The Effects of Gene Recruitment on the Evolvability and Robustness of Pattern-Forming Gene Networks

Chapter *in* Lecture Notes in Electrical Engineering · September 2008

DOI: 10.1007/978-1-4020-8919-0_3

2 authors:

Alexander V Spirov
Stony Brook University
93 PUBLICATIONS **1,009** CITATIONS
SEE PROFILE

David M Holloway
British Columbia Institute of Technology
106 PUBLICATIONS **512** CITATIONS
SEE PROFILE

Chapter #

THE EFFECTS OF GENE RECRUITMENT ON THE EVOLVABILITY AND ROBUSTNESS OF PATTERN-FORMING GENE NETWORKS

Alexander V. Spirov[1] and David M. Holloway[2]

[1]Applied Mathematics and Statistics, and Center for Developmental Genetics, State University of New York, CMM Bldg, Rm481, South Loop, SUNY at Stony Brook, Stony Brook, NY 11794-5140, USA (corresponding author to provide phone: 631-632-8221; fax: 631-632-1692; e-mail: Alexander.Spirov@sunysb.edu). [2]Mathematics Department, British Columbia Institute of Technology, Burnaby, B.C., Canada (e-mail: David_Holloway@bcit.ca), and with the Biology Department, University of Victoria, B.C., Canada.

Abstract: Gene recruitment or co-option is defined as the placement of a new gene under a foreign regulatory system. Such re-arrangement of pre-existing regulatory networks can lead to an increase in genomic complexity. This reorganization is recognized as a major driving force in evolution. We simulated the evolution of gene networks by means of the Genetic Algorithms (GA) technique. We used standard GA methods of point mutation and multi-point crossover, as well as our own operators for introducing or withdrawing new genes on the network. The starting point for our computer evolutionary experiments was a 4-gene dynamic model representing the real genetic network controlling segmentation in the fruit fly *Drosophila*. Model output was fit to experimentally observed gene expression patterns in the early fly embryo. We compared this to output for networks with more and less genes, and with variation in maternal regulatory input. We found that the mutation operator, together with the gene introduction procedure, was sufficient for recruiting new genes into pre-existing networks. Reinforcement of the evolutionary search by crossover operators facilitates this recruitment, but is not necessary. Gene recruitment causes outgrowth of an evolving network, resulting in redundancy, in the sense that the number of genes goes up, as well as the regulatory interactions on the original genes. The recruited genes can have uniform or patterned expressions, many of which recapitulate gene patterns seen in flies, including genes which are not explicitly put in our model. Recruitment of new genes can affect the evolvability of networks (in general, their ability to produce the variation to facilitate adaptive evolution). We see this in particular with a 2-gene subnetwork. To study robustness, we have subjected the networks to experimental levels of variability in maternal regulatory patterns. The majority of networks are not robust to these

perturbations. However, a significant subset of the networks do display very high robustness. Within these networks, we find a variety of outcomes, with independent control of different gene expression boundaries. Increase in the number and connectivity of genes (redundancy) does not appear to correlate with robustness. Indeed, removal of recruited genes tends to give a worse fit to data than the original network; new genes are not freely disposable once they acquire functions in the network.

Key Words: Complexification of gene networks, gene co-option, gene recruitment, pattern formation, modeling of biological evolution by Genetic Algorithms, redundancy and robustness of gene networks.

1. INTRODUCTION

Early in metazoan evolution, gene networks specifying developmental events in embryos may have consisted of no more than two or three interacting genes. Over time, these were augmented by incorporating new genes and integrating originally distinct pathways[1]. While it may initially be thought that new functions require novel genes, whole genome sequencing has shown that apparent increases in developmental complexity do not correlate with increasing numbers of genes[2]: the number of genes in the human genome is somewhat higher than in fruit flies and nematodes, but lower than in pufferfish and cress and rice plants. Therefore, evolution of developmental pathways may most commonly proceed by recruitment of preexisting external genes into preexisting networks, to create novel functions and novel developmental pathways[3]; developmental evolution may act primarily on genetic regulation[4, 5].

Specifically, gene recruitment may occur through mutational changes in the regulatory sequences of a gene in an established pathway, enabling a new transcriptional regulator (or regulators) to bind. This regulator may be from a newly evolved gene (say via duplication and subsequent change), in which case it simply adds to the existing pathway, or it may have already been part of a pre-existing pathway, in which case the two pathways become integrated. In either case, the developmental function of the pathway may be significantly altered. Similarly significant alterations can arise by inserting regulatory sequences for an existing gene at new loci, transferring transcriptional control of the original gene to other members of the genome[1, 6].

In insects, two distinct modes of segmenting the body have evolved. In primitive insects, such as the grasshopper, the short germ band mode lays out body segments sequentially. Many more highly derived insects, such as flies, use the long germ band mode to establish all body segments

simultaneously. This simultaneous mechanism must act quickly during development; it has been proposed that it evolved by co-option of new genes to the short germ band mechanism, in order to maintain accurate regulation of patterned gene transcription over the whole embryo in a condensed time frame 1. The invertebrate segmentation network is one of the best-studied gene ensembles, in which the amount of diverse experimental data provides a unique opportunity for studying known and hypothetical scenarios of its evolution in detail. In particular, the level of detail for the segmentation gene network for the fruit fly (*Drosophila melanogaster*) has made it for many years the most popular object for computer simulations of its function and evolution[7, 8, 9, 10, 11, 12].

In this publication, we investigate the interrelations between redundancy (addition of extra genes to a network), evolvability (ability of a network to change), and robustness (ability of a network to remain fit in a variable environment). We use an *in silico* approach to simulate evolution of a dynamic model of the gap gene network, central to fly segmentation (specifically). This model (adapted from[9, 14]) is a system of differential equations describing the regulatory interactions of 4 gap genes (*giant*, *gt*; *hunchback*, *hb*; *Krüppel*, *Kr*; *knirps*, *kni*), under the control of gradients of maternal proteins (Bicoid, Bcd, in our basic model; plus maternally-derived Hb (Hb_{mat}), Caudal (Cad), and Tailless (Tll) in our extended model). Fig. 1A shows the integrated (averaged) spatial patterns of the gap genes along the antero-posterior (A-P; head to tail) axis of the fly embryo in early nuclear cleavage cycle 14A (*even-skipped*, *eve*, is a pair-rule gene, regulated by the maternal and gap genes). Fig. 1B shows the gap patterns slightly later in development, at mid cleavage cycle 14A. Fig. 1C shows the patterns of the maternal input factors. Model parameters for gene interaction strengths are varied and solutions selected by a Genetic Algorithms method (details below) based on how well they fit the gap gene data. This produces networks describing particular interactions (and quantitative strengths) between the component genes (e.g., Fig. 1D). In this way, we can use a model of our current understanding of fly segmentation to study the evolutionary dynamics of how the segmentation network may have arisen, and how this might reflect on its current characteristics.

In particular, we are interested in what genetic mechanisms are necessary for recruiting (co-opting) new genes to small networks, what characteristics these recruits have (e.g., spatial patterns, regulatory interactions), and how they might change the behavior of the network. There is currently much discussion in evolutionary biology on these topics, and it is expected that the outgrowth of preexisting networks through gene recruitment should cause structural (genes duplicating existing ones) and functional (development of compensatory pathways) redundancy of the networks[13]. Cases of such

redundancy have been found in many genetic ensembles in many organisms[13]. One of the common conclusions from these cases is that the redundancy could affect such key species characteristics as evolvability or robustness to perturbations and variability during development.

