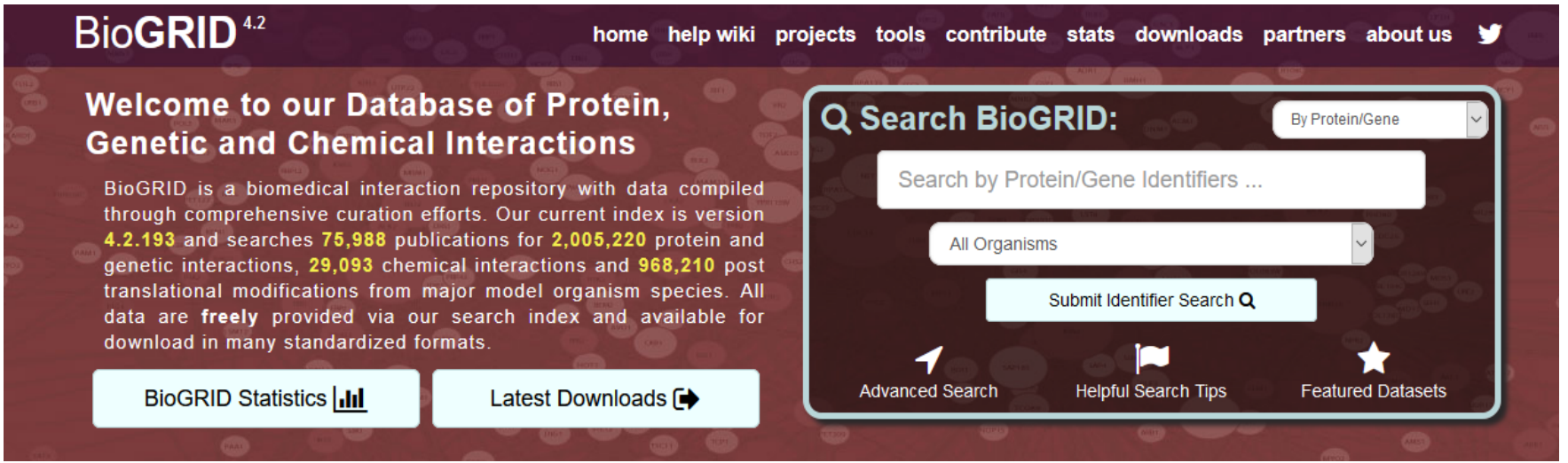
Graph-based methods for modeling biomolecular networks

Gatis Melkus

Supervisor: Dr. sc. comp. Juris Vīksna

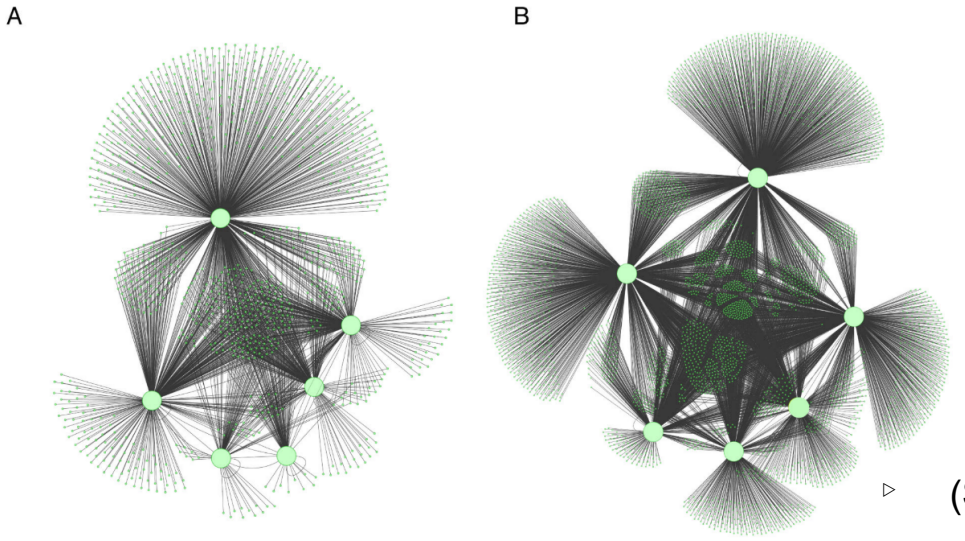
The definition of biomolecular data

- **Biomolecular data** is generated through molecular biology experiments, particularly in high-throughput experimental methods.
- Current methods, including **single-cell** and **-omics** experiments, generate data that absolutely requires extensive machine interpretation.



