

2008

Soybean Genomics: Developments through the Use of Cultivar “Forrest”

David A. Lightfoot

Southern Illinois University Carbondale, ga4082@siu.edu

Follow this and additional works at: http://opensiuc.lib.siu.edu/psas_articles

Published in *International Journal of Plant Genomics*, doi:10.1155/2008/793158

Recommended Citation

Lightfoot, David A. "Soybean Genomics: Developments through the Use of Cultivar “Forrest”." (Jan 2008).

This Article is brought to you for free and open access by the Department of Plant, Soil, and Agricultural Systems at OpenSIUC. It has been accepted for inclusion in Articles by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

Review Article

Soybean Genomics: Developments through the Use of Cultivar “Forrest”

David A. Lightfoot

Department of Plant Soil and General Agriculture, Center for Excellence in Soybean Research, Teaching and Outreach,
Southern Illinois University at Carbondale, Carbondale, IL 62901-4415, USA

Correspondence should be addressed to David A. Lightfoot, ga4082@siu.edu

Received 25 April 2007; Accepted 26 December 2007

Recommended by P. Gupta

Legume crops are particularly important for all cropping systems, world over, due to their ability to support symbiotic nitrogen fixation, a key to sustainable crop production and reduced carbon emission. Among legumes, however, soybean (*Glycine max*) has a special position as a major source of increased production in the common grass-legume rotation. The soybean crop has high seed protein content (~40%), good seed oil content (~20%) and is broadly tolerant to many diseases and stresses. In the past, attempts for genetic improvement to increase soybean seed yield largely relied on selection from the existing variability among cultivars. Often selection is focused on increased resistance to various diseases to avoid yield losses due to these diseases. The “Forrest” variety of soybean was named for the abilities of a southern Civil War General to arrange a defense. The cultivar “Forrest” and other Forrest-derived lines like “Hartwig” and “Ina” have saved US growers billions of dollars in crop losses due to resistances programmed into the genome of Forrest cultivar. Moreover, since Forrest grows well in the north-south transition zone, breeders have used this cultivar as a bridge to introduce a great deal of quantitative genetic variation from the southern to the northern US gene pool. Over the past decade, investment in Forrest genomics resulted in the development of the following resources for further genomics research: (i) a genetic map, (ii) three RIL populations (96 > n > 975), (iii) ~200 NILs, (iv) 115 220 BACs and BIBACs, (v) a physical map, (vi) 4 different minimum tiling path (MTP) sets, (vii) 25 123 BAC end sequences (BESs) that encompass 18.5 Mbp spaced out from the MTPs, (viii) a map of 2408 regions each found at a single position in the genome and 2104 regions found in 2 or 4 similar copies at different genomic locations (each of > 150 kbp), (ix) a map of homoeologous regions among both sets of regions, (x) a set of transcript abundance measurements that address biotic stress resistance, (xi) methods for transformation, (xii) methods for RNAi, (xiii) a TILLING resource for directed mutant isolation, and (xiv) analyses of conserved synteny with other sequenced genomes. Genes isolated from Forrest-derived BACs include candidates for resistance to nematode (*Rhg4* and *rhg1*), resistance to *Phytophthora sojae* (*Rps5*), resistance to *Pseudomonas syringae* (*Rps1*) and resistance to *Fusarium virguliforme* (*Rfs2*). These resources also assisted in the genomic analysis of soybean nodulation (*GmNark* and *GmNod*). Additional loci for seed yield, seed composition as well as resistances to 3 biotic stresses, 4 fungal species and 3 nematode species have been identified (unpublished reports). In combining desired characters, the structure of chromosomes appears to be pivotal to the special qualities of the Forrest genome. Genes underlying many quantitative and qualitative loci are targeted for isolation in the laboratories of the worldwide collaboration group. Data on the Forrest genome are provided to the scientific community through SoyGD, LIS, Soybase, and GenBank. The SoyGD portal has been particularly useful for the analysis of important biological processes. Sufficient numbers of BACs have already been sequenced to create a dense public database of new genetic markers for soybean breeders. The SoyGD portal at <http://soybeangenome.siu.edu> integrates the chromosome map with the whole genome shotgun sequence, a new community resource that identifies complete genes, a partial genome annotation and many thousands of SNP candidates in introns and promoters of protein-coding genes.

Copyright © 2008 David A. Lightfoot. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

1 The soybean cultivar “Forrest,” a product of a USDA
2 breeding program, represents a determinate, Southern

germplasm [1]. It was the first cultivar to possess soybean
cyst nematode (SCN) resistance associated with high yield,
and is believed to have played a key role in saving billions

of US dollars during 1970s and 1980s that would have otherwise been lost, either due to SCN or due to the poor agronomic performance of earlier SCN resistant cultivars (see [2] and references therein). Forrest was an important parent of modern cultivars, “Hartwig,” “Ina” and many others that have an improved SCN resistance gene from PI437654 introgressed into their genome [3–5]. Forrest was also central to an understanding of the genetics of resistance to sudden death syndrome, an important new disease of soybean [6–9].

Forrest is also one of the two cultivars (the other being “Williams 82”), providing the majority of genomic tools for soybean, available in USA (Figure 1) [10, 11]. These two cultivars provide models for soybean genomics research in the same way as are the cultivars *Col* and *Ler* in *Arabidopsis thaliana* or Mo17 and B73 in *Zea mays*. However, since the genomics of “Williams 82” was recently reviewed [11], its inclusion in this article would be repetitive. The other cultivars, which represent the worldwide germplasm variation for soybean genomics, include the following: (i) “Noir 1,” a Korean plant introduction (PI) [12], (ii) “Misuzudaizu,” a Japanese cultivar [13], and (iii) “Suinong14,” a Chinese cultivar [14]. The soybean community is committed to advance the genomics of all these cultivars, which have been used in the past as resources for genomics research. However, the intent of this review is to present an overview of the genomic resources derived from Forrest; these genomics resources enable a wide range of analyses that address several fundamental questions, like the following: (i) what is the source of genetic variation in soybean improvement? [15]; (ii) what is the role of variation in regions of genome duplication in paleopolyploid species? [16]; (iii) how does the nodulation of legumes work? [17]; (iv) why are protein and oil contents of seed inversely related? [18, 19]; (v) why are seed yield and disease resistance so hard to combine? [4, 5, 15, 20]; (vi) why is seed isoflavone content limited below 6 mg/kg? [18, 21–24]; (vii) how does partial resistance to disease work [6–9, 18]? It is believed that the development and use of genomics tools derived from Forrest will help soybean researchers to provide answers to these questions.

2. GENETIC VARIATION BETWEEN FORREST AND OTHER CULTIVARS

An important question that received the attention of soybean researchers in the past is how much sequence variation one can expect between Forrest and other cultivars, if many are to be sequenced. This variation is extensive (about 1 bp difference per 100–300 bp), when judged by using the criteria like the following: (i) the coefficient of parentage [25], (ii) the number of shared RFLP bands [26], (iii) polymorphism among microsatellite markers [27], and (iv) DNA sequence comparisons (Figure 2). In soybean, the degree of linkage disequilibria among loci is high, extending over distances that range from 50 kbp to 150 kbp [28]. Few meioses have occurred within these regions to reshuffle the gene or DNA sequences, because soybean is largely an inbreeding crop. In recent times, only seven or eight crosses have been made, starting from the time when the PIs were collected

to the development of most modern US cultivars (Figure 3). Therefore, in different parts of the genome, LD encompasses large segments and sets of genes.

2.1. The Essex × Forrest population

A soybean recombinant inbred line (RIL) mapping population (Reg. no. MP-2, NSL 431663 MAP) involving Forrest was recently developed from the cross “Essex MAP” (PI 636326 MAP) × “Forrest MAP” (PI 636325 MAP) [10]. This RIL population was used for constructing a genetic map [9, 24, 30] that has been used extensively for an analysis of marker-trait associations [7–9, 24, 30–38]. The genetic marker data encompass thousands of polymorphic markers and tens of thousands of sequence-tagged site (STS) that were collected at SIUC by Dr. Lightfoot’s group (Table 1) [10]. The genetic maps of E × F94 will continue to be enriched [27, 39]. The registration of this population [10] has allowed public access to the population and data generated from it worldwide.

A key feature of the above mapping population is that Essex (registered in 1973 [10]) was derived from the same southern US germplasm pool to which Forrest (registered in 1972 [1]) belongs. Consequently the RILs share identity across about 25% of their genomes, the portion that was monomorphic in both of the parents (Figure 3) [25, 26]. Further, the two cultivars were selected under similar conditions and, therefore, appear rather similar in most environments [6–10, 15–20, 30–38]. However, detailed records of maturity dates are important, since even a single day variation in maturity may influence the results of QTL analysis for many other traits [10, 41]. Since morphological and developmental traits differ very little in the population, the RILs have been used extensively to map those genes which control biochemical and physiological traits (Table 2). For example, the parents of the mapping population differ by resistance traits, which exhibit both qualitative and quantitative inheritance (Table 3).

A major limitation in using E × F population in genomics research is the small population size ($n = 100$) that could preclude fine mapping [10]. To overcome this problem, populations of near isogenic lines (NILs; $n = 40$; Figure 3) were developed from each RIL [10, 37, 38, 43]. The NIL populations are listed in Table 1. The residual heterozygosity present in the F5 seed was largely fixed and captured in these NILs. The heterogeneity across the RILs has been measured to be 8%, which is more than the 6.25% expected among F5 lines [7, 24]. That increased heterogeneity appears to be caused by selection, since rare heterozygous plants still exist in some RILs and NILs [37, 38, 40]. Each locus that segregates in the RIL population is expected to segregate in about eight NIL populations. Therefore, each region in the genome will be segregating in about 420 lines ($100 + 8 \times 40$), quite sufficient to create fine maps of 0.25 cM resolution (Table 4). A 0.25 cM interval represents 25–100 kbp on the physical map [16], sufficient for candidate gene identification [37, 38].

Consequent to the development of the NILs, the E × F population was used to study the genetics of a large number

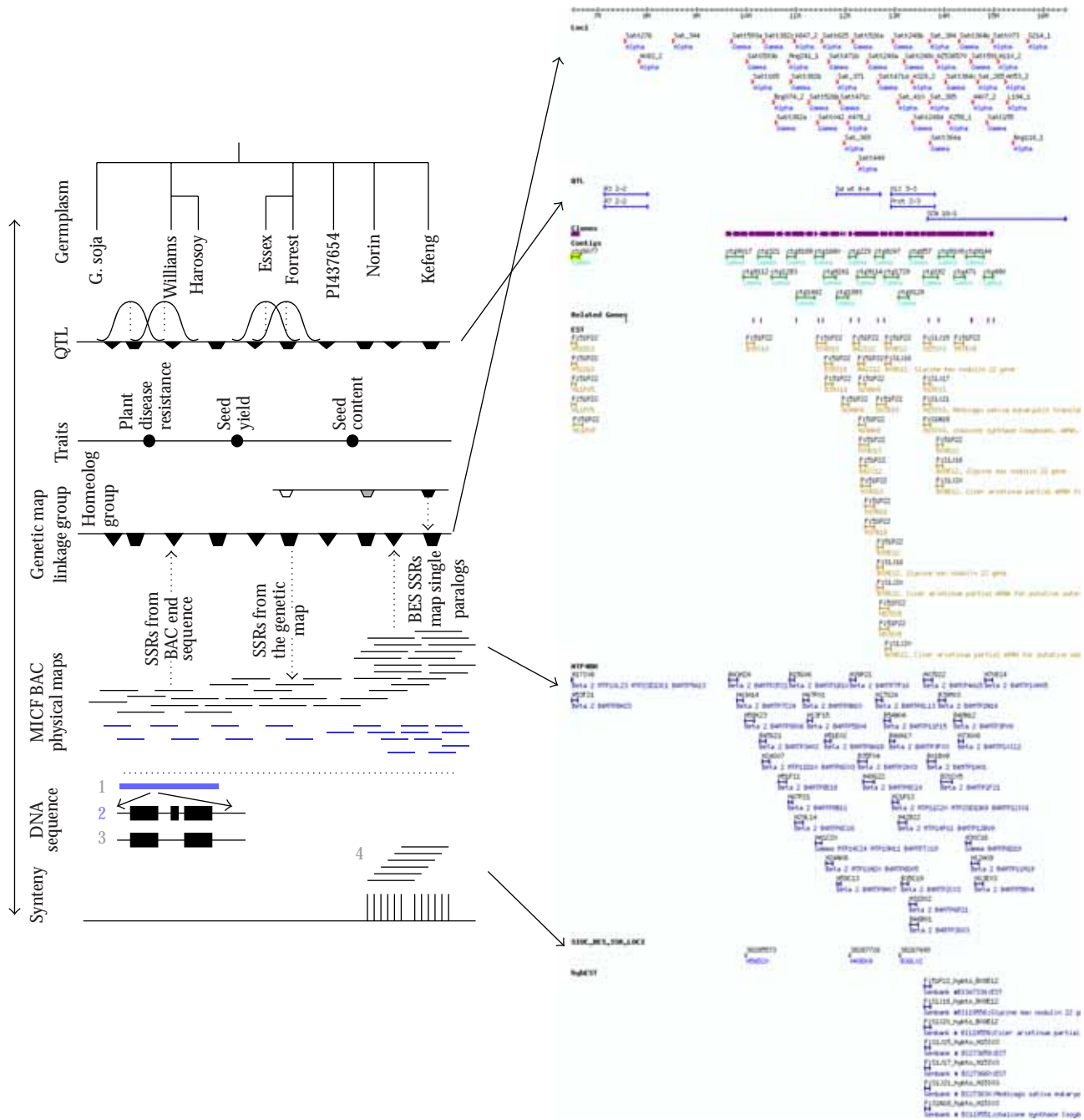


FIGURE 1: Soybean genomic resources and products schematic for Forrest (A) compared to the SoyGD representation (B). Panel A. Germplasm that are exemplars of soybean genetic diversity are shown. Selected germplasm encompass in mapped QTL a wide variety of traits placed on the composite genetic map. BAC libraries exist for many of the germplasm sources. Forrest BACs (shown in black) form the basis of an MICF physical map with 6-fold coverage. A region of conserved duplication (12-fold coverage) is shown on the right of the figure. In this region, fingerprinted clones from two homoeologous linkage groups coalesce. Genetic markers identified in, or derived from, **BESs** will separate some of the duplicated conserved regions. Genetic markers anchored from map to BAC are of little use in conserved duplicated regions. BACs from diverse germplasm are shown as blue bars. There are 3 levels of DNA sequence envisioned. At level 1, **BESs** provide a sequence every 10–15 kbp with which to identify gene rich regions for later complete sequence determination (level 2). Arrayed BAC end sequences will be used to identify conserved syntenic regions in the genomes of model plant species. This information will also separate some of the duplicated conserved regions in soybean. Panel B. Shown are the chromosome (cursor), DNA markers (top row of features, red); QTL in the region (second row, blue); coalesced clones (purple) comprising the anchored contigs (third row, green); BAC end sequences (fourth row black); **BESs** encoding gene fragments (fifth row, purple); EST hybridizations to MTP2BH (sixth row gold); MTP4BH clones (seventh row, dark blue); **BESs**-derived SSR (eighth row, green).