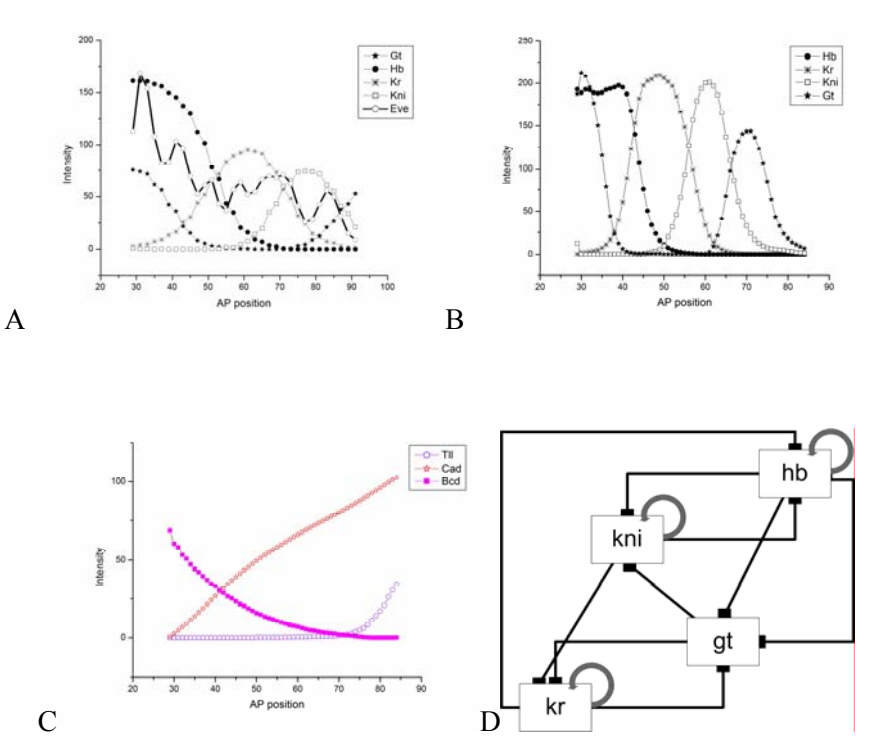


Figure 1. Biological data used to fit ODE model by GA.

A. Integrated gene expression profiles for early cleavage cycle 14A. Vertical axis represents relative protein concentrations (proportional to intensity), horizontal axis represents position along the anteroposterior (A-P) embryo axis (where 0% is the anterior pole). Data from the FlyEx database[26].

B. Integrated gene expression profiles for mid cleavage cycle 14A.

C. Integrated gene expression profiles for the external, maternal inputs used in this paper, from the very beginning of cleavage cycle 14A.

D. Overview of the gap gene network (After[14]). The gap genes are represented as boxes. Repressive interactions are represented by T-bar connectors. Looped arrows mean self-regulation.

The segmentation network lays down the spatial order of the developing embryo, so the fitness of any network depends on how reliably it establishes spatial position. In our computations, we establish this type of fitness by scoring model solutions on how well they reproduce experimental pattern. By doing hundreds of simulations, we generate a large sample of networks

for studying the mechanisms of gene recruitment and how these relate to evolvability and robustness (in particular for making reproducible output in the face of biological levels of variability in the upstream maternal control gradients).

We investigate the mechanisms of gene recruitment through the Genetic Algorithms (GA) technique. Run on our fly segmentation model, it is a simulation of how this network may have evolved in nature. We use standard GA operators (mutation and crossover), as well as our own operators for introducing and removing new genes on the networks.

In computing evolutionary searches, we have found that the standard operator for point mutations, in combination with the gene introduction operator, is enough to support recruitment of new genes to pre-existing networks. This is in contrast to a mainstream view in evolutionary biology, that the main mechanism facilitating recruitment is the sophisticated shuffling of genetic material, such as unequal crossover (recombination), or the activity of transposons[15]. A computational approach allows us to systematically compare recruitment by these different mechanisms, specifically point mutation versus one- and multi-point crossover.

Our results indicate how complexification or "outgrowth" of gene networks can proceed, by recruiting new genes to make new connections between old and new members of the network. We have characterized the structure of the evolved networks, as well as the possible influence of gene recruitment on evolvability. In particular, we found that for a 2-gene subnetwork evolvability is clearly raised via co-option of new genes. Evolvability was not so clearly raised when starting with the 4-gene networks.

We also studied the effects of gene recruitment on the robustness of the computer networks to variability in external control parameters. Specifically, we simulated variability in maternal morphogenetic factors which are upstream (in terms of regulatory control) of the simulated networks. The gap gene network has been shown to be quite robust to this sort of variability, spurring a great deal of interest in the biology community on how embryos might filter maternal or environmental variability or noise[16, 17, 18, 19, 20, 21]. We tested robustness in our basic 4-gene network (Bcd control only, which has been most extensively studied experimentally) and our extended 4-gene network (additional control by $Hb_{maternal}$, Cad and Tll). By simulating variability in each of the upstream factors we can see which computed solutions have experimentally observed levels of robustness, or better, and whether networks evolve with more robustness to particular factors. Computation allows us to understand the experimentally well-characterized factors, such as Bcd, and extend results to the other, less well-characterized maternal factors.

Analyzing several hundred high-scoring solutions of the three variants of our model, we found very diverse ways for the networks to solve the pattern fit, and these had quite different levels of robustness to variability in maternal factors. We did not, however, find a clear correlation between the types of new connections in the evolved networks and robustness. Many of the recruited genes are, however, spatially patterned like known *Drosophila* genes. These patterns can either be like those for members of our 4-gene model, or our evolutionary searches also recruit genes with patterns like real segmentation genes that aren't one of the original model genes.

2. METHODS AND APPROACHES

2.1 The segmentation gene network and its modeling

Four gap genes, *Kr*, *gt*, *kni* and *hb*, are the core elements in our segmentation model. In *Drosophila*, these are transcriptionally activated by the maternal Bcd protein gradient in a concentration dependent manner, a classic example of a morphogen as characterized by Wolpert[22]. Three other gradients, Hb_{mat}, Cad, and Tll, help determine the positions of the gap genes. The combination of this upstream specification and gap-gap cross-regulation results in sharp and precise gap patterns.

We model these genes (and proteins) and their interactions using the gene circuit framework[23, 24], to produce A-P concentration patterns (fitting data such as in Figs. 1A-B). The model is computed for a one-dimensional row of nuclei, between 30 and 94% A-P position (where 0% is the anterior pole) during nuclear cleavage cycles 13 and 14A. The gap gene proteins (Kr, Gt, Kni and Hb) are variables in the model, with the rates of change of their concentrations dv_i^a/dt (for each gene product a in each nucleus *i*) defining a system of *number of proteins* times *number of nuclei* ODEs (Ordinary Differential Equations) given by

$$\begin{aligned} dv_i^a/dt &= R_a g\left(u^a\right) + \\ &\quad D^a\left[\left(v_{i-1}^a - v_i^a\right) + \left(v_{i+1}^a - v_i^a\right)\right] - \lambda_a v_i^a . \end{aligned} \tag{1}$$

The main terms on the right hand side of (1) represent protein synthesis (R_a), diffusion (D^a) and decay (λ_a). $g\left(u^a\right)$ is a sigmoid regulation-expression function. For values u^a below -1.5 and above 1.5 $g\left(u^a\right)$ rapidly approaches zero. u^a is given by $u^a = \sum_b T^{ab} v_i^b + m^a v_i^{Bcd} + h^a$. Parameters T^{ab} constitute a genetic interconnectivity matrix, representing activation of gene a by the product of gene b (with concentration v_i^b) if positive, repression if negative, and no interaction if close to zero. v_i^{Bcd} represents the concentration of Bcd in nucleus *i*, which is constant in time. m^a describes the regulatory input of Bcd to each gene. Bcd is a general activator for all four gap genes considered. h^a represents regulatory input from ubiquitous factors. Our extended model includes Hb_{mat}, Cad, and Tll in a similar manner to Bcd, as time-independent parameters.

