GMR
The Original

EVOLUTIONARY CONSERVATION AND DIVERGENCE OF REGULATORY GENE NETWORKS

Dr. Nallusamy Duraisamy[1], Swetha[2], Dr. Sreenivasan Vadivelu[3], Tarang Bhatnagar[4], Dr. Manoj A. Vora[5]

[1.] Scientist, Department of Research Meenakshi Academy of Higher Education and Research, India, Email: nallu@maher.ac.in
[2.] Assistant Professor Meenakshi College of Allied Health Sciences, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, India, Email: swethaahs@maher.ac.in
[3.] Associate Professor, Department of Respiratory Medicine, ORCID: https://orcid.org/0009-0008-3712-074X
[4.] Centre of Research Impact and Outcome, Chitkara University Rajpura – 140417, Punjab, India,
Email: tarang.bhatnagar.rop@chitkara.edu.in, ORCID: https://orcid.org/0009-0004-6880-8716
[5.] Assistant Professor, Faculty of Science Gokul Global University, Sidhpur, Gujarat, India,
Email: manojvora.chem.gsc@gokuluniversity.ac.in, ORCID: 0000-0002-9603-4576

ABSTRACT

Regulatory gene networks are key players in the regulation of complex biological characteristics by coordinating the expression of genes via transcription factors, enhancers, promoters and other cis-regulatory factors. These networks do not exist in isolation but they change throughout the history of the species and influence both the general biological processes conserved and species-specific changes. Nevertheless, it is still not fully understood how much regulatory architectures are conserved or diverged throughout evolution especially at the systems level. This paper seeks to explore evolutionary conservation and divergence of regulatory gene networks through an integrative computational genomics framework. The method is a combination of comparative genomics, ortholog mapping, GWAS and eQTL integration, transcription factor binding analysis, regulatory annotation data, and network reconstruction methods to simulate regulatory interactions across species.

These findings indicate that there are conserved central regulatory modules that are preserved despite evolutionary distances and therefore propose intense selection on key biological activities. Conversely, they find species-specific regulatory hubs and network restructuring events, suggesting that evolutionary breakage is majorly caused by non-coding regulatory regions changes and not protein-coding sequence changes. In general, the paper reveals that the regulatory gene networks have a modular structure with a few functional cores and peripheral components, which is rewired during evolution. It gives a systems level insight into evolution of regulatory programs to generate fundamental biological stability and species-specific phenotypic diversity.

KEYWORDS: Evolution, regulatory gene networks, transcription factors, conservation, divergence, comparative genomics, eQTL, enhancer, promoter interactions, ortholog mapping, network biology.

1. INTRODUCTION

Gene regulatory networks (GRNs) are the complicated interaction networks that regulate gene expression in space, time and context. The transcription factors are the main regulators of these networks, binding to regulatory sequences like promoters and enhancers, and thereby regulating genes [11]. It is through this regulated control that GRNs establish cellular identity, developmental programs, physiological reactions, and organism phenotypes [3]. The interaction of regulatory gene networks within different species is a basic question of evolutionary biology. Evolution does not just act on protein-coding sequences but also has a profound impact at the level of non-coding regulatory regions causing gene expression to change without necessarily having an effect on protein structure [2], [4]. Consequently evolutionary conservation of GRNs indicates biological functions that are vital and well-defended, so as divergence indicates adaptations that lead to species-specific traits and environmental reactions [5], [6].

Although there have been notable advances in genomics, a big issue still lies in comprehending how systems regulatory networks emerge. Although some studies have examined individual genes or individual regulatory elements, most studies are restricted to one species or one isolated feature of the genome [1], [9], [10]. This discontinuous perspective renders it hard to embrace the organization of regulatory systems worldwide throughout evolution. Specifically, the patterns of transcription factor binding, enhancerpromoter interactions, expression quantitative trait loci (eQTLs) have all been performed on a case-by-case basis, and seldom in a combined evolutionary framework [7], [8]. Besides, current comparative genomics analyses tend to concentrate on sequence conservation incompletely integrating functional regulatory interactions and network topology [12].