DX409547SOYFK12TH LargeInsertSoybeanGenLibBuild4 Glycine max genomic clone H53F21:Build4MTP8A23, genomic			
Survey sequence	117	GGC TTTGATTGAGGCTTCTTTCCCTTGATTTCTGCCATTCTTACTAGCTTATTTCAATTGT	176
1587408390FFYA466822.x2	392	451
1645481618FFOF353160.b1	29	88
1558430135BIWS948569.x1	486	545
159432350FFYA560607.y2	517	458
1559927967BXC212749.g1	169	. . . T . . . G . . . T . - A . C A . . . AT . . . C C	227
1580476593BXC552535.g1	245 G T . - A . C A T C G	303
1559913219BXC315601.g1	822	. . . T . . . G . . . T . - A . C A AT C C	764
1564454727FFYA110813.b1	341 G T . A . C A AT C T	399
158038383BXC524105.g1	463 G T . - A . C A AT C C	522
1580748770FFYA346301.g1	677	. . . T . . . G . . . T . - A . C T T C C	619
1315473315BIWS198915.y3	253	. . . T . . . G . . . T . - A . C A T C C	311
1315592865BIWS253663.x2	809 G T . - A . C A AT C T	751
1597304772FGNN95100.g1	108 G T . A . C A T C G	166
1547297263BIWU102117.b1	902 G T . - A . C A T C G	844
1563485876BXC282850.b1	469	. . . T . . . G . . . T . - A . C A AT C C	411
1315286149BIWS119045.b1	709	. . . T . . . G . . . T . - A . C A AT C G	651
1587384578FFYA424540.x2	684	. . . T . . . G . . . T . - A . C A AT C G	626
1680112410FFOF435772.g1	743	. . . T . . . G . . . T . - A . C A AT C C	686
1576170959FFYA331684.y2	529	. . . T . . . G . . . T . - A . C . . A A T C G	471
1547040313BIWS586771.y1	302 G T . - A . C A T C G	360
1553727822BXC30662.g1	319 G T . - A . C A T C G	260
1564467352FFYA129198.g1	317 G T . A A C G	376

(a)

CG825374SOYBA22TV LargeInsertSoybeanGenLib Glycine max genomic clone B47P08:MTP7C19, genomic survey			
Sequence	65	AGGGACAGGGGAATGTGGTCTTTTCTTGATCCTCAGGAGCATTATGAAGGGGGAAAGAAG	124
1546960481BIWS535643.y2	57	116
1547840359BXC54981.g1	32 T	91
1607454735FGNN173762.b1	613 T	554
1597034748FGNN43751.b1	521 T	462
1610955864FGNN203808.b1	373 T	432

CG825374SOYBA22TV LargeInsertSoybeanGenLib Glycine max genomic clone B47P08:MTP7C19, genomic survey			
Sequence	125	GAGAAAACGATGAAGAAAAGAGTAAGGAGACTTAGCTGTCAAGCGCTCAAGCATTTGAT	184
1546960481BIWS535643.y2	117	C C C C T C	176
1547840359BXC54981.g1	92	C C C C T C	151
1607454735FGNN173762.b1	553	C C C C T C	494
1597034748FGNN43751.b1	461	C C C C T C	402
1610955864FGNN203808.b1	433	C C C C T C	492

(b)

H53F21		A T C T T A C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A T G
E22P03	A T T C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A T G	
E05A01	A T T C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A T G	
H07C13	A T T C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A T G	
H53H14	A T T C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T G C C T A T C T T T G A A A T G	
H65P05	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
H20I07	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
H39K22	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
E66B10	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
H65D04	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
H17I08	A T C C T T A C T A G C T T A T T T C A A T T G T G A C T C T A A T G T G T A T G T T C C T A T C T T T G A A A A G	
ISO56K20	A T C C T T A C T A G C T T A T T T A A C T G T G A C T C T A A T G C T T A T G C T C C T A T C T T T G A A A A G	

(c)

FIGURE 2: Comparison of MegaBlast analysis of an unduplicated region and a twice duplicated region as inferred by the fingerprint physical map (a). Analysis of the BESs from H53F21 in quadruplicated contig 9077. These BESs contained a very common repeat with 400 copies per haploid genome. Sequence analysis supported the inferred by a four copies of the region per haploid genome detected by fingerprints (a). MegaBlast of H53F21 (Build4MTP8A23, gi89261445) against 7.3 million reads with repeated masking gave 7 identical matches among 24 homoeologous sequences. Cluster 1 was composed of traces ending in ...822,...160,...569,...607,...662,...749, and ...105 that shared A at position 172 (circled). Homoeolog specific variations (polymorphisms) were evident among the 4 clusters inferred. Cluster 2 was composed of clones ending in 749, 850, and 601 that shared C at position 172. Cluster 3 was composed of clones ending in 100, 117, and 535 that shared G at position 172. Cluster 4 also had G at that position. TreeCluster analysis showed the most similar homoeologs clustered into 4 separate sets as expected for regions duplicated twice (circled) (b). Analysis of the BES from B47P08 in contig 321 from an unduplicated region. Sequence analysis supported the inferred with an unduplicated region detected by fingerprints at 90% sequence identity (c). The sequences found among BACs resequenced from contig 9077 showing a set of SNHs (HSVs) separated two groups of the four inferred to be present: the A cluster and the G cluster (adapted from [29]).

of quantitative traits (QTs), leading to the identification of quantitative trait loci (QTL; Table 2) underlying more than seventy different traits [24, 39, 40, 42, 44–46]. Biochemical and physiological traits included resistance to soybean sudden death syndrome (SDS) [caused by *Fusarium virguliforme*] in the US and Argentina, resistance to soybean cyst nematode (SCN; *Heterodera glycine* Ichinohe), seed yield,

seed quality traits, agronomic traits, water use efficiency, manganese toxicity, aluminum toxicity, partial resistance to *Phytophthora sojae*, and insect herbivory. However, new opportunities abound because dozens of traits for resistance to pests and pathogens segregate in the population but were not yet mapped [10]. Further, the concentrations of many secondary metabolites among lines vary widely during

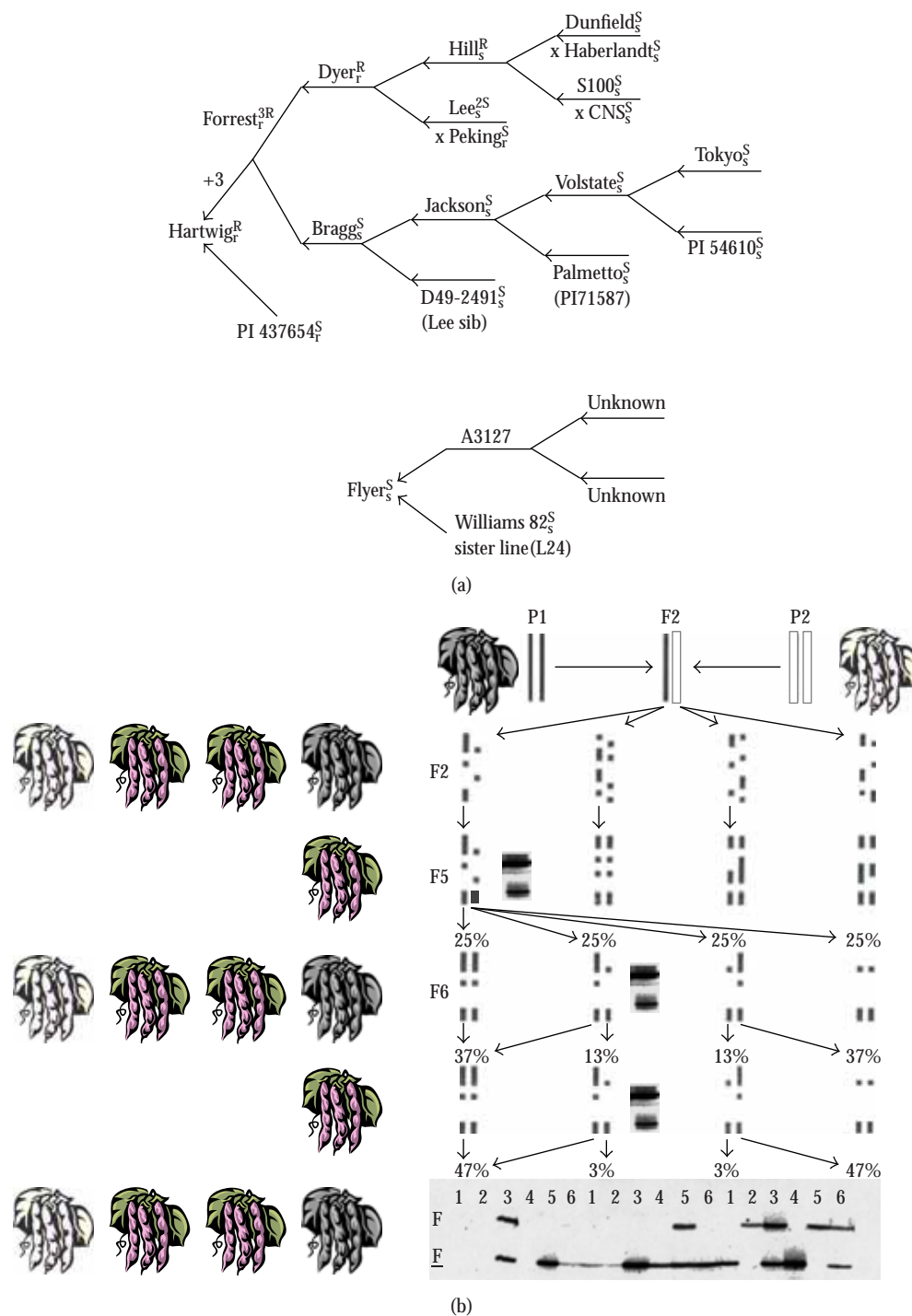


FIGURE 3: Genetic systems used with Forrest germplasm and the inbred soybean crop (a). The ancestry of Forrest and Hartwig showing the known cultivars that were crossed and the relationship between Flyer and Williams 82 (b). A diagram showing how NILs derived from RILs fix most loci but allow the continued segregation of heterozygous regions in inbred crops like soybean. The effect is to Mendelize a few of the loci contributing to QT while causing the majority to be fixed. A dark pod parent is crossed with a light colored pod parent of the F1 heterozygous type (shown as purple pods) selfed and F2 progeny advanced to the F5. A heterozygous plant at any time or heterogeneous RIL at F_{5,7} or later identified is shown as purple pods. Single plants are extracted and seed increased. NILs that result may fix the heterogeneous region to the parent 1 allele, the parent 2 allele, or are still heterogeneous. Occasionally heterozygous plants are found within some heterogeneous NILs even at the F_{5,15} and the progeny of such plants can be used to find new recombination events. Shown are the results with Satt309 and NIL11 plant 3 and eighteen of the progeny collected from it (adapted from [40]).

TABLE 1: Description of 20 linkage groups mapped in the Essex \times Forrest mapping population. The map distances and markers distribution for the linkage groups were generated from analysis of the 100 F₅-derived progeny from E \times F.

Linkage group	NIL ^(a) populations	Map		No. of markers					
		Distance (cM)	Total	SSR	RFLP	RAPD	BESs ^(b)	EST ^(b)	BESs ^(c) SSR
A1	6	73.8	14	4	3	7	458	13	4
A2	8	259.0	22	10	8	4	757	0	7
B1	4	164.0	16	11	2	3	234	7	5
B2	5	53.4	12	7	1	4	156	3	6
C1	4	150.1	13	10	0	3	136	0	9
C2	8	213.2	30	19	4	7	565	14	4
D1a + Q	9	140.0	17	14	0	3	625	30	3
D1b+W	8	87.4	14	8	1	5	124	1	3
D2	7	245.4	19	15	0	4	122	0	4
E	6	97.4	9	6	0	3	362	11	5
F	4	219.9	29	16	5	8	369	0	2
G	12	242.5	37	19	12	6	1126	33	5
H	8	98.3	9	6	1	2	427	9	4
I	9	116.9	16	11	0	5	192	6	3
J	7	40.7	7	3	1	3	577	3	2
K	9	150.9	18	13	0	5	590	1	4
L	8	103.8	12	9	0	3	91	3	2
M	6	105.2	10	6	1	3	87	9	4
N	3	145.1	21	9	2	10	156	0	3
O	2	116.4	13	10	0	3	566	9	0
Total	100	2823.4	337	206	41	90	7720	152	79
Unlinked	(2007)	0	0	0	0	0	10529	485	10

^(a) NIL populations segregate for 2 or more regions on different chromosomes.

^(b) ESTs and BESs may appear at 2 or more locations on the linkage map if they appear in homoeologous regions of different linkage groups.

^(c) BESs-SSR placed on the genetic map, many more are placed in SoyGD by inference from marker anchored contigs.

development and among different organs [47]. Pesticide uptake, metabolism and degradation rates also vary among lines (unpublished). Preliminary studies have shown the link between the genome, proteome, and metabolome (the interactome), which can be further explored in these segregating populations [48]. Therefore, E \times F will eventually be used to map thousands of QTL for hundreds of QT.

Importantly, the NILs that have been developed from each RIL for fine mapping also allow confirmation of QTL detected in the RIL population. For instance, *cqSDS001* was assigned to a QTL confirmed by NILs derived from Ripley [49], but earlier detected through RILs derived from Flyer [50] and "Pyramid" [6, 33]. The QTL have also been renamed under the new rules for QTL adopted by the Soybean Genetic Committee in 2006 [51], as a result of which *cqRfs1*, *cqRfs2*, and *cqRfs4* were renamed as *cqSDS003*, *cqSDS002*, and *cqSDS004*, respectively.

The molecular linkage map, the RILs, and the NILs were used during the positional cloning of *nts1*, *GmNARK* [50], *Rpg1* [17, 35], *Rhg1*, [38] *Rhg4*, [52], and *Rfs2* [37]. Many opportunities for further gene isolations exist. Tables 2 and 3 list some of the known phenotypes that differ between the parents and segregate among the lines and that are candidates for gene isolation. The RIL and NIL populations provide sets of recombination events that can be used to identify the

positions of genes underlying QT [10]. Since all the lines self-fertilize, the populations can be used to provide an immortal resource, if seed germination ability can be regenerated every five years. This type of resource is particularly important for soybean because the draft genome sequence will be released in April 2008 (unpublished). Combining knowledge of locus positions with a comprehensive knowledge of gene content will lead to the rapid isolation of many new and economically important genes [16].

Selected lines from the E \times F population that contrast for mapped QTL were also used for a variety of studies including the following: (i) to validate assays of pathogenicity [32, 53–55], (ii) to examine the effects of resistance genes on gene expression [34, 56, 57], (iii) to analyze components of drought tolerance [24, 31, 36, 42, 46, 58], (iv) to validate methods of marker assisted selection [6, 31, 59–62], and (v) to provide for germplasm releases (Figure 4) and cultivars [6, 63]. New cultivars and new methods for selection of improved soybean genotypes are among the most important spin-offs from the genomics research involving Forrest soybean. Among the selected lines, E \times F78 later became LS-G96 [63] and then "Gateway 512" (Gateway Seeds, Nashville, Ill, USA). This line together with the line E \times F55 was used as parents that combined moderate resistance (carrying resistance alleles at six loci) to SDS with high yield. The

TABLE 2: Ranges and means of selected mean traits measured across multiple locations and years using the RIL population and the “Essex” and “Forrest” parents. For traits 1–35 see [24]; traits 36–79 were from [39, 42] and or unpublished.

