Developing molecular tools and insights into the *Penstemon* genome using genomic reduction and next-generation sequencing

Dockter *et al.*
Developing molecular tools and insights into the *Penstemon* genome using genomic reduction and next-generation sequencing

Rhyan B Dockter¹, David B Elzinga¹, Brad Geary¹, P Jeff Maughan¹, Leigh A Johnson², Danika Tumbleson¹, JanaLynn Franke¹, Keri Dockter¹ and Mikel R Stevens¹*

**Abstract**

**Background:** *Penstemon*’s unique phenotypic diversity, hardiness, and drought-tolerance give it great potential for the xeric landscaping industry. Molecular markers will accelerate the breeding and domestication of drought tolerant *Penstemon* cultivars by, creating genetic maps, and clarifying of phylogenetic relationships. Our objectives were to identify and validate interspecific molecular markers from four diverse *Penstemon* species in order to gain specific insights into the *Penstemon* genome.

**Results:** We used a 454 pyrosequencing and GR-RSC (genome reduction using restriction site conservation) to identify homologous loci across four *Penstemon* species (*P. cyananthus*, *P. davidsonii*, *P. dissectus*, and *P. fruticosus*) representing three diverse subgenera with considerable genome size variation. From these genomic data, we identified 133 unique interspecific markers containing SSRs and INDELs of which 51 produced viable PCR-based markers. These markers produced simple banding patterns in 90% of the species × marker interactions (~84% were polymorphic). Twelve of the markers were tested across 93, mostly xeric, *Penstemon* taxa (72 species), of which ~98% produced reproducible marker data. Additionally, we identified an average of one SNP per 2,890 bp per species and one per 97 bp between any two apparent homologous sequences from the four source species. We selected 192 homologous sequences, meeting stringent parameters, to create SNP markers. Of these, 75 demonstrated repeatable polymorphic marker functionality across the four sequence source species. Finally, sequence analysis indicated that repetitive elements were approximately 70% more prevalent in the *P. cyananthus* genome, the largest genome in the study, than in the smallest genome surveyed (*P. dissectus*).

**Conclusions:** We demonstrated the utility of GR-RSC to identify homologous loci across related *Penstemon* taxa. Though PCR primer regions were conserved across a broadly sampled survey of *Penstemon* species (93 taxa), DNA sequence within these amplicons (12 SSR/INDEL markers) was highly diverse. With the continued decline in next-generation sequencing costs, it will soon be feasible to use genomic reduction techniques to simultaneously sequence thousands of homologous loci across dozens of *Penstemon* species. Such efforts will greatly facilitate our understanding of the phylogenetic structure within this important drought tolerant genus. In the interim, this study identified thousands of SNPs and over 50 SSRs/INDELs which should provide a foundation for future *Penstemon* phylogenetic studies and breeding efforts.

**Keywords:** Breeding domesticated *Penstemon*, Genome reduction, Homologous sequences, LTR retroelements, Plantaginaceae, Pyrosequencing, Repetitive elements

* Correspondence: mikel_stevens@byu.edu

1Plant and Wildlife Sciences Department, Brigham Young University, Provo, UT 84602, USA

Full list of author information is available at the end of the article

© 2013 Dockter et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Background
Interest is increasing in drought tolerant landscape plants due to water shortages experienced by many municipalities, especially in the Southwestern US [1,2]. However, the increased use of drought tolerant species also carries concerns regarding the introduction of non-native and potentially invasive species [3,4]. One way to address both issues is to landscape with native xeric flora [3]. *Penstemon* Mitchell (Plantaginaceae) has excellent potential for xeric landscapes and some *Penstemon* cultivars, adapted to mild climates, are already used throughout Europe as landscape plants [5-10]. Despite its potential, few *Penstemon* cultivars are used in xeric landscapes and there has been little to no drought or cold tolerant breeding programs for such landscapes [6-8,10-12]. *Penstemon*, with over 270 species, is one of the largest and most diverse plant genera of those that are strictly indigenous to North and Central America. This genus features a diverse in morphology, including a broad assortment of colors, flowers, and leaf structures. *Penstemon*’s putative center of origin is the arid Intermountain West of the United States [13,14] and has frequently been discussed as an untapped resource for xeric landscape cultivar development [5-7,9-11,15-17]. Because domestication and cultivar development, of any species, is slow, costly, and time consuming, few in the landscape industry have invested in native species breeding. However, given the recent and dramatic decrease in costs and relative ease of genotyping, we anticipate the wider utilization of marker assisted selection to accelerate breeding programs of native species, including drought tolerant *Penstemon* [18-20].

PCR-based markers are now essential tools to facilitate plant domestication, plant breeding, germplasm conservation, phylogenetics, and genetic mapping studies [19-22]. Not surprisingly, little molecular or traditional genetic work has been reported for *Penstemon* [23]. To achieve broad resolution of the genome with three of the most efficient markers, SSRs (simple sequence repeats or microsatellites), INDELS (insertions/deletions), and SNPs (single nucleotide polymorphisms), vast amounts of DNA sequence are needed, particularly for SNPs where sufficient read depth is needed to distinguish true polymorphisms from sequence noise [24-26]. With the development of next-generation sequencing (e.g., Roche 454-pyrosequencing) the cost of high-throughput marker discovery has been dramatically reduced [18].

Additionally, Maughan et al. [25] described a simple genome reduction method, known as GR-RSC (genome reduction using restriction site conservation), which reduces the genome by >90% thereby, making it feasible to redundantly sequence the remaining genome with next-generation sequencing technologies. This process is repeated across multiple cultivars or species, with comparisons identifying many inter- and intraspecific homologous loci. Genomic reduction techniques consistently identify homologous loci between related species [20,27], and GR-RSC has enabled the identification and development of interspecific homologous SNPs [20].

We utilized GR-RSC to identify homologous sequences in four diploid (2n = 2x = 16) *Penstemon* species chosen to represent a range of taxonomic and genome size diversity [5,14]. Included in our analysis are two closely related species from the subgenus *Dasanthera* (P. davidsonii Greene and P. fruticosus (Pursh) Greene var. fruticosus), one from the subgenus *Habroanthus* (P. cyananthus Hook. var. cyananthus), and one (P. dissectus Elliot) from the monophyletic subgenus *Dissecti*, which is phenotypically divergent from all other *Penstemon* species. This experimental design allowed us to make broad inter- and intra-subgenera comparisons in *Penstemon*. The objectives of our study were three-fold: First, identify homologous SSR and INDEL markers from the four diverse species and test their conservation across 93, mostly xerophilic, *Penstemon* taxa. Second, identify conserved homologous sequences for SNPs for use in future interspecific studies. Third, assess observed variation in the GR-RSC sequences to gain insights into the *Penstemon* genome and possible reasons for the large size variation previously identified among the diploid taxa [5].

Methods
Plant material and DNA extraction
DNA from *P. cyananthus*, *P. davidsonii*, *P. dissectus*, and *P. fruticosus* leaf tissue was extracted using the CTAB purification method [28] with modifications [29] for the GR-RSC technique. The source localities and identification of these plants have been reported previously [5]. A single sample from each species with the highest quality and DNA concentration, as determined using a ND-1000 spectrophotometer (NanoDrop Technologies Inc., Montchanin, DE), was selected to provide the 500 ng of DNA necessary for the genome reduction protocol.

For the molecular marker experiments, we used 93 *Penstemon* taxa. Leaf tissue was collected mostly from wild populations in the United States Intermountain West (Table 1). Each field-collected sample was identified to species and (or) variety using taxonomic keys specific to the area [30,31]. We extracted DNA using Qiagen DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA), and concentrations were diluted to 25–35 ng/μL.

Genome reduction, barcode addition and 454 pyrosequencing
Genome reduction followed Maughan et al. [25]. Briefly, for each sample, EcoRI and Bfai were used for the initial restriction digest, after which a biotin-labeled adapter was ligated to the EcoRI restriction site and a non-labeled adapter was ligated independently to the Bfai restriction
| Species                        | County | Marker sizes in bp |
|-------------------------------|--------|--------------------|
|                               |        | PS004  | PS011  | PS012  | PS014  | PS017  | PS032  | PS034  | PS035  | PS048  | PS052  | PS053  | PS075  |
| **Subgenus Dasanthera**       |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| *P. davidsonii*               | Purchased⁴ | 460    | 500    | 360    | 370    | 700    | 370    | 320, 950 | 520    | 440    | 220    | 320    | 140    |        |
| *P. fruticosus* v. *fruticosus* | Purchased | 460    | 500    | 360    | 410    | 700    | 340    | 360    | 520    | 420    | 220    | 320    | 140    |        |
| *P. montanus* v. *montanus*   | Utah   | 480    | 430    | 390    | 370    | 450    | 310    | 340, 310 | 470    | 430    | 200    | 390    | 115    |        |
| **Subgenus Dissecti**         |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| *P. dissectus*                | Purchased | 440    | 860    | 370    | 380    | 750    | 370    | 320    | 920    | 380    | 220    | 320    | 140    |        |
| **Subgenus Habroanthus**      |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| *P. ammophillus*              | Kane   | 480    | 800    | 400    | 430    | 470    | 300    | 1250, 340 | 470    | 420    | 230    | 360    | 125    |        |
| *P. barbatus* v. *torreyi*    | Garfield | 420    | 800    | 400    | 490    | 500    | 320    | 340    | 500    | 410    | 200    | 360    | 110    |        |
| *P. barbatus* v. *trichander* | San Juan | 650, 480 | 850    | 420    | 500, 490 | 520    | 310    | 310    | 500    | 450, 410 | 200    | 370    | 130    |        |
| *P. compactus*                | Garfield | 650, 480 | 850    | 420    | 490    | 470    | 330, 310 | 310    | 500    | 430    | 200    | 360    | 125    |        |
| *P. cyanoanthus* v. *cyananthus* | Wasatch | 420    | 860    | 400    | 410    | 750    | 370    | 310, 340 | 630    | 420    | 220    | 160    | 320    | 160    |
| *P. cyanoanthus* v. *subglaber* | Box Elder | 440, 420 | 850    | 400    | 490, 470 | 500, 450 | 340, 310 | 310, 280 | 520    | 410    | 210    | 390    | 360    | 125    |
| *P. cyanocaulis*              | Emery   | 440    | 310    | 420    | 490, 470 | 520    | 330, 320 | 320    | 480    | 420    | 210    | 350    | 120    |        |
| *P. eatonii* v. *eatonii*     | Utah    | 420    | 800    | 400    | 490    | 450    | 320    | 300    | 500    | NM⁴    | 210    | 350    | 135, 125 |        |
| *P. eatonii* v. *undosus*     | Washington | 420    | 850    | 420    | 470    | 420    | 320    | 290    | 650, 500 | 410    | 210    | 340    | 125    |        |
| *P. fremontii*                | Uintah  | 480    | 850    | 400    | 430    | 490    | 320, 310 | 340    | 500    | 420    | 220    | 370    | 130    |        |
| *P. gibbensii*                | Daggert | 480, 440 | 850    | 420    | 490    | 420    | 320    | 300    | 480    | 430    | 220    | 360    | 130    |        |
| *P. idahoensis*               | Box Elder | 440    | 800    | 400    | 410    | 470    | 310    | 340    | 500    | 430    | 250    | 340    | 130    |        |
| *P. laevis*                   | Kane    | 440    | 850    | 400    | 470    | 470    | 310    | 350, 320 | 500    | 420    | 220    | 360    | 125    |        |
| *P. leiophyllus* v. *leiophyllus* | Iron   | 480    | 850, 490 | 420    | 430    | 450    | 310    | 340    | 480    | 430    | 220    | 350    | 120    |        |
| *P. longiflorus*              | Beaver  | 440    | 800    | 420, 400 | 470 | 470, 450 | 330, 310 | 310    | 500    | 450    | 230    | 350, 220 | 125    |        |
| *P. navajop*                  | San Juan | 480    | 800    | 400    | 490    | 550    | 330, 300 | 360, 340 | 500    | 450    | 230    | 410    | 135, 130 |        |
| *P. parvus*                   | Garfield | 480    | 850    | 450    | 500, 490 | 490    | 320    | 300    | 500    | 430    | 210    | 380    | 360    | 130    |
| *P. pseudoputus*              | Garfield | 480    | 800    | 420, 400 | 430 | 490, 420 | 320    | 340    | 480    | 450    | 230, 220 | 350    | 130    |        |
| *P. scaricosus* v. *albidifluis* | Uintah | 440    | 850    | 400    | 490    | 490    | 310    | 320    | 480    | 410    | 210    | 370    | 115    |        |
| *P. scaricosus* v. *cyanomontanus* | Uintah | 440    | 850    | 400    | 490    | 490    | 330    | 310    | 500    | 420    | 210    | 360    | 115    |        |
| *P. scaricosus* v. *garrettii* | Duchesne | 490, 480 | 850    | 420    | 430    | 490    | 320    | 420, 340 | 520    | 430    | 230    | 360    | 125    |        |
| *P. scaricosus* v. *scariosus* | Sevier  | 480    | 1500, 1300 | 400    | 470 | 520    | 340, 310 | 310    | 500    | 430    | 210    | 360    | 130    |        |
| *P. speciosus*                | Box Elder | 440    | 800    | 420    | 500, 490 | 490    | 320    | 340, 310 | 520    | 310    | 210    | 360    | 120    |        |
Table 1 Penstemon taxa (with collection counties) utilized in the 12 marker analysis with respective marker sizes (Continued)