Biomolecular interaction data

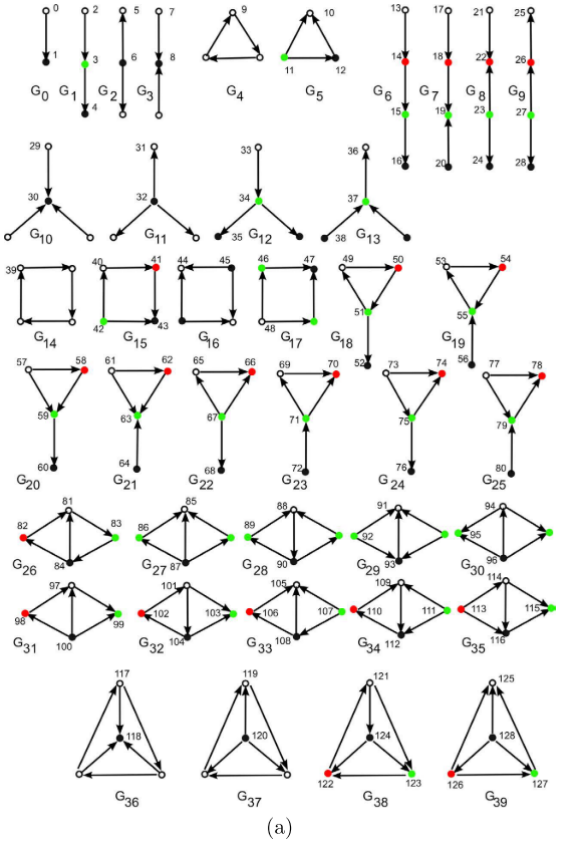
- Molecular methods are increasingly capable of producing not only part lists of nucleic acids and proteins, but entire **interactomes**.
- Systematizing interaction data into functional **networks** is crucial to understanding the actual functions of the cell's molecular machinery.



- **PROBLEM:** even with a lot of data, networks remain challenging to meaningfully analyze.
- (Sorrells et al. 2015)

Graph-based studies of networks

- **Network biology** is focused on the analysis of graphs generated by linking either recorded or inferred interaction data.
- **Biomolecular networks** generated from this data can then be studied further with methods such as:
  - Topological examination via graph properties
  - Integrative comparisons with other data

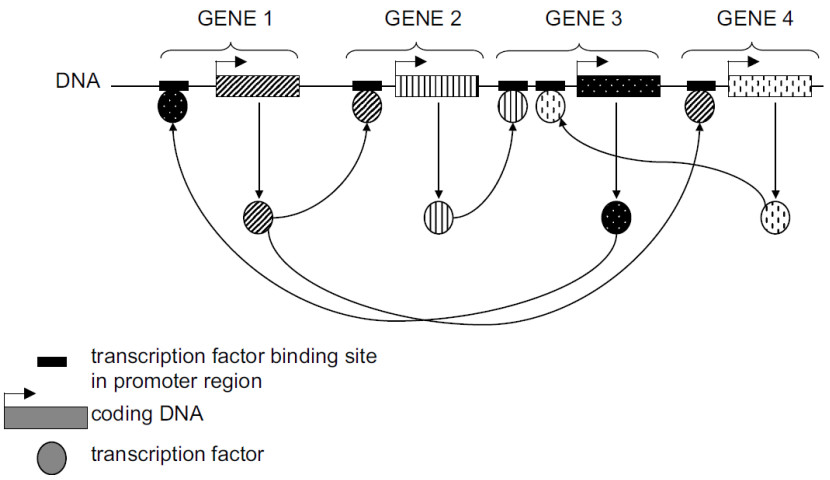


(a)

(Sarajlić et al. 2016)

Gene regulatory networks

- **Gene regulatory networks or GRNs** are constructed from data about transcriptional regulatory interactions, with genes serving as nodes and interactions as edges.
- Gene expression (transcription) is one of the fundamental mechanisms of life and is regulated extensively through biological mechanisms.



- (Schlitt et al. 2015)

Motifs in regulatory networks

- Topological analysis in regulatory networks often relies on overrepresented subgraphs, or **network motifs**.
- Despite early enthusiasm, generally insufficient to derive consistent conclusions about network function.

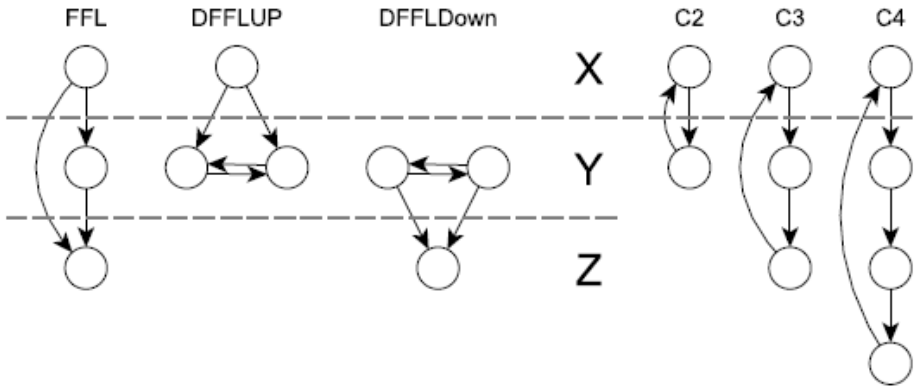


Fig. 1. Generalized forms of commonly considered network motifs.

