



**HAL**  
open science

# Life's Attractors Continued: Progress in Understanding Developmental Systems Through Reverse Engineering and In Silico Evolution

Anton Crombach, Johannes Jaeger

► **To cite this version:**

Anton Crombach, Johannes Jaeger. Life's Attractors Continued: Progress in Understanding Developmental Systems Through Reverse Engineering and In Silico Evolution. *Evolutionary Systems Biology: Advances, Questions, and Opportunities*, 2021, 978-3-030-71736-0. 10.1007/978-3-030-71737-7\_4 . hal-03128711v2

**HAL Id: hal-03128711**

**<https://hal.science/hal-03128711v2>**

Submitted on 21 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Life’s Attractors Continued: Progress in Understanding Developmental Systems Through Reverse Engineering and In Silico Evolution\*

Anton Crombach<sup>†</sup>

Johannes Jaeger<sup>‡</sup>

## Abstract

We present a progress report on our efforts to establish a new research program for evolutionary systems biology, based on reverse engineering and *in silico* evolution. The aim is a mechanistic understanding of the genotype-phenotype map and its evolution. Our review focuses on the case study of the gap gene network in dipteran insects (flies and midges). This network is the top regulatory tier of the segmentation gene hierarchy, generating a pattern of overlapping expression domains that subdivide the embryo during early embryogenesis. It is one of the best-understood developmental regulatory networks today. We have studied this system in a comparative way, across three species: the vinegar fly, *Drosophila melanogaster*, the scuttle fly, *Megaselia abdita*, and the moth midge, *Clogmia albipunctata*. In this context, we discuss methodological challenges concerning data processing and model-fitting, consider different functional decompositions of the gap gene network, and highlight novel insights into network evolution by compensatory developmental system drift. Finally, we discuss the prospect of simulating the phylogenesis of the gap gene network using *in silico* evolution. We conclude by arguing that our case study is a first step toward a more systematic empirical investigation into the principles of network evolution.

## 1 Introduction

A few years ago, we proposed a comparative research program into the function and evolution of developmental systems based on reverse engineering and *in silico* evolution [46].<sup>1</sup> The program’s main objective is to move beyond static descriptions of regulatory network *structure*, toward a more processual perspective, based on the analysis and comparison of network *dynamics* arising from the interactions between genetic and non-genetic factors [28, 29]. This objective stems from our desire to shed light upon a fundamental question in biology: how phenotypic variability is generated and shaped by developmental processes. The variability of phenotypic traits among individuals in a population provides the raw material for evolution by natural selection. In the famous words of Hugo De Vries, we are interested not in the survival but the *arrival* of the fittest [110].

Developmental processes not only constrain or bias but, more importantly, *generate* phenotypic variability and determine the variational properties of morphological traits [84, 111, 112]. In abstract terms, development can be represented as a mapping from genotype to phenotype (the G–P map), which relates variation at the genotypic level to the variability of phenotypes [2, 15, 75, 90]. Mathematically, G–P maps may either be interpreted as purely correlational maps (e.g., in quantitative genetics) or as characterizing causal-mechanistic processes (e.g., in evolutionary developmental and systems biology)

---

\*This manuscript is a chapter in the book *Evolutionary Systems Biology: Advances, Questions, and Opportunities* to be published with Springer-Nature.

<sup>†</sup>Inria, Antenne Lyon La Doua, 56 Boulevard Niels Bohr, 69603 Villeurbanne, France, anton.crombach@inria.fr

<sup>‡</sup>Complexity Science Hub (CSH) Vienna, Josefstädter Straße 39, A-1080 Vienna, Austria, jaeger@csh.ac.at

<sup>1</sup>While our argument focuses on developmental processes in multicellular organisms, it is directly extendable to other ontogenetic processes (metabolic, physiological, etc.) that constitute the life cycle of uni- and multicellular organisms.

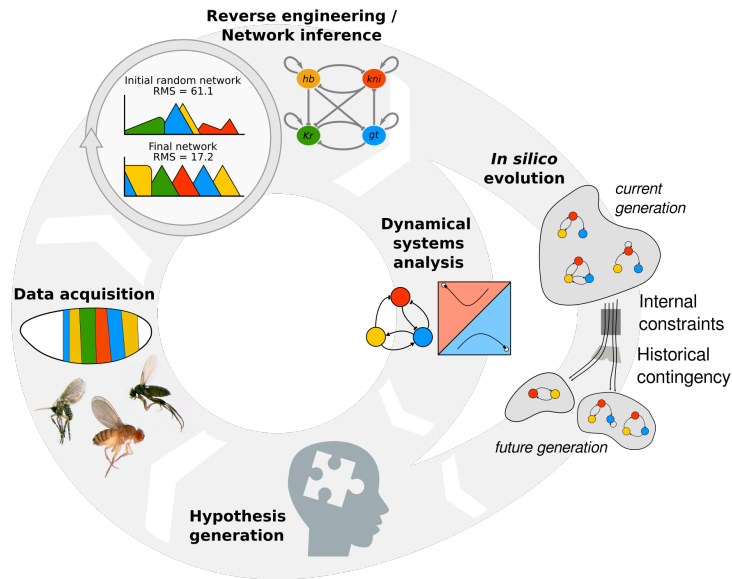
[28]. In the latter sense, G–P maps are thought to be best captured by models of developmental gene regulatory networks (GRNs) [26, 111]. If required, such networks can be expanded to include non-genetic regulatory factors.

Continuing on this path, we need empirical studies to establish the structure and dynamics of developmental regulatory networks, if we are to understand the origins of phenotypic variability [46]. This entails the decomposition of a system into building blocks using genetic and molecular methods and to *recompose* it to show how the building blocks and their interactions generate the orchestrated overall behavior of the process [8, 9, 10, 11, 13, 14, 28, 29]. Considering the complexity of development in even the simplest multicellular organisms, this recombination poses a formidable challenge.

To manage the regulatory complexity involved in characterizing developmental systems, we need dynamic mathematical models. These models integrate the components of a system by defining how these parts interact, which in turn leads to the characteristic dynamics of the system. We previously presented four distinct approaches that yield such models of developmental regulatory networks and their evolution [46]. These consist of (1) *ensemble approaches*, (2) *forward modeling*, (3) *reverse engineering*, and (4) *in silico evolution*. Ensemble modeling involves the simulation of entire categories of dynamic networks to reveal their shared general properties. Forward modeling is the classical bottom-up approach to biophysical modeling, where models are formulated in physicochemical terms and parameters are determined by measurement. In contrast, our proposal for a research program in evolutionary systems biology is based on reverse engineering and *in silico* evolution, which are employed together in a complementary fashion (Fig. 1) [46].

Over the last decade, much progress has been made in deciphering the evolution and functioning of developmental systems. Examples range from exploratory, computational studies [41, 54, 55, 106, 107, 108] to approaches combining theory and experiment [6, 78, 83, 123] (see also the chapter by Onimaru and Marcon). Here, we primarily focus on describing our own recent work on the gap gene system of dipteran insects (flies, midges, and mosquitoes), which we already used as a case study in [46] (see Fig. 1). A concise overview on this work can be found in [45]. It shows how reverse engineering can be used to analyze the dynamics and evolution of a complex regulatory network involving multiple regulatory genes—illustrating the high-dimensional and nonlinear nature of the G–P map and its influence on the evolutionary process. In particular, our work reveals how we can explain the phenomenon of developmental system drift in a mechanistic manner. Moreover, it demonstrates the multifaceted nature of a complex regulatory process and how alternative ways of subdividing it help in understanding its overall behavior. All of these aspects together yield a much deeper level of insight into the dynamics of development and evolution than could be achieved by genetic and molecular experimental processes alone.

In the rest of this chapter, we will reflect on a number of methodological and conceptual issues in reverse engineering and *in silico* evolution, which should be of general interest to practitioners in evolutionary systems biology. In particular, we document the major challenges we have faced, and the progress we have made to overcome them, during our efforts to reverse engineer the gap gene system across three different evolutionary lineages of dipteran insects. The chapter is composed of four sections. In the first, we focus on methodological bottlenecks imposed by the reverse engineering approach. In the second, we discuss a bottom-up strategy to decompose and recombine a complex dynamical system such as the gap gene network in order to generate mechanistic explanations of the underlying regulatory process. We also briefly touch upon a complementary top-down strategy [105]. Due to space constraints, we refer to the chapter by Jaeger and Monk for an in-depth account of this approach. In the third part, we connect these explanations to network evolution. Finally, we conclude with an outlook on future work using *in silico* evolution that will shed light on the evolutionary transitions that have occurred in insect body segmentation.



**Figure 1: The systems biology cycle for reverse engineering evolving developmental processes** (modified after [61]). Researchers *generate hypotheses* about regulatory mechanisms underlying observable developmental dynamics (e.g., gene expression patterns). *Data are acquired* that capture all relevant features of those developmental dynamics, shown as expression domains on a schematic fly embryo. For our comparative study, we used data from three species of dipteran insects: the moth midge *Clogmia albipunctata*, the vinegar fly *Drosophila melanogaster*, and the scuttle fly *Megaselia abdita* (left to right). Dynamical models (gene circuits) are then fitted to the data in order to *infer the regulatory structure of the underlying network*. Root mean square (RMS) deviation indicates quality of the fit (lower is better). The best circuits and the topology of their phase spaces (trajectories, attractors, and their basins, shown schematically for a bistable switch network) are then analyzed using the tools of *dynamical systems theory*. Comparative analyses distinguish between essential and accidental features of the system. In addition, we can use fitted circuits as start and end points of *in silico evolutionary simulations*, which reveal possible evolutionary pathways between mechanisms, and how much they depend on internal constraints or historical contingency. Gap genes: *hunchback* (*hb*), yellow; *Krüppel* (*Kr*), green; *knirps* (*kni*), red; *giant* (*gt*), blue

## 2 Reverse Engineering with Gene Circuits

Let us first provide a brief outline of our reverse engineering approach. It requires (1) a suitable model of the regulatory system, (2) an efficient and reliable method for inferring the parameters of the model by fitting it to (3) data that suitably represent the observable dynamics of the system. (4) Finally, we need a conceptual framework and analytical tools to extract biological insights from the resulting set of model fits. We will discuss each of these steps in turn.

We use gene circuits as models of the gap gene network [68, 79]. Gene circuits are dynamical models, historically derived from recurrent Hopfield neural networks [43]. In mathematical terms, a gene circuit is defined as a set of coupled ordinary differential equations (ODEs). Each equation describes the change in concentration  $g$  over time  $t$  of the product (mRNA or protein) of gene  $a$  in nucleus  $i$ :

$$\frac{dg_i^a}{dt} = R^a \Phi(u^a) - \lambda^a g_i^a + D^a(n)(g_{i-1}^a + g_{i+1}^a - 2g_i^a) \quad (1)$$

with  $R$ ,  $\lambda$ , and  $D(n)$  representing rates of production, decay, and diffusion, respectively. Diffusion rates depend on nuclear density, i.e., the number of nuclear divisions  $n$  that the system has undergone at time  $t$ .

