



Comparative Mapping of the Wild Perennial *Glycine latifolia* and Soybean (*G. max*) Reveals Extensive Chromosome Rearrangements in the Genus *Glycine*

Sungyul Chang¹, Carrie S. Thurber¹, Patrick J. Brown¹, Glen L. Hartman^{1,2}, Kris N. Lambert¹, Leslie L. Domier^{1,2*}

1 Department of Crop Sciences, University of Illinois, Urbana, Illinois, United States of America, **2** United States Department of Agriculture, Agricultural Research Service, Urbana, Illinois, United States of America

Abstract

Soybean (*Glycine max* L. Mer.), like many cultivated crops, has a relatively narrow genetic base and lacks diversity for some economically important traits. *Glycine latifolia* (Benth.) Newell & Hymowitz, one of the 26 perennial wild *Glycine* species related to soybean in the subgenus *Glycine* Willd., shows high levels of resistance to multiple soybean pathogens and pests including *Alfalfa mosaic virus*, *Heterodera glycines* Ichinohe and *Sclerotinia sclerotiorum* (Lib.) de Bary. However, limited information is available on the genomes of these perennial *Glycine* species. To generate molecular resources for gene mapping and identification, high-density linkage maps were constructed for *G. latifolia* using single nucleotide polymorphism (SNP) markers generated by genotyping by sequencing and evaluated in an F₂ population and confirmed in an F₅ population. In each population, greater than 2,300 SNP markers were selected for analysis and segregated to form 20 large linkage groups. Marker orders were similar in the F₂ and F₅ populations. The relationships between *G. latifolia* linkage groups and *G. max* and common bean (*Phaseolus vulgaris* L.) chromosomes were examined by aligning SNP-containing sequences from *G. latifolia* to the genome sequences of *G. max* and *P. vulgaris*. Twelve of the 20 *G. latifolia* linkage groups were nearly collinear with *G. max* chromosomes. The remaining eight *G. latifolia* linkage groups appeared to be products of multiple interchromosomal translocations relative to *G. max*. Large syntenic blocks also were observed between *G. latifolia* and *P. vulgaris*. These experiments are the first to compare genome organizations among annual and perennial *Glycine* species and common bean. The development of molecular resources for species closely related to *G. max* provides information into the evolution of genomes within the genus *Glycine* and tools to identify genes within perennial wild relatives of cultivated soybean that could be beneficial to soybean production.

Citation: Chang S, Thurber CS, Brown PJ, Hartman GL, Lambert KN, et al. (2014) Comparative Mapping of the Wild Perennial *Glycine latifolia* and Soybean (*G. max*) Reveals Extensive Chromosome Rearrangements in the Genus *Glycine*. PLoS ONE 9(6): e99427. doi:10.1371/journal.pone.0099427

Editor: Matthew N. Nelson, The University of Western Australia, Australia

Received: February 21, 2014; **Accepted:** May 14, 2014; **Published:** June 17, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This study was supported by funding from the National Sclerotinia Initiative and the United States Department of Agriculture-Agricultural Research Service. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: leslie.domier@ars.usda.gov

Introduction

Soybean (*Glycine max* L. Mer.) is a major source of dietary protein and oil in animal production and for human consumption worldwide [1]. With increasing utilization of soybean for animal feed in countries like China, there is added demand for soybean production [2]. Most of the increased demand for soybean products has been met by expanding the land area devoted to soybean production [3]. However, it is not clear if the expansion of soybean production areas alone will be able to keep pace with this growing demand. In addition, the global movement of soybean pathogens and pests and the emergence of new pathogens, as illustrated by the recent identification of soybean aphids (*Aphis glycines* Matsumura), soybean rust (*Phakopsora pachyrhizi* Syd. & P. Syd) and Soybean vein necrosis virus in North America [4–6], necessitates the identification of novel genes that will enable producers to meet the ever increasing demand for soybean production in the face of changing abiotic and biotic stresses.

Because of its narrow genetic base, soybean, like many cultivated crops, lacks diversity found in some of its wild relatives. The genus *Glycine* consists of 28 species split between two subgenera, *Glycine* Willd. and *Soja* (Moench) F. J. Hermann. The subgenus *Soja* contains two annual species, *G. max*, the domesticated species in the genus, and *G. soja* Sieb. & Zucc., both of which are native to Asia. The *Glycine* subgenus contains 26 species, including *G. latifolia* (Benth.) Newell & Hymowitz, that are native to Australia and surrounding islands, and have been shown to possess genes for agronomically valuable traits, such as resistance to *Heterodera glycines* Ichinohe and tolerance to *Sclerotinia sclerotiorum* (Lib.) de Bary. and drought [7–13].

To date however, it has not been possible to utilize genes from the perennial *Glycine* species for soybean improvement even though *G. max* and the perennial *Glycine* species share a relatively recent whole genome duplication that occurred between 5 and 13 million years ago [14,15]. Attempts to hybridize *G. max* and *Glycine* perennials have, with the exception of *G. tomentella* (2n = 78) Hayata, been unsuccessful, even with *in vitro* embryo rescue

[16,17]. Cytogenetic observations have shown aberrant chromosome pairing in F_1 hybrids leading to embryo abortion [18]. Sequence data from *G. latifolia* and *G. tomentella* suggest that differences in intergenic and pericentromeric sequences, including sequences of widely dispersed retrotransposons, have reduced chromosomal pairing during hybridization [19–21].

Advances in molecular biology provide tools to circumvent the genetic barriers to capturing the biological diversity present in the perennial relatives of *G. max*. In addition to the genome sequence of *G. max* [15], the genome sequence of the wild annual species *G. soja* was recently determined [22], and shed light on similarities and differences between the two interfertile species. Using the genomic information, high-throughput sequencing and virus-based gene silencing techniques, multiple genes have been identified in soybean [23–26]. Even though gene mapping resources developed for *G. max* have not been directly useful in perennial *Glycine* species, the identification of large syntenic blocks between *G. max* and other legume species [27–30] suggests that high levels of synteny will be observed between annual and perennial *Glycine* species. The development of methods for cost-efficient discovery and mapping of single nucleotide polymorphism (SNP) markers through methods like genotyping by sequencing (GBS) [31,32] have made it possible to fine map genes in plant species for which genetic resources are lacking, as is the case with the perennial *Glycine* species. Here, we describe the construction of high-density linkage maps for a perennial relative, *G. latifolia*, of cultivated soybean and compare the orders of mapped SNP markers to their positions in the genome sequences of *G. max* and *Phaseolus vulgaris* L.