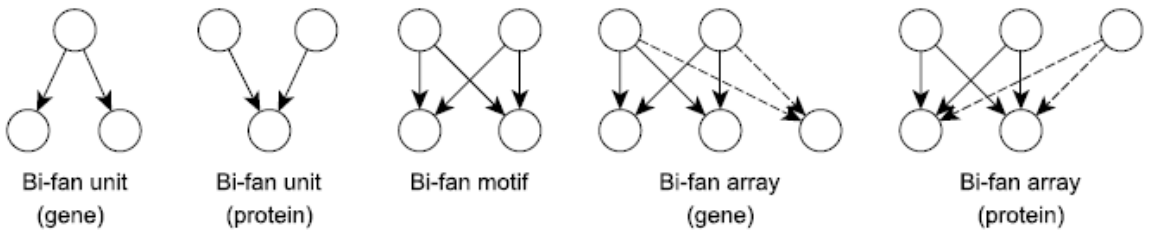


Fig. 2. The bi-fan unit, bi-fan motif and bi-fan array, with variants for gene and protein pairs.

Motifs as applied to paralogs

- **Paralogs** are genes with a shared origin, duplicated at some point in their evolutionary history.
- The **bi-fan** motif is noticeably enriched for paralogs in GRNs, indicating that evolutionary relatedness might lead to symmetry in regulatory networks.

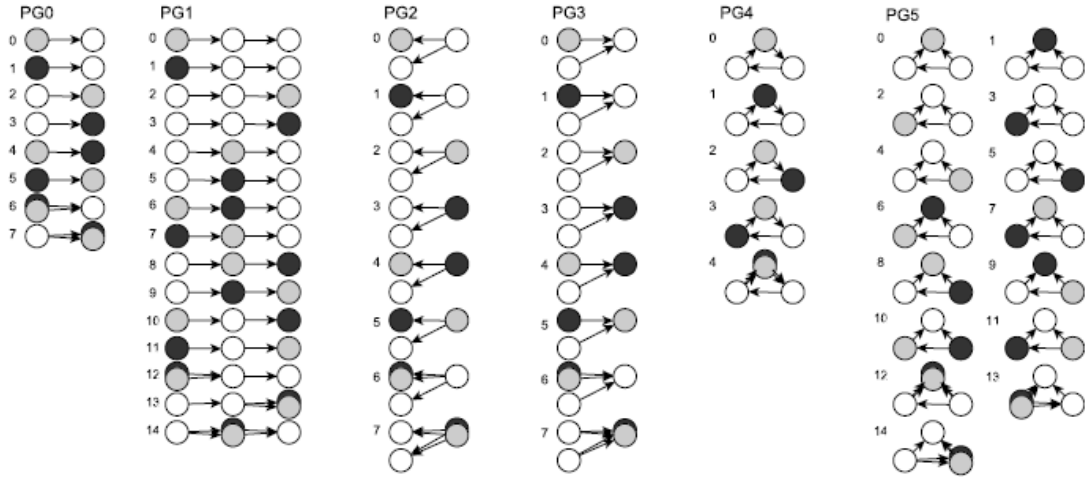


Fig. 3. All possible 3-vertex motifs in a directed graph with respect to a gene pair. Black and gray vertices denote two members of a given gene pair in a particular orbit of the motif. Variations also include 4-vertex motifs with "twinned" positions where a bi-fan unit is embedded, denoted by a gray vertex overlapping a black vertex. The motifs shown include FFLs (PG5), feedback loops (PG4), bi-fan motifs (PG2-7 and PG3-7) as well as bi-fan units (PG0-6 and PG0-7).

Motifs as applied to paralogs

- Observing the symmetry in feed-forward loops in yeast showed a noticeable increase of symmetrical motifs among paralogous pairs compared to randomly selected gene pairs.
- The difference was especially pronounced in a set of paralogs traceable to a whole genome duplication.

