Combinatorics of \textit{cis}-regulatory elements in osmotic stress response of Ascomycetes.

Aurian García-González
Mentors: Sushmita Roy, Jay Konieczka, Dawn Thompson
Broad Institute- Summer Research Program 2011
August 3, 2011
Question:

- Where does phenotypic diversity come from?
Question:

- If their gene content is so similar, why are they so different?
How do we study evolution of gene regulation?

Model system

~ 300 MYA

~ 95 MYA

How do we study evolution of gene regulation?

Model system

Phenotype: osmotic stress

How do we study evolution of gene regulation?

Model system

Phenotype: osmotic stress

- Readily tractable experimentally.
- Thoroughly described in *S. cerevisiae* as Hog1-directed.
Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles
Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**

Clusters with similar expression – regardless of gene content.
Characterizing osmotic stress response

1) Perform stress experiments to obtain expression profiles

2) Cluster genes to identify phenomenologs
Clusters with similar expression – regardless of gene content.

3) What are the cis-regulatory elements present in these phenomenologs?
Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**
   Clusters with similar expression – regardless of gene content.

3) What are the **cis**-regulatory elements present in these phenomenologs?
Characterizing osmotic stress response

1) Perform stress experiments to obtain expression profiles

2) Cluster genes to identify phenomenologs
   Clusters with similar expression – regardless of gene content.

3) What are the cis-regulatory elements present in these phenomenologs?
Use ModuleDigger to find CRMs

ModuleDigger:
Uses motif data for genes in a cluster and computes hierarchical scores for modules.

Use ModuleDigger to find CRMs

ModuleDigger:
Uses motif data for genes in a cluster and computes hierarchical scores for modules.

Geneset specificity score \( f_g \) for the motifs in each gene

\[ \text{S. cerevisiae} \]

orf 1
orf 2
orf 3
orf 5
orf 6
Use ModuleDigger to find CRMs

ModuleDigger:
Uses motif data for genes in a cluster and computes hierarchical scores for modules.

Studying combinations of motifs in CRMs
Studying combinations of motifs in CRMs
Studying combinations of motifs in CRMs

S. cerevisiae

More frequent

Less frequent
Studying combinations of motifs in CRMs
Studying combinations of motifs in CRMs

Are they conserved across all species?
Studying combinations of motifs in CRMs

Are they conserved across all species?

NO
Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

Most induced

\( RCS1, SIP4 \)

Most repressed

\( RCS1, SIP4, MBP1, TEC1, MBP1, PDR3 \)
Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

Most induced

RCS1, SIP4

Most repressed

RCS1, SIP4
MBP1, TEC1
MBP1, PDR3

Post-WGD

Most induced

MIG1, MSN2/MSN4
MIG1, ADR1
MSN2/MSN4, ADR1

Most repressed

MIG1, MSN2/MSN4
MIG1, ADR1
MSN2/MSN4, ADR1

Motif combinations are conserved in clades
Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

Most induced

RCS1, SIP4

Most repressed

RCS1, SIP4

MBP1, TEC1

MBP1, PDR3

Post-WGD

Most induced

MIG1, MSN2/MSN4

MIG1, ADR1

MSN2/MSN4, ADR1

Most repressed

RCS1, SIP4

CRZ1, SIP4

Pathogens

Most induced

RCS1, SIP4

MIG1, MSN2/MSN4

MIG1, ADR1

MSN2/MSN4, ADR1

Most repressed

RCS1, SIP4
Stress-related regulatory elements co-occur with glucose-related elements

Pre-WGD
- *S. cerevisiae*
- *S. bayanus*
- *C. glabrata*
- *S. castelli*
- *K. lactis*
- *C. albicans*
- *S. pombe*

- RCS1
- SIP4
- MBP1
- TEC1
- MBP1
- PDR3

Post-WGD
- *S. cerevisiae*
- *S. bayanus*
- *C. glabrata*
- *S. castelli*
- *K. lactis*
- *C. albicans*
- *S. pombe*

- MIG1
- MSN2/MSN4
- ADR1
- RCS1
- SIP4
- CRZ1
- SIP4

Pathogens
- *S. cerevisiae*
- *S. bayanus*
- *C. glabrata*
- *S. castelli*
- *K. lactis*
- *C. albicans*
- *S. pombe*

- Stress
- Glucose Metabolism
- Cell-cycle progression
- Metabolite transport
Conclusions

• Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.
Conclusions

• Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.

• The most induced clusters showed more motif pairs than the repressed clusters.
Conclusions

• Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.
• The most induced clusters showed more motif pairs than the repressed clusters.
• Glucose responsive elements occur with stress elements.
Future work

• Explore higher order combinations.
Future work

• Explore higher order combinations.
• Develop more sophisticated algorithms to discover significant CRMs.
Future work

• Explore higher order combinations.
• Develop more sophisticated algorithms to discover significant CRMs.
• Correlate trans interactions with observed phenotype.
Acknowledgements

Mentors:

Sushmita Roy
Jay Konieczka
Dawn Thompson
Aviv Regev
Regev Lab

SRPG:

Eboney Smith
Bruce Birren
Lori Thomae
Nicole Edmonds
Broad staff