To overcome these shortcomings, this paper suggests a system-wide scheme to study the evolutionary conservation and divergence of regulatory gene networks. The goal is achieved by incorporating comparative genomics, ortholog mapping, regulatory annotations and network reconstruction methods to discover conserved regulatory cores and divergent network modules

2. RELATED WORK

Regulatory gene networks have been studied widely using both computational, experimental and evolutionary approaches. The initial computational methods like clustering algorithms [1] and reverse engineering techniques [9], [10] were initial tools used in deducing a gene network architecture. Bayesian models further developed the subject by providing probabilistic models of gene expression data [7], [8], with powerful approaches to estimating networks and making functional inferences. These computational inventions provided the foundation of systems level studies of gene regulation.

The governing networks have been shown to be biologically relevant in varied situations in the experimentations. As an example, Bruex et al. [3] modeled a gene regulatory network of Arabidopsis root epidermis differentiation, and Sauka-Spengler and Bronner-Fraser [11] recently have shown orchestration of neural crest formation by regulatory modules. Proteomic and regulatory responses to stress have been examined, Brion et al. [4] demonstrating species-specific responses to stress in yeasts, and Ballarar [2] highlighting light-regulated responses to stress in plants. These results highlight the interaction between environmental cues and regulatory network dynamics.

In evolutionary terms, research has looked into the aspects of robustness and adaptability in regulatory systems. Ciliberti et al. [5] have shown that the strength may be built up through time in complex networks, whereas de Vos et al. [6] emphasized the evolutionary optimality by synthetic biology knowledge. The more recent developments make use of machine learning to integrate single-cell data, which allows mapping regulatory interactions in high-resolution [12]. Together, these studies provide a solid base of reasoning on the conservation of functional cores to be counterbalanced by divergent adaptive modules, and are thus directly relevant to the systems-level context of this paper.

3. MATERIALS AND METHODS

3.1 Study Design and Analytical Framework

This work builds a multi-layer framework of computational biology to study evolutionary conservation and divergence of regulatory gene networks. The complete design unites the comparative genomics, functional genomics, and systems biology to assemble a single model of gene regulation in species. The workflow is designed in such a way that it proceeds through the sequence level conservation, then regulatory interaction networks, and ultimately systems level interpretation.

There are four large steps in the analysis pipeline: ortholog mapping, regulatory element conservation analysis, transcription factor binding conservation and network reconstruction. The stages add a different level of biological interpretation, making sure that both the sequence level and functional-level evolutionary signals are represented. This hierarchical design is designed to compare across species e.g. human, mouse, and zebrafish systematically. The framework provides commonality between datasets through genomic coordinates, annotation formats, and regulatory evidence sources standardization. Cross-omics integration can also be used to discover conserved regulatory architectures and lineage-specific rewiring events. This systematic methodology will make cross-species regulatory network analysis scaleable and reproducible.

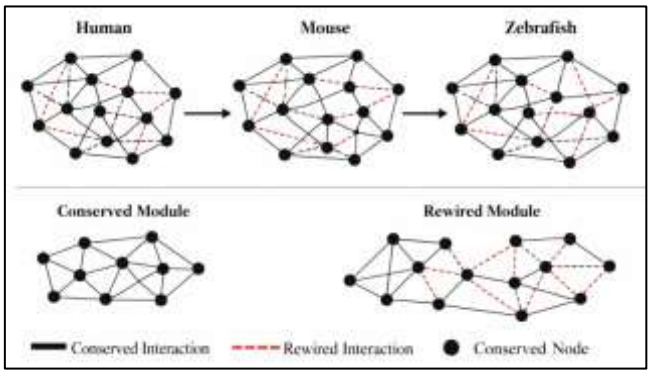

Fig 1: Cross-species Evolutionary Gene Regulatory Network Framework

3.2 Data Sources and Genomic Resources

To be as inclusive as possible in collecting regulatory and genetic information, genomic and functional data were gathered using numerous publicly accessible databases. Genome assemblies and gene annotations were acquired at NCBI and Ensembl, databases that offer standardized genome cross-species references. Ortholog recognition and genomic coordinate mapping were done with the help of these databases.