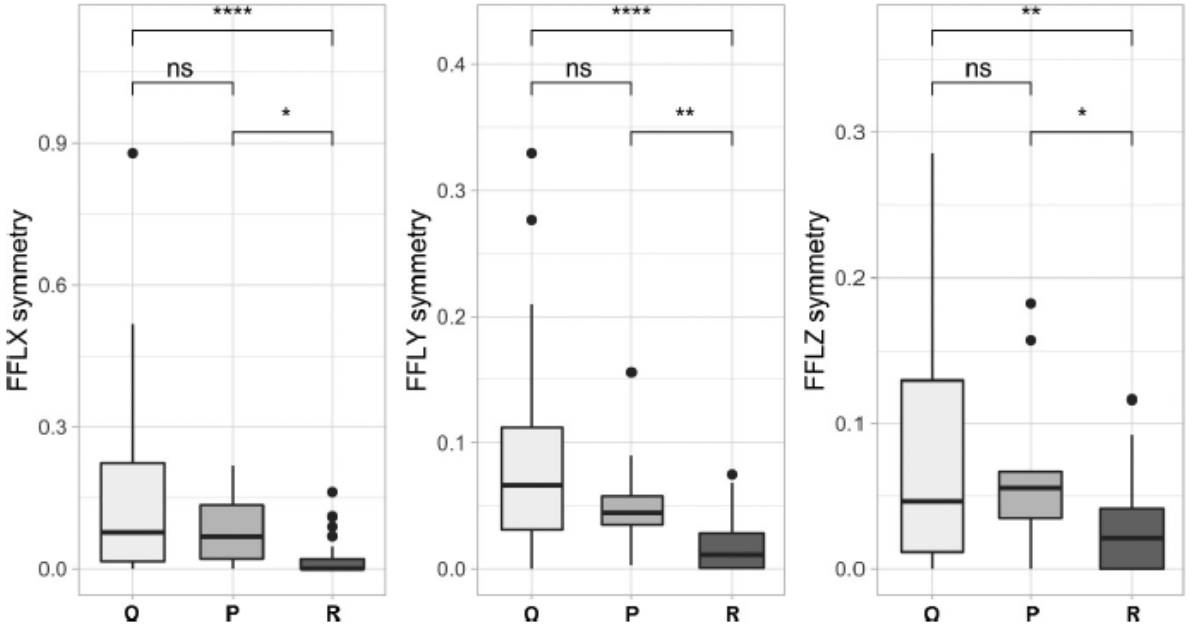


Fig. 6. The ratio of symmetrical feed-forward loops to asymmetrical loops as observed in ohnologue ($\mathbf{O}, n = 29$), nonWGD paralogue ($\mathbf{P}$, $n = 9$) and unrelated ($\mathbf{R}, n = 38$) pairs of TFs in yeast. Ratios are considered separately for symmetry at X, Y and Z positions. Differences in means were assessed for statistical significance by pairwise Wilcoxon tests corrected for multiple comparisons.

Motifs as applied to paralogs

- Furthermore, when examining symmetry within paralog classes, paralogs within the species for baker's yeast as well as *C. elegans* showed increased symmetry compared to genes which only shared a protein family.

