BAC sequencing analyses in the hexaploid

*Spartina maritima* (Poaceae):

Homoeolog divergence and microsynteny

in the grass family

Carine Charron¹, Mathieu Rousseau-Gueutin², Julie Ferreira de Carvalho², Houda Chelaifa², Julien Boutte², Gaëtan Dro¹, Joëlle Fourment³, Arnaud Bellec³, Hélène Berges³, Julie Poulain⁴, Arnaud Couloux⁴, Sophie Mangenot⁴, Patrick Wincker⁴, Armel Salmon², Angélique D’Hont¹ and Malika Ainouche²

¹CIRAD, UMR AGAP, Montpellier, France; ²Université de Rennes 1, UMR CNRS 6553, Rennes, France; ³INRA, CNRGV, Castanet Tolosan, France
⁴Genoscope, Evry, France

Réunion annuelle du Groupe Polyploïdie-Cytogénétique

Clermont-Ferrand, France, 22 mai 2014
Chloridoideae: a poorly investigated lineage in the Poaceae family

Christin et al., 2009
**Spartina**: Recurrent hybridization and polyploidy

(Rousseau-Guetin M., Bellot S. *et al.*, In prep)

![Diagram of Spartina phylogeny](image)

*Spartina anglica* C. E. Hubbard:
A natural model system for analysing early evolutionary changes that affect allopolyplid genomes.
(Ainouche *et al.*, 2004)
**Spartina maritima**: Evolutionary and ecological interests

2n = 6x = 60  
2C = 3.7 pg, c.a. 3.7 Gb

**Old-World native**

**Distribution:**
- Species in regression
- Local distribution
- Slow lateral extension

**Biological and genetic diversity:**
- Low seed-set
- Vegetative propagation
- Genetically depauperate populations
- Male parent of S. x townsendii et S. anglica
The *Spartina* challenge: no diploid reference genome

**Strategy:**
Analysis of homoeologous in the hexaploid parents

- **Development of *S. maritima* genomic resources**

  Reference transcriptome assembly
c. a. 17,000 annotated genes
  Ferreira *et al.*, 2013

  Genomic DNA from 454 Roche pyrosequencing
c. a. 100 Mb (A. Salmon)

  BAC library construction
  40,641 high-quality BAC-End Sequences
  Ferreira *et al.*, 2013

- **Micro-analysis of homoeologous sequences, two complementary approaches:**
  - Sequencing of a set of homoeologous BAC clones
  - Intragenomic SNP analyses (NGS data): Julien Boutte, Malika Ainouche and Armel Salmon
Objectives

- To analyze homoeologous regions in the hexaploid genome
- To perform comparative analyses between *Spartina* (Chlorideae) and representatives from other grass subfamilies

**Analysis of hom(oe)ologous chromosome segments (BAC clones) from two regions of the *S. maritima* genome**

**Orthologous region to that containing the *Alcool dehydrogenase 1* gene in *Sorghum* and *Saccharum***

**Region containing the *cinnamyl alcohol dehydrogenase* gene, involved in the lignin biosynthetic pathway**

*S. maritima*  
\( x = 10; 2n=6x; 2C = 3700 \text{ Mb} \)
Translocation containing *Adh1* gene previously detected in the Panicoids (*Sorghum, Saccharum* and *Zea*)

San Miguel et al., 1996 – Science; Tikhonov et al., 1999 – PNAS; Ilic et al., 2003 – PNAS; Jannoo et al., 2007 – Plant Journal
S. maritima BAC sequencing analyses

1. Identification of hom(oe)ologous BAC clones

**S. maritima** BAC library
Coll. H. Berges (CNRGV Toulouse)

70 000 BAC clones (c.a. 110 kb) => 7.7 Gb
= 2x coverage of hexaploid genome

**Probe design**

- **Oryza sativa**
- **Zea mays**
- **Sorghum bicolor**
- **Saccharum hybrid**

BACs hybridizing to a minimum of two probes were identified and extracted

Hybridization of probes corresponding to the target region

Fingerprinting to check BACs
2. Sequencing and assembly of BAC clones – Genoscope

3. Automatic and manual annotation of BAC clones

4. Intra- and inter-species comparisons:

Between hom(oe)logous
*S. maritima* sequences

With orthologous
Poaceae region

http://southgreen.cirad.fr/

Artemis, annotation tool
Adh1 region: Two divergent homoeologs identified

*S. maritima* homeologs

<table>
<thead>
<tr>
<th>BAC</th>
<th>Size</th>
<th>Exons</th>
<th>Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>111 kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>97 kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>98 kb</td>
<td>100 %</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>93 kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>151 kb</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

