Paradigm Academic Press
Innovation in Science and Technology
ISSN 2788-7030
NOV. 2025 VOL.4, NO.10

Theoretical Modeling of Gene Regulatory Networks in Non-Model Organisms

Miguel Duarte[1]

[1] University of Coimbra, Portugal

Correspondence: Miguel Duarte, University of Coimbra, Portugal.

doi:10.63593/IST.2788-7030.2025.11.006

Abstract

Theoretical modeling of gene regulatory networks (GRNs) in non-model organisms represents a vital frontier in systems biology, where data scarcity and incomplete genomic annotation challenge conventional empirical methods. This study presents a comprehensive theoretical exploration of how mathematical and computational frameworks can be applied to reconstruct, analyze, and interpret GRNs in species beyond the classical model systems. By integrating network theory, information-theoretic inference, and dynamical systems modeling, this paper articulates a conceptual foundation for understanding gene regulation as a process of information exchange and systemic organization. It argues that the abstraction of biological networks into formal models enables the discovery of underlying principles of control, robustness, and adaptability that are conserved across evolution, even in the absence of experimental validation. The paper develops a multilayer theoretical synthesis encompassing topological network structures, dynamic feedback regulation, and probabilistic inference strategies. Through examples from fungal, plant, and microbial systems, it demonstrates how systems-level integration and computational innovation can uncover hidden regulatory logic in underexplored taxa. The study concludes that theoretical modeling is not a substitute for empirical biology but a necessary complement that extends the reach of biological reasoning into domains where direct experimentation remains impractical. The work thus positions theoretical GRN modeling as both a methodological framework and a philosophical approach to understanding life's organizational complexity across the full spectrum of biodiversity.

Keywords: gene regulatory networks, theoretical modeling, non-model organisms, systems biology, network inference, computational biology

1. Introduction

Understanding how genes regulate one another forms the foundation of systems biology and quantitative genetics. The complexity of gene interactions defines the architecture of life, shaping how cells interpret signals, adapt to environmental conditions, and maintain homeostasis. A gene regulatory network (GRN) represents the intricate web of molecular interactions in which transcription factors, noncoding RNAs, and epigenetic elements coordinate gene expression. In classical molecular biology, knowledge of these networks has largely arisen from empirical studies in well-characterized model organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Drosophila melanogaster*. These systems provide controlled experimental frameworks, abundant genomic data, and established molecular tools that enable direct observation of regulatory dynamics. However, this empirical bias toward a few species has limited the theoretical generalization of gene regulatory mechanisms across the vast diversity of life.

Non-model organisms comprise the overwhelming majority of biodiversity and include species from extreme environments, emerging pathogens, and ecologically vital taxa without established genetic systems. The scarcity of functional genomic resources for such organisms constrains direct experimental inference of their regulatory networks. In these contexts, theoretical modeling becomes not only a methodological alternative but an epistemological necessity. Theoretical models enable researchers to formalize hypotheses about gene regulation in species where laboratory validation is impractical or impossible. These models rely on mathematical abstractions and computational simulations that translate limited genomic and transcriptomic information into structured hypotheses about regulatory relationships. They extend the reach of biology from data-rich organisms to those accessible only through computational inference.

Theoretical modeling of GRNs in non-model organisms operates at the intersection of systems theory, network science, and evolutionary biology. It seeks to reconstruct gene interactions from patterns in data using probabilistic, dynamical, or topological formalisms. Approaches such as information theory, Bayesian inference, and differential equation systems allow for prediction of regulatory dependencies even in the absence of complete annotation. These frameworks embody a shift from descriptive biology toward predictive modeling, where the goal is not only to catalog molecular components but to understand how their coordinated behavior gives rise to biological function. The emphasis moves from empirical replication to theoretical coherence, privileging consistency and explanatory power over direct measurement.

In the broader view, theoretical models of GRNs also act as conceptual bridges that connect individual species to universal biological principles. They provide tools for comparing the architecture and dynamics of regulatory systems across evolutionary lineages, revealing how gene networks evolve, reorganize, and adapt. The modeling of non-model organisms thus contributes not only to specific genomic understanding but to the development of a more unified theoretical biology. It illustrates how mathematical formalization can extend scientific inquiry beyond the boundaries of direct experimentation, transforming the study of gene regulation into a discipline that integrates computation, abstraction, and evolution into a coherent framework.

2. Conceptual Foundations

A gene regulatory network represents a conceptual framework that describes the collective behavior of genes within a cell as a system of interdependent components. Each gene is both an actor and a target within a web of molecular communication that governs transcriptional activity. The central premise of modeling such a network is that gene expression patterns are not random but arise from a structured interplay among regulatory elements, transcription factors, and signaling cascades. Theoretical modeling attempts to capture this structure through abstractions that translate biological relationships into formal mathematical representations. These representations permit reasoning about how local interactions give rise to global behavior, how stability and variability emerge from feedback loops, and how the system adapts to environmental or genetic perturbations.

