

Chapter 1

Gene Regulatory Models for Plant Development and Evolution

E.R. Alvarez-Buylla, M. Benítez, M. Aldana, G.J. Escalera-Santos, Á. Chaos, P. Padilla-Longoria, and R. Verduzco-Vázquez

1.1 Introduction: the Need for Mathematical Models to Understand Plant Development

During development, complex interactions amongst genetic and non-genetic elements give rise to robust spatiotemporal patterns. Moreover, an important feature of biological systems is the nontrivial flow of information at several scales. When we consider the scale determined by the cell, we observe that it integrates information coming from gene regulatory networks (GRNs), biochemical pathways, and other microscopic processes. If we consider larger scales, then intercellular communication, mechanical and geometric effects (such as growth, shape, and size), and environmental influences have to be taken into account. This is why understanding how patterns arise during development requires the use of formal dynamical models able to follow the concerted action of so many elements at different spatiotemporal scales.

E.R. Alvarez-Buylla, M. Benítez, M. Aldana, G.J. Escalera-Santos, Á. Chaos, P. Padilla-Longoria, and R. Verduzco-Vázquez

C3, Centro de Ciencias de la Complejidad, Cd. Universitaria, UNAM, México, D. F., México

E.R. Alvarez-Buylla, M. Benítez, G.J. Escalera-Santos, and Á. Chaos

Departamento de Ecología Funcional, Instituto de Ecología, Universidad Nacional Autónoma de México, 3er Circuito Exterior Junto a Jardín Botánico, 04510 Distrito Federal, Coyoacán, CU, Mexico

e-mail: eabuylla@gmail.com, ealvarez@ecologia.unam.mx

M. Aldana

Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Campus Cuernavaca, Morelos, 62210, Mexico

P. Padilla-Longoria

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, 04510, Distrito Federal, Mexico

R. Verduzco-Vázquez

Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, 62210, Mexico

The fact that biological entities and scales often interact nonlinearly makes mathematical modeling of biological systems, and in particular of gene regulatory networks, a nontrivial problem. From the mathematical point of view, the incorporation of all these interactions can be taken into account only by implementing hybrid models, that is, by incorporating both discrete and continuous elements, as well as deterministic and stochastic frameworks. In fact, depending on the specific space-time scale at which a process is being observed, it might appear discrete or continuous, deterministic or random. For instance, the levels of gene expression might be taken as discrete (the gene is “on” or “off”) when seen at rough space-time scales, but when observed with a finer gauge, these levels appear as continuously varying.

Mathematical models of GRNs provide an integrative tool, a systematic way of putting together and interpreting experimental information about the concerted action of gene activity. They also offer new insights on the mechanisms underlying biological processes, in particular developmental ones, as well as a means to make informed predictions on the behavior of such complex systems.

1.2 Dynamic GRN Models

Today, one of the most important challenges in systems biology is to relate the gene expression patterns of an organism with its observed phenotypic traits. Since these patterns result from the mutual activation and inhibition of all the genes in the genome in a coordinated way, the above problem is equivalent to relating the dynamical properties of the underlying genetic network with the organism’s phenotype (Hasty et al. 2001; Levine and Tjian 2003). In order to achieve this goal, one must decide first how to model the dynamics of the genetic network.

Amongst the several theoretical approaches that have been proposed to model the genetic dynamics, two stand out, namely, the *continuous* and the *discrete* (Smolen et al. 2000; Bower and Bolouri 2001). The continuous approach is based on systems of coupled nonlinear differential equations that describe the temporal evolution of the concentration of the chemicals involved in the gene regulation processes (proteins, enzymes, transcription factors, metabolites). This description is particularly suitable when the systems under consideration consist of a small number of components (Arkin et al. 1998; Vilar et al. 2003). However, large-scale genome analysis has revealed that the coordinated expression of dozens, or even hundreds of genes is required for many cellular processes to occur, such as cell division or cell differentiation (Whitfield et al. 1992; Rustici et al. 2004). For such processes, the continuous approach becomes intractable due to the great number of components and equations involved.

The discrete approach to model the dynamics of genetic networks was first introduced by Kauffman to describe, in a qualitative way, the processes of gene regulation and cell differentiation (Kauffman 1969). This approach focuses on the state of expression of the genes, rather than on the concentration of their products.