Materials and Methods

Plant materials

Reciprocal crosses were performed between *G. latifolia* plant introduction (PI) 559298 and PI 559300 (obtained from the USDA Soybean Germplasm Collection in Urbana, Illinois; <http://www.ars-grin.gov/npgs/urbana.html>) as previously described [19]. Populations were advanced by selfing the F_2 generation to the F_5 generation.

GBS mapping

DNA was extracted from leaf tissue of PI 559298, PI 559300, 146 F_2 plants and 89 F_5 plants using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA). DNA samples were digested with *Bfa*I and *Pst*I-HF restriction enzymes (New England Biolabs, Ipswich, MA) as described by Thurber *et al.* [33]. For these experiments, *Bfa*I was selected because it did not produce strong banding patterns in preliminary restriction enzyme digestions of *G. latifolia* DNA and *Pst*I was selected because *G. latifolia* sequence data [19] contained an intermediate number of *Pst*I recognition sites. For example, the previously determined *G. latifolia* sequence data were predicted to contain 2.1×10^4 *Mlu*I sites, 8.9×10^4 *Pst*I sites and 3.1×10^5 *Hind*III sites. Up to 96 samples were sequenced per lane of a HiSeq2000 (Illumina Inc., San Diego, CA) at the W. M. Keck Center at the University of Illinois, Urbana, IL, USA to produced 100-nt single-end reads. In both experiments, DNA from each of the parental lines was independently processed twice to serve as a control for SNP identification. The barcode splitter from TASSEL [34] was used to assign sequence reads to individual lines and remove barcode sequences, which produced 90-nucleotide sequence reads that were analyzed for SNPs. The parsed sequence data for the F_2 and F_5 populations have been deposited in the NCBI Short Read Archive as part of project SRP013346. Next, three Perl scripts were used to analyze the sequence reads for the bi-parental

populations. First, sequence reads for each individual/line in the F_2 and F_5 populations and from the parental lines, PI 559298 and PI 559300, were aligned using Bowtie [35] to a *G. latifolia* pseudo-reference sequence, which was generated by sequencing 180-bp, 500-bp paired-end and 3-kb, 8-kb, and 15-kb mate-pair libraries prepared from *G. latifolia* PI 559298 DNA, sequenced on an Illumina HiSeq 2000, and *de novo* assembled using ALLPATHS-LG [36] (Chang *et al.*, manuscript in preparation). The resulting assembly contained 16,423 scaffolds representing 1,069 Mbp, with an N50 of 235 Kb. Bowtie2 [37], which allows for insertions and deletions (indels), was also evaluated for read mapping, but at the high stringencies for matching employed, few indels were detected and the output from Bowtie was parsed more directly to SNP calls than output from Bowtie2. Second, SNPs were called when at least three reads from both replications of PI 559298 differed from both replications of PI 559300. Finally, Bowtie output files for each individual/line were used to assess allelic frequencies for each SNP using a custom Perl script, which ignored SNPs in low quality sequence reads (average quality scores of 40 or less). Based on allelic frequencies at each locus for each line, the Perl script then created a genotype matrix file for linkage analysis. Markers with less than 30% missing data and whose segregation did not differ significantly ($P > 0.05$) from expected segregation ratios were selected for *de novo* linkage map construction.

Linkage maps were constructed using MSTMap [38] with a P -value = 1.0^{-9} , and visualized using MapDraw [39]. Consensus linkage maps were constructed for *G. latifolia* from the F_2 and F_5 data using MergeMap [40]. A weight of 5.0 was assigned to the F_5 linkage maps and a weight of 1.0 to F_2 linkage maps to reflect the higher confidence in the quality of the maps because of the reduced potential for errors in calling of heterozygous genotypes in the F_5 population relative to the F_2 population. To assess the synteny between *G. latifolia* linkage groups (LGs) and *G. max* chromosomes, SNP-containing sequences from *G. latifolia* were aligned to the *G. max* genome sequence [15] using BLAST [41]. For comparisons with *P. vulgaris* chromosomes, *G. latifolia* SNP-containing sequences and *G. max* gene models [15] were aligned to *P. vulgaris* chromosomes (<http://www.phytozome.net/commonbean.php>) using BLAST and visualized with MizBee [42]. When *G. latifolia* sequences aligned at more than one location, the most syntenic location was chosen for these analyses.

Results

Mapping GBS SNP markers in *G. latifolia* populations

Genotyping by sequencing of the F_2 population produced a total of 4.00×10^8 100-nt reads, of which 1.70×10^8 passed all quality controls and uniquely aligned to PI 559300 sequences. After barcodes were removed, 90 nt were used for SNP discovery. In the F_2 population, 5,160 markers could be reliably scored between the parental lines PI 559298 and PI 559300. Linkage maps constructed from that initial data set represented over 13,000 centimorgans (cM), which was significantly larger than *G. max* (2,296 to 2,550 cM) and previous *G. latifolia* (1972 cM) linkage maps [19,43–46] and likely resulted from errors in calling heterozygous genotypes because of low coverage at some loci. The data set was reprocessed to exclude markers with more than 30% missing data and with segregation ratios that differed significantly from 1:2:1 ($P > 0.05$), which resulted in 2,377 markers (Table S1). The average depth of coverage for the selected SNPs was 32 reads per locus and ranged from 0 to 270 reads. The markers formed 20 large LGs (Figure S1), with an average of 119 markers per LG (Table 1), and a total length of 2,305 cM. To confirm marker orders, an F_5 population was analyzed by the

Table 1. Comparison of the numbers of SNP markers in F₂, F₅, merged and Chang *et al.* [19] maps.

