

# GA Simulation of Evolution of the Hierarchical Module Structure on Gene Networks

Shinji Nakashima\* and Koji Kurata\*\*

\*Graduate School of Engineering and Science, University of the Ryukyus

\*\*Faculty of Engineering, University of the Ryukyus

1 Senbaru, Nishihara, Okinawa. 903-0213

(Tel:098-895-8638 ; Fax:098-895-3636)

(shinji@mibai.tec.u-ryukyu.ac.jp)

**Abstract:** The animal body plan is controlled by the genetic networks of the hierarchical modular structure. For example, homeotic selector genes function at many levels in the regulatory hierarchy, so that homeotic mutations result in the transformation of one body segment into another, which is recognized by the misplaced development of structures that are normally present elsewhere in the embryo. The purpose of this research is to search for the mechanism of the evolution of such genetic networks. We modeled simplified genetic networks, and simulated the evolution of the genetic networks by GA. The simulation results show that the environmental perturbation possibly gave rise to the evolution of the hierarchical modular structure.

**Keyword:** Hierarchical Modular Structure, Genetic Networks, Genetic Algorithm, Evolution, Environmental perturbation

## 1 Introduction

The animal body plan is controlled by the genetic networks of the hierarchical modular structure. The homeotic selector genes function at multi levels in the regulatory hierarchy of the developmental genetic networks. Therefore, homeotic mutations result in the transformation of one body segment into another. For example, an *antennapedia* mutant, whose antennae are converted into leg structures caused by mutation in the *Antennapedia* gene [1-3]. Davidson and Erwin (2006, 2009) [4, 5] assumed the genetic networks that control the early development of animal embryos and proposed a hierarchical modular structure on genetic networks, which can be described by a hierarchy with four types of modules, *kernel*, *plug-ins*, *I/O switches* and *batteries*. The kernels are top components and the batteries are bottom components of hierarchy [5, 6]. If the genetic network were a nearly random network, any change to the network would result in drastic difference in the body plan because each gene may regulate or be regulated by several other genes, and the effects will spread out to the whole network [7].

The purpose of our research is to search for the mechanism of the evolution of such genetic networks of hierarchical modular structure. Why do genetic networks have hierarchical modular structure? We assumed one of the causes is the *environmental perturbation*, namely, the environmental perturbation is advantageous for hierarchical modular structure.

We modeled simplified genetic networks, and simulated the evolution of the genetic networks by GA.

The simulation results show that the environmental perturbation possibly gave rise to the evolution of the hierarchical modular structure.

## 2 The model of genetic networks and phenotype

The various model of genetic networks [8-11] has been proposed, but these models are too complex for our simulation. Thus, we propose a simplified model of genetic networks. The model of a genetic network and phenotypic expression used for our simulation is similar to feedforward neural network with linear units, which is given as follows. A gene in our model can make other genes expressed. Then, those genes can also make other genes expressed. Fig.1 is a conceptual diagram of the model of genetic networks and phenotype. The genes are linearly-ordered and numbered. Let us call genes of larger (smaller) number as *lower* (*higher*) genes. We assume, for simplicity, a lower gene doesn't regulate a higher gene. Thus, the feedback is not considered in this model, and the generality of this model is restricted. But emergence of the hierarchical modular structure are expressible on this model. Therefore, the model is sufficient for the purpose or observation of evolution of the hierarchical modular structure. Each individual consists of genotype  $G$  and phenotype vector  $\mathbf{p}$ . The dimension of  $\mathbf{p}$  is  $n_p$ , and  $G$  consists of  $n_g$  components. Gene  $g_i$ , the  $i$ -th component of  $G$ , is given by the following definition for