The first conceptual step in theoretical modeling is defining the nature of interactions. In biological systems, regulatory relationships between genes can be activating, repressing, or conditionally dependent on specific molecular states. This diversity of interactions challenges any attempt to describe the system with a single mathematical framework. Continuous models based on differential equations describe regulatory influence as quantitative changes in gene expression over time. These models emphasize dynamics and enable simulation of temporal processes such as oscillations, feedback stabilization, or switch-like transitions. In contrast, discrete models such as Boolean networks abstract regulation into binary or categorical states, focusing on qualitative transitions that capture the logic of gene activation patterns. Bayesian networks introduce probabilistic reasoning, where each gene's expression is treated as a random variable influenced by a set of regulators, allowing uncertainty to be explicitly modeled.

Each of these frameworks embodies a philosophical stance about what constitutes a sufficient description of biological reality. Deterministic formulations assume that the governing laws of regulation can be precisely specified by parameters and equations, while stochastic or probabilistic models acknowledge the inherent variability of molecular interactions. This theoretical plurality reflects a central tension in biology: whether living systems can be represented as mechanistic machines or as statistical ensembles governed by emergent properties. The study of GRNs through theoretical modeling occupies a middle ground, where both precision and uncertainty coexist within the same formal system.

In non-model organisms, the challenge of incomplete data elevates the importance of inference over direct observation. Theoretical modeling becomes an act of reconstruction, where the structure of a network must be deduced from partial information such as gene expression matrices, comparative genomics, or motif predictions. The process of network inference draws on principles from information theory, graph theory, and machine learning. Mutual information, for example, quantifies how much the state of one gene reduces uncertainty about another, offering a nonparametric measure of association that does not rely on linear assumptions. This measure has been widely adopted in reconstructing regulatory relationships from transcriptomic data, especially in organisms lacking experimental perturbation data. Bayesian network inference extends this logic by constructing directed graphs that capture conditional dependencies, allowing the identification of potential causal regulators under constraints of limited data.

The conceptual foundation of these methods lies in the recognition that gene regulation is fundamentally a process of information transfer. A regulatory interaction can be viewed as a channel through which information about one molecular state influences another. From this perspective, the cell becomes an information-processing system, and its gene network functions as a distributed computational structure. Theoretical models thus bridge molecular biology with principles of communication theory and dynamical systems. This perspective encourages the study of network motifs, recurrent substructures such as feedback loops or feedforward circuits, which act as basic computational units performing filtering, amplification, or memory functions within the network.

The structure of GRNs can also be interpreted through graph-theoretic concepts such as connectivity, centrality, and modularity. Highly connected nodes, often corresponding to transcription factors, act as regulatory hubs that integrate signals and coordinate responses. Modules of densely connected genes tend to represent co-regulated pathways or functional units, suggesting a hierarchical organization where local clusters operate semi-independently within a larger network architecture. Theoretical modeling of such organization provides insight into robustness and evolvability, explaining how networks maintain stability despite mutations or environmental variation. The concept of redundancy within these networks illustrates how multiple regulators can compensate for one another, ensuring that critical cellular processes persist even under perturbation.

Incorporating dynamics into GRN models introduces another layer of conceptual complexity. Temporal models account for delays in transcription and translation, feedback loops that stabilize expression, and stochastic fluctuations arising from molecular noise. Continuous deterministic models describe how concentrations of mRNA or proteins evolve over time through coupled differential equations, revealing steady states or limit cycles that correspond to biological phenotypes. Stochastic models such as the Gillespie algorithm treat these processes as probabilistic events, providing insight into phenomena like bistability or stochastic gene switching. These theoretical constructs reveal that biological regulation is not only structured but also dynamic, with the capacity to encode memory, oscillate rhythmically, or switch between distinct functional states.

For non-model organisms, integrating these dynamic principles requires creative adaptation. When empirical data are scarce, models must rely on generic assumptions about regulatory mechanisms derived from evolutionary conservation or physical constraints. Comparative network modeling provides a path forward, where homologous genes across species are mapped onto inferred networks, and evolutionary algorithms are used to optimize parameters consistent with known biological principles. Such theoretical transfer leverages the shared architecture of gene regulation across life forms while allowing the exploration of unique adaptations in under-studied species. The emphasis shifts from reproducing known behaviors to predicting plausible mechanisms consistent with evolutionary logic.

The conceptual foundation of theoretical GRN modeling also extends into questions of inference validation and epistemology. Unlike empirical studies where hypotheses are tested through experiment, theoretical models must be evaluated through internal coherence, predictive accuracy, and correspondence with indirect evidence. In non-model organisms, success may be measured by the ability of a model to generate testable predictions that guide limited empirical investigation. This approach transforms modeling into a heuristic process, where theoretical constructs inform experimental design and discovery rather than serving as post hoc explanations. In this sense, theoretical models act as scaffolds for knowledge construction, shaping the trajectory of inquiry in data-limited contexts.