Linkage Group	Markers				Length (cM)			
	F ₂	F ₅	Merged	Chang <i>et al.</i> [19]	F ₂	F ₅	Merged	Chang <i>et al.</i> [19]
1	123	185	206	17	116.7	92.2	139.8	117.44
2	152	192	229	15	142.6	162.6	220.4	196.71
3	87	127	157	10	91.4	103.6	137.7	46.96
4	107	138	158	13	121.0	95.8	153.4	124.51
5	139	155	202	8	137.5	94.3	180.3	62.18
6	157	207	240	13	159.5	119.4	183.1	112.52
7	80	128	144	15	85.4	93.8	129.7	93.96
8	150	175	214	13	129.9	141.8	175.2	151.24
9	126	175	201	14	129.6	118.4	153.7	94.25
10	102	149	174	11	112.7	132.0	171.7	98.99
11	106	128	164	13	114.4	104.0	132.3	103.77
12	94	103	130	16	91.9	100.9	118.4	116.93
13	115	169	193	15	96.9	94.0	130.6	156.41
14	85	75	125	12	79.1	39.5	95.8	51.73
15	114	164	185	12	104.4	98.3	125.9	103.64
16	119	155	182	9	102.8	95.9	144.7	64.38
17	136	167	202	11	126.9	90.6	147.5	67.76
18	130	177	199	19	124.4	98.8	156.1	90.93
19	131	179	214	11	121.8	94.4	147.7	44.11
20	124	162	191	11	116.3	136.5	170.3	73.8
Total	2377	3110	3710	258	2305.2	2106.8	3014.3	1972.2

doi:10.1371/journal.pone.0099427.t001

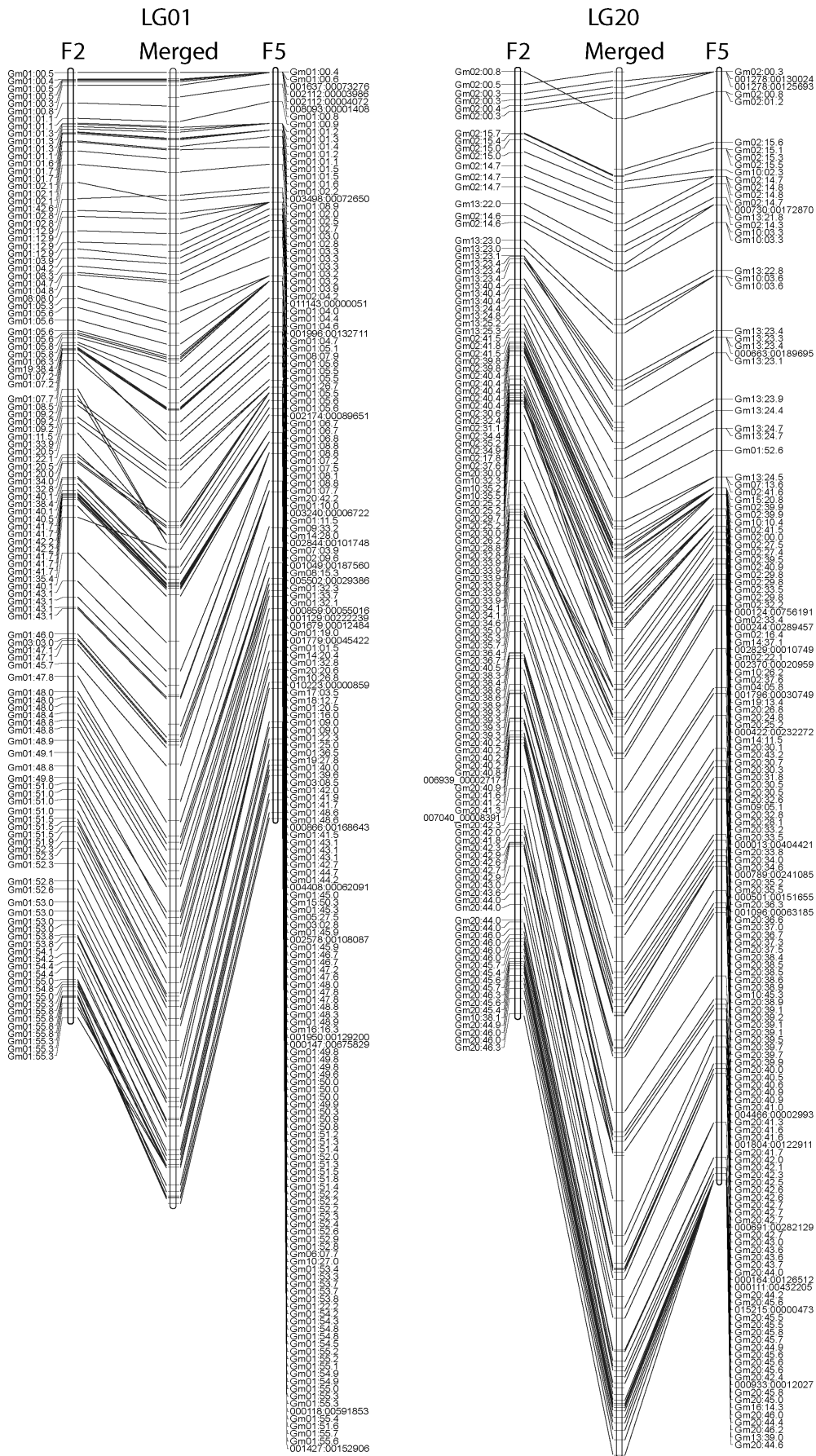


Figure 1. Comparison of F₂, F₅ and merged linkage maps for GBS SNP markers for *Glycine latifolia* linkage groups 1 and 20. Orders of SNP markers were very similar between the F₂ and F₅ populations. In some cases, markers that segregated in the F₂ population co-localized in the F₅ population, which may have resulted from errors in calling heterozygous loci in the F₂ population. While linkage group 1 showed a high level of collinearity with *G. max* chromosome 1, linkage group 20 had regions of collinearity with multiple *G. max* chromosomes. Even so, there was good agreement in marker order between the F₂ and F₅ populations for linkage group 20. Markers were named for the *G. max* chromosome and the nucleotide position on the chromosome ($\times 10^{-6}$) to which the SNP-containing sequences aligned. Markers that did not align to a *G. max* chromosome were named for the *G. latifolia* scaffold containing the SNP.
doi:10.1371/journal.pone.0099427.g001

same procedures. The analysis produced a total of 1.92×10^8 100-nt reads, of which 1.05×10^8 passed all quality controls and uniquely aligned to PI 559300 sequences. The data produced 7,081 SNPs between the parental lines, from which 3,110 GBS markers (Table S2) were selected using similar criteria and analyzed in an F₅ population. Average depth of coverage for the selected SNPs was 21 reads and ranged from 0 to 264 reads. As with the F₂ population, most of the markers formed 20 large LGs (Figure S2), with an average of 155 markers per LG and a total map length of 3,110 cM. A total of 1,777 markers were shared between the two populations with 600 markers unique to the F₂ population and 1,333 markers unique to the F₅ population. The orders of shared markers were very similar in linkage maps constructed from the F₂ and F₅ populations (Figure 1). In some cases, markers that appeared to segregate in the F₂ population did not segregate in the F₅, presumably caused by errors in calling heterozygous loci in the F₂ population. The shared markers were used as a framework to construct consensus linkage maps for *G. latifolia* (Figures 1 & S3). The merged consensus maps contained 3,710 markers (Table 1).

