Soybean Biotechnology, and End-User Benefits

Peter M. Gresshoff
www.cilr.uq.edu.au
p.gresshoff@uq.edu.au

Bundaberg March 2007
Thanks:

Arief Indrasumunar
Attila Kereszt
Ning Nontachaiyapoom
Qunyi Jiang
Yu-Hsiang Lin
Bernard Carroll
Akira Miyahara
Miki, Snow-Li, Dr Ning, Cuc, Aura, Sean, Satomi, Meng
Artem Men, Tripty Hirani, Pick-Kuen Chan
All other CILR colleagues, especially Uli Mathesius and Michael Djordjevic (ANU)

ARC CoE funding, Qld Smart State Initiative, UQ Strategic Funds
Protein- and oil-rich Seeds

Cell division
Nodule induction
Colonisation
Endosymbiosis
Root hair deformation and curling
Nitrogen fixation
Protein- and oil-rich Seeds
- The soybean genome
  - molecular maps
  - molecular markers

- Gene discovery
  - positional cloning
  - TILLING

- Gene expression profiling
  - quantitative target methods
  - gene chips (Affymetrix)

- Gene transfer
  - Agrobacterium-based
Legume phylogeny
Legumes are closely related based on molecular markers

How Many Manhattan Phone Books Does It Take to Print the Soybean Genome?

In small font, closely spaced

- **Human** -- 160 phone books
  - 160 phone books

- **Fruit Fly** -- 6 books
  - 6 books

- **Roundworm** -- 5 books
  - 5 books

- **Yeast** -- 1 book
  - 1 book

- **E. coli** -- ¼ book
  - ¼ book

- **HIV** -- 1 page
  - 1 page

(1000 pages per phone book)
(22 million “letters” per phone book)
The soybean genome?

- 20 chromosomes
- 1,100,000,000 base pairs (ATGC)
- about 36,000 genes
- many duplicated

- US DOE Soybean Genome Project (JGI; 2006 onwards)
- Japanese Soybean Project (2007 onwards)
How Many Genomes?

• 1995
  – First genome published
    \((H. \ influenzaa)\)

• Today
  – Archaea: 26 complete, 26 draft/in progress
  – Bacteria: >293 complete, 561 draft/in progress
  Eukaryotes: 20 complete, 204 draft/in progress
Legume Comparative Map

Basic Genome Info.

- **Mt**: *M. truncatula*, x=8, 500Mbp
- **Ms**: *Lotus japonicus*, x=6, 500Mbp
- **Ps**: *Pea*, x=7, 500Mbp
- **Ca**: *Chickpea*, x=8, 750Mbp
- **Lj**: *L. japonicus*, x=6, 500Mbp
- **Vr**: *Common bean*, x=11, 520Mbp
- **Pv**: *Mungbean*, x=20, 1120Mbp
- **Gm**: *Soybean*, x=20, 1120Mbp

**Galegoid**

**Phaseoloid**
454 Life Sciences “Pyro” Sequencing – the latest in DNA sequencing

• Margulies et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380

• PCR amplification of single DNA molecules and sequencing in picoliter ($10^{-12}$ litre) reactors

• 100-fold more efficient than traditional automated sequencing

• no DNA cloning and plasmid purification required

Now available at UQ Brisbane
### Generation next. Companies racing for the $1000/human genome sequence strive simultaneously for low cost, high accuracy, the ability to read long stretches of DNA, and high throughput.

<table>
<thead>
<tr>
<th>Company</th>
<th>Format</th>
<th>Read Length (bases)</th>
<th>Expected Throughput (million bases/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>454 Life Sciences</strong></td>
<td>Parallel bead array</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Agencourt Bioscience</td>
<td>Sequencing by litigation</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>Applied Biosystems</td>
<td>Capillary electrophoresis</td>
<td>1000</td>
<td>304</td>
</tr>
<tr>
<td><strong>LI-COR Biosciences</strong></td>
<td>Electronic microchip</td>
<td>20,000</td>
<td>14,000</td>
</tr>
<tr>
<td>Microchip Biotechnologies</td>
<td>Parallel bead array</td>
<td>850-1000</td>
<td>7</td>
</tr>
<tr>
<td>NimbleGen Systems</td>
<td>Map and survey microarray</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Solexa</td>
<td>Parallel microchip</td>
<td>35</td>
<td>500</td>
</tr>
<tr>
<td>VisiGen Biotechnologies</td>
<td>Single-molecule array</td>
<td>NA</td>
<td>1000</td>
</tr>
</tbody>
</table>
Use of Mutants to Discover Genes

