Worldwide utilization of Chinese soybean germplasm

Li Juan Qiu

Institute of Crop Science
Chinese Academy of Agricultural Sciences
Historical contribution of Chinese soybean germplasm
Soybean Importance

- Soybean is grown for food, feed, fat, industrial materials and bio-energy worldwide.
- Soybean has both high nutritional and health value
  - Seed protein (40%) and oil (20%) content are higher
  - Storage protein (~70%) is related to soybean processing
  - It has biochemical activity protein (~30%), including trypsin inhibitor, lipoxygenase, α-amylase, etc.
The classification of soybean

**Leguminosea, Papilollateae, **Glycine**

G. max, G. soja

2n = 40

subg. Glycine

2n = 38, 40, 78, 80

---

Harlan and deWet (1971)
The distribution of genus *Glycine*

**subgenera *Glycine***
- Twenty six species in Australia
- Three species in China
  - *G. tomentella*
  - *G. tabacina* Benth
  - *G. dolichocarpa*

**subgenera *Soja***
- *G. soja*: East Asia
- *G. max*: Worldwide
Cultivated soybean is native to China

- The history of soybean cultivation are more than 4500 years since agricultural ancestor Houji, who planted five crops including soybean.
- The earliest soybean was named as ‘shu’ in the book of Odes.
- The name of soybean was translated bases on the shu in the other countries.
Landraces were distributed to Asia, Europe and America

The first century and the Age of Discovery

Chang, World Agriculture, 1989
Soybean production in the world

- Total production in the world is 251mt
- China is the fourth country among ~50 countries
• **North American**

  50% of the nuclear DNA and 83% of the cytoplasmic DNA of USA cultivars originated in China (Gizlice et al., 1994)

- Cultivars in northern US were mostly developed from spring-sown types introduced from Northeast China. **Mandarin, Richland, S-100, Mukden**
- Southern US cultivars were developed mainly from spring-sown type introductions from southern China. **CNS, S-100**

---

**Most of leading ancestors of North American germplasm were from China**

Carter et al., 2004; Chang, 1989
• South American
  • 11 Chinese soybean germplasm accessions contributed 89% of genetic make-up of Brazil cultivars (Hiromoto and Vello, Brazil J Genetics, 1986)
  
• The soybean output of Argentina is based largely on US and Brazilian materials, and thus indirectly from China

⚠️ Chinese cultivars have also provided an important genetic base for cultivar development in Japan
Partial elite accessions used in soybean production

- **Resistance / Tolerance:**
  - **SCN:** Peking (Pickett, Custer, Forrest, Mack, Centennial, Hartwig, etc), PI88788(Bedford, Fayette, etc)
  - **SMV:** PI486355
  - **Phytophthora sojae:** A.K, Mukden, CNS
  - Water logging

- **High yield:**
  - Huangbaozhu (Changyebai, Xiaodaobaimu) and Zihuasihao (Xiaojinbai, Zaoshengxiaojin, etc) in Japan
  - **Monsanto** patent high yield QTL from Chinese wild soybean

---

Forrest is believed to have played a key role in saving billions of US dollars during 1970s and 1980s
US and Brazil Soybean Germplasm Collections

There are 170,000 accessions in the world. 45,000 accessions are unique.

50 countries: <10 accessions
28 countries: 10-100 accessions
16 countries: >100 accessions
Williams (7) X Kinwa (Rps-1k)

Wayne

L49-4091 X Clark

[ Lincoln (2) X Richland] F3

[ Lincoln (2) X Richland] [CNS] F1

Lincoln: Unknown

Clemson

Lincoln (2) X Richland

Lincoln (2) X Richland

Lincoln (2) X Richland

Illini X Dunfield

1920 from A.K.