ENCODE and Roadmap Epigenomics projects were merged to provide regulatory and functional genomic data, and these projects have enhancer, promoter, and chromatin accessibility annotations. Human eQTL data were found in the GTEx consortium and could be used to map tissue-specific regulators. GWAS Catalog data was

added to connect regulatory networks with loci that are related to a trait. In order to supplement the regulatory and interaction-level analysis, protein-protein interaction networks were retrieved through the STRING database. Comparative genomic conservation tracks were obtained at the UCSC Genome Browser. All data sets were harmonized to a single genomic reference system so that they are compatible across various levels of analysis.

Table 1: Data Sources and Their Functional Roles

Data Source	Purpose
NCBI / Ensembl	Genome and ortholog mapping
GTEx	Tissue-specific eQTL data
ENCODE / Roadmap	Regulatory annotations
GWAS Catalog	Trait association mapping
UCSC Genome Browser	Conservation tracks
STRING Database	Protein interaction networks

3.3 Ortholog Mapping and Gene Conservation Analysis

Ortholog mapping was conducted to determine the evolutionary conserved genes in the selected species with human, mouse, and zebrafish being the ones. Ensembl Compara and reciprocal best-hit methods were used to define gene orthology relationships to provide high confidence gene pairings. This step will be used to identify sets of conserved genes that will most probably maintain key biological processes. This was done through the classification of genes into conserved orthologs and lineage-specific genes through sequence similarity and evolutionary distance. Conserved genes are fundamental biological processes, and lineage specific genes with species-specific adaptation. Comparison of downstream regulatory networks is based on this classification.

Ambiguous mappings and paralogous gene conflicts were filtered to enhance reliability. Regulatory network was reconstructed using only high confidence one to one orthologs. This guarantees that comparative analysis is a factual evolutionary conservation as opposed to artifacts of gene duplication.

3.4 Regulatory Element Conservation Analysis.

Conservation of non-coding regulatory DNA was evaluated by analyzing regulators and promoters across species. Genome-wide comparative tracks were used to align sequences of regulatory regions in order to determine conserved regulatory regions. These factors also were examined with regard to functional activity in terms of epigenomic annotations.

Phylogenetic conservation measures like phastCons and phyloP scores were used as conservation scoring, to measure evolutionary constraint on a nucleotide scale. Elements were assigned functional conservation status based on high conservation scores in regulatory regions and low conservation scores in regulatory regions. Regulatory elements were classified, based on these scores, into highly conserved regulatory regions, weakly conserved or lineage-specific regulatory elements. This classification gives an insight into the evolution of gene regulation beyond the protein-coding sequences and is involved in the phenotypic diversity.

3.5 Transcription Factor Binding and Motif Evolution Analysis

The conservation of transcription factor binding based on knownTF binding motifs in genomic sequences was studied using curated motif databases. PWMs were used to determine the possible location of binding sites in various genomes of species. This enabled the comparison of TF binding potential with evolutionary distances.

In order to determine preserved and broken transcription factor binding sites, motif conservation analysis was conducted. Gain and loss of regulatory motifs were monitored to learn about regulatory rewiring events. These modifications denote the way transcriptional regulation is modified without modifying coding sequences.

3.6 Gene Regulatory Network Construction

Interactions between the transcription factors and the genes were used to assemble species-specific gene regulatory networks. Any TF-gene connections were established using evidence of motif binding, enhancer-promoter relationships, and eQTL relationships. This integration of evidences enhances trust in the interactions between the regulators.

Every network unites several layers of regulation, such as direct TF binding, distal enhancer regulation, and expression-based associations. There were assigned edge weights in regard to strength of supporting evidence to be certain that it quantitatively represented the regulatory influence. The species were placed on separate networks so that they could be compared. Such networks are the foundation on which conserved regulatory structures and lineage-independent rewiring events in downstream analysis are found.

