Using A Minimum Tile Path For Plant Transformations Encompassing the Entire Soybean Genome

Hidayat Ullah\textsuperscript{1,2}, Mohammed Jasim\textsuperscript{1,3} and David A Lightfoot\textsuperscript{1*}

\textsuperscript{1} Department of Plant, Soil and Agricultural Systems, Genomics and Biotechnology Core Facility, Centre for Excellence, The Illinois Soybean Centre, Southern Illinois University, Carbondale IL, 62901, USA; \textsuperscript{2} Faculty of Agriculture, Abdul Wali Khan University, Mardan, Khyber Pakhtunkhwa, Pakistan; \textsuperscript{3} Department of Biotechnology, Al-Anbar University, Ranadi, Iraq

Received: October 10, 2011 / Accepted: January 20, 2012

Abstract

Genomes like \textit{Glycine max} (soybean) that have been highly conserved following increases in ploidy represent a frontier for genome analysis. Many soybean QTL analyzed to date have been composed of gene clusters each with contributing a portion of the trait rather than alleles of single genes. At the Soybean Genome Database (SoyGD) \url{http://soybeangenome.siu.edu} the genome browser that integrates and served the publicly available soybean physical map, BAC fingerprint database and genetic map associated genomic data shows a minimum tile of transformation ready BIBAC-like clones in pCLD04541 (pV41; oriV; tra; bom). Sequence resources made available through the DOE genome sequencing project have allowed the minimum tile to be revised and new functional analyses to be made. There are 3,840 MTP clones that appeared to encompass 90\% of the genome (see \url{http://soybeangenome.siu.edu/cgi-bin/gbrowse/BES_scaffolds}). The BIBAC-like clones (tetR) from \textit{E. coli} DH10 B were transferred en masse to \textit{Agrobacterium tumefaciens} by triparental matings with EHA105 (rifR) mediated by pRK2013 (oriP) in DH10B (kanR) in 384 well plates. Although not necessary the extra helper plasmid boosted efficiency 10 fold. Individual \textit{A. tumefaciens} rifampicin and tetracyclin resistant strains were used for transformation of \textit{Arabidopsis thaliana} flowers in 384 well arrays. Initially kanamycin selection was used to isolate transgenic plants. Because the BACs were already tetR the recA mutants of \textit{A. tumefaciens} could not be used (Tn3 insertions). Consequent to this and partial transconjugation events only some inserts are transferred completely while other transformed lines contain a substitution series of deleted inserts anchored on the Ti-left border (LB). These are maintained as kanR mixtures of seed. Phenotypes found for lines transgenic for particular BACs that were repeated in-clude seed composition (protein, oil), development (growth, senescence) and disease resistance (sudden death syndrome (SDS) and soybean cyst nematode (SCN).

Introduction

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 (Lightfoot, 2008). Among legumes, however, soybean (\textit{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 so named for the abilities of a southern Civil War General to arrange a defense (Hartwig and Epps, 1973). 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 (Wrather et al., 1997; 2003). 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 (Lightfoot, 2008).

Over the past decade, investment in Forrest genomics resulted in the development of the many resources for r genomics research (Lightfoot, 2008). Resources included: (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 (BES) that encompass 18.5 Mbp spaced out from the MTPs; (viii) a map of 2,408 regions each found at a single position in the genome and 2,104 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. Data on the Forrest genome and resources are provided to the scientific community through SoyGD, LIS, Soybase and GenBank (Wu et al., 2004a,b; Shultz et al., 2006a,b; 2007a,b; Saini et al., 2008).

In combining desired characters, the structure of loci and even chromosomes appears to be pivotal to the special qualities of the Forrest genome (Ruben et al., 2006; Afzal et al., 2009; Karangula et al., 2009). Genes underlying many quantitative and qualitative loci are targeted for isolation in the laboratories of the worldwide collaborating groups. What was missing until recently (Hamilton et al., 1997) was a methods to test the function of BACs or genes within BACs in planta. Transformation with large insert clones has become a limiting need. Here that earlier work was extended to BACs from soybean.

In preliminary work genes isolated from Forrest-derived BACs were used for transformation of A thaliana by the floral drop method (Lightfoot and Ullah, 2009). BACs used included those encoding candidates for resistance to nematode (Rgh4 and rhg1), resistance to Phytophthora sojae (Rps5), resistance to Pseudomonas syringae (Rps1) and resistance to Fusarium virguliforme (Rf52). 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). Here is reported for the first time the transformation of the entire soybean genome into A. thaliana as the minimum tile path of BACs from the Forrest physical map and genome sequence.

Materials and Methods

There are 3,840 MTP clones that appeared to encompass 90% of the genome (see http://soybeangenome.siu.edu/cgi-bin/gbrowse/BES_scaffolds; Shultz et al., 2006a,b; 2007a,b). The BIBAC-like clones from E. coli DH10B were transferred en masse to A. tumefaciens EHA105 by triparental matings (Figure 1) mediated by pRK2013 in 384 well plates on solid TY media (no selection). Individual A. tumefaciens strains were selected for rif and kan resistance twice, once on solid media then in liquid media.