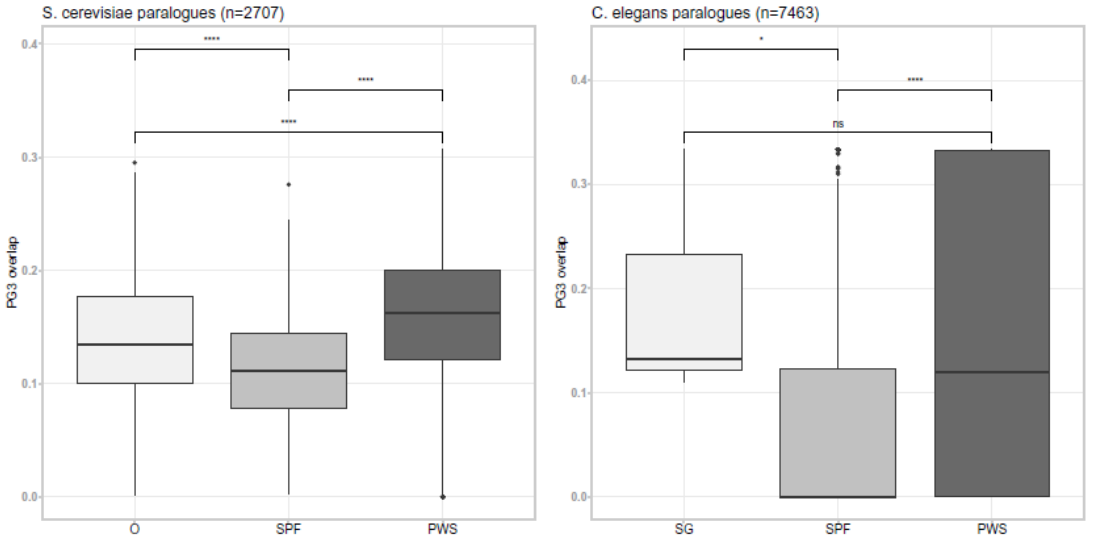


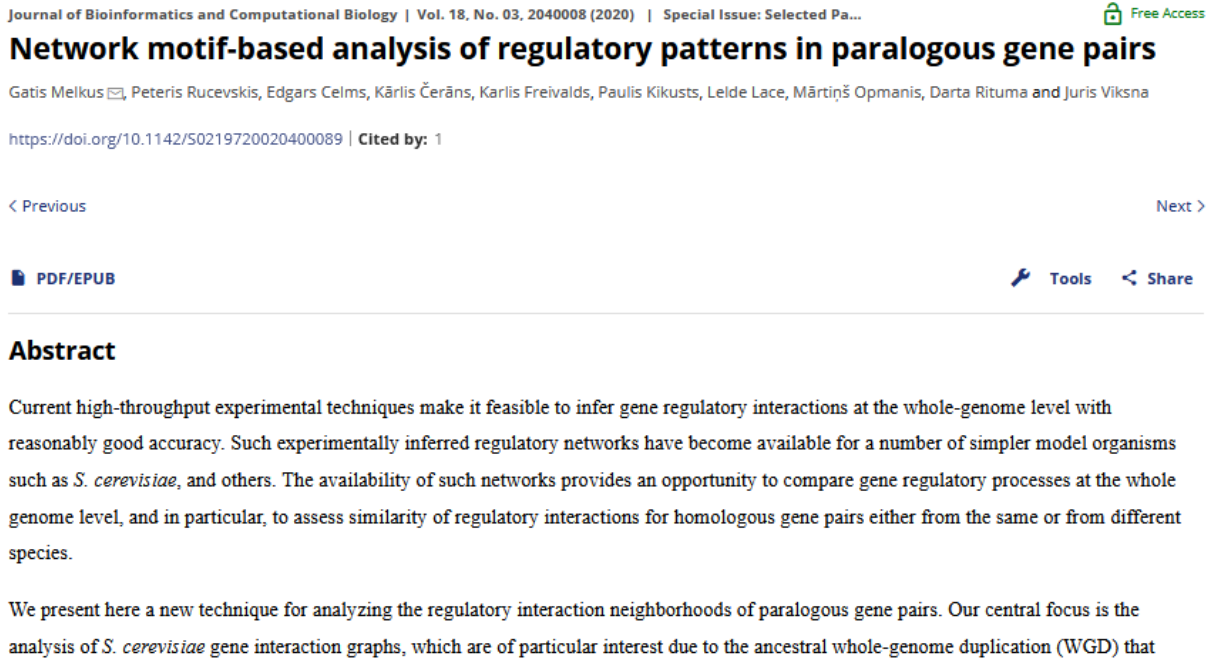
Fig. 7. Interaction overlap for PG3 motifs (including bi-fan motifs) in worm and yeast networks for different paralogue classes. Differences in means were assessed for statistical significance by pairwise Wilcoxon tests corrected for multiple comparisons.

Notes: Abbreviations: **O** — ohnologue, **SPF** — same protein family, **PWS** — paralogue within species, **SG** — split gene.

Topological symmetry due to relatedness?

- While functional implications for pure network topology are hard to pinpoint, linking topology to ancestry directly through symmetry could form the basis of a new network analysis approach.

- Our findings were published in the Journal of Bioinformatics and Computational Biology Vol. 18, No. 3.



Journal of Bioinformatics and Computational Biology | Vol. 18, No. 03, 2040008 (2020) | Special Issue: Selected Pa...

Free Access

Network motif-based analysis of regulatory patterns in paralogous gene pairs

Gatis Melkus, Peteris Rucevskis, Edgars Celms, Kārlis Čerāns, Karlis Freivalds, Paulis Kikusts, Lelde Lace, Mārtiņš Opmanis, Darta Rituma and Juris Viksna

https://doi.org/10.1142/S0219720020400089 | Cited by: 1

< Previous Next >

PDF/EPUB Tools Share

Abstract

Current high-throughput experimental techniques make it feasible to infer gene regulatory interactions at the whole-genome level with reasonably good accuracy. Such experimentally inferred regulatory networks have become available for a number of simpler model organisms such as *S. cerevisiae*, and others. The availability of such networks provides an opportunity to compare gene regulatory processes at the whole genome level, and in particular, to assess similarity of regulatory interactions for homologous gene pairs either from the same or from different species.

We present here a new technique for analyzing the regulatory interaction neighborhoods of paralogous gene pairs. Our central focus is the analysis of *S. cerevisiae* gene interaction graphs, which are of particular interest due to the ancestral whole-genome duplication (WGD) that

Further questions on regulatory networks

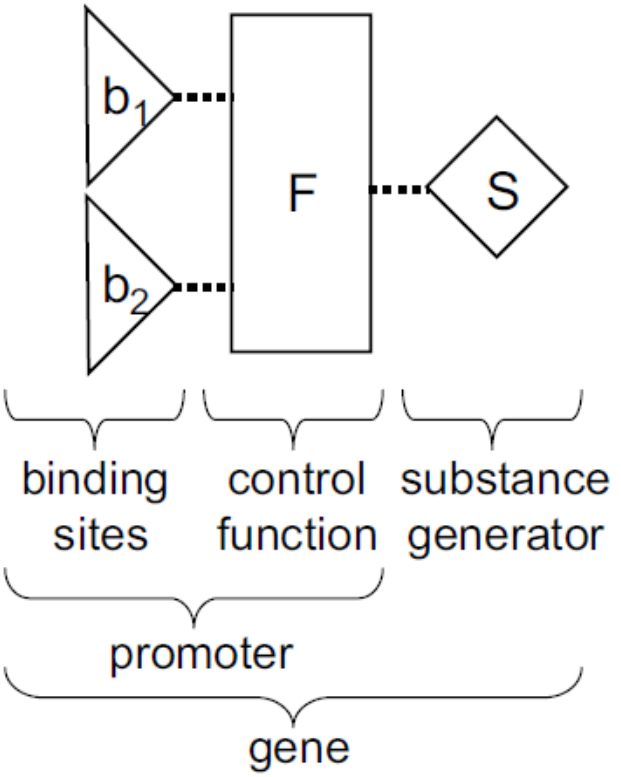
- Extending basic knowledge of interaction topology.
- Regulatory changes across time periods and within disparate spaces.
- Efficient integration of various experiments into a single knowledge base.
- Deriving conclusions from biological networks useful in predictive applications.