1913 from NE China Baimei

1926 from Jilin, China

1927 from Nanjing, China

Reference genome

Williams 82

China

L57-0034
Potential usage of genetic diversity of Chinese soybean germplasm for a more prosperous future
Soybean in China

- Hometown of soybean (G. max)
- 4th major food crops
- Consuming 58 mt
- Producing 16-18 mt
- Reserving >30,000 accessions
- Developing >1800 varieties
- Processing >100 edible soybean products
- Major plant protein source of the food
- Major protein source of feeding industry
Soybean Planting Regions in China

I

II

III

IV

V

VI

(Sp) Single Cropping

(Sp, Su) Double Cropping

(Sp, Su, Au, Wi) Multiple Cropping

IIASA-LUC GIS Database, 2001
The genetic gain of soybean yield is 1.07%
Zhong huang 13: High-yield and wide-adapbility
The 1st prize of scientific and technologic progress

Accumulative area (mha)

2082.5kg/h

1666.5kg/h

Hybrid Soybean: Heterosis exploitation
2nd prize of national science and technology invention

CMS lines: higher outcrossing rate
Pollinators: alfalfa leaf cutting bee, bee, natural insects
Producing seed: >1003kg/ha, 1M:2F
**Yield:** 2800~3000kg/ha in regional test
- Zhonghuang35 got 5577kg/ha in Xinjian
- Liaodou14 got 4908kg/ha in Liaoning
- Hefeng55 got 4258kg/ha at 852 farm in 40 hectare field.

**Quality**
- Protein: 183 cultivars > 45%, including 24 cultivars> 48% (Gandou 2 for 51.62%)
- Oil: 121 cultivars >22% (Jihuanguang13 for 24.10%)
- High ratio of 11S/7S (5.31)
- Lipoxygenase null
- High isoflavone (3.7 mg/g)
- Trypsin inhibitor null
Utilization of soybean germplasms is limited

- 1,300 soybean cultivars released during 1923-2005
- Traced back to 670 ancestors

<table>
<thead>
<tr>
<th>Type of end ancestor</th>
<th>No. Ancestor</th>
<th>Ratio (%)</th>
<th>Nuclear genetic contribution</th>
<th>Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved cultivar</td>
<td>47</td>
<td>7.01</td>
<td>101.29</td>
<td>7.79</td>
</tr>
<tr>
<td>Landrace</td>
<td>346</td>
<td>51.64</td>
<td>991.74</td>
<td>76.29</td>
</tr>
<tr>
<td>Breeding line</td>
<td>257</td>
<td>38.36</td>
<td>194.09</td>
<td>14.93</td>
</tr>
<tr>
<td>Wild soybean</td>
<td>17</td>
<td>2.54</td>
<td>7.06</td>
<td>0.54</td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
<td>0.45</td>
<td>5.89</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>670</strong></td>
<td><strong>1.300</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ref: Xiong et al, 2008
China owns the most of soybean germplasm

- Systematically collecting germplasm 3 times in national wide
- More than 23,000 cultivated and 7,000 wild accessions are conserved in Chinese National Gene bank (CNGB).
Diversity in Chinese soybean germplasm

Phenotype

- Inflorescence
- Hypocotyls
- Root nodule
- Pod
- Seed
Accessions with favorite traits has been identified

<table>
<thead>
<tr>
<th></th>
<th>Yield</th>
<th>Growth period</th>
<th>Maturity period</th>
<th>Seed weight</th>
<th>Plant height</th>
<th>Growth habit</th>
<th>Determinant</th>
<th>Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>23562</td>
<td>23562</td>
<td>23522</td>
<td>21946</td>
<td>22544</td>
<td>22598</td>
<td>7011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>312</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12,375</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16969</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8654</td>
</tr>
<tr>
<td></td>
<td>Resistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drought</td>
<td>7011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.7%</td>
</tr>
<tr>
<td></td>
<td>Cold</td>
<td>312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3%</td>
</tr>
<tr>
<td></td>
<td>Salt</td>
<td>10298</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43.7%</td>
</tr>
<tr>
<td></td>
<td>SCN</td>
<td>12,375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>52.5%</td>
</tr>
<tr>
<td></td>
<td>SMV</td>
<td>16969</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71.9%</td>
</tr>
<tr>
<td></td>
<td>Rust</td>
<td>8654</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36.7%</td>
</tr>
<tr>
<td></td>
<td>Morphological</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seed color</td>
<td>22637</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96.0%</td>
</tr>
<tr>
<td></td>
<td>Hilum color</td>
<td>15822</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>67.1%</td>
</tr>
<tr>
<td></td>
<td>Cotyledon color</td>
<td>22611</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95.9%</td>
</tr>
<tr>
<td></td>
<td>Pubescence color</td>
<td>22538</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95.6%</td>
</tr>
<tr>
<td></td>
<td>Flower color</td>
<td>22607</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95.8%</td>
</tr>
<tr>
<td></td>
<td>Seed shape</td>
<td>22608</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95.8%</td>
</tr>
<tr>
<td></td>
<td>Leaf shape</td>
<td>21941</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>93.0%</td>
</tr>
</tbody>
</table>
Accessions with favorite quality