Forward Genetics
Wild type soybean root

Nodulation zone (was AON minus)

No-Nodulation zone (now AON plus)

Nodulation is internally regulated by Autoregulation of Nodulation (AON) leading to restricted nodule development
NODULATION requires the *Rhizobium*-produced Nod Factor

- Root hair deformation and curling
- Cell division
- Infection
Induction of CELL DIVISION

ROOT HAIR/EPIDERMIS

‘ACTIVATED STATE’

Rhizobium

NODULE

PRIMORDIUM

Maintenance of CELL DIVISION

Soybean

Q

SDI

LCO

NARK

leaf
Isolation and properties of soybean [Glycine max (L.) Merr.] mutants that nodulate in the presence of high nitrate concentrations

(nitrate inhibition/nts mutants/nitrate-tolerant symbiosis/ethyl methanesulfonate mutagenesis/nitrogen fixation)

Bernard J. Carroll, David L. McNeil*, and Peter M. Gresshoff†
Nodule Autoregulation Receptor Kinase

Long-Distance Signaling in Nodulation Directed by a CLAVATA1-Like Receptor Kinase

Iain R. Searle,1,2* Artem E. Men,3† Titeki S. Laniya,3
Diana M. Buzas,3 Inaki Iturbe-Ormaetxe,2‡ Bernard J. Carroll,1,2§ Peter M. Gresshoff3

Proliferation of legume nodule primordia is controlled by shoot-root signaling known as autoregulation of nodulation (AON). Mutants defective in AON show supernodulation and increased numbers of lateral roots. Here, we demonstrate that AON in soybean is controlled by the receptor-like protein/kinase GmNARK (Glycine max nodule autoregulation receptor kinase), similar to Arabidopsis CLAVATA1 (CLV1). Whereas CLV1 functions in a protein complex controlling stem cell proliferation by short-distance signaling in shoot apices, GmNARK expression in the leaf has a major role in long-distance communication with nodule and lateral root primordia.
Sureeporn Nontachaiyapoom
MPMI 2007 (in press)
GmNARK promoter analysis

- pGmNARK contains two functional TATA boxes
- No control by inoculation
- Phloem Regulatory Region was defined

Sureeporn Nontachaiyapoom
MPMI 2007 (in press)
GmNARK localisation in transgenic soybean roots

- A: control (K599)
- D: pGmNARK::GFP
- I: pAtSUC2::GFP
- H: pGmNARK::GUS

Sureeporn Nontachaiyapoom
MPMI 2007 (in press)
**GmNARK** promoter model:

TF1 or TF3 may be LjKlavier, MtLSS or PsSYM28

Sureeporn Nontachaiyapoom  
(CILR, Brisbane), MPMI (in press)
Induction of CELL DIVISION

ROOT HAIR/EPIDERmis

‘ACTIVATED STATE’

Maintenance of CELL DIVISION

PRIMORDIUM

A complex loop of control

leaf

NARK

SDI

NODULE

Rhizobium

Soybean

LCO

Q
The soybean drip system

Petiole Feeding of Liquid Extracts

Blue food dye tracer

Time = 0

8 – 10 hrs later
<table>
<thead>
<tr>
<th>Source material from leaf of</th>
<th>nts1007 Nodule Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>nts1007 (+Bj)</em></td>
<td>high</td>
</tr>
<tr>
<td>WT (- Bj)</td>
<td>high</td>
</tr>
<tr>
<td>WT (+Bj)</td>
<td>low</td>
</tr>
<tr>
<td>WT (+Bj) heat treated</td>
<td>low</td>
</tr>
</tbody>
</table>
Nod factors perception limits
nodule number
Non-Nodulation Mutants of Soybean

Non-nodulation mutants made with EMS
Nodulation mutants are ontogenically blocked in different places.
LysM1 LysM2 LYSM3
Extra-cellular

Intracellular ‘Kinase’ Domain

Ancestral insertion

GmNFR5α

nod139 nn5

GmNFR5β

LysM1 LysM2 LYSM3
Extra-cellular
nod139 contains a naturally induced mutation (insertion) found in many related cultivars but not ancestral genotypes.