| Subgenus Penstemon | P. abietinus | Sevier | 440 | AD⁴ | 390 | 400 | 520 | 320 | 340 | 500 | 430 | 230 | 350 | 125 |
|---------------------|--------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| P. aculeiflorus      | Daggett      | 420    | 400 | 430 | 470 | 320 | 350 | 480 | 420 | 220 | 350 | 120 |
| P. ambiguus v. laevismus | Washington | 520 | 850 | 390 | 490 | 470 | 320 | 1250, 340 | 500 | 400 | 220 | AD | 120 |
| P. angustifolius v. dulcis | Millard | 440 | 850, 600 | 400 | 490 | 520 | 370, 150 | 310 | 520 | 420 | 220 | 360 | 125 |
| P. angustifolius v. venosus | San Juan | 480 | 310 | 390 | 470 | 470 | 320, 150 | 340 | 550 | 450 | 220 | 360 | 135 |
| P. angustifolius v. vernalensis | Daggett | 480 | 800 | 390 | 430 | 470 | 370, 150 | 350 | 500 | 420 | 220 | 380 | 125 |
| P. atwoodii          | Kane         | 440 | 800 | 420 | 490, 390 | 470 | 300 | 310 | 480 | 400 | 180 | 360 | 280 | 115 |
| P. balscatus         | Garfield     | 440 | 850 | 400 | 500 | AD | 330, 310 | 320 | 520 | 420 | 230 | 380 | 125 |
| P. breviculus        | San Juan     | 650, 480 | 190 | 400 | AD | 470 | 500, 220 | 320 | 480 | 430 | 210 | 350 | 125 |
| P. caespitosus v. caespitosus | Uintah | 440 | 850 | 390 | 390 | 490 | 320 | NM | 190 | NM | 210 | 390, 370 | 115 |
| P. caespitosus v. desertipicta | Washington | 440 | 230 | 390 | 470, 370 | 470, 360 | 330 | 350 | 1000, 300 | 430 | 210 | 400, 380 | 130 |
| P. caespitosus v. petbrevis | Wasatch | 420 | 490 | 390 | 430, 400 | 470 | 320 | 350 | 520 | 380 | 220 | 340 | 120 |
| P. carnosus          | Emery        | 440 | 850 | 420 | 490 | 490 | 330, 300 | 310 | 500 | 430 | 220 | 350 | 130, 120 |
| P. concinnus         | Beaver       | 440 | 800 | 420 | 430, 400 | 500 | 480 | 350 | 480 | 420 | 190 | 700, 360 | 120 |
| P. confusus          | Washington   | 480 | 850 | 420 | 490 | 490 | 520 | 300 | 320 | 480 | 450 | 220 | 350 | 125 |
| P. crandallii v. atratus | San Juan | 420 | 490 | 390 | 500 | 450 | 370 | 320 | 280 | 400 | 190 | 350 | 120 |
| P. crandallii v. crandallii | San Juan | 420 | 340, 190 | 390 | 500 | 450 | 370, 340 | 310 | 280 | 380 | 190 | 350 | 115 |
| P. devius v. pedicellatus | Teton | 420 | 850 | 420 | 430 | 550 | 340 | 320 | 550 | 340 | 230 | 370 | 130 |
| P. dolius v. dolius | Millard      | NM | 710 | 400 | 400 | 490, 320 | 530, 300 | 320 | 480 | 450, 420 | 180 | 340 | 105 |
| P. dolius v. duchesnensis | Duchesne | 420 | AD | 420 | 400 | 500 | 340 | 320 | 480 | 410 | 180 | 360, 340 | 140, 120 |
| P. emantherus v. cleburnii | Daggett | 420 | 850 | 420 | 410 | 450 | 480 | 300 | 500 | 490, 420 | 190 | 390, 360 | 140, 130 |
| P. florensis         | Uintah       | 480 | AD | 420, 400 | 490, 430 | 470 | 300 | 350 | 520 | 420 | 220 | 360 | 125 |
| P. franklinii        | Iron         | 480 | 800 | 400 | 430 | 470 | 320, 300 | 350 | 520 | 420 | 240 | 380 | 125 |
| P. goodrichii        | Uintah      | 420 | 650 | 390 | 400 | 490 | 480 | 310 | 480 | 400 | 200 | 370, 350 | 135 |
| P. graharnii         | Uintah      | 420 | 850 | 400 | 390 | 470 | 530, 320 | 350 | 500 | 420 | 230 | 500, 370 | 120 |
| P. humilis v. brevifolius | Cache | 390 | 850 | 370 | 500 | 450 | 340, 320 | 350 | 480 | 500 | 220 | 280 | 115 |
| Taxon  | County  | Marker 1 | Marker 2 | Marker 3 | Marker 4 | Marker 5 | Marker 6 | Marker 7 | Marker 8 | Marker 9 | Marker 10 | Marker 11 | Marker 12 | Marker 13 | Marker 14 | Marker 15 | Marker 16 | Marker 17 | Marker 18 |
|--------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| P. humilis v. humilis | Box Elder | 420 | 850 | 390 | 410 | 520 | 330, 310 | 360 | 500 | 500, 470 | 220 | 350 | 120 |
| P. humilis v. obtusifolius | Washington | 420 | 800 | 390 | 520, 490 | AD | 330 | 340 | 480 | 470 | 200 | 350 | 120 |
| P. immanifestus | Millard | 480 | 710 | 420 | 490 | 380 | 300 | 320 | 480 | 410, 380 | 220 | 400 | 360 | 120 |
| P. lentus v. albilitorus | San Juan | NM | 430 | 390 | 430 | 450 | 320 | 300 | 500 | 400 | 210 | 470 | 370 | 140 |
| P. lentus v. lentus | San Juan | 480 | 850 | 400 | 430 | 470 | 300 | 310 | 500 | 410 | 210 | 400 | 370 | 145 |
| P. linarioides v. sili | Washington | 420 | 850 | 370 | 490, 390 | 470 | 330, 310 | 350 | 470 | 400 | 210 | 370 | 125 |
| P. marcusii | Emery | 390 | 800 | 450 | 370 | 490 | 310 | 340, 320 | 500 | NM | 200 | 390 | 360 | 120 |
| P. moffatii | Grand | 390 | 800 | 420 | 390 | 490 | 300 | 330 | 480 | 430 | 430 | 290, 200 | 380, 350 | 140 |
| P. nanus | Millard | 480 | 800 | 420 | 390 | 500 | 410, 380 | 220 | 340 | 500 | 410 | 320 | 470 | 380 | 120 |
| P. ophianthus | Sevier | 520 | 850 | 420 | 370 | 900, 750 | 330, 310 | 310 | 480 | 420 | 190 | AD | 115 |
| P. pachyphyllus v. congestus | Kane | 480 | 850 | 400 | 430 | 470, 380 | 320 | 310 | 520 | 410 | 250 | 370 | 170 |
| P. pachyphyllus v. mucronatus | Daggett | 440 | 800 | 390, 370 | 430 | 500 | 320 | 300, 280 | 520, 500 | 430 | 220 | 350 | 120 |
| P. pachyphyllus v. pachyphyllus | Duchesne | 480 | 850 | 390 | 410 | 490 | 370, 330 | 340, 240 | 400, 190 | 500, 430 | 290, 230 | 380, 220 | 125 |
| P. palmeri v. palmeri | Washington | 440 | 850 | 400 | 500, 490 | 520, 490 | 330 | 310 | 500 | 430 | 210 | 380 | 125 |
| P. petiolatus | Washington | 420 | 1000 | 400 | 500, 490 | 500 | 330 | 300 | 480 | 420 | 210 | 380 | 145 |
| P. pinorum | Washington | 480 | 800 | 420 | 610 | 500 | 480 | 310 | 480 | 470 | 200 | 390 | 125 |
| P. proceraus v. abemans | Garfield | 440 | 1000 | 850 | 450, 370 | 520 | 520 | 330 | 360 | 480 | 410 | 220 | 370 | 115 |
| P. proceraus v. proceraus | Iron | 420 | 850, 550 | 370 | 490 | 470 | 340, 310 | 360 | 470 | 470 | 220 | 340 | 120 |
| P. radicosus | Daggett | 420 | AD | 420 | 490 | 470 | 330, 310 | 310 | 500 | 450 | 200 | 360 | 125 |
| P. rydbergii v. aggregatus | Box Elder | 420 | 850 | 400 | 520 | 500 | 340 | 360 | 520 | 470 | 210 | 380 | 115 |
| P. rydbergii v. rydbergii | Rich | 420 | 710 | 400 | 520 | 500 | 370 | 320 | 500 | 470, 140 | AD | 390 | 115 |
| P. thompsoniae | Kane | 420 | AD | 370 | 500 | 450 | 340, 320 | 340 | 500 | 410 | 220 | 390, 370 | 130 |
| P. tusharenis | Beaver | 420 | 1300, 230 | 370 | 430 | 450 | 320 | 320 | 500, 300 | 410 | 230 | 340 | 120 |
| P. utahensis | San Juan | 480 | 410 | 420 | 430 | 490 | 300 | 310 | 500 | 410 | 220 | 370 | 125 |
| P. watsonii | Sevier | 420 | AD | 370 | 490 | 470 | 320 | 350 | 480 | 490 | 220 | 350 | 120 |
| P. whippleanus | Iron | 420 | 800 | 400 | 370 | 450 | 310 | 350 | 500 | 430 | 210 | 370 | 105 |
| P. yampaensis | Daggett | 570 | 710 | 400 | 430, 390 | 490 | 500, 320 | 310 | 480 | 410 | 260, 230 | 340 | 120 |

**Subgenus Saccanthera**

| Taxon  | County  | Marker 1 | Marker 2 | Marker 3 | Marker 4 | Marker 5 | Marker 6 | Marker 7 | Marker 8 | Marker 9 | Marker 10 | Marker 11 | Marker 12 | Marker 13 | Marker 14 | Marker 15 | Marker 16 | Marker 17 | Marker 18 |
|--------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| P. leonardi v. Higginsi | Washington | 390 | 1300 | 420, 400 | 490, 430 | 550, 520 | 320 | 310 | 480 | 470 | 250 | AD | 125 |
| P. leonardi v. leonardi | Utah | 440 | 800 | 420 | 430 | 490 | 320 | 340, 320 | 480 | 500 | 240 | 370 | 120 |
| P. leonardi v. patricus | Tooele | 440 | 850 | 370 | 470 | AD | 370 | 310 | 550, 520 | 470 | 230 | 380 | 115 |
| P. platyphyllus | Salt Lake | 420 | 800 | 400 | 430 | 470 | 330 | 310 | 520 | 430 | 240 | AD | 135 |
| P. rafniflorus | Washington | 420 | 1100, 430 | 400 | 410 | 420 | 320, 300 | 290 | 500 | 490, 430 | 470 | 370 | 120 |
| P. sepalus | Utah | 420 | 800 | 400 | 470 | 450 | 330 | 310 | 500 | 430 | 230 | 390 | 130 |
Table 1 *Penstemon* taxa (with collection counties) utilized in the 12 marker analysis with respective marker sizes (Continued)

|                      | 9  | 10 | 12 | 14 | 12 | 13 | 11 | 11 |
|----------------------|----|----|----|----|----|----|----|----|
| Total unique molecular weight bands | 6  | 7  |    |    |    |    |    |    |
| Total pairs of dual molecular weight bands | 7  | 7  | 16 | 11 | 28 | 14 | 7  | 8  |
| Total monomorphic markers | 85 | 80 | 86 | 76 | 80 | 65 | 78 | 86 |
| Total NM              | 2  | 0  | 0  | 0  | 0  | 1  | 0  | 4  |
| Total AD              | 0  | 6  | 0  | 1  | 2  | 0  | 0  | 0  |

1 All counties are in Utah except Teton, Co. which is in Wyoming.
2 Purchased = *P. davidsonii* and *P. fruticosus* were purchased from nurseries in Utah Co., Utah while *P. dissectus* was purchased from a nursery in Aiken Co., South Carolina.
3 NM = no marker.
4 AD = ambiguous data (usually multiple bands and or smearing).
site. Next, a non-labeled size exclusion step using Chroma Spin + TE-400 columns (Clontech Laboratories, Inc., Mountain View, CA) and magnetic biotin-streptavidin separation (Dynabeads M-280 Streptavidin, Invitrogen Life Science Corporation, Carlsbad, CA) was performed. Unique multiplex identifiers (MID) barcodes were added independently to each species using primers complementary to the adapter and cut sites (Table 2). Preliminary amplification was performed using 95°C for 1 min., 22 cycles of 95°C for 15 s, 65°C for 30 s, and 68°C for 2 min. PCR products were loaded into a 1.2% agarose Flashgel DNA Cassette (Lonza Corporation; Rockland, ME) to verify smearing and adequate amplification in preparation for pyrosequencing.