Gene regulation is modeled by a sigmoid regulation-expression function  $\Phi$ . This function captures

the coarse-grained dynamics of eukaryotic transcriptional regulation, that is switch-like activation with a saturating response:

$$\Phi(u^a) = \frac{1}{2} \left( \frac{u^a}{\sqrt{(u^a)^2 + 1}} + 1 \right), \quad (2)$$

$$u^a = \sum_{b \in G} w^{ba} g_i^b + \sum_{m \in M} e^{ma} g_i^m + h^a \quad (3)$$

Regulatory interactions are represented by an interconnectivity matrix of regulatory weights  $w \in W$ , and external input weights  $e \in E$  for regulators that are not themselves regulated by gap genes. In our case study, internal regulators include the trunk gap genes  $G = \{hb, Kr, gt, kni\}$  and external inputs are provided by maternal gradients and the products of terminal gap genes  $M = \{Bcd, Cad, Tll, Hkb\}$ . Each weight  $w$  and  $e$  determines whether a specific regulatory interaction exists (non-zero weights) and, if so, whether it is activating (positive) or repressing (negative). These matrices abstract from the complex (and largely unknown) biochemical details of eukaryotic transcriptional regulation.  $h$  is a threshold parameter that captures the transcriptional state of a gene in the absence of any spatially specific regulators.

Reverse engineering means fitting gene circuits to spatiotemporal gene expression data. This is framed as an optimization problem where values need to be estimated for the parameters  $W, E, h, R, D$ , and  $\lambda$ , such that the simulated model reproduces the expression data as closely as possible. The two weight matrices  $W$  and  $E$  are central to this challenge. They represent the regulatory structure of the network, which cannot be measured directly, but must be inferred from data. In other words, gene circuits are used as an analytical tool to understand the genetic regulatory mechanisms driving the observed dynamics of gene expression.

To determine the difference between model output and gene expression data we can use different residuals or cost functions. Originally, we used the ordinary least squares (OLS) approach [49, 50, 66, 67], which was later replaced by weighted least squares (WLS), which take the variability of the data into account (see section 3.2). Thus, our cost function is defined as:

$$\text{cost} = \sum_{a \in G} \sum_{t \in T} \sum_{i \in N(n)} v_i^a(t) (g_i^a(t) - \text{data}_i^a(t))^2 \quad (4)$$

The value of this function depends on the number of data points used for fitting. To make the score of a fit comparable between different simulation settings, we use the root mean square deviation (RMS; see Fig. 1):

$$\text{RMS} = \sum_{a \in G} \sum_{t \in T} \sum_{i \in N(n)} \sqrt{\frac{1}{N_{\text{data}}} (g_i^a(t) - \text{data}_i^a(t))^2} \quad (5)$$

Since the RMS still depends on the scale of the data, we scale mRNA data to match protein data [23]. The result is an RMS which is a normalized measure for the differences between model output and observed expression levels.

We find optimal parameter sets by minimizing the cost function (Eq. 4). This poses a challenging nonlinear global optimization problem. Several algorithms have been used to solve this challenge. Initial studies using gene circuits used a global optimization approach called parallel Lam simulated annealing (pLSA) [18, 64]. pLSA is a robust optimization method, that is computationally costly. If high-performance computing facilities are available, it remains our preferred fitting method. Several alternatives, mostly based on evolutionary computation and scatter search, have been proposed to reduce computational cost [1, 32, 59, 74]. Until recently, however, none of these alternatives were able to match the robustness and reliability of pLSA on our particular problem. This appears to have changed with a new method called FIGR, which converts the problem of fitting nonlinear gene circuit models (based

on sigmoid regulation-expression functions) to a classification problem on models with purely on-off regulatory dynamics (Heaviside step function), which can be solved using standard machine learning approaches [31]. This method leads to significant speedup without a loss of robustness, which effectively removes the computational bottleneck of fitting models to data (see section 3).

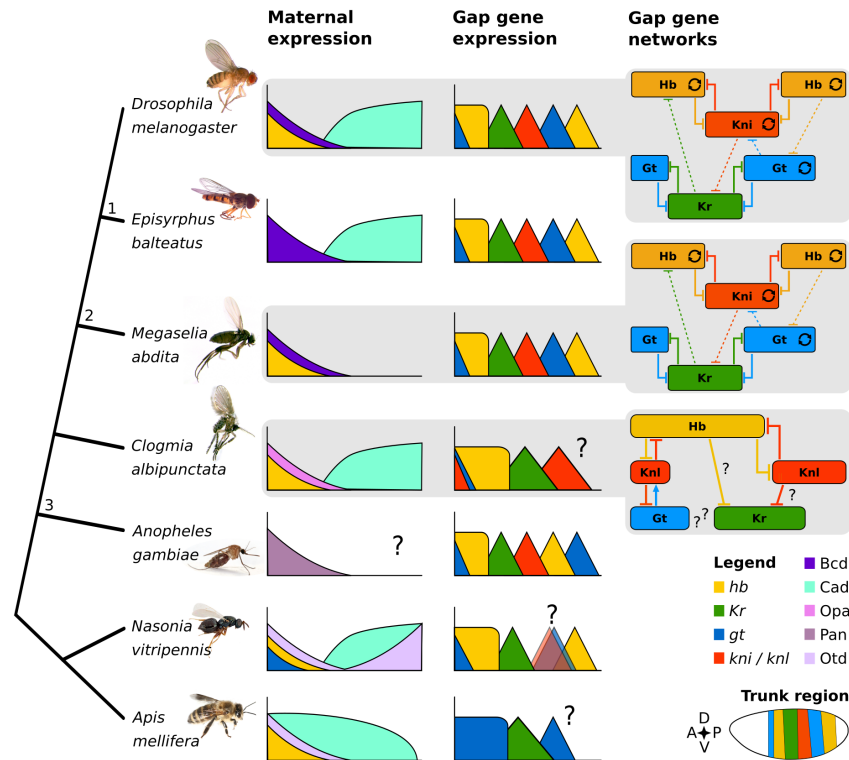
After optimization is performed, model fits with an RMS below a certain threshold are carefully inspected for visible expression defects and selected for analysis. We also assess the determinability of the inferred parameter values using statistical methods [3, 4, 5, 23]. In the best case, this yields a consensus network structure, where the qualitative nature of each interaction (activation or repression) is shared between a large majority of the selected fits. Alternatively, we end up with a set of different network variants that can all explain the observed expression dynamics. In both cases, solutions must be compared to and validated with experimental evidence. From the set of validated best-fitting solutions, we choose representative examples for further analysis of regulatory mechanisms and phase space topology [66, 102, 103, 104]. These analyses result in mechanistic and structural explanations for the dynamical behavior of the system (Fig. 1) [28].

### 3 Challenges in reverse Engineering Gap Gene Networks

The gap gene network is active in early embryogenesis, during the blastoderm stage, before the onset of gastrulation [44]. Pioneering genetic work in *D. melanogaster* established that gap genes form the top-most zygotic layer of the hierarchical gene regulatory network governing segment determination. They encode short-lived transcription factor proteins that diffuse through the syncytial blastoderm embryo. Here, we consider the four trunk gap genes *hunchback* (*hb*), *Krüppel* (*Kr*), *giant* (*gt*), and *knirps* (*kni*). They generate a set of broad, overlapping expression domains subdividing the embryo along its main (or anteroposterior, A–P) body axis (Fig. 2). No tissue growth is involved in this process. Gap genes are activated by maternal gradients of Bicoid (Bcd), Hb, and Caudal (Cad) proteins, and cross-regulate each other. Together with the maternal factors, they activate the pair-rule genes, whose periodic two-segment patterns are exemplified by the seven stripes of *even-skipped* (*eve*) expression. Pair-rule genes, in turn, regulate the molecular pre-pattern of the segment polarity genes, expressed in 14 single-cell wide stripes, which demarcate the boundaries of the embryonic parasegments that form later in development.

The molecular pre-pattern of the segment polarity genes is widely conserved across insects and other arthropods, but the peculiar way by which *D. melanogaster* arrives at it is not [17, 21, 27, 81]. The simultaneous subdivision of the pre-gastrulation embryo observed in *D. melanogaster* is called long-germband segment determination. Outside the dipteran lineage, it also occurs in some groups of beetles and the Hymenoptera (ants, bees, and wasps; see Fig. 2). In contrast, other insect lineages determine most of their segments later in development, after gastrulation has occurred, through sequential addition and tissue growth. This sequential mode is called short-germband segment determination.

To better understand how the derived long-germband mode of segment determination in *D. melanogaster* originated and evolved, we set out to perform a comparative analysis of the mechanisms underlying gap gene expression across several distantly related species of dipteran insects. To achieve this, we reconstructed the gap gene system *in silico* in two non-model species. The first is the phorid scuttle fly *Megaselia abdita*, which belongs to a basally branching cyclorrhaphan lineage [115]. *M. abdita* possesses a Bcd gradient, but no maternal Cad expression (Fig. 2) [62, 91, 92, 93]. Gap gene expression in this species is very conserved compared to *D. melanogaster*, which allows us to study how compensatory evolution maintains patterning output in the presence of changing maternal inputs [120, 121]. In addition, we attempted to reverse engineer the gap gene system of the psychodid moth midge *Clogmia albipunctata* (Fig. 2). In this basally branching dipteran lineage [56], there is no *bcd* gene, and posterior gap gene expression differs markedly from that of both other species [37, 53, 80].



**Figure 2: Evolution of the gap gene system in holometabolan insects.** A simplified phylogenetic tree is shown, featuring selected holometabolan insect lineages with (partially) known gap gene expression and regulation. The top branch of the tree represents the dipteran lineage (flies, midges, and mosquitoes) and the bottom branch the hymenopteran lineage (bees and wasps). The vinegar fly *Drosophila melanogaster*, hoverfly *Episyrphus balteatus*, and scuttle fly *Megaselia abdita* represent cyclorrhaphan lineages; the moth midge *Clogmia albipunctata* and the malaria mosquito *Anopheles gambiae* depict nematoceran lineages within the Diptera. *Apis mellifera* is the honeybee and *Nasonia vitripennis* the jewel wasp. Key evolutionary events are numbered on the tree: 1 and 3, loss of maternal Hunchback (Hb); 2, loss of maternal Caudal (Cad) expression. The second column shows maternal morphogen gradients [122], while column three depicts the relative position of gap gene expression domains along the A–P axis [44, 63, 65, 69, 76, 116, 117]. Column four shows the structure of the three reverse engineered gap gene networks described in the text, with inhibitory interactions indicated by T-bars and self-activation by circular arrows (activation by maternal gradients is omitted for simplicity). Question marks indicate incomplete or missing information: *C. albipunctata* may have an additional posterior gap gene, and several of its gap gene interactions remain undetermined [24, 37]. The posterior morphogen of *A. gambiae* is unknown. Data is ambiguous regarding the ordering of posterior *gt* and *kni* in *N. vitripennis* [69]. The schematic embryo in the bottom-right corner shows the location of gap gene expression patterns in the trunk region of cyclorrhaphan flies (anterior is to the left; dorsal is up). Gene/protein names in the legend: *hunchback* (*hb*), *Krüppel* (*Kr*), *giant* (*gt*), *knirps* (*kni*), *knirps-like* (*knl*), Bicoid (Bcd), Caudal (Cad), Odd-paired (Opa), Pangolin (Pan), Orthodenticle (Otd). Image sources: all dipterans by Wotton *et al.* [120], except *A. gambiae* by Muhammad Mahdi Karim (Wikimedia Commons), *A. mellifera* (*carnica*) from Makro Freak (Wikimedia Commons), and *N. vitripennis* from the New Zealand Arthropod Collection (flickr.com).