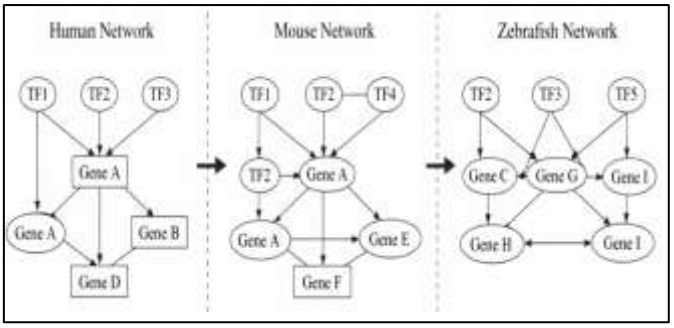

Fig 2: Species-specific gene regulatory network architecture

3.7 Comparative Network Analysis and Module Detection

To determine preserved and divergent interactions between regulators, cross-species network alignment was done. Network comparison aimed at detecting shared hubs, rewired nodes and structural variations between connectivity patterns. This assists in the comprehension of regulatory system evolutionary stability. The Louvain algorithm was used to perform community detection on each regulatory network to identify modular structures in the community. Modules are sets of co-regulated, functionally related genes. Conserved biological programs can be identified by comparative analysis of modules.

Functional differences were studied on the conserved and species-specific modules to comprehend evolutionary adaptation. The conserved modules are generally the core biological processes, and the divergent modules are generally specialized functions.

4. RESULTS AND DISCUSSION

4.1 Ortholog Conservation Across Species

Orthologous gene mapping in human, mouse and zebrafish showed that there was a robust core of conserved genes which remain intact across the evolutionary gap. These genes are basic biological processes without which survival and cell survival is impossible. A great percentage of the genes were found to be one-to-one orthologs, which means that there were high evolutionary pressures.

Gene sequences that showed the highest scores in conservation were essential genes that were involved in various processes that included DNA replication, transcriptional regulation, and protein synthesis. There was low sequence divergence of these genes, and this indicated high purifying selection. This implies that fundamental cellular machinery is very stable with respect to cross-species. On the contrary, the less conserved genes were related primarily either to species-specific functions or environmental adaptation. These results prove that the pressure of evolution massively keeps those genes intact that are part of the basic cellular functions and, conversely, diverges those pathways that are specialized.

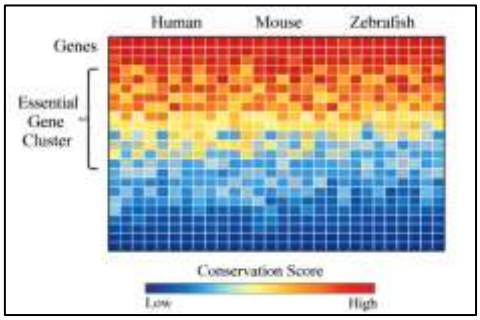

Fig 3: Ortholog conservation heatmap across human–mouse–zebrafish

4.2 Conservation of Regulatory Elements

Regulatory Elements Genome-wide analysis revealed that about 30-50 percent of regulatory regions are shared across species, but that this depends on the evolutionary separations. The conservation was higher in promoter regions than in enhancers which meant that there were stronger evolutionary restrictions on proximal regulation.

Enhancers had much smaller conservation levels, showing their importance in species-specific gene regulation and phenotypic diversity. There was a rapid turnover of many distal regulatory elements, implying flexibility of long-range regulation of genes. These findings underscore an important evolutionary principle that the fundamental mechanisms of gene expression are conserved but regulatory fine-tuning can be more plastic and changeable.

4.3 Transcription Factor Binding Conservation

Transcription factor binding The key regulatory protein families evaluated included CTCF, ETS and GATA and were found to be highly conserved across the species. These transcription factors are key players in chromatin organization, developmental regulation and the stability of gene expression.

Although there was an overall conservation, some lineage-specific changes in binding were noted. These involve both the gain and loss of binding sites, which is an evolutionary rewiring of transcriptional regulation. These changes were witnessed more often in the non-coding regions compared to the coding regions. This implies that it is regulatory network change, and not transcription factor changes, that leads to evolutionary divergence between regulatory networks.