Conceptualizing gene regulatory networks as mathematical systems offers more than a methodological toolkit. It proposes a vision of biology in which the essence of life is seen through the lens of relationships, dependencies, and dynamical organization. In non-model organisms, this theoretical lens uncovers the universal patterns that underlie biological complexity, revealing how diverse forms of life achieve coordinated control through variations on common regulatory principles. The study of GRNs thus becomes a philosophical exploration of how information, structure, and function intertwine within living systems. Theoretical modeling transforms the problem of missing data into an opportunity to seek general laws of organization that transcend the boundaries of individual species. Through abstraction, it provides not only predictive frameworks but also conceptual clarity about what it means for a genome to regulate itself, adapt, and persist as an integrated system.

3. Integrating Systems-Level Approaches

The integration of systems-level approaches into the theoretical modeling of gene regulatory networks transforms isolated molecular observations into a coherent representation of cellular organization. At this level of abstraction, the cell is conceptualized not as a collection of individual genes but as a dynamic network of interactions that collectively determine biological function. Theoretical modeling seeks to describe this network in mathematical terms that can capture both structure and behavior. For non-model organisms, this systems-level perspective is essential because it allows researchers to extrapolate from limited or fragmented datasets toward comprehensive frameworks of regulation. The emphasis shifts from studying individual genes to understanding the architecture of the regulatory system as a whole.

Systems-level modeling begins by conceptualizing biological processes as interconnected networks that link genetic, transcriptional, and metabolic layers. The idea of the interactome, which includes all known physical and functional associations between biomolecules, provides the foundation for constructing theoretical models that reflect how genes influence one another indirectly through shared pathways or signaling intermediates. In the absence of complete experimental maps, graph theory provides a mathematical language for representing such systems. Each gene or protein is represented as a node, and each regulatory relationship as an edge connecting them. The topology of this network—its connectivity, degree distribution, and clustering—encodes the structural logic of cellular organization. Modeling these topological features enables predictions about control points, stability, and modularity even when empirical measurements are incomplete.

At the systems level, topology is not merely descriptive but predictive. Highly connected nodes, often called hubs, are associated with essential genes or master regulators that exert broad influence across the network. Their removal can destabilize network behavior, a property known as fragility, which has implications for understanding genetic robustness and evolutionary adaptation. In contrast, modularity describes the presence of semi-independent clusters of genes that co-regulate specific biological processes. Theoretical models that account for modular structure can identify functional units within non-model organisms and predict how local changes in one module may influence others through shared regulators. This hierarchical organization of modules within networks parallels the organization of biological functions, where complex traits emerge from the coordination of smaller subsystems.

Dynamic systems theory extends the topological model by incorporating time-dependent processes and feedback mechanisms. Theoretical frameworks using coupled differential equations, Boolean networks, or stochastic simulations allow for the study of how regulatory signals propagate through the network over time. Feedback loops can stabilize gene expression, create oscillations, or generate bistable switches that underlie cell differentiation and response plasticity. The modeling of such dynamics at the systems level reveals how cells achieve balance between stability and adaptability. In non-model organisms, where temporal data are often limited, theoretical models approximate dynamics through inferred causal relationships, leveraging statistical dependencies and conserved motifs to simulate likely behaviors under different environmental or developmental contexts.

Integrating systems-level approaches also requires incorporating information from multiple omic layers. Transcriptomic data capture gene expression patterns, proteomic data describe interactions among proteins, and metabolomic data reveal downstream consequences of regulation. Theoretical frameworks that integrate these heterogeneous datasets aim to construct multilayered regulatory networks that reflect the flow of information from genome to phenotype. Such integration is particularly valuable for non-model species, where direct experimental validation across all levels of organization is infeasible. By linking data across scales, models can infer how genetic variation leads to functional outcomes, offering insights into the systems biology of adaptation, resilience, and ecological interaction.

Hybrid approaches combining data-driven inference with prior biological knowledge have emerged as a powerful strategy for modeling non-model systems. In this paradigm, theoretical models incorporate constraints derived from evolutionary conservation, literature-based interactions, or known biochemical mechanisms. These constraints act as guiding principles that reduce the space of possible network configurations, allowing robust predictions even from limited data. Knowledge-based modeling provides a scaffold upon which computational inference operates, ensuring that the resulting networks maintain biological plausibility. Linde et al. demonstrated that integrating curated knowledge with transcriptomic data enhances model stability and predictive accuracy, enabling genome-scale modeling even in species with sparse annotations. Such hybrid systems-level approaches exemplify how theoretical abstraction can compensate for empirical gaps without compromising the integrity of biological interpretation.

In systems-level theoretical modeling, the concept of emergent behavior plays a central role. Emergence refers to the collective properties that arise from interactions among components but cannot be attributed to any single element. At the scale of gene regulatory networks, emergent phenomena include homeostasis, differentiation, and adaptive response. Theoretical models aim to identify the conditions under which such properties appear, using simulation and analysis to reveal how global patterns arise from local rules. These insights have profound implications for non-model organisms, where evolutionary pressures may have sculpted unique regulatory architectures that produce similar emergent behaviors through alternative network configurations. Modeling such systems provides a theoretical basis for understanding evolutionary convergence and diversity in gene regulation.