### 3.1 Bottleneck No. 1: Quantitative Data

reverse engineering complex regulatory systems poses significant methodological challenges, particularly if attempted in non-model organisms, where robust and efficient experimental protocols are often lacking. The first major challenge consists in the generation of suitable data for model fitting and model validation. This is the principal bottleneck, especially if we are aiming to understand processes of pattern

formation or morphogenesis that require spatially resolved time series of gene expression data.

The quality of a reverse engineered model crucially depends on the quality of the data used for fitting. However, the amount of work and the technical complications involved in generating high-quality spatial gene expression data can be daunting. Progress in high-throughput methodology for image bioinformatics is being made, but is slower than in other areas, such as sequencing. For instance, it took over five years to generate the gap protein expression data used to fit the initial set of gene circuit models for *D. melanogaster* [5, 49, 50, 66, 67]. These data were based on immunofluorescence staining protocols and confocal microscopy, combined with a quantitative data-processing pipeline to measure the concentration of maternal gradients and gap gene proteins at high spatiotemporal and cellular resolution [95, 96].

Clearly, for our comparative project, we had to find a good compromise between data quality and effort. As we struggled to raise suitable antibodies for immunofluorescence in non-model dipteran species (for a partial success story in *C. albipunctata*, see [53]), we chose instead to demonstrate that post-transcriptional regulation of gap genes is not essential for the establishment and dynamics of their expression domains in *D. melanogaster* [12]. This implied that we could use gap gene mRNA expression data for model fitting. Such data are much easier to acquire in non-model systems than protein expression data.

We proceeded to develop a fast and robust approach to generate semiquantitative mRNA expression data based on colorimetric *in situ* hybridization protocols, imaging by widefield microscopy, and a standardized, interactive pipeline for “medium-throughput” data processing and consistent staging of embryos into homologous developmental time classes [22]. Our measurements are semiquantitative in the sense that they only capture the timing and location of gene expression precisely, while accurate levels of expression cannot be assayed due to the nonlinear nature of the staining protocol. Expression data are available through a database called SuperFly, which contains expression data from over 1500 embryos from all 3 species (<http://superfly.crg.eu>) [19]. This unique data set of spatiotemporal gene expression patterns provided the basis for our comparative study of gap gene regulation dynamics across dipteran species. Once pipeline and database were in place, it only took a few months for three researchers to generate and process all the data, which means that the approach is scalable to much larger systems.

A related challenge, that tends to receive little attention, is appropriate embryo staging and the establishment of homology between developmental stages in different species. The staging scheme used in *D. melanogaster* resolves expression patterns down to intervals of about 7 min in the late blastoderm (the whole blastoderm stage lasts for about 90 min) [96]. This scheme is based on morphological features, such as the progression of membrane invagination during cellularization, that are not necessarily conserved in other species. For this reason, we had to establish equivalent staging schemes for *M. abdita* and *C. albipunctata*. This required the identification of suitable morphogenetic landmarks: the occurrence of specific nuclear divisions and nuclear movements, the appearance of the head furrow, pole cell formation, and the onset and progression of cellularization during the blastoderm stage. The task was complicated by the possibility of heterochronic shifts in developmental timing between landmarks and by the presence of extraembryonic membranes (amnion and serosa) in *M. abdita* and *C. albipunctata*, which are heavily reduced in *D. melanogaster* [87]. Despite considerable differences between species, our efforts resulted in a consistent sequence of homologizeable developmental stages, which put us in the exceptional position to draw detailed comparisons between the dynamical properties of gap gene expression between all three species [57, 119].

Next, we ensured that our semiquantitative data from *D. melanogaster* would yield the same kind of fitted gene circuit models as the (more accurate and fully quantitative) original set of protein expression data [23]. Not surprisingly, these efforts yield model solutions with more dispersed parameter values compared to the original protein fits [5]<sup>2</sup>. Moreover, parameter determinability analysis indicates

---

<sup>2</sup>We also pruned data and time points from our data set to test the minimal requirements for success. This led to the insight that the data should have the right spatiotemporal resolution to capture the expression features one wants to explain, such as transitions between different types of dynamics or the appearance of expression boundaries and domains



that the level of statistical confidence generally decreases when we use semiquantitative mRNA data. These shortcomings can be overcome, however, using experimental approaches such as RNA interference (RNAi) to empirically validate the nature and strength of a given regulatory interaction. More importantly, in *D. melanogaster* at least, we can still reconstruct a consensus network structure that includes all the relevant developmental mechanisms identified in gap gene circuits fit to protein expression data [5, 23, 50].

All of this is encouraging news for attempts at fitting gene regulatory network models to spatiotemporal gene expression data in general. We expect many developmental systems to be more sensitive to alterations in the timing and location of gene expression than to the precise level of expression for specific factors. This needs to be confirmed case by case, of course. But if it is a general trend, then fast semiquantitative approaches will be sufficient to yield robust and consistent results for reverse engineering in many developmental systems.

### 3.2 Bottleneck No. 2: Model Fitting

Another bottleneck for reverse engineering lies in the global optimization procedure required for fitting nonlinear dynamical models to expression data. We have already mentioned above that computational efficiency is no longer the problem today than it was a few decades ago. This is mainly due to the incessant increase in available computational power, but also because of more effective optimization algorithms (e.g., scatter search and FIGR) becoming available (see section 2). However, a number of other problems remain, which have to do with the determinability of parameter values. As discussed in the previous section, fits to mRNA expression data result in higher dispersion and lower determinability. Another factor that affects determinability is the presence of correlations between parameter values [3, 4]. Such correlations are known to exist between rates of production, diffusion, and degradation in kinetic systems. Moreover, correlations can arise from redundancies in the regulatory network. For instance, auto-activation can compensate for ubiquitous activation of gene expression, and vice versa, or higher rates of maternal activation can be compensated by stronger rates of cross-repression between gap genes.

Other parameter correlations are not so obvious to detect. They require parameter determinability analysis to be uncovered. One of these cases affects repression of trunk gap genes by terminal gap genes *tll* and *hkb*. It turns out that all but one of these regulatory interactions—the repression of *hb* by Hkb—are functionally redundant [5]. This not only corroborates experimental observations [44], but also allows us to eliminate these parameters from optimization, since their inferred values would not carry useful regulatory information. In turn, this leads to the practical benefits of speeding up the fitting procedure and reducing the dispersion of other parameter values. Model fitting, therefore, is most efficient if informed by functional considerations, without unnecessarily constraining the range of possible parameter values for the interactions for which we can extract information from the expression data. Naturally, acquiring these insights is heavily problem- and context-dependent, and must be done case by case in an empirical manner.

Two other aspects of model fitting are worth noting. First, fitting quality and biological accuracy of inferred parameter values improved greatly when we introduced a weighted least squares (WLS) cost function, which takes both average expression patterns and their standard deviations into account [3, 5] (see section 3). As a general rule, protein expression shows variances that are roughly proportional to expression levels [95]. For this reason, we introduced artificial variances to levels of mRNA expression in our semiquantitative data, which mimicked the pattern observed for protein data [23]. This greatly improved the determinability of parameter estimates, by putting a high penalty on gene expression in areas of the embryo where they should not be expressed. As a second step for quality control, two independent experts inspected each circuit for remaining small domains of ectopic expression. In our experience, these circuits tend to lead to artifacts in the inferred parameter values and need to be removed as outliers that confuse the analysis of network structure.

---

[23].

In conclusion, we recommend an iterative procedure for reverse engineering where the values of inferred parameters are increasingly constrained through empirical considerations and determinability analysis. More generally, good model fits require data that capture the timing and location of relevant expression features, and take variability of gene expression into account. Getting the exact levels of gene expression right is not crucial, at least for the gap gene system (see also [101]). Our semiquantitative approach with mRNA instead of protein data is robust across model and non-model species and requires much less time and effort than methods based on immunofluorescence. A trained technician or postdoc is able to gather a data set for model fitting in about 3–4 months.

Taken together, this means we have overcome the two main bottlenecks involved in reverse engineering systems of spatial pattern formation: the acquisition of microscopy data and the fitting of a consistent set of network models. This establishes the reverse engineering approach as a method that is widely applicable in evolutionary systems biology.

## 4 The Art and Science of Network Decomposition

Once we have obtained a set of validated gap gene circuits with a consensus network structure (or a small set of network variants), we must find a way to analyze these models to extract biological insight from them. What we need to do to understand the G–P map as a causal regulatory process, is to reconstruct how parts of the system govern specific features of the developmental dynamics that propagate the process from its initial to its final state. Importantly, this can be done in different ways that are all equally valid. Here we discuss a bottom-up approach that enables a first recomposition of parts of the network. This approach is complemented by a top-down analysis that we briefly elaborate on here and that is described in more detail in the chapter by Jaeger and Monk.

Our starting point is the consensus network structure of the gap gene system in *D. melanogaster* (Figs. 2 and 3). Simply by looking at this representation of the network, we can infer certain regulatory principles. For instance, the two double-negative (hence positive) feedback loops between pairs of gap genes with complementary expression patterns are easy to notice: *hb* and *kni*, as well as *Kr* and *gt* (Fig. 3b). The bistable switching behavior typically driven by such positive feedback loops explains that only one of the genes in each pair can be present in an embryonic nucleus at a given time.

In other cases, however, visual inspection and mental simulation are not sufficient. For example, what do the weak repressive interactions between overlapping gap domains do (Fig. 3c)? Upon their discovery, these interactions elicited some controversy [44], and no specific regulatory role for them could be determined based on qualitative evidence from genetics. This only changed with quantitative, high-resolution data and gene circuit models, which revealed that these interactions are involved in dynamic positional shifts of posterior gap domains toward the anterior [50, 95]. These domain shifts are an important expression feature that can neither be explained by genetic experiments (since the shifts are an emergent property of the system and do not depend on a specific mutation) nor mental simulation of the system (since the interactions that generate them are too numerous and interconnected to simply “think them through”). This is why we need dynamical models for network analysis and mechanistic explanations [28, 29].