- protein content
- oil content
- fatty acid
- amino acid
- 28k allergic protein null
- protein subunit null
- trypsin inhibitor
- lypoxygenase null
- isoflavone 400~7000mg/g
- high soluble protein
Establishment of core collections

Whole collection

Primary core collection

Core collection

SCN core collection
integrated applied core collection

WANG et al, 2006, Euphytica
Qiu et al, 2003, Acta Agrona Sin
Ma et al. 2006, J Inter Plant Biol
Qiu et al, 2009, Acta Agrona Sin

Song et al 2010 Chinese Agron Sinica
Guo et al 2013 The Journal Crop
Qiu et al 2013 Plant Molecular Biology
Population structure

In accordance with geographic origin in cultivated and wild soybeans respectively

1863 landraces; 59 SSR

112 wild soybean; 99 SSR, 554 SNP

LI et al. New Phytologist, 2010; LI et al. TAG, 2008
Diversity distribution at molecular level

- An increase in allele size in landrace relative to wild soybean
- Cultivated soybean domesticated along the Yellow River valley

### Average standardized allele size

<table>
<thead>
<tr>
<th></th>
<th>total</th>
<th>Random sampled (50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.soja</td>
<td>-0.404</td>
<td>-0.406</td>
</tr>
<tr>
<td>NESpM</td>
<td>0.008</td>
<td>0.013</td>
</tr>
<tr>
<td>NSpM</td>
<td>-0.113</td>
<td>-0.121</td>
</tr>
<tr>
<td>HSuM</td>
<td>-0.041</td>
<td>-0.042</td>
</tr>
<tr>
<td>SSpM</td>
<td>0.035</td>
<td>0.037</td>
</tr>
<tr>
<td>SSpSM</td>
<td>0.052</td>
<td>0.047</td>
</tr>
<tr>
<td>SSuM</td>
<td>0.078</td>
<td>0.076</td>
</tr>
<tr>
<td>SSuSM</td>
<td>0.024</td>
<td>0.024</td>
</tr>
</tbody>
</table>

*LI et al, 2013, Genet Res and Crop Evol*
Resequencing selected diverged accessions

• Integrated with 17 wild soybeans and 14 cultivated soybeans (Lam et al. 2010)
• Identified 5,102,244 SNPs, in which 25.5% was special to our 25 accessions
2.99% of the whole genomic regions was probably impacted by artificial selection for preferred agronomic traits.
A large variation in extent of LD among chromosomes

Ø A dramatic increase in LD across the whole genome from wild soybean towards cultivated soybean

Ø The LD pattern of the landraces slightly from modern soybeans compared to wild soybeans

The density of genotyping should be designed specific to genomic regions in association mapping.
Clarify the interspecies genetic difference in whole genome

De Novo sequencing wild soybean

Plant
Seed
Photoperiodic reaction
Physiological trait

G.max
G.soja

STRUCTURE

Li et al. New Phytologist, 2010
Variation between *G. soja* and *GmaxW82*

- SNPs 3.63 ~ 4.72 million
- indels 0.50 ~ 0.77 million
- CNV 1,978 genes
- *G. soja*-specific PAV
- *G. max*-specific PAV

