Domain Cooccurrence
Distribution of Genetic Regulation Activation from an Evolutionary Perspective

Arli Parikesit
Chair of Bioinformatics, Department of Computer Science
University of Leipzig
The complexity of gene regulatory mechanism

The complexity of gene regulation is due to a variety of mechanisms that are conceptually different. They are:

- Chromatin regulation: Facilitate DNA organization and prevent DNA aggregation and tangling which is important for replication, segregation, and gene expression.
- Transcription factor: Controls the transfer (transcription) of genetic information from DNA to RNA.
- MicroRNA: Emerged as an abundant class of regulatory genes that regulate protein expression.
The Necessity of Gene complexity

- More regulatory states.
- Higher stability (precision).
- More fine grained regulation.
Difference in the programmes of gene expression is epigenic.

Development is, by definition, epigenetic. Reik (2007) hypothesizes, that differences in the programmes of gene expression that result in the development of different organs and tissues occur without changes to the sequence of our DNA.
Fundamental and Functional unit of protein structure is the domain.

Independent of neighboring sequences, a domain folds into a distinct structure and mediates a biological functionality.
Research Objective

- Description of the functional evolution of the regulatory mechanisms. This will allow us to have insights into major innovations, adaptive changes, and the concepts of regulatory network evolution.

- Description of the occurrence of the gene regulation domain. Ongoing research: Micro RNA and Transcription factor activator domains. Evolutionary relation
Protein components involved in regulation by miRNA
Micro RNA Regulation Domains
RISC Protein

The cofactors beside AGO are still unknown
**Method**

Domain Distribution on Phylogenetic trees:
- Extract the number of genes with domain X for every species.
- Compute the domain cooccurrence.
- Map domain cooccurrence into the phylogenetic tree.

Tools used: Biofuice (Trans factor) and Sonja's SUPER script
Map GO-term “transcription factor activator” into gene-ids

- Used by mediator for mapping/operator execution

- **Domain model** indicates available object types and relationships
## Result and Discussion

<table>
<thead>
<tr>
<th>No</th>
<th>Protein</th>
<th>Found (max 912)</th>
<th>Missing (max 912)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Argonoute</td>
<td>156</td>
<td>756</td>
</tr>
<tr>
<td>2</td>
<td>DGCR8</td>
<td>0</td>
<td>912</td>
</tr>
<tr>
<td>3</td>
<td>Dicer</td>
<td>821</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>TRBP</td>
<td>912</td>
<td>0</td>
</tr>
</tbody>
</table>

- Only TRBP can be found in all domains of life
- The DGCR8 related domain was not found yet
- Still search for Drosha's domains
Domain interaction Matrix: PIWI-PAZ interaction (AGO) and WW domain (DGCR8) is available in every Euchariotic clades. RIBOc (Dicer and Drosha) and RIBO-DSRM (Dicer and Drosha) are roughly available as well.
Functional annotation: Proline Binder is available in every clades. RNA Binder, RNA degradation-RNA Binder, and RNA degradation are roughly available.
Protein Annotation Problem: Only Argonout which finely annotated.
The Transcription Factor Phylogenetic trees and Domain Cooccurrence are still on-going research
Conclusion

- TRBP has complete phylogenetic annotation.
- No phylogenetic tree for DGCR8.
- PIWI-PAZ interaction (AGO) and WW domain (DGCR8) is available in every Euchariotic clades. RIBOc (Dicer and Drosha) and RIBO-DSRM (Dicer and Drosha) are roughly available as well.
- Proline Binder is available in every clades. RNA Binder, RNA degradation-RNA Binder, and RNA degradation are roughly available.
- Protein annotation problem: only argonoute which finely annotated
Outlook

• Computing Domain Cooccurrence of Translation and Post Translation Domain activator.
• Improving the annotation technique
Reference

Reference

THANKS TO...

Sonja Prohaska
Axel Wintsche
Toralf Kirsten
Peter Stadler
DAAD
THANK YOU FOR YOUR ATTENTION