### 4.1 Bottom-up Decomposition into Regulatory Mechanisms

Our validated gene circuit models allow us to directly track the influence of each specific regulatory interaction in the gap gene network, in each nucleus separately, and at any time point during segmental patterning. To achieve this, effective regulatory contributions are calculated as regulator concentration  $g_t^a$  multiplied by regulatory weights from matrices  $W$  and  $E$  (see Eq. (3)) and plotted against each other for visual inspection (see Fig. 6C in [23]). Regulatory inputs can be integrated over an embryonic region and time interval of interest. This enables us to precisely characterize the relative importance of individual contributions to every observable expression feature that is correctly reproduced by the model.

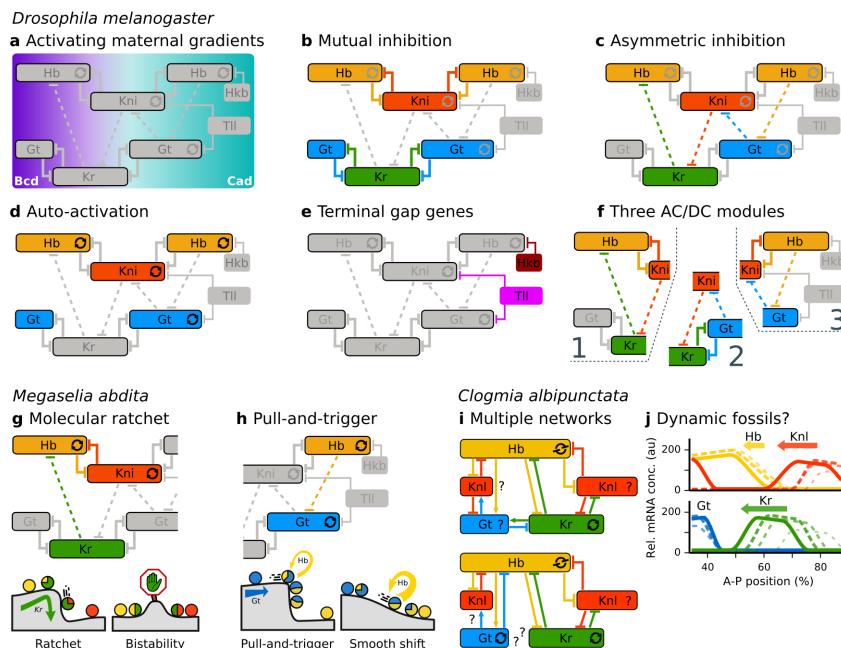
Using this kind of graphical analysis, we were able to distinguish five basic developmental mechanisms involved in gap gene expression in *D. melanogaster* (Fig. 3a–e) [23, 49, 50]:

1. Gap genes are predominantly activated by maternal factors Bcd (in the anterior) and Cad (in the posterior region of the embryo), especially during early stages when gap protein levels are still low (Fig. 3a) [42, 51]. Bcd acts in a graded spatially distributed manner, and Cad mainly contributes to activation in the posterior region of the embryo, where its early concentration levels are uniformly high. This activating mechanism does not correspond to classical morphogen-based patterning, since maternal inputs are not specific enough on their own to precisely position gap domain boundaries [47, 48, 52]. As Bcd and Cad concentrations reduce over time, maternal activation diminishes, yet its role remains important as it contributes to the dynamics of gap domain shifts [103].
2. Strong mutual repression between complementary gap domains of *hb* and *kni*, as well as *Kr* and *gt* establishes the basic staggered arrangement of gap gene expression domains (Fig. 3b) (see above). This double-negative (hence positive) feedback mechanism lies at the core of gap gene regulation and is conserved among dipteran species [24, 25] (Fig. 2). In addition, it is required for the sharpening of gap domain boundaries over time [44].
3. Gap genes with overlapping expression domains weakly repress each other. These interactions show a posterior bias such that Hb represses *gt*, Gt represses *kni*, Kni represses *Kr*, and Kr represses *hb*, but not the other way around (Fig. 3c). As mentioned above, this asymmetric cascade of repressive interactions causes the temporal shift of posterior gap domains toward the embryo's anterior.
4. Auto-activation boosts expression levels and contributes to the sharpening of expression domain boundaries of *hb*, *Kr*, and *gt* at later stages, when gap protein levels are sufficiently high (Fig. 3d). However, auto-activation does not influence the determination of positional information, and is dispensable for gap gene patterning in principle [74]. Our models indicate that auto-activation is present in all three species [23, 24, 25]. However, we cannot analyze its precise regulatory role, since auto-regulation is mainly involved in controlling levels of expression, which are not accurately captured by our semiquantitative data. We do know that in *D. melanogaster* *gt* expression levels are governed by a temporal switch from regulatory elements that mediate broad early expression in response to maternal gradients, to domain-specific elements later on [42].
5. Repression by terminal gap genes *tll* and *hkb* excludes the expression of gap genes from the posterior pole, restricting them to the embryo's trunk region (Fig. 3e). The terminal gap genes mediate the influence of the terminal maternal system on segmentation gene expression [44]. Most of the regulatory inputs from these two genes are redundant with the exception of repression of *hb* by Hkb, which establishes the posterior boundary of the posterior *hb* domain [5].

## 4.2 Top-down Decomposition into Dynamical Modules

Decomposing a network by tracking single regulatory interactions is but one possible way to analyze the dynamics of a complex regulatory system. We call it the bottom-up approach, as it examines how individual contributions combine to yield the overall behavior of the network. In contrast, a top-down approach considers dynamics at the level of the whole system and dissects it into modular components (see the chapter by Jaeger and Monk). This approach is based on dynamical systems theory, specifically phase space analysis [94]. For our gene circuit models this means to identify attractors, their basins, and the separatrices between them. In time-variant systems, like the gap gene network, we also need to consider the possible transitions (bifurcations) between different dynamical regimes and transient behavior along trajectories that stay far from any steady state [102].

Phase space analysis of gap gene circuits reveals two fundamentally different patterning regimes: stationary domains in the anterior, generated by multi-stability, versus shifting domains in the posterior,



**Figure 3: Comparative analysis of dipteran gap gene networks.** **a–e** The bottom-up decomposition of the *D. melanogaster* gap gene network reveals the five basic regulatory mechanisms shown here. **f** A complementary, top-down decomposition reveals three dynamical modules. Each module has the structure of an AC/DC circuit [105] (see the chapter by Jaeger and Monk). **g** A molecular ratchet involving *hb*, *Kr*, and *kni* can explain the shift of the anterior *hb* domain boundary in *M. abdita* and the lack of such a shift in *D. melanogaster*, where this boundary is set through bistability instead. **h** In *M. abdita*, posterior *Gt* activates *hb*, which triggers *hb* auto-activation once *Hb* concentration has sufficiently built up (“pull-and-trigger” mechanism), while *D. melanogaster* exhibits a smooth shift based on *hb* auto-activation only. **i** reverse engineering the *C. albipunctata* gap gene system resulted in a set of four different network structures, two of which are shown here. **j** Gap domain shifts are much more pronounced in *C. albipunctata* than in *D. melanogaster* or *M. abdita*, suggesting that these shifts may be a dynamic vestige (or “fossil”), a remnant inherited from the ancestral short-germband segmentation process. Inhibitory interactions are indicated by T-bars, self-activation by circular arrows, and self-inhibition by circular T-bars. For gene names, see Fig. 2. See main text for details.

generated by a damped oscillator [104]. The presence of such different regimes allows us to identify which subsets of gap genes contribute to each of them (Fig. 3f) [105], even if the gap gene network shows no signs of modularity in its structure, being very densely connected (see Fig. 2). In the anterior trunk region, for example, only *gt*, *hb*, and *Kr* are expressed. Their interactions generate switch-like multi-stable behavior leading to stationary domain boundaries for these gap genes. In the posterior trunk region, *Kr*, *kni*, and *gt* are expressed, where they generate oscillatory behavior and shifting domain boundaries. Surprisingly, the central region of the embryo contains a third dynamical module composed of *hb*, *Kr*, and *kni*, which drives two different kinds of behavior. It straddles the bifurcation point between nuclei that show switch-like and oscillatory behavior. Thus a non-modular network produces modular, multifunctional behavior. These subsets of gap genes are not structural but *dynamical* modules.

Strikingly, all three dynamical modules of the gap gene network exhibit exactly the same network structure (topology) (Fig. 3f), only differing in strength of regulatory interactions. This canonical structure corresponds to the AC/DC circuit, a known minimal motif for producing both switch-like and oscillatory behaviors [71]. This explains how a single subcircuit can produce two different dynamical behaviors depending on the maternal input it receives.

In summary, we demonstrate that one can decompose a regulatory network in multiple ways: bottom-up, by plotting the contributions of individual regulatory interactions; and top-down, by decomposing

the system based on different kinds of behavior observed in the system’s dynamics. Both of these approaches are valid ways to decompose a network. Both describe how specific subsets of interactions between gap genes contribute to specific aspects of gene expression. Yet, each decomposition has its peculiar focus. In the bottom-up approach, we classify contributions to specific expression features, such as the positioning, sharpening, or shift of particular domain boundaries. In the top-down approach, we focus on more general dynamical behaviors. Even though gap domains in the anterior and the posterior of the embryo look similar at first sight, they are generated by fundamentally distinct processes. These analyses may not easily add up to a unified picture of the network. Instead, they complement and contextualize each other, which makes it easier to recognize the uses and limitations of either approach. In short, these perspectives synergize to give us a deeper understanding of the system [118].

## 5 Drifting Shifts: The Evolution of the Gap Gene System

A comparative analysis of the gap gene network beyond the drosophilid lineage illuminates the origin and evolution of long-germband segment determination. Moreover, it allows us to distinguish conserved aspects of the patterning mechanisms from contingent regulatory features that are evolutionarily labile. More generally, our reverse engineering approach grounds the search for principles of network evolution in empirical evidence.

The gap gene network, in turn, is an ideal model system in which to establish methodology for comparative modeling of regulatory evolution. We can start our analysis with a common set of candidate genes, since gap genes are highly conserved across dipteran lineages [44]. We can track gene expression using mRNA data, as post-transcriptional regulation is not required for the positioning of gap domains [12]. We can ignore morphological changes and focus on gene expression only, because there is no growth nor tissue arrangement during the blastoderm stage. We can homologize specific developmental stages between species, as dipteran embryos are morphologically very similar to each other [57, 119]. Moreover, we can compare the resulting dynamic patterns at high spatiotemporal resolution [19]. And we can validate our fitted models with genetic perturbation methods, such as RNAi, that establish independent empirical evidence on gap gene regulation [58, 120]. All of this renders the dipteran gap gene network exceptionally tractable as a paradigm system for metazoan patterning. The combination of experimental decomposition and *in silico* recomposition of the gap gene system was particularly successful in the phorid scuttle fly *Megaselia abdita* (Fig. 2).