Thus, the level of expression of a given gene is represented by a discrete variable g that usually takes the values $g=0$ if the gene is not expressed, and $g=1$ if the gene is fully expressed. The genome is considered then as a set of N discrete variables, $g_1, g_2, \dots, g_N$, their values changing in time according to:

$$g_n(t+1) = F_n[g_{n_1}(t), g_{n_2}(t), \dots, g_{n_{k_n}}(t)] \quad (1.1)$$

In this equation, $(g_{n_1}, g_{n_2}, \dots, g_{n_{k_n}})$ are the k_n regulators of the gene g_n , and F_n is a discrete function (also known as a logical rule) constructed according to the nature of the regulators. The advantage of the discrete model is that it can incorporate a much larger number of components than the continuous models. Furthermore, recent work shows evidence that, in spite of the simplicity of the discrete approach, it is able to reproduce the gene expression patterns observed in several organisms (Huang and Ingber 2000; Mendoza and Alvarez-Buylla 2000; Albert and Othmer 2002; Espinosa-Soto et al. 2004; Davidich and Bornholdt 2008). This evidence has been obtained for relatively small genetic networks for which both the regulators and the logical rules are known for each gene.

Accumulated data on molecular genetics and current high-throughput technology (see next section) have made available a great amount of data regarding GRNs, yet information for all the regulators and logical rules in entire genomes is not available yet for any organism. Nonetheless, it is important to emphasize that, for the small genetic modules or sub-networks that have been thoroughly documented experimentally, the discrete approach gives accurate predictions.

Arguably, one of the most important results of the discrete model is the existence of *dynamical attractors*. Starting out from an initial state $[g_1(0), g_2(0), \dots, g_N(0)]$ in which some genes are active and some others inactive, Eq. (1.1) generates dynamics in which each gene goes through a transient series of active/inactive states until the whole network enters into a periodic pattern of expression (Fig. 1.1). Some genes reach a constant value that does not change in time anymore, whereas some others keep “blinking” in a periodic way. This periodic state of expression of the entire network is the dynamical attractor. The set of all the possible initial states that after a transient time fall into the same attractor is called the *basin of attraction* of that attractor. Each attractor is uniquely identified by its set of active genes. In other words, particular sets of genes are expressed in different attractors, and this is precisely the characteristic that identifies the different functional states of the cell. For this reason, Kauffman formulated the hypothesis—confirmed experimentally—that the dynamical attractors of the genetic network correspond to the different cell types or cell fates observed in the organism.

Since the level of expression of each gene is discretized into a finite number of values, the total number, Ω , of dynamical states in which the network can be found is also finite, and is given by $\Omega = \prod_{n=1}^N m_n$, where m_n is the number of discrete values that g_n can acquire. Under Eq. (1.1), the dynamical space of the network (i.e., the possible Ω states) is partitioned into disjoint sets consisting of the attractors and their corresponding basins of attraction (Fig. 1.1). This structure of the dynamical