After the initial PCR, concentrations of each of the four species samples were determined fluorometrically using PicoGreen® dye (Invitrogen, Carlsbad, CA). Samples were then pooled using approximately equal molar concentrations of each species except for Penstemon cyananthus (genome size = 1C = 893 Mbp), where the molar concentration was doubled to maintain a similar genomic representation compared to the other three species with smaller genome sizes (P. dissectus, 1C = 462 Mbp; P. davidsonii, 1C = 483 Mbp; and P. fruticosus, 1C = 476 Mbp; [5]). DNA fragments between 500–600 bp were selected following Maughan et al. [25]. Sequencing was performed by the Brigham Young University DNA Sequencing Center (Provo, UT) using a half 454-pyrosequencing plate, Roche-454 GS GLX instrument, and Titanium reagents (Brandord, CT).

Sequence assembly

Sequence data were sorted by species using their unique MID species barcode (Table 2) by means of the software package CLC Bio Workbench (v. 2.6.1; Katrinebjerg, Aarhus N, Denmark). Following sorting (Table 2), assemblies were performed using Roche’s de novo assembler; Newbler (v. 2.6), which yields consensus sequences (contigs) of all individual reads, from each independent species, for use in subsequent analyses.

A full assembly (all individual reads of all four species pooled together) was performed by Newbler with “complex genome” parameter set and a trim file with MID barcodes specified; all other parameters were left to their defaults. For all subsequent species assemblies (all individual reads of one species), these same parameters were used with a few added conservative options selected: an expected depth of ‘10’ (20 default), a minimum overlap length of ‘50’ (40 default), and a minimum overlap identity of 95% (90% default).

Repeat element identification

Assembled sequences from all four species were masked for possible genome wide repetitive elements using a combination of RepeatModeler and RepeatMasker [32]. RepeatModeler is a de novo repeat element family identification and modeling algorithm that implements RECON [33] and RepeatScout [34]. RepeatModeler scanned all contigs from the four Penstemon species assemblies and produced a predicted repeat element library of predictive models to find repeat elements. Using this reference library, RepeatMasker then scanned the four species to filter out repetitive elements. Singletons were omitted from the analysis. To assess possible repetitive element biases with RepeatMasker when implementing a de novo library from RepeatModeler, we analyzed the GR-RSC data from Arabidopsis RIL’s (recombinant inbred lines) Ler-O and Col-4 from Maughan et al.’s. [35] study, compared to the Arabidopsis non-reduced genome downloaded from TAIR (The Arabidopsis Information Resource) [36].

Marker development, verification, and use

To identify SSRs, INDELs, and SNPs, we used software MISA and SNP_Finder_Plus (custom Perl-script), respectively [25,37,38]. RepeatMasker was used to identify and mask transposable elements. MISA parameters were set as follows: di-nucleotide motifs had a minimum of eight repeats, tri-nucleotide motifs had a minimum of six repeats, tetra-nucleotide motifs had a minimum of five repeats, and 100 bp was set as the interruption

---

**Table 2 The four multiplex identifiers (MID) barcodes (adapter) primers used for the genomes of Penstemon cyananthus, P. dissectus, P. davidsonii, and P. fruticosus**

| Species          | MID ID # | EcoR1 MID primer1 | Bfa1 MID primer2 |
|------------------|----------|-------------------|------------------|
| P. cyananthus    | MID 1    | 5'- ACGAGTGCGTGACTGGTACCAATTC | 5'- AGACGTGCGTGATGAGTCCTGAGTA |
| P. dissectus     | MID 2    | 5'- ACGCTCGACAGACTGGTACCAATTC | 5'- AGCTCAGACAGATGAGTCCTGAGTA |
| P. davidsonii    | MID 3    | 5'- AGACGCACCTCGACTGGTACCAATTC | 5'- AGACGCACGTGAGTCCTGAGTA |
| P. fruticosus    | MID 4    | 5'- AGACCTGTAGGAGCTGGTACCAATTC | 5'- AGACCTGTAGGAGCTGGTACCAATTC |

1 The "AATTC" at the 3′ end the primer was where adapters complement the enzyme EcoR1 cut site and the preceding "C" is where the base was changed to avoid further enzymatic cleavage of the fragment.

2 The "TA" at the 3′ end of the primer was where adapters complement the enzyme Bfa1 cut site and the preceding "G" is where the base was changed to avoid further enzymatic cleavage of the fragment.
(max difference between two purported SSR alleles). For the comparison of SSR frequency and repeat motifs across species, “unmasked” assembly files were used to remove bias caused by masking low complexity reads. The following parameters were used to define the heuristic thresholds for SNP_Finder_Plus: 8× minimum read depth for the SNP, 30% proportion of the reads representing the minor allele and 90% identity (an indication of homozygosity within a single species used in a dual-species assembly) required for each SNP locus. These parameters also helped compensate for sequencing and assembly errors, which allow greater confidence in calling base pair discrepancies as actual SNPs in the dual-species assemblies and the confident identification of heterozygosity in the individual assemblies. For both individual assemblies and dual species assemblies SNPs reported are those conforming to the aforementioned parameters.

All genomic sequences matching the above criteria were used for marker development. Primer3 v2.0 [39] was used to identify primers for amplifying these markers, with the following parameters: optimal primer size = 20 (range = 18–27); product size range = 100–500 base bp; Tm range = 50–60°C with 55°C optimum; and maximum polynucleotide = 3. Allowing PCR products greater than 200 bp greatly increased the possibility of INDELS in the PCR products.

The PCR (SSR/INDEL) markers were validated using the original four species as template DNA. Each 10 μl PCR reaction had ~ 30 ng genomic DNA, 0.05 mM dNTPs, 0.1 mM cresol red, 1.0 μl of 10X PCR buffer (Sigma-Aldrich, St. Louis, MO), 0.5 units of JumpStart Taq DNA Polymerase (Sigma-Aldrich, St. Louis, MO) and 0.5 μM (each) of the forward and reverse primers. The thermal cycler (Mastercycler® Pro; Eppendorf International; Hamburg, Germany) was set as follows: 94°C for 30 s, 45 cycles of 92°C for 20 s, (primer specific annealing temperature)/C for 1 min. 30 s, 72°C for 2 min., and 72°C for 7 min. (final extension). Following PCR reactions, DNA was loaded into 3% Metaphor® agarose (Lonza Corporation; Rockland, ME) and run using a gel electrophoresis box at 100 V for 2 h. Optimal annealing temperatures for each SSR/INDEL marker were selected based on clarity of bands produced over varying annealing temperatures. Only SSR/INDEL markers with one or two reproducible bands are reported in the marker studies (Tables 1 and 3). The same conditions used for marker validation were used in the SSR/INDEL marker studies, except gel electrophoresis times were increased to 4 h at 100 V.

The gels were evaluated and scored as: 1 = marker present; 0 = marker absent based upon molecular weight. The results were then analyzed to assess the strength of hierarchical signal in these data using 10,000 replications of fast bootstrapping as implemented in PAUP* v. 4.0b10 [40].

Our interspecific SNP genotyping was accomplished using Fluidigm (Fluidigm Corp., South San Francisco, CA) nanofluidic Dynamic Array Integrated Fluidic Circuit (IFC) Chips [40] on the EP-1TM System (Fluidigm Corp., South San Francisco, CA) and competitive allele-specific PCR KASPar chemistry (KBioscience Ltd., Hoddesdon, UK). A 5 μl sample mix, consisting of 2.25 μl genomic DNA (20 ng μl–1), 2.5 μl of 2x KBiosciences Allele Specific PCR (KASP) reagent Mix (KBioscience Ltd.), and 0.25 μl of 20x GT sample loading reagent (Fluidigm Corp., South San Francisco, CA) was prepared for each DNA sample. Similarly, a 4 μl 10x KASP Assay, containing 0.56 μl of the KASP assay primer mix (allele specific primers at 12 μM and the common reverse primer at 30 μM), 2 μl of 2x Assay Loading Reagent (Fluidigm Corp., South San Francisco, CA), and 1.44 μl DNase-free water was prepared for each SNP assay.

The two assay mixes were added to the dynamic array chip, mixed, and then thermal cycled using an integrated fluidic circuit Controller HX and FC1 thermal cycler (Fluidigm Corp., South San Francisco, CA). The thermal cycler was set as follows: 70°C for 30 min; 25°C for 10 min for thermo mixing of components followed by hot-start Taq polymerase activation at 94°C 15 min then a touchdown amplification protocol consisting of 10 cycles for 94°C for 20 sec, 65°C for 1 min (decreasing 0.8°C per cycle), 26 cycles of 94°C for 20 sec, 57°C for 1 min, and then hold at 20°C for 30 sec. Five end-point fluorescent images of the chip were acquired using the EP-1TM imager (Fluidigm Corp., South San Francisco, CA), once after the initial touchdown cycles were complete and then after each additional run on “additional touchdown cycles.” The extra cycles were run four times, with an analysis of the chip after each run.

The determination of each SNP allele was based on a minimum of at least two of three SNP genotyping experiments. The primers were then analyzed for functionality using the results from each of the five stops for each chip, which were compared to determine the most accurate call. Functionality was determined by number of calls versus no calls, and consistency.

Cross species sequencing verification
To evaluate the DNA sequence homology and polymorphism type (SSR or INDEL) at specific marker amplicons (Table 1) across the Penstemon genus, DNA samples from each of five species (P. cyananthus, P. davidsonii, P. dissectus, P. fruticosus, and P. pachyphyllus) were amplified and Sanger sequenced. We accomplished the PCR amplification using Qiagen HotStarTaq Plus Master Mix (Valencia, CA, USA) according to the manufacturer’s recommendations. The amplification protocol consisted
Table 3 Summary of marker characteristics including the primary SSR motif identified in the original GR-RSC (genome reduction using restriction site conservation) sequence, primer sequences, EFL (expected fragment length), total bands, and fragment sizes

| Marker name | Primary motif | Forward primer (5’-3’) | GenBank accession ID | EFL | Total unique bands | Fragment size |
|-------------|---------------|------------------------|----------------------|-----|-------------------|--------------|
| PS003       | (AT)_8        | TGCCCTGTCTTTACATCCCA   | JQ966997             | 217 | 3                 | 360          |
|             |               | CATGAACCTGTGAAATCCCA   |                      |     |                   |              |
| PS004       | (AT)_8        | GTTCCCAATGCTGCCACAT    | JQ951613             | 476 | 3                 | 420          |
|             |               | TTGTCTGTGAAACCGTTAGGT |                      |     |                   |              |
| PS005       | (GAA)_8      | GCCCAACTTCCGTAATGGAA   | JQ966998             | 303 | 3                 | 260          |
|             |               | AACTGCTGCGACTCGACTC   |                      |     |                   |              |
| PS009       | (GAA)_8      | ACCTCGAACACGCGTCTCC   | JQ966996             | 466 | 4                 | 370          |
|             |               | TCTGAGGGAAACCAAGGG    |                      |     |                   |              |
| PS011       | (GAA)_8      | AAGTGGCCACTGGATGCTCTT | JQ951614             | 435 | 2                 | 860          |
|             |               | GCAGCTGAGCTCCGAAAT    |                      |     |                   |              |
| PS012       | (TA)_8       | TCCATATTGAAACCAATGACTG| JQ951615             | 402 | 3                 | 400          |
|             |               | TGAATGGCAACCGTAAATCA  |                      |     |                   |              |
| PS013       | (TA)_8       | GAGAATGTATTTAAAACAGTGA | JQ967000             | 399 | 2                 | 400          |
|             |               | TCAGTACGTGAAACCTTTGCAATAA |             |     |                   |              |
| PS014       | (TGA)_8      | CGATTTTGTATGTTGATAGCGA| JQ951616             | 409 | 3                 | 410          |
|             |               | CCKTCATGCCCGGCGTATGT |                      |     |                   |              |
| PS015       | (TCG)_8     | GCCGAGCTGAGGAAAGCAA   | JQ967001             | 409 | 2                 | 490          |
|             |               | AATTCAAGCCTGCCACGC     |                      |     |                   |              |
| PS016       | (CT)_8      | CATGCGCCTTTCACACT      | JQ967002             | 447 | 3                 | NM²          |
|             |               | GAGCGGAGGCGTACGAGATCTA |                      |     | 1,100             | 1,060        |
| PS017       | (AG)_8       | GAAGGCTGATGACATATTTGC| JQ951617             | 455 | 2                 | 750          |
|             |               | ATTAGGCTCCACGAACAAA    |                      |     | 700               | 750          |
| PS019       | (AG)_8       | AATTCACGAGCCATACAA     | JQ967003             | 473 | 1                 | 380          |
|             |               | TGAATGGAGTCTTACACGGCTT |                      |     | 380               | 380          |
| PS021       | (CT)_8      | CCTAGCTTACGCGATACAGG   | JQ967004             | 386 | 3                 | 350          |
|             |               | AGATTTGCTGCAGTCAGTCACTTA |             |     | 450               | 450          |
| PS023       | (AG)_8      | GCTGGAGAATAACATGCGCG   | JQ967005             | 469 | 4                 | 310          |
|             |               | CCCTTCTGCAATCCCATACG   |                      |     | 480               | 120          |
| PS024       | (CTG)_8     | CCTCTTGCGCGTGCCTCT     | JQ967006             | 403 | 2                 | 430          |
|             |               | CCACCCCAACACAAACACAC   |                      |     | 430               | 400          |
| PS025       | (TC)_8      | GCACATGGAATGGAAGGATGC  | JQ967007             | 440 | 3                 | 440          |
|             |               | AGCATGCTGGAAGGAAACCA   |                      |     | 410               | 440          |
Table 3. Summary of marker characteristics including the primary SSR motif identified in the original GR-RSC (genome reduction using restriction site conservation) sequence, primer sequences, EFL (expected fragment length), total bands, and fragment sizes (Continued)