Hybrid modeling approaches

- Most GRN models are Boolean: an interaction exists or it does not.
- Actual regulatory functionality is dynamic: regulatory proteins bind to specific sites in particular concentrations dictated by many interdependent systems, and activities are often concentration-dependent.
- Fully dynamic models for complex networks are impractical to model, which creates a need for **hybrid models**: their more straightforward elements are Boolean while dynamic modeling is used at decisive points.

Finite state linear models

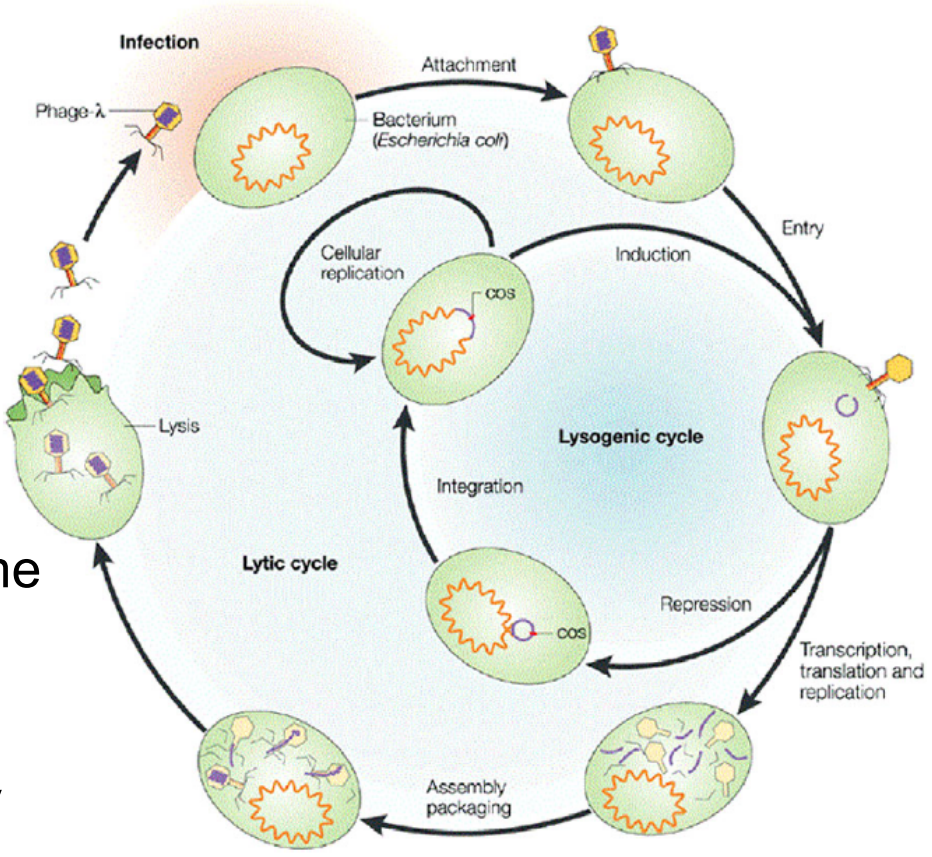
- The finite state linear model is built on several elements which represent a gene and its corresponding product:
  - Binding sites
  - Substance generator
  - Control function
- Each element is adaptable to permit either a boolean or a dynamic implementation of parts of the network.



- (Schlitt et al. 2007)

Bacteriophage λ

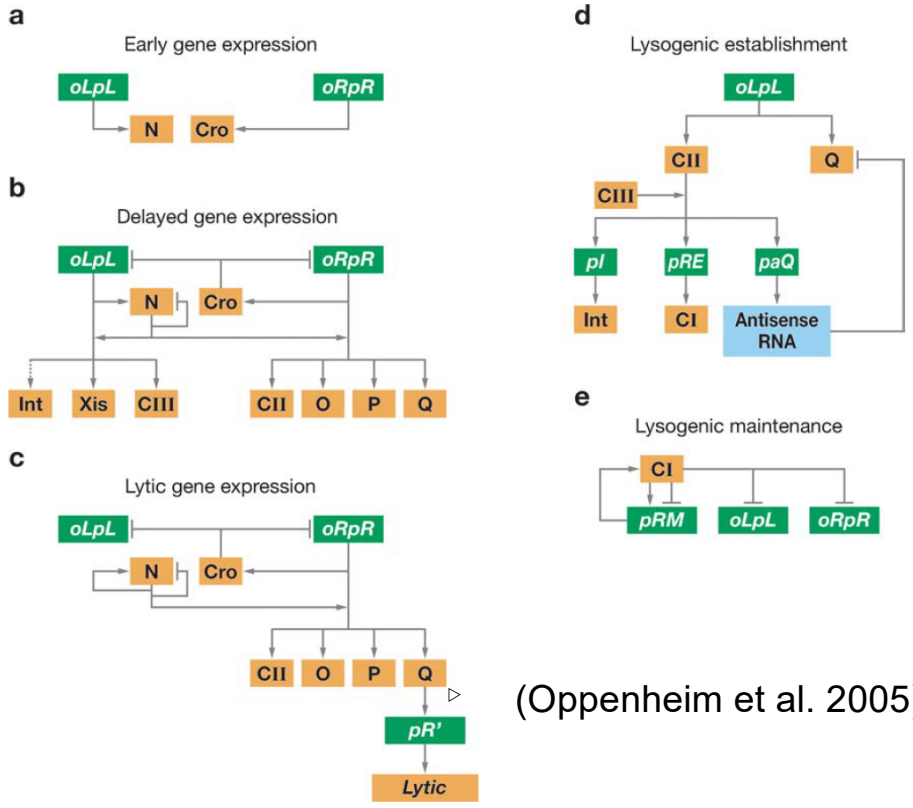
- A well-understood virus that can be practically modeled on its own.
- Acts as a temperate bacteriophage, either destroying its E. coli host (*lysis*) or integrating itself into the host genome and laying dormant (*lysogeny*).
- Possesses a well-characterized molecular switch that determines entry into lysis or lysogeny.