Another crucial element of systems-level integration is the application of statistical mechanics and control theory to biological regulation. The cell can be viewed as a system seeking equilibrium under continuous perturbation, where regulatory feedback maintains functional stability. Theoretical models inspired by control theory describe how feedback circuits detect deviations and implement corrective responses. These frameworks explain phenomena such as metabolic homeostasis and stress tolerance in terms of network stability criteria. In non-model organisms, applying such theoretical tools allows predictions about robustness and resilience, qualities that are often key to survival in fluctuating environments. This approach also informs synthetic biology, where insights from theoretical systems modeling guide the design of artificial regulatory networks with desired behaviors.

Network inference at the systems level also raises important computational challenges. The number of possible network configurations grows exponentially with the number of genes, creating a vast search space that cannot be exhaustively explored. To address this complexity, theoretical models employ heuristic and optimization algorithms that balance computational feasibility with biological realism. Techniques such as sparse regression, information-theoretic selection, and Bayesian sampling reduce dimensionality while preserving essential network features. These algorithms allow researchers to infer large-scale GRNs in non-model organisms without the need for complete data, making it possible to approximate the regulatory landscape from limited expression or sequence information.

The integration of systems-level approaches also invites a reconsideration of what constitutes biological explanation. In classical molecular biology, causation is often described in linear terms: one gene regulates another through a direct molecular mechanism. Systems-level modeling replaces this reductionist view with a relational one, where causation emerges from network context rather than isolated interactions. A theoretical model that accurately predicts system behavior can be considered explanatory even if it does not specify every molecular detail. This shift from mechanism to pattern aligns with a broader philosophical transition in biology, from studying isolated causes to understanding systemic organization. For non-model organisms, such a framework is indispensable, since detailed mechanisms are often inaccessible, but global regulatory principles can still be inferred and analyzed.

The value of integrating systems-level approaches in theoretical GRN modeling extends beyond scientific understanding to practical applications. Predictive models of regulatory systems can guide conservation efforts by identifying genes associated with environmental adaptability or resilience. In agriculture, theoretical models of plant regulatory networks can predict traits related to stress tolerance or yield optimization in species that have not been domesticated. In microbiology and ecology, they can help elucidate how microbial communities regulate collective functions, such as nutrient cycling or symbiosis, through distributed regulatory coordination. Each of these applications relies on the same theoretical foundation: the capacity to infer, simulate, and interpret regulation as a system-level phenomenon.

At its core, systems-level integration represents a shift in how biology conceives complexity. Theoretical modeling transforms genetic data into an analytical language capable of expressing organization, adaptation, and resilience. For non-model organisms, this transformation is not merely technical but epistemological. It redefines what it means to know a biological system when direct observation is limited. Through abstraction and synthesis, theoretical systems biology reveals that understanding life requires seeing it as an interdependent whole, where patterns of regulation, not isolated parts, define the essence of function. The integration of systems-level approaches into GRN modeling thus marks a convergence of mathematical reasoning and biological insight, offering a framework through which the hidden logic of non-model organisms can be uncovered and understood.

4. Theoretical Innovations and Computational Inference

Theoretical innovation in modeling gene regulatory networks arises from the need to transform sparse or incomplete biological data into coherent representations of regulatory architecture. Computational inference serves as the bridge between theory and data, enabling the extraction of structure, dynamics, and function from high-dimensional genomic information. The modeling of gene regulatory networks in non-model organisms depends critically on such innovation because direct empirical evidence is limited. The challenge lies in constructing theoretical systems capable of inferring relationships that are biologically plausible, computationally tractable, and generalizable across species. Theoretical advances in mathematics, computer science, and statistical learning have expanded the tools available for this task, allowing complex biological systems to be represented as inferable networks rather than as opaque molecular collections.

The foundation of computational inference lies in the recognition that gene expression data, when properly analyzed, contain traces of the underlying regulatory structure. Each transcript reflects a combination of regulatory influences, environmental inputs, and stochastic fluctuations. Theoretical models treat the relationships among these variables as structured dependencies that can be reconstructed mathematically. Information theory introduced the concept of mutual information as a measure of association between genes, capturing both linear and nonlinear dependencies. This principle was adapted into algorithms such as ARACNe and CLR, which infer regulatory edges by quantifying information transfer between gene pairs. Such models became foundational for non-model organisms because they require only expression data and make minimal assumptions about specific molecular mechanisms. Their theoretical grounding in information theory allows them to function across diverse biological contexts where mechanistic data are unavailable.

Bayesian inference represents another major theoretical innovation in the modeling of gene regulatory networks. In Bayesian frameworks, each gene is represented as a random variable whose state depends probabilistically on the states of its regulators. The network structure corresponds to a directed acyclic graph, and the goal of inference is to identify the configuration of edges that best explains the observed data. The Bayesian approach allows the explicit incorporation of prior knowledge, such as sequence motifs or conserved pathways, and provides a principled mechanism for managing uncertainty. In non-model organisms, where empirical validation is difficult, this probabilistic reasoning becomes particularly powerful because it formalizes uncertainty rather than ignoring it. Bayesian networks can be extended to dynamic Bayesian models, which capture temporal dependencies and feedback regulation, allowing the inference of time-ordered causal relationships from sequential data. These theoretical frameworks provide a means of simulating how regulatory states evolve and interact across developmental or environmental gradients.