| Marker | Motif | Sequence | Accession | EFL (bp) | Bands | Fragment Sizes |
|--------|-------|----------|-----------|----------|--------|----------------|
| PS026  | (CTT)₆ | ACTTTAATGCTCCTCCTTGTGCA TCCGCAACCTGTTTATGTA | JQ967008 | 465      | 1      | 460 NM 460    |
| PS028  | (AC)₉ | GGGAGGCAGTAACAACAAAGA TACCTGTCGGCACACTGGATT | JQ967009 | 316      | 4      | 950, 400 460 320 400 460 |
| PS029  | (TA)₈ | ACCAAGTTGATGATTTGTTGG GGTGTTGGAATGAGCTTGGAGAAGA | JQ967010 | 440      | 3      | 840, 500 500 420 |
| PS032  | (GT)₄ | ACAAAAGCTCCTCAATCGCC GCATGTCGGACACACCT | JQ951618 | 328      | 2      | 370, 370 370 340 |
| PS034  | (CT)₉ | CAAACAAATCAAACAGCACT CATCGGAATCTAATGTCTCAAA | JQ951619 | 322      | 5      | 310, 340, 320, 950 320 360 |
| PS035  | (TC)₈ | TTTGACAGCTCTTTGGCAT ATCTGTTCAAGGCATGGAAT | JQ951620 | 486      | 3      | 630, 520 920 520 |
| PS036  | (TA)₉ | TTTCAATTTGTTGAGTTCAATCTC TCCGAGAATCTAATGCTCAAT | JQ967011 | 405      | 3      | 770, 770, 820 590 770 |
| PS038  | (TC)₉ | GTAATTACTTCGGCAGTTTTGTTAATT GGTCGGACCTAATACGGTTATAT | JQ967012 | 100      | 1      | NM 100 100 NM |
| PS040  | (CA)₉ | TAAAGAGGCTTTAAGGCGG ACCCTGAAGGCTTGGGAAGA | JQ957013 | 399      | 3      | 380, 390 410 390 |
| PS041  | (TA)₉ | TTTCCGCAAGAAGAGACAGCAT CTGTTGCAAGATTCCTTGCAT | JQ967014 | 249      | 3      | 270, 270 420 240 |
| PS045  | (CT)₈ | GCCACATACATGAAGACGTGAA CGAATCTCCTTGTCTTCTTCT | JQ967015 | 366      | 4      | 460, 440 120, 400 |
| PS047  | (AC)ₙ | ACACGACATCTTTCGCAAGA GCGTATTGGAGAGATTGGGA | JQ967016 | 428      | 3      | 470, 510 440 470 470 |
| PS048  | (CA)₉ | GCATTAGATCCGGAATATCTCACA TGCCGTAGTGGTGATTTCCTT | JQ951621 | 436      | 3      | 420, 440 380 420 |
| PS049  | (AG)₉ | CCCATCAATAAGAAGAAGAAGAAGA GGTGAAACCTGCTTCTAAACC | JQ967017 | 460      | 460    | 1,000 460 |
| PS050  | (AT)₉ | GTAACCTCTCGAACAAGTCTGAC TCGAGTACACATGCTTCATT | JQ967018 | 434      | 2      | 480, 460 480 460 |
| PS051  | (TG)₉ | GTAAACGACAATTATTACTCTTTCA CGAAGACTTTTCCCGAGAACC | JQ967019 | 352      | 1      | 280, 280 280 280 |
Table 3 Summary of marker characteristics including the primary SSR motif identified in the original GR-RSC (genome reduction using restriction site conservation) sequence, primer sequences, EFL (expected fragment length), total bands, and fragment sizes (Continued)

| Marker | Motif | Primer | EFL | Total Bands | Bands | Size |
|--------|-------|--------|-----|-------------|-------|------|
| PS052  | (c,da,di,f) | (AC)9 | CGCGGTCAATCTTGAAATCT JQ951622 | 206 | 1 | 220 220 220 220 |
| PS053  | (c,da) | (AC)8 | AATCATAGTCGAGCGGT JQ951623 | 410 | 3 | 160, 320 320 320, 450 450 |
| PS054  | (di) | (AC)8 | TCGCTAACGCTCTGGAGGC JQ967020 | 192 | 3 | 200 200 180 190 |
| PS055  | (di) | (AG)8 | TGGTACGCTCTGCATAAAC JQ967021 | 412 | 4 | 960 500 1,040 470 |
| PS056  | (c,da,f) | (GA)8 | TCGTTAAGCAATCTCGGAGC JQ967022 | 319 | 3 | 200 180 190 190 |
| PS057  | (di) | (AT)8 | TGCCTAATGGACCTGATCCT JQ967023 | 402 | 2 | 570 350 290 290 |
| PS058  | (di) | (AT)9 | TGGATGCCCTATGGGTACA JQ967024 | 484 | 2 | 560 470 470 470 |
| PS059  | (di) | (AT)8 | CATGAGAAGTAGATGACTGGGA JQ967025 | 320 | 2 | 350 350 350 350 |
| PS060  | (c,di) | (AT)8 | TCCATTCGCTGTCTCGCTTTCA JQ967026 | 484 | 2 | 450, 530 450, 530 450, 530 450 |
| PS061  | (c,di) | (AT)9 | CGACCAATCATCAACCAAACA JQ967027 | 453 | 3 | 480 480, 530 450, 530 450, 530 |
| PS062  | (c,di) | (AT)8 | TGAGAGGGTACGAGAATGTC GC JQ967028 | 320 | 2 | 350 350 350 350 |
| PS063  | (c,di) | (AT)9 | CGAGGCTGAGATGTTGACA JQ967029 | 437 | 4 | 490 500, 680 470 470 |
| PS064  | (c,di) | (AT)9 | ATGATGCTCCTAGTGTTGAC JQ967030 | 434 | 4 | 250 480 250, 480 250, 480 |
| PS065  | (c,di) | (AT)9 | TGGGATGAGCCTCGTGTTG JQ967031 | 463 | 4 | 500 500 500 480 |
| PS066  | (c,di) | (AT)9 | CATTGAGATCGAGTGTGTTGJQ967032 | 309 | 4 | 220 210 210 210 |
| PS071  | (c,di) | (AT)9 | AAGATGCGCTGCTGCGGCGTG JQ967033 | 446 | 1 | NM NM NM NM |
| Marker | Motif | Species | Primer Sequences | JQ Number | EFL | Total Bands | Band Sizes |
|--------|-------|---------|------------------|-----------|-----|-------------|------------|
| PS074  | (AAG)₆ | (c,da,di,f) | AGAAATCTCGCTTCCAGA CGCAACCTTAGGTCACGTTC | JQ967034 168 1 | 170 170 170 170 |
| PS075  | (TA)₈  | (c,di)  | CACCACTTGGCAGCATTAA CAAATACATTATTGATGGAAACG | JQ951624 120 2 | 160 140 140 140 |
| PS076  | (GTG)₆ | (c,da)  | CTGACAGCAACATGAAATGAA CAATCTTTGCCAATTTCCCA | JQ967035 161 1 | 170 170 170 170 |

1 Parentheses indicates the species possessing sequence from which primers were designed (c = P. cyananthus, da = P. davidsonii, di = P. dissectus, f = P. fruticosus).  
2 NM = no marker.
of an initial denaturation step of 5 min at 95°C, followed by 40 cycles of amplification consisting of 30 sec denaturation at 94°C, 30 sec for primer annealing at 55°C and 1 min of extension at 72°C. PCR products were separated on 1% agarose gels run in 0.5X TBE and visualized by ethidium bromide staining and UV transillumination. PCR products were purified using a standard ExoSAP (Exonuclease I/Shrimp Alkaline Phosphatase) protocol and sequenced directly as PCR products. DNA sequencing was performed at the Brigham Young University DNA Sequencing Center (Provo, UT, USA) using standard ABI Prism Taq dye-terminator cycle-sequencing methodology. DNA sequences were analyzed, assembled and aligned using Geneious software (Biomatters, Auckland, New Zealand).

Gene ontology
We used BLASTX [41] on assembled sequences of all four species to compare with the GenBank refseq-protein database [42] with a threshold of <1.0e-15. Blast2GO (v2.4.2) was used to map the blast hits and annotate them to putative cellular components, biological processes, and molecular functions found in the blast database [43]. For species comparisons, the GO level 3 was used for cellular components and level 2 was used for both biological processes and molecular functions.

Assembled sequences of all four species were also compared to all available Antirrhinum and Mimulus (genera more or less related to Penstemon) genes on GenBank (downloaded 23 June 2011). Comparisons were made using BLASTN [41] with an e-value threshold of <1.0e-15.

Results and discussion

Genome reduction, pyrosequencing and species assemblies
Given that a full 454 pyrosequencing plate using Titanium reagents is capable of producing 1.3 million reads averaging ~400 bp each [25], we expected a half plate to produce approximately 250 Mbp from 650,000 reads. Our reaction produced 287 Mbp from 733,413 reads, 20% more than expected, with an average read length of 392 bp. In total, 93.8, 46.4, 48.8, and 53.3 Mbp were sequenced from P. cyananthus, P. dissectus, P. davidsonii and P. fruticosus, respectively, closely resembling the 2:1:1:1 ratio of DNA pooled from each species for sequencing (Table 4). Likewise, from our de novo assemblies, we identified nearly twice as many contigs, 9,714 in P. cyananthus than the 4,777 found in P. fruticosus, for example, which was expected because we sequenced approximately twice as much DNA from P. cyananthus than the other three species. There was 0.6% of P. cyananthus genome represented compared to 0.5% average coverage of the other three species (Table 4); thus, essentially an equal genome representation from each species was realized using the GR-RSC technique by pooling approximately equal genome molar concentrations in the sequencing reaction. The contigs of this study have been deposited at DDBJ/EMBL/GenBank as a Whole Genome Shotgun project under the accessions AKKG00000000 (P. cyananthus), AKKH00000000 (P. dissectus), AKKI00000000 (P. davidsonii), and AKKJ00000000 (P. fruticosus). The version described in this paper is the first version for each accession, XXXX01000000.

DNA sequences produced by the GR-RSC technique represent a broad sample of the genome. With this sample, we can begin to estimate genome-wide characteristics, such as GC content, frequency of repeat elements, and so forth. From the genome reduction, GC content was measured to be 36.4%, 34.5%, 35.3%, and 35.15% for P. cyananthus, P. dissectus, P. davidsonii and P. fruticosus, respectively (Table 4), matching the average 35% GC content reported for dicots [44]. Using the dicot average GC content a priori, we estimated a theoretical frequency of the Bfai and EcoRI recognition sites. The theoretical GC content in combination with estimated genome sizes of the four species [5] suggested the GR-RSC should have rendered a 104 fold reduction of the genome of each species. With a reduced genome of these species, the 650,000 reads that were sequence suggest an average of 11× coverage; however the observed read depth was 8.5×, 22.7% less than expected (Table 4). This lighter coverage is partly due to the lower than expected specificity of reads. An average of 48.2% of the reads were matched to contigs with the other 51.8% either too short or lacking in homology to successfully match to a contig (Table 4).

The full assembly of all four Penstemon, using the Newbler de novo assembler, produced a total of 44,966 contigs, representing 16.4 Mbp, or 5.7% of our total sequence. In the individual species assemblies of P. cyananthus, P. dissectus, P. davidsonii, and P. fruticosus, a total of 9,714, 5,364, 4,882, and 4,777 contigs were created representing 4.6, 2.6, 2.4, and 2.3 Mbp of assembled bases respectively. These contigs represent, on average, 0.5% of the total genomes being sequenced (Table 4).