Caution: Not all soybean lines have the same background.

Gs : Glycine soja
PI : PI437.654
B* : Bragg wild type segregant
139 : nod139
49 : nod49
W : Williams
NN5 : NN5
nts : nts382, nts1007, nts1116
Cl : Clark
rj1 : rj1
GmNFR1 Mutant Alleles Detection

NFR1α and β receptor components of soybean

NFR1 receptors are Lys M receptor kinases
Agrobacterium-based complementation of the *nod49* mutant

- **nod49+35SGmNFR1α**
- **nod49+Empty Vector**
- **nod49+nPGmNFR1α**

Nodule Interval
**GmNFR1α** complements the *nod49* mutation and its over-expression increases nodule number.

Nodulation of transgenic roots of different soybeans

<table>
<thead>
<tr>
<th></th>
<th>+Empty Vector</th>
<th>+35S Promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>nod49</em></td>
<td>0</td>
<td>279 ± 46</td>
</tr>
<tr>
<td>Bragg</td>
<td>97 ± 25</td>
<td>166 ± 30</td>
</tr>
<tr>
<td>Clark</td>
<td>116 ± 9</td>
<td>236 ± 38</td>
</tr>
</tbody>
</table>

*Number of nodules per root ± standard error; p< 0.05*
Improved **nitrogen status** of composite plants transformed with an overexpressing *GmNFR1 α* gene construct

<table>
<thead>
<tr>
<th>Total N mg/plant</th>
<th>+Empty vector</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nod49</strong></td>
<td>4.2 ± 0.4 a</td>
</tr>
<tr>
<td><strong>Bragg (WT)</strong></td>
<td>54.5 ± 5.6 c</td>
</tr>
</tbody>
</table>

*numbers followed by the same letter for the same measured parameter are not significantly different at P<0.05.*

Plants were generated by *A. rhizogenes* K599 transformation.
Improved **nitrogen status** of composite plants transformed with an over-expressing *GmNFR1α* gene construct

<table>
<thead>
<tr>
<th></th>
<th>Total N mg/plant</th>
<th>+Empty vector</th>
<th>+35S <em>GmNFR1α</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nod49</strong></td>
<td>4.2 ± 0.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>126.5 ± 8.2&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td><strong>Bragg (WT)</strong></td>
<td>54.5 ± 5.6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>74.2 ± 7.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

*numbers followed by the same letter for the same measured parameter are not significantly different at P<0.05. ± standard error

Plants were generated by *A. rhizogenes* K599 transformation.
Increased *Bradyrhizobium* Response (NN/plant) of composite plants transformed with an over-expressing *GmNFR1α*.

<table>
<thead>
<tr>
<th></th>
<th><em>B. japonicum</em> cfu.ml⁻¹</th>
<th>Empty vector (no <em>GmNFR1α</em>)*</th>
<th><em>GmNFR1α</em> + Native 3.4 kb promoter</th>
<th><em>GmNFR1α</em> + 35S promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nod49</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10²</td>
<td>0.0 a*</td>
<td>6.3 ± 3.6 b</td>
<td>134.0 ± 25.4 d</td>
<td></td>
</tr>
<tr>
<td>10⁵</td>
<td>0.0 a</td>
<td>46.5 ± 9.5 c</td>
<td>570.0 ± 40.1 f</td>
<td></td>
</tr>
<tr>
<td>10⁷</td>
<td>0.0 a</td>
<td>97.7 ± 25.4 d</td>
<td>565.3 ± 54.6 f</td>
<td></td>
</tr>
<tr>
<td><strong>rj1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10²</td>
<td>0.0 a*</td>
<td>2.0 ± 1.2 b</td>
<td>41.0 ± 15.3 c</td>
<td></td>
</tr>
<tr>
<td>10⁵</td>
<td>0.0 a</td>
<td>151.3 ± 34.2 d</td>
<td>316.5 ± 28.5 e</td>
<td></td>
</tr>
<tr>
<td>10⁷</td>
<td>0.0 a</td>
<td>143.0 ± 27.0 d</td>
<td>296.0 ± 35.8 e</td>
<td></td>
</tr>
</tbody>
</table>
Increased *Bradyrhizobium* Response (NN/plant) of composite plants transformed with an over-expressing *GmNFR1 α*