Genes affected by PAV, CNV or large-effect SNPs/indels may be useful for future crop improvement.
Genomic basis of agronomic traits

Case
Flowering and maturity time
**Evolution of the *G. max* / *G. soja* species complex**

- *G. soja* diverged from *G. max* more than 0.8 mya
- Nearly 3 times older than a previous estimate of 0.27 mya based on re-sequencing of a single *G. soja* genome
Mapping the important traits

- Domesticating traits
- Genetic improvement traits

**Diversity Materials**
- Selective signal
- 2094 candidate genes

**Natural Population**
- Association
- 158 markers

**Genetic Seg population**
- Linkage
- 30 QTL

**Pedigree population**
- IBD
- 17 repetitive QTL

- SCN
- Seed color
- Fatty acids
- Plant height
- Salinity
- Scn
- Protein
- Content etc
- Yield
- Quality

**BMC Genomics, 2013**
**Mol Breeding, 2009**
**TAG, 2013**
**Conserv Genet, 2011**
**Euphytica, 2006**
**Plant Genome, submitted**
**Euphytica, 2011**
**Plant Breeding, 2009**
Artificial selection of *GmTFL1* in minicore collection

- 4 SNPs nonsynonymous change resulting podding habit change

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of accession</th>
<th>Determinate (%)</th>
<th>Semi-determinate (%)</th>
<th>Indeterminate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Tfl1</em>-b</td>
<td>39</td>
<td>—</td>
<td>2 (5.13)</td>
<td>37 (94.87)</td>
</tr>
<tr>
<td><em>Tfl1</em>-a</td>
<td>22</td>
<td>—</td>
<td>0 (0)</td>
<td>22 (100)</td>
</tr>
<tr>
<td><em>tfl1</em>-ab</td>
<td>20</td>
<td>14 (70)</td>
<td>4 (20)</td>
<td>2 (10)</td>
</tr>
<tr>
<td><em>tfl1</em>-tb</td>
<td>43</td>
<td>35 (81)</td>
<td>6 (13.95)</td>
<td>2 (4.65)</td>
</tr>
<tr>
<td><em>tfl1</em>-bb</td>
<td>14</td>
<td>9 (64)</td>
<td>0 (0)</td>
<td>5 (35.71)</td>
</tr>
<tr>
<td><em>tfl1</em>-ta</td>
<td>24</td>
<td>20 (83)</td>
<td>2 (8.33)</td>
<td>2 (8.33)</td>
</tr>
</tbody>
</table>

Tian et al. PNAS, 2010
Identified new alleles with core collection

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of Germplasm</th>
<th>Yellow (%)</th>
<th>Green (%)</th>
<th>Brown (%)</th>
<th>Black (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmF3'H &amp; GmF3'5'H-a</td>
<td>120</td>
<td>15 (12.5%)</td>
<td>4 (3.3%)</td>
<td>9 (7.5%)</td>
<td>78 (65.0%)</td>
</tr>
<tr>
<td>GmF3'H &amp; GmF3'5'H-b</td>
<td>48</td>
<td>9 (18.8%)</td>
<td>7 (14.6%)</td>
<td>17 (35.4%)</td>
<td>11 (22.9%)</td>
</tr>
<tr>
<td>gmf3'5'h-b &amp; gmf3'5'h</td>
<td>22</td>
<td>20 (90.9%)</td>
<td>2 (9.1%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Guo et al., TAG, 2013

![Seed Coat Colors](image)
CC-based backcross introgression lines

- **Released cultivar**
  - $A \times B$
  - $F_1 \times A$
  - $F_2 \text{BC}_1 \times A$
  - $F_3 \text{BC}_2 \times A$

- **Core collection**
  - Elite lines, Var.
  - RILs, NILs

- Recurrent parent(s): new released adapted varieties
- Screening BIL with favorite traits for mapping genes and breeding
Challenging for high protein breeding in soybean
About 15 proteins were shown to be recognized by sera of soybean-sensitive patients with atopic dermatitis.

