High Incidence of Biallelic Mutation of Floral Genes in CRISPR/Cas9 Transgenic Poplars

Components of Apomixis Workshop
Estefania Elorriaga
Oregon State University

CRISPR might provide robust containment

Strauss lab goal
To develop robust containment technologies for forest trees – facilitate their use in field research and commercially

Thesis project goal
To determine the mutation efficiency of CRISPR-Cas9 nucleases on the poplar floral genes PtLEAFY and PtAGAMOUS

- The ability to reliably generate knock-out mutants in non-model plants has been far from reach for many years
- CRISPR-Cas nucleases might be the answer!

Destroying protein function of essential flowering genes

1. **PtLEAFY (PtLFY)**
   - Transcription factor active in early floral development
   - Determines meristem identity
   - Single copy gene
   - 2 alleles

2. **PtAGAMOUS (PtAG)**
   - MADS box transcription factor
   - Controls differentiation of stamens and carpels
   - Two genes (one being a paralog)
   - Two alleles per genes
   - Before CRISPR Cas nucleases, generating knock-out mutations in two genes simultaneously would have been a daunting prospect

Large-scale field trial provided many informative knock-down events

- 3,414 poplar trees planted in 9 acres
- 3 clones x 2 blocks each
  - 2 female, one male
- 23 constructs, 1-35 events, 1-6 ramets
- Mostly RNAi
- Most targeting LFY and/or AG
- 96% survival since planting in 2011
- Flowering initiated in 2014

Overview of experimental methods

**PtLFY and PtAGs** knock-out plants are expected to produce no flowers or sterile flowers in poplar

1st Generation field trial, 9 acres planted 2011
**Experimental constructs**

**Control construct**

**Nuclease constructs**

**Mutagenesis strategy LFY: Two target sites within one gene**

Target site for LFY1C

>\text{PtLFY1C} \ldots \text{CATGCACCAGTGAAA} \\
GATCACAGAGAGAGACAA \\
GGGGGCAGATAGAT \\
ATG \\
GATCCGGAGGCTTT \\
CACGGCGAGTTTGTTCAAATGGGACACGAGAGCAATGGTGCCACATCCTAACCGTCTGCTTGAAATGGTGCCACCCCCTGCTCAGCAGCCAC \\
CGGCTGCGGCGTTTGCTGTAAGGCCAAGGGAGCTATGTGGGCTAGAGG \\
AGTTGTTTCAAGCTTATGGTATTAGGTACTACACGGCAGCGAAAATAGCTGAACTCGGGTTCACAGTGA \ldots \\
>\text{PtLFY2C}

Target site for LFY2C

**Mutagenesis strategy AG: Two target sites that hit both homologs**

Target site for AG1C

>\text{PtAG1C} \ldots \text{GGATCAGCTAGCTAGACTGCAGCT} \\
ATG \\
GAATATCAAAATGAATCCCTTGAGAGCTCCCCCCTGAGGAAGC \\
TAGGAA \\
GGGGAAAGGTGGAGATCAAG \\
CGGATCGAGAACACCACCAATC \\
GCCAAGTCACTTTCTGCAAA \\
AGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATGCTGAGGTTGCACTCATCG \ldots \\
>\text{PtAG2C}

Target site for AG2C

**Many transgenic events being studied**

<table>
<thead>
<tr>
<th>Construct</th>
<th>No. of regenerated shoots</th>
<th>No. of putative transgenic events</th>
<th>No. of events analyzed molecularly</th>
<th>Percent analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFY1C</td>
<td>1,307</td>
<td>153</td>
<td>131</td>
<td>100%</td>
</tr>
<tr>
<td>LFY2C</td>
<td>1,205</td>
<td>131</td>
<td>46</td>
<td>35%</td>
</tr>
<tr>
<td>LFY1-LFY2C</td>
<td>1,673</td>
<td>107</td>
<td>60</td>
<td>50%</td>
</tr>
<tr>
<td>AG1C</td>
<td>1,149</td>
<td>294</td>
<td>54</td>
<td>18%</td>
</tr>
<tr>
<td>AG2C</td>
<td>1,376</td>
<td>125</td>
<td>12</td>
<td>10%</td>
</tr>
<tr>
<td>AG1-AG2C</td>
<td>1,903</td>
<td>185</td>
<td>79</td>
<td>40%</td>
</tr>
<tr>
<td>Total</td>
<td>8,663</td>
<td>1,006</td>
<td>382</td>
<td>38%</td>
</tr>
<tr>
<td>Cas9 control</td>
<td>970</td>
<td>95</td>
<td>14</td>
<td>15%</td>
</tr>
</tbody>
</table>

**PCR amplification in LFY for amplicon sequencing**

PCR product ~723 bps

- Exon 1
- Exon 2
- Exon 3

- LFY1C target
- LFY2C target

Our PCR-based assay should pick up both alleles, unless there is a very large deletion

**PCR amplification in AG genes for amplicon sequencing**

PCR product ~350 bps

- Exon 1
- Exon 2
- Exon 3

- AG1C target
- AG2C target
Homozygous mutants have “clean” peaks in electropherogram