No. of trait and symbol	Unit	RIL population	
		Average	Range
1. SDS disease incidence	Score	48.5	4.4–94
2. SDS disease severity	Score	1.5	1.1–2.3
3. SDS disease index	Score	9.3	1.1–23.9
4. Soybean cyst nematode IP	(%)	53	0–100
5. Yield during SDS	Kg·ha ⁻¹	3.3	2.9–3.76
6. Seed daidzein content	μg·g ⁻¹	1314	874.5–2181
7. Seed genistein content	μg·g ⁻¹	996.8	695.5–1329
8. Seed glycitein content	μg·g ⁻¹	206.1	116–309
15. Total seed isoflavone content	μg·g ⁻¹	2516.8	1774.2–3759
21. Resistance to manganese toxicity	Scale 0–5	2.02	1.1–4.5
32. Seed yield	Kg ha ⁻¹	3.44	2.64–4.13
33. Leaf trigonelline content (irrigated)	μg·g ⁻¹	98.85	59.87–126.96
34. Leaf trigonelline content (rain-fed)	μg·g ⁻¹	417.94	245.95–618.18
35. Flower color (white: purple)	color	43:47	na
38. Mean SDS DX in Argentina	Scale 1–10	1.6	0.1–3.1
43. Tolerance to aluminum toxicity	(%)	14	–20–37
47. Seed protein content	(%)	39.5	37.5–41.5
51. Seed oil content	(%)	18.9	18.0–20.1
55. Resistance to insect herbivory (IP)	(%)	22.3	13.0–32.5
60. Seedling root growth	mm	8.3	6–11

RIL E × F23 was released as SD-X for very high resistance to SDS [34] and good yield potential under license from Access Plant Technologies (Plymouth, Ind, USA), because it contained beneficial alleles at all eight known resistance loci. In contrast, E × F85 is susceptible to SDS as it contained no beneficial alleles at the known resistance loci. It makes a great entry for sentinel plots. For animal feed and human food, E × F52 has been used as a parent to provide very high phytoestrogen contents to progeny (unpublished), since it contained beneficial alleles at all the known loci underlying phytoestrogen content. Low phytoestrogen contents are also required for estrogen sensitive consumers; E × F89 and E × F92 were used as parents to provide parents for low phytoestrogen in the progeny (unpublished).

2.2. Related populations flyer by hartwig (F × H) and Resnik by Hartwig (R × H)

The F × H and R × H populations are integrated with E × F96 [10], since Forrest was the recurrent parent used to develop Hartwig (Figure 3) [62] and Essex shares many alleles with the Flyer and Resnik [15, 27]. Flyer and Resnik were sister lines derived from a cross between a Williams 82 sister line and a commercial cultivar [64]. The F × H has 92 RILs and R × H has 952 RILs that have been used to confirm QTL detected in E × F96 and for fine mapping of these QTL [4, 5, 15, 50, 52]. Flyer and Resnik each contains many genes conferring resistance against *P. sojae*. Both these populations can be used to map genes underlying additional biochemical, physiological, and some agronomic traits that

include the following: (i) resistance against *Phytophthora* root rot, soybean sudden death syndrome (SDS) caused by *F. virguliforme* and soybean cyst nematode (SCN), *Heterodera glycine* Ichinohe, (ii) seed yield [15, 50, 52], and (iii) seed quality traits. These RILs were also used to develop SSR markers that anchor contigs and sequence scaffolds (<http://soybeangenome.siu.edu>) to the physical map [27].

3. PHENOTYPIC VARIATION BETWEEN FORREST AND OTHER CULTIVARS

One major limitation using the resources based on Forrest was the low amount of genetic variation detected in the populations based on this cultivar [65]. The implication was that the alleles detected in E × F would not be weaker variants of the major gene effects found in weedy plant introductions (PIs). It was hypothesized that, instead, the loci detected in the E × F population and in the material derived from this population perhaps represented other gene systems of lower hierarchical position and therefore lower value. Consideration of a few examples of the locations of QTL underlying phenotypic variation between Forrest and other cultivars has been informative regarding this issue. The results to date all infer that the alleles underlying QTs in Forrest are variations in the same genes as the PI alleles, if weaker in effects on QTs.

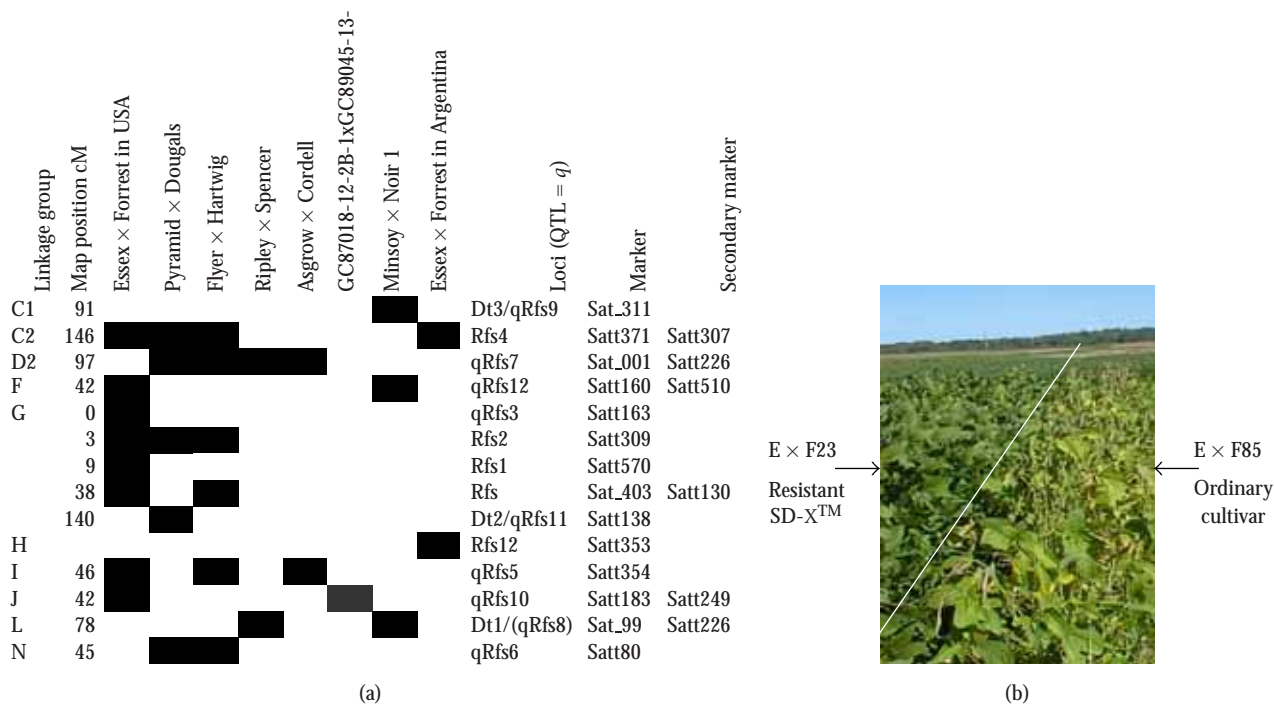


FIGURE 4: An example of the use of Forrest genomics resources for soybean germplasm improvement (a). Summary of the map locations of the known loci for resistance to SDS. A black rectangle indicates that the allele is segregating in that population. Nonsegregating alleles may be either fixed to the resistance or susceptibility forms (b). An example of quantitative variation for disease resistance identified in lines derived from Forrest. The resistant line RIL23, left of the line, has beneficial alleles for six QTL for resistance to *Fusarium virguliforme*. The leaf scorch associated with the fungal infection is evident in the neighboring RIL80 to the right of the white line.

3.1. The genetics of phytoestrogen content

The phytoestrogen content of soybeans seed mainly consists of daidzein (60%) and genistein (~30%) with small proportion of glycitein (~10%) [66]. Analysis of germplasm and elite cultivars (18, 21–24, 67–69) indicated that phytoestrogen concentrations in some elite cultivars (~2 mg/kg) were higher than those in many of the ancestors of cultivated soybean (~1 mg/kg). Phytoestrogen content and profile varied with environment (year and location effect) and genotype. However, the final seed content was largely controlled by the genotype (40–60% of the variation) and is controlled by a set of about 6–12 loci [18, 24, 67]. If the content of each phytoestrogen component was controlled independently, improvements in content by genetic selection should be possible. For instance, raising glycitein content to the same amount as that of daidzein could double the total phytoestrogen content. However, because heritability of phytoestrogen content is moderate at about 40–60%, direct selection (without DNA markers) has not been very effective. Through marker-assisted selection (MAS), the phytoestrogen amounts were raised to 3.6 mg/kg, well above the amounts found in elite cultivars or weedy PIs. Here, the variation programmed by the alleles segregating in E × F population was greater than that among the entire germplasm collection.

Recently, crosses have been made between lines from southern Illinois and Canada having the highest phytoe-

strogen contents [23] and, separately, the lines having the lowest phytoestrogen content [67]. MAS exercised in the segregating populations (at F4 in 2007) should lead to improvement in phytoestrogen content. Opportunities for collaborative studies exist with sets of RILs in maturity groups that are not adapted to be grown in southern Illinois or Canada.

3.2. The genetics of seed yield, protein and oil content

The overall average increase of 1–2% per year in soybean yield witnessed during 1960–1999 was only half the yield advances achieved in corn and other out crossing crops, where genetic diversity was not limiting [68]. As one would expect, there are hundreds of loci controlling yield in soybean [69]. In view of this, half of the yield loci detected in E × F population were those which were earlier detected in other crosses [24]. These loci could each boost seed yield by 0.2 Mg/Ha. In contrast, substantial gains (0.9–1.1 Mg/Ha) can be made in soybean yield by identifying unique alleles in weedy PIs and introgressions into elite cultivars [70]. The nature of the genes altering seed yield will be an interesting product from fine map analysis and positional cloning.

The major components of soybean seed yield include the following: (i) protein (~40%), (ii) oil (~20%), (iii) structural carbohydrates (~6%), (iv) water (~13%), (v) soluble carbohydrates (~14%), and (vi) other metabolites (~7%) [71]. Metabolic changes during development driven

TABLE 3: Disease resistance that segregates among the RIL and NIL population.

Disease resistance in	Causal agent
A. Forrest	
Soybean cyst nematode	<i>Heterodera glycines</i> HG type 0; races 3
Root-knot nematode	<i>Meloidogyne incognita</i>
Bacterial pustule	<i>Xanthomonas glycines</i>
Wildfire	<i>Pseudomonas syringae</i> subsp. <i>tabaci</i>
Target spot	<i>Alternaria</i> sp
Partial <i>Phytophthora</i> root rot	<i>Phytophthora sojae</i>
SDS root rot	<i>Fusarium virguliforme</i>
SDS leaf symptoms	Toxin
B. Essex	
Bacterial pustule	<i>Xanthomonas glycines</i>
Downy mildew	<i>Peronospora manshurica</i>
Frogeye leaf spot	<i>Cercospora sojae</i>
Purple seed stain disease	<i>Cercospora kikuchii</i>
Partial <i>Phytophthora</i> root rot	<i>Phytophthora sojae</i>
SDS leaf symptoms	Toxin
C. Hartwig	
Soybean cyst nematode	All HG Types from 1.2.3.4.5.6.7.
Root-knot nematode	<i>Meloidogyne incognita</i>
Reniform nematode	<i>Rotenlenchulus reniformis</i>
Bacterial pustule	<i>Xanthomonas axonopodis</i> pv. <i>glycines</i>
Wildfire	<i>Pseudomonas syringae</i> pv. <i>tabaci</i>
Target spot	<i>Corynespora cassiicola</i>
Partial <i>Phytophthora</i> root rot	<i>Phytophthora sojae</i>
SDS root rot	<i>Fusarium virguliforme</i>
SDS leaf symptoms	Toxin
D. Flyer	
Powdery mildew	caused by <i>Microspora diffusa</i>
Purple seed stain disease	<i>Cercospora kikuchii</i>
Pod and stem blight	<i>Diaporthe phaselorum</i>
Multirace <i>Phytophthora</i> root rot	<i>Phytophthora sojae</i>
SDS leaf symptoms	Toxin

by gene expression underlie the seed composition and yield [72]. Seed yield and composition are under polygenic control with different genes active at different stages of seed development. Seed traits are also associated with significant genotype \times environment ($G \times E$) interactions as observed in $E \times F$ population (see [15, 18, 19]). Again, the $G \times E$ interactions significantly reduce the effectiveness of visual selection based on the phenotype alone.

At harvest, seed protein content is inversely related to seed oil content and seed yield in $E \times F$ population [18, 19]

as also in other germplasm (see [68]). While some loci are implicated in all the three traits, there are others which influence only one or two of the three traits. Several QTL underlying soybean yield, protein, and oil content have been mapped in both the $E \times F$ and the $F \times H$ RIL populations [5, 18]. They do correspond with loci detected in crosses between high protein weedy types and low protein adapted cultivars. Three QTL on linkage groups A1, A2 and linkage group E have been fine-mapped and localized within 0.25 cM using substitution mapping to identify the underlying genes. Isolation of these genes will partly explain the molecular basis of the genetic control of yield and its component traits. However, a danger here is that because different genes are active at different stages of seed development, one would generally map only a composite trait, based on a mean of the action of several loci. Isolation of genes by position would not be successful in this circumstance.

3.3. The genetics of *Phytophthora* root rot resistance

The annual soybean yield loss suffered from the root and stem rot disease caused by the oomycete pathogen, *Phytophthora sojae* is valued at about \$273 million in the US [73]. Monogenic resistance due to a series of *Rps* genes has been providing a reasonable protection to the soybean crop against the pathogen over the last four decades [74]. Several mapped *Rps*-genes are known to occur in Flyer and Resnik [50, 64]. Partial, rate-reducing resistance to many races of *P. sojae* is found also in Forrest, Essex, and Hartwig. The loci providing this partial resistance were not mapped by 2007.

3.4. The genetics of SCN resistance

Soybean cyst nematodes (*Heterodera glycines* L.) are the most damaging pests of soybean worldwide [73]. Development of resistant cultivars is the only viable control measure [75]. Resistance genes have been found to be located on 17 of the 20 chromosomes by 2007. A combination of recessive genes is necessary to provide resistance against SCN populations because many are known to be capable of overcoming all known single resistance genes. SCN populations can be classified into 16 broad races or up to 1024 biotypes (HG Types) [76] based on the host responses of 8 weedy indicator lines. SCN resistance in many other adapted and weedy cultivars [9, 31] shared the same loci underlying bigenic inheritance in $E \times F$ [20]. The $E \times F$ population was used to isolate candidate genes for those two loci (*rhg1* and *Rhg4*; Table 4) that control resistance against SCN race 3 (HG Type 0). Alleles of the candidate genes were identified in many PIs through association studies [38, 77]. Paralogs of both these genes were found at new locations in BAC libraries and whole genome shotgun (WGS) sequences [78, 79]. They appear to be part of multigene families showing homoeology and intragenomic conserved synteny.