$i = 1, 2, \dots, n_g$ :

$$g_i = (x_i; \rho_i, \sigma_i, \gamma_i), \quad (1)$$

$$\rho_i = (\rho_{1i}, \rho_{2i}, \dots, \rho_{i-1i}), \quad (2)$$

$$\sigma_i = (\sigma_{i+1i}, \sigma_{i+2i}, \dots, \sigma_{n_g i}), \quad (3)$$

$$\gamma_i = (\gamma_{1i}, \gamma_{2i}, \dots, \gamma_{n_p i}), \quad (4)$$

where  $x_i (\geq 0)$  is an *amplification coefficient*,  $\gamma_{ji}$  is a *causal coefficient*. The amplification coefficient  $x_i$  is the strength of influence for the lower gene. the *causal coefficient*  $\gamma_{ji}$  is the Boolean representation of the causal relation from the gene  $g_j$  to the phenotype  $p_j$ .  $\sigma_{ij}$  and  $\rho_{ji}$  are Boolean *causal factors* to determine the influence of gene  $g_j$  to gene  $g_i$ . For two genes  $g_i$  and  $g_j$  ( $j < i$ ), we suppose gene  $g_j$  makes gene  $g_i$  expressed only when  $\sigma_{ij}\rho_{ji} = 1$ , so that the causal relation between  $g_i$  and  $g_j$  can change depending on the mutation in either of the two genes. And the expression level  $y_i$  of gene  $g_i$  is defined as follows for  $i = 1, 2, \dots, n_g$ :

$$y_i = x_i \sum_{j=1}^{i-1} \sigma_{ij}\rho_{ji}y_j, \quad y_1 = 1. \quad (5)$$

Note that  $\sigma_{ij}$  belongs to  $g_j$  and  $\rho_{ji}$  belongs to  $g_i$ .  $g_1$ , which is the highest gene, is defined as the trigger for the whole network, which exists only to make some other genes expressed, and so  $y_1 = 1$ ,  $\gamma_{i1} = 0$  ( $i = 1, 2, \dots, n_p$ ).

$p_k$  is the component of phenotype vector  $\mathbf{p}$ , which is calculated shown as follows for  $k = 1, 2, \dots, n_p$ :

$$p_k = \sum_{j=2}^{n_g} \gamma_{kj}y_j. \quad (6)$$

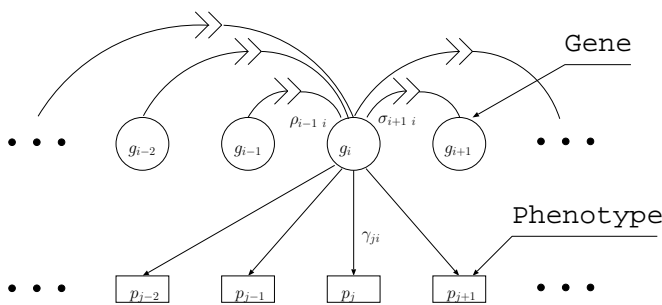


Fig. 1: Conceptual diagram of the model of genetic networks and phenotype

### 3 Method of individual's evolving

We applied genetic algorithm [12] (GA) to simulate the evolution of the genetic networks. Each genetic operations is given as follows.

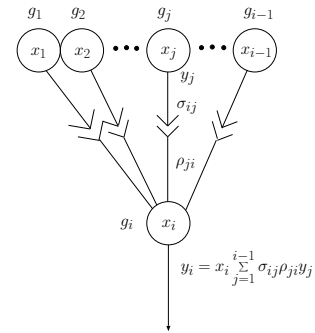


Fig. 2: Conceptual diagram of model of genetic networks

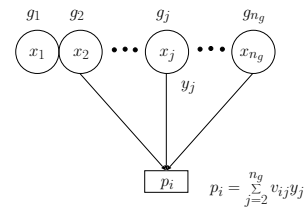


Fig. 3: Phenotype vector

- Method of crossover  
We apply the uniform crossover. Individual genes in the genotype are compared between two parents. The genes are swapped with a fixed probability of 0.5.
- Method of mutation  
When the mutation of gene  $g_i$  occurs with a certain probability, Gaussian mutation occurs in amplification coefficient  $x_i$ , and then, the mutation in causal factors and coefficient  $\rho_{ji}$ ,  $\sigma_{ji}$  and  $\gamma_{ji}$  occurs in a certain probability.
- Method of selection  
We apply the elite preservation strategy.
- The fitness function  $f$  of an individual is defined as follows:

$$f(\mathbf{t}; \mathbf{p}) = -\|\mathbf{t} - \mathbf{p}\|^2, \quad (7)$$

where  $\mathbf{p}$  is the phenotype vector of the individual and  $\mathbf{t} = (t_1, t_2, \dots, t_{n_p})$  is the best phenotype vector for a given environment, and we call  $\mathbf{t}$  *target phenotype vector*.

