

**International Soybean Genome Consortium**  
**20 April 2007**

I would like to first thank our Japanese hosts for the organizing the first meeting of the International Soybean Genome Consortium at the NIAS in Tsukuba. We would also like to congratulate them on their recent award for sequencing and development of genetic tools in soybean.

**Attendees**

**Japan**— Teruo Ishige , Kyuya Harada, Takuji Sasaki, Takashi Matsumoto, Baltazar Antonio, Yuichi Katayose, Akito Kaga, Atsunori Fukuda.

**Korea**—Suk-Ha Lee, Ik-Young Choi, Kyujung Van

**China**—Junyi Gai, Shouyi Chen, Lijuan Qiu, Jinsong Zhang, Deyue Yu, Yang Gao

**U.S.**—Perry Cregan, Scott Jackson Randy Shoemaker

**Election of Chair**

Scott Jackson was elected chair for a one-year term.

**Next Meetings**

The next meeting will be held in conjunction with the Plant and Animal Genome Meeting in San Diego January 12-16, 2008 followed by a meeting April 13-18, 2008 coinciding with the International Crop Science Congress Meeting in Jeju Island, Korea. Information on both meetings can be found at:

<http://www.intl-pag.org>

<http://www.cropscience2008.com/>

**Part I: General Discussion.**

This was a discussion by country of ongoing and planned soybean genomics projects. Part of this discussion was standardization of chromosome/linkage group nomenclature. Drs. P. Cregan and R. Shoemaker agreed to spearhead this effort and will report on progress at the January meeting.

**China**—J. Zhang and S. Chen reported on fairly extensive analysis of soybean ESTs and development of a soybean-specific unigene set. Also predict ~63,000 genes in soybean.

W. Dong reported work at BGI that includes: 2 BAC libraries (Kefang No. 1 and Suinong No. 14). BAC libraries have been pooled for PCR screening. Plans include putting ~2000 ESTs on genetic map via EST-SSRs.

L. Qiu reported on genotyping the germplasm collection held by the CAAS. More than 23,000 accessions are maintained with a primary core set of 2,794, a secondary core of ~600 and a mini-core with 200 accessions. Genotyping with ~60 SSRs has been performed.

L. Qiu also reported that SuiNong No. 14 has been selected as the representative genotype for sequencing. It is high yielding and has a mixture of US and Japanese cultivars in its pedigree. A FLcDNA library has been made for SuiNong 14 and ~400 have been sequenced and QCd.

**Korea**—Suk-Ha Lee reported on ongoing work sequencing around the RXP locus. Using BACs from the Williams 82 genotype in collaboration with R. Shoemaker (U.S.). BACs from an internally duplicated region have been sequenced and show a lot of colinearity between the paralogous regions.

They are attempting to sequence by pooling BACs and using advanced sequencing technologies (454).

**Japan**—K. Harada reported on their recent success in securing money for sequencing of the genotype Enrei (Maturity Grp IV). Included in this is the development of a set of genomic/genetic tools for breeding such as linkage and physical maps, identification of agriculturally important genes (especially those specific to paddy-grown soybeans, low yield and flooding/logged stress) and isolation of genes and development of DNA markers.

Populations **to be** developed: Enrei x Peking (F2 and RILs) and Enrei x Williams 82 (F2).

Two BAC libraries for Enrei have developed (Hind III and BamHI), 100,000 clones in each library and 50,000 clones from each library will be end-sequenced.

Enrei will be sequenced to 10x sequence coverage using advanced sequencing technologies (454 GSFLX).

~40,000 FLcDNAs for Nourin No.2 are already end-sequenced at RIKEN.

A 44,000 feature Agilent microarray chip developed using FLcDNAs and ESTs will be available soon.

A. Kaga reported on development of high-density linkage map using EST-SSRs and now SSRs from genomic sequence.

Y. Katayose reported further on the BAC libraries for Enrei. They are cloned in pIndigoBAC5. They are designated GMJENa (HindIII) and GMJENb (BamHI) (100,000 clones each).

**United States**—S. Jackson reported on the ongoing genome projects in the genotype Williams 82. The ultimate goal of the US soybean genome community is a reference genome, at least for the gene-rich, euchromatic portions of the genome.

BAC libraries: three BAC libraries with three enzymes (HindIII, BstYI and BamHI) have been made, also known as GM\_Wba, GM\_Wbb and GM\_Wbc. Subsets of each of the three libraries have been end-sequenced and the traces and sequences deposited in GenBank.

All three libraries have been fingerprinted via HICF and a physical map produced (map is actively curated). The physical map can be seen at soybase.org. More than 2000 markers (overgos and PCR pool based) have been placed in the physical map (all available via soymap.org)

JGI is doing a shotgun sequence and at this point in time has ~4.3X coverage, all available via the trace archive at GenBank. Targeted completion date for the shotgun sequence is 2008.

P. Cregan and many collaborators have placed ~3,000 SNP-based markers on the genetic map (many of these have been placed on the physical map via overgo hybridizations). All mapping data and marker information is available at <http://bfgl.anri.barc.usda.gov/soybean/>.