Three cultivars including Peking, PI437654, and Hartwig encoded 2–4 additional genes that provide additional resistances to SCN [52, 80, 81]. Peking has alleles for resistance to races 1 and 5 that were not transferred to Forrest [20]. Hartwig and PI437654 have complete resistance against all

TABLE 4: Saturation mapping with markers on chromosome 18 in the 2–4 Mbp encompassing *Rhg1*, *Rfs1*, and *Rfs2* (SDS) loci with leaf and root phenotype classes shown.

Geno type	Satt214	Sat1	TMD1	Satt309	Sat185	CGG5	OI03	CTA13	Bng122	Leaf	Root
1	E	F	E	E	F	E	F	E	F	S	R
2	E	E	E	E	E	E	E	E	F	R	S
3	E	E	E	H	E	E	E	E	F	R	S
4	E	E	F	F	E	E	E	E	E	R	S
5	E	F	F	F	F	E	E	E	E	R	S
6	F	F	F	F	E	F	F	F	F	R	R
7	F	F	F	F	E	E	E	F	F	R	S
8	F	F	F	F	F	F	F	F	E	R	R

rices of SCN except race 0, HG Type 1.2.3.4.5.6.7.8. The location of SCN resistance loci in $F \times H$ and $R \times H$ agreed with those found in crosses between PIs and adapted germplasm [81, 82]. Therefore, the resistance to SCN traits that are introgressed from PIs to Forrest-based germplasm is useful and the underlying genes can be isolated from Forrest.

3.5. The genetics of SDS resistance

Soybean sudden death syndrome caused by *Fusarium virguliforme* (e.g., *solani* f. sp. *glycines*) is among the most damaging syndrome of diseases affecting soybean in the US and worldwide [73]. The syndrome is composed of a root rot disease and a leaf scorch disease [53]. Development of resistant cultivars is the only viable control measure. Twelve resistance loci have now been found on 8 chromosomes (Figure 4), eight segregate in $E \times F$ [24, 44] and two additional loci segregate in $F \times H$ [5, 50]. A combination of loci is needed to provide resistance to both root rot (2 or more loci) and leaf scorch (all loci). Loci for resistance to SDS were named *Rfs* to *Rfs11* [39]. Using NILs (Table 4), a set of candidate genes for the *Rfs2* locus were identified [37]. Among these genes, a receptor like kinase [38] and a laccase [83] are being tested for their ability to provide resistance following transformation and mutation (unpublished). However, the presence of a pair of syntenic genes on linkage group O with similar DNA sequences (84%) and encoding nearly identical amino acid sequences (98%) complicates the analysis following reverse genetics approach.

One of the two loci underlying root rot resistance is encoded in the DNA sequence around marker OI03₅₁₄ that lies between AFLP derived SCARs, CGG5, and CTA13 on linkage group G [37]. However, the root rot resistance locus (Table 4) lay in a region not well represented among BAC libraries [84, 85], so that the gene isolation was delayed until the local genome sequence could be assembled. Transcript analysis showed that the fungus attempts to prevent gene transcription in the target roots [34, 55, 56]. Resistant cultivars prevent the poisoning of transcription by inducing stress and defense genes that produce fungicidal metabolites within 2 days of contact with the fungus. However, the induced genes do not appear to map to the loci that control the SDS resistance response [57]. Instead, genes of a higher hierarchical position in the interactome were found in this

interval (unpublished). One of these genes is expected to underlie root resistance to SDS.

For the fungus, *F. virguliforme* causing SDS, no races are known so far in the US [86]. When lines from $E \times F$ have been used to look at variations in pathogenicity between strains, no convincing evidence for a host differential response was observed (unpublished). However, different *Fusarium* species that are capable of causing SDS are found in South America [86]. $E \times F$ was planted in Argentina since 2004, and it was shown that the SDS pathogen(s) invoked responses that mapped to different resistance loci [39]. Therefore, the fungus does have the potential to form races that vary in their pathogenicity. Hence, soybean breeders should be cautious in using the available resistance genes and should realize that stacking of all the twelve genes for full resistance would not be wise because it would select for mutants in the pathogen populations that could lead to the development of races.

In conclusion, a variety of approaches including QTL analysis, fine map development for some loci, and analysis of isolated genes have revealed that the alleles detected in $E \times F$ are variants of the same major genes found in weedy plant introductions (PIs) [5, 24, 41, 53]. Only few loci detected in the $E \times F$ population and in the other materials derived from this cross seem to represent other gene systems at a lower hierarchical position [57]. Identification of the lower tier of genetic control may require intercrosses among NILs or assays that relate to development, time, position, or cell type.

4. STRUCTURAL GENOMICS RESOURCES

Soybean (*Glycine max* L. Merr.) has a genome size of 1115 Mbp/1C [87]. The soybean genome is the product of a diploid ancestor ($n = 11$), that underwent aneuploid loss ($n = 10$), allo- and autopolyploidization events separated by millions of years ($n = 40$) with reversion to a lower ploidy after one of those two events ($n = 20$) [88]. Evidence that two genome duplications occurred, 40–50 MYA and 8–10 MYA, was supported by RFLP analysis suggesting 4–8 homoeologous loci for most probes [89] and discontinuous variation among paralogous EST sequences [90–92]. Even PCR-based markers that can amplify single loci from genomic DNA amplify multiple amplicons from BAC pool DNA (Figure 2).

The duplicated regions have been segmented and reshuffled to **polyploidization** events [16, 93–95].

Recently, a systematic measurement of DNA sequence divergence between homoeologous regions was made possible by comparing Forrest BAC end sequences with 7 million reads from the WGS sequences of Williams 82 [29, 93]. MegaBlast searches distinguished some regions, resolving up to 10% nonidentity between homoeologs over a 60 bp window (Figure 2). This implied that significant sequence divergence has occurred at about half the loci tested, as predicted from the gene-family size distribution observed in the physical map [57] (Figure 5). Conversely, highly conserved regions (>90% identity) exceeding about 150 kbp (the size of a large insert clone) have been inferred in certain regions [29]. Within these regions, 2 or 4 homoeologs can be distinguished by single nucleotide variants that correspond to the duplicated regions of a paleopolyploid genome or recently polyploid genome. These variants have been described as single nucleotide polymorphisms among homoeologs (SNHs) [93] though they are commonly called homoeologous sequence variants (HSVs) (see, e.g., [91]).

Overlain on the segmented regions found in 2 or 4 copies, the soybean genome is a composite of dispersed and contiguous euchromatic regions [88]. The short arms of four chromosomes are entirely heterochromatic, but in the remaining 16 chromosomes with potentially gene rich euchromatic arms, the heterochromatin is restricted to pericentromeric regions. Euchromatin represents 64% of the soybean genome, with a range of 40–85% on an individual chromosome. Due to these features and the following other reasons, analysis of soybean genome has been a challenge: (i) large genome size, (ii) serial duplication of regions, (iii) small proportion of unique DNA, and (iv) highly conserved repeated DNA. One reasonable prediction would be that many of the duplicated regions would be silenced in heterochromatin. However, a comparison of the genetic map and physical map [93–95] has shown that duplicated segments are neither clustered nor restricted to heterochromatic arms. Further, the gene-rich islands are not separate from the duplicated regions. Therefore, new models to explain gene regulation that include duplicated conditions must be developed. Lessons learned from this exercise will help in the analysis of some legume and many dicotyledonous crop genomes, where genome duplication is believed to have often accompanied speciation. Breeders, who develop new cultivars through selection from the available variation within a cultivar, will also utilize this information and will develop new selection methods through an understanding of the effects and benefits of partial, segmented, genome duplication.

4.1. BAC libraries and physical maps

Construction of fingerprint-based physical maps in soybean relied **on the and availability** of deep-coverage high-quality large insert genomic libraries, and a number of such public-sector large insert libraries are available in four different plasmid vectors, providing >45-fold genome coverage. BAC libraries are available not only for Forrest and PI437654,

but also for some *G. soja* PIs and the wild relatives of *G. max* [84, 85, 96, 97]. Among these libraries, there are three “Forrest” BAC libraries [84, 85], available in two different plasmid vectors with different *oris* and different selectable markers (Table 5). Despite the availability of these rich BAC resources, there are still a few regions of the genome that are not well represented across the above set of BAC libraries. New libraries without involving restriction digestion may help solve this problem (unpublished).

A double-digest-based physical map for the soybean genome is now nearing completion. For this purpose, soybean BACs from five libraries belonging to three cultivars were fingerprinted and assembled [98] using a moderate information content fingerprint method (MICF) and FPC. The available BACs presently include 1182 Faribault BACs (~130 kbp, *EcoRI* inserts, **0.125x**), 860 Williams 82 BACs (~130 kbp, *HindIII* inserts, 0.1x) and 78 001 Forrest BACs that were selected from the three libraries (125–157 kbp *EcoRI*, *HindIII*, and *BamHI* inserts, 9x). Cultivar sequence variation did not appear to cause incorrect binning of BACs by FPC. However, the first release (build 3) [98] had many problems (Table 6), since many individual contigs appeared to contain noncontiguous genomic regions, and in some cases, different contigs contained the same region of the genome. Also, the available set of contigs encompassed a space that was 300 Mbp more than the size of the soybean genome. Clone contamination caused many of these problems, so that new methods to identify and eliminate contaminated clones were developed [99].

Subsequently, the publicly available soybean BAC fingerprint database was used to create build 4 [16] with the following specific aims: (i) to increase the number of genetic markers in the map, (ii) to reduce the frequency of clone contamination, (iii) to rebuild the physical map at high stringency, (iv) to examine clone density per contig, and (v) to examine the effectiveness of the generic genome browser in representing duplicated homoeologous regions (Table 6). Clones suspected of contamination were listed, **fingerprints were examined, and contaminated clones** removed from the FPC database. Many (7134 about 10%) well-to-well contaminated clones were removed from the fingerprint database. The edited database produced 2854 contigs and encompassed 1050 Mbp. In addition, homoeologous regions that might cause separate contigs to coalesce were detected in several ways. First, contigs with high clone density (23%) were inferred to represent two copy (240) or four copy (406) conserved genomic regions per haploid genome (Table 6). If the polyploid regions could all be split using HSVs (Figure 1) [29], there would be 1624 regions with two copies and 480 regions with four copies in the soybean genome. A second proof of this genome structure was that pairs of separate contigs that contained the same marker anchors (69%) were inferred to represent homoeologous but diverged genomic regions (Figure 6) [16]. A third proof came from EST hybridizations to BAC libraries where gene families with 1, 2, 4, and 8 members were more common than those with 3 or 5 members [57]. Finally, similarity search within the whole genome sequence at 90% similarity showed that the sequences that map to the contigs with duplicated regions do

TABLE 5: Progress in the soybean physical map builds 2 to 5.

	Automated Build 2 Sept. 2001	Manual Build 3 Oct 2002	Manual Build 4 Oct 2003 Total	Judged by BACs/unique band to be (pploid) [unique]	Manual Build 5 Jan 2008
BAC clones in FPC database	81,024	83,026	78,001		78,001
BACs used in contig assembly	75,568	78,001	72,942		72,837
Number of singletons	5,884	4954	27,1812		17,942
Marker anchored singletons	0	0	120		63
Clone in contigs (fold genome)	69,684	73,069	45,135		58,765
Fold genome in contigs	8.7	9.1	5.6		62
Number of contigs	5,488	2,907	2,854	(646) [2208]	521
Anchoring Markers	0	385	404	(280) [124]	1,523
Anchored Contigs	0	781	742	(181) [223]	455
Contlgs contain: >25 clone	220	921	477	(268) [209]	335
10–25 clones	3,038	920	1,458	(433) [1025]	110
3–9 clones	1,845	850	820	(0) [820]	43
2 clones	385	216	99	(0) [99]	33
Unique bands in the contigs	396,843	345,457	#258,240	(64,560)	257,356
Length of the contigs (Mb)	1,667*	1,451*	1.037	(0.258) [0.769]	1.034

* Based on 4.00 kbp per unique hand. # Based on 4.05 kbp per unique band, for 2,854 contigs containing ~68 unique bands in 15 clones, 264 duplicated region contigs containing ~68 unique bands in 30 clones 15,840 unique bands and 406 highly repeated region contigs containing ~68 unique bands in 60 clones, 48,720 unique bands.

TABLE 6: Summary of sequence coverage of the three minimum tile paths (MTPs) used for BAC end sequencing made from three BAC libraries. To calculate the percentage of the soybean genome covered by the clones (clone coverage) in our *EcoRI*- (MTP4E) and *BamHI* or *HindIII* insert libraries (MTP2BH and MTP4BH), the genome size of soybean was assumed to be 1130 Mb. The BAC libraries were each constructed from DNA derived from twenty five seedlings of an inbred cultivar Forrest.

	MTP4E	MTP4BH	MTP2 BH	Totals
Vector	pBeloBAC11	pCLD04541		na
Insertion site	<i>EcoRI</i>	<i>BamHI</i> or <i>HindIII</i>		na
Duplicates/region	1	1	1–4	na
Number of clones	3840	4608	576	17 088
Mean insert size (kbp)	175 ± 7	173 ± 7	140 ± 5	685
Clone coverage	0.7	0.8	1.4	3.1
BESs good reads	3 324	6772	924	25 123
BESs coverage (Mbp)	2.9	5.0	0.7	18.5

have homoeologs in the sequence, whereas sequences from single copy regions do not (Figure 2) [29, 93].

To deal with duplicated regions, SoyGD was adapted to distinguish homoeologous regions by showing each contig at all potential anchor points, spread laterally, rather than as overlapping [16]. Therefore, it should be realized that the genes in such regions have duplicates in other regions of the genome (Figure 6). This information will prove useful in future for gene isolation by positional cloning following a reverse genetics approach, where aneuploidy pathways regularly cause wide-spread failures [100–102] due to inability to predict phenotypes reliably.

In build 5, DNA sequence scaffolds (unpublished) have been used to cluster groups of neighboring contigs. This, however, does not solve the problems faced due to genome duplication. In many cases, (60–80%), homoeologous vari-

ants may help separation of coalesced regions [29], but this would require **BESs** for every fingerprinted BAC clone. In a minority of regions (20–40%), sequences longer than **BESs** may be needed to correctly separate BAC clones into contigs.