Another theoretical frontier is the integration of machine learning into GRN inference. Classical approaches relied on pairwise correlations or small-scale statistical models, but advances in computational power have enabled the application of deep learning, ensemble methods, and kernel-based approaches. Neural network architectures such as autoencoders and graph neural networks can model complex nonlinear dependencies without requiring explicit specification of functional forms. These models learn latent representations that capture regulatory modules and hierarchical organization within gene expression data. In non-model organisms, deep learning enables transfer learning, where models trained on data from well-characterized species are adapted to infer networks in related but under-studied taxa. This transfer of learned representations embodies a new kind of theoretical inference that leverages evolutionary similarity to overcome data scarcity.

Theoretical innovation also extends to integrating physical and biochemical constraints into computational inference. The concept of constraint-based modeling, originally developed for metabolic networks, has been adapted to gene regulation. In these models, the space of possible network configurations is restricted by known biological laws, such as conservation of mass, transcriptional kinetics, or thermodynamic feasibility. By embedding these constraints into the inference process, theoretical models avoid biologically implausible predictions and achieve greater interpretability. In non-model systems, these constraints often derive from comparative genomics or universal properties of regulatory motifs, allowing inference grounded in general biological principles rather than species-specific data.

An emerging theoretical direction involves the use of network sparsity and regularization techniques. Gene regulatory networks are typically sparse, meaning that each gene is influenced by a limited number of regulators. Sparse modeling techniques such as LASSO regression and elastic net regularization exploit this property to reduce overfitting and enhance interpretability. These methods identify a minimal set of predictors that best explain the expression of each gene, thereby inferring the most likely regulatory connections. The theoretical basis of sparsity aligns with biological efficiency, reflecting the economy of regulation observed in real systems. For non-model organisms, sparse inference provides a practical advantage because it limits the complexity of the network even when data dimensionality exceeds sample size, a common situation in ecological or field-derived datasets.

Theoretical developments in dynamical systems have also transformed how computational inference approaches gene regulation. By modeling the temporal evolution of gene expression as a system of coupled nonlinear equations, researchers can explore attractor landscapes that correspond to stable cellular states. These attractors represent theoretical predictions of differentiation, adaptation, or homeostasis. Dynamical models capture how perturbations propagate through the network, offering insight into resilience and critical transitions. In non-model organisms, such models can be applied to predict adaptive responses under changing environmental conditions, even when only static or partial data are available. Computational inference identifies the parameters that reproduce observed expression patterns, while theoretical analysis interprets these parameters in terms of biological function and stability.

A critical aspect of theoretical innovation is the reconciliation of inference and validation. In the absence of extensive experimental verification, theoretical frameworks rely on internal consistency, cross-validation with orthogonal data types, and simulation-based testing. Synthetic networks serve as benchmarks for evaluating algorithmic performance, ensuring that inferred networks reflect plausible biological behavior. For non-model organisms, where ground truth is limited, theoretical validation may involve testing whether inferred networks reproduce known evolutionary or ecological patterns. This conceptual approach treats predictive accuracy and explanatory coherence as dual criteria of model reliability. Theoretical inference becomes an iterative process, where models are continually refined through comparison with indirect evidence and general biological principles.

Computational inference also benefits from the incorporation of evolutionary theory. Gene regulatory networks evolve under constraints of robustness and adaptability, and their architectures bear traces of these evolutionary pressures. Comparative modeling uses homologous genes and conserved network motifs to infer ancestral regulatory structures and predict unobserved connections in related species. Evolutionary algorithms mimic natural selection to optimize network configurations according to fitness criteria derived from data or theoretical expectations. These methods embody the principle that biological networks are products of adaptive optimization rather than random assembly. In non-model organisms, integrating evolutionary reasoning allows inference guided by phylogenetic relationships and functional conservation, linking theoretical modeling with the broader narrative of biological diversity.

The rise of high-throughput sequencing has accelerated the development of scalable computational inference. Algorithms that once handled hundreds of genes now process entire genomes, integrating transcriptomic, epigenetic, and chromatin accessibility data. Theoretical innovations in parallel computing, probabilistic graphical modeling, and matrix factorization have made it possible to infer networks involving thousands of nodes with reasonable computational cost. Theoretical progress has thus paralleled technological advances, turning data abundance in some domains into transferable insights for those where data remain scarce. The interplay between theoretical generalization and computational efficiency has become central to systems biology, defining the capacity to extend inference across diverse organisms.