Marker analysis
We utilized assembly contigs from genomic sequence of all four species with “masked” multiple repeats, such as transposons, to identify SSRs. Penstemon cyananthus, P. dissectus, P. davidsonii, and P. fruticosus had 97, 113, 49, and 58 SSRs identified respectively (Table 5). There were more SSRs identified in P. dissectus than P. cyananthus, which has a 1.9 times larger genome and a higher representation of sequence than P. dissectus (Table 5). This inverse relationship between genome size and SSRs content agrees with observations in other plant genomes [45]. Some SSRs were found as putative homologs in
Table 4 Summary data from 454-pyrosequencing and Newbler de-novo assembly (v.2.0.01) of Penstemon cyananthus, P. dissectus, P. davidsonii, and P. fruticosus

| Assembly                  | Genome size (Mbp)¹ | GC content | Reads | Bases² | % Reads assembled | % Bases assembled | Contigs created | Bases in assembly | % Genome represented | Average coverage | Bases shared between assemblies |
|---------------------------|-------------------|------------|-------|--------|-------------------|-------------------|-----------------|-------------------|----------------------|----------------|---------------------------|
| P. cyananthus            | 893               | 36.4%      | 199,329 | 87,753,792 | 53.1%             | 50.0%             | 9,714           | 4,623,755 | 0.5%                 | 7.7X            |                           |
| P. dissectus             | 462               | 34.5%      | 98,868  | 43,304,550 | 52.8%             | 50.9%             | 5,364           | 2,629,819 | 0.6%                 | 8.2X            |                           |
| P. davidsonii            | 483               | 35.3%      | 103,963 | 45,599,742 | 45.8%             | 43.5%             | 4,882           | 2,376,141 | 0.5%                 | 9.1X            |                           |
| P. fruticosus            | 476               | 35.2%      | 113,146 | 49,786,980 | 41.0%             | 38.8%             | 4,777           | 2,322,606 | 0.5%                 | 8.9X            |                           |
| P. cyananthus × P. dissectus | 298,197       | 53.0%      | 131,058,342 | 69,150,079 | 50.1%             | 14,523           | 6,915,079 | 338,495 |
| P. cyananthus × P. davidsonii | 303,292       | 49.9%      | 133,353,534 | 67,570,234 | 46.9%             | 14,254           | 6,705,536 | 242,873 |
| P. cyananthus × P. fruticosus | 312,475       | 47.8%      | 137,540,772 | 67,05,536  | 44.9%             | 14,134           | 6,705,536 | 240,825 |
| P. dissectus × P. davidsonii | 202,831       | 48.3%      | 88,004,292  | 48,554,491 | 46.1%             | 10,053           | 4,855,491 | 150,469 |
| P. dissectus × P. fruticosus | 212,014       | 45.7%      | 93,091,530  | 47,744,539 | 43.5%             | 9,873            | 4,774,539 | 177,886 |
| P. davidsonii × P. fruticosus | 217,109       | 44.0%      | 95,386,722  | 44,421,194 | 41.7%             | 9,184            | 4,442,194 | 256,553 |
| Full Penstemon Assembly  | 730,215         | 47.9%      | 265,987,500 | 163,635,589 | 46.4%             | 44,966           | 16,363,589 |                    |

¹ The diploid (2n = 2x = 16) genome size as reported by Broderick et al. [5].
² Bases denotes the total number of bases used to create the assembly and not the total number of bases sequenced.
multiple species; after eliminating redundancies, we tallied 133 unique SSRs (Table 3). We generated primer pairs surrounding 77 of these SSRs large enough to potentially capture INDELs, of these, 51 produced 1 or 2 reproducible bands with no or few faint superfluous bands. From those 51, there was an overall success rate of 94% with 42 (82%) being polymorphic between the four species (Table 3).

To assess the possibility of utilizing these markers in interspecific plant improvement studies, 12 of the 51 SSR/INDEL markers (Table 3) were tested on 93 mostly xeric Penstemon taxa (72 species [Table 1]) representing five of six subgenera recognized in the genus [14]. The overall success rate of the markers was 98% with 100% being polymorphic across the 93 taxa. Without sequencing each band and/or doing inheritance studies on each marker it is not possible to clearly determine if a polymorphism of a given marker is a variant of an allele or a new locus. However, we did amplify and sequence the amplicon produced at 11 of these markers in five Penstemon species (P. cyananthus, P. davidsonii, P. dissectus, P. fruticosus, and P. pachyphyllus). P. pachyphyllus var. pachyphyllus represents the largest subgenus (Penstemon) in the genus. These five species represented four of the presently classified six Penstemon subgenera. Of the 55 attempted sequences, 60% produced high quality sequences results which could be compared to the original 454 contigs containing the microsatellites. Using BLASTN (v2.2.25+) [41] we found that 33 sequences matched the respective microsatellite-containing contigs from which the SSR/INDEL markers were derived with an e-value of no more than 1.0e-36. An example of the types of polymorphism (SSRs and INDEL) found at these loci across the various species is represented graphically for the marker PS035 (Figure 1). For 22 (40%) of the 55 attempted sequences, we were unable to obtain high quality sequence information. In the majority of these cases (94%) the lack of high quality data was clearly due to the amplification of multiple amplicons (seen as multiple bands in gel electrophoresis).

![Figure 1 An example of SSR and INDEL found in the comparisons of four Penstemon species in the sequences of marker Pen035.](image-url)
which impeded the sequencing of the PCR reaction. The source of the multiple amplicons may be from heterozygosity at the locus or from the amplification of paralogous loci.

Both the sequence data (Figure 1) as well as the marker size data (Tables 1 and 3) are clear evidence of sequence conservation, and probable homologous loci, in many of the SSR/INDEL markers. Marker PS012, the apparent most conserved marker, had six unique molecular weight bands and was present in all 93 taxa. The marker with the most diversity in its molecular weights was PS011 which had 18 variants and was not readable in seven of the 93 taxa. Of the 1,116 possible marker × taxa interactions, 22 (2.0%) did not produce reliable data. Seven of those 22 (0.5%) were absent of any product with the remaining 15 producing multiple bands (reported as ambiguous data). Clearly readable double bands were found in 135 of the 1,116 (12.1%) marker × taxa interactions (Table 1).

Our data suggest a high degree of sequence conservation across the genus, favoring the present hypothesis of a recent and rather rapid evolutionary radiation of the genus [13,14]. Furthermore, our data agree with Morgante et al. [45] who suggest that SSR presence in non-coding sequence are highly conserved and predate recent genome expansions of many plants. Some of our markers differed in length by as much as 570 bp (Tables 1 and 3) suggesting the presence of INDELs and possibly additional SSRs (Table 3). We confirmed the presence of INDELs in the sample of 11 markers which we sequenced (Figure 1). In some instances, these large fragment length variances may be amplifying a different locus, which is a recognized concern when using SSR based markers above the species level [46,47]. INDELs are useful as PCR based markers since they, like SSRs, are codominant and abundant in the genome and are commonly used in genetic mapping [26]. By combining the SSRs we identified in the source sequence for each of these markers with potential INDELS, alleles will be easily and inexpensively identified by gel electrophoresis.

To assess the possibility of phylogenetic (i.e., hierarchical) structure of the variation within these SSR/INDEL data at the broad taxonomic scale of our survey, we analyzed the 12 marker data set (Table 1) with PAUP*. Fast bootstrapping recovered a largely unresolved topology suggesting rampant homoplasy. Or one or more of these markers represent more than one locus. These results are similar to what others have reported about SSR type markers. SSRs have demonstrated utility for population and intraspecific relationships, such as cultivar differentiation; however, they can be problematic when used to reconstruct relationships above the species level where length differences are expected to poorly reflect homology [47,48]. Nonetheless, with over 96% of these SSR/INDEL regions being conserved across Penstemon, these markers have potential for studies of interspecific hybridization and cultivar development.

Interspecific Penstemon breeding is complex [7,11,15,49]; thus, having a set of inexpensive and easily used SSR/INDEL markers, which amplify across the genus, will have utility in understanding the results of some wide crosses. Empirical studies of various Penstemon interspecific crosses have ranged from a clearly recognizable intermediate phenotype of the two parents, to the F1 essentially mimicking one of the two parents, usually mirroring the female parent. Furthermore, in some instances the F2′s and additional generations continue to mimic the female parent to the point that Viehmeyer began to question if apomixis was involved. An example of this phenomenon was a ‘Flathead Lake’ × P. cobaea interspecific cross. It was not until the hybrid progeny of this cross was crossed with other interspecific hybrids when the progeny gave a much wider range of phenotypes [49]. A probable reason for this phenomenon is “unequal segregation” which has been described in other wide crosses [50,51]. Thus through the use of these SSR/INDEL markers, regions of the genome can be identified which are unusual genotypic combinations, for that specific cross, and selections made accordingly [51-54]. Thus increasing the number of unique genotype/phenotype plants to be grown out to maturity from thousands of seedlings. Since many Penstemon require two years before their first anthesis, using markers to identify the greatest number of genotypic diverse plants is potentially very useful in the breeding of this crop.

Beyond amplification ability, we also assessed the composition and trends of all SSRs identified. On average, adenine and thymine rich repeat motifs were the most common repeat type in the di-, tri-, and tetra-nucleotide repeat motifs (Figure 2). In general, AT motifs are the most common motifs in noncoding regions of most plant genomes [45]. More variation was observed in the repeat motifs in the tetra-nucleotide repeats across the four species. Even closely related P. fruticosus and P. davidsonii had completely distinct tetra-nucleotide repeat motifs (Figure 2). This is likely due, in part, to the rarity of the motifs and high number of possible nucleotide combinations. Several studies have found that the hypothetical origins of some SSRs are retrotransposition events [48,55,56] and, as such, may be useful in developing part of a unique “fingerprint” for a given species.

SNP analysis
Using our SNP discovery parameters of an 8× minimum coverage, and 30% representation of the minor allele, we identified an average of one SNP per 2,890 bp across the four species ranging from P. cyananthus (1/1,855 bp) to
The three species with similar genome sizes all had similar SNP frequencies (Table 6). As reported in other plant species [57,58], we found that the frequencies of bp transitions (A ↔ G or C ↔ T) were more common compared to transversions (A ↔ T, A ↔ C, G ↔ C, G ↔ T) in *Penstemon* by an average factor of 1.5 (Table 6). This is close to the 1.4 factor in *Arabidopsis* [35]. In the dual species assemblies, using the same parameters and a 90% SNP identity, the average transition to transversion mutation rate was lower at 1.2 (Table 6).

Table 6 SNP type and distributions along with SNP comparisons of sequences found within and between species (homologous sequence comparisons) using SNP_Finder_Plus (8X min. coverage, 30% min. minor allele, 90% min. identity)

| Species assembly          | SNP       | Average coverage | SNPs/assembly length | SNP distribution |
|---------------------------|-----------|------------------|----------------------|------------------|
|                           |           |                  |                      |                  |
| *P. cyananthus*           | 2,493     | 16.4             | 0.000539 (~1/1855 bp)|                  |
| *P. dissectus*            | 737       | 14.3             | 0.000720 (~1/3568 bp)|                  |
| *P. davidsonii*           | 713       | 14.4             | 0.000300 (~1/3333 bp)|                  |
| *P. fruticosus*           | 615       | 12.4             | 0.000265 (~1/3777 bp)|                  |
| *P. cyananthus × P. dissectus* | 3,253     | 10.6             | 0.000610 (~1/104 bp)  |                  |
| *P. cyananthus × P. davidsonii* | 1,958     | 10.7             | 0.000862 (~1/124 bp) |                  |
| *P. cyananthus × P. fruticosus* | 2,013     | 10.6             | 0.008367 (~1/119 bp) |                  |
| *P. dissectus × P. davidsonii* | 2,348     | 10.8             | 0.015605 (~1/64 bp)  |                  |
| *P. dissectus × P. fruticosus* | 2,133     | 10.0             | 0.011991 (~1/83 bp)  |                  |
| *P. davidsonii × P. fruticosus* | 2,156     | 10.1             | 0.008404 (~1/119 bp) |                  |

1 Assembly length is bases shared between assemblies (see Table 4).
of any two of the four species. The frequency of SNPs between homologous sequences of P. dissectus and P. davidsonii was the highest at 1/64 bp, with the lowest being between P. cyananthus and P. davidsonii at 1/119 bp. These results are in line with previous molecular based studies [5,14]. Penstemon davidsonii and P. fruticosus both belong to subgenus Dasanthera, while P. cyananthus and either P. davidsonii or P. fruticosus homologous sequences had fewer SNPs at 1/124 and 1/119, respectively. All homologous sequence comparison involving P. dissectus had the highest density of SNPs (Table 6) suggesting that P. dissectus is the most evolutionary distant of the four species.

It is important, for a high degree of confidence in the results, when the “SNP identity” parameter in SNP_Finder_Plus to have two or more independent samples from the same species. This requirement was not met for each of the species assemblies, thus, introducing a weakness in our interspecific SNP comparisons. Although with the parameters of a minimum 8× coverage and minor allele frequency set at least 30%, a putative SNP must be present in at least three of the eight contig reads, thus providing some protection from mislabeling a sequencing and/or assembly error as a SNP. Furthermore, when doing across species comparisons the average SNP coverage was actually 14.4× (Table 6). Therefore, on average, five identical putative SNPs represented the minor allele.