The rif/tetR strains were used for transformation of A. thaliana flowers following 2 fold dilution with 0.5 MS 0.005% (w/v) Sillwet50 and 384 clone-to-pot arrays (Figure 2). Kanamycin selection with Sillwet50 adjuvant was used to isolate transgenic plants by applications on three occasions, 7 days apart, to T1 lines (7,14 and 21 dag). Seed were collected in eppendorf tubes at T1 to ease reuse and distribution. T2 seed were selected for KanR but collected into 96 well plates to allow for robotic methods of analysis.

Transformed plants and their progenitor strains were tested with the appropriate BES and marker amplicons by PCR to assess the extent of insert stability in A. tumefaciens and the proportion of the insert transferred to the plants (Figure 3).

Phenotypes tested for among transgenic plants included altered flower color, disease resistances, seed yield, seed protein, seed oil, seed isoflavone content, nitrogen use efficiency, water use efficiency and herbicide resistances.

Results

Assessing the Efficiency of Transformation

PCR amplicons showed about 50% of KanR plants had complete inserts and 50% had deletion derivatives (Figure 3). The deletion derivatives appeared to derive from instability in the A. tumefaciens rather than the plant. This might be expected since the KanR derived from nptII in pCLD04541 is between the left

![Figure 1. Triparental mating schema to mobilize soybean BACs into transformation competent Agrobacterium tumefaciens.](image-url)
Figure 2. Transformation and selection of *A. thaliana* transgenic with BACs.

Figure 3. Hybridization of right border probes to single colony derived strains of *A. tumefaciens* (A) and PCR amplifications from single plants transgenic (B) with BAC clones derived from the rhg1 region. Panel A shows some clones have right border sequences and so are full length clones but about half do not and so are deletions. Panel B shows an ideogram of the overlapping BAC clones B73P6 and PCR amplifications of three markers contained in the clone from *A. thaliana*’s DNA extracted from transformed plants (lanes 1-4) compared to the alleles in Essex (E) and Forrest (F) soybean cultivars and 4 RILs (lanes 5-8).
border and the multicloning site, causing the selectable marker to be the first gene transferred. This partial deletion series can be advantageous. With whole insert transformation each phenotype may be associated with the 20-35 genes encoded by an individual BAC but with the deletion derivatives a phenotype might be associated with a smaller number of genes.

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* mediated transformation of cultured cells with Forrest BAC clones has been successfully achieved using previously described protocols involving the T-DNA vector pCLD04541. In this protocol, *npt II* gene was used as a plant selectable marker, and kanamycin as used as a selective agent. Screen-able markers are available in some BAC clones (Table 1). 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 first set of clones selected for transformation are listed in Table 1, and provided for complementation of easily score-able 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 or a large insertion/deletion (Ashfield et al., 2003; Tri-

### Table 1. Some of the BACs, mutant and non-mutant soybean lines to be transformed for complementation.

<table>
<thead>
<tr>
<th>SoyGD BIBAC Clone names</th>
<th>Phenotypes</th>
<th>Scaffold BIBAC</th>
<th>Insert Size kbp</th>
<th>Dominant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gm-SIU1-B100B10</td>
<td><em>Rhg</em>4 bigenic resistance to SCN</td>
<td>B11G08</td>
<td>140</td>
<td>Yes</td>
</tr>
<tr>
<td>Gm-SIU1-B73P06</td>
<td><em>rhg</em>1 bigenic resistance to SCN and Rfs2 for SDS</td>
<td>B2704</td>
<td>79</td>
<td>Co-</td>
</tr>
<tr>
<td>Gm-SIU2-H050N07</td>
<td><em>Rpg</em>1-b resistance to bacterial pustule</td>
<td>H78C14</td>
<td>110</td>
<td>Yes</td>
</tr>
<tr>
<td>Gm-SIU1-B54E07</td>
<td><em>T</em> Tawny Pubescence; Flavonoid-3-monooxygenase</td>
<td>H62108</td>
<td>82</td>
<td>Yes</td>
</tr>
<tr>
<td>Gm-SIU2-H40008</td>
<td><em>W1</em> White Flower and Black Hila Color</td>
<td>H33M18</td>
<td>153</td>
<td>No</td>
</tr>
<tr>
<td>Gm-SIU2-H82C08</td>
<td><em>Rfs</em>1 root resistance to SDS</td>
<td>H37G19/B4</td>
<td>130</td>
<td>Yes</td>
</tr>
<tr>
<td>Gm-SIU1-H49N14</td>
<td><em>Rps</em>4 resistance to Phytophthora root rot</td>
<td>H8H15</td>
<td>120</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*a* *Rhg*4 and *rhg*1 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* *Rpg*1-b encodes a nucleotide binding leucine rich repeat protein. *c* *T* encodes flavonoid-3 monooxygenase (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. (Lightfoot, IJP 2008:1-22. doi:10.1155/2008/793158).