## Part II: Discussion of Specific Aspects of Collaboration

It was agreed that the name of the organization would be:

**The International Soybean Genome Consortium (ISGC)**

### Available Resources:

BAC libraries

3 - Williams82 BAC libraries

2 - Enrei BAC libraries

200,000 *Hind*III clones, average size = 140 kbp, BES available 2007

*Mbo*I in preparation, BES available 2008

? - Suinong #14 BAC libraries

### Mapping Populations:

China: Kefeng No. 1 x Suinong #14

Dr. Gai Junyi has 8 mapping populations

Japan Misuzudaizu x Moshidou Gong 503 –156RILs

Enrei x Peking-192 RILs in the near future

Enrei x Williams 82-382 F2

1000 Backcross introgression lines of Enrei<sup>3</sup> x Peking in the near future

USA Minsoy x Noir 1 - 89 RILs at this time

Minsoy x Archer - 89 RILs at this time

Evans x Peking – 104 RILs

Williams 82 x Forrest – Univ. of Missouri –As many as 1000 RILs  
available in the near future

Williams 82 x *G. soja* PI 468916: 450 RILs available late 2007

Korea: Pureunkong x Jinpumkong 2 – 92 RILs

Hwanggeumkong x IT182932 – 105 RILs

### Markers:

SSR and SNPs – See SoyBase.org and the USDA Beltsville Website  
(<http://bfgl.anri.barc.usda.gov/soybean/index.html>)

The Japanese have many hundreds of EST-derived SSRs and other fragment  
length polymorphisms that will be mapped in Misuzudaizu x Moshidou Gong  
503 (Kazusa DNA Research Institute)

### FL cDNAs

10,000 in GenBank

TIGR (Chris Towne) will produce and 25,000 in the next two years

Pioneer Hibred (DuPont) and Monsanto have many ESTs but these are not yet fully available  
Fosmid library that has been used in the JGI sequencing will be made available

**Microarrays:**

Affymetrix Chip: 37,000 soybean genes + 13,000 *Phytophthora infestans* and soybean cyst nematode genes  
cDNA arrays are available through Lila Vodkin at the Univ. of Illinois, Urbana, IL USA  
Long oligo arrays are also available through Lila Vodkin at the Univ. of Illinois, Urbana, IL USA

**Sequencing Strategy, Quality and Assembly\*\***

U.S.: JGI sequence will be an 8X Sanger shotgun sequence – the ultimate goal is an aligned reference sequence of at least the gene rich regions.  
Depending upon results of the assembly of the current 4.3X sequence, JGI may decide to use some type of filtering (Cot or methyl filtration) to target gene-rich regions in the remaining 3.5X sequence  
All data including traces of the JGI initial 4.3X Williams 82 sequence are available. The initial assembly of this 4.3X will not be made available.  
BAC by BAC Sanger sequencing of a few regions of Williams 82 have been completed in order to verify the integrity of the shotgun sequence assembly.  
Japan: A 10X sequence of Enrei will be completed using the 454 BioSciences sequencer. The current sequence assembly software available from 454 will not handle a whole genome shotgun sequence as large as soybean.

**Annotation**

The following community standards must be developed and agreed upon:  
Definition of a gene  
Gene naming conventions  
Possible annotation models: Rice or Medicago (MGAC?) or other  
To develop annotation standards it was agreed that a committee be established for this purpose with a representative(s) from each of the member nations. The following nominations were put forth:  
Japan: Takeshi Ito (taitoh@affrc.go.jp) (Takashi Matsumoto can provide contact info. for Takeshi Ito  
U.S.A.: Chris Towne, Steve Cannon  
China: ???  
Korea: Beom Soon Choi (bschoi@nicem.snu.ac.kr)  
A brief discussion of gene prediction programs followed.

Linkage Group designations will be altered. The current system which uses letters will be replaced by a system of numbers from 1 to 20. Cregan will consult with Ted Hymowitz about this and report back to the group.

**Databasing:**

All BES, complete BAC sequence and genome shotgun sequence should go to NCBI and the trace archives (the JGI Williams 82 reads and traces are available at NCBI)

Williams 82 BAC contigs are available on SoyBase

Steve Cannon (USDA, Ames in the Shoemaker Lab) will begin integrating the Williams 82 sequence and physical map

Eventually all Williams 82 sequence contigs plus the physical and genetic maps will all be available in CMap on SoyBase.

The ultimate goal: A “BioMoby” type system that will connect all national soybean genome databases to permit answers to inquiries that can seamlessly acquire information from the various national databases.

### **Data Release Policy**

USA: All reads go into the NCBI database, both FASTA and trace files  
JGI builds will not be immediately available

Japan: Enrei shotgun sequence and BES sequence will be placed in the DDBJ (DNA DataBank of Japan)

China: Not sequencing at the moment

Korea: A number of Williams 82 BAC clones are being sequenced. The question arose about the possible duplication of effort with the JGI sequencing of 500 Williams 82 BACs. It was agreed that a database be created on SoyBase that would provide a listing of all the Williams 82 BACs that were being sequenced by both the JGI and the group at the Seoul National Univ.

### **Timeline:**

The JGI Williams 82 and the Japanese Enrei sequencing should both be completed sometime in 2008-2009.

### **Patent Issues:**

No patenting of sequences.

### **Private Sector Collaboration/Policy**

Private organizations have a large amount of soybean sequence (Pioneer and Monsanto) but it is not clear how or if it can be released. Monsanto ?? has released a large amount of maize sequence.

### **Progress Reports**

An ISGC site will be established on SoyBase where each member country can periodically indicate progress on assembly, new markers, anchoring of the sequence to the genetic and physical map, etc.

### **Publication**

A publication should be drafted that describes the goals of the ISGC, its members, its structure. The drafting of the publication was deferred to a committee of national representatives as follows:

U.S.A: Jackson  
Korea – Suk-Ha Lee  
Japan: Kyuya Harada  
China: Lijuan Qiu

Additional member nations should be considered including

Brazil: Abdelnoor  
Canada:  
India:  
Australia:

\*\* re-affirmation of the importance of a reference sequence was discussed as it is vital for the approaches being taken by several other countries where a draft is produced and a reference sequence is used to help assemble the draft and look for polymorphisms.