Table 2: Conserved vs Divergent Transcription Factors

Category	Examples	Conservation Level	Function
Conserved TFs	CTCF, ETS, GATA	High	Chromatin organization, transcription control
Moderately conserved TFs	AP-1, SP1	Medium	Signal response regulation
Divergent TFs	Lineage-specific TFs	Low	Species-specific adaptation

4.4 Regulatory Network Architecture

Rebuilding of gene regulatory networks revealed a core set of structure across all species experimented. This central network was composed of well-linked transcription factors and vital genes which were required in basic biological activities.

But there was great divergence in peripheral parts of the network. These areas contained genes that regulate the environment, immune signaling and tissue-specific regulation. This means that evolutionary changes tend to be more within peripheral regulatory layers, and not in the core network architecture. On the whole, the regulatory networks have a core periphery structure with the core being evolutionary stable and the periphery being highly dynamic.

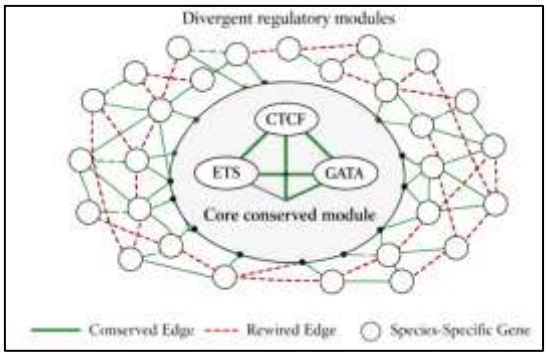

Fig 4: Core–Periphery Evolutionary Regulatory Network with Rewiring

4.5 Module-Level Conservation and Divergence

Community detection analysis revealed specific regulatory modules which corresponded to biological processes. The developmental pathways, cell cycle regulation and chromatin organization were found to be enriched in core conserved modules, showing that there are strong evolutionary conservation of important biological systems.

Conversely, divergent modules were rich in variation of immune responses, and specialization in the neural system and metabolic adaptation pathways. Species-specific regulatory dynamics and increased diversity in gene composition were observed in these modules.Such a modular organization implies that evolution can operate on the level of regulatory programs, as opposed to each gene, and that functional modules can be maintained, and functional rewiring can occur in specialized systems.

4.6 Hub Gene and Key Regulator Analysis.

The analysis of the network topology revealed that there were conserved hub genes with high centrality scores in all the species. Such hubs mostly dealt with transcriptional regulation, chromatin remodeling, and the signaling pathways. They are very connected thus show high centrality in network stability.

Divergent hub genes were also discovered, that were specific to individual species. It is probable that lineage-specific adaptations and functional specialization are due to these genes. These had variably connected connectivity and less cross-species conservation compared to conserved hubs. The findings indicate that conserved hubs are the cause of evolutionary stability, whereas peripheral hub nodes emergence or rewiring causes adaptation.

Table 3: Hub Gene Analysis Across Species

Gene Type	Examples	Centrality Score	Functional Role
Conserved hubs	TP53, MYC, CTCF	High	Core regulatory control
Divergent hubs	Species-specific regulators	Medium–Low	Adaptive traits
Bridge hubs	ETS family genes	High betweenness	Cross-module regulation

4.7 Evolutionary Rewiring of Regulatory Networks

Analysis of regulatory interactions based on evolutionary analysis showed considerable gain and loss of transcription factorgene interactions across species. Such rewiring was also mainly in non-coding control regions notably enhancers and distal control elements. Most rewiring events did not lead to a change in the content of genes; most of them were a change in the regulatory connections. This implies that reconfiguration of networks than gene innovation is the cause of evolutionary adaptation. In general, the findings indicate that networks of regulatory genes can be modified by both conservative core stability and rewiring of external regulatory associations.

4.8 Discussion

A basic core-periphery evolutionary architecture is pointed out by the comparative analysis of the regulatory gene networks in species. The central backbone of regulatory stability is comprised of conserved transcription factors including CTCF, ETS and GATA that is critical to the maintenance of fundamental biological functions such as transcriptional regulation and chromatin structure. These hubs serve to anchor evolutionary adaptation within the network among species. Peripheral modules, on the other hand, rewire substantially in response to alterations in enhancer activity, transcription factor binding and non-coding regulatory elements. This evolutionary give and take gives organisms the capacity to adjust to environmental stress and produce their own lineage- specific characteristics without undermining the integrity of the core.

