**Abstract**

The Evolutionary Genomics of *Magnaporthe oryzae* (GEMO) project is an attempt to identify the genomic determinants and evolutionary events involved in pathogenesis, host specificity and adaptation. Ten closely related genomes of *M. oryzae* strains and one of the sister species *M. grisea* with different main host and host range were sequenced and analysed [1].

We focused here on the detection and analysis of potentially horizontally acquired DNA regions that we characterized using compositional methods.

We examined the general content, the functional profiles and the potential donors of the putative acquired genes. We reviewed the acquired regions and investigated a few target candidates for pathogenicity.

**Methods**

*Horizontal Transfer detection* [2]

- Tetranucleotide composition in sliding windows, 5kb long, 100bp step
- Kullback-Leibler divergence of composition: windows vs whole genome
- Parametric determination of threshold

*Horizontal Transfer filtering*

- To avoid false positive detection from known repeats, we filtered out:
  - Transposable elements, annotated with REPET base and blast.
  - Homopolymers >100bp, Nannotator, PC
  - Simple repeats, annotated with RepeatMasker [4]

*Potential origin identification*

- Inference with GOHTAM [1], a database of species composition
- Neighbor species with a distance over 160 arbitrary units were discarded

*Functional profile*

- Functional annotation was predicted with InterProScan [5]
- Simple repeats, annotated with RepeatMasker [4]
- Potential origin identification
- Inference with GOHTAM [1], a database of species composition
- Neighbor species with a distance over 160 arbitrary units were discarded

**Potential origin**

<table>
<thead>
<tr>
<th>Class</th>
<th># of regions</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td>3</td>
<td>Gorilla gorilla cyanomaculata 2.1 (x2) / Sinam adenosa 15</td>
</tr>
<tr>
<td>Fungi</td>
<td>6</td>
<td>Humicola grisea / Colletotrichum higginsii (4x) / Cryphonectria hulcenate</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>3</td>
<td>Gnomonia perspectiva / Lysites mucronata / Rhodococcus tenuis</td>
</tr>
<tr>
<td>Algae</td>
<td>3</td>
<td>Pyrenula pyrensissima / Agardhelia percula / Chroomonas sp.</td>
</tr>
<tr>
<td>Chloroplasts</td>
<td>2</td>
<td>Monosiga ovata</td>
</tr>
</tbody>
</table>

**Table 1: Genomes and statistics**

<table>
<thead>
<tr>
<th>Genomes</th>
<th>7B-15</th>
<th>BR29</th>
<th>BR32</th>
<th>CD156</th>
<th>FR13</th>
<th>TH12</th>
<th>PH14</th>
<th>TH16</th>
<th>US71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main host</td>
<td>rice</td>
<td>rice</td>
<td>wheat</td>
<td>Elusine</td>
<td>rice</td>
<td>rice</td>
<td>rice</td>
<td>rice</td>
<td>rice</td>
</tr>
<tr>
<td>Size (Mb)</td>
<td>40.9</td>
<td>40.9</td>
<td>41.9</td>
<td>42.7</td>
<td>43.1</td>
<td>46.3</td>
<td>48.5</td>
<td>49.8</td>
<td>39.1</td>
</tr>
<tr>
<td>Predicted genes</td>
<td>12,827</td>
<td>12,616</td>
<td>14,781</td>
<td>14,415</td>
<td>15,055</td>
<td>20,621</td>
<td>19,811</td>
<td>20,067</td>
<td>13,725</td>
</tr>
<tr>
<td>HT DNA (Mb)</td>
<td>1.22</td>
<td>1.23</td>
<td>2.59</td>
<td>1.55</td>
<td>0.14</td>
<td>0.37</td>
<td>0.67</td>
<td>0.52</td>
<td>0.47</td>
</tr>
<tr>
<td>HT DNA %</td>
<td>2.97</td>
<td>3.0</td>
<td>6.18</td>
<td>3.62</td>
<td>0.32</td>
<td>0.79</td>
<td>1.38</td>
<td>1.05</td>
<td>1.19</td>
</tr>
<tr>
<td>HT Genes</td>
<td>592</td>
<td>374</td>
<td>948</td>
<td>438</td>
<td>41</td>
<td>229</td>
<td>252</td>
<td>175</td>
<td>176</td>
</tr>
</tbody>
</table>

1. The reference strain Mo 70-15 is sequenced with Sanger technology. YB29 is a strain from the M. grisea species.

**Figure 1:** Phylogeny of strains. NeibhorNet network from 6,878 orthogroups. *sativa* US71

**Figure 2:** Functional profile

**Figure 3:** GO enrichment

**Figure 4:** Potential donors

**Table 2: Most credible donors**

**Candidate(s)**

The putative transferred region is in red.

The composition of this region is very similar to the whole genome composition of another fungus. Colletotrichum higginsii, a plant pathogen.

The gene is a repeat assembly of 3 protein domains.

The condensation domain is found in many enzymes which synthesise peptide antibiotics. It catalyses a reaction to form peptide bonds in non-ribosomal peptide biosynthesis. We hypothesise an NRPS activity.

**Conclusion & perspectives**

The genomes of *Magnaporthe oryzae* contain between 0.14 (0.3%) to 3.59 (6.2%) of putative horizontal transfers, likely coming from various kingdoms.

The functional profiling suggests their implication in various processes, including pathogenicity, regulation and cell motion. Early analyses yielded interesting candidates but further work is required for other functional classes. The strain specific transfers will be analysed in the future to assess their implication in the variable host range of the strains. These investigations are expected to enhance the understanding of the evolutionary origins of the features driving pathogenicity phenotype and host specificity.

**Acknowledgments**

Funding: ANR 2009-GENM-029-01. Travel grant from the SFBL many many thanks! Special thanks to Sophie Schbath and Philippe Leroy.