The ultimate aim of theoretical innovation in GRN inference is to move from descriptive reconstruction to predictive understanding. A successful theoretical model does not merely replicate observed data but anticipates unseen behaviors, predicts the outcome of perturbations, and identifies potential regulatory principles. Computational inference transforms biology into a predictive science, where theory informs experimentation and discovery proceeds through cycles of simulation and validation. For non-model organisms, this predictive capacity has transformative implications. It enables hypothesis generation about adaptation, stress response, or developmental regulation in species that cannot be studied experimentally. It reveals the underlying logic of life in contexts where observation is limited to fragments of data, converting those fragments into coherent theoretical wholes.

The integration of theoretical innovation and computational inference redefines the study of gene regulation as a dialogue between mathematics and biology. Theories of information flow, dynamical stability, and network topology become tools for interpreting the complexity of living systems. Each model becomes an experiment in abstraction, an attempt to uncover general laws governing the architecture of life. In the realm of non-model organisms, this approach provides the only feasible path toward understanding systems that resist direct experimentation. Theoretical innovation transforms computational inference into an instrument of exploration, capable of revealing the hidden regularities that organize life across scales and species. In this synthesis of mathematics, computation, and biology lies the foundation of a new theoretical understanding of gene regulation that transcends the limitations of data and reaches toward universal principles of biological organization.

5. Philosophical and Practical Challenges

Theoretical modeling of gene regulatory networks in non-model organisms raises questions that reach beyond computational technique or biological description. At its core, this endeavor engages with the philosophy of scientific representation, the nature of explanation, and the limits of knowledge in biological inquiry. It invites reflection on what it means to model life when direct empirical observation is constrained, and how theory itself functions as a form of evidence. The modeling of systems that cannot be fully observed challenges traditional hierarchies between empirical science and theoretical abstraction. It requires acknowledgment that understanding in biology often emerges not from direct measurement but from conceptual coherence and the ability to generate meaningful predictions about unobservable structures.

The distinction between model organisms and theoretical models is epistemologically significant. Model organisms, such as *Drosophila melanogaster* or *Arabidopsis thaliana*, serve as experimental proxies, allowing scientists to manipulate and observe biological processes under controlled conditions. Their epistemic authority derives from empirical reproducibility. Theoretical models, in contrast, operate within a different epistemic regime. They do not reproduce phenomena but simulate relationships, encoding hypotheses about causation and interaction into formal systems. The theoretical model stands not as a representation of a specific organism but as an abstract framework that captures essential structural or dynamic properties common across species. This distinction underlines a philosophical tension in biology: whether understanding arises from the observation of particular instances or from the construction of generalizable conceptual systems.

In the context of non-model organisms, this tension becomes especially visible. The absence of experimental tractability forces theorists to rely on inference and abstraction. Knowledge about these organisms often comes in fragmented forms, such as partial genomes, limited expression datasets, or comparative sequence alignments. Theoretical modeling transforms these fragments into structured hypotheses about regulation and function. From a philosophical perspective, this act of construction raises the question of whether a model that cannot be directly verified still constitutes scientific knowledge. One argument suggests that scientific understanding does not depend solely on verification but on the model's capacity to organize observations, reveal hidden relationships, and generate testable expectations. Theoretical models of gene regulatory networks achieve this by articulating principles of organization that can later guide empirical research when new data become available.

A deeper philosophical challenge lies in the issue of representation. A theoretical model simplifies a complex biological system to make it intelligible. Simplification involves selective abstraction: the omission of detail to highlight structure. This process risks distorting reality if essential features are excluded or misrepresented. In non-model organisms, where the underlying system is poorly understood, the risk of misrepresentation is heightened. Yet the very act of abstraction can also be viewed as a creative scientific act. Theoretical biologists construct idealized systems that express the logic of biological regulation in formal terms. These idealizations are not literal depictions of molecular mechanisms but conceptual instruments that make reasoning about complex phenomena possible. In this sense, theoretical models do not compete with empirical descriptions; they complement them by providing frameworks through which complexity can be navigated.

The practical challenge associated with such modeling lies in data limitation and quality. Non-model organisms often lack comprehensive genomic annotation, curated databases, or high-resolution temporal datasets. These deficiencies constrain the ability to parameterize or validate models. Theoretical approaches must therefore rely on general biological principles such as conservation of motifs, energy optimization, or statistical regularities in gene expression. This reliance introduces uncertainty into inference. The practical response to this uncertainty has been the development of probabilistic and ensemble-based models, where predictions are expressed as distributions rather than fixed outcomes. Philosophically, this shift toward probabilistic representation reflects an acknowledgment that biological knowledge is inherently uncertain and context-dependent. Theoretical modeling thus becomes a discipline of reasoning under uncertainty rather than a pursuit of deterministic truth.

Another challenge arises from the question of causation. In experimental biology, causation is established through intervention and manipulation. In theoretical modeling, causation must be inferred from structure and correlation. The distinction between correlation and causation becomes blurred in this context. Theoretical models infer causal relationships by assuming that regulatory influence manifests as statistical dependence or directed information flow. These assumptions rely on formal analogies between mathematical and biological causality. Philosophers of science have debated whether such analogies can legitimately replace experimental proof. Some argue that theoretical causation serves as a provisional construct, valid insofar as it yields accurate predictions and coherent explanations. Others maintain that without empirical grounding, causal claims remain hypothetical. For non-model organisms, where experimental validation may be permanently unattainable, theoretical causation occupies a pragmatic role. It becomes a working hypothesis that structures future inquiry rather than a definitive statement about biological mechanism.