Figure 4. BAC transformed *A. thaliana* with 3 clones from a region underlying soybean seed yield. Panel A shows the differences in biomass, growth, and vigor among the lines. Panel B shows an ideogram of the QTL region and the BAC clones positions in that interval.
Phenotypes were found for clones predicted to underlie seed yield, oil content, root development, resistance to SCN and bacterial pustule. For seed yield the interval had be reduced to about 0.7 cM or 300 kbp by comparative mapping in two populations (Yuan et al., 2002). The three BACc transformed into A.thaliana gave three different phenotypes (Figure 4). The lines transgenic with H27o04 were very high yielding giving 3 fold the seed of the average lines transformed at the same time in the same growth chamber. The lines transgenic with H34m04 were typical of the average. However, lines transformed with H60c03 were suppressed in yield four fold compared to lines grown at the same time in the same growth chamber. That two clones from the same QTL could have opposite effects was predicted from fine mapping studies where QTL can often be separated into different loci. Loci of opposite effects on seed yield form part of the basis of hybrid vigor in outbreeding crops. However, in soybean they are undesirable and need to be separated by selection of recombination events that have positive yield effect in phase.

Disease resistance genes had effect on A. thaliana development in all three cases tested. With lines transgenic with the clone H50n07 there was marked auto-necrosis (Figure 5). The genes encoded by the BAC clone include the NB-LRR proteins that underlie the resistance of soybean to bacterial pustule. The genes alone used as subclones do not cause the effect shown here (Ashfield et al., 2004; 2006) though they do provide A. thaliana resistance to Pseudomonas syringae pv. glycines. The whole BAC may cause the genes encoded be regulated in a distinct manner.
causing an inappropriate hypersensitive reaction.

In contrast to Rpg1, the disease resistance loci on BACs encompassing rhg1/Rfs2 and Rfs1 did not cause necrosis, so this is not an effect of all loci (Figure 6). The rhg1 clones stunted plant development and produced a clavata-like phenotype. The genes encoded by the BAC include a RLK in the same family as clavata 1. This protein does bind CLE domains, including that in nematode secretions. The Forrest allele of the rhg1 locus has been implicated in reducing seedling root growth, germination and seedling vigor, serious negative agronomic effect of SCN resistance. The disease resistance loci on BACs encompassing Rfs1 did not cause any negative effects and the vigor of plants was greater when challenged with Fusarium. The fine map of showing the clone 73p06 encompassed part of the Rhg1 locus (Figure 3) and B30m24 appeared to encompass part of the Rfs1 locus.

There were no phenotypic changes from clones predicted to encode flower color, pubescence color or Phytophthora resistance. Clearly not all clones will have an effect on A. thaliana development. The BACs will be introducing 20-30 new genes per line, but these will be regulated from endogenous promoter rather than overexpressed. It will be interesting to assay the extent of genetic perturbation in these lines by microarray.

**Massively Parallel Plant Transformation with BACs**

To date (mid 2012) the transformation of the entire MTP to Arabidopsis thaliana had reached the half was point. Five of the twelve 384 well plates of MTP4BH had been used to generated T2 seed. The efforts were limited by time and greenhouse space. Relatively few phenotypic effects were found in the T2 selections suggesting that the phenotype observed with mapped clones was the effect of enrichment.

**Conclusions**

The non-redundant part of the soybean genome can be transferred to Arabidopsis by minimum tile of 5,836 transformations. Genes underlying known QTL identified for protein and oil on linkage group A1 (Chromosome 5) were separated on two BACs. Seed yield QTL on linkage group K (Chromosome 9) could be divide into 3 BACs with opposite effects on growth. Two BACs from linkage group G (chromosome 9) encode rhg1/Rfs2 that stunts growth and Rfs1 that resists Fusarium. The BAC encompassing disease resistance gene Rpg1 appeared to cause auto-apoptosis.

Isolating genes underlying many QTL might be more efficient by mass transformation than fine mapping. There are a lot of QTL mapped in soybean, about 1,500 at the last count. Many more will be discovered by new mapping techniques. There are 79 traits and 307 QTL mapped in ExF92 alone (Lightfoot 2008). Fine mapping is very expensive and slow depending on the generation of massive populations to find rare recombinants. In contrast it takes about 40 BACs to encompass 10 cm (4 Mbp), the mean interval for a coarsely mapped QTL. Forty transformation for a thousand QTL would be a massive undertaking. Therefore, whole genome transformation with 5-6 thousand clones and development of an immortal seed bank will be more efficient than transformations QTL by QTL.

The en masse transformation method might be extended to soybean by hairy root transformation of composite traits. For whole soybean transformation kanR is a difficult selectable marker to rely upon. Therefore, a suicide plasmid was developed to introduce the part selectable marker to the right border region by recombination. The use of the plasmid will be reported.

**References**


Kazi S (2005) Identification of twenty three loci conditioning agronomic traits in recombinant inbred and near isogenic soybean lines derived from Flyer by Hartwig. MS Thesis SIUC Carbondale IL, USA, pp. 212.