Synteny between *G. latifolia* linkage groups and *G. max* chromosomes

Because little information is available on *G. latifolia* chromosomes, or chromosomes of any other perennial *Glycine* species, *G. latifolia* LGs were named for the *G. max* chromosomes to which *G. latifolia* SNP-containing sequences predominantly aligned. When mapped orders of *G. latifolia* SNPs were compared to positions of their sequences in the *G. max* genome, *G. latifolia* LGs 1, 3, 4, 6, 9, 10, 11, 12, 14, 15, 17, and 18 showed a high degree of collinearity with the corresponding *G. max* chromosomes and no interchromosomal rearrangements (Figure 2). In contrast, the remaining eight chromosomes each had at least one interchromosomal rearrangement relative to *G. max* (LG 13: one rearrangement; LG16: two rearrangements; LGs 2, 5, and 8: three rearrangements; LGs 7 and 20: four rearrangements and LG 19: five rearrangements). For example, *G. latifolia* LG 13 contained regions syntenic with *G. max* chromosomes 13 and 20 and *G. latifolia* LG 2 contained regions syntenic to *G. max* chromosomes 1, 2, 8, 13 and 19 (Figure 2). *Glycine latifolia* chromosome 8 appeared to be the product of two translocations between *G. max* chromosomes 2 and 8. Similarly, *G. latifolia* LG 20 appeared to contain a reciprocal translocation between *G. max* chromosomes 2 and 20. Syntenic blocks in rearranged LGs corresponded to between 0.3 Mb and 30 Mb (between *G. latifolia* LG 8 and *G. max* chromosome 8) in the *G. max* genome with an average of 6.9 Mb. Even though we described the structure of *G. latifolia* linkage groups as products for rearrangements relative to *G. max*, the structures of the ancestral chromosomes is not known. Hence, in some cases, *G. latifolia* linkage groups may have under gone fewer rearrangements than *G. max* chromosomes. Singleton markers (single *G. latifolia* SNP markers that aligned to a *G. max* chromosome without at least a second proximal collinear marker) were ignored for these analyses.

As with molecular markers in *G. max*, comparison of the genetic distances between GBS markers in *G. latifolia* and the physical

distances between positions to which the SNP-containing sequences aligned on *G. max* chromosomes, indicated that there was reduced recombination in regions that corresponded to *G. latifolia* centromeres (Figure 3). Points deviating from the main line may represent mis-aligned, or mis-mapped sequences or intrachromosomal rearrangements. No *G. latifolia* markers were identified that aligned to 13.7 Mb and 15.0 Mb in the central regions of *G. max* chromosomes 5 and 20, respectively, which may have been caused by low GBS marker density (i.e., lack of *Pst*I cutting sites) or low sequence conservation in highly repetitive pericentromeric regions.

Similarities between *G. latifolia* linkage groups and *Phaseolus vulgaris* chromosomes

The progenitor of *P. vulgaris* diverged from the *Glycine* about 19 million years ago [47,48]. While 3,664 SNP-containing sequences from *G. latifolia* aligned to the *G. max* genome, 2,063 *G. latifolia* sequences aligned to *P. vulgaris* chromosomes. For *G. max*, the sequences of 30,327 gene models aligned to *P. vulgaris* chromosomes. As reported for comparisons between *G. max* and *P. vulgaris* [49], extensive blocks of synteny were observed between *G. latifolia* LGs and *P. vulgaris* chromosomes (Figure 4). McClean *et al.* observed that the pericentromeric regions of *G. max* chromosomes 10, 12, 14, 17, 18, and 20 had extensive syntenic blocks with *P. vulgaris* chromosomes. *Glycine latifolia* LGs 7 and 20 appeared to have larger syntenic blocks with single *P. vulgaris* chromosomes than with *G. max* chromosomes. Both *G. latifolia* LGs 10 and 20 showed large blocks of synteny with *P. vulgaris* chromosome 7, as did *G. max* chromosome 10. The results suggest that *G. max* chromosome 7 and 20 have been reorganized after the whole genome duplication event, but *G. max* chromosome 10 and *G. latifolia* LGs 7, 10 and 20 appear to have retained gene orders more similar to shared ancestral chromosomes.

Discussion

Glycine latifolia is a perennial relative of soybean that is native to eastern Australia with a trailing or twining growth habit [50]. Like *G. max*, the genome of *G. latifolia* contains $2n = 40$ morphologically similar chromosomes [50]. In this study, we used GBS to construct high-density linkage maps for *G. latifolia* and showed that eight of the 20 *G. latifolia* LGs were rearranged relative to *G. max* chromosomes. Linkage maps were constructed *de novo* from F₂ and F₅ populations to confirm marker orders. Genotyping by sequencing has been applied to several plant species with and without reference genome sequences. For example, Sonah *et al.* [51] reported the application of GBS to a set of eight diverse soybean genotypes and identified from 4,028 to 5,807 SNPs between pairs of soybean lines by aligning sequence reads to the *G. max* genome sequence. Similar to results reported here, Ward *et al.* [52] produced genetic maps for raspberry (*Rubus idaeus* L.) without a reference sequence consisting of 2,391 and 4,521 markers. Russell *et al.* [53] used 1,901 SNPs identified using GBS without a reference sequence to map quantitative trait loci in blackcurrant (*Ribes nigrum* L.). Similar to this study, Ma *et al.* [54] identified 3,745 SNP by GBS in *Miscanthus sinensis* Anderss, a potential