### 5.1 The Scuttle Fly *Megaselia abdita*: Compensatory Evolution

Evidence from systematic RNAi knockdown experiments and fitted gene circuit models indicates that the qualitative structure of the gap gene network is highly conserved between *M. abdita* and *D. melanogaster* (Fig. 2) [25, 120]. Both species exhibit the same pattern of repressive interactions among gap genes, with strong double-negative feedback between complementary domains, and much weaker (and posteriorly biased) repression between gap genes with overlapping expression. All five basic regulatory mechanisms observed in *D. melanogaster* (see Fig. 3a–e) are also present in *M. abdita*. However, there are qualitative differences in the way maternal factors activate gap genes in either species, as well as quantitative differences in the strength of gap–gap cross-repressive interactions.

One experimentally observable difference between the two species is that there is no maternal expression of *cad* in *M. abdita*, while zygotic expression is present (Fig. 2) [93]. In addition, *bcd* mRNA shows a much broader anterior localization pattern than in *D. melanogaster* [62, 91, 92], indicating that the Bcd protein gradient extends further to the posterior in this species. Despite these expression differences, the activating role of both of these maternal factors in gap gene expression is conserved [121]. This results in a broader initial distribution of anterior gap domains (activated by Bcd), and a delayed activation of *gt* and *hb* in the posterior of the embryo (because there is no maternal Cad) [120]. In addition, the maternal Hb gradient does not contribute to A–P polarity in *M. abdita*, in contrast to *D. melanogaster* [121]. Moreover, experimental evidence from the syrphid hoverfly *Episyrphus balteatus*

shows no maternal expression of *hb* [63]. This suggests a qualitative rewiring of maternal inputs on the gap gene network in the evolutionary lineage leading up to *D. melanogaster* where maternal Hb acquired a role in setting up embryo polarity.

These differences in early maternal activation are later compensated by gap–gap cross-repression in *M. abdita*, such that the patterning output of the system—the position of gap domains at the onset of gastrulation—is nearly identical between the two species [120]. In particular, *M. abdita* shows a marked anterior shift of the posterior boundary of the anterior *hb* domain, a boundary which is stationary in *D. melanogaster* and was thought to be a stable organizing center in the central region of the embryo (see [25, 120] and references therein). We explain this difference in mechanistic terms through altered repressive inputs from Kr and Kni on *hb*. While these repressors act in a redundant manner in *D. melanogaster*, they are both required for correct positioning of the *hb* boundary in *M. abdita* [120]. Our models show that weak repression by Kr gradually diminishes the concentration of *hb* in the boundary region until *kni* can become expressed, which strongly represses *hb*, downregulating it in an irreversible manner (Fig. 3g) [25]. In summary, the *hb* boundary is controlled by a multi-stable switch in *D. melanogaster* and by a molecular ratchet in *M. abdita*.

Another difference is that the abdominal domain of *kni* and the posterior domain of *gt* form later and in a more posterior position in *M. abdita* than in *D. melanogaster*. This delayed onset of expression is caused by the absence of maternal Cad (Fig. 2). As a consequence, both domains initiate their anterior shifts with a delay, which they compensate by a higher shift velocity at later stages [120]. The main regulatory difference between the two species lies in the effect of Gt on *hb*: in *M. abdita*, Gt weakly activates *hb*, setting the pace for the early buildup of Hb protein in the posterior of the embryo. This buildup is driven predominantly by strong *hb* auto-activation in *D. melanogaster*, where it leads to a gradual accumulation of Hb protein in the posterior of the embryo, repressing *gt* and causing its smooth shift toward the anterior. In contrast, *hb* auto-activation is much weaker in *M. abdita*, and is only triggered at a later stage, once enough Hb protein has accumulated. This provides a mechanistic explanation for the accelerated domain shift at late stages. We termed this the “pull-and-trigger” mechanism for the observed biphasic shifts in *M. abdita* (Fig. 3h) [25].

An important point to make here is that higher-level insights are rarely possible in biology without careful consideration of (mechanistic) detail. Understanding compensatory evolution in regulatory networks requires a detailed understanding of the underlying mechanisms. Only in this way can we arrive at lineage explanations—plausible scenarios for sequences of regulatory changes [16]—that explain the evolution of a gene network. Such explanations are the foundation for a broader understanding of the principles of network evolution. Before we discuss such general insights, we review our work in the psychodid moth midge *C. albipunctata*.

## 5.2 The Moth Midge *Clogmia albipunctata*: Shifts as Dynamic Fossils?

For several reasons, our reverse engineering approach was less successful in *C. albipunctata*. First of all, it was difficult to establish an efficient protocol for RNAi knockdown in this species, which hindered systematic model validation. Second, *C. albipunctata* has no maternal Bcd gradient. Instead, it uses a maternal gradient of the pair-rule gene product Odd-paired (Opa) as its anterior determinant [122]. Since this was unknown at the time, we assumed an unknown anterior gradient to be present. Finally, we strongly suspect that we miss at least one regulatory factor involved in gap gene regulation in *C. albipunctata*: Gt protein is not detectable in the posterior of the embryo [24, 37, 53], and a candidate gene that is expressed in the relevant region shows no gap-like phenotype upon knockdown [58]. Therefore, a region in the posterior of the embryo lacks an expression domain at the late blastoderm stage (Fig. 2). Missing regulators are a severe problem for reverse engineering, since they can lead to defects and artifacts in models fitted to incomplete data. Together, these obstacles impeded us from extracting a consensus network. Instead, we end up with four candidate networks for *C. albipunctata* that remain to be tested against experimental evidence (Fig. 3i) [24]. This limits any detailed and robust conclusions about the mechanisms of gap gene regulation in this species, though our analysis still allowed

us to gain some general insights into the evolution of long-germband segmentation.

One such insight is drawn directly from our semiquantitative gene expression data: not only posterior gap genes but also *Kr* and *hb* show extensive positional shifts toward the anterior, which are more pronounced than the shifts observed in *D. melanogaster* and *M. abdita* (Fig. 3j) [24, 37, 53]. Gap gene patterning in *C. albipunctata* is very dynamic. Interestingly, shift mechanisms are based on weak repression with posterior bias between overlapping gap domains (where we can resolve them with our models) [24]. In this sense, they are similar in *C. albipunctata* and *M. abdita*.

Another remarkable feature of gap gene regulation in *C. albipunctata* concerns the nature of maternal regulation: our models indicate that *Cad* acts as a repressor of *gt*, which explains the absence of its posterior domain [24]. The absence of a posterior *gt* domain, in turn, relaxes the requirement for mutual repression between *Kr* and *gt*, as their patterns are no longer complementary. In fact, *gt* may not even be a proper gap gene in this species as it lacks a gap phenotype upon RNAi knockdown [58]. This is probably connected to the presence of additional, yet-to-be-identified factors involved in gap gene regulation in *C. albipunctata*, but the precise role and importance of such gene recruitment and loss for gap gene network evolution remain to be determined once the missing regulators are identified.

The observation that a basally branching dipteran shows more pronounced gap domain shifts supports the view that shifts are an ancestral feature. Still, the functional significance of these shifts for segment patterning remains unclear. They may add robustness to the process [66]. In addition, they are required for the correct placement of pair-rule expression stripes—the primary regulatory targets of the gap gene system—in *D. melanogaster* [20]. Ultimately, however, their presence may reflect evolutionary inertia more than functional conservation. Gap domain shifts—and the oscillator-based mechanisms that generate them [103, 104]—may represent a *dynamic fossil*, a remnant of the ancestral short-germband mechanism of segment determination. This hypothesis is supported by the fact that the damped oscillatory mechanisms that produce domain shifts in the gap gene system are functionally similar to the limit-cycle oscillators of sequential short-germband segmentation [20, 86, 105]. But the evidence remains preliminary. Future studies based on a larger sample of evolutionary lineages, are required to resolve this issue.

### 5.3 Evolving Mechanisms and Network Drift

What general insights into network evolution do we gain from our comparative analysis of the gap gene system? Our first conclusion involves the evolution of the five basic gap gene regulatory mechanisms. We have shown that these core principles are largely conserved across dipteran lineages: we find broad activation by maternal factors, strong mutual repression between complementary gap genes, shifts through weak repression with a posterior bias among overlapping gap domains, gap gene auto-activation, and repression by terminal gap genes in all species we have examined [24, 25, 120, 121]. However, none of these mechanisms are perfectly conserved and some evolved more rapidly than others. In particular, maternal inputs are more variable than any of the regulatory mechanisms specific to gap genes. This is consistent with an hourglass pattern of developmental variation in dipteran insects, where patterning becomes increasingly more canalized as we progress downward through the layers of the segmentation gene network [30, 60, 66, 73, 77, 85, 88, 89, 120].

The degree of conservation also varies among mechanisms at the level of gap–gap cross-regulation. This is because mutations in different mechanisms have different consequences. The core repressive feedback loops between complementary gap domains are particularly conserved. They are only partially disrupted in *C. albipunctata*, where additional regulatory factors may have acquired their function. In contrast, our work in *M. abdita* shows how alterations in maternal inputs can be buffered by alterations in the timing and extent of gap domain shifts, implicating the weak interactions between overlapping domains [25, 120, 121]. A similar explanation likely applies to the pronounced shifts observed in *C. albipunctata* with its distinct set of maternal regulators [24, 122]. Taken together, this suggests that compensatory evolution is an important factor driving changes in gap domain shifts in dipterans.

More broadly speaking, we are providing a detailed mechanistic explanation for regulatory network

evolution by developmental system drift, also called phenogenetic drift [39, 40, 72, 100, 113, 114]. In this mode of network evolution, regulatory interactions change without affecting the overall output of the system. In this way, regulatory systems can evolve along so-called genotype networks, which are meta-networks—networks of networks that produce the same phenotypic outcome, connected to each other through single mutational steps [110]. Our analysis reveals that evolution of the gap gene network occurs along such a genotype (meta-)network, predominantly driven by mutations that fine-tune the strength of regulatory interactions influencing the extent and velocity of gap domain shifts [25].

We emphasize that the developmental system drift we observe in the gap gene system is based on quantitative changes in interaction *strengths* [120]. This stands in stark contrast to earlier accounts of system drift, which focus on qualitative changes that alter the components, and the number or kind of interactions between them (see, e.g., [100, 110, 113]). Considering the highly redundant nature of eukaryotic transcriptional regulation and the high rate of evolutionary turnover for transcription factor binding sites, we expect quantitative changes in regulatory interactions to be much more widespread than qualitative changes in a system [120]. In addition, quantitative changes tend to be less disruptive than qualitative ones and are thus less likely to be eliminated by selection. For these reasons, we predict that quantitative developmental system drift is a common mode of regulatory network evolution.

## 6 Outlook: *in silico* Evolution

We have presented a significant advance toward a lineage explanation—a sequence of mutational transformations—for evolution in dipteran insects [16]. Our new insights into system drift suggest possible evolutionary paths and transitions between lineages. However, our analysis does not provide any information about the actual sequence of regulatory changes. The main question is if there are many possible evolutionary paths, or if epistatic effects in the network constrain them to a few specific sequences of changes. Of course, the actual evolutionary pathway of the gap gene network can be empirically resolved through additional sampling of dipteran lineages. This requires reverse engineering the gap gene network in many other species, which will cost a large amount of time and effort. Meanwhile, we can test possible constraints on the sequence of transitions through *in silico* evolution.