The full extended model involves heavy computation, which can greatly delay evolutionary searches and the generation of large samples of networks. For this reason, we have focused on reduced networks to study robustness and evolvability, either the basic 4-gene model (under Bcd control only), or a 2-gene subnetwork of Hb and Kr, with Bcd, Cad and Tll maternal control. This subnetwork can serve as a core for the whole gap network, and allow us to compare 4-gene networks evolved from this core to experimental results and results from the original (non-evolved) 4-gene models.

After networks are created through the Genetic Algorithm selection, we can analyze the robustness of each solution to maternal variability (in Bcd, Hb_{mat}, Cad and Tll). For this, we take any particular parameter set (network) and rerun the solution many times with different Bcd, Hb_{mat}, Cad and Tll gradients. The gradient variability is biological: the different gradients are data obtained from individual embryos. We used 89 individual Bcd gradients, 38 Cad gradients, 35 Hb_{mat}, and 27 Tll ones.

2.2 Experimental data for fitting

The data we used to fit our models is the result of a large-scale project we are engaged in, aimed at collecting, processing and analyzing the expression of the *Drosophila* segmentation genes[17, 20, 25]. Most of this dataset is now available publicly[26]. In this paper, we use expression data from early and mid cleavage cycle 14 (prior to full cellularization). This period of development is the stage during which segmentation patterns become mature, and also progressively more complicated, due to activation of more and more genes

that interact with the four gap genes in our model (this is called the mid-blastula transition). It is unknown precisely how many newly activated genes begin to interact with our 4 gap genes, nor do we know the precise activation times of these new genes. Therefore, new genes which are recruited in our simulations may shed light on the spatial patterns and regulatory features of real genes activated during the mid-blastula transition.

We have found that our models have faster and better fits to early patterns than to later, more mature ones. We believe this reflects that early gap patterns are chiefly under the control of the genes and maternal factors explicitly included in the model, while later, more complicated patterns begin to reflect interactions with other, newly activated genes recruited to the basic network.

We have also found, with quantitative data analysis[20, 25,] that segmentation patterns become more precise and robust from early to mid cycle 14. Hence, it is instructive to fit our models separately to early patterns and to mid cycle 14 patterns, to see if the robustness of the solutions reflects the trend in the data.

2.3 GA to Simulate Evolution of Gene Networks

The set of ODEs (1) was solved numerically by Euler's method[27]. We minimized the following cost function E by adjusting parameters T^{ab} in equation (1):

$$E = \sum (v_i^a(t)_{model} - v_i^a(t)_{data})^2. \qquad (2)$$

For the remaining parameters, m^a and h^a were found in preliminary runs and then used as fixed parameters; R_a, D^a, λ_a were determined similarly for the core 4- and 5- gene networks, but were found by GA in the reduced 2-gene networks; for the extended 4-gene model, R_a was found by GA and the rest were fixed.

Our approach followed the general scheme of population dynamics, by using repeated cycles of mutation, selection and reproduction. This is common to both GA[28] and general simulations of biological evolution.

Following the standard GA approach, the program generates a population of floating-point chromosomes, one chromosome for each gene a. The value of a given floating-point array a (chromosome a) at index b corresponds to a

T^{ab} value (see eqs. (1)). The task of the evolutionary search is to optimize the T^{ab} to fit to the experimental patterns (e.g. Fig. 1A, B).

The initial chromosome values are generated at random. The program then calculates the ν_i by eqs. (1) and scores each chromosome set (T matrix) by the cost function E (eq. 2). An average score is then calculated for all the chromosome sets run. Chromosome sets with worse-than-average scores are replaced by randomly-chosen chromosome sets with better-than-average scores. A proportion (from 5-25%, depending on computation) of the chromosomes are then selected to reproduce, undergoing the standard operations of mutation and crossover (defined below; 1/10 of these operations are crossover), giving changes to one or more of the T^{ab} values. The complete cycle of ODE solution, scoring, replacement of below-average chromosome sets, and mutation and crossover is repeated until the E score converges below a set threshold, typically 4000 – 5000 generations.

In GA, mutation is a genetic operator used to maintain genetic diversity from one generation of a population of chromosomes to the next, analogous to biological mutation. Point mutation in GA involves a probability that a T^{ab} value on a chromosome will be changed from its original state (compared to changing a nucleotide in biological point mutation).

GA crossover is a genetic operator used to vary chromosomes from one generation to the next, by swapping strings of values between chromosomes, analogous to crossover in biological reproduction. In one-point crossover, a point on a parent chromosome is selected. All data beyond that point is swapped between two parent chromosomes. Two-point crossover calls for two points to be selected on the parent strings. Everything between the two points is then swapped between the parent strings. Multi-point crossover is defined by analogy with the two-point case.

The model is implemented in Delphi (Windows) and GNU Pascal (Linux) and available from the authors upon request. Each run of the algorithm requires about 3 h CPU time on a Dell workstation (Intel Xeon CPU 2.80 GHz).

2.3.1 Introduction and withdrawal of new genes

In biology, one can imagine at least two scenarios for how new genes could become available for recruitment into a network[3]. First, a new gene could appear in the genome by the process of gene duplication. Second, a given gene from another network could become available for recruitment. In our model we do not distinguish these two cases, but introduce a Gene Introduction operator which adds a new gene to the network (at a rate of 5 – 10% per generation, depending on computation). Specifically, this adds a

new row and column to the T^{ab} matrix, which can be then be operated on by mutation and crossover. To study the importance of this one-way process forcing networks to recruit new genes, we introduced a Gene Withdrawal operator which removes a row and column from the T^{ab} matrix (at a rate of 2 - 10% per generation, depending on computation). Gene Withdrawal does not operate if the network is minimal (N = 4 genes).

3. RESULTS AND DISCUSSION

Recruitment of new genes into the preexisting network is typical for our model. We have found that even with point mutation alone, the network will recruit small numbers (from one to four) of new genes by the time it converges below the threshold *E* score. If mutation is reinforced with crossover, the number of recruits increases slightly (but statistically significantly). Increasing the rate of crossover leads to continual recruitment up to convergence, with some dozens of genes in the final networks.

We also find that cooption of a new gene can facilitate the evolutionary search, i.e. increase evolvability, in the sense of giving faster and better fits to the data. This was most apparent for a subnetwork 2-gene model.

Recruited genes can be uniform or spatially patterned. These patterns can either recapitulate patterns of existing network genes, or introduce patterns novel to the model network, but like patterns seen in the full biological network.

When we test the evolved networks against variability in maternal control factors, we find a significant minority display high robustness. These network solutions are varied, and show that robustness can be local to particular gene pattern boundaries. We did not find, however, that gene recruitment was associated with the robustness of the networks.