4.2. Minimum tile paths

The creation of minimally redundant tile paths (MTP) from contiguous sets of overlapping clones (contigs) in physical maps is a critical step for structural and functional genomics [95]. The first minimum tiling path (MTP) developed (from builds 2 and 3) contained 2 fold redundancy of the haploid genome (2,100 Mbp). MTP2 was 14 208 clones (mean insert size 140 kbp) that were picked from the 5597 contigs of build 2. MTP2 was constructed from three BAC libraries (*BamHI* (B), *HindIII* (H) and *EcoRI* (E) inserts), encompassing the

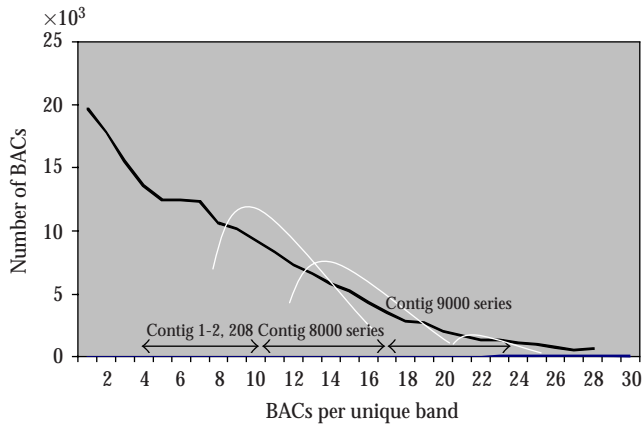


FIGURE 5: Quality estimate for the physical maps build 4 showing measurements of BAC clones per unique band. Three sets of distributions were inferred, representing the diverged DNA and the conserved DNA following the two genome duplications (shown as white lines). The 2208 single copy contigs (labeled 1–3500 after merges and splits) encompassed diverged DNA and are each inferred to contain clones from a single region. Contigs in the 8000 series are inferred to contain clones from two homoeologous regions. Contigs in the 9000 series are inferred to contain clones from four homoeologous regions. Clearly, some contigs in each set will be missplaced, hybrid contigs will occur, and ranges will overlap.

10 contigs of build 3 that were derived from build 2 by a series of contig merges, but does not distinguish regions by degree of duplication, so that many regions are redundant. The
 11 MTP2 is used in two parts, MTP2BH and MTP2E (Table 6) because they are largely redundant and overlap each other. Also, the vectors differ in the antibiotic resistance conferred. Consequently, only the MTP2BH was used for development of EST map [57].

The third and fourth MTPs, called MTP4BH and MTP4E (Table 6), were each based on build 4 [95]. Each was selected as a single path through each of the 2854 contigs. MTP4BH had 4608 clones with a mean size 173 kbp in the large (27.6 kbp) T-DNA vector pCLD04541, which is suitable for plant transformation and functional genomics. Plates 1–8 contained clones from the contigs belonging to the single copy regions of the genome. Plates 9 and 10 were picked from the duplicated and quadruplicated regions without redundancy, so that an individual clone represented either 2 or 4 regions per haploid genome. Plates 11 and 12 contained the marker anchored clones also used in MTP2BH. Plate 13 of MTP4BH was developed from just 6 contigs from regions with four copies by redundant picking. This set of clones should resolve into 48 regions, if methods to separate them can be developed as the genome sequencing is completed [93]. This set of 13 plates was used for HICF fingerprinting by the same methods that were used for Williams 82 [11] and PI437654 BACS [79, 96]. The BACs used for HICF will form a bridge to other physical maps and a resource to test the ability of HICF to correctly separate duplicated regions, particularly in the contigs in plate 13.

MTP4E was designed to be 4608 BAC clones with large inserts (mean 175 kbp) in the small (7.5 kbp) pECBAC1 vector [57, 85]. However, only 3840 clones were picked to date. Sequencing efficiency was low on this MTP and reracking will be needed [103]. The vector is suitable for DNA sequencing and these clones will be used for sequencing across gaps in the WGS sequence.

MTP4BH and MTP4E clones each encompassed about 800 Mbp before duplicate regions were considered. The single copy regions represented 700 Mbp [57]. In addition there were 50 Mbp from the duplicate and 50 Mbp from the quadruplicate regions in the MTP. Because those regions were duplicate and quadruplicate they encompass another 300 Mbp in total. MTP2BH, MTP4E, and MTP4BH were each used for BAC-end sequencing and microsatellite integration into the physical map [27, 39]. MTP2BH was used for EST integration to the physical map [16, 57]. MTP4BH was used for high information content fingerprinting for integration with the Williams 82 physical map [11, 104]. In conclusion, it appears like each MTP and the derived **BESs** will be useful to deconvolute and finish the whole genome shotgun sequence of soybean while the whole genome sequence will help complete the physical map. A complete MTP5BH would be a useful tool for functional genomics because clones from these libraries were constructed in a T-DNA vector and are ready for plant transformation. About four thousand transgenic lines made from BACs would be enough to transfer every soybean sequence to another plant.

4.3. BAC end sequences (**BESs**)

BAC end sequences (**BESs**) anchored to a robust physical map constitute an important tool for genome analysis, and have been developed from BACs belonging to three available MTPs including MTP2BH, MTP4BH, and MTP4E [95, 103]. Therefore, three sets of **BESs** were available, of which the first set consisted of 13 474 good **BESs** derived from 8064 clones of MTP2BH (Table 5). Enquiries to GenBank nr and pat databases identified 7260 potentially genic homologs, and an analysis of the locations of inferred genes suggested presence of gene-rich islands on each chromosome [37]. In addition, 42 **BESs** showed homology (extending over a length of 80–341 bp at e^{-30} to e^{-300}) with DNA markers (10 RFLPs, 20 microsatellites) that were already genetically mapped [95]. This amounts to homology with about 2% of the markers, whose sequences are available in GenBank. Available **BESs** also carried as many as 1053 new SSR markers [27, 37] that are described further in the next section.

The second set of **BESs** consisted of 7700 good **BESs** reads from clones of MTP4BH (Table 5) of which 4147 had homologs in the GenBank nr and pat databases [57]. The clones in plates 11 and 12 were resequenced and so have 2 records for each BAC end in GenBank. Resequenced clones help determine the sequence error rate and greatly facilitate SNP detection [18, 19]. Twenty additional genetic anchors were detected in this second set of **BESs** (6 RFLPs, 14 microsatellites), which represented about 1% of the soybean markers with sequences in GenBank. This second set of **BESs** carried 625 SSR markers [27, 37] that are described further

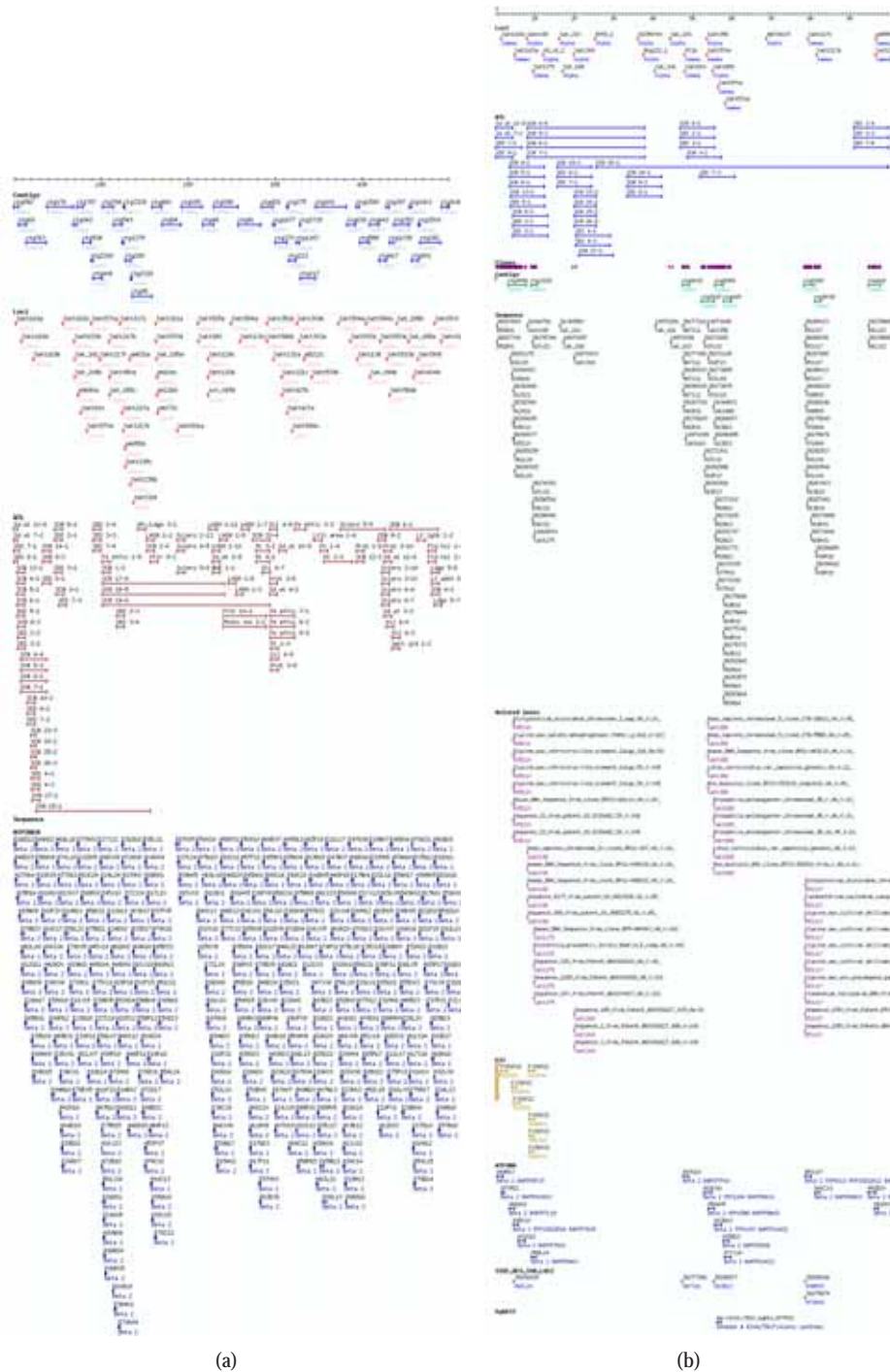


FIGURE 6: Description of chromosome 18 resources at SoyGD (a). The current GMOD representation of 50 Mbp of the 51.5 Mbp chromosome 18 (linkage group G) in SoyGD (a). shows the build 3 version of the chromosome (cursor), anchored contigs (top row, blue), DNA markers (second row of features, red), QTL in the region (third row, burgundy), MTP2 clones (B, H, and E fourth row, dark blue). Not shown here were BAC clones, ESTs, BAC end sequences, and gene models (b) shows the build 4 representation of 10 Mbp of the 51.5 Mbp chromosome 18 in SoyGD. Shown are the chromosome (cursor), DNA markers (top row of features, red); QTL in the region (second row, blue); coalesced clones (purple) comprising the anchored contigs (third row, green); BAC end sequences (fourth row black); **BESS** encoding gene fragments (fifth row, puce); EST hybridizations to MTP2BH (sixth row gold); MTP4BH clones (seventh row, dark blue); **BESS** derived SSR (eighth row, green); EST hybridizations inferred on build 4 from clones also in MTP2BH (ninth row, blue); WGS trace file matches from MegaBlast (tenth and last row, light blue). It is recommended for readers to visit updated site <http://bioinformatics.siu.edu> to see a full detailed color version and a build 5 view. The gaps between contigs will be filled in build 5 by contig merges suggested by **BESS**-SSRs and contig end overlap data.

in the next section. The third set of **BESs** from MTP4E have recently been released and are only partly analyzed (Table 6).

The above builds of physical map representing recently duplicated regions of the genome can be further improved with existing databases and tools. In particular, this can be achieved by increasing the number of reliable genetic anchors derived from **BESs** [27, 37] and separating BACs from homoeologous regions with diagnostic SNPs (Figure 2) before contigs were formed [93].

4.4. Genetic map and SSR markers derived from **BESs**

The molecular genetic map for soybean genome can be improved further through several approaches including (i) addition of **BESs** markers on the available genetic map [27, 37], (ii) bioinformatics analysis of contig data [16] and (iii) through the use of novel approaches to error detection [99]. The composite genetic map of soybean at SoyGD (in 2007) contained 3073 DNA markers [16, 27], which included 1019 class I SSRs, each with >10 di- or trinucleotide repeat motifs (BARC-SSR markers; Song et al., 2004), and a few class II SSRs with <10 di- to pentanucleotide repeats that were mostly SIUC-SSR markers. Forrest **BESs** helped in increasing the number of class I and II SSR markers for the soybean genome, and allowed integration of BAC clones into the soybean physical map.

SSRs were mined separately from the two sets of **BESs** described above. As mentioned above, the first set of 10 Mbp of BAC end sequences (**BESs**) derived from 13 474 reads of 7050 clones constituting MTP2BH, had 1053 SSRs (333 class I + 720 from class II), and the second set of 5.7 Mbp **BESs** derived from 7700 reads from 5184 clones constituting MTP4BH, had 620 SSRs (150 class I + 480 class II). Potential markers are shown on the MTP_SSR track at SoyGD (Figure 6). About 530 primer pairs were designed for both the sets of SSRs. These primers were 20–24 mers long with a T_m of $55 + 1^\circ\text{C}$, and provided amplicons that were 100–500 bp long. As many as 123 of these primers belonging to duplicated regions gave multiple amplified products, and therefore should be avoided.

Different possible motifs were not randomly distributed among the above SSRs, with AT rich motifs being more frequent [27]. Compound SSRs having tetranucleotide repeats clustered with di- and trinucleotide motifs were also found. About 75% of class I and 60% of class II SSR markers were polymorphic among the parents of four recombinant inbred line (RIL) populations. Most of the **BESs**-SSRs were located on the soybean genetic map in regions with few BARC-SSR markers [27, 39]. Therefore, **BESs**-SSRs represent a useful tool for the improvement of the genetic map of soybean.

4.5. SNP markers derived from **BESs** to WGS

The soybean genome has been shown to be composed of ~8000 short interspersed regions of one, two, or four copies per haploid genome, as shown by RFLP analysis, SSR anchors to BACs and by BAC fingerprints [16]. Recently, the genome has been sequenced by WGS sequencing of 4 kbp inserts in pUC18 [105]. When the extent and homogeneity

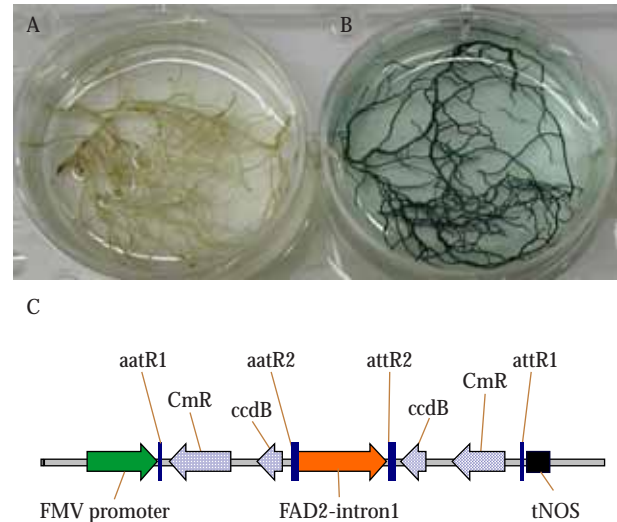


FIGURE 7: Evidence for RNAi silencing of GUS gene in 35S::GUS soybean plants. Panel A. GUS expression in composite plant roots expressing and RNAi from the gene encoding GFP. Panel B. GUS expression in composite plant roots expressing RNAi from the gene encoding GUS. Panel C. The transformation cassette used (thanks to Dr. C. G. Taylor, Danforth Center, unpublished data).

of duplications within contigs was examined using BAC end sequences (**BESs**) derived from minimum tile paths (MTP2BH and MTP4BH; Figure 2) [29], a strong correlation was found between the fold of duplication inferred from fingerprinting and that inferred from WGS matches. Duplicated regions were identified by BAC fingerprint contig analysis using a criterion of less than 10% mismatch across a trace with a window size of 60 bp. Previously, simulations had predicted that fingerprints of clones from different regions would coalesce, if sequence variation was less than 2%. Hopefully, the HSVs among contigs from duplicated regions can be used to separate clone sets from different regions. Ironically, improvements for contig building methods will result from the whole genome sequence! However, many duplicated regions with less than 1% sequence divergence were found [29, 93]. The implication for bioinformatics and functional annotation of the soybean genome (and other paleopolyploid or polyploid genomes) is that reverse genetics with many genes will be nearly impossible without tools to simultaneously repress or mutate several gene family members.