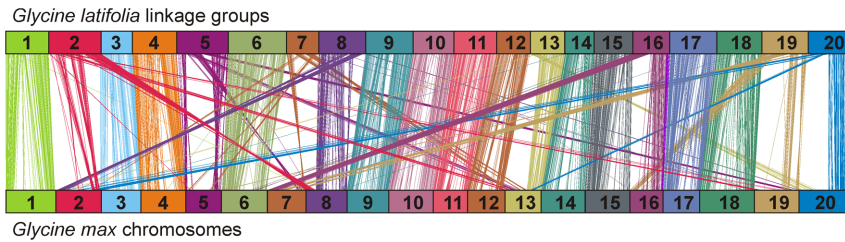


Figure 2. Synteny between *Glycine latifolia* linkage groups and *G. max* chromosomes. Sequences containing mapped *G. latifolia* SNPs were aligned to the *G. max* genome sequence. *Glycine latifolia* linkage groups (top) and physical maps for each *G. max* chromosome (bottom) are displayed as linear arrays. Vertical and diagonal lines connect genetic and physical locations of SNP markers between the two species. doi:10.1371/journal.pone.0099427.g002

bioenergy crop [55], and used them to compare the genome organization of *M. sinensis* to those of *Brachypodium distachyon* L., *Oryza sativa* L., *Sorghum bicolor* (L.) Moench, and *Zea mays* L. by aligning the SNP-containing sequences to the heterologous genome sequences.

Mahama et al. [56] genetically and cytologically identified seven different chromosomal translocation lines in soybean and reported that crossing a soybean line homozygous for a single translocation

with a wild-type soybean line resulted in significant levels of pollen abortion, ovule abortion, and reduction in seed set. Hence, it is not surprising that crosses between *G. latifolia* and *G. max* that involve eight chromosomes with multiple translocations do not produce fertile progeny [57]. Following embryo rescue and colchicine treatment to double chromosome numbers [58], it may be possible to recover lines that retain unrearranged *G. latifolia* chromosomes from wide-hybridization experiments between *G. latifolia* and *G. max*, but lines containing chromosomes with multiple translocations would likely not be fertile.

Using fluorescence *in situ* hybridization-based karyotyping of the seven soybean translocation lines, Findley et al. [59] identified reciprocal translocations between *G. max* chromosomes 1 and 8; 2 and 8; 2 and 11; 4 and 13; and 5 and 13. It is interesting that *G. latifolia* LGs also showed reciprocal exchanges between chromosomes 2 and 8, which may indicate the presence of recombination hotspots in the rearranged chromosomes. It has been difficult to identify translocations in soybean because of its small and morphologically similar chromosomes [60]. Consequently, it is possible that other recombination hotspots remain to be identified.

Based on hybridization success, hybrid seed viability, fertility of F_1 plants in intra- and interspecific hybrids and degree of meiotic chromosome pairing, species within the genus *Glycine* have been assigned genome types [61]. *Glycine latifolia*, along with *G. microphylla* Tindale, and *G. tabacina* (Labill.), Benth (all $2n = 40$) contain B genome types. In interspecific crosses, *G. latifolia*, *G. microphylla* and *G. tabacina* produce vigorous F_1 plants with normal seed set [61–64], which suggests that *G. microphylla* and *G. tabacina*, other than paracentric inversions [65], have chromosome structures very similar to those found in *G. latifolia*. In contrast, F_1 plants from crosses between D-genome perennial species (i.e., *G. tomentella*) show seedling lethality [50,63]. It has been possible to recover plants from crosses between *G. tomentella* and *G. max*, using *G. tomentella* lines with 78 chromosomes [16,17]. The $2n = 78$ *G. tomentella* lines may contain a set of 20 chromosomes that have fewer rearrangements than *G. latifolia* that facilitate chromosome pairing during crossing with *G. max*.

Generally, nuclear genomes of closely related species show high degrees of collinearity that degrade with increasing phylogenetic distance, but the rates at which chromosomes diverge vary widely among taxa [66]. In comparisons of genetic maps between *Arabidopsis thaliana* (L.) Heynh. and *Brassica nigra* (L.) W.D.J. Koch, Lagercrantz [67] estimated that there had been about three chromosomal rearrangements per million years since the two species had diverged 11 to 35 million years ago [68]. In contrast, Koch et al. [69] estimated that *Arabidopsis lyrata* L. and *Capsella rubella* Reut. had undergone less than 0.09 rearrangements per million years since the two lineages diverged 10 to 14 million years ago. Like *G. tomentella*, *G. latifolia* diverged from the progenitors of

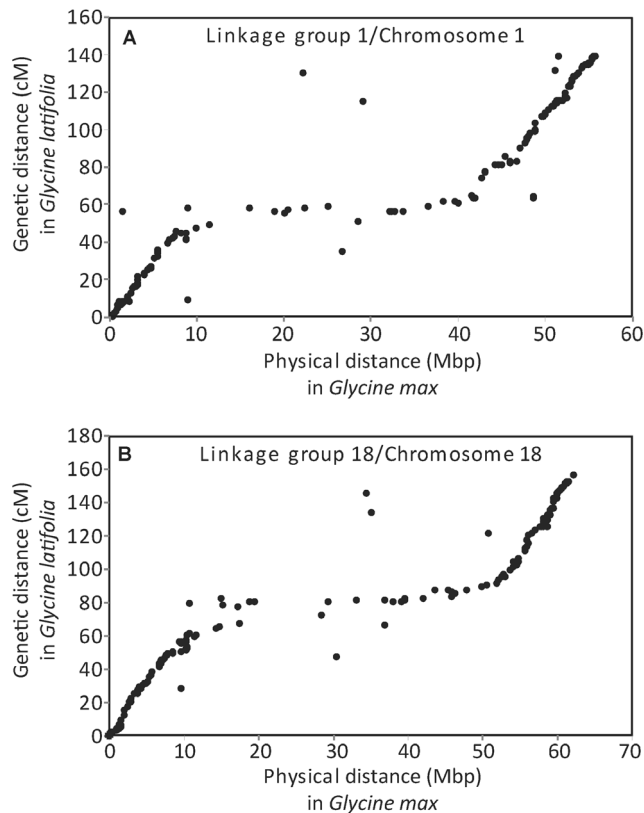


Figure 3. Comparison of genetic distances in *Glycine latifolia* to physical distances in *G. max* for linkage groups/chromosomes 1 and 18. *Glycine latifolia* linkage groups 1 and 18 showed a high degree of collinearity with the corresponding *G. max* chromosomes. The genetic distances in *G. latifolia* were plotted against the physical locations of the SNP markers on the *G. max* chromosomes 1 and 18. As in *G. max*, the predicted ratios of genetic and physical distances varied along *G. latifolia* linkage groups. The slopes were steeper near the ends of linkage groups and flatter near the center in regions predicted to correspond to centromeres, where recombination is lower. doi:10.1371/journal.pone.0099427.g003