The ethical and epistemic implications of such modeling also deserve attention. When theoretical models are used to predict gene function or network behavior in species that cannot be experimentally manipulated, the results may influence conservation decisions, agricultural practices, or ecological management. The potential consequences of error become significant. This creates a philosophical responsibility to reflect on the limits of inference and the conditions under which theoretical predictions should inform practice. The use of computational models as surrogates for direct experimentation highlights the broader issue of how science navigates between theoretical elegance and empirical accountability. Responsible modeling in this context requires transparency about assumptions, explicit communication of uncertainty, and iterative refinement as new data emerge.

From a broader epistemological standpoint, theoretical modeling in non-model organisms challenges traditional demarcations between theory and observation. It reveals that theory is not merely a product of empirical data but an active process of shaping how data are interpreted. Incomplete datasets do not merely limit understanding; they stimulate the creation of models that reimagine what is knowable. This creative aspect aligns theoretical biology with fields such as cosmology or theoretical physics, where empirical access is limited and models serve as both hypotheses and explanatory frameworks. In such disciplines, validation is achieved not by direct observation but by coherence, predictive success, and integration within a larger theoretical system. The same criteria apply to GRN modeling in non-model organisms, suggesting that biological theory can achieve rigor even without experimental replication.

The practical side of modeling also involves technological and methodological constraints. Building and analyzing large-scale gene regulatory networks requires computational resources and algorithmic sophistication that may exceed current capabilities. The choice of algorithms, priors, and parameterization strategies can influence outcomes, raising questions about the objectivity of model inference. Theoretical innovation must therefore be accompanied by methodological reflexivity, acknowledging that models are products of both biological insight and computational design. In non-model systems, where uncertainty is amplified, this reflexivity becomes part of the scientific method itself. The act of modeling becomes self-referential, with each iteration serving as both an exploration of biology and an examination of the assumptions embedded in computational reasoning.

Philosophically, the endeavor also invites reconsideration of what counts as explanation in biology. A model may explain by showing how patterns arise from rules, by unifying disparate phenomena under a single framework, or by providing an intelligible narrative of process. Theoretical models of GRNs achieve explanation through structure and dynamics rather than molecular detail. They show how regulatory coherence emerges from the interplay of feedback loops and control hierarchies. In doing so, they shift explanation from the mechanistic level to the systemic. This systemic explanation aligns with the idea that biological understanding is not only about identifying causes but about grasping organization. Theoretical modeling of non-model organisms thus participates in a philosophical transformation of biology from an empirical to a structural science.

The interplay between philosophical reflection and practical application defines the frontier of theoretical modeling. The limitations that make non-model organisms challenging subjects also make them fertile ground for conceptual innovation. By forcing reliance on theory rather than direct measurement, they reveal the creative and interpretive dimensions of scientific modeling. The philosophical challenge lies in embracing abstraction without losing contact with reality, while the practical challenge lies in implementing abstraction without generating arbitrary speculation. Together these challenges define a new epistemology of systems biology, one that values models not as substitutes for experiments but as instruments of reasoning that extend the reach of empirical science.

Theoretical modeling of gene regulatory networks in non-model organisms becomes both a scientific and a philosophical enterprise. It transforms the unknown into a structured field of inquiry, where questions about representation, causation, and explanation are as significant as questions about data and computation. The limitations that accompany these models do not diminish their value; they clarify the conditions under which knowledge is possible. In a discipline defined by complexity and partial understanding, theoretical modeling demonstrates that abstraction can be a form of insight, and that the boundaries of empirical science are not barriers but invitations to deeper reflection on how life can be understood through theory.

6. Conclusion

The theoretical modeling of gene regulatory networks in non-model organisms stands as one of the most significant developments in modern theoretical biology. It represents an intellectual synthesis of mathematics, computation, and evolutionary thought, offering a conceptual framework that transcends the boundaries of traditional experimental biology. The value of such modeling lies in its ability to generate structured understanding where empirical data are sparse, transforming the unknown into a field of inquiry defined by formal relationships and predictive logic. In doing so, theoretical approaches reveal that the essence of biology is not confined to direct observation but is deeply rooted in the capacity to reason abstractly about systems of life.

The study of gene regulatory networks reflects a broader philosophical movement in science toward understanding complexity through systems thinking. Living organisms can no longer be understood as simple assemblies of molecular components. They are dynamic systems whose behavior arises from the coordination of interactions across multiple levels of organization. For non-model organisms, whose biology remains largely unexplored, theoretical modeling offers a way to infer the rules that govern these interactions by analyzing patterns of information, structure, and function. Each network becomes an expression of how life organizes itself, adapting to constraints and responding to the environment through mechanisms that theory can describe even when experiment cannot verify.