% identity between homeologous genes:
- Exons: 96 %
- Introns: 91 %

Differential TE insertion pattern between the 2 divergent homoeologous sequences (haplotypes I and II).
Adh1 region - Comparison with orthologous Poaceae chromosome segments

- O. sativa
- B. distachyon
- S. maritima Haplotype I
- S. maritima Haplotype II
- S. italica
- Z. mays Hap I
- S. bicolor Hap II

Haplotype I

Haplotype II

S. spontaneum origin
S. officinarum origin

10 kb
Adh1 region – 18 genes in all Poaceae chromosome segments
Adh1 region – 7 genes absent in S. maritima sequences
Adh1 region – no Adh1 gene in *S. maritima* sequences

- *O. sativa*
- *B. distachyon*
- *S. maritima*
  - Haplotype I
  - 100
  - Haplotype II
- *S. italica*
- *Z. mays*
- *S. bicolor*
- *S. spontaneum*
- *S. officinarum*

**Adh1** region:

- **Adh1 region** – no Adh1 gene in *S. maritima* sequences
- **10 kb**
- **0.05**

**Haplotype**:

- Haplotype I
- Haplotype II
Adh1 region – Translocation after divergence Panicoicoideae/Chloridoideae

- Translocation containing Adh1 gene previously detected in the Panicoids (Sorghum, Saccharum and Zea)
  
  Jannoo et al., 2007; Illic et al., 2003; Tikhonov et al., 1999, San Miguel et al., 1996

- Translocation not shared with Spartina

Translocation occurred after the divergence between Panicoideae and Chloridoideae 45-60 Mya

Oryza sativa

Spartina maritima

Zea mays
Homoeologous region 1
Homoeologous region 2

Sorghum bicolor

Saccharum hybrid
Phylogenetic analysis of the CAD family in monocot plants

*S. maritima* region containing the CAD6 gene
S. maritima homeologs

**Haplotype I**
- BAC 1: 101 kb
  - 100% identity
- BAC 2: 86 kb
- BAC 3: 129 kb
- BAC 4: 98 kb
- BAC 5: 102 kb
  - lots of TE

**Haplotype II**
- 100% identity
- lots of TE

<table>
<thead>
<tr>
<th>Genes</th>
<th>Collinear homeologous genes</th>
<th>Transposable elements (TEs)</th>
</tr>
</thead>
</table>

- **Structural rearrangements:** 2 genes to only one of the two haplotypes
- **Differential TE insertion pattern between the two haplotypes**
Important divergence between the two CAD6 homoeologs:

Insertion/deletion of 9 amino-acids

What about evolution of the *CAD6* gene in other *Spartina* species?

Next steps:

- Primer design
- PCR amplification on gDNA from other *Spartina* species
- Phylogeny reconstruction based on CAD6 sequences
Lot of rearrangements between the orthologous sequences
Two divergent homoeologs detected in both regions

<table>
<thead>
<tr>
<th></th>
<th>% identity exons</th>
<th>% identity introns</th>
<th>Ka</th>
<th>Ks</th>
<th>Ka/Ks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adh1 region</td>
<td>96</td>
<td>91</td>
<td>0.016</td>
<td>0.030</td>
<td>0.54</td>
</tr>
<tr>
<td>CAD6 region</td>
<td>96</td>
<td>92</td>
<td>0.023</td>
<td>0.058</td>
<td>0.47</td>
</tr>
</tbody>
</table>

But 3 homoeologs expected in hexaploids:

- Gene loss – fractionnation?
- Technical bias?
- Incomplete BAC sampling?
**NGS data**

*In silico* haplotype detection from transcriptomic NGS reads (454 + Illumina)

13 genes from the Adh1 region screened

5 genes were retrieved (expressed in cDNA)

42/46 homoeosnps validated

7/10 were 100% validated haplotypes

from 3 to 18 haplotypes / alignment

Haplotype I NGS

Haplotype II NGS

Haplotype III NGS

Gene #13

Boutte *et al.* in prep
First homoeologous genomic regions investigated in *Spartina*

Insights into evolution of complex genomes in *Spartina* genus
**MOB team**
Mathieu Rousseau-Gueutin
Julie Ferreira de Carvalho
Houda Chelaifa
Julien Boutte
Armel Salmon
Malika Ainouche

http://ecobio.univ-rennes1.fr/

**SEG team**
Carine Charron
Angélique D’Hont

**Bioinformatics team**
Gaëtan Droc

http://umr-agap.cirad.fr

**Arnaud Couloux**
Julie Poulain
Sophie Mangenot
Patrick Wincker

http://www.genoscope.cns.fr/spip/