We expand on these findings in detail in the subsections below.

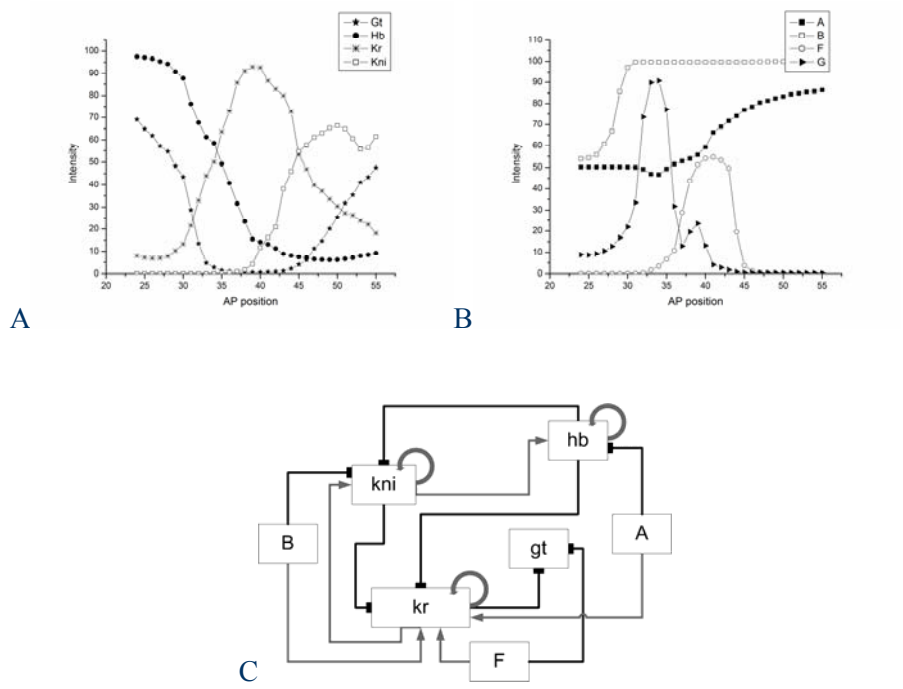


Figure 2. An example of a redundant gene network selected by Genetic Algorithms: 12 (A-L) genes have been recruited to the original 4 model genes.

A. Representative patterns for the 4 obligatory genes.

B. Patterns for some (A,B,F,G) of the genes recruited upstream of the 4 obligatory genes (in A).

C. Overview of the gene network in A-B, showing some of the interactions of the recruited genes. The genes are represented as boxes. Repressive interactions are represented by T-bar connectors. Looped arrows mean self-regulation. Cf. with Fig. 1C.

In this simulation, 4000 networks were generated for each generation; the point mutation rate was 18% per generation, plus 2% crossover rate; 20% of individuals with the best scores were marked for reproduction; and the rate for new gene recruitment was 5% per generation.

3.1 Point mutations are enough to recruit new genes

In our first series of runs, we studied recruitment events in detail and checked if crossover can raise the efficacy of recruitment. Several sets of runs under different conditions (point mutations only; point mutations + multi-point crossover; etc.) were performed, with each set including ~200 runs. For runs with both mutation and crossover, the mutation rate was adjusted so that total change per generation stayed comparable to runs with mutation only (e.g. if crossover, with rate 2% per generation was added to mutation, which had run at a 20% rate, the mutation rate would be adjusted

to 18%). Runs with E (eqn. (2)) scores below a threshold level were picked as winners. The threshold was established by visual inspection of the quality of fits to the expression patterns, and resulted in about half of the runs being winners. These winners were analyzed further to see what qualities they had.

We found new genes recruited to the network formed two distinct types of pattern. In the first type, recruits formed flat or nearly flat patterns (uniform distributions); they were incorporated into the network as ubiquitous activators or inhibitors. In the second type, recruits produced monotonic gradients, or even more sophisticated spatial patterns, influencing the patterns of the obligatory, minimal 4 genes of the network (*gt, hb, Kr & kni*). Fig. 2 shows a representative example of such a network. The obligatory 4 genes all fit well to the experimental data in Fig. 1. All good-score networks studied (112 point mutation only + 94 also with crossover) included at least one new recruit acting upstream of the obligatory genes (i.e., the obligatory genes were regulatory targets of the recruits). Nearly all networks studied included at least one (but usually more) upstream recruit that formed an AP gradient, such as Bcd. But most networks also included one or more upstream recruits that formed an opposing, postero-anterior gradient (Fig. 2B, patterns A, B). This is especially interesting because the minimal 4-gene ensemble we fitted in these runs did not posses such postero-anterior gradients. Hence, recruitment produced a kind of compensation for this lack of essential external output: in real fly embryos postero-anterior gradients of proteins such as *caudal* and *nanos* are essential for early segmentation.

In some cases, upstream recruits formed not simple monotonic gradients, but more sophisticated patterns with sub-domains (Fig. 2B, patterns F, G). These patterns are reminiscent of the mature patterns of *Drosophila* gap genes and demonstrate how recruitment could supply new gap genes for an evolving segmentation network (as in the transition from short to long germ band mechanisms).

We found that the point mutation operator is enough to recruit at least one new gene to the network; i.e., not one of the evolved high-score networks had just the obligatory 4 genes (Table 1). The mean number of recruits was around 3, while the average number of recruits upstream of (controlling) the obligatory 4 genes was about 2. As shown in Table 1, it appears that crossover selects networks with a slightly better score, gives a higher average number of upstream recruits and a lower number of these recruits have uniform distributions (all differences statistically significant). These upstream recruits more often form gradients or more complicated patterns (highly statistically significant). So, while we find crossover facilitates recruitment, point mutation is certainly sufficient for this, in

contrast to a mainstream view in evolutionary biology, that complex recombination of genetic material is required for recruitment[15].

Table 1. Outgrowth of networks by evolutionary search, with point mutations only and point mutations plus crossover*

	N runs	mean score	recruits, in toto	recruits upstream of 4 obligatory genes	upstream recruits expressed ~ubiquito-usly	upstream recruits forming patterns
point mutations	112	188.00±69.61	2.99±0.93	1.98±0.75	0.33±0.49	1.65±0.57
point mutations + crossover	94	171.18±63.70	2.95±0.60	2.22±0.92	0.20±0.43	2.02±0.98

* Results are mean±standard deviation.

3.2 Addition and subtraction of new genes

A simple explanation for why the number of recruits rises during evolution of a network is that addition of new potential recruits to the system creates an implicit pressure facilitating that recruitment. More specifically, a new recruit becomes incorporated into a network as its T matrix values begin to deviate from zero. Holding the T values at their initial zero state would involve a cost to the existing network, hence the presence of a new recruit causes pressure to evolve its T values and become incorporated into the network. With this tendency towards incorporation, the mean number of recruits should depend on the introduction rate. To test this, we introduced the Gene Withdrawal operator into our computations, as a way to control the net introduction rate. In conditions where addition is higher than subtraction, mutation and crossover operators still ensure recruitment. However, if the subtraction rate is equal to or greater than the addition rate, then recruitment is reduced compared to the Table 1 results; due to the random nature of the mutation, some networks can still gain recruits under these conditions. Hence, by using the Gene Introduction and Withdrawal operators to control net addition, we can show that addition of a new recruit creates implicit pressure for incorporation, facilitating recruitment.

