Recruiting New Genes in Evolving Genetic Networks: Simulation by the Genetic Algorithms Technique

Article · October 2007

Source: DOAJ

CITATION
1

READS
14

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

Recruiting New Genes in Evolving Genetic Networks: Simulation by the Genetic Algorithms Technique

Alexander V. Spirov and David M. Holloway

***Abstract*—Gene recruitment or co-option is defined as the placement of a 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 minimal 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 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. Gene recruitment causes outgrowth of an evolving network, resulting in structural and functional redundancy. Such redundancies can affect the robustness and evolvability of networks.**

***Index Terms*—Complexification of gene networks, gene co-option, modeling of biological evolution by Genetic Algorithms, redundancy and robustness of gene networks.**

I. 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 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 [6], [1].

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].

Here, we investigate the mechanisms of co-option, with direct application to how this might occur in insect segmentation networks. We use the Genetic Algorithms (GA) technique to model biological evolution on a minimal

Manuscript received July 22, 2007. This work was supported by the Joint NSF/NIGMS BioMath Program grant R01-GM072022.

A. Spirov is with the 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).

D. Holloway is with the Mathematics Department, British Columbia Institute of Technology, Burnaby, B.C., Canada (e-mail: David_Holloway@bcit.ca).

4-gene model network of early fly development (adapted from [11], [7]). In our simulations, we used standard GA operators (mutation and crossover) as well as our own operators for introducing and removing new genes on the minimal network. Simulation output is tested against experimentally observed segmentation gene expression patterns.

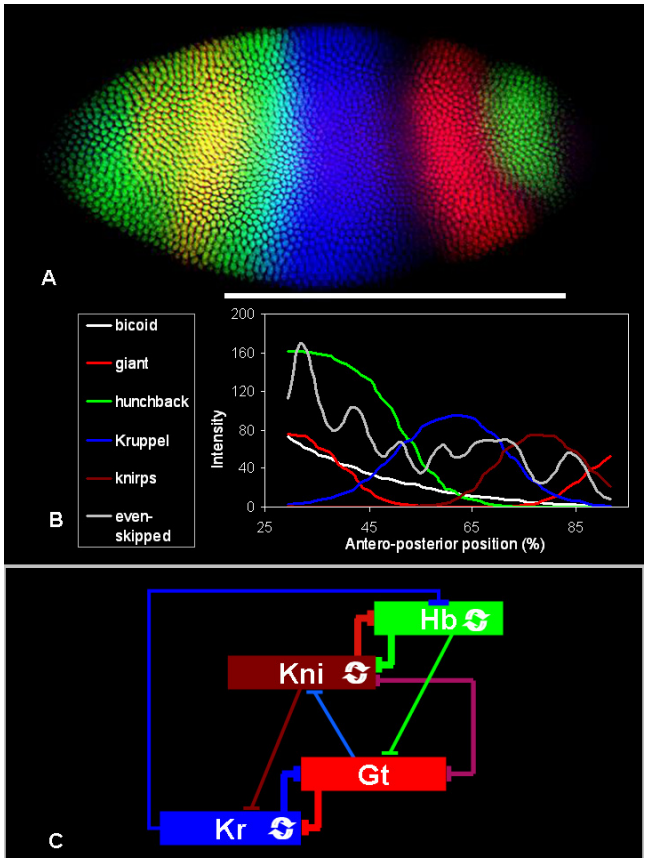

Figure 1. Biological data used to fit ODE model by GA.

A. *Drosophila melanogaster* early (blastoderm stage) embryo at nuclear cleavage cycle 14A, immunostained for *Gt, Hb,* and *Kr* proteins (FlyEx database embryo; http://flyex.ams.sunysb.edu/flyex/). Anterior is to the left, dorsal is up. White bar indicates the region included in the model.

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

C. Overview of the gap gene network (After [7]). The gap genes are represented as boxes. Repressive interactions are represented by T-bar connectors. Looped arrows mean self-regulation.

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. By contrast, many evolutionary biologists believe that the main mechanism facilitating recruitment is the sophisticated shuffling of genetic material, such as unequal crossing over, or the activity of transposons [8]. By using a computational approach, we can test the significance of different mechanisms for recruitment. Specifically, we test point mutation against one- and multi-point crossover (recombination).

