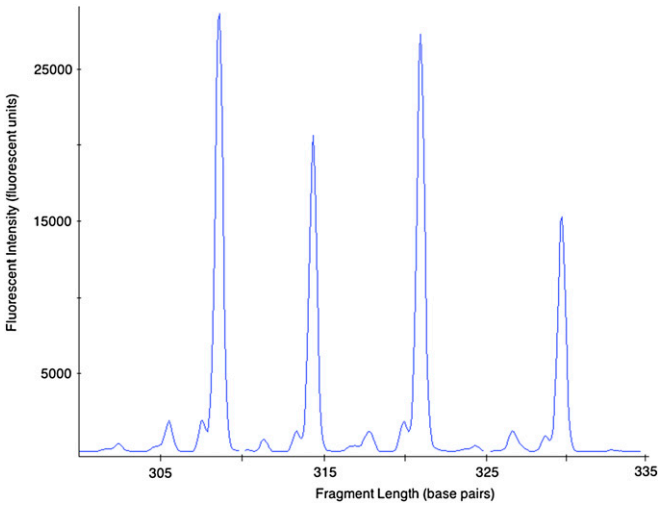


Fig. 1. Electropherogram showing four alleles from an individual using the marker Pcon312b.

TABLE 2. Observed heterozygosity and allelic diversity for the microsatellite markers developed for *Plectritis congesta.*

| Locus | $H_o$ | No. of alleles |
|---|---|---|
| Pcon202 | 0.769 | 6 |
| Pcon276 | 0.761 | 6 |
| Pcon315 | 0.893 | 11 |
| Pcon349 | 0.718 | 9 |
| Pcon151 | 0.816 | 7 |
| Pcon391 | 0.855 | 6 |
| Pcon194a | 0.933 | 7 |
| Pcon231b | —[a] | 9 |
| Pcon222b | 0.801 | 8 |
| Pcon228 | 0.781 | 8 |
| Pcon312b | 0.858 | 8 |

*Note*: $H_o$ = observed heterozygosity.
[a] AUTOTET was not able to process the observed heterozygosity for Pcon231b, possibly due to the nature of variation within that locus.

The 454 sequences had an average length of 324 bp. Post-processing of the sequence data resulted in the loss of 38 137 of 262 079 (15%) of the original reads. Isolating reads with microsatellite DNA resulted in 8872 sequences, which were assembled into 547 contigs with an average of 312 bp length and 11.3 reads. There were 288 nonredundant candidate primers developed from the assembled contigs. Of the 48 primers that were initially tested, 11 reliably amplified, were polymorphic, and did not require optimization. Some markers were apparently polyploid, having up to four alleles represented from a single individual (Table 1, Fig. 1). Cross-amplification in *P. macrocera* was successful for all of the microsatellite primers. Allele range, observed heterozygosity, and the number of alleles for polyploid loci were assessed with AUTOTET (Thrall and Young, 2000). Allele range, observed heterozygosity, and the number of alleles were assessed for the diploid markers using GENEPOP (Raymond and Rousset, 1995; Rousset, 2008). All markers were polymorphic, with the number of alleles ranging from six to 11 and the observed heterozygosity ranging from 0.718 to 0.933 between all three *P. congesta* populations (Table 2). The size range of the markers ranged from 132 to 409 bp (Table 1).

## CONCLUSIONS

Cloning techniques are still commonly used in marker development, and are comparatively costly and difficult to develop primers from. Using 454 sequencing on enriched microsatellite DNA, we were able to generate 228 independent microsatellite markers. Selecting 48 of these markers yielded 11 polymorphic microsatellite loci that had good amplification profiles and required no optimization beyond initial primer screening. High-throughput sequencing has a number of advantages over cloning techniques for microsatellite development, and we recommend the continued use of 454 sequencing for marker development.

Purging the linkers from the data resulted in a loss of 684 sequences (38 137 after removing trimmed sequences less than 200 bp) and increased the number of steps required to prepare the data for primer development. The success of 454 sequencing for identifying microsatellite sites for primer development on a relatively inexpensive platform in our study and others demonstrates the utility of this platform for marker development (Csencsics et al., 2010). However, high-throughput sequencing may not always be the best option for marker development, particularly for laboratories that are already set up for inexpensive marker development via cloning techniques. Additionally, the reliability and quality of markers generated by 454 sequencing has not been assessed in relation to markers developed by cloning techniques. The decision to use 454 sequencing for marker development will therefore be dependent on the resources available to the laboratory and the species being investigated.

Approximately half of the microsatellite markers we developed appeared to be duplicated. While this is consistent with previous studies that have suggested a polyploid history in *Plectritis* (i.e., chromosome counts from Engel [1976]), it is possible that only a few loci have been duplicated in this species. The availability of both diploid and duplicated markers and the cross-amplification of these markers within the genus provides a toolset for further study of gene or genome duplication within this clade, as well as many other ecological questions in this highly variable species.

## LITERATURE CITED


BELL, C. D., AND M. J. DONOGHUE. 2005. Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. *Organisms, Diversity & Evolution* 5: 147–159.

BLANCA, J., AND B. CHEVREUX. 2010. sff_extract. COMAV Institute, Universidad Politécnica, Valencia, Spain.

CSENCSICS, D., S. BRODBECK, AND R. HOLDEREGGER. 2010. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *Journal of Heredity* 101: 789–793.

ENGEL, K. 1976. Beirtrage zur systematik der Valerianaceae unter besonderes Berücksichtigung cytosystematischer ergebnisee. Dissertation, Justus Liebig-Universität Giessen, Giessen, Germany.

FAIRCLOTH, B. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8: 92–94.

GLENN, T. C., AND N. A. SCHABLE. 2005. Isolating microsatellite DNA loci. *Methods in Enzymology* 395: 202–222.

GORDEN, A. 2010. FASTX Toolkit. Howard Hughes Medical Institute, Cold Spring Harbour, New York, USA.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

ROUSSET, F. 2008. GENEPOP'007: A complete reimplementation of the GENEPOP software for Windows and Linux. *Moleculary Ecology Resources* 8: 103–106.

ROZEN, S., AND H. J. SKALETSKY. 2000. Primer3 on the WWW for general users and for biologist programmers. Humana Press, Totowa, New Jersey, USA.

THRALL, P. H., AND A. YOUNG. 2000. AUTOTET: A program for analysis of autotetraploid genotypic data. *Journal of Heredity* 91: 348–349.