To understand the viability of our interspecific SNP as markers, we utilized the 1,958 P. davidsonii × P. cyananthus and 2,348 P. davidsonii × P. dissectus SNPs identified in the 14,254 and 10,053 respective homologous contig parings (Tables 4 and 6). After removing contigs absent of identifiable SNPs, putative repetitive elements, and nonnuclear plastid DNA, 431 remained. Of these contigs, 99 were homologous across all three species (P. cyananthus, P. davidsonii and P. dissectus) another 164 were only in the P. davidsonii × P. cyananthus comparisons while the remaining 168 were in the P. davidsonii × P. dissectus contigs. Of those 431 contigs, we selected the first 192 for SNP marker development, 86 from each of the species comparisons. These contigs were utilized for competitive allele-development, 86 from each of the species comparisons. Of those 431 contigs, 28.5%, 16.8%, 17.4% and 16.1% of the respective sequence from P. cyananthus, P. dissectus, P. davidsonii, and P. fruticosus as repeat elements using RepeatModeler and RepeatMasker. Of these elements, 3.0-7.8% were identified as LTR (long terminal repeat) retroelements, 0.3-1.0% transposons and the remainder were unclassified (Table 5). Since RepeatModeler utilizes RECON and RepeatScout to create a de novo model in RepeatMasker in place of the Arabidopsis model, details about the subcategories of LTRs and transposons which are included in the model could not be addressed. Maughan et al. [35] utilized GR-RSC on the Arabidopsis lines Ler-0 and Col-4. Utilizing RepeatModeler, then RepeatMasker on their sequence data from these lines, we found an average of 6.2% were identified as repetitive elements, of which 4.4% were identified as LTR retroelements and 0.4% were transposons. By way of comparison, the downloaded full “non-genome reduced” sequence of Arabidopsis line TAIR10 had a similar 7.4% of the sequence identified as repeat elements of which 3.0% were LTR retroelements and 0.2% were transposons (Table 5 and Figures 3 and 4). These data suggest that the GR-RSC method reflects, at least for repetitive elements, similar proportions as to that found in the full sequence of Arabidopsis.

Broderick et al. [5] hypothesized that the broad range found in Penstemon genome sizes, of the same ploidy, may be explained by retrotransposons. Lynch [60] detailed a relationship between genome size and repeat elements suggesting a linear relationship between the number of elements and genomes size [60-62]. The four Penstemon species used in this study provide insufficient evidence to establish a linear relationship between genome size and repeat elements in Penstemon. However, the three smaller, similar sized, Penstemon genomes possess comparable quantities of repetitive elements whereas P. cyananthus (the largest genome) has nearly double the number of repeat elements compared to the other three species (Figure 3).

Not only do repetitive elements largely influence genome size, but they are also likely to evolve more rapidly than do low-copy sequence [62,63]. Thus, repetitive elements of a species take on unique “fingerprints” which become valuable in phylogenetic relationship studies [64,65]. Thus, our limited four Penstemon species genomic data set suggest agreement with the two hypotheses that firstly, repetitive elements are a major component of the genome size variation identified by Broderick et al. [5]. Secondly, these elements are variable between the species we tested suggesting the possibility of identifying species
| Name          | Contig       | SNP Allele Source 1 | SNP Allele GenBank Accession #2 | SNP Type | SNP Allele Specific A1 Forward (5′→3′) 3 | Common Allele Specific Reverse (5′→3′) 3 | P. davidsonii | P. cyananthus | P. dissectus | P. fruticosus | P. pachyphyllus | P. cyananthus + P. davidsonii | P. dissectus + P. davidsonii |
|---------------|--------------|---------------------|---------------------------------|----------|------------------------------------------|------------------------------------------|--------------|--------------|--------------|--------------|----------------|-------------------------------|-----------------------------|
| PenSNP00001   | 00336CD      | A/G                 | PenSNP00001                     |          | AAGATTGCA TGGAGAGGA AATGGATT AGATTTGCA | CGATCCAAA TGGCAGATG CGAGAAA               | X            | Y            | X            | X            | Y              | Y                             |                             |
| PenSNP00002   | 00405CD      | C/T                 | PenSNP00002                     |          | AGCGAGTAT AAGATTTGG TTTTCTTC            | CCAACACTT CCGCAGAAA CTCTTAA              | Y            | X            | Y            | X            | Y              | H                             | H                           |
| PenSNP00003   | 02625CD      | A/T                 | PenSNP00003                     |          | AAAAGCTCC CAAACATGA CTAAGAAGT          | AATTCCTCGA CACATTGAGA GAGCGTAA          | Y            | X            | Y            | X            | Y              | H                             | H                           |
| PenSNP00004   | 02857CD      | A/C                 | PenSNP00004                     |          | ATCAAAATGA ACTTGTCCT ATGAGGCT          | GCAAACAGGT GCAAAAAATT GTAGCGTAA         | X            | Y            | X            | X            | H              | H                             | H                           |
| PenSNP00005   | 03943CD      | A/G                 | PenSNP00005                     |          | ACTACAAAA ACTACCTTT CTCTTA             | GGGGTACAGA GTTGAAAGA AGGAA              | X            | Y            | X            | X            | H              | H                             | H                           |
| PenSNP00006   | 04420CD      | A/C                 | PenSNP00006                     |          | TCGCTCTCAA ATCGATATG ATGAGGCT          | GTGGTTCTTC CCCCTTAA GGCATT              | Y            | X            | Y            | X            | Y              | H                             | H                           |
| SNP Marker Name | Accession ID | Polymorphism Type | Primer Sequences | A1 | A2 | Common Allele Specific Reverse |
|-----------------|-------------|-------------------|------------------|----|----|-----------------------------|
| PenSNP00007     | JX649984    | A/T               | GGCAACATC        | X  | X  | X                           |
|                 |             |                   | CTCAGCAGA        | X  | X  | X                           |
|                 |             |                   | GACA             | X  | X  | X                           |
|                 |             |                   | GGCAACAT         | X  | X  | X                           |
|                 |             |                   | CCTCAGCA         | X  | X  | X                           |
|                 |             |                   | GAGACT           | X  | X  | X                           |
| PenSNP00008     | JX649985    | C/G               | GGTTGGTA         | X  | X  | X                           |
|                 |             |                   | TTGGTTAC         | X  | X  | X                           |
|                 |             |                   | TTTATGGG         | X  | X  | X                           |
|                 |             |                   | GGTTGGTAT        | X  | X  | X                           |
|                 |             |                   | TGGTTACT         | X  | X  | X                           |
|                 |             |                   | TTATGGC         | X  | X  | X                           |
| PenSNP00009     | JX649986    | C/T               | ACAATTTTG        | X  | X  | X                           |
|                 |             |                   | ATAATTCCATT      | X  | X  | X                           |
|                 |             |                   | CTCAGTGGT       | X  | X  | X                           |
|                 |             |                   | CACAATATTI       | X  | X  | X                           |
|                 |             |                   | GATAATTCCAT      | X  | X  | X                           |
|                 |             |                   | TCTCAGTGGCA      | X  | X  | X                           |
| PenSNP00010     | JX649987    | A/C               | AGCTGATTAT       | X  | X  | X                           |
|                 |             |                   | TCCCTAAAAAC      | X  | X  | X                           |
|                 |             |                   | CCAATT           | X  | X  | X                           |
|                 |             |                   | GCCTGATTAT       | X  | X  | X                           |
|                 |             |                   | CCCCTAAAAC       | X  | X  | X                           |
|                 |             |                   | CCAATG        | X  | X  | X                           |
| PenSNP00011     | JX649988    | C/T               | TTTGGGACCT       | X  | X  | X                           |
|                 |             |                   | GCAAATGAC       | X  | X  | X                           |
|                 |             |                   | CATC             | X  | X  | X                           |
|                 |             |                   | CTGGGACAC        | X  | X  | X                           |
|                 |             |                   | TGGACTGAC       | X  | X  | X                           |
|                 |             |                   | CATT             | X  | X  | X                           |
| PenSNP00012     | JX649989    | A/G               | AGATAGAC         | X  | X  | X                           |
|                 |             |                   | GTGGTTTTCT       | X  | X  | X                           |
|                 |             |                   | TTCAAGCA         | X  | X  | X                           |
|                 |             |                   | AGATAGACG        | X  | X  | X                           |
|                 |             |                   | TGTTATTTCT       | X  | X  | X                           |
|                 |             |                   | TCACGGG          | X  | X  | X                           |
| PenSNP00013     | JX649990    | A/T               | TATTTCTT         | X  | X  | X                           |
|                 |             |                   | TCTGCAAATC       | X  | X  | X                           |
|                 |             |                   | TCACATTG        | X  | X  | X                           |
|                 |             |                   | ATTTTCCTT        | X  | X  | X                           |
|                 |             |                   | CTGAATCT        | X  | X  | X                           |
|                 |             |                   | CAACATTG        | X  | X  | X                           |
| PenSNP marker name | GenBank dbSNP accession ID | Polymorphism type | KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued) |
|-------------------|-----------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PenSNP00014       | JX649991                    | A/C               | AGGCCTGTGG CTGACTTTGCA GGCCGTGTT CTGACTTTGCA                                                                                                                                                    |
|                   |                             |                   | GGCATATCT TGCCCGGT TCCACAA                                                                                                                                                                    |
| PenSNP00015       | JX649992                    | A/C               | AAATGCTC CCTCATTITG ACCATATGA ATGCTCCCT CATTITGC CATTGTC                                                                                                                                   |
|                   |                             |                   | GTCAACGG ATTTGGAAG AGTCGGTA                                                                                                                                                                  |
| PenSNP00016       | JX649993                    | C/G               | TGAAAAATTC AGATTTATGG AAAAAACAGTC GAAAAATTICA GATTIAAGAA CAAAACAGTC                                                                                                                        |
|                   |                             |                   | AGACTTTGAA CAAATTTCCT GGGTCAAAA                                                                                                                                                              |
| PenSNP00017       | JX649994                    | A/G               | TGACCAAGGA ATCTGTCAGAG AAGCTT GACCAAGGA ATCTGTCAGA GAACCT                                                                                                                                |
|                   |                             |                   | CTCTACTGTG GCTGGTTACC TCTA                                                                                                                                                                   |
| PenSNP00018       | JX649995                    | G/T               | TACCTCCAAT TGTGATGCA ACATTAG CTACTCCTCCA ATTTGATGTC AACATTAG                                                                                                                             |
|                   |                             |                   | CTAAGTGAA GAACACAG AGGA                                                                                                                                                                    |
| PenSNP00019       | JX649996                    | G/T               | ATCTCCTCC TCTTGAGCT AAAGC CATTCCCT TCTTTTGAC ATCAAAGA                                                                                                                                        |
|                   |                             |                   | GAGCCCA CCTCGACT GCTCTATTT                                                                                                                                                                  |
| PenSNP00020       | JX649997                    | A/G               | AAGGACTG AGTACCAA GACAGACTG GGACTGAG TACCAAGA CAGATCC                                                                                                                                        |
|                   |                             |                   | GCCAGGGTA CTGAACTG TCGTTA                                                                                                                                                                  |
Table 7 Penstemon SNP marker, GenBank dbSNP accession ID, polymorphism type, KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued)

| SNP Name          | GenBank ID | Polymorphism Type | Primer Sequences | Y | X | Y | H | H |
|-------------------|------------|-------------------|------------------|---|---|---|---|---|
| PenSNP00021       | JX649998   | C/T               | AGCATATTG AAAAAAGATC AGTGGCATAG AAAGCATAT AGAAAAGATC CAGTCGATTAA | Y | X | Y | Y | H |
| PenSNP00022       | JX649999   | A/G               | AAATACCT GAGCTTCT GCCTTTGT ACCTT GAGTCGCATAG CATTCGGCC TCTTGCG | X | Y | Y | Y | H |
| PenSNP00023       | JX650000   | C/G               | ACCATTACG GTAATATT CCAAAAGGC ACCATTACG GTAATATT CCAAAAGGC | Y | X | Y | Y | H |
| PenSNP00024       | JX650001   | A/G               | GTCAATTGT CAAGTGTTGA TTTTTCTTACATA ACATTGCA AGTGTTGATTT TCTTACAGT | Y | X | Y | Y | H |
| PenSNP00025       | JX650002   | A/T               | ATCCGATTCT TCCTCCG TCTTGGACCGT CCAAAGAAGGAA | X | Y | X | Y | H |
| PenSNP00026       | JX650003   | A/G               | TCTCTTCG TCTTT TCTTCTCT CCTCCTCG TCTTCC TCTTCC TCTC | Y | X | Y | Y | H |
| PenSNP00027       | JX650004   | A/G               | TCGACC CAAACCTG TCACA CTTGCTTGGCT TCTGGAAGAG | Y | X | Y | Y | H |
Table 7 Penstemon SNP marker name, GenBank dbSNP accession ID, polymorphism type, KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued)

| SNP Name     | Accession | Type   | KASPar™ Primer Sequences |
|--------------|-----------|--------|--------------------------|
| PenSNP00028  | JX650005  | C/T    | TGGTATCTTTGGTTGAAAAGTCTACCAAC, TGGTTTGAAACCTTTGTTGTTGAA, TACCTCTAAACTCTCGGAT |
| PenSNP0029   | JX650006  | A/G    | TGGTCCTGTTCCTTACCATACACATACATGGAT, CTTTACCATTCTTGTGATCTGTTGTTGAA, TAAAGCTATGAGGGAAGTT |
| PenSNP0030   | JX650007  | A/G    | AGTAGTACAGAATACTTAAGAACATACACCA, AGTAGTACAGAATACTTAAGAACATACACCA, AGTAGTACAGAATACTTAAGAACATACACCA, AGTAGTACAGAATACTTAAGAACATACACCA |
| PenSNP0031   | JX650008  | C/T    | AGTGGATGGTGTGGTACGAGTCGCGGCTTTGA, TCCGATCTTGAATTCTA, TACTACAAAAGGTTAAGTGCAATTCATAGTTGATTCTA |
| PenSNP0032   | JX650009  | A/T    | GTACCAGGCCGTTGATATGAGCTCCACCAAAATCCGGCGGCAATTAAATCTCTAAATCACCTGGT |
| PenSNP0033   | JX650010  | G/T    | GTTTGATTCTAAGTTGATCTTGA, TCTTGGATTGAGATTTG, CAGATCTTAAATCTCTTAAATCACCTGGT |
| PenSNP0034   | JX650011  | C/G    | ACATTAGGGATTCAATGTTCAAATACTTGCGTCAATCACCATAGCTTAAATCTCTTAAATCACCTGGT |