Our results also 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 studied the structural and functional redundancy of the evolved networks, as well as the possible influence of the redundancy on their evolvability and robustness.

II. Methods and Approaches

A. The segmentation gene network and its modeling

The fruit fly's segmentation gene network is one of not too many fully characterized genetic ensembles, making it for many years the most popular object for computer simulations of its function and evolution [9], [10], [11], [12], [13], [14].

The maternal Bicoid (Bcd) protein in the blastoderm stage *Drosophila* embryo is a classic example of a morphogen [15]: distributed in a spatial gradient, it transcriptionally activates its targets, such as the gap genes *Krüppel* (*Kr*), *giant* (*gt*), *knirps* (*kni*) and *hunchback* (*hb*) (Fig. 1), in a concentration-dependent manner at distinct positions. Subsequent cross-regulation of the gap-genes makes segmentation patterns sharper and more precise.

The gene circuit framework [16], [17] was used to model the antero-posterior (AP) pattern formation shown in Fig. 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 model includes the *bcd*, *hb*, *Kr*, *kni* and *gt* gene products (see Fig. 1). The rates of change in protein concentration dv_i^a/dt for each regulated gene product a in each nucleus i are given by a system of *number of proteins* times *number of nuclei* ODEs defined by

$$dv_i^a/dt = R_a g\left(\sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{Bcd} + h^a\right) + D^a\left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\right] - \lambda_a v_i^a . \quad (1)$$

The main terms on the right hand side of (1) represent protein synthesis, diffusion and decay, respectively. R_a, D^a, and λ_a are parameters representing the rates of maximum synthesis, diffusion and decay. $g\left(u^a\right)$ is a

sigmoid regulation-expression function. For values u^a below -1.5 and above 1.5 $g(u^a)$ 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 the gene system. Bcd is a general activator for all four gap genes considered. h^a represents regulatory input from ubiquitous factors.

B. GA to Simulate Evolution of Gene Networks

The set of ODEs (1) was solved numerically by Euler's method [18]. We minimized the following cost function *E* by adjusting parameters T^{ab} in equation (1):

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

The rest of the parameters, $R_a, m^a, h^a, D^a, \lambda_a$, were found in preliminary runs and then used as fixed parameters in the following computer experiments.

Our approach followed the general scheme of population dynamics, by using repeated cycles of mutation, selection and reproduction. This is common to both GA [19] 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. The value of a given floating-point array a (chromosome a) at index b corresponds to a T^{ab} value (See (1)). The task of the evolutionary search is to optimize the T^{ab} to fit to the experimental patterns (Fig. 1B).

The initial chromosomes are generated at random. The program then calculates the v_i and scores every chromosome by the cost function *E*, and calculates an average score. Chromosomes with worse-than-average scores are replaced by chromosomes with better-than-average scores. (The choice of the chromosomes is random.) A proportion of the chromosomes then undergoes standard operations of mutation and crossover (defined below), giving changes to one or more of the T^{ab} values. The complete cycle of evaluation, scoring, replacement of loser chromosomes, and mutation and crossover is repeated multiple times to simulate evolution.

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 one.

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

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 from time to time. 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. Gene Withdrawal does not operate if the network is minimal (N = 4).

III. 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 in 5,000 generations. If mutation is reinforced with crossover, the number of recruits increases somewhat (but statistically significantly). Increasing the rate of crossover leads to continual recruitment up to convergence, with some dozens of genes in the final population.