3.3 Network redundancy and evolvability

The minimal, obligatory 4-gene network fits experimental pattern with good quality. Introduction of new recruits to this network does not generally raise the quality of the fits. In this sense, the new interactions with the

recruited gene can be considered redundant. However, this is not what is frequently called structural redundancy, in which repeated elements (genes) can substitute for lost elements, providing a type of robustness in networks. We find that withdrawal of a recruited gene from a good-scoring network (solution) makes its fit worse. Therefore, recruitment tends to alter the interactions of the original network; it is not advantageous to remove a gene once it has acquired functionality in the network. In this and the next section, we evaluate how recruitment affects a network's properties of evolvability and robustness. In terms of evolvability, we investigate whether recruitment of additional genes aids a network's capacity to evolve further. In particular, we can see if recruitment leads to faster (less generations) or better fits of the network to the data.

3.3.1 Evolvability of the four-gene models

To begin to investigate what potential role these added interactions provide, we tested whether they might help a network recruit more new genes. In these runs, we constrained the model to keep 5 obligatory genes: *gt, Kr, kni, hb,* and one new recruit. We first fit the model to the usual gap gene data of Fig. 1, during which process new genes were recruited. Once a good fit was attained, the fit criteria were changed to require the model to fit an expression pattern for 5 genes, by including the pattern for the primary pair-rule gene *even-skipped*. The 5th pattern could be fit by any of the newly recruited genes. Our expectation was that higher redundancy of networks could facilitate the evolutionary search for the gene to fit this pattern. We performed runs with point mutations only, and with point mutations and crossover (Table 2). The parameters for these runs were exactly as for section 3.3 (see caption for Fig. 2). To our surprise, we did not find any difference in efficacy between these runs and the previous runs of Table 1, as measured by the average number of recruits. We did find, however, that the average number and character of the recruits upstream of the *even-skipped* gene were significantly different: upstream recruits are far fewer in Table 2, and recruits form far fewer patterns. As in Table 1, crossover still tends to favor patterned recruits, compared with mutation alone.

Hence, gene networks with four obligatory genes do not show evident correlation of evolvability and the extent of redundancy.

Table 2. Efficacy of evolutionary search with redundant networks

	N runs	mean score (averaged)	recruits, in toto (averaged)	recruits upstream of *even-skipped* (*eve*) (averaged)	recruits upstream *eve*, expressed ~ubiquitously (averaged)	recruits upstream *eve*, forming gradients (averaged)
point mutations	98	235.23±65.68	3.06±0.96	0.50±0.56	0.40±0.49	0.10±0.39
point mutations + crossover	73	224.35±76.45	2.92±0.57	0.67±0.67	0.23±0.43	0.44±0.52

3.3.2 Evolvability of the two-gene model

To further investigate how the number of genes in the network might affect evolvability, we did a similar study, but starting from a 2-gene network, with *hb* and *Kr* only. As a control, we ran 500 simulations with this simple network, and computed an average score for how well Hb and Kr fit the biological patterns (both the mid cycle 14A patterns and the early cycle 14A ones, see Methods). Then, we did a series of 579 test simulations, in which two new genes were added at the onset of the evolutionary computation. Further addition/withdrawal operators were not used during the course of the computations. Again, the test networks were only required to fit the Hb and Kr patterns, but we wanted to see whether the two introduced genes would be incorporated into the network in such a way as to affect these pattern fits. Using the average score of the test computations, we found that the added genes significantly improved the fitting of the Hb and Kr pattern, both for early and mid cycle 14A, with the mid 14A difference being more dramatic. On average, the tests had scores of 128.005±71.703, and the controls had scores of 165.073±32.809.

For the 2-gene model, we find that redundancy serves as a mechanism to find not only better solutions, but also usually to find these solutions faster, in less generations; recruitment significantly raises the efficacy of the evolutionary search. Hence for a small fragment of the network, which could be treated as an "ancestral" primitive primary gene ensemble, redundancy via co-option could substantially facilitate evolutionary searches. To improve pattern, evolution causes such a primitive network to enlarge.

We wondered what kind of spatial patterns are generally made by the recruits, particularly whether they tended to mimic the missing two members of the real 4-gene gap network, *gt* and *kni*. We found that the patterns of the co-opted genes are usually reminiscent of anterior Hb or Gt domains; that is, simple S-shaped patterns, but often with reversed orientation (Fig. 3 A). We also found cases where recruits had Kr-like patterns (Fig. 3 B).

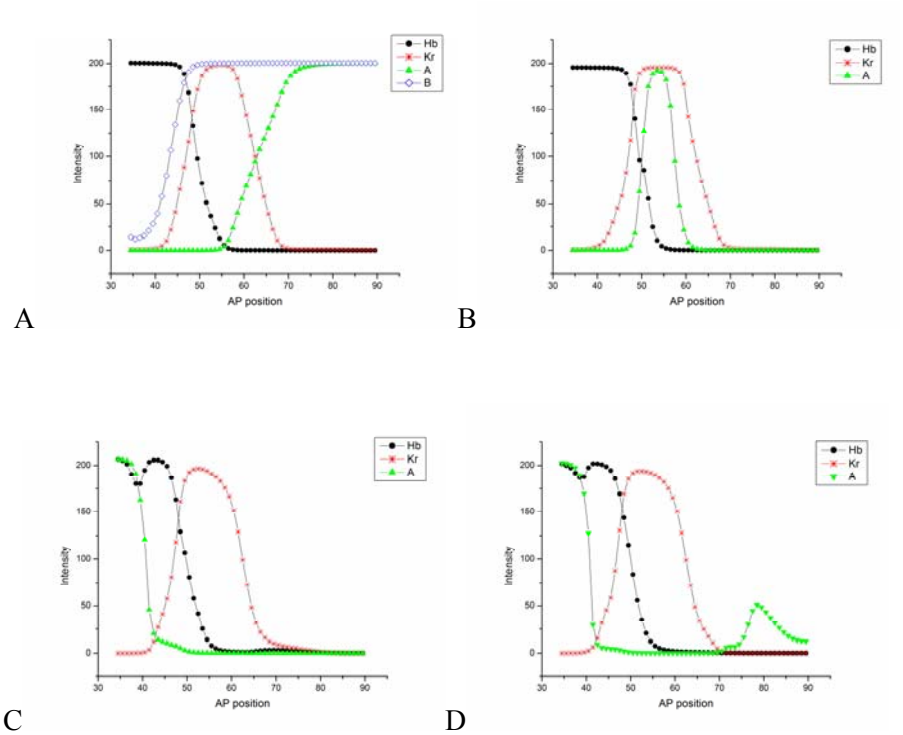


Figure 3. Representative examples of 2-gene models with two recruits.

A. Recruit patterns are similar to Gt and Hb (reverse orientation; Cf. Fig. 1A).

B. Recruit pattern is similar to Kr.

C. Recruit pattern looks like anterior Gt (Cf. Fig. 1A).

D. Recruit pattern is reminiscent of real Gt, with anterior and posterior domains.

We found several cases when one of the co-opted genes formed pattern similar to anterior Gt (Fig. 3 C). We also saw more complicated patterns, reminiscent of real two-domain gap patterns (Fig. 3 D). It could be that the evolutionary search is tending to fill in the missing gap patterns to generate the structure of the real, complete gap network. However, these two-domain *gt*-like patterns were relatively rare, and we did not find any *kni*-like patterns.