*In silico* evolution is a simulation-based approach that models the evolutionary process (for reviews and example studies, see [7, 33, 34, 107, 108]; see also the chapters by Beslon et al. and by Hogeweg). It comes in many variations, yet the general idea is as follows: we simulate populations of “digital organisms” that have a genotype, which is translated into a phenotype via a G–P map. Each phenotype is assigned a fitness value based on a predefined fitness function, which is used to select individuals to create offspring. Reproduction consists in copying an individual with a certain rate of mutational changes to its genotype. In this way, populations evolve. The whole procedure is similar to optimization by evolutionary computation (see section 2), except that we are less interested in the final outcome of the simulation and more in *how* the population has evolved: which evolutionary trajectories are taken? What kind of network structures are changed in the process? What intermediate stages occur? In the case of the gap gene system, the aim is to take the gap gene networks of *C. albipunctata* or *M. abdita* as initial conditions, and to trace their evolution back to the common ancestor and along the phylogenetic tree to *D. melanogaster*.

In contrast to comparative empirical studies, the *in silico* approach provides several advantages for the study of evolutionary dynamics. It enables us to rapidly test and explore many scenarios, to modify any of the simulation parameters (e.g., mutational operators, rates, population sizes, etc.), and to trace parent–offspring relationships, thus creating a perfect “fossil record” [7]. Especially the latter is important, as it gives us the exact sequence of mutations and their precise phenotypic effects along the lineage that leads to the fittest individual(s). Thus by running many simulations and by searching for parallel mutational trajectories, we can separate contingency from general trends in the evolutionary process [7].

We have not yet carried out a definitive set of simulations of gap gene evolution. One reason is that the simulation setup must be designed to be robust against many of our necessarily ad hoc modeling



choices. Not only is the problem underdetermined in terms of which network modeling formalism to use but also concerning population size and structure and the kind of mutations that are implemented (and at what rate). One example is whether to simulate gene duplication, deletion, and recruitment, since it is currently unknown if such events occurred in the evolution of the gap gene system (see section 5). Another issue concerns the use of gene circuits: these models feed all regulatory inputs to a gene into a single sigmoid function, which means that if they are sensitive to changes in one input, they will be sensitive to all. This is clearly not an accurate representation of transcriptional regulation, which requires individual thresholds for each regulatory interaction. Hence, the latter type of model is usually preferred for *in silico* evolutionary studies [35, 36, 97, 98].

Another sensitive choice is what fitness function to use, since this strongly influences if a simulation gets stuck at a local fitness peak (our unpublished results). Residual functions measuring squared differences between model and data may be useful for model optimization, but they may overconstrain the evolutionary process. Indeed, the actual selective pressure(s) on the gap gene system may be less stringent than getting the exact spatiotemporal expression pattern right. Though we do not know what is selected for, distinct dipteran lineages show significant differences in the dynamics of gap gene expression. Thus the most relevant feature may be a regular and ordered set of segment polarity stripes by the onset of gastrulation. For this reason, we currently experiment with an indirect approach, inspired by a recent *in silico* evolution study that explores how *D. melanogaster* and the nematoceran malaria mosquito, *Anopheles gambiae*, swapped the order of their posterior *gt* and *hb* domains during evolution (Fig. 2) [82]. Instead of defining fitness directly on gap gene expression, they define it at the next level of the segmentation gene hierarchy, basing it on the number and position of pair-rule expression stripes. In other words, fitness is defined by the output of the gap genes.

Given the above setup, we aim to answer two questions. The first concerns compensatory evolution of gap domain shifts in response to altered maternal inputs between *M. abdita* and *D. melanogaster*. The second focuses on the gain/loss of the posterior *gt* and *hb* expression domains between *C. albipunctata* and *D. melanogaster* (Fig. 2). For both questions, we take advantage of the perfect “fossil record” provided by the simulations to compare possible mutational trajectories, and to search for trends or fixed sequences of mutations that lead to changes in gap gene expression across simulations. If such trends exist, they indicate the presence of epistatic constraints, which lead to a small set of key mutations being crucial for the transition from the initial to the target phenotype. On top of that, our simulations would also provide the epistatic network *context* these mutations require to exert their effects. To the best of our knowledge, this would be the first such study where evolutionary simulations are explicitly constrained by expression data.

Such simulations open up the exciting prospect of analyzing the evolution of a developmental gene regulatory network at the level of its phase space structure. We could trace changes in gene expression dynamics to the bifurcations that cause them. Such a study would actualize the pioneering insights of C. H. Waddington, René Thom, and the process structuralists, who proposed decades ago that the best way to understand the evolution of organic form, and the systems that generate it, is to understand how these systems move through the space of possible configurations [2, 38, 70, 99, 109]. We have never been as close to realizing this vision as we are now.

## 7 Conclusion

This chapter provides a progress report, following up on the research program we proposed in [46]. We have demonstrated its usefulness and its potential for evolutionary systems biology. It is an approach based on reverse engineering and *in silico* evolution, which strives to strike a compromise between the accuracy and rigor of forward modeling and the generality of the ensemble approach. We have discussed our accomplishments, the challenges overcome, and those we still have to meet. We focused on the methodological bottlenecks of data acquisition and model fitting, on issues of network decomposition, and on our comparative analysis to understand the mechanisms underlying compensatory network drift in the dipteran gap gene system. This work is complemented by the account of dynamical network

modularity presented in the chapter by Jaeger and Monk. Together with [45], these chapters present a comprehensive review of our efforts to understand the function and evolution of the gap gene network across dipteran lineages.

Despite a number of open challenges—such as establishing a robust set up for *in silico* evolution—we are optimistic about the prospects of our proposed approach. It enables us to elevate our piecemeal understanding of genetic regulatory mechanisms to a more integrated view of network evolution at the systems level. We have demonstrated that we can use dynamical models to recompose the orchestrated behavior of whole evolving regulatory networks and that we can do this in a rigorous, detailed, and empirically grounded way. Such recomposition is essential if we are looking for a mechanistic understanding of regulatory network evolution. The general idea is to accumulate more case studies of this kind—in different species and different developmental contexts. Ultimately, the ambition is to reveal the regularities, or even principles, underlying the evolution of regulatory networks. Here, we have documented a small but important first step on this fascinating journey.

**Acknowledgements** It is nearly impossible to thank all the people whose work contributed to this review. We would like to thank all members of the Jaeger lab at the CRG in Barcelona, especially Karl Wotton, Eva Jiménez-Guri, Berta Verd, Hilde Janssens, Astrid Hörmann, Damjan Cicin-Sain, and Kolja Becker for essential contributions to the project. JJ would like to thank John Reinitz and David H. Sharp, who got it all started; Michael Akam and Urs Schmidt-Ott and members of their labs for the help and support required to get the comparative work off the ground; James Sharpe for mentorship and friendship, and Nick Monk for our wonderful long-term collaboration. AC would like to thank the Sharpe lab, the EMBL/CRG Systems Biology Research Unit, and colleagues at the CRG for their support, insightful discussions, and an extraordinary postdoc experience.

## References

- [1] A. M. Abdol, D. Cicin-Sain, J. A. Kaandorp, and A. Crombach. Scatter search applied to the inference of a development gene network. *Computation*, 5(2):22, June 2017.
- [2] P. Alberch. From genes to phenotype: dynamical systems and evolvability. *Genetica*, 84(1):5–11, 1991.
- [3] M. Ashyraliyev, J. Jaeger, and J. G. Blom. Parameter estimation and determinability analysis applied to *Drosophila* gap gene circuits. *BMC Syst. Biol.*, 2:83, Sept. 2008.
- [4] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom. Systems biology: parameter estimation for biochemical models. *FEBS J.*, 276(4):886–902, Feb. 2009.
- [5] M. Ashyraliyev, K. Siggens, H. Janssens, J. Blom, M. Akam, and J. Jaeger. Gene circuit analysis of the terminal gap gene *huckebein*. *PLoS Comput. Biol.*, 5(10):e1000548, Oct. 2009.
- [6] R. Bailleul, C. Curantz, C. Desmarquet-Trin Dinh, M. Hidalgo, J. Touboul, and M. Manceau. Symmetry breaking in the embryonic skin triggers directional and sequential plumage patterning. *PLoS Biol.*, 17(10):e3000448, Oct. 2019.
- [7] B. Batut, D. P. Parsons, S. Fischer, G. Beslon, and C. Knibbe. *In silico* experimental evolution: a tool to test evolutionary scenarios. *BMC Bioinform.*, 14 Suppl 15:S11, 2013.
- [8] W. Bechtel. Mechanism and biological explanation. *Phil. Sci.*, 78:533–557, 2011.
- [9] W. Bechtel. Understanding endogenously active mechanisms: a scientific and philosophical challenge. *Eur. J. Phil. Sci.*, 2:233–248, 2012.

- [10] W. Bechtel and A. Abrahamsen. Explanation: a mechanist alternative. *Stud. Hist. Phil. Biol. & Biomed. Sci.*, 36:421–441, 2005.
- [11] W. Bechtel and A. Abrahamsen. Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Stud. Hist. Phil. Sci.*, 41:321–333, 2010.
- [12] K. Becker, E. Balsa-Canto, D. Cicin-Sain, A. Hoermann, H. Janssens, J. R. Banga, and J. Jaeger. Reverse-engineering post-transcriptional regulation of gap genes in *Drosophila melanogaster*. *PLoS Comput. Biol.*, 9(10):e1003281, Oct. 2013.
- [13] I. Brigandt. Systems biology and the integration of mechanistic explanation and mathematical explanation. *Stud. Hist. Phil. Biol. & Biomed. Sci.*, 44:477–492, 2013.
- [14] I. Brigandt. Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P.-A. Braillard and C. Malaterre, editors, *Explanation in Biology*, pages 135–173. Springer, Dordrecht, 2015.
- [15] J. Burns. The synthetic problem and the genotype-phenotype relation in cellular metabolism. In C. H. Waddington, editor, *Towards a Theoretical Biology*, volume III, pages 47–51. Edinburgh University Press, London, 1970.
- [16] B. Calcott. Lineage explanations: explaining how biological mechanisms change. *Br. J. Philos. Sci.*, 60(1):51–78, Mar. 2009. ISSN 0007-0882.
- [17] A. D. Chipman. The evolution of the gene regulatory networks patterning the *Drosophila* blastoderm. *Current Topics in Developmental Biology*, 139:297–324, 2020.
- [18] K.-W. Chu, Y. Deng, and J. Reinitz. Parallel simulated annealing by mixing of states. *J. Comput. Phys.*, 148(2):646–662, Jan. 1999.
- [19] D. Cicin-Sain, A. H. Pulido, A. Crombach, K. R. Wotton, E. Jiménez-Guri, J.-F. Taly, G. Roma, and J. Jaeger. SuperFly: a comparative database for quantified spatio-temporal gene expression patterns in early dipteran embryos. *Nucleic Acids Res.*, 43(Database issue):D751–755, Jan. 2015.
- [20] E. Clark. Dynamic patterning by the *Drosophila* pair-rule network reconciles long-germ and short-germ segmentation. *PLoS Biol.*, 15(9):e2002439, Sept. 2017.
- [21] E. Clark, A. D. Peel, and M. Akam. Arthropod segmentation. *Development*, 146(18), Sept. 2019.
- [22] A. Crombach, D. Cicin-Sain, K. R. Wotton, and J. Jaeger. Medium-throughput processing of whole mount in situ hybridisation experiments into gene expression domains. *PLoS ONE*, 7(9):e46658, 2012.
- [23] A. Crombach, K. R. Wotton, D. Cicin-Sain, M. Ashyraliyev, and J. Jaeger. Efficient reverse-engineering of a developmental gene regulatory network. *PLoS Comput. Biol.*, 8(7):e1002589, 2012.
- [24] A. Crombach, M. A. García-Solache, and J. Jaeger. Evolution of early development in dipterans: reverse-engineering the gap gene network in the moth midge *Clogmia albipunctata* (Psychodidae). *BioSystems*, 123:74–85, Sept. 2014.
- [25] A. Crombach, K. R. Wotton, E. Jiménez-Guri, and J. Jaeger. Gap gene regulatory dynamics evolve along a genotype network. *Mol. Biol. Evol.*, 33(5):1293–1307, 2016.
- [26] E. H. Davidson and D. H. Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311(5762):796–800, Feb. 2006.