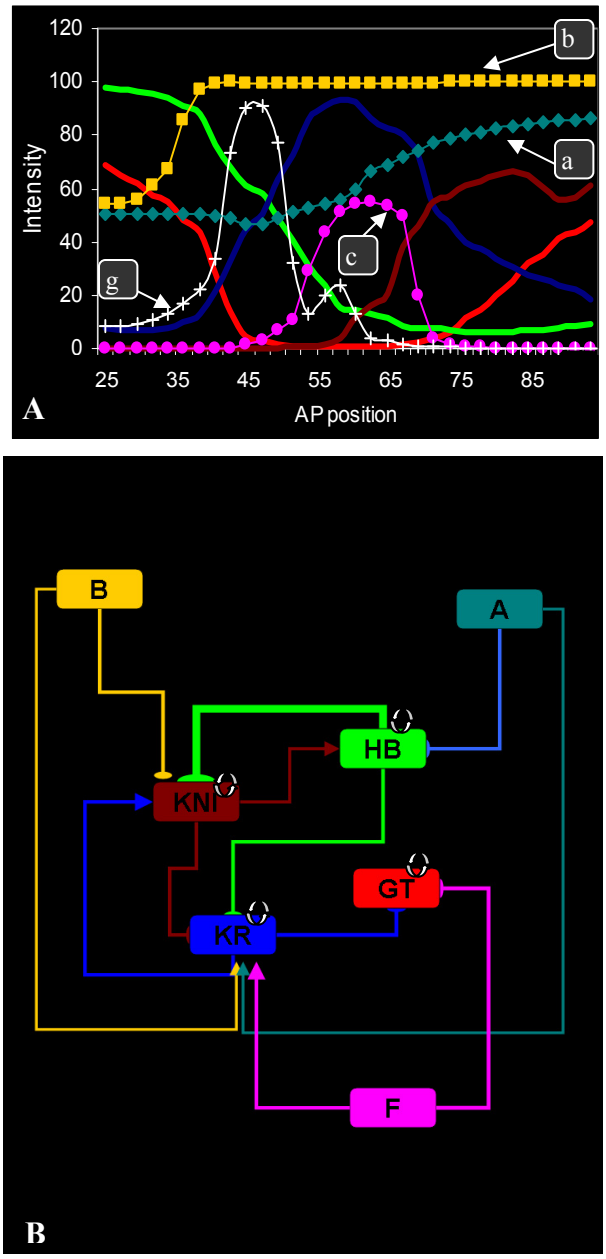

Figure 2. An example of a redundant genetic network (16 genes in toto) selected by genetic algorithms.
The size of the population is 4000; mutation rate is 18% per generation; 20% of individuals with higher scores are marked for reproduction (truncation strategy); rate for new genes recruitment is 5% per generation; rate for crossover action is 2% per generation.

A. Representative patterns (lines with points, a, b, c, & g) for genes recruited upstream of the 4 obligatory genes (shown in solid lines: *gt*, red; *hb*, green; *Kr*, blue; *kni*, brown).
B. Overview of the gene network in A. The genes are represented as boxes. Repressive interactions are represented by T-bar connectors. Looped arrows mean self-regulation. Cf. with Fig. 1C.

A. Point mutations are enough to recruit new genes

In our first series of runs, we studied in detail the recruitment events and checked if crossover can raise the efficacy of recruitment. Several sets of runs under different conditions (mutations only; 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. Runs with *E* (see (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.

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 (ubiquitous distribution); they were incorporated into the network as ubiquitous activators or inhibitors. In the second type, recruits produced monotonic gradients or even more sophisticated 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 + 94) included at least one new recruit acting upstream of the obligatory genes: the obligatory genes were 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 opposing, postero-anterior gradients (Fig. 2A, 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 gradients, but more sophisticated patterns with sub-domains (Fig. 2A, patterns c, 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 only 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, crossover helps to select networks with a slightly better score, and a slightly higher average number of recruits. Crossover shifts selection in such a way that recruits upstream of the obligatory 4 genes less often have ubiquitous distributions, and more often form gradients or more complicated patterns.

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

	N runs	mean score (averaged)	recruits, in toto (averaged)	recruits upstream of 4 obligatory genes (averaged)	upstream recruits expressed ~ubiquitously (averaged)	upstream recruits forming patterns (averaged)
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

B. Addition and subtraction of new genes

A simple explanation for why the number of recruits rises during evolution is that addition of new potential recruits to the system creates an implicit pressure that facilitates recruitment. To study this effect, we introduced the Gene Withdrawal operator into our computations. 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. Implicit pressure by the Gene Introduction operator does facilitate recruitment.

C. 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, withdrawal of a recruited gene from a good-scoring network (solution) makes its fit worse. 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 a 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 search of the next new recruit. 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 3.1 (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.

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

ISBN:978-988-98671-6-4

D. *Redundancy and robustness of gene networks*

Above, we have shown that our model of evolution does account for recruitment of new genes and the selection of functionally 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 [20], [21], [22], [23], [24], [25]. 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). If we take one of these networks, and now run it on the individual, and varying, Bcd gradients in our database, we get a picture of how robust the network's gap gene patterning is, and how this compares with the observed biological robustness. We can compare network robustness for the minimal 4-gene model, as well as for the evolved redundant models.