In summary, we have found that for the case of small fragments of gene ensembles, the co-option of new genes really does facilitate the evolutionary search. We can speculate that similar mechanisms acted during early evolution of primitive ancestral gene ensembles, while for evolutionarily

more mature and larger gene networks this tendency has become less pronounced.

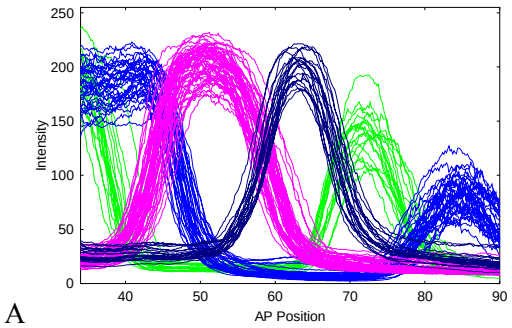
3.4 Redundancy and robustness of gene networks

Above, we have shown that our models of evolution, both the 2-gene and 4-gene ones, do account for recruitment of new genes and the selection of redundant networks. Here, we investigate the influence of redundancy on network robustness. A case of robustness that has received much attention in *Drosophila* segmentation is the robustness to variability in the shape of the Bcd morphogen gradient [16 - 21]. We can use our GA model to study this kind of robustness. The networks in the previous sections were selected on an averaged Bcd gradient (average profile of the real Bcd gradients in the FlyEx database[26]). If we take one of these networks, and now run it on the individual, and varying, Bcd gradients in our database (Fig. 4B), we get a picture of how robust the network's gap gene patterning is, and how this compares with the observed biological robustness (Fig. 4A). We can compare network robustness for the starting 2-gene and 4-gene models, as well as for the evolved redundant models.

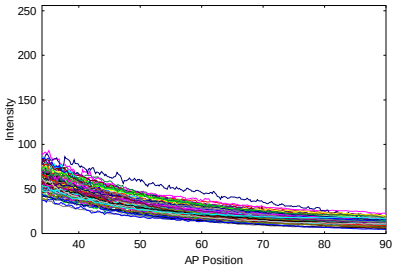
3.4.1 Robustness of two-gene networks

We start our investigation of robustness on the 2-gene model described in section 3.3.2, with the same test and control simulations described there. With this simplified model, the effect of variability in Bcd input on robustness is especially evident.

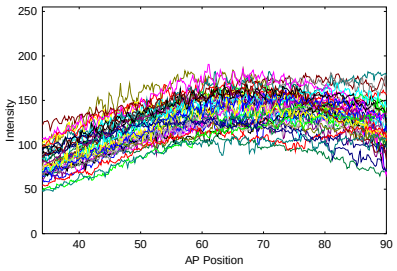
First, we found that the 2-gene solutions can be very robust to Bcd variability. Some solutions are substantially more robust than the robustness level observed for real *Drosophila* segmentation genes (Fig. 5A). However, good or even very good solutions (according to fitting score) can show no robustness to Bcd variability. These non-robust solutions can give Hb and Kr variability as high as that for Bcd (Fig. 5B), in contradiction to the several-fold drop in variability seen in the data (Fig. 4; [20]). The best-fit solutions span from highly robust, capable of filtering out Bcd variability nearly completely, to solutions unable to filter variability at all.



A



B



C

Figure 4. The real, biological variability in the gap gene patterns and the maternal inputs (See [20, 25]; Fly Ex DB).

A. The between-embryo variability of the gap gene patterns (gt , hb, Kr, kni; Cf. with Fig. 1B) for mid cleavage cycle 14A. Much of this variability is probably due to between-embryo variability in the maternal gradients, such as Bcd (B), and Cad (C).

B. The between-embryo variability of the maternal morphogenetic gradient Bcd, 89 embryos, 13th cleavage cycle.

C. The between embryo variability of the maternal factor Cad, 38 embryos, 13th cleavage cycle.

It is biologically established that the position of each domain border of each gap gene pattern is under the control of different combinations of regulatory inputs from the other members of the segmentation ensemble. In the case of the 2-gene model, we have one border for Hb and two borders (anterior & posterior) for Kr. Even for good-scoring solutions, there are cases when Hb is robust but Kr is less robust, or even non-robust (Fig. 5C). In many cases, the anterior Kr border is more robust then the posterior one (Fig. 5D), but we also saw some cases with the opposite (Fig. 5E). Our results show that robustness can evolve relatively independently at each border. Hence, the positional error for each border can be relatively independent.

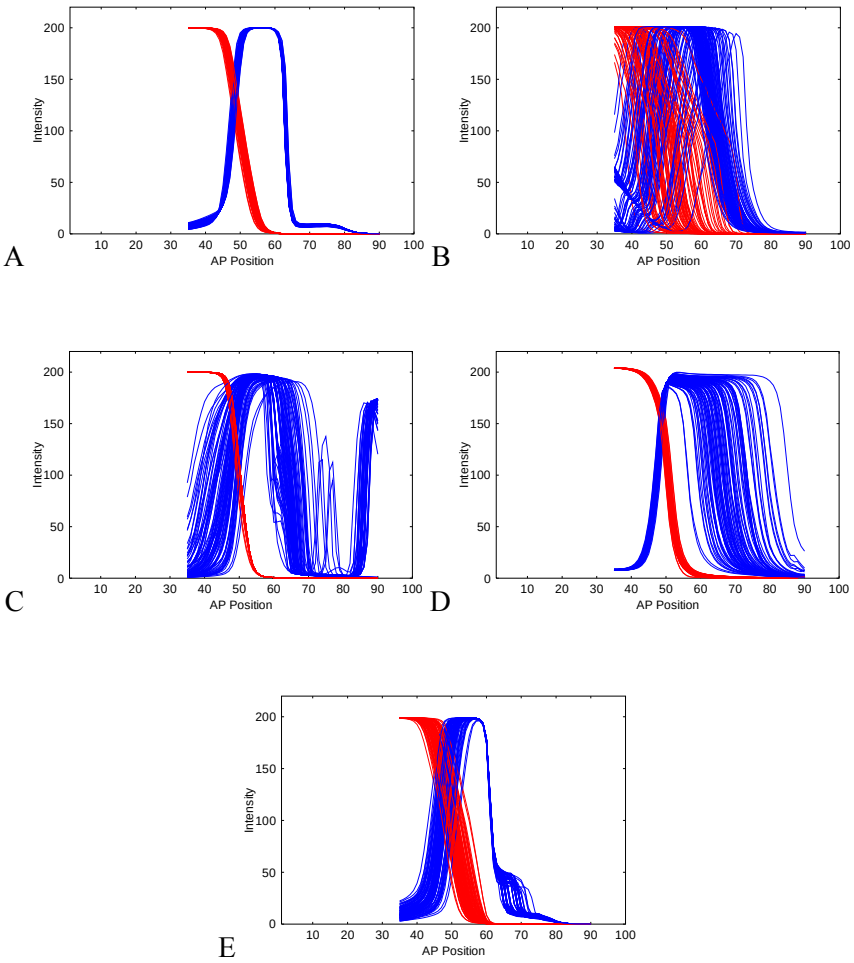


Figure 5. 2-gene model: robustness to Bcd variability. Bcd – green; Hb – red; Kr – blue.

A. Highly robust.

B. Not robust.

C. Hb is robust, but Kr is not robust.

D. All borders are robust, except for posterior Kr.

E. The posterior Kr border is the most robust.

Detailed analysis of the dynamics underlying this robustness to Bcd variability in the networks, both for the *hb-Kr* pair and the rest of gap ensemble, will be presented in another paper [29].