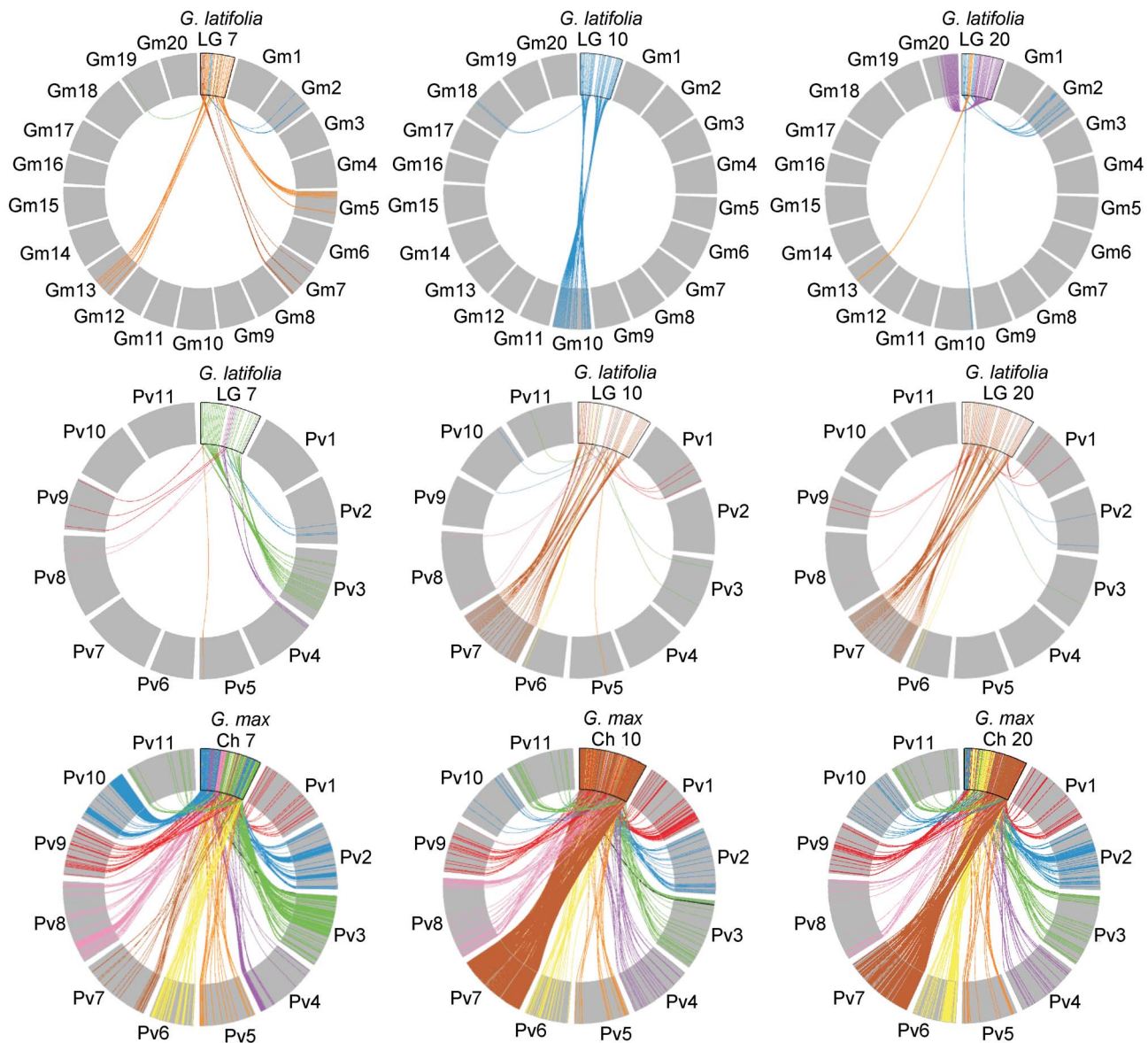


Figure 4. Comparison of synteny of individual *Glycine latifolia* linkage groups 7, 10 and 20 with *G. max* and *Phaseolus vulgaris* chromosomes. *Glycine latifolia* linkage groups (A & B) and *G. max* chromosomes (C) are placed at the top of each circle with colored lines connecting positions of *G. latifolia* SNP markers (A & B) or *G. max* gene model (C) sequences to positions in *G. max* (Gm01 – Gm20) or *P. vulgaris* (Pv01 – Pv11) chromosomes, represented by gray boxes. *Glycine latifolia* linkage groups 7 and 20 appeared to have larger syntenic blocks with single *P. vulgaris* chromosomes than *G. max* chromosomes. Both *G. latifolia* linkage groups 10 and 20 showed large blocks of synteny with *P. vulgaris* chromosome 7, as did *G. max* chromosome 10. Synteny maps were constructed using MizBee [42]. doi:10.1371/journal.pone.0099427.g004

G. max between 5 and 7 million years ago [20,70], which would mean that the genomes of *G. latifolia* and *G. max* have undergone up to five interchromosomal rearrangements per million years. This number is higher than in the studies mentioned above and may indicate a higher rate of chromosomal instability or simply that a higher density of markers was used which permitted the detection of a larger number of chromosomal rearrangements. In addition to being geographically isolated from the progenitors of *G. max*, *G. latifolia*, like other $2n = 40$ perennial *Glycine* species, often produces seed from cleistogamous flowers [60], which reduces its opportunities for outcrossing and possibly removing some of the selection pressure against chromosomal rearrangements. As a consequence, more translocations may have been preserved in

Glycine species than in other non-*Glycine* species examined that outcross more readily.

As functions are assigned to *G. max* genes, the perennial *Glycine* species become important sources of new alleles that can be isolated and moved into *G. max* by standard transformation or developing gene replacement technologies [71–73]. Jones et al. [74] recently demonstrated the feasibility of this approach by using *Agrobacterium*-mediated transformation to transfer a functional gene for resistance to late blight (caused by *Phytophthora infestans* (Mont.) DeBary) from *Solanum venturii* Hawkes & Hjerting to cultivated potato (*Solanum tuberosum* L.). Gene identification in perennial *Glycines* species would be greatly aided by determining the genomic sequences for at least a subset of the species. Because many of the

chromosomes of perennial *Glycine* species are collinear with their *G. max* homologues, the *G. max* genome sequence could be used as a reference to assemble genome sequences of perennial *Glycine* species. Using the heterologous *G. max* reference sequence will reduce the depth of sequence coverage needed for genome assembly compared to *de novo* genome sequencing [75,76]. Even though genetic hybridization may not be possible by standard means, methods for high-throughput gene mapping and identification afforded by next-generation sequencing provide tools to capture the variation present in the wild species.