- (Jamal et al. 2018)

The lytic/lysogenic switch

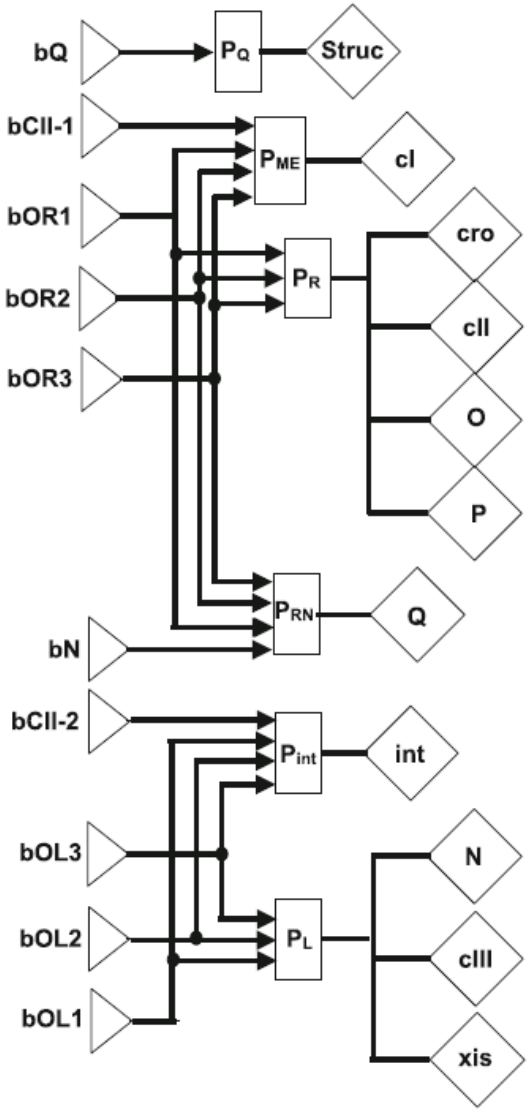
- Dependent on fewer than 20 genes to operate characterized by competition between two key factors:
  - Q antiterminator, the lytic determinant,
  - cII, the lysogenic determinant.
- The simplicity of the switch allows us to efficiently diagram the entire state space of the regulatory network.



(Oppenheim et al. 2005)

An FSLM for λ

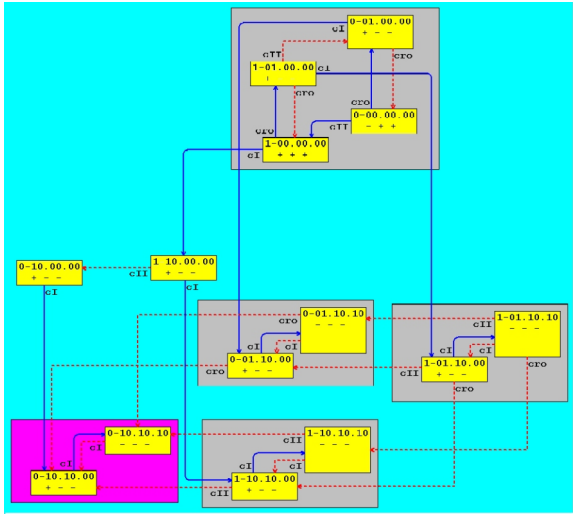
- The lytic/lysogenic switch can be modeled entirely with an FSLM, with the primary dynamic elements being the complex binding process for the promoters pR and pL.
- Our FSLM allows for various analyses of the network, including the possibility for a variety of starting conditions that could model, for example, multiplicity of infection.