5. FUNCTIONAL GENOMICS TOOLS

Unequivocal identification and map-based cloning of genes underlying quantitative traits have been a challenge for soybean genomics research. Gene redundancy, gene action, and low transformation efficiencies seriously hampered positional cloning [16]. Therefore, a variety of approaches need to be used for soybean functional genomics research. Two major areas of soybean genomics research include (i) annotation of genomic sequences (genes with unknown functions) and (ii) analysis of genome sequences of “Forrest”

TABLE 7: Some of the BACs, mutant and nonmutant soybean lines to be transformed for complementation.

BIBAC clone names	Phenotypes	Insert size kbp	Dominant?
Gm-SIU1-B100B10	<i>Rhg4</i> bigenic resistance to SCN ^(a)	140	Yes
Gm-SIU1-B73P06	<i>rhg1</i> bigenic resistance to SCN and <i>Rfs2</i> for SDS ^(a)	79	Co-
Gm-SIU2-H050N07	<i>Rpg1-b</i> resistance to bacterial pustule ^(b)	110	Yes
Gm-SIU1-B54E07	<i>T</i> tawny pubescence; flavonoid-3-monoxygenase ^(c)	82	Yes
Gm-SIU2-H04P03	<i>W1</i> White flower and black hila color ^(d)	153	No
Gm-SIU2-H82CO8	<i>Rfs1</i> root resistance to SDS	130	Yes
Gm-SIU1-TBD	<i>Rps4</i> resistance to <i>Phytophthora</i> root rot	120	Yes

^(a) *Rhg4* and *rhg1* each encodes transmembrane receptor-like kinase. Resistant and susceptible alleles differ by 3–6 amino acid changes and 23 base changes. There are mutant lines derived from Forrest.

^(b) *Rpg1-b* encodes a nucleotide binding leucine rich repeat protein.

^(c) *T* encodes flavonoid-3 monoxygenase (EC1.13.14.21). The recessive genes differ from the dominant by deletion of a single C nucleotide. There are mutant lines.

^(d) *W* encodes an unknown enzyme, probably a glycosidase.

for synteny with the genomes of other dicotyledonous genera and with those of other soybean cultivars.

5.1. Annotation of genome sequences

The three methods that proved useful for annotation of the genome sequences of Forrest and related germplasm include (i) mutant complementation using transformation, (ii) gene silencing through RNAi, and (iii) targeted mutations. Each will be briefly discussed.

- (i) *Mutant complementation using transformation.* A popular approach for the study of gene function is mutant complementation, which involves transformation of mutants with the wild alleles. Therefore, development of transformation protocols is an essential component of functional genomics research. In soybean, *A. tumefaciens* and *A. rhizogenes*-mediated transformation of cultured cells with Forrest BAC clones has been successfully achieved using previously described protocols involving the T-DNA vector pCLD04541 [84]. In this protocol, *npt II* gene is used as a plant selectable marker, and kanamycin as used as a selective agent [106–109]. Screenable markers are available in some BAC clones (Table 7). Whole BAC transformation is important because fine maps locating loci at genetic distance of 0.25 cM that is equivalent to 50–150 kbp were earlier prepared using RILs and NILs. The clones selected for transformation are listed in Table 7, and should provide for complementation of easily scoreable phenotypes in mutants. For instance, dominant mutant phenotypes of traits like pubescence, color, and disease resistances should be evident in the very first products of transformation. BAC transformation with sets of overlapping clones will be the best approach in situations where an individual locus represents a cluster of genes [37, 38].
- (ii) *Gene silencing using RNAi.* The composite plant system for RNAi has been tested in NILs derived from Forrest, and has been validated by Dr. C. G. Taylor at the Danforth Center (St. Louis, Mo, USA) [110] through expression of gene-specific dsRNA constructs.

Using this system, shoots from stable transgenic soybean plants showing constitutive expression of *uidA* (GUS) were transformed with dsRNA constructs (Figure 7) that were designed using a modified pKannibal vector [111], with the 35S promoter replaced by the figwort mosaic virus (FMV) promoter. The 600 bp homologous sequences of the GUS or green fluorescent protein (GFP) gene were introduced in an antisense and sense orientation separated by the pKannibal intron (spacer) sequence. These constructs were designed to produce transcripts with a stem loop secondary structure that would be recognized by the plant cell machinery and activate RNAi. The dsRNA constructs placed in a binary vector, introduced into *A. rhizogenes*, were used for composite plant production [112]. GUS-specific RNAi construct silenced, while non-GUS RNAi (GFP) construct failed to silence GUS expression in hairy roots produced on shoots of transgenic soybean plants. These results show that the hairy roots can be used to produce dsRNAs. Further, the RNAi machinery in soybean hairy roots is fully functional in a sequence-specific manner. Thus, RNAi technology will allow the rapid analysis of sets of candidate genes for alleles underlying variation [38].

- (iii) *Study of gene function through TILLING.* Two soybean mutagenized M2 libraries are already available for TILLING [113], from which ~3000 of the 6000 available M2 lines were phenotyped visually. A soybean mutant database has been developed to track and sort these mutants (www.soybeantilling.org). While developing a database that would allow search for “TILLED” genes a search engine was developed, so that the database can be searched for both phenotype and gene. The mutations occurred at a rate of ~1 mutation/170 kbp, so that a screening of 6150 M2 families may provide a series of up to 40 to 60 alleles within each 1.5 kbp fragment of a target gene. This approach led to the identification of a putative mutant for a soybean leucine rich repeat receptor like kinase gene *Gm-Clavata1A* (AF197946; Figure 8). In future, TILLING and crosses among TILLED mutants [100–

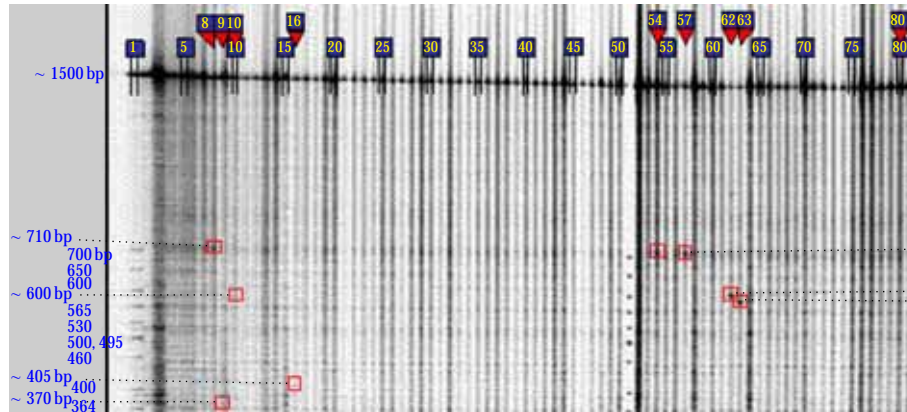


FIGURE 8: Soybean Tilling gel image of *Gm_clavata1* pool ps33 screening, representing 768 individuals, 8 individuals per pool (LI-COR 700 channel mutations are marked in red boxes; blue boxes represent corresponding band in 800 channel) from www.soybeantilling.org (thanks to Dr. K. Meksem and Dr. B. Liu SIUC, unpublished data).

102] will allow the testing of candidate genes and will provide new genetic variation that may lead to germplasm enhancement.

5.2. Analyses of conserved synteny

12 Forrest genome sequences have also been used for a study of their synteny with genomes of other dicotyledonous genera/species and also with the genomes of other soybean cultivars. For this purpose, cross-species transferable genetic markers are available in the data-based legumeDB1 [114], and can be used to compare the linear order of markers/genes, which are either species specific or conserved across genera [115–124]. For instance, genes for resistance to pathogens will often appear as new genes or gene clusters inserted in regions, which otherwise exhibit conserved synteny across genera [35, 115, 122]. Synteny extends beyond genes into repeat DNA, as exemplified by the distributions of 15 bp sequences that provide sequence-specific genome fingerprints [94]. Interestingly the fingerprints do not show the same patterns of relatedness between species found in gene sequence. Therefore, genome fingerprinting will help identify good candidates for cross species markers in repeat DNA such as microsatellite markers.

Conserved synteny has also been observed among the genomes involved in the constitution of the allo- and autotetraploids hypothesized for soybean. It has been shown that about 25–30% of the genome has extensive conservation of gene order in otherwise shuffled blocks of 150–300 kbp [16]. Consequently, blocks of 3–10 genes are repeated at 2 or 4 locations per haploid genome [38, 79]. There are also genomic regions, where synteny among genomes of different cultivars has been shown to break down. Several interesting features including the following have been observed in these nonsyntenic regions: (i) in some cases, a loss of conserved synteny between cultivars is associated with a gene introgressed from a Plant Introduction [38]. (ii) In another case, a moderately repeated sequence common in one cultivar is absent in another cultivar [29]. (iii) In still

another case, a sequence inserted in one cultivar appears to alter the expression of a neighboring gene (unpublished). It is thus apparent that genome analysis involving study of an association of these nonsyntenic sequence tracts in otherwise syntenic regions, with phenotypes will be an active area of research, when genome sequences from a number of soybean cultivars are available.

6. CONCLUSIONS

The soybean genomics resources developed through the use of cultivar Forrest have been used and will continue to be used in future leading to significant advances in soybean genomics knowledge base. The soybean genome shows evidence of a paleopolyploid origin with regions, encompassing gene-rich islands that were highly conserved following duplication [16]. In fact, it was estimated that 25–30% of the genome was highly conserved after both duplications. Implications of this feature are profound. First, a map of homoeology and an associated map of duplicated regions had to be developed. Second, an estimate of sequence conservation among the duplicated regions was necessary. Third, the implications for functional genomics were considered. Given that all soybean genes have been duplicated twice during recent evolution, and that most plant genomes encode functionally redundant pathways, it is not surprising that TILLING, RNAi-mediated silencing and overexpression of several genes often did not lead to phenotypic changes [101, 102, 110, 113]. In future, the E × F population will continue to be used for (i) an analysis of functions of a number of gene families, (ii) patenting of inventions based on useful genes [6, 77, 124–126], (iii) manipulation of soybean seed composition including protein, oil [19] and bioactive factors [127–129], and (iv) an analysis of the protein interactome [130]. In summary, the newly released E × F population and the other associated genomic resources developed through the use of cultivar “Forrest” will provide tremendous opportunities for further research in the field of genomics research.

ACKNOWLEDGMENTS

This research summarized in this review was funded in part by grants from the NSF 9872635, ISA (95-122-04; 98-122-02; 02-127-03), and USB 2228-6228. The physical map was based upon work supported by the National Science Foundation under Grant no. 9872635. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The continued support of SIUC, College of Agriculture, and Office of the Vice Chancellor for Research to the Genomics Facility is appreciated. The author thanks Dr. Q. Tao and Dr. H. B. Zhang for assistance with fingerprinting; Dr. C. Town and Dr. C. Foo at TIGR for the **BESS**; Dr. K. Meksem for the Tilling figure; Dr. C. G. Taylor for an RNAi figure; Dr. Z. Zhang for transformation of Forrest; and the SIUC team for their rigor in addressing a novel problem for genomics. All team members are thanked for their continued collaborations. Thanks are also due to P. K. Gupta, the guest editor, for critical readings, which led to significant improvement of the manuscript.