Table continued...
| PenSNP marker name | GenBank dbSNP accession ID | Polymorphism type | KASPar™ primer sequences (A1, A2 and common allele specific reverse) | (Continued) |
|-------------------|----------------------------|------------------|-------------------------------------------------|-----------|
| PenSNP00035       | 08352DD                    | A/T              | AGTACAAGGA AAAACCTTTA TTAGTAAGTATA AGTACAAGGA AAAACCTTTTAT TAGTAAGTATT | CTGACACAA ACCCATCTA ATATGACCAA |
| PenSNP00036       | 08488DD                    | A/T              | GTGTGGGAG AGCCAGGT GCCA | GTATTGAGGAT CATTCTGACAA AAAACATA |
| PenSNP00037       | 08608DD                    | C/T              | GTAGATAAG TTGATGGCGA GACG | CCAAAACAAT GCACCAACATT CTCCTT |
| PenSNP00038       | 08831DD                    | A/T              | TTGAACCGC CATGGAAAAT GTTTTTAGA TTGAACGTC ATGAAAAGTT GTTTTAGT | ATTTTGACCA AGGAGCCTATC AGAGG |
| PenSNP00039       | 08947DD                    | A/T              | GGGATCGTAA AACTCAGGAA AAATGA | TCGAATACTC GTGGGCTCTT CGATT |
| PenSNP00040       | 08959DD                    | A/G              | AGAGAATGAG AAGGAGAAGGA AGAAA GAAATGAGAA AGAGAAGGA AGAGA | CT CCTACGG TTGACTATC GTAGTA |
| PenSNP00041       | 09272DD                    | A/T              | TTCTACAAAC AATCAGAGTC ATCATT | TCGACACCTT TTGCCCTATC TTGAA |

Dockter et al. BMC Genetics 2013, 14:66
http://www.biomedcentral.com/1471-2156/14/66
Page 24 of 33
Table 7 Penstemon SNP marker, GenBank dbSNP accession ID, polymorphism type, KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued)

| SNP ID     | GenBank ID | Polymorphism Type | Primer Sequences (A1, A2 and common allele specific reverse) |
|------------|------------|-------------------|-------------------------------------------------------------|
| PenSNP00042| JX650019   | C/T               | GTTTTATACG CATCCATATAC ATAATAATAG GTTTTATACGC ATCCATATAC ATAATAAAG GTTTTATACGC ATCCATATAC AATAATAAG |
|            |            |                   | GGTTCACCTCT CCAGAAAATAA AATCTTATAT                          |
|            |            |                   | Y X Y Y H X                                               |
| PenSNP00043| JX650020   | A/G               | AATTCACAGTC AAATGCAAG GTTGCA CAACGTCAAATT GGAAGGGTGGCG |
|            |            |                   | TTCACATAC CGCCTGAGT TGGCAT                                 |
|            |            |                   | Y X Y Y H H                                               |
| PenSNP00044| JX650021   | A/G               | TTTTTAATAAT ATCCGTGGTAG AATTTAT TTTTTAATAAT GTTTTATACGC ATCCATATAC AATAATAAAG |
|            |            |                   | AAATGAGT GGATGGCTA GGAAGACTAA                             |
|            |            |                   | X Y X Y X Y X H H                                          |
| PenSNP00045| JX650022   | A/T               | AGATCTGGAG ACTAAAT AGATCTGGAG ACTAAAT                     |
|            |            |                   | CGAAGAGTT TGGGAGGC CGAT                                   |
|            |            |                   | X Y X Y Y                                                |
| PenSNP00046| JX650023   | C/T               | GTCCGACGGT ACAAAGCAGC CTGGCCGAGT GACAGTCAGT               |
|            |            |                   | CGCCGTCAAA AGAGACCTT TGGGAGTG                              |
|            |            |                   | Y X Y Y H H                                               |
| PenSNP00047| JX650024   | C/T               | AGGAGATTCCT CTGCGTGGAGCT GGAGAGATTCCT CTGCGTGGAGCT       |
|            |            |                   | TCTTCACATG TTTGAGTTA GGCTGAAT                              |
|            |            |                   | X Y X Y X X H H                                           |
| PenSNP00048| JX650025   | A/G               | ACCTCCATGAGA GACCATAAA CTAGGCTCGATGG AGGACCATAGA         |
|            |            |                   | GCTGCTCTCC TGGAGAA CTTCT                                  |
|            |            |                   | X Y X X H H                                               |
| PenSNP00049| JX650026   | G/T               | AAAATGCATGTA GTTTGGTTTACG AAAATGCATGTA GTTTGGTTTACG   |
|            |            |                   | CACACCCCA AAGGCAAG AATGACAT                                 |
|            |            |                   | X Y X Y H H                                               |
| SNP Name     | GenBank Accession | Polymorphism Type | Primer Sequences                                      | KASPar™ Primer Sequences |
|--------------|-------------------|-------------------|-------------------------------------------------------|--------------------------|
| PenSNP00050  | 13159DD           | C/T               | TGAATGTACCTTT TCAATGATAGA GAACG GTTGAATGTAC TTTTCATTGATA GAGAACA | Y Y X Y Y Y H           |
| PenSNP00051  | 13463DD           | G/T               | GCCTTTGACG GCCGAAGGAT TTC CGCCTTTGA CCGGGAAG GATTGA | X Y X X H H             |
| PenSNP00052  | 14334DD           | A/G               | AGAAACAAC AAATACGAA TAAATCAGGCA AAAAAACA AAATACGGA TAATTACCCCA | X X Y H X X Y           |
| PenSNP00053  | 00290DD03373CD    | C/T               | TGCCCTTTGGG CGGCCCACAAAGCCTAAGAGA TTGGCTGACTCTA CTTGTTTA | Y X Y Y Y H H           |
| PenSNP00054  | 00354DD04637CD    | A/G               | GCAAAACGG GACACCTCA TTGATT CAAAAGGGA ACCCTCAG GTTACTTCGTC | X X X X X H             |
| PenSNP00055  | 01161DD11697CD    | A/G               | ACTGTAAA TACACTAGG TTACAGT CTGTGAAA TACACTAC GTTACAGC | Y Y X X Y Y Y H         |
| PenSNP00056  | 01323DD15501CD    | A/G               | ACTTGAAGA ATTTTCTCAC TACTTCGT CCTGAAAGAA TTGTTTACACT ACGCGTTC | X X X H H             |
| SNP Name        | Accession | GenBank Accession | Polymorphism Type | Forward Primer (A1) | Reverse Primer | Heterozygous | Homozygous A1 | Homozygous A2 | Common Allele Specific Reverse |
|-----------------|-----------|-------------------|-------------------|---------------------|----------------|--------------|---------------|---------------|-----------------------------|
| PenSNP00057     | 01541DD   | JX650034          | A/G               | AATTAGAAC           | GGAGCCCAA      | Y            | X             | Y             | H                          |
|                 | DD02481CD |                   |                   | CAACATCCAC          | CTTTTACATT     |              |               |               |                     |
|                 |           |                   |                   | TGATCCCAA          | CTTTTCTA       |              |               |               |                     |
|                 |           |                   |                   | AGAACCACA          |                |              |               |               |                     |
|                 |           |                   |                   | TCCACTGAT          |                |              |               |               |                     |
|                 |           |                   |                   | TCCAG             |                |              |               |               |                     |
| PenSNP00058     | 02019DD   | JX650035          | A/G               | GTGATTGTTA         | GTACGAGGCT     | Y            | Y             | X             | H                          |
|                 | DD0312CD  |                   |                   | AATCTGAATA         | TCGAAAAAGA     |              |               |               |                     |
|                 |           |                   |                   | TATAATTTCTTTT      | CCAGAT         |              |               |               |                     |
| PenSNP00059     | 02851DD   | JX650036          | A/G               | AAGAGGTTGA         | GAAGAAAAATC    | X            | Y             | X             | H                          |
|                 | DD1719CD  |                   |                   | TCCTAAGTTA         | ATGTCCAC       |              |               |               |                     |
|                 |           |                   |                   | TCGAGG             | ATCTCGGTA      |              |               |               |                     |
|                 |           |                   |                   | AGAGGTTGAT         |                |              |               |               |                     |
|                 |           |                   |                   | CTTAAGGTAT        |                |              |               |               |                     |
|                 |           |                   |                   | CGAGG            |                |              |               |               |                     |
| PenSNP00060     | 03089DD   | JX650037          | C/T               | TTTCAGAGTC         | GATTTCTTG      | X            | X             | Y             | H                          |
|                 | DD14703CD |                   |                   | ACTAATGTTC         | TCCATCTCTT     |              |               |               |                     |
|                 |           |                   |                   | TCAG             | CAAGATGTA      |              |               |               |                     |
|                 |           |                   |                   | GTTTTCAGAG        |                |              |               |               |                     |
|                 |           |                   |                   | TCACAAATGT        |                |              |               |               |                     |
|                 |           |                   |                   | TCTCACA           |                |              |               |               |                     |
| PenSNP00061     | 03423DD   | JX650038          | A/C               | AATTCTTCTTA        | TATCTTAGA      | Y            | X             | Y             | H                          |
|                 | DD25897CD |                   |                   | CTCATCATATTG       | CATGGACAT      |              |               |               |                     |
|                 |           |                   |                   | ATCCGGA          | GGAATAATGGA    |              |               |               |                     |
|                 |           |                   |                   | ATCTACGT         |                |              |               |               |                     |
|                 |           |                   |                   | CCATTGTGAT        |                |              |               |               |                     |
|                 |           |                   |                   | CGAGG            |                |              |               |               |                     |
| PenSNP00062     | 04632DD   | JX650039          | A/T               | AAATGGGT          | CTCTCTTTAC     | Y            | Y             | X             | H                          |
|                 | DD19186CD |                   |                   | CAGCTGCAA         | TCTGTTTTTCT    |              |               |               |                     |
|                 |           |                   |                   | ATTTCCGCA        | TCTTTT         |              |               |               |                     |
|                 |           |                   |                   | AAATGGGT          |                |              |               |               |                     |
|                 |           |                   |                   | CAGCTGAA         |                |              |               |               |                     |
|                 |           |                   |                   | ATTTCCGCT        |                |              |               |               |                     |
| PenSNP00063     | 05160DD   | JX650040          | C/G               | TGATCGTGTG        | GATCCCATATA    | Y            | X             | Y             | H                          |
|                 | DD08243CD |                   |                   | AAAATGATAAT       | GACCTCTTTT     |              |               |               |                     |
|                 |           |                   |                   | TGATACAA         | AAGATCTTAA     |              |               |               |                     |
|                 |           |                   |                   | CGATCGTGTG        |                |              |               |               |                     |
|                 |           |                   |                   | AAAATGATAA       |                |              |               |               |                     |
|                 |           |                   |                   | TTGATACAAC       |                |              |               |               |                     |
| Table 7 Penstemon SNP marker name, GenBank dbSNP accession ID, polymorphism type, KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued)
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PenSNP00064 | 06332DD03627CD | JX650041 | A/C | ATCAAATGCC ATAGATCCCTG CAGATTAG CAAATGCCA TAGATCCCTG CAGATTAG | ACATTTCCATAC ACCAACCTTCTT CCACCTA | X | X | Y | X | X | H |
| PenSNP00065 | 08748DD13630CD | JX650042 | C/G | AGCTGTTC AGGAGGT TCATGAATGCCA CACCATGTA AGGAGGTTCATGAATC | CACCAGTGAA ACCAACACTTATTGTCATT | X | Y | X | X | H | H |
| PenSNP00066 | 09773DD14323CD | JX650043 | A/G | TCATGCCCATACTCCAC CATGCCACATGCAGCATGA | CCTGATATGA ACATGGGGAGGTAT | Y | Y | X | Y | Y | Y |
| PenSNP00067 | 10248DD06150CD | JX650044 | C/G | TGTGTGATTGAAATCTG CCCCCTGCTGTTTCATATCT | GTTTGTATCTACCCTTGAGCATTTGAAA | Y | Y | X | Y | Y | Y | Y | H |
| PenSNP00068 | 10624DD11358CD | JX650045 | A/G | GTGCCAGT GTGAAACTGCA GCATCA | GTTTTTTCCTGGGTGCTAAGGTTCAT | Y | X | Y | Y | Y | H | H |
| PenSNP00069 | 11267DD06273CD | JX650046 | A/C | ACCAAATATACTATCTGAGTCACCAGAAGCA CCAAAATACATTTAGCTCCTCACAGTGAC | GACTGGAAGATGGTGTGCAAAGGAC | Y | X | Y | Y | Y | H | H |
| PenSNP00070 | 11564DD17128CD | JX650047 | C/T | TGGACGTTC GACATGGGA ACACAAAGATC AATTGGGAGACATGGAATT | ATATGAAA CTCCCCAC AAGAAA | Y | X | Y | X | H | H |
| PenSNP marker name | GenBank dbSNP accession ID | Polymorphism type | KASPar™ primer sequences (A1, A2 and common allele specific reverse) for all 75 functional SNP assays (Continued) |
|--------------------|---------------------------|------------------|------------------------------------------------------------------------------------------------|
| PenSNP00071        | 11647DD17264CD            | C/G              | GCACGAGCC AAAATCTT GAGC  
GCACGAGCC AAAATCTT GAGC  
ATTTGCATG TGATCCT GTGTTGGGA |
| PenSNP00072        | 11671DD17144CD            | C/T              | GTGCAGCA ACCTTATT CATGAC  
ATTTGCAG CAACCCCT ATCTAGAT |
| PenSNP00073        | 12915DD17470CD            | A/G              | AAGAAAAG GGTGGACAA ATTAAACCGT  
AAGAAAAG GGTGGACAA ATTAAACCGT  
CAGAACAC ATCATACCTG ATAAAATCTCTT |
| PenSNP00074        | 13828DD14937CD            | C/T              | GTAAAGATAT GCTGCCAGA TGG  
GTAAAGATAT GCTGCCAGA TGG  
CCTGTAAGAA GTTTTGTCC CTAGCTA |
| PenSNP00075        | 14286DD18608CD            | G/T              | GTATTGAG AGCCCTT ACCCG  
GTATTGAG AGCCCTT ACCCG  
CCACTTGAAT TGTTGAAGA GTGTTGGGA |