Supporting Information

Figure S1 Genetic linkage maps of GBS markers in a *Glycine latifolia* F₂ population. (TIF)

Figure S2 Genetic linkage maps of GBS markers in a *Glycine latifolia* F₅ population. (TIF)

Figure S3 Consensus genetic linkage maps for *Glycine latifolia* derived from 3710 GBS markers. (TIF)

References

- Hartman GL, West ED, Herman TK (2011) Crops that feed the World 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. *Food Security* 3: 5–17.
- Simelton E (2011) Food self-sufficiency and natural hazards in China. *Food Security* 3: 35–52.
- Brown LR (2012) Full Planet, Empty Plates: The new geopolitics of food scarcity. New York, New York: W. W. Norton & Company.
- Schneider RW, Hollier CA, Whitam HK, Palm ME, McKemy JM, et al. (2005) First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. *Plant Disease* 89: 774–774.
- Ragsdale DW, Voegtlin DJ, O'Neil RJ (2004) Soybean aphid biology in North America. *Annals of the Entomological Society of America* 97: 204–208.
- Zhou J, Kantartzi SK, Wen RH, Newman M, Hajimorad MR, et al. (2011) Molecular characterization of a new tospovirus infecting soybean. *Virus Genes* 43: 289–295.
- Lim SM, Hymowitz T (1987) Reactions of perennial wild species of genus *Glycine* to *Septoria glycines*. *Plant Disease* 71: 891–893.
- Riggs RD, Wang S, Singh RJ, Hymowitz T (1998) Possible transfer of resistance to *Heterodera glycines* from *Glycine tomentella* to soybean. *Journal of Nematology* 30: 547–552.
- Hartman GL, Gardner ME, Hymowitz T, Naidoo GC (2000) Evaluation of perennial *Glycine* species for resistance to soybean fungal pathogens that cause Sclerotinia stem rot and sudden death syndrome. *Crop Science* 40: 545–549.
- Hartman GL, Wang TC, Hymowitz T (1992) Sources of resistance to soybean rust in perennial *Glycine* species. *Plant Disease* 76: 396–399.
- Burdon JJ, Marshall DR (1981) Evaluation of Australian native species of *Glycine* for resistance to soybean rust. *Plant Disease* 65: 44–45.
- Horlock CM, Teakle DS, Jones RM (1997) Natural infection of the native pasture legume, *Glycine latifolia*, by alfalfa mosaic virus in Queensland. *Australasian Plant Pathology* 26: 115–116.
- Jones RM, Brown AHD, Coote JN (1996) Variation in growth and forage quality of *Glycine latifolia* (Benth.) (Newell and Hymowitz). *Genetic Resources Communication* No. 26. Australia: Division of Tropical Crops and Pastures, CSIRO.
- Doyle JJ, Egan AN (2010) Dating the origins of polyploidy events. *New Phytologist* 186: 73–85.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
- Singh RJ, Kollipara KP, Hymowitz T (1998) Monosomic alien addition lines derived from *Glycine max* (L) Merr and *G. tomentella* Hayata: production, characterization, and breeding behavior. *Crop Science* 38: 1483–1489.
- Singh RJ (2010) Methods for producing fertile crosses between wild and domestic soybean species. In: USPO, editor. USA.
- Hymowitz T, Singh RJ, Kollipara KP (1998) The genomes of the *Glycine*. *Plant Breeding Reviews* 16: 289–231.
- Chang S, Hartman GL, Singh RJ, Lambert KN, Hobbs HA, et al. (2013) Identification of high-quality single-nucleotide polymorphisms in *Glycine latifolia* using a heterologous reference genome sequence. *Theoretical and Applied Genetics* 126: 1627–1638.
- Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, et al. (2008) Differential accumulation of retroelements and diversification of NB-LRR

Table S1 GBS SNP Genotype data for the F₂ mapping population. (DOCX)

Table S2 GBS SNP Genotype data for the F₅ mapping population. (DOCX)

Acknowledgments

We thank Nancy McCoppin for assistance in maintenance of mapping populations. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the United States Department of Agriculture or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Author Contributions

Conceived and designed the experiments: SC PJB GLH LLD. Performed the experiments: SC CST GLH. Analyzed the data: SC LLD PJB. Contributed reagents/materials/analysis tools: SC CST PJB GLH KNL LLD. Wrote the paper: SC LLD GLH PJB CST KNL.

- disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. *Plant Physiology* 148: 1740–1759.
- Wawrzynski A, Ashfield T, Chen NWG, Mammadov J, Nguyen A, et al. (2008) Replication of nonautonomous retroelements in soybean appears to be both recent and common. *Plant Physiology* 148: 1760–1771.
- Kim MY, Lee S, Van K, Kim TH, Jeong SC, et al. (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences of the United States of America* 107: 22032–22037.
- Meyer JDF, Silva DCG, Yang C, Pedley KF, Zhang C, et al. (2009) Identification and analyses of candidate genes for *Rpp4*-mediated resistance to Asian soybean rust in soybean. *Plant Physiology* 150: 295–307.
- Cook DE, Lee TG, Guo XL, Melito S, Wang K, et al. (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338: 1206–1209.
- Liu SM, Kandath PK, Warren SD, Yeckel G, Heinz R, et al. (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492: 256–+.
- Xia ZJ, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, et al. (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proceedings of the National Academy of Sciences of the United States of America* 109: E2155–E2164.
- Boutin SR, Young ND, Olson TC, Yu ZH, Shoemaker RC, et al. (1995) Genome conservation among three legume genera detected with DNA markers. *Genome* 38: 928–937.
- Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. *Theoretical and Applied Genetics* 103: 765–773.
- Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, et al. (2009) A consensus genetic map of cowpea *Vigna unguiculata* (L) Walp. and synteny based on EST-derived SNPs. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18159–18164.
- Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, et al. (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520–524.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* 6: e19379.
- Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America* 108: 6893–6898.
- Thurber CS, Ma JM, Higgins RH, Brown PJ (2013) Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biology* 14: 12.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.

35. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
36. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1513–1518.
37. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.
38. Wu YH, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLOS Genetics* 4: e1000212.
39. Liu RH, Meng JL (2003) MapDraw: a Microsoft Excel macro for drawing genetic linkage maps based on given genetic linkage data. *Hereditas* 25: 317–321.
40. Wu Y, Close TJ, Lonardi S (2008) On the accurate construction of consensus genetic maps. *Comput Syst Bioinformatics Conf* 7: 285–296.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
42. Meyer M, Munzner T, Pfister H (2009) MizBee: a multiscale synteny browser. *Iccv Transactions on Visualization and Computer Graphics* 15: 897–904.
43. Choi IY, Hyten DL, Matukumalli LK, Song QJ, Chaky JM, et al. (2007) A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics* 176: 685–696.
44. Hyten DL, Choi IY, Song QJ, Specht JE, Carter TE, et al. (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. *Crop Science* 50: 960–968.
45. Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, et al. (2004) A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics* 109: 122–128.
46. Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, et al. (1999) An integrated genetic linkage map of the soybean genome. *Crop Science* 39: 1464–1490.
47. Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* 54: 575–594.
48. Stefanović S, Pfeil BE, Palmer JD, Doyle JJ (2009) Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. *Systematic Botany* 34: 115–128.
49. McClean PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11: 184 doi:10.1186/1471-2164-1111-1184.
50. Newell CA, Hymowitz T (1980) A taxonomic revision in the genus *Glycine* subgenus *Glycine* (Leguminosae). *Brittonia* 32: 63–69.
51. Sonah H, Bastien M, Iquira E, Tardivel A, Legare G, et al. (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLOS ONE* 8: e54603.
52. Ward JA, Bhangoo J, Fernandez-Fernandez F, Moore P, Swanson JD, et al. (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* 14: 2 doi:10.1186/1471-2164-14-2.
53. Russell J, Hackett C, Hedley P, Liu H, Milne L, et al. (2013) The use of genotyping by sequencing in blackcurrant (*Ribes nigrum*): developing high-resolution linkage maps in species without reference genome sequences. *Molecular Breeding* doi: 10.1007/s11032-11013-19996-11038.
54. Ma XF, Jensen E, Alexandrov N, Troukhan M, Zhang LP, et al. (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLOS ONE* 7: e33821.
55. Graham-Rowe D (2011) Agriculture: Beyond food versus fuel. *Nature* 474: S6–S8.
56. Mahama AA, Deaderick LM, Sadanaga K, Newhouse KE, Palmer RG (1999) Cytogenetic analysis of translocations in soybean. *Journal of Heredity* 90: 648–653.
57. Chung GH, Kim KS (1991) Obtaining intersubgeneric hybridization between *Glycine max* and *Glycine latifolia* through embryo culture. *Korean Journal of Plant Tissue Culture* 18: 39–45.
58. Singh RJ, Kollipara KP, Hymowitz T (1993) Backcross (BC2-BC4)-derived fertile plants from *Glycine max* and *Glycine tomentella* intersubgeneric hybrids. *Crop Science* 33: 1002–1007.
59. Findley SD, Cannon S, Varala K, Du JC, Ma JX, et al. (2010) A fluorescence *in situ* hybridization system for karyotyping soybean. *Genetics* 185: 727–744.
60. Hymowitz T (2004) Speciation and cytogenetics. In: Boerma HR, Specht JE, editors. *oybeans: Improvement, Production, and Uses*. Third ed. Madison, WI: ASA-CSSA-SSSA. pp. 97–136.
61. Singh RJ, Hymowitz T (1985) Intra- and interspecific hybridization in the genus *Glycine*, subgenus *Glycine* Willd.: chromosome-pairing and genome relationships. *Zeitschrift für Pflanzenzüchtung* 95: 289–310.
62. Singh RJ, Hymowitz T (1985) An intersubgeneric hybrid between *Glycine tomentella* Hayata and the soybean, *Glycine max* (L.) Merr. *Euphytica* 34: 187–192.
63. Singh RJ, Kollipara KP, Hymowitz T (1988) Further data on the genomic relationships among wild perennial species (2n = 40) of the genus *Glycine* Willd. *Genome* 30: 166–176.
64. Doyle JJ, Doyle JL, Grace JP, Brown AHD (1990) Reproductively isolated polyploid races of *Glycine tabacina* (Leguminosae) had different chloroplast genome donors. *Systematic Botany* 15: 173–181.
65. Singh RJ, Hymowitz T (1985) The genomic relationships among six wild perennial species of the genus *Glycine* subgenus *Glycine* Willd. *Theoretical and Applied Genetics* 71: 221–230.
66. Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. *American Journal of Botany* 91: 1709–1725.
67. Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150: 1217–1228.
68. Muller J (1981) Fossil pollen records of extant angiosperms. *Botanical Review* 47: 1–&.
69. Koch MA, Kiefer M (2005) Genome evolution among cruciferous plants: A lecture from the comparison of the genetic maps of three diploid species - *Capsella rubella*, *Arabidopsis lyrata* subsp *Petraea*, and *A. thaliana*. *American Journal of Botany* 92: 761–767.
70. Gill N, Findley S, Walling JG, Hans C, Ma JX, et al. (2009) Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiology* 151: 1167–1174.
71. Dinkins RD, Collins GB (2008) *Agrobacterium*-mediated genetic transformation of soybean. In: Kirti PD, editor. *Handbook of New Technologies for Genetic Improvement of Legumes*. Boca Raton, Florida: CRC Press. pp. 89–102.
72. Rech EL, Vianna GR, Aragao EJM (2008) High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nature Protocols* 3: 410–418.
73. Weinthal DM, Taylor RA, Tzfira T (2013) Nonhomologous end joining-mediated gene replacement in plant cells. *Plant Physiology* 162: 390–400.
74. Jones JDG, Witek K, Verweij W, Jupe F, Cooke D, et al. (2014) Elevating crop disease resistance with cloned genes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.
75. Kim J, Larkin DM, Cai QL, Asan, Zhang YF, et al. (2013) Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1785–1790.
76. Schneeberger K, Ossowski S, Ott F, Klein JD, Wang X, et al. (2011) Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proceedings of the National Academy of Sciences of the United States of America* 108: 10249–10254.