- [27] G. K. Davis and N. H. Patel. Short, long, and beyond: molecular and embryological approaches to insect segmentation. *Annu. Rev. Entomol.*, 47:669–699, 2002.
- [28] J. DiFrisco and J. Jaeger. Beyond networks: mechanism and process in evo-devo. *Biology & Philosophy*, 34(6):54, Nov. 2019.
- [29] J. DiFrisco and J. Jaeger. Genetic causation in complex regulatory systems: an integrative dynamic perspective. *BioEssays*, 42(6):1900226, 2020.
- [30] D. Duboule. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development*, 1994(Supplement):135–142, Jan. 1994. ISSN 0950-1991, 1477-9129. Publisher: The Company of Biologists Ltd.
- [31] D. A. Fehr, J. E. Handzlik, Manu, and Y. L. Loh. Classification-based inference of dynamical models of gene regulatory networks. *G3 (Bethesda)*, 9(12):4183–4195, 2019.
- [32] Y. Fomekong-Nanfack, J. A. Kaandorp, and J. Blom. Efficient parameter estimation for spatio-temporal models of pattern formation: case study of *Drosophila melanogaster*. *Bioinformatics*, 23(24):3356–3363, Dec. 2007.
- [33] P. François. Evolution *in silico*: from network structure to bifurcation theory. *Adv. Exp. Med. Biol.*, 751:157–182, 2012. ISSN 0065-2598.
- [34] P. François. Evolving phenotypic networks *in silico*. *Semin. Cell Dev. Biol.*, 35:90–97, Nov. 2014. ISSN 1096-3634.
- [35] P. François and E. D. Siggia. Predicting embryonic patterning using mutual entropy fitness and *in silico* evolution. *Development*, 137(14):2385–2395, July 2010. ISSN 1477-9129.
- [36] P. François, V. Hakim, and E. D. Siggia. Deriving structure from evolution: metazoan segmentation. *Mol. Syst. Biol.*, 3:154, 2007. ISSN 1744-4292.
- [37] M. García-Solache, J. Jaeger, and M. Akam. A systematic analysis of the gap gene system in the moth midge *Clogmia albipunctata*. *Dev. Biol.*, 344(1):306–318, Aug. 2010.
- [38] B. C. Goodwin. Development and evolution. *J. Theor. Biol.*, 97(1):43–55, July 1982. ISSN 0022-5193.
- [39] E. S. Haag. Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. *Genetica*, 129(1):45–55, Jan. 2007.
- [40] E. S. Haag and J. R. True. Developmental system drift. In L. Nuño de la Rosa and G. B. Müller, editors, *Evolutionary Developmental Biology*. Springer, Dordrecht, 2018.
- [41] P. F. Hagolani, R. Zimm, M. Marin-Riera, and I. Salazar-Ciudad. Cell signaling stabilizes morphogenesis against noise. *Development*, 146(20), Oct. 2019.
- [42] A. Hoermann, D. Cicin-Sain, and J. Jaeger. A quantitative validated model reveals two phases of transcriptional regulation for the gap gene *giant* in *Drosophila*. *Dev. Biol.*, 411(2):325–338, Mar. 2016.
- [43] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.A.*, 81(10):3088–3092, May 1984.
- [44] J. Jaeger. The gap gene network. *Cell. Mol. Life Sci.*, 68(2):243–274, Jan. 2011.

- [45] J. Jaeger. Shift happens: the developmental and evolutionary dynamics of the gap gene system. *Curr. Opin. Syst. Biol.*, 11:65–73, Oct. 2018.
- [46] J. Jaeger and A. Crombach. Life's attractors : understanding developmental systems through reverse engineering and *in silico* evolution. *Adv. Exp. Med. Biol.*, 751:93–119, 2012.
- [47] J. Jaeger and J. Reinitz. On the dynamic nature of positional information. *BioEssays*, 28(11):1102–1111, 2006.
- [48] J. Jaeger and B. Verd. Dynamic positional information: patterning mechanism versus precision in gradient-driven systems. In S. Small and J. Briscoe, editors, *Current Topics in Developmental Biology*, volume 137 of *Gradients and Tissue Patterning*, pages 219–246. Academic Press, Jan. 2020.
- [49] J. Jaeger, M. Blagov, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, S. Surkova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, 167(4):1721–1737, Aug. 2004.
- [50] J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz. Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, 430(6997):368–371, July 2004.
- [51] J. Jaeger, D. H. Sharp, and J. Reinitz. Known maternal gradients are not sufficient for the establishment of gap domains in *Drosophila melanogaster*. *Mech. Dev.*, 124(2):108–128, Feb. 2007.
- [52] J. Jaeger, D. Irons, and N. Monk. Regulative feedback in pattern formation: towards a general relativistic theory of positional information. *Development*, 135(19):3175–3183, Oct. 2008.
- [53] H. Janssens, K. Siggins, D. Cicin-Sain, E. Jiménez-Guri, M. Musy, M. Akam, and J. Jaeger. A quantitative atlas of Even-skipped and Hunchback expression in *Clogmia albipunctata* (Diptera: Psychodidae) blastoderm embryos. *EvoDevo*, 5(1):1, Jan. 2014.
- [54] A. Jiménez, J. Cotterell, A. Munteanu, and J. Sharpe. Dynamics of gene circuits shapes evolvability. *Proc. Natl. Acad. Sci. U.S.A.*, 112(7):2103–2108, Feb. 2015.
- [55] A. Jiménez, J. Cotterell, A. Munteanu, and J. Sharpe. A spectrum of modularity in multi-functional gene circuits. *Mol. Syst. Biol.*, 13(4):925, Apr. 2017.
- [56] E. Jiménez-Guri, J. Huerta-Cepas, L. Cozzuto, K. R. Wotton, H. Kang, H. Himmelbauer, G. Roma, T. Gabaldón, and J. Jaeger. Comparative transcriptomics of early dipteran development. *BMC Genomics*, 14:123, Feb. 2013.
- [57] E. Jiménez-Guri, K. R. Wotton, B. Gavilán, and J. Jaeger. A staging scheme for the development of the moth midge *Clogmia albipunctata*. *PLoS ONE*, 9(1):e84422, 2014.
- [58] E. Jiménez-Guri, K. R. Wotton, and J. Jaeger. *tarsal-less* is expressed as a gap gene but has no gap gene phenotype in the moth midge *Clogmia albipunctata*. *R. Soc. Open Sci.*, 5(8):180458, Aug. 2018.
- [59] L. Jostins and J. Jaeger. Reverse engineering a gene network using an asynchronous parallel evolution strategy. *BMC Syst. Biol.*, 4(1):17, Mar. 2010.
- [60] A. T. Kalinka, K. M. Varga, D. T. Gerrard, S. Preibisch, D. L. Corcoran, J. Jarrells, U. Ohler, C. M. Bergman, and P. Tomancak. Gene expression divergence recapitulates the developmental hourglass model. *Nature*, 468(7325):811–814, Dec. 2010.
- [61] H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, Nov. 2002.

- [62] S. Lemke, M. Stauber, P. J. Shaw, A. M. Rafiqi, A. Prell, and U. Schmidt-Ott. Bicoid occurrence and Bicoid-dependent *hunchback* regulation in lower cyclorrhaphan flies. *Evol. Dev.*, 10(4):413–420, Aug. 2008.
- [63] S. Lemke, S. E. Busch, D. A. Antonopoulos, F. Meyer, M. H. Domanus, and U. Schmidt-Ott. Maternal activation of gap genes in the hover fly *Episyrphus*. *Development*, 137(10):1709–1719, May 2010.
- [64] Z. Lou and J. Reinitz. Parallel simulated annealing using an adaptive resampling interval. *Parallel Computing*, 53:23–31, Apr. 2016.
- [65] J. A. Lynch, A. E. Brent, D. S. Leaf, M. A. Pultz, and C. Desplan. Localized maternal *orthodenticle* patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature*, 439(7077):728–732, Feb. 2006.
- [66] Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A.-R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Comput. Biol.*, 5(3):e1000303, Mar. 2009.
- [67] Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A.-R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz. Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. *PLoS Biol.*, 7(3):e1000049, Mar. 2009.
- [68] E. Mjolsness, D. H. Sharp, and J. Reinitz. A connectionist model of development. *J. Theor. Biol.*, 152(4):429–453, Oct. 1991.
- [69] E. C. Olesnick, A. E. Brent, L. Tonnes, M. Walker, M. A. Pultz, D. Leaf, and C. Desplan. A *caudal* mRNA gradient controls posterior development in the wasp *Nasonia*. *Development*, 133(20):3973–3982, Oct. 2006.
- [70] G. Oster and P. Alberch. Evolution and bifurcation of developmental programs. *Evolution*, 36(3):444–459, May 1982. ISSN 1558-5646.
- [71] J. Panovska-Griffiths, K. M. Page, and J. Briscoe. A gene regulatory motif that generates oscillatory or multiway switch outputs. *J. R. Soc. Interface*, 10(79):20120826, Feb. 2013. ISSN 1742-5662.
- [72] M. Pavlicev and G. P. Wagner. A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol. Evol.*, 27(6):316–322, June 2012.
- [73] M. L. Perkins. Dynamic positioning and precision of bistable gene expression boundaries through diffusion and morphogen decay. *bioRxiv*, page 2020.12.17.423202, Dec. 2020.
- [74] T. J. Perkins, J. Jaeger, J. Reinitz, and L. Glass. Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput. Biol.*, 2(5):e51, May 2006.
- [75] M. Pigliucci. Genotype-phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philos. Trans. R. Soc. B*, 365(1540):557–566, Feb. 2010.
- [76] M. A. Pultz, L. Westendorf, S. D. Gale, K. Hawkins, J. Lynch, J. N. Pitt, N. L. Reeves, J. C. Y. Yao, S. Small, C. Desplan, and D. S. Leaf. A major role for zygotic *hunchback* in patterning the *Nasonia* embryo. *Development*, 132(16):3705–3715, Aug. 2005.
- [77] R. A. Raff. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. University of Chicago Press, Chicago (IL), 1996.