<table>
<thead>
<tr>
<th></th>
<th>Empty vector (no <em>GmNFR1a</em>)*</th>
<th><em>GmNFR1a</em> + Native 3.4 kb promoter</th>
<th><em>GmNFR1a</em> + 35S promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nod49</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10²</td>
<td>0.0 a*</td>
<td>6.3 ± 3.6 b</td>
<td>134.0 ± 25.4 d</td>
</tr>
<tr>
<td>10⁵</td>
<td>0.0 a</td>
<td>46.5 ± 9.5 c</td>
<td>570.0 ± 40.1 f</td>
</tr>
<tr>
<td>10⁷</td>
<td>0.0 a</td>
<td>97.7 ± 25.4 d</td>
<td>565.3 ± 54.6 f</td>
</tr>
<tr>
<td><strong>rj1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10²</td>
<td>0.0 a*</td>
<td>2.0 ± 1.2 b</td>
<td>41.0 ± 15.3 c</td>
</tr>
<tr>
<td>10⁵</td>
<td>0.0 a</td>
<td>151.3 ± 34.2 d</td>
<td>316.5 ± 28.5 e</td>
</tr>
<tr>
<td>10⁷</td>
<td>0.0 a</td>
<td>143.0 ± 27.0 d</td>
<td>296.0 ± 35.8 e</td>
</tr>
</tbody>
</table>
Transcriptional Profiling

The first Soybean EST Microarray (Maguire et al, 2002) 4092 genes
Affymetrix Soybean Microarray

# Transcripts
35,611 Soybean
15,421 Phytophthora
7,431 Heterodera (cyst nematode)
Affymetrix Soybean Microarray

# Transcripts
35,611 Soybean
15,421 Phytophthora
7,431 Heterodera (cyst nematode)

22 squares x 58,000 genes
Expression of Gma.736 (Tryptophan Synthase) in Root

- major effect

GmNARK-dependent Leaf Transcriptome Profiling of Systemic Signal Effects after Root Inoculation

Mark Kinkema, CILR 2007
TILLING (Reverse Genetics)

Targeting Induced Local Lesions In Genomes

Using molecular genetics without TRANSGENICS
The TILLING procedure

ENDO 1 is like Cel1: endonclease (patented); cuts single strand mismatch
GmClavata1A predicted protein domain structure and amplicon location

The protein domains prediction was realized at http://smart.embl-heidelberg.de/ using the SMART program (Simple Modular Architecture Research Tool).
Position of the identified mutation within the GmCLV1A amplicon (ps33 screen)
### GmClavata1A and GmNARK mutants