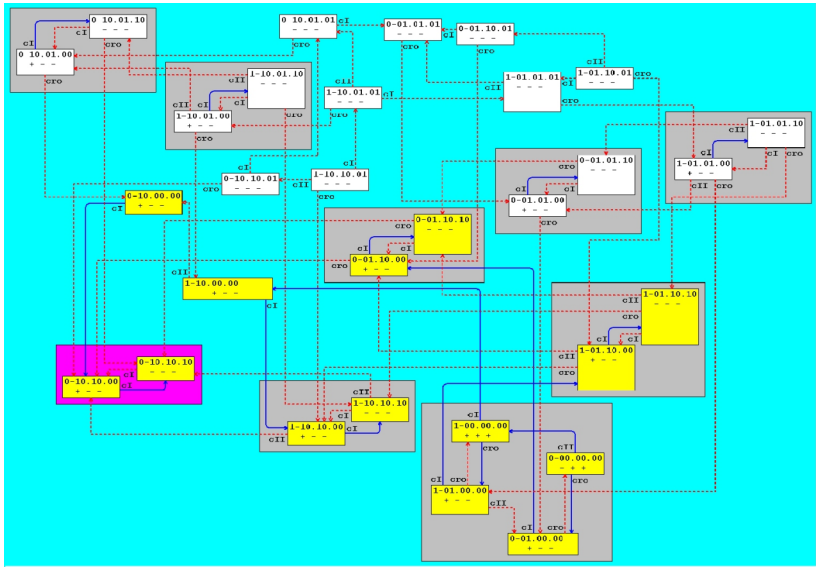


- (Rukliša et al. 2019)

Insights from the λ FSLM

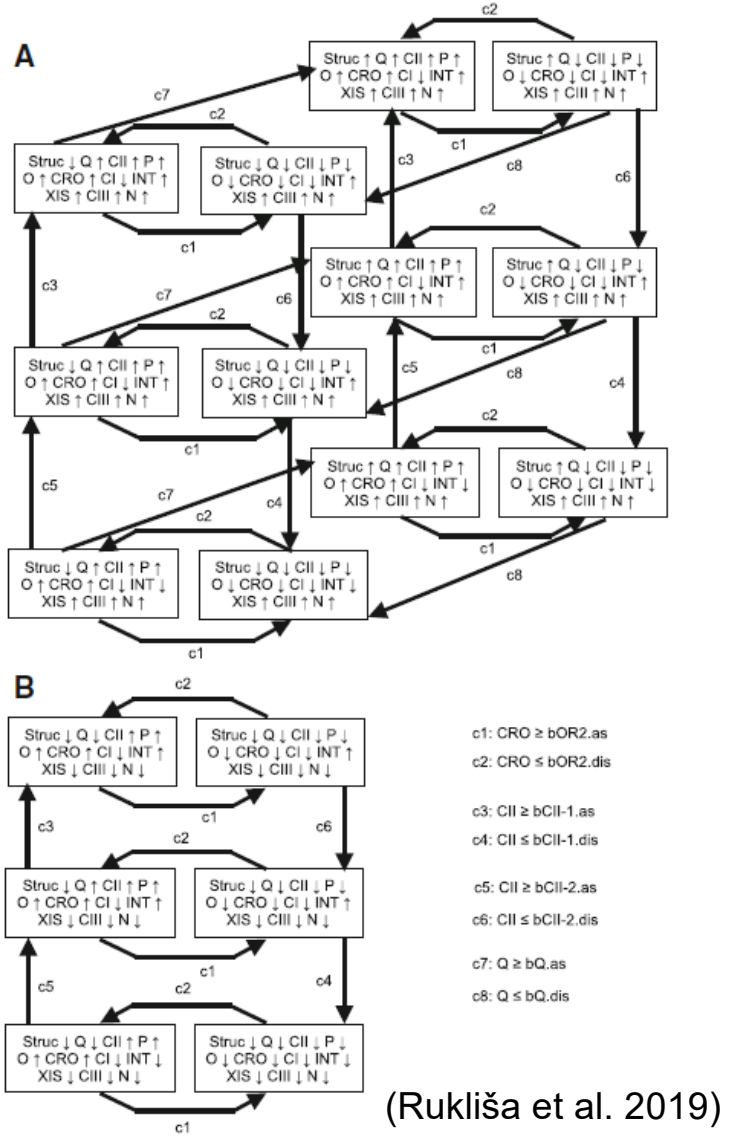
- Our modeling approach allows for several interesting observations.
- Notably, the FSLM allows us to find attractors in a biomolecular network which in the case of the bacteriophage map well to the lytic and lysogenic states.
- These attractors can be studied, which allows us to evaluate the key conditions required for their establishment.





Insights from the λ FSLM

- Furthermore, our model allows us to detect additional states potentially outside of a strict lytic/lysogenic dichotomy.
- The model we use is additionally quite receptive to additional information, allowing us to model further variables according to our needs.
- The FSLM can generate testable predictions about phage behavior possible to confirm in a practical setting.



(Rukliša et al. 2019)

Hybrid modeling: where next?

- We intend to continue updating our phage model and adapting it to new contexts and viruses such as the Mu bacteriophage.
- Using a flexible enough set of starting conditions and modeling adaptations, we could feasibly model phage “voting”, a process which is known to occur within a bacterium infected with multiple phages.
- At least one publication on the topic should be forthcoming by June.

Summary

- Graph-based methods for the interpretation of biomolecular data are broadly useful in deriving systematic molecular knowledge.
- Topology-based analyses, while limited, can still be mapped to actual biological properties.
- For more in-depth modeling, development of hybrid models that can handle the logic and complexity of biiological networks is crucial.
- Viral models, being largely autonomous subunits inside of a more complicated host, are ideal for hybrid model development.

Thank you!