- [78] J. Raspopovic, L. Marcon, L. Russo, and J. Sharpe. Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science*, 345(6196):566–570, Aug. 2014. ISSN 1095-9203.
- [79] J. Reinitz and D. H. Sharp. Mechanism of eve stripe formation. *Mech. Dev.*, 49(1-2):133–158, Jan. 1995.
- [80] K. B. Rohr, D. Tautz, and K. Sander. Segmentation gene expression in the mothmidge *Clogmia albipunctata* (Diptera, psychodidae) and other primitive dipterans. *Dev. Genes Evol.*, 209(3):145–154, Mar. 1999.
- [81] M. I. Rosenberg, J. A. Lynch, and C. Desplan. Heads and tails: evolution of antero-posterior patterning in insects. *Biochimica Et Biophysica Acta*, 1789(4):333–342, Apr. 2009.
- [82] J. B. Rothschild, P. Tsimiklis, E. D. Siggia, and P. François. Predicting ancestral segmentation phenotypes from *Drosophila* to *Anopheles* using *in silico* evolution. *PLoS Genet.*, 12(5):e1006052, 2016.
- [83] A. Sadier, M. Twarogowska, K. Steklikova, L. Hayden, A. Lambert, P. Schneider, V. Laudet, M. Hovorakova, V. Calvez, and S. Pantalacci. Modeling *Edar* expression reveals the hidden dynamics of tooth signaling center patterning. *PLoS Biol.*, 17(2):e3000064, Feb. 2019.
- [84] I. Salazar-Ciudad. Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development. *J. Exp. Zool. B Mol. Dev. Evol.*, 306(2):107–125, Mar. 2006.
- [85] K. Sander. *Development and Evolution*. Cambridge University Press, Cambridge, 1983.
- [86] A. F. Sarrazin, A. D. Peel, and M. Averof. A segmentation clock with two-segment periodicity in insects. *Science*, 336(6079):338–341, Apr. 2012.
- [87] U. Schmidt-Ott, A. M. Rafiqi, and S. Lemke. Hox3/zen and the evolution of extraembryonic epithelia in insects. *Advances in Experimental Medicine and Biology*, 689:133–144, 2010.
- [88] F. Seidel. Körpergrundgestalt und Keimstruktur. Eine Erörterung über die Grundlagen der vergleichenden und experimentellen Embryologie und deren Gültigkeit bei phylogenetischen Überlegungen. *Zoologischer Anzeiger*, 164:245–305, 1960.
- [89] J. M. Slack, P. W. Holland, and C. F. Graham. The zootype and the phylotypic stage. *Nature*, 361(6412):490–492, Feb. 1993. ISSN 0028-0836.
- [90] O. S. Soyer and M. A. O'Malley. Evolutionary systems biology: what it is and why it matters. *BioEssays*, 35(8):696–705, Aug. 2013.
- [91] M. Stauber, H. Jäckle, and U. Schmidt-Ott. The anterior determinant *bicoid* of *Drosophila* is a derived Hox class 3 gene. *Proc. Natl. Acad. Sci. U.S.A.*, 96(7):3786–3789, Mar. 1999.
- [92] M. Stauber, H. Taubert, and U. Schmidt-Ott. Function of *bicoid* and *hunchback* homologs in the basal cyclorrhaphan fly *Megaselia* (Phoridae). *Proc. Natl. Acad. Sci. U.S.A.*, 97(20):10844–10849, Sept. 2000.
- [93] M. Stauber, S. Lemke, and U. Schmidt-Ott. Expression and regulation of *caudal* in the lower cyclorrhaphan fly *Megaselia*. *Dev. Genes Evol.*, 218(2):81–87, Feb. 2008.
- [94] S. H. Strogatz. *Nonlinear Dynamics and Chaos; With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 2nd edition, 2015.

- [95] S. Surkova, D. Kosman, K. Kozlov, Manu, E. Myasnikova, A. A. Samsonova, A. Spirov, C. E. Vanario-Alonso, M. Samsonova, and J. Reinitz. Characterization of the *Drosophila* segment determination morphome. *Dev. Biol.*, 313(2):844–862, Jan. 2008.
- [96] S. Surkova, E. Myasnikova, H. Janssens, K. N. Kozlov, A. A. Samsonova, J. Reinitz, and M. Samsonova. Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images. *Fly*, 2(2):58–66, Apr. 2008.
- [97] K. H. W. J. ten Tusscher. Mechanisms and constraints shaping the evolution of body plan segmentation. *Eur. Phys. J. E*, 36(5):54, May 2013. ISSN 1292-895X.
- [98] K. H. W. J. ten Tusscher and P. Hogeweg. Evolution of networks for body plan patterning; interplay of modularity, robustness and evolvability. *PLoS Comput. Biol.*, 7(10):e1002208, Oct. 2011. ISSN 1553-7358.
- [99] R. Thom. *Structural Stability and Morphogenesis*. Advanced Books Classics Series. Perseus Books Group, 1976.
- [100] J. R. True and E. S. Haag. Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.*, 3(2):109–119, Apr. 2001.
- [101] M. Uzkudun, L. Marcon, and J. Sharpe. Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. *Mol. Syst. Biol.*, 11(7):815, July 2015. ISSN 1744-4292.
- [102] B. Verd, A. Crombach, and J. Jaeger. Classification of transient behaviours in a time-dependent toggle switch model. *BMC Syst. Biol.*, 8:43, Apr. 2014.
- [103] B. Verd, A. Crombach, and J. Jaeger. Dynamic maternal gradients control timing and shift-rates for *Drosophila* gap gene expression. *PLoS Comput. Biol.*, 13(2):e1005285, 2017.
- [104] B. Verd, E. Clark, K. R. Wotton, H. Janssens, E. Jiménez-Guri, A. Crombach, and J. Jaeger. A damped oscillator imposes temporal order on posterior gap gene expression in *Drosophila*. *PLoS Biol.*, 16(2):e2003174, 2018.
- [105] B. Verd, N. A. Monk, and J. Jaeger. Modularity, criticality, and evolvability of a developmental gene regulatory network. *Elife*, 8, 2019.
- [106] R. M. A. Vroomans, P. Hogeweg, and K. H. W. J. ten Tusscher. Segment-specific adhesion as a driver of convergent extension. *PLoS Comput. Biol.*, 11(2):e1004092, Feb. 2015. ISSN 1553-7358.
- [107] R. M. A. Vroomans, P. Hogeweg, and K. H. W. J. ten Tusscher. *In silico* evo-devo: reconstructing stages in the evolution of animal segmentation. *EvoDevo*, 7:14, 2016. ISSN 2041-9139.
- [108] R. M. A. Vroomans, P. Hogeweg, and K. H. W. J. ten Tusscher. Around the clock: gradient shape and noise impact the evolution of oscillatory segmentation dynamics. *EvoDevo*, 9:24, 2018. ISSN 2041-9139.
- [109] C. H. Waddington. *The Strategy of the Genes*. George Allen & Unwin, London, 1957.
- [110] A. Wagner. *The Origins of Evolutionary Innovations*. Oxford University Press, July 2011.
- [111] G. P. Wagner. *Homology, Genes, and Evolutionary Innovation*. Princeton University Press, Apr. 2014. ISBN 978-0-691-15646-0.
- [112] G. P. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976, June 1996.
- [113] K. M. Weiss. The phenogenetic logic of life. *Nat. Rev. Genet.*, 6(1):36–45, Jan. 2005.



- [114] K. M. Weiss and S. M. Fullerton. Phenogenetic drift and the evolution of genotype-phenotype relationships. *Theor. Popul. Biol.*, 57(3):187–195, May 2000.
- [115] B. M. Wiegmann, M. D. Trautwein, I. S. Winkler, N. B. Barr, J.-W. Kim, C. Lambkin, M. A. Bertone, B. K. Cassel, K. M. Bayless, A. M. Heimberg, B. M. Wheeler, K. J. Peterson, T. Pape, B. J. Sinclair, J. H. Skevington, V. Blagoderov, J. Caravas, S. N. Kutty, U. Schmidt-Ott, G. E. Kampmeier, F. C. Thompson, D. A. Grimaldi, A. T. Beckenbach, G. W. Courtney, M. Friedrich, R. Meier, and D. K. Yeates. Episodic radiations in the fly tree of life. *Proc. Natl. Acad. Sci. U.S.A.*, 108(14):5690–5695, Apr. 2011.
- [116] M. J. Wilson and P. K. Dearden. Diversity in insect axis formation: two *orthodenticle* genes and *hunchback* act in anterior patterning and influence dorsoventral organization in the honeybee (*Apis mellifera*). *Development*, 138(16):3497–3507, Aug. 2011.
- [117] M. J. Wilson, M. Havler, and P. K. Dearden. *giant*, *krüppel*, and *caudal* act as gap genes with extensive roles in patterning the honeybee embryo. *Dev. Biol.*, 339(1):200–211, Mar. 2010.
- [118] W. C. Wimsatt. *Re-Engineering Philosophy for Limited Beings*. Harvard University Press, Cambridge, MA, 2007.
- [119] K. R. Wotton, E. Jiménez-Guri, B. García Matheu, and J. Jaeger. A staging scheme for the development of the scuttle fly *Megaselia abdita*. *PLoS ONE*, 9(1):e84421, 2014.
- [120] K. R. Wotton, E. Jiménez-Guri, A. Crombach, H. Janssens, A. Alcaine-Colet, S. Lemke, U. Schmidt-Ott, and J. Jaeger. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly *Megaselia abdita*. *Elife*, 4, Jan. 2015.
- [121] K. R. Wotton, E. Jiménez-Guri, and J. Jaeger. Maternal co-ordinate gene regulation and axis polarity in the scuttle fly *Megaselia abdita*. *PLoS Genet.*, 11(3):e1005042, Mar. 2015.
- [122] Y. Yoon, J. Klomp, I. Martin-Martin, F. Criscione, E. Calvo, J. Ribeiro, and U. Schmidt-Ott. Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. *Elife*, 8, 2019.
- [123] M. Zagorski, Y. Tabata, N. Brandenberg, M. P. Lutolf, G. Tkačik, T. Bollenbach, J. Briscoe, and A. Kicheva. Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science*, 356(6345):1379–1383, June 2017. ISSN 1095-9203.