1These contigs have been deposited at DDBJ/EMBL/GenBank as a Whole Genome Shotgun project under the accessions AKKG00000000 (P. cyananthus), AKKH00000000 (P. dissectus), AKKI00000000 (P. davidsonii), and AKKJ00000000 (P. fruticosus). The version described in this paper is the first version for each accession, XXXX01000000.

2The GenBank accession identification for the full sequence for each allele with the specific SNP bp identified.

3KASPar™ primers: A1 and A2 primers are SNP allele specific. All A1 Forward primers had the follow universal primer GAAGGTCGAGCTCAACGGATTTGCAATGCT added to the 5' end of the allele specific sequence. All A2 Forward primers had the follow universal primer GAAGGTCGAGCTCAACGGATTTGCAATGCT added to the 5' end of the allele specific sequence.

4H = heterozygous compared to either homozygous condition for either “X” or “Y.”
specific repetitive elements. However, without further comparisons we were unable to identify specific repetitive elements associated with the four *Penstemon* species used in this study.

**Gene ontology**

Using BLASTX we identified an average of 21.5% of the contigs across the four species as putative genes with an average of 13.9% annotated by Blast2GO (Table 5). These putative genes were compared and contrasted in a more detailed study by Dockter [23]. Furthermore, he compared the *Penstemon* sequences to known genes from the related genera *Antirrhinum* and *Mimulus*, and identified nine putative *Penstemon* genes from *Antirrhinum* and 14 from *Mimulus* with an e-value below 1.0e-13. Three genes (NADH dehydrogenase from *M. aurantiacus*, ribosomal protein L10 from *M. guttatus*, and ribosomal protein subunit 2 from *M. aurantiacus, M. szechuanensis*, and *M. tenellus var. tenellus*) were perfect hits (e-value = 0.0).

**Conclusions**

*Penstemon* are recognized for their phenotypic variation and their adaptation to multiple environments...
[6-8,13,14,17,30,31]. Broderick et al. [5] found that this diversity is reflected by a wide range in their genome sizes. Nevertheless, even with this demonstrated plasticity we have identified evidence that there is a high level of sequence conservation across the genus. This apparent sequence conservation is in harmony with the hypothesis that Penstemon has rapidly irradiated to its variety of species rather recently in evolutionary time [13,14]. Furthermore, our study identified evidence that the genome size variation in Penstemon is rooted in the amount of repetitive elements in each species.

Despite the large differences in Penstemon’s genome size, the finding that the genus has a great deal of sequence conservation is invaluable for the development of interspecific markers. The further development and mapping of a number of conserved markers will facilitate the domestication of xeric Penstemon cultivars via interspecific hybridization which are largely unexploited largely due to crossing barriers [6-8,10-12,15]. Viehmeyer [16] hypothesized that it might be possible to develop Penstemon breeding lines that would facilitate the indirect interspecific hybridization of any two species within the genus. He and others have used traditional breeding techniques to develop a number of interspecific hybrids [7,11,15,17,66]. Clarifying the phylogenetic relationships within the genus should facilitate these objectives [67]. In the largest Penstemon phylogenetic study conducted to date, Wolfe et al. [14] sequenced the ITS and two chloroplast genes in 163 species. They concluded that many species are polyphyletic in their origins thus making them difficult to discriminate between one another; thus, requiring additional molecular studies to more accurately define taxonomic relationships.

We tested 51 SSR/INDEL based markers (Table 3), and identified several thousand inter- and intraspecific SNPs (Table 6), all of which have potential as both inter- and intraspecific markers. Of the 51 SSRs/INDELS we selected 12 to test across 93 Penstemon taxa. The resulting data was used to more clearly define the phylogenetic relationships of those taxa but our results were incoherent. It is possible that some of these markers may represent more than one locus in the Penstemon genome. This situation has been identified by others as a potential weakness in using SSR based markers in interspecific phylogenetic studies [46,47]. A major reason for the vagary in Penstemon’s phylogeny is that it appears to have quite recently evolved and rapidly radiated leaving weak species boundaries [13,14]. Furthermore, there are a number of reports of speciation via natural interspecific hybridization found within the genus [14,68-73]. Therefore, like Wolfe et al. [14], we concluded that better marker data sets will be required to reduce present phylogenetic ambiguity.

To gain clearer insights into the relationships of Penstemon it will take carefully designed large scale sequencing studies. There are methods which are showing promise to do such studies economically. One example would be to utilize GR-RSC or similar methods which will sample large quantities of homologous sequence of a genome at ever decreasing costs [18,20,74]. Since our SSR/INDEL, sequence, and SNP data have demonstrated broad applicability across Penstemon it becomes evident that further studies utilizing this same GR-RSC protocol and downstream analysis on additional species would allow broader comparisons of putative genes, repeat elements, SNPs and SSRs, facilitating a much better understanding of the genus. Furthermore, using this technique on carefully selected parents and their segregating progeny would allow Penstemon genetic mapping studies which would greatly enhance the ability to do breeding and domestication studies within the genus. Historically, studies of this nature would have been unthinkable; however, mass homologous loci sequence studies are rapidly becoming feasible [18,20,74]. In the interim it is possible to take the data we report here and further test the 75 SNPs we have reported here along with others not yet developed and for around US$0.05/data point [18,20] do a much broader study. Studies on homologous SNPs across many Penstemon taxa, similar to the Amaranthus study of Maughan et al. [20], should assist in developing improved insights into Penstemon phylogenetic relationships and produce high quality genetic maps from carefully designed segregating Penstemon populations.

Competing interests
The authors declare no competing interests.

Authors’ contributions
Rhyan B Dockter, David B Elzinga, Brad Geary, P Jeff Maughan, Leigh A Johnson, Danika Tumblison, Jana Lynn Franke, Keri Dockter, and Mikel R Stevens. RBD preformed the GR-RSC technique and either carried out or oversaw the all other steps of the study and participated in all planning and design of all experiments as well as their analysis and did the initial drafting of the manuscript. DBE did or assisted in all bioinformatics performed in this study. BG participated in the design of all aspects of the study as well as advised RBD and was involved in the editing and revising of the manuscript. PJM advised and assisted in the GR-RSC technique as well as advised RBD in relevant issues of the bioinformatics of the study and was involved in the editing and revising of the manuscript. LAJ advised and assisted RBD and MRS in the taxonomy related issues of the study and was involved in the editing and revising of the manuscript. DT, JF, and KD carried out all aspects, including basic analysis, of the marker studies reported. MRS was the senior advisor of RBD and was intricately involved in all aspects of the study and the manuscript. All authors both read and approved the final manuscript.

Acknowledgements
We acknowledge Shaun Broderick, a graduate student, Tiffany Austin, and Aaron King, undergraduates, for their laboratory assistance and Robert Byers a graduate student and Scott Yourstone, an undergraduate, for their bioinformatic assistance, all from Brigham Young University. This research was funded in part by an Annaley Naegle Redd Assistantship from the Brigham Young University Charles Redd Center for Western Studies and a
51. Eshed Y, Zamir D. A genomic library of Lycopersicon pennellii in L. esculentum: A tool for fine mapping of genes. Euphytica 1994; 79:175–179.

52. Robbins MD, Masud MAT, Panthee DR, Gardner RG, Francis DM, Stevens MR: Marker assisted selection for coupling phase resistance to Tomato spotted wilt virus and Phytophthora infestans (late blight) in tomato. HortScience 2010, 45:1424–1428.

53. Canady MA, Meglic V, Chetelat RF. A library of Solanum lycopersicoides introgression lines in cultivated tomato. Genome 2005, 48:685–697.

54. Canady MA, Yi Y, Chetelat RF: Homoeologous recombination in Solanum lycopersicoides introgression lines of cultivated tomato. Genetics 2006, 174:1775–1788.

55. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Carinhour S, McCouch S: Computational and experimental analysis of microsatellites in rice (Oryza sativa L): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001, 11:1441–1452.

56. Parida SK, Kalia SK, Kaul S, Dalal V, Hemaprabha G, Selvi A, Pandit A, Singh A, Gaikwad K, Sharma TR, Srivastava PS, Singh NK, Mohapatra T: Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theor Appl Genet 2009, 118:327–338.

57. Zhang FK, Zhao ZM: The influence of neighboring-nucleotide composition on single nucleotide polymorphisms (SNPs) in the mouse genome and its comparison with human SNPs. Genomics 2004, 84:785–790.

58. Morton BR, Bi IV, McMullen MD, Gaut BS. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. Genetics 2006, 172:569–577.

59. Schmutz H, Meister A, Horres R, Bachmann R: Genome size variation among accessions of Arabidopsis thaliana. Ann Bot 2004, 93:317–321.

60. Lynch M: The Origins of Genome Architecture. Sunderland, MA: Sinauer Associates, Inc; 2007.

61. Lynch M, Conery JS: The origins of genome complexity. Science 2003, 302:1401–1404.

62. Kidwell MG: Transposable elements and the evolution of genome size in eukaryotes. Genetica 2002, 115:49–63.

63. Roskinia O, Barber JC, Neu E, Belavyev A: Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res 2008, 120:351–357.

64. Kolano B, Gardunia BW, Michalska M, Bonifacio A, Fairbanks D, Maughan PJ, Coleman CE, Stevens MR, Jellen EN, Maluszynska J: Chromosomal localization of two novel repetitive sequences isolated from the Chenopodium quinoa genome. Genome Res 2011, 21:710–717.

65. Kubis S, Schmidt T, Heslop-Harrison JS: Repetitive DNA elements as a major component of plant genomes. Ann Bot 1998, 82:Suppl A:455–55.

66. Meyers B: A summary of Bruce Meyers’ Penstemon hybridizations. Bul Amer Penstemon Soc 1998, 57:2–11.

67. Friedt W, Snowdon RJ, Ordon F, Atemeyer J: Plant breeding: assessment of genetic diversity in crop plants and its exploitation in breeding. Prog Bot 2007, 68:151–178.

68. Wolfe AD, Elsens WJ: Diploid hybrid speciation in Penstemon (Scrophulariaceae) revisited. Amer J Bot 1993, 80:1082–1094.

69. Wolfe AD, Elsens WJ: Nuclear ribosomal DNA restriction site variation in Penstemon section Peltanthera (Scrophulariaceae): an evaluation of diploid hybrid speciation and evidence for introgression. Amer J Bot 1994, 81:1627–1635.

70. Wolfe AD, Elsens WJ: Evidence of chloroplast capture and pollen-mediated gene flow in Penstemon sect. Peltanthera (Scrophulariaceae). Syst Bot 1995, 20:395–412.

71. Datwyler SL, Wolfe AD: Phylogenetic relationships and morphological evolution in Penstemon subg. Dosanthera (Veronicaceae). Syst Bot 2004, 29:165–176.

72. Wolfe AD, Xiang Q-Y, Kephart SR: Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. Mol Ecol 1998, 7:1107–1125.

73. Wolfe AD, Xiang Q-Y, Kephart SR: Diploid hybrid speciation in Penstemon (Scrophulariaceae). Proc Natl Acad Sci USA 1998, 95:5112–5115.

74. Elshrie R, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 2011, 6:e19379.