Because the 2-gene model is under the control of not one, but three external inputs (Bcd, Cad & Tll), we could also study its robustness to the variability of these other factors. Cad displays an even higher variability than Bcd (Fig. 4C; Cf. [20]). We have found a set of solutions that can filter this Cad variability to a degree comparable to Bcd filtering. This small set of solutions can filter Cad variability substantially better than real embryos can (data not shown). A typical result from this set has very precise Hb and anterior Kr borders, but un-precise posterior Kr (Fig. 6A). This situation is not unexpected, because the more posterior the domain position, the higher Cad intensity level and the higher the Cad positional (horizontal) variability (Fig. 4C; Cf. [20]).

We found cases where Cad variability induced not only quantitative, but qualitative changes in the Hb and Kr profiles. For instance, Cad variability can cause or at least highly amplify a second, posterior domain of Hb (Fig. 6B). Interestingly, this new posterior domain really does form in embryos during cycle 14, but later than the early stage patterns we used to fit the model.

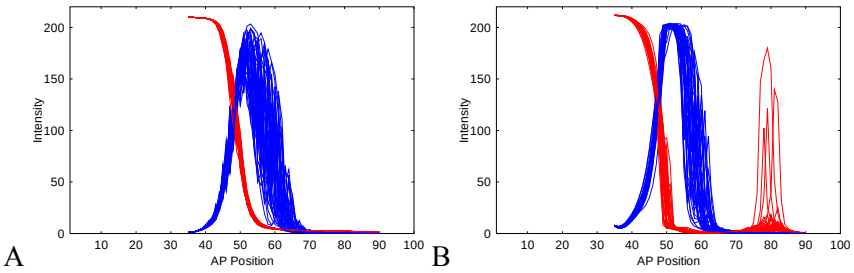


Figure 6. 2-gene model: Cad noise filtration. Hb – red; Kr – blue.

A. Hb and anterior Kr borders are very precise, but posterior Kr is not.

B. Cad variability can cause, or amplify, the second, posterior domain of Hb.

In summary, the simple 2-gene model, an elementary module of the gap network, shows all possible combinations of robust versus non-robust behavior, including a significant subset of solutions which are very robust to upstream variability. We also found that robustness of the three domain borders can be independently controlled.

We did not find any significant correlation between the fitting scores of solutions and their robustness, either for the control 2-gene networks or redundant solutions with one or two recruits. Robust solutions constitute about 10% or less of the total for 2-gene model, and non-robust solutions constitute a similar proportion.

3.4.2 Robustness of the four-gene networks

To see if the external noise filtration we found in the elementary 2-gene model works in more complicated gene networks, we tested the robustness of 4-gene network solutions to biological variability in a similar way. In these cases, the model has 10 domain borders, which makes systematic analysis more difficult. However, we again found that there is (1) no evident correlation between fitting scores and robustness, (2) different borders of different domains can display quite different levels of positional precision, (3), both robust and non-robust solutions are relatively rare, and (4) there were no evident differences in robustness between the control 4-gene networks and redundant networks evolved from these.

By using our extended 4-gene model, which includes control by all (or at least most) of the known maternal factors, Bcd, Hb_{mat}, Cad and Tll, we can investigate the networks' abilities to filter a more complete set of external variabilities, or to test their robustness to combinations of these factors.

We performed a series of runs to fit this extended version of the 4-gene model, for both control conditions and test runs with recruitment of new members to the core 4-gene ensemble. This extended 4-gene model is several times more computationally intensive, so for this case we have obtained several dozen networks with an appropriate level of fitting to the experimental data.

We performed a detailed analysis of the 18 best-fit solutions (control and test runs) for robustness to variability in all four external factors (Bcd, Hb_{mat}, Cad and Tll), one by one, and in pairs of the factors. For Bcd variability alone, we found that the behavior of the extended gene network is similar to that observed for 2-gene and minimal 4-gene models. In most cases (Fig. 7A; control runs shown in Fig. 7, but test runs give the same qualitative results), Hb and Gt tend to be highly robust, while Kr and kni are less robust (but they are comparable to the biologically observed robustness). In other cases, all the gap domains (with the exception of posterior Hb) can show similar, and relatively high, levels of positional variability (not shown). Finally, we see some cases of autonomy in robustness to Bcd variability: the Kr domain can show very high precision, while the other genes do not (not shown).

With variable Cad input, we have not found robust solutions (Fig. 7B). The only precise domain in the case of Fig. 7B is the most anterior Gt one. Robustness of this domain can be expected because it is chiefly under control of Bcd and relatively independent of Cad regulation. With variable Tll (also a posterior gradient), the picture is similar (Fig. 7C); the extended 4-gene model is largely not robust to this. Only the most anterior borders of

Gt, Hb and Kr are robust, and again these are largely under Bcd control and are relatively independent of Tll.

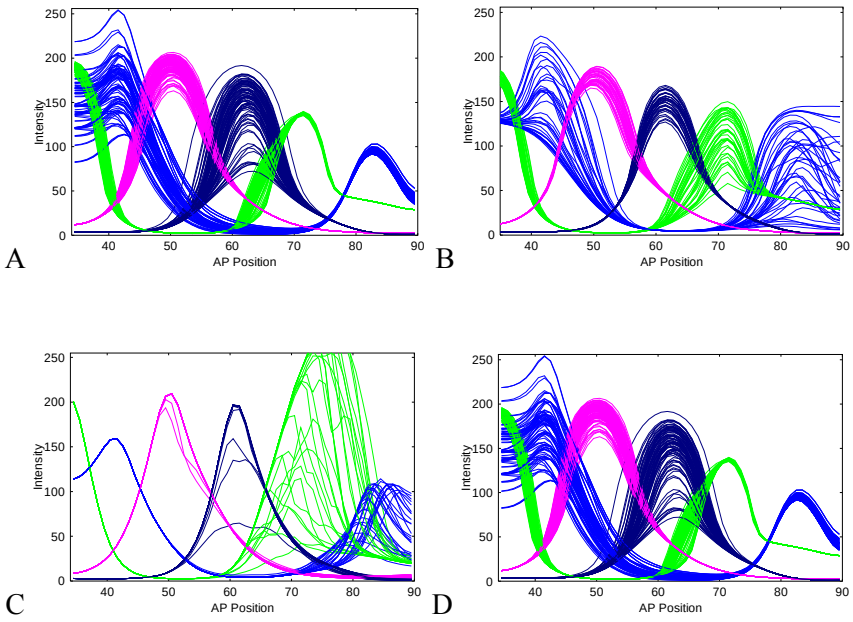


Figure 7. Extended 4-gene model (Bcd, Cad, Tll maternal control). Colors as in Fig. 4.

A. Robustness to Bcd variability.

B. Non-robustness to Cad variability.

C. Non-robustness to Tll variability.

D. Robustness to double Hb + Bcd noise.

Looking at pairs of external factors, the most interesting case was the pair of Bcd and Hb_{mat}. It is biologically established that this pair of anteroposterior gradients cooperatively control patterning of gap genes in the anterior half of the *Drosophila* embryo[30]. We found that the extended 4-gene model is capable of decreasing not only Bcd or Hb_{mat} variability separately, but can filter both these variabilities together (Fig. 7D).