- A non-mutant transgenic event
- A 2bp deletion homozygous event

Heterozygous mutants have double peaks in electropherogram

- A non-mutant transgenic event
- A heterozygous mutant event

The rate of mutation was high for most populations

<table>
<thead>
<tr>
<th>Construct</th>
<th>GE events sequenced</th>
<th>Type of mutation</th>
<th># of events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single LFY1C</td>
<td>131</td>
<td>Homozygous</td>
<td>40 (30%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>69 (53%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>22 (17%)</td>
</tr>
<tr>
<td>Single LFY2C</td>
<td>46</td>
<td>Homozygous</td>
<td>15 (32%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>28 (62%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>Double LFY1C-LFY2C</td>
<td>60</td>
<td>Homozygous</td>
<td>11 (18%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>44 (74%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>5 (8%)</td>
</tr>
</tbody>
</table>

The rate of homozygous mutation was high across constructs

One construct targeting PtAG was not very active

No mutations in control

<table>
<thead>
<tr>
<th>Construct</th>
<th>GE events sequenced</th>
<th>Type of mutation</th>
<th># of events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single LFY1C</td>
<td>131</td>
<td>Homozygous</td>
<td>40 (30%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>69 (53%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>22 (17%)</td>
</tr>
<tr>
<td>Single LFY2C</td>
<td>46</td>
<td>Homozygous</td>
<td>15 (32%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>28 (62%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>Double LFY1C-LFY2C</td>
<td>60</td>
<td>Homozygous</td>
<td>11 (18%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>44 (74%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>5 (8%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct</th>
<th>GE events sequenced</th>
<th>Type of mutation</th>
<th># of events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single AG1C</td>
<td>54</td>
<td>Homozygous</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>7 (13%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>47 (87%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct</th>
<th>GE events sequenced</th>
<th>Type of mutation</th>
<th># of events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas (empty vector)</td>
<td>14</td>
<td>None</td>
<td>14 (100%)</td>
</tr>
</tbody>
</table>
Large majority of events with single or double mutations

<table>
<thead>
<tr>
<th>Construct</th>
<th>GE events sequenced</th>
<th>Type of mutation</th>
<th># of events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>382</td>
<td>Homozygous</td>
<td>92 (24%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heterozygous</td>
<td>194 (51%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td>96 (25%)</td>
</tr>
</tbody>
</table>

Small indels are frequent for single target nucleases

Most mutations are small indels

Large mutations are common among active double nucleases

Most homozygous mutants for AG1-AG2C had small indels

About 25% of homozygous mutants for AG1-AG2C had DSBs at both target sites

All the homozygous mutants for LFY1-LFY2C had DSBs at both target sites

Most mutations have completely disturbed final protein

Early stop codons due to frameshift indels

Similar disturbances in AGAMOUS mutants

Wild type Wild type sequence

Homozygous mutants

Wild type Wild type sequence

Homozygous mutants

Partial LEAFY peptide sequence

Early stop codons

Large deletions (not pictured) will remove ~40 amino acids!
Early stop codons that disturb the protein function

Wild type

Homozygous mutants

Wild type transgenic

Partial AGAMOUS peptide sequence

Large deletions (not pictured) will remove ~10 and 13 amino acids!

Such high homozygous mutation rates? Are they reliable? Do they indicate “asexual gene drive?”

• To determine allelic mutation frequency, we need to know if the homozygous mutants are real
• We also need to know if the heterozygous mutants have different mutations in both alleles or just one allele
• To answer, we cloned and sequenced both alleles of a subset of our homozygous and heterozygous mutant events

LFY amplicon has four SNPs that can help us identify separate alleles

Topo cloning was used to isolate separate alleles from 17 events

• Selected events at random from LFY1C
  – Heterozygous: 12 events from 69
  – Homozygous: 5 events from 40
• Amplified LFY gene sequence
• Ligated PCR product into TOPO vector
• Sequenced individual colonies for each event
  – Four (p=88%) to ten (p~100%) colonies per event

Most heterozygous mutants have mutations in both alleles

• Out of the 12 heterozygous mutants:

<table>
<thead>
<tr>
<th>Heterozygotes</th>
<th>No. of events</th>
<th>SNPs present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt/Mt1</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Mt1/Mt2</td>
<td>7</td>
<td>Yes</td>
</tr>
<tr>
<td>Mt1/Mt2/Mt3</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Totals</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

• Out of the 5 homozygous mutants:

<table>
<thead>
<tr>
<th>Homozygotes</th>
<th>No. of events</th>
<th>SNPs present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mt1/?</td>
<td>2</td>
<td>Just one haplotype</td>
</tr>
<tr>
<td>Mt1/Mt2</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Totals</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Is “asexual gene drive” happening?

• According to literature, DSBs in recombination partners (sister chromatid or homologous chromosome) strongly increases the frequency of homologous recombination
  – CRISPR literature also cites homozygous mutants in several plant studies
• Is this the source of our high rate of homozygous mutants?
• Need to determine mutant allele frequency and also convergent (same mutation in different events) mutation frequency with high precision
  – Need to sequence both alleles for the entire LFY1C population to get adequate statistical power
In summary: CRISPR-Cas9 nucleases work very well in poplar

- Three of four nucleases were highly active
  - Putative homozygous mutation frequency was ~ one-quarter
- Constructs with a single guide RNA generally lead to small indel mutations (1-5 bp)
- Two active guide RNAs targeting the same locus induced much larger deletions
- Transgenic controls with only the Cas9 gene showed no mutations
- Very promising means for efficient bisexual containment in vegetatively propagated poplar
  - Reproductive phenotyping underway using early flowering FT system

Acknowledgments

Steve Strauss, PI
Cathleen Ma, transformation
Clark Embleton, PCR and sequencing
Analeslie Martinez, tissue culture
Melissa Meyhoff, tissue culture
Amy Klocko, mentoring through entire project
Haiwei Lu, Flower images