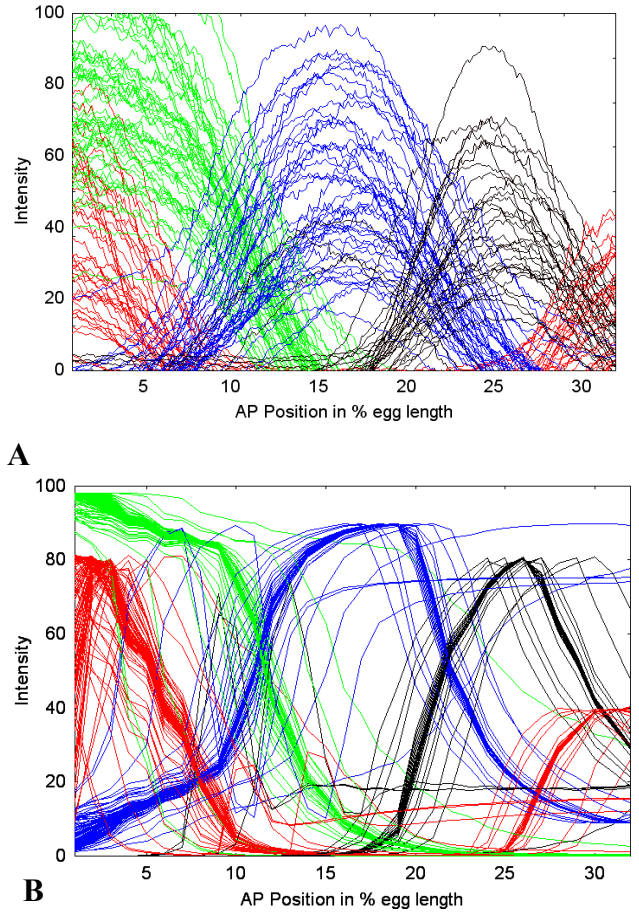

Figure 3. The variability of early gap gene patterns in reality (See [24]; Fly Ex DB), and in computer experiments.

A. The between-embryo variability of early gap gene patterns (*gt* – red, *hb* – green, *Kr* – blue, *kni* – brown) for early cleavage cycle 14A. This variability probably is caused by between-embryo variability in the shape of the primary morphogenetic gradient Bcd.

B. The between-embryo variability simulated by using individual, real Bcd gradients with different shapes as external input for a model of the same 4 genes.

Fig. 3 shows the real experimentally observed variability of early gap gene patterns (Fig. 3A) compared to one of the typical redundant models (Fig. 3B). The simulated gap gene network appears at least as noisy as the real one. The same level of noisiness was observed for the non-redundant minimal 4-gene models fitted with recruiting off (data not shown). That is, the redundancy gained in our computer experiments does not increase robustness to the variability of the Bcd gradient, but does produce noise levels comparable to the real ones.

IV. Conclusions

- Recruitment (co-option) of new genes into pre-existing gene networks was observed in simulations of network evolution using Genetic Algorithms (GA).
- The GA operator for point mutations was sufficient to produce recruitment. Crossover (recombination) is not required for gene recruitment.
- The GA operator for multi-point crossover facilitates co-option of new recruits upstream of the original network genes.
- Recruitment makes networks structurally and functionally redundant.
- The functional redundancy achieved by recruitment does not influence the robustness and evolvability of networks (Cf. [26]).

ACKNOWLEDGMENTS

The authors thank an anonymous reviewer for helpful comments on the manuscript.

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, JR, and SB Carroll., (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 Vol. 3, No. 7, e245.

[6] Davidson, E.H., (2001). *Genomic regulatory systems: development and evolution*, Academic, San Diego.

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

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

[9] 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.

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

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

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

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

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

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

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

[17] Reinitz J. and Sharp D.H., (1995). "Mechanism of eve stripe formation," *Mech Dev* 49:133-158.

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

[19] Schwefel, H.-P., (1981). *Numerical optimazaion of computer model*s, Wiley, Chichester.

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

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

[22] 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.

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

[24] 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.

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

[26] Wagner, A., (2005). "Distributed robustness and redundancy as causes of mutational robustness," *Bioessays* 27:176-188.

ISBN:978-988-98671-6-4