Notably, the results underline the idea that evolutionary innovation comes as a result of regulatory reconfiguration instead of a result of gene invention. Transcription factor-gene interactions, particularly in distal regulatory regions, are gained and loss, re-shaping expression patterns and functional modules, which add to phenotypic diversity. This hierarchical structure, based on modules, shows the balance of evolution between robustness and adaptability: changes allowed in peripheral networks, but necessary regulatory cores needed to be conserved. A systems-level view gives more understanding of the role of genomic architecture in contributing both biological stability and evolutionary flexibility to the complexity of species.

5. CONCLUSION

This paper gives a holistic view of the evolution of regulatory gene networks at a systems level in the form of a precarious equilibrium between conservation and divergence. The comparative genomics with transcription factor binding analysis, and network construction involving human, mouse and zebrafish indicates a modular structure composed of a functional core that is conserved and a peripheral region rich in adaptability.

The central modules, which are rich in key biological mechanisms, including transcriptional regulation, chromatin structuring, and cell cycle regulation, are evolutionarily conserved, indicating high selection pressure in maintaining core cellular processes. Conversely, peripheral modules reveal a wide-scale rewiring of networks, which is mostly prompted by modifications of non-coding regulatory factors and transcription factor binding specificities. These rewiring processes make species-specific adaptations and phenotypic diversity possible without interfering with essential biological stability. On balance, the results have shown that evolutionary innovation in gene regulation, is not due to new gene but rather as a result of restructuring of old gene regulatory links. The evolutionary logic of regulatory networks is characterized by this dual mechanism, core conservation, and peripheral divergence. The system defined here provides a generalizable system to examine the regulatory evolution between species and offer important perspectives on the way that genomic architecture promotes both biological and adaptive plasticity.

REFERENCES

[1] A. Ben-Dor, A., & Yakhini, Z. (1999, April). Clustering gene expression patterns. In Proceedings of the third annual international conference on Computational molecular biology (pp. 33–42).
[2] Ballaré, C. L. (2014). Light regulation of plant defense. Annual review of plant biology, 65, 335–363.
[3] Bruex, A., Kainkaryam, R. M., Wieckowski, Y., Kang, Y. H., Bernhardt, C., Xia, Y., ... & Schiefelbein, J. (2012). A gene regulatory network for root epidermis cell differentiation in Arabidopsis. PLoS genetics, 8(1), e1002446.
[4] Brion, C., Pflieger, D., Souali-Crespo, S., Friedrich, A., & Schacherer, J. (2016). Differences in environmental stress response among yeasts is consistent with species-specific lifestyles. Molecular Biology of the Cell, 27(10), 1694–1705.
[5] Ciliberti, S., Martin, O. C., & Wagner, A. (2007). Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS computational biology, 3(2), e15.
[6] de Vos, M. G., Poelwijk, F. J., & Tans, S. J. (2013). Optimality in evolution: new insights from synthetic biology. Current opinion in biotechnology, 24(4), 797–802.
[7] Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000, April). Using Bayesian networks to analyze expression data. In Proceedings of the fourth annual international conference on Computational molecular biology (pp. 127–135).
[8] Imoto, S., Goto, T., & Miyano, S. (2001). Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. In Biocomputing 2002 (pp. 175–186).
[9] Liang, S., Fuhrman, S., & Somogyi, R. (1998, January). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In Biocomputing (Vol. 3).
[10] Liang, S., Fuhrman, S., & Somogyi, R. (1998, January). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In Biocomputing (Vol. 3).
[11] Sauka-Spengler, T., & Bronner-Fraser, M. (2008). A gene regulatory network orchestrates neural crest formation. Nature reviews Molecular cell biology, 9(7), 557–568.
[12] Uhler, C., & Shivashankar, G. V. (2022). Machine learning approaches to single-cell data integration and translation. Proceedings of the IEEE, 110(5), 557–576.