REFERENCES

- [1] E. E. Hartwig and J. M. Epps, "Registration of 'Forrest' soybeans," *Crop Science*, vol. 13, no. 2, p. 287, 1973.
- [2] G. J. Carbone, L. O. Mearns, T. Mavromatis, E. J. Sadler, and D. Stooksbury, "Evaluating CROPGRO-soybean performance for use in climate impact studies," *Agronomy Journal*, vol. 95, no. 3, pp. 537-544, 2003.
- [3] O. Myers Jr. and S. C. Anand, "Inheritance of resistance and genetic relationships among soybean plant introductions to races of soybean cyst nematode," *Euphytica*, vol. 55, no. 3, pp. 197-201, 1991.
- [4] R. R. Prabhu, V. N. Njiti, B. Bell-Johnson, et al., "Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers," *Crop Science*, vol. 39, no. 4, pp. 982-987, 1999.
- [5] S. Kazi, J. L. Shultz, R. Bashir, et al., "Identification of loci underlying seed yield and resistance to soybean cyst nematode race 2 in 'Hartwig,'" *Theoretical and Applied Genetics*. In press.
- [6] D. A. Lightfoot, K. Meksem, and P. T. Gibson, "Soybean sudden death syndrome resistant soybeans, soybean cyst nematode resistant soybeans and methods of breeding and identifying resistant plants," DNA markers, US Patent no. 6300541, October 2001.
- [7] N. Hnetkovsky, S. J. C. Chang, T. W. Doubler, P. T. Gibson, and D. A. Lightfoot, "Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS)," *Crop Science*, vol. 36, no. 2, pp. 393-400, 1996.
- [8] S. J. C. Chang, T. W. Doubler, V. Kilo, et al., "Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS)," *Crop Science*, vol. 36, no. 6, pp. 1684-1688, 1996.
- [9] S. J. C. Chang, T. W. Doubler, V. Y. Kilo, et al., "Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3," *Crop Science*, vol. 37, no. 3, pp. 965-971, 1997.
- [10] D. A. Lightfoot, V. N. Njiti, P. T. Gibson, M. A. Kassem, J. M. Iqbal, and K. Meksem, "Registration of Essex x Forrest recombinant inbred line mapping population," *Crop Science*, vol. 45, no. 4, pp. 1678-1681, 2005.
- [11] S. A. Jackson, D. Rokhsar, G. Stacey, R. C. Shoemaker, J. Schmutz, and J. Grimwood, "Toward a reference sequence of the soybean genome: a multiagency effort," *Crop Science*, vol. 46, supplement 1, pp. S55-S61, 2006.
- [12] J. H. Orf, K. Chase, T. Jarvik, et al., "Genetics of soybean agronomic traits—I: comparison of three related recombinant inbred populations," *Crop Science*, vol. 39, no. 6, pp. 1642-1651, 1999.
- [13] N. Yamanaka, S. Ninomiya, M. Hoshi, et al., "An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion," *DNA Research*, vol. 8, no. 2, pp. 61-72, 2001.
- [14] Y. S. Dong, L. M. Zhao, B. Liu, Z. W. Wang, Z. Q. Jin, and H. Sun, "The genetic diversity of cultivated soybean grown in China," *Theoretical and Applied Genetics*, vol. 108, no. 5, pp. 931-936, 2004.
- [15] J. Yuan, V. N. Njiti, K. Meksem, et al., "Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance," *Crop Science*, vol. 42, no. 1, pp. 271-277, 2002.
- [16] J. L. Shultz, D. Kurunam, K. Shopinski, et al., "The soybean genome database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*," *Nucleic Acids Research*, vol. 34, database issue, pp. D758-D765, 2006.
- [17] P. M. Gresshoff, J. Stiller, T. L. D. McGuire, et al., "Integrating functional genomics to define the plants function in symbiotic nodulation," in *Nitrogen Fixation: Global Perspectives*, T. Finan, M. O'Brien, D. Layzell, K. Vessey, and W. Newton, Eds., pp. 95-98, CAB International, New York, NY, USA, 2001.
- [18] C. R. Yesudas, J. L. Shultz, and D. A. Lightfoot, "Identification of loci underlying resistance to Japanese beetle herbivory, in soybean," *Theoretical and Applied Genetics*. In press.
- [19] C. R. Yesudas, R. Bashir, J. Shultz, S. Kazi, and D. A. Lightfoot, "QTL for seed isoflavones, protein, oil and Japanese beetle (*Popillia japonica*, Newman) resistance in soybean [*Glycine max* (L.) Merr.]," in *Plant & Animal Genome XV Conference*, p. 411, San Diego, Calif, USA, January 2007.
- [20] K. Meksem, P. Pantazopoulos, V. N. Njiti, L. D. Hyten, P. R. Arelli, and D. A. Lightfoot, "'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci," *Theoretical and Applied Genetics*, vol. 103, no. 5, pp. 710-717, 2001.
- [21] V. N. Njiti, K. Meksem, D. A. Lightfoot, W. J. Banz, and T. A. Winters, "Molecular markers of phytoestrogen content in soybeans," *Journal of Medicine and Food*, vol. 2, pp. 165-167, 2000.
- [22] K. Meksem, V. N. Njiti, W. J. Banz, et al., "Molecular markers of phytoestrogen content in soybeans," *Journal of Biomedicine and Biotechnology*, vol. 1, no. 1, pp. 38-44, 2001.
- [23] M. A. Kassem, K. Meksem, M. J. Iqbal, et al., "Definition of soybean genomic regions that control seed phytoestrogen amounts," *Journal of Biomedicine and Biotechnology*, vol. 2004, no. 1, pp. 52-60, 2004.
- [24] M. A. Kassem, J. L. Shultz, K. Meksem, et al., "An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits," *Theoretical and Applied Genetics*, vol. 113, no. 6, pp. 1015-1026, 2006.
- [25] Z. Gizlice, T. E. Carter Jr., T. M. Gerig, and J. W. Burton, "Genetic diversity patterns in North American public

- soybean cultivars based on coefficient of parentage," *Crop Science*, vol. 36, no. 3, pp. 753–765, 1996.
- [26] P. Keim, W. Beavis, J. Schupp, and R. Freestone, "Evaluation of soybean RFLP marker diversity in adapted germplasm," *Theoretical and Applied Genetics*, vol. 85, no. 2-3, pp. 205–212, 1992.
- [27] J. L. Shultz, J. S. Kazi, R. Bashir, A. J. Afzal, and D. A. Lightfoot, "Development of microsatellite markers from the soybean sequence ready physical map," *Theoretical and Applied Genetics*, vol. 114, no. 6, pp. 1081–1090, 2007.
- [28] Y. L. Zhu, Q. J. Song, D. L. Hyten, et al., "Single-nucleotide polymorphisms in soybean," *Genetics*, vol. 163, no. 3, pp. 1123–1134, 2003.
- [29] D. A. Lightfoot, J. Shultz, and N. Saini, "Reannotation of the physical map of *Glycine max* for ploidy by BAC end sequence driven whole genome shotgun read assembly," in *Proceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP '07)*, pp. 65–72, Las Vegas, Nev, USA, June 2007.
- [30] M. J. Iqbal, K. Meksem, V. N. Njiti, M. A. Kassem, and D. A. Lightfoot, "Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex x Forrest RILs," *Theoretical and Applied Genetics*, vol. 102, no. 2-3, pp. 187–192, 2001.
- [31] K. Meksem, E. Ruben, D. L. Hyten, M. E. Schmidt, and D. A. Lightfoot, "High-throughput genotyping for a polymorphism linked to soybean cyst nematode resistance gene *Rhg4* by using TaqmanTM probes," *Molecular Breeding*, vol. 7, no. 1, pp. 63–71, 2001.
- [32] V. N. Njiti, J. E. Johnsona, T. A. Torto, L. E. Gray, and D. A. Lightfoot, "Inoculum rate influences selection for field resistance to soybean sudden death syndrome in the greenhouse," *Crop Science*, vol. 41, no. 6, pp. 1726–1731, 2001.
- [33] V. N. Njiti, K. Meksem, M. J. Iqbal, et al., "Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 294–300, 2002.
- [34] M. J. Iqbal, S. Yaegashi, V. N. Njiti, R. Ahsan, K. L. Cryder, and D. A. Lightfoot, "Resistance locus pyramids alter transcript abundance in soybean roots inoculated with *Fusarium solani* f. sp. *glycines*," *Molecular Genetics and Genomics*, vol. 268, no. 3, pp. 407–417, 2002.
- [35] T. Ashfield, A. Bocian, D. Held, et al., "Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes," *Molecular Plant-Microbe Interactions*, vol. 16, no. 9, pp. 817–826, 2003.
- [36] M. A. Kassem, K. Meksem, C. H. Kang, et al., "Loci underlying resistance to manganese toxicity mapped in a soybean recombinant inbred line population of 'Essex' x 'Forrest'," *Plant and Soil*, vol. 260, no. 1-2, pp. 197–204, 2004.
- [37] K. Triwitayakorn, V. N. Njiti, M. J. Iqbal, S. Yaegashi, C. Town, and D. A. Lightfoot, "Genomic analysis of a region encompassing *QRfs1* and *QRfs2*: genes that underlie soybean resistance to sudden death syndrome," *Genome*, vol. 48, no. 1, pp. 125–138, 2005.
- [38] E. Ruben, J. Jamai, A. J. Afzal, et al., "Genomic analysis of the *Rhg1* locus: candidate genes that underlie soybean resistance to the cyst nematode," *Molecular Genetics and Genomics*, vol. 276, no. 6, pp. 503–516, 2006.
- [39] R. Bashir, "Minimum tile derived microsatellite markers improve the physical map of the soybean genome and the Essex by Forrest genetic map," M.S. thesis, Southern Illinois University at Carbondale, Carbondale, III, USA, 2007.
- [40] K. Meksem, T. W. Doubler, K. Chanchaoenchai, et al., "Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines," *Theoretical and Applied Genetics*, vol. 99, no. 7-8, pp. 1131–1142, 1999.
- [41] V. N. Njiti and D. A. Lightfoot, "Genetic analysis infers *Dt* loci underlie resistance to *Fusarium solani* f. sp. *glycines* in indeterminate soybeans," *Canadian Journal of Plant Science*, vol. 86, no. 1, pp. 83–90, 2006.
- [42] A. Sharma, "The identification of QTL underlying resistance to aluminum toxicity in a soybean recombinant inbred line population of Essex x Forrest," M.S. thesis, Southern Illinois University at Carbondale, Carbondale, III, USA, 2007.
- [43] V. N. Njiti, T. W. Doubler, R. J. Suttner, L. E. Gray, P. T. Gibson, and D. A. Lightfoot, "Resistance to soybean sudden death syndrome and root colonization by *Fusarium solani* f. sp. *glycine* in near-isogenic lines," *Crop Science*, vol. 38, no. 3, pp. 472–477, 1998.
- [44] M. A. Kassem, K. Meksem, A. J. Wood, and D. A. Lightfoot, "A microsatellite map developed from late maturity germplasm 'Essex' by 'Forrest' detects four QTL for soybean seed yield expected from early maturing germplasm," *Reviews in Biology & Biotechnology*, vol. 6, no. 1, pp. 7–12, 2007.
- [45] A. Alcivar, J. Jacobson, J. Rainho, K. Meksem, D. A. Lightfoot, and M. A. Kassem, "QTL underlying seedling root traits mapped in the 'Essex' by 'Forrest' soybean RIL population," *Annals of Botany*. In press.
- [46] A. Alcivar, J. Jacobson, J. Rainho, K. Meksem, D. A. Lightfoot, and M. A. Kassem, "Genetic analysis of soybean plant height, hypocotyl and internode lengths," *Journal of Agricultural, Food, and Environmental Sciences*, vol. 1, no. 1, pp. 40–50, 2007.
- [47] Y. Cho, V. N. Njiti, X. Chen, et al., "Quantitative trait loci associated with foliar trigonelline accumulation in *Glycine max* L," *Journal of Biomedicine and Biotechnology*, vol. 2, no. 3, pp. 1–7, 2002.
- [48] J. A. Afzal, S. Fasi, R. Mungur, D. B. Goodenowe, and D. A. Lightfoot, "Comparisons of metabolic profiles by FT-ICR-MS and GC-MS using near-isolines that contrast for resistance to SCN and SDS," in *Plant & Animal Genome XV Conference*, San Diego, Calif, USA, January 2007.
- [49] A. L. de Farias Neto, R. Hashmi, M. E. Schmidt, et al., "Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes 'PI 567374' and 'Ripley'," *Molecular Breeding*, vol. 20, no. 1, pp. 53–62, 2007.
- [50] S. Kazi, V. N. Njiti, T. W. Doubler, et al., "Registration of the flyer x hartwig recombinant inbred line mapping population," *Journal of Plant Registrations*, vol. 1, no. 2, pp. 175–178, 2007.
- [51] Soybean Genetics Committee Report, <http://soybase.agron.iastate.edu/resources/QTL.php>.
- [52] S. Kazi, J. L. Shultz, R. Bashir, A. J. Afzal, V. N. Njiti, and D. A. Lightfoot, "Identification of loci underlying resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer,'" *Theoretical and Applied Genetics*. In press.
- [53] V. N. Njiti, R. J. Suttner, L. E. Gray, P. T. Gibson, and D. A. Lightfoot, "Rate-reducing resistance to *Fusarium solani* f. sp. *phaseoli* underlies field resistance to soybean sudden death syndrome," *Crop Science*, vol. 37, no. 1, pp. 132–138, 1997.
- [54] Y. Luo, O. Myers Jr., D. A. Lightfoot, and M. E. Schmidt, "Root colonization of soybean cultivars in the field by

- Fusarium solani* f. sp. *glycines*," *Plant Disease*, vol. 83, no. 12, pp. 1155–1159, 1999.
- [55] M. J. Iqbal, A. J. Afzal, S. Yaegashi, et al., "A pyramid of loci for partial resistance to *Fusarium solani* f. sp. *glycines* maintains myo-inositol-1-phosphate synthase expression in soybean roots," *Theoretical and Applied Genetics*, vol. 105, no. 8, pp. 1115–1123, 2002.
- [56] M. J. Iqbal, S. Yaegashi, R. Ahsan, K. L. Shopinski, and D. A. Lightfoot, "Root response to *Fusarium solani* f. sp. *glycines*: temporal accumulation of transcripts in partially resistant and susceptible soybean," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1429–1438, 2005.
- [57] K. L. Shopinski, M. J. Iqbal, J. L. Shultz, D. Jayaraman, and D. A. Lightfoot, "Development of a pooled probe method for locating small gene families in a physical map of soybean using stress related paralogues and a BAC minimum tile path," *Plant Methods*, vol. 2, pp. 20–28, 2006.
- [58] Y. Cho, D. A. Lightfoot, and A. J. Wood, "Survey of trigonelline concentrations in salt-stressed leaves of cultivated *Glycine max*," *Phytochemistry*, vol. 52, no. 7, pp. 1235–1238, 1999.
- [59] K. Meksem, E. Ruben, D. Hyten, K. Triwitayakorn, and D. A. Lightfoot, "Conversion of AFLP bands into high-throughput DNA markers," *Molecular Genetics and Genomics*, vol. 265, no. 2, pp. 207–214, 2001.
- [60] B. B. Bell-Johnson, G. R. Garvey, J. E. Johnson, K. Meksem, and D. A. Lightfoot, "Methods for high-throughput marker assisted selection for soybean," *Soybean Genetics Newsletter*, vol. 25, pp. 115–118, 1998.
- [61] D. A. Lightfoot and M. J. Iqbal, "Marker assisted selection for soybean," in *Agricultural Biotechnology and Genomics*, V. Krishna, Ed., pp. 15–51, American Scientific, Stevenson Ranch, Calif, USA, 2003.
- [62] M. J. Iqbal and D. A. Lightfoot, "Application of DNA markers: soybean improvement," in *Molecular Marker Systems in Plant Breeding and Crop Improvement*, L. Horst and W. Gerhard, Eds., p. 475, Springer, New York, NY, USA, 2004.
- [63] M. E. Schmidt, R. J. Suttner, J. H. Klein III, P. T. Gibson, D. A. Lightfoot, and O. Myers Jr., "Registration of LS-G96 soybean germplasm resistant to soybean sudden death syndrome (SDS) and soybean cyst nematode race 3," *Crop Science*, vol. 39, no. 2, p. 598, 1999.
- [64] B. A. McBlain, R. J. Fioritto, S. K. St. Martin, et al., "Registration of 'Flyer' soybean," *Crop Science*, vol. 30, no. 2, p. 425, 1990.
- [65] G. Stacey, A. Dorrance, H. Nguyen, et al., "SoyCAP: roadmap for soybean translational genomics," white paper. USB, USDA, Beltsville, Md, USA, 2004.
- [66] H.-J. Wang and P. A. Murphy, "Isoflavone composition of American and Japanese soybeans in Iowa: effects of variety, crop year, and location," *Journal of Agricultural and Food Chemistry*, vol. 42, no. 8, pp. 1674–1677, 1994.
- [67] V. S. Primomo, V. Poysa, G. R. Ablett, C.-J. Jackson, M. Gijzen, and I. Rajcan, "Mapping QTL for individual and total isoflavone content in soybean seeds," *Crop Science*, vol. 45, no. 6, pp. 2454–2464, 2005.
- [68] J. E. Specht, D. J. Hume, and S. V. Kumudini, "Soybean yield potential—a genetic and physiological perspective," *Crop Science*, vol. 39, no. 6, pp. 1560–1570, 1999.
- [69] J. E. Specht, "Report of yield QTL—coordination topic," in *Soybean Breeders' Workshop*, St. Louis, Mo, USA, February 2005.
- [70] V. C. Concibido, B. La Vallee, P. Mcclair, et al., "Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars," *Theoretical and Applied Genetics*, vol. 106, no. 4, pp. 575–582, 2003.
- [71] J. W. Burton, "Quantitative genetics: results relevant to soybean breeding," in *Soybeans: Improvement, Production, and Uses*, J. R. Wilcox, Ed., pp. 211–247, ASA, CSSA, and SSSA, Madison, Wis, USA, 2nd edition, 1987.
- [72] D. Sun, Y. Du, Z. Zhang, et al., "GXE interactions and development influence loci underlying seed quality traits in soybean," *Theoretical and Applied Genetics*. In press.
- [73] J. A. Wrather, S. S. Koenning, and T. R. Anderson, "Effect of diseases on soybean yields in the United States and Ontario (1999 to 2002)," *Plant Health Progress*, pp. 1–16, 2003.
- [74] D. Sandhu, H. Gao, S. Cianzio, and M. K. Bhattacharyya, "Deletion of a disease resistance nucleotide-binding-site leucine-rich-repeat-like sequence is associated with the loss of the *Phytophthora* resistance gene *Rps4* in soybean," *Genetics*, vol. 168, no. 4, pp. 2157–2167, 2004.
- [75] V. C. Concibido, B. W. Diers, and P. R. Arelli, "A decade of QTL mapping for cyst nematode resistance in soybean," *Crop Science*, vol. 44, no. 4, pp. 1121–1131, 2004.
- [76] T. L. Niblack, P. R. Arelli, G. R. Noel, et al., "A revised classification scheme for genetically diverse populations of *Heterodera glycines*," *Journal of Nematology*, vol. 34, no. 4, pp. 279–288, 2002.
- [77] K. Meksem and D. A. Lightfoot, "Novel polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and methods of use thereof," US Patent no. 09772134, January 2001.
- [78] A. J. Afzal and D. A. Lightfoot, "Soybean disease resistance protein RHG1-LRR domain expressed, purified and refolded from *Escherichia coli* inclusion bodies: preparation for a functional analysis," *Protein Expression and Purification*, vol. 53, no. 2, pp. 346–355, 2007.
- [79] N. S. Nelsen, A. L. Warner, D. A. Lightfoot, B. F. Matthews, and H. T. Knap, "Duplication of a chromosomal region from linkage group A2 involved in soybean cyst nematode resistance in soybean," *Molecular Breeding*. In press.
- [80] S. C. Anand, "Registration of 'Hartwig' soybean," *Crop Science*, vol. 32, no. 4, pp. 1069–1070, 1992.
- [81] D. M. Webb, B. M. Baltazar, A.P. Rao-Arelli, et al., "Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654," *Theoretical and Applied Genetics*, vol. 91, no. 4, pp. 574–581, 1995.
- [82] I. Schuster, R. V. Abdelnoor, S. R. R. Marin, et al., "Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*)," *Theoretical and Applied Genetics*, vol. 102, no. 1, pp. 91–96, 2001.
- [83] R. Ahsan, M. J. Iqbal, A. J. Afzal, A. Jamai, K. Meksem, and D. A. Lightfoot, "Analysis of the activity of the soybean laccase encoded within the *Rfs2/Rhg1* locus," *Theoretical and Applied Genetics*. In press.
- [84] K. Meksem, K. K. Zobrist, E. Ruben, et al., "Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping," *Theoretical and Applied Genetics*, vol. 101, no. 5–6, pp. 747–755, 2000.
- [85] C.-C. Wu, P. Nimmakayala, F. A. Santos, et al., "Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 1041–1050, 2004.
- [86] T. Aoki, K. O'Donnell, Y. Homma, and A. R. Lattanzi, "Sudden-death syndrome of soybean is caused by two morphologically and phylogenetically distinct species within