4. CONCLUSIONS

In this work, we have presented the results of a computational simulation of evolution of the segmentation gene network, controlling spatial patterning

in early fly embryonic development. We used Genetic Algorithms (GA) methods to evolve the parameters of a differential equation model for the segmentation proteins, tested against fitness for matching the biological data for the protein patterns.

We simulated recruitment (co-option) of new genes to existing networks, by introducing gene-addition and gene-removal operators, on top of standard GA techniques. We found that recruitment occurred in all our simulations, even for those in which only point mutations were operating. Crossover aided recruitment, but was not necessary. The recruited genes were either ubiquitous or formed spatial patterns, many of which were similar to real, biological gene patterns, including patterns for genes not in our core starting networks.

With our generated networks, we tested whether recruitment affected evolvability or robustness to variability in external factors. We found the evolvability was especially aided in a 2-gene subnetwork, possibly representing the process by which the ancestral short-germ band segmentation mechanism evolved into the long-germ band mechanism of flies. For robustness, we tested the networks for their ability to filter variability in upstream, regulatory maternal factors. This apparent filtration has been the subject of a great deal of attention in developmental biology in recent years, and we find that a significant subset of our evolved networks (with and without recruitment) have the capacity to filter this maternal variability, i.e. generate domain boundaries with greater precision than the maternal regulatory gradients. Robust networks display a variety of behaviors, demonstrating that domain boundaries can be regulated independently with respect to spatial precision. The model was very successful for filtering variability in anterior gradients, such as Bcd and Hb_{mat} (individually, or as a pair), but less so with posterior gradients, such as Cad and Tll. There is no apparent correlation between the quality of a network's fit to the data and its capacity for robustness to maternal variability.

It has been suggested that redundancy, either structural, in which new genes can substitute for existing genes, or functional, in which new genes create compensatory pathways, can provide robustness to networks[13]. We do not find that recruitment aids robustness to maternal variability, but our recruited genes are not strictly structurally or functionally redundant. Rather, removal of recruited genes results in fitting scores lower than the starting network: genes can not be freely disposed once they have been integrated into the functionality of the network.

Our work demonstrates that relatively simple evolutionary operators can account for network outgrowth. The evolved networks display a number of features of the biological system of interest, such as recruitment of genes

from ancestral modules, and robustness to regulatory variability, shedding light on the evolutionary and functional dynamics of this developmental network.

5. ACKNOWLEDGMENTS

This work was supported by the Joint NSF/NIGMS BioMath Program grant R01-GM072022. The authors thank an anonymous reviewer for helpful comments on an earlier version of the manuscript.

6. REFERENCES

[1] Wilkins, A. S., (2002). *The Evolution of Developmental Pathways*, Sinauer Associates, Sunderland, MA.

[2] Duboule, D., and Wilkins, A., (1998). The evolution of bricolage, *Trends Genet.* 14:54–59.

[3] True, J.R., and Carroll S.B., (2002). Gene co-option in physiological and morphological evolution, *Annu. Rev. Cell Dev. Biol.* 18:53–80.

[4] Carroll, S.B., Grenier, J.K., and Weatherbee, S.D., (2001). *From DNA of Diversity: Molecular Genetics and the Evolution of Animal Design*, Malden, MA: Blackwell Science.

[5] Carroll, S.B., (2005). Evolution at two levels: on genes and form, *PLoS Biology* 3(7): e245.

[6] Davidson, E.H., (2001). *Genomic Regulatory Systems: Development and Evolution*, Academic, San Diego.

[7] Goodwin, B.C. and Kauffman, S.A., (1990). Spatial harmonics and pattern specification in early Drosophila development. Part I. Bifurcation sequences and gene expression, *J Theor Biol*. 144:303-19.

[8] Reinitz, J. and Sharp, D.H., (1995). Mechanism of formation of eve stripes, *Mechanisms of Development*, 49:133-158.

[9] Jaeger, J., et al., (2004). Dynamic control of positional information in the early Drosophila blastoderm, *Nature* 430:368-371.

[10] Hunding, A., Kauffman, S.A., and Goodwin, B.C., (1990) Drosophila. segmentation: supercomputer simulation of prepattern hierarchy, *J. Theor. Biol*. 145:369–384

[11] Burstein, Z., (1995). A network model of the developmental gene hierarchy, *J. Theor. Biol*. 174:1–11.

[12] Sánchez, L. and Thieffry, D., (2001). A logical analysis of the Drosophila gapgene system, *J. Theor. Biol*. 211:115-141.

[13] Wagner, A. (2005). Distributed robustness versus redundancy as causes of mutational robustness, *BioEssays*, 27: 176-188.

[14] Jaeger, J., et al. (2004). Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster, *Genetics* 167:1721-1737.

[15] Carroll, R.L., (2002). Evolution of the capacity to evolve, *Journal of Evolutionary Biology*, 15:911-921.

[16] Houchmandzadeh, B., Weischaus, E., and Leibler, S., (2002). Establishment of developmental precision and proportions in the early Drosophila embryo, *Nature* 415:798-802.

[17] Holloway, D.M., Reinitz, J., Spirov, A.V., and Vanario-Alonso, C.E., (2002). Sharp borders from fuzzy gradients, *Trends Genet* 18:385-387.

[18] Spirov, A.V. and Holloway, D.M., (2003). Making the body plan: precision in the genetic hierarchy of Drosophila embryo segmentation, *In Silico Biology* 3:89-100.

[19] Houchmandzadeh, B., Wieschaus, E. and Leibler, S. (2005). Precise domain specification in the developing Drosophila embryo, *Phys Rev* E 72, 061920.

[20] Holloway, D.M., Harrison, L.G., Kosman, D., Vanario-Alonso, C.E., and Spirov, A.V., (2006). Analysis of pattern precision shows that Drosophila segmentation develops substantial independence from gradients of maternal gene products, *Developmental Dynamics* 235:2949–2960.

[21] Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the Limits to Positional Information, *Cell*, 130:153-164.

[22] Wolpert, L., (1969). Positional information and the spatial pattern of cellular differentiation, *J. Theor. Biol*. 25:1–47.

[23] Mjolsness, E., Sharp, D.H., and Reinitz, J., (1991). A connectionist model of development, *J Theor Biol.* 152:429-453.

[24] Reinitz, J. and Sharp, D.H., (1995). Mechanism of eve stripe formation, *Mechanisms of Development* 49:133-158.

[25] Surkova, S., Kosman, D., Kozlov, K., Manu, Myasnikova, E., Samsonova, A.A., Spirov, A., Vanario-Alonso, C.E., Samsonova, M., and Reinitz, J. (2008). Characterization of the Drosophila segment determination morphome. *Developmental Biology* 313: 844-862.

[26] Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M., and Reinitz, J. (2004). A database for management of gene expression data in situ. *Bioinformatics* 20: 2212-2221.

[27] Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., (1988). *Numerical Recipes*, Cambridge University Press, Cambridge.

[28] Schwefel, H.-P., (1981). *Numerical Optimization of Computer Model*s, Wiley, Chichester.

[29] Manu, Surkova, S., Spirov, A.V., Gursky, V., Janssens, H., Kim, A., Radulescu, O., Vanario-Alonso, C.E., Sharp, D.H., Samsonova, M., and Reinitz, J. (submitted). Canalization of gene expression in the Drosophila blastoderm by dynamical attractors.

[30] Simpson-Brose, M., Treisman. J., and Desplan, C., (1994) Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell* 78:855–865.