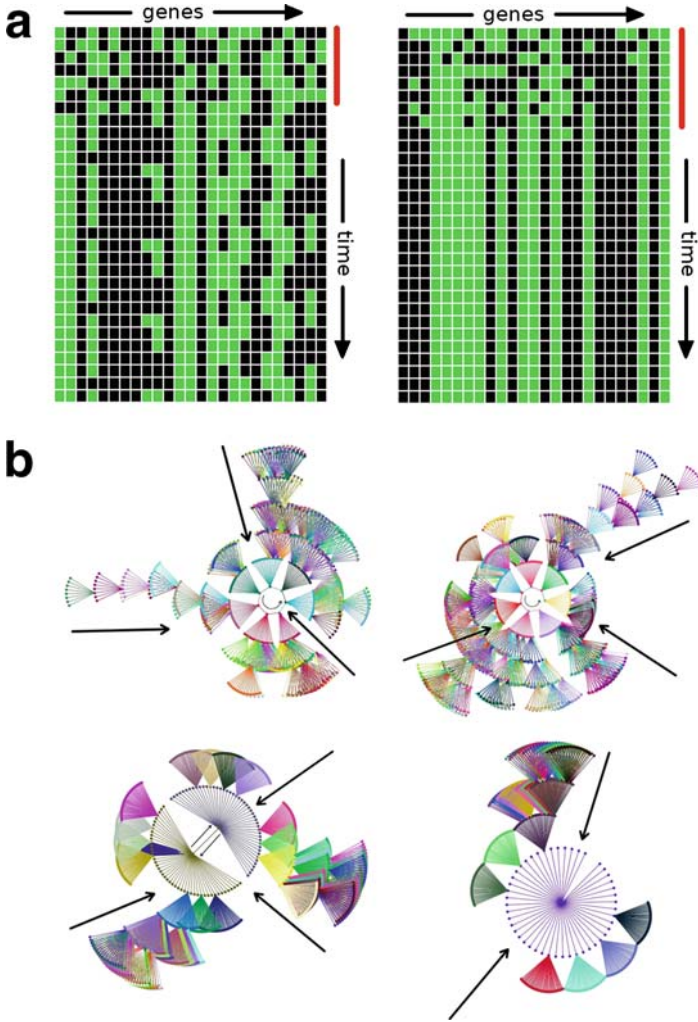


Fig. 1.1 Attractors and attractor basin in a GRN. **(a)** Visual representation of the dynamical attractors of a genetic network. Each square represents a gene, in *gray* if it is expressed, and in *black* if it is not. The genes are lined up horizontally so that each row represents the state of expression of the entire genome at a given time. Time flows downward. After a transient time (indicated with a *vertical line*), the whole network reaches a periodic pattern of expression, which is the dynamical attractor. As shown, two different initial states (the *uppermost rows*) can lead to different attractors. The attractor on the *left* has period six, whereas the attractor on the *right* consists of only one state. **(b)** Visual representation of the attractor landscape for a randomly constructed network with $N=12$ genes. Each *dot* represents a dynamical state of the network (i.e., one of the rows in **a**), and the *lines* represent discrete time steps. Two dots are connected if they are successive states under the dynamics given by Eq. (1.1). The fan-like structures reflect the fact that many states can have the same successor in time (the dynamics are dissipative). The *arrows* indicate the direction of the dynamical flow. In this particular example, the state-space of possible dynamical states organizes into four disjoint sets consisting of the attractors and their respective basins of attraction

space is called the *attractor landscape*, and constitutes a representation of the epigenetic landscape conceived by Waddington (1957) to qualitatively understand the different functional states of the cell. It has been shown recently for several cases that many important phenotypic traits of the organism, such as the cell type or the cell cycle, are encoded in the entire attractor landscape.

1.3 Inference of GRN Topology from Microarray Experiments

GRN architecture inference is the process by means of which information on the regulators is obtained from experimental data. In some cases, network structure has been inferred from thorough data of molecular genetics experiments, enabling novel insights and predictions for particular developmental systems (Mendoza and Alvarez-Buylla 2000; Albert and Othmer 2002; Espinosa-Soto et al. 2004). Nevertheless, the current available technology enables the generation of large sets of genomic information, commonly acquired from microarray experiments. This experimental technique allows observing the expression pattern of a set of genes at different sample points in time or under different experimental conditions, and has generated a vast data base.

Although powerful, microarray experiments and their data have two difficulties. First, an enormous number of experiments are necessary in order to confidently infer all the logical rules in a given genome. Second, the data obtained are very noisy, which is why uncovering structural or dynamic information is anything but trivial. We briefly introduce some of methods and approaches that have addressed the need of formal frameworks in this area.

Reverse engineering is the process of discovering the functional principles of a device, object, or system through analysis of its structure, function, or operation. In the context of GRNs, it constitutes the process of network structure inference from the analysis of experimental data on gene expression under diverse conditions, often derived from microarray experiments. Despite the particular method to be used to analyze microarray data, the overall goal of GRN reverse engineering is to find mathematical evidence supporting the proposition of an interaction between the nodes of the network.

Two main classes of methods have been proposed to infer GRN architectures via reverse engineering methods. The first class relies on probability theory, and its objective is to find the most probable network architecture given a genetic expression pattern, or to quantify the existent correlation between pairs of genes. Bayesian networks, both traditional and their dynamic variant, fall into the first approach, while mutual information methods fall into the second one. The second class of methods is based on continuous analysis. It involves ordinary differential equations (ODEs), and is supported by the theory and methods from stability analysis of dynamic systems.

Plant Developmental Biology - Biotechnological
Perspectives

Volume 1

Pua, E.C.; Davey, M.R. (Eds.)

2010, XXII, 497 p., Hardcover

ISBN: 978-3-642-02300-2