- the *Fusarium solani* species complex: *F. virguliforme* in North America and *F. tucumaniae* in South America," *Mycologia*, vol. 95, no. 4, pp. 660–684, 2003.
- [87] K. Arumuganathan and E. D. Earle, "Estimation of nuclear DNA content of plants by flow cytometry," *Plant Molecular Biology Reporter*, vol. 9, no. 3, pp. 229–241, 1991.
- [88] R. J. Singh and T. Hymowitz, "The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis," *Theoretical and Applied Genetics*, vol. 76, no. 5, pp. 705–711, 1988.
- [89] R. C. Shoemaker, K. Polzin, J. Labate, et al., "Genome duplication in soybean (*Glycine* subgenus *soja*)," *Genetics*, vol. 144, no. 1, pp. 329–338, 1996.
- [90] R. Shoemaker, P. Keim, L. Vodkin, et al., "A compilation of soybean ESTs: generation and analysis," *Genome*, vol. 45, no. 2, pp. 329–338, 2002.
- [91] G. Blanc and K. H. Wolfe, "Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes," *The Plant Cell*, vol. 16, no. 7, pp. 1667–1678, 2004.
- [92] A.-G. Tian, J. Wang, P. Cui, et al., "Characterization of soybean genomic features by analysis of its expressed sequence tags," *Theoretical and Applied Genetics*, vol. 108, no. 5, pp. 903–913, 2004.
- [93] C. R. Yesudas, J. L. Shultz, H. B. Zhang, G. K.-S. Wong, and D. A. Lightfoot, "A catalog of duplicated regions from marker amplicon homologs and BAC DNA sequence analysis in soybean, a paleopolyploid genome," in *Plant & Animal Genome XIV Conference*, p. 33, San Diego, Calif, USA, January 2006.
- [94] J. L. Shultz, J. D. Ray, and D. A. Lightfoot, "A sequence based synteny map between soybean and *Arabidopsis thaliana*," *BMC Genomics*, vol. 8, article 8, pp. 1–8, 2007.
- [95] J. L. Shultz, C. R. Yesudas, S. Yaegashi, J. A. Afzal, S. Kazi, and D. A. Lightfoot, "Three minimum tile paths from bacterial artificial chromosome libraries of the soybean (*Glycine max* cv. 'Forrest'): tools for structural and functional genomics," *Plant Methods*, vol. 2, article 9, pp. 1–10, 2006.
- [96] J. P. Tomkins, R. Mahalingam, H. Smith, J. L. Goicoechea, H. T. Knap, and R. A. Wing, "A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with cyst nematode resistance," *Plant Molecular Biology*, vol. 41, no. 1, pp. 25–32, 1999.
- [97] Y. Chen, Y. Wang, K. Meksem, and D. Wang, "Construction of a partial physical map for wild soybean," in *Plant & Animal Genome XI Conference*, p. 483, San Diego, Calif, USA, January 2003.
- [98] C. Wu, S. Sun, P. Nimmakayala, et al., "A BAC- and BIBAC-based physical map of the soybean genome," *Genome Research*, vol. 14, no. 2, pp. 319–326, 2004.
- [99] J. L. Shultz, K. Meksem, and D. A. Lightfoot, "Evaluating physical maps by clone location comparisons," *Genome Letters*, vol. 2, no. 3, pp. 109–118, 2003.
- [100] N. Bouché and D. Bouchez, "Arabidopsis gene knockout: phenotypes wanted," *Current Opinion in Plant Biology*, vol. 4, no. 2, pp. 111–117, 2001.
- [101] T. Nawy, J.-Y. Lee, J. Colinas, et al., "Transcriptional profile of the *Arabidopsis* root quiescent center," *The Plant Cell*, vol. 17, no. 7, pp. 1908–1925, 2005.
- [102] K. Meksem, "Soybean tilling project," 2007, <http://www.soybeantilling.org/tilling.jsp>.
- [103] J. L. Shultz, C. R. Yesudas, S. Yaegashi, et al., "Sequencing of BAC ends associated with an updated ECBAC1 minimum tile path of soybean (*Glycine max*)," GenBank ER962965 to ER966289 (3,324 sequences), 2007.
- [104] M. C. Luo, J. Dvorak, J. L. Shultz, and D. A. Lightfoot, "Enhancement of the Forrest physical map with four color fluorescent fingerprints of a minimum tile path," In press.
- [105] DOE Soybean Project, <http://www.jgi.doe.gov/sequencing/cspseqplans2006.html>.
- [106] P. M. Olhoft, L. E. Flagel, C. M. Donovan, and D. A. Somers, "Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method," *Planta*, vol. 216, no. 5, pp. 723–735, 2003.
- [107] M. M. Paz, H. Shou, Z. Guo, Z. Zhang, A. K. Banerjee, and K. Wang, "Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant," *Euphytica*, vol. 136, no. 2, pp. 167–179, 2004.
- [108] P. Zeng, D. A. Vadnais, Z. Zhang, and J. C. Polacco, "Refined glufosinate selection in *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill]," *Plant Cell Reports*, vol. 22, no. 7, pp. 478–482, 2004.
- [109] K. Triwitayakorn, "Dissection of gene clusters underlying resistance to *Fusarium solani* F. sp. *glycines* (*Rfs* loci) and *Heterodera glycines* (*Rhg* loci) in soybean," Ph.D. thesis, PLB, Southern Illinois University at Carbondale, Carbondale, Ill, USA, 2002.
- [110] U. Z. Hammes, E. Nielsen, L. A. Honaas, C. G. Taylor, and D. P. Schachtman, "AtCAT6, a sink-tissue-localized transporter for essential amino acids in *Arabidopsis*," *The Plant Journal*, vol. 48, no. 3, pp. 414–426, 2006.
- [111] S. V. Wesley, C. A. Helliwell, N. A. Smith, et al., "Construct design for efficient, effective and high-throughput gene silencing in plants," *The Plant Journal*, vol. 27, no. 6, pp. 581–590, 2001.
- [112] R. Collier, B. Fuchs, N. Walter, W. K. Lutke, and C. G. Taylor, "Ex vitro composite plants: an inexpensive, rapid method for root biology," *The Plant Journal*, vol. 43, no. 3, pp. 449–457, 2005.
- [113] A. Jamai, L. Shiming, X. H. Liu, M. Goellner-Mitchum, H. Ishihara, and K. Meksem, "Molecular and functional analysis of the *Rhg4* locus conferring resistance to the soybean cyst nematode," in *Plant & Animal Genome XIV Conference*, p. 781, San Diego, Calif, USA, January 2007.
- [114] P. Moolhuijzen, M. Cakir, A. Hunter, et al., "LegumeDB1 bioinformatics resource: comparative genomic analysis and novel cross-genera marker identification in lupin and pasture legume species," *Genome*, vol. 49, no. 6, pp. 689–699, 2006, erratum in *Genome*, vol. 49, no. 9, pp. 1206–1207, 2006.
- [115] H. H. Yan, J. Mudge, D.-J. Kim, B. C. Shoemaker, D. R. Cook, and N. D. Young, "Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*," *Genome*, vol. 47, no. 1, pp. 141–155, 2004.
- [116] D. Grant, P. Cregan, and R. C. Shoemaker, "Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 8, pp. 4168–4173, 2000.
- [117] J. A. Schlueter, P. Dixon, C. Granger, et al., "Mining EST databases to resolve evolutionary events in major crop species," *Genome*, vol. 47, no. 5, pp. 868–876, 2004.
- [118] M. G. Francki and D. J. Mullan, "Application of comparative genomics to narrow-leaved lupin (*Lupinus angustifolius* L.) using sequence information from soybean and *Arabidopsis*," *Genome*, vol. 47, no. 4, pp. 623–632, 2004.
- [119] T.-Y. Hwang, J.-K. Moon, S. Yu, et al., "Application of comparative genomics in developing molecular markers

tightly linked to the virus resistance gene *Rsv4* in soybean," *Genome*, vol. 49, no. 4, pp. 380–388, 2006.

- [120] H.-W. Wang, J.-S. Zhang, J.-Y. Gai, and S.-Y. Chen, "Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1086–1097, 2006.
- [121] M. D. Gonzales, E. Archuleta, A. Farmer, et al., "The legume information system (LIS): an integrated information resource for comparative legume biology," *Nucleic Acids Research*, vol. 33, database issue, pp. D660–D665, 2005.
- [122] J. Mudge, S. B. Cannon, P. Kalo, et al., "Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*," *BMC Plant Biology*, vol. 5, article 15, pp. 1–16, 2005.
- [123] J. L. Shultz, S. Ali, L. Ballard, and D. A. Lightfoot, "Development of a physical map of the soybean pathogen *Fusarium virguliforme* based on synteny with *F. graminearum* genomic DNA," *BMC Genomics*, vol. 8, Article ID 262, 7 pages, 2007.
- [124] V. N. Njiti, K. Meksem, D. A. Lightfoot, W. J. Banz, and T. A. Winters, "A method for breeding and genetically manipulating phytoestrogen content in soybeans," US Patent no. 10/008789, August 2000.
- [125] D. A. Lightfoot, "Isolated polynucleotides and polypeptides relating to loci underlying seed protein and oil content and methods of use thereof," US Patent no. 07/2007, December 2007.
- [126] D. A. Lightfoot, K. Meksem, and P. T. Gibson, "Method of determining soybean sudden death syndrome resistant in a soybean plant," US Patent no. 7288386, October 2007.
- [127] W. J. Banz, M. P. Williams, D. A. Lightfoot, and T. A. Winters, "The effects of soy protein and soy phytoestrogens on symptoms associated with cardiovascular disease in rats," *Journal of Medicine and Food*, vol. 2, no. 3-4, pp. 271–273, 1999.
- [128] J. A. Greer-Baney, W. J. Banz, D. A. Lightfoot, and T. A. Winters, "Dietary soy protein and soy isoflavones: histological examination of reproductive tissues in rats," *Journal of Medicinal Food*, vol. 2, no. 3-4, pp. 247–249, 1999.
- [129] M. J. Iqbal, S. Yaegashi, R. Ahsan, D. A. Lightfoot, and W. J. Banz, "Differentially abundant mRNAs in rat liver in response to diets containing soy protein isolate," *Physiological Genomics*, vol. 11, no. 3, pp. 219–226, 2003.
- [130] Y. Cho, V. N. Njiti, X. Chen, D. A. Lightfoot, and A. J. Wood, "Trigonelline concentration in field-grown soybean in response to irrigation," *Biologia Plantarum*, vol. 46, no. 3, pp. 405–410, 2003.

Composition Comments

1. Please check the correctness of the address.
2. Please reduce the abstract to not more than 200 words as per the journal's style.
3. Please rephrase the highlighted part for the abstract should be self-contained as per the journal's style.
4. We reduced the font size of labels in Figure 2 and some labels in Figures 5 and 8. Please check.
5. We redrew some parts of Figures 3(b) and 8. Please check.
6. Please clarify the missing part in the highlighted part in the caption of Table 2.
7. We changed "SDX" to "SD-X" in Figure 4(b). Please check.
8. We rephrased the highlighted part for the sake of clarity. Please check all highlighted changes throughout.
9. Should we remove "and" in the highlighted part? Otherwise, please insert the missing part.
10. We multiplied the y-axis of Figure 5 by factor "10³." Please check.
11. We reduced the enumeration on the x-axis of Figure 5. Please check.
12. Please check labels of Figure 8 since they are not clear.
13. Please update the information of these references [5, 18, 45, 52, 72, 79, 83, 104] if possible.
14. We changed the title of these references [15, 43, 78, 84] Please check.
15. Please provide the volume number of this reference. If possible.