#### GmCLAVATA 1A

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sequence</th>
<th>Effect</th>
<th>SIFT (IC)</th>
<th>Zygouscity DNA Stock</th>
<th>M2 Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>s17C2</td>
<td>g1021r</td>
<td>E765=</td>
<td>0.01 (3.11)</td>
<td>Hetero SB1208</td>
<td>3137</td>
</tr>
<tr>
<td>s19H7</td>
<td>g903a</td>
<td>G726E</td>
<td>0.25 (3.11)</td>
<td>Homo SB1407</td>
<td>301</td>
</tr>
<tr>
<td>s22D6</td>
<td>c306t</td>
<td>T527I</td>
<td>0.00 (3.11)</td>
<td>Hetero SB1680</td>
<td>1172</td>
</tr>
<tr>
<td>s23E8</td>
<td>c365y</td>
<td>P547S</td>
<td>0.00 (3.11)</td>
<td>Homo SB1765</td>
<td>1418</td>
</tr>
<tr>
<td>s23G7</td>
<td>c403y</td>
<td>F559=</td>
<td>0.00 (3.11)</td>
<td>Hetero SB1758</td>
<td></td>
</tr>
<tr>
<td>s24F5</td>
<td>g516a</td>
<td>G597D</td>
<td>0.00 (3.11)</td>
<td>Homo SB1813</td>
<td></td>
</tr>
<tr>
<td>s7G8</td>
<td>g821a</td>
<td>G699R</td>
<td>0.00 (3.11)</td>
<td>Homo SB511</td>
<td></td>
</tr>
<tr>
<td>s9E5</td>
<td>g883r</td>
<td>R719=</td>
<td>0.00 (3.11)</td>
<td>Hetero SB789</td>
<td></td>
</tr>
</tbody>
</table>

**Outcome:** 3 equal sense mutants 5 missense mutants 0 non-sense mutants

#### GmNARK

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sequence</th>
<th>Effect</th>
<th>SIFT (IC)</th>
<th>Zygouscity DNA Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>s13B1</td>
<td>g1072r</td>
<td>M754I</td>
<td>0.00 (2.96)</td>
<td>Hetero SB889</td>
</tr>
<tr>
<td>s13C3</td>
<td>c643y</td>
<td>N611=</td>
<td>0.00 (2.96)</td>
<td>Hetero SB908</td>
</tr>
<tr>
<td>s15D4</td>
<td>c1177t</td>
<td>H789=</td>
<td>0.00 (2.96)</td>
<td>Homo SB1066</td>
</tr>
<tr>
<td>s15D4</td>
<td>t1243g</td>
<td>H811Q</td>
<td>0.00 (2.96)</td>
<td>Homo SB1066</td>
</tr>
<tr>
<td>s17B2</td>
<td>g226a</td>
<td>G472=</td>
<td>0.00 (2.96)</td>
<td>Homo SB1207</td>
</tr>
<tr>
<td>s18B5</td>
<td>g158a</td>
<td>E450K</td>
<td>0.00 (2.96)</td>
<td>Homo SB1305</td>
</tr>
<tr>
<td>s19B2</td>
<td>c305t</td>
<td>P499S</td>
<td>0.00 (2.96)</td>
<td>Homo SB1345</td>
</tr>
<tr>
<td>s19C3</td>
<td>g853r</td>
<td>A681=</td>
<td>0.00 (2.96)</td>
<td>Hetero SB1354</td>
</tr>
<tr>
<td>s20A4</td>
<td>c305y</td>
<td>P499S</td>
<td>0.00 (2.96)</td>
<td>Hetero SB1454</td>
</tr>
<tr>
<td>s5E7</td>
<td>g695r</td>
<td>D629N</td>
<td>0.00 (2.96)</td>
<td>Hetero SB405</td>
</tr>
</tbody>
</table>

**Outcome:** 4 equal sense mutants 6 missense mutants 0 nonsense mutants
Soybean TILLING Results

GmCLV1A

GmNARK

Soybean TILLING Results
Conclusion/hypothesis: CLAVATA1A in soybean controls juvenile nodal integrity, and thus suppresses branching at cotyledonary and unifoliate nodes.

NARK evolved after duplication. CLV1A and NARK may have other functions in the root.
Comparison of LTP gene expression in $LTP_{RNAi}$ plants vs. controls
Advantages of transgenics:

a) correct promoter
b) correct genome position
c) political sustainability
d) allelic series for structure : function
e) speedy if M2 DNA available and pooled
f) easily applied to existing seed collections
Conclusions

Soybean genomics allows structural understanding and prediction

Nitrogen fixation and nodulation can be increased by transgenic and non-transgenic means

We know how to monitor soybean gene function

We can discover, isolate, change and transfer soybean genes
End-user benefits

- Increased nitrogen and phosphorus input
- Improved flowering, rooting, plant structure
- Altered products (phytoestrogens, taste)
- Improved pest resistance
- Altered protein to oil ratios
- Biofuel (biodiesel)