The integration of mathematical abstraction with biological interpretation has redefined the practice of modeling itself. Equations and algorithms serve not only as computational tools but as representations of biological principles. They formalize hypotheses about regulation, feedback, and stability in ways that can be analyzed, compared, and generalized. This formalization gives theoretical models an autonomy that extends beyond specific datasets. A well-constructed model captures the logical architecture of life, demonstrating how universal laws of organization can manifest differently across species. For non-model organisms, this theoretical generality becomes essential. It provides the means to infer shared regulatory patterns that link the molecular ecology of diverse life forms into a coherent framework of understanding.

Theoretical modeling also deepens the conceptual unity between evolution and regulation. Gene networks evolve through processes of duplication, divergence, and selection, and their structure encodes the history of adaptation. Modeling these networks in non-model species allows the reconstruction of evolutionary logic in systems that cannot be experimentally probed. It reveals how robustness and flexibility co-exist, how novel traits arise from the reconfiguration of existing modules, and how stability in gene expression emerges from dynamic complexity.

In this sense, theoretical modeling does not only describe regulatory processes; it interprets them as evolutionary narratives inscribed in network topology and system behavior.

The role of theory in biology is often misunderstood as secondary to empirical discovery, yet in the study of non-model organisms, it becomes foundational. Theoretical models create the conceptual scaffolding upon which future experimentation can be built. They define the questions worth asking, predict relationships worth testing, and propose organizational patterns that guide empirical verification. This iterative relationship between theory and observation transforms biology from an inductive science into one that is both predictive and explanatory. Theoretical modeling of gene regulation thus exemplifies a new epistemic mode in which understanding is achieved not through accumulation of facts but through the articulation of coherent structures that give meaning to data.

The extension of theoretical frameworks into the domain of biodiversity has implications that reach beyond academic theory. Modeling gene regulatory networks in previously unstudied organisms allows the exploration of how life functions in extreme environments, how ecological interactions shape regulation, and how molecular systems adapt to diverse evolutionary pressures. These insights enrich not only basic science but also applications in conservation, agriculture, and biotechnology. Theoretical models become instruments for anticipating biological behavior, guiding decision-making in contexts where experimentation is limited or impossible. Theoretical modeling invites a reconsideration of what it means to understand life. To model a regulatory network is to engage with the logic of self-organization, to describe how complexity arises from simplicity through the interplay of structure and process. It is to view biology as a field where mathematics and imagination converge to reveal patterns that empirical methods alone cannot uncover. For non-model organisms, this convergence is especially significant. It transforms the absence of direct evidence into an opportunity for conceptual innovation, making theory itself a means of discovery.

The future of theoretical modeling in biology will depend on maintaining this balance between abstraction and realism. Models must remain grounded in biological plausibility while continuing to expand the boundaries of what can be inferred. The success of this approach will not be measured only by accuracy but by its capacity to inspire new forms of inquiry and to unify understanding across the diversity of life. The study of gene regulatory networks in non-model organisms embodies this vision. It shows that theoretical reasoning can illuminate the hidden order of living systems and that abstraction, when guided by rigor and imagination, can reveal truths about life that extend beyond the reach of observation. Through this union of theory and biology, the scientific imagination continues to transform uncertainty into knowledge, giving structure and meaning to the endless complexity of the natural world.

References

Bonasio, R. (2015). The expanding epigenetic landscape of non-model organisms. *Journal of Experimental Biology, 218*(1), 114–122.

Gelfand, M. S. (2023). Bioinformatics and evolution of non-model organisms. Institute of Molecular Genetics and Genetic Engineering (IMGGE) Belgrade Proceedings.

Levy, A., & Currie, A. (2015). Model organisms are not (theoretical) models. *The British Journal for the Philosophy of Science, 66*(2), 327–348.

Linde, J., Buyko, E., Altwasser, R., Hahn, U., & Guthke, R. (2011). Full-genomic network inference for non-model organisms: a case study for the fungal pathogen Candida albicans. *International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering, 5*(8), 379–389.

Linde, J., Schulze, S., Henkel, S. G., & Guthke, R. (2015). Data- and knowledge-based modeling of gene regulatory networks: an update. *EXCLI Journal, 14*, 346–378.

Rodgers-Melnick, E., Culp, M., & DiFazio, S. P. (2013). Predicting whole genome protein interaction networks from primary sequence data in model and non-model organisms using ENTS. *BMC Genomics, 14*, 608.

Sharma, R. (2013). Prediction of regulatory networks for non-model organisms. University of New Brunswick.

Williams, T. D., Turan, N., Diab, A. M., & Wu, H. (2011). Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. *PLoS Computational Biology, 7*(8), e1002126.

Xu, Y., Guo, M., Zou, Q., Liu, X., Wang, C., & Liu, Y. (2014). System-Level Insights into the Cellular Interactome of a Non-Model Organism: Inferring, Modelling and Analysing Functional Gene Network of Soybean (Glycine max). *PLoS ONE, 9*(12), e113907.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).