Three of them were identified as major allergens.

<table>
<thead>
<tr>
<th>Type</th>
<th>Protein family</th>
<th>Exp. Masses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly m Bd 28K</td>
<td>Cysteine protease C1 Pfam PF00112</td>
<td>26 kDa</td>
</tr>
<tr>
<td>Gly m Bd 30K</td>
<td>Cupin Pfam PF00190</td>
<td>30 kDa</td>
</tr>
<tr>
<td>Gly m Bd 60K</td>
<td>7S Cupin Pfam PF00190</td>
<td>68 kDa</td>
</tr>
</tbody>
</table>
## Soybean allergens

[http://fermi.utmb.edu/SDAP/](http://fermi.utmb.edu/SDAP/)

<table>
<thead>
<tr>
<th>No</th>
<th>Allergen Description</th>
<th>No</th>
<th>Allergen Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P34, 34 kDa maturing seed vacuolar thiol protease; Hydrophobic seed protein; EC</td>
<td>14</td>
<td>Glycinin (legumin, 11S)</td>
</tr>
<tr>
<td>2</td>
<td>Gly m IA</td>
<td>15</td>
<td>Glycinin (legumin, 11S), G1 subunit</td>
</tr>
<tr>
<td>3</td>
<td>Gly m IB</td>
<td>16</td>
<td>Glycinin (legumin, 11S), G2 subunit</td>
</tr>
<tr>
<td>4</td>
<td>Hull allergen; Cytokinin</td>
<td>17</td>
<td>Glycinin (legumin, 11S), G3 subunit</td>
</tr>
<tr>
<td>5</td>
<td>Profilin; GmPRO1</td>
<td>18</td>
<td>Glycinin (legumin, 11S), G4 subunit</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>19</td>
<td>Glycinin (legumin, 11S), G5 subunit</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>20</td>
<td>MP27/MP32 homologue</td>
</tr>
<tr>
<td>8</td>
<td>SAM22, PR-10 protein</td>
<td>21</td>
<td>conglycinin, 7S seed storage protein</td>
</tr>
<tr>
<td>9</td>
<td>β-conglycinin (vicilin, 7S)</td>
<td>22</td>
<td>glycinin, 11S seed storage protein</td>
</tr>
<tr>
<td>10</td>
<td>β-conglycinin (vicilin, 7S); α subunit</td>
<td>23</td>
<td>glycinin, 11S seed storage protein</td>
</tr>
<tr>
<td>11</td>
<td>β-conglycinin (vicilin, 7S); α subunit</td>
<td>24</td>
<td>lectin, agglutinin</td>
</tr>
<tr>
<td>12</td>
<td>β-conglycinin (vicilin, 7S); β subunit</td>
<td>25</td>
<td>Kunitz trypsin inhibitor</td>
</tr>
<tr>
<td>13</td>
<td>β-conglycinin (vicilin, 7S); β subunit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenging

• To develop stable high protein varieties
• To develop high protein varieties in high latitude regions
• To combine high protein and high yield
• To balance the nutrition and processing quality since different components have different functions on the processing product
Some successful cases have proved that Chinese soybean germplasm were powerful materials for gene discovery, allele mining, genomic study, marker development, and breeding.

The platform of soybean core, mini core and integrated applied core collections is helpful for enhancing utilization of soybean genetic resources in sustainable crop improvement for food security.

The efficient utilization of the platform of core collections in the future is relying on accurate phenotyping methods, abundant functional markers, high-throughput genotyping platforms, and effective breeding programs.

Qiu et al. Plant Molecular Biology, 2013
Acknowledgements

Institute of Crop Science, CAAS

University of Georgia  Dr. Scott Jackson
Purdue University Dr. Jianxin Ma
BGI-Shenzhen  Dong Li  Shancen Zhao  Jun Wang  Jian Wang
Novogene Bioinformatics Tech  Dr. Rui-Qiang Li  Guang Yu  Zhou

Funding: State Key Basic Research and Development Plan of China (973)