## 4 Environmental perturbation and hierarchical modular structure

### 4.1 Environmental perturbation

The environmental perturbation is expressed by changing the target phenotype vector. For generation  $s$ ,  $i$ -th

component of the target phenotype vector is given as follows for  $i = 1, 2, \dots, n_p$ :

$$t_i(s) = \prod_{j=1}^m a_j(s)^{\alpha_j^i} \quad (j = 1, \dots, m) \quad (8)$$

where  $a_j(s)$  is an *environmental factor*, which varies depending on generation  $s$ . Parameter  $\alpha_j^i$  is the Boolean representation of the influence of  $a_j$  to  $t_i$ . Let  $T_j$  be the set of components influenced by environmental factor  $a_j$ . We call  $T_j$  a *functional module*. For instance, a hand is useful only when the fingers are in good proportion. The environmental perturbation may require larger or smaller hands but the proportion of the fingers will remain the same. In this example, the sizes of the fingers are considered to form a functional module. Parameters  $\alpha_j^i$  is set, so that the following relations hold:  $T_j \subseteq T_k$  or  $T_j \supseteq T_k$  or  $T_j \cap T_k = \emptyset$ .

## 4.2 Hierarchical modular structure

The functional module in the components of phenotype vector  $\mathbf{p}$  is the set  $P_j$  of components corresponding set  $T_j$  of target phenotype vector components. We say that gene  $g_i$  *regulates* functional module  $P_j$ , when all the components in functional module  $P_j$  are changed at the same rate and all the other components stay unchanged by the mutation in amplification coefficient  $x_i$  of gene  $g_i$ . We say a genetic networks has *hierarchical modular structure* when for any functional module there is one or more genes which regulate it.

## 5 Method of simulation

Each parameter was set as follows.

Table 1: The value of each parameter

Number of gene $n_g$	7
Dimension of phenotype vector $n_p$	5
Max generation	500000
Population size	800
Offspring size	3200
Standard deviation in Gaussian mutation	0.15
Probability of mutation of each gene	0.08
Conditional probability that mutation of each causal coefficient of the gene occurs when mutation occurs in the gene	0.03

We prepared three environmental factors of  $a_1(s)$ ,  $a_2(s)$ , and  $a_3(s)$ . The environmental factors change respectively as shown in Fig.4, with the period of 306 generation.

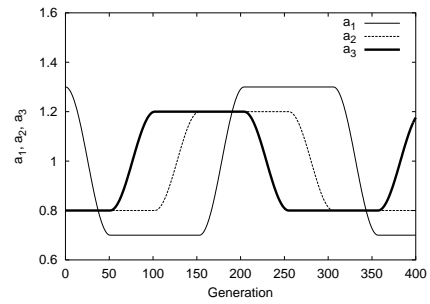


Fig. 4: Transition of the environmental factors

We put  $\alpha_1^1, \alpha_1^2, \alpha_1^3, \alpha_1^4, \alpha_1^5, \alpha_2^1, \alpha_2^2, \alpha_2^3, \alpha_3^4, \alpha_3^5$  equal to 1 and the other equal to 0 in Eq. (8). Then Eq. (8) rewritten as:

$$t_1, t_2, t_3 = a_1(s)a_2(s) \quad (9)$$

$$t_4, t_5 = a_1(s)a_3(s) \quad (10)$$

In this case, there were three function modules,  $P_1 = \{p_1, p_2, p_3, p_4, p_5\}$ ,  $P_2 = \{p_1, p_2, p_3\}$ , and  $P_3 = \{p_4, p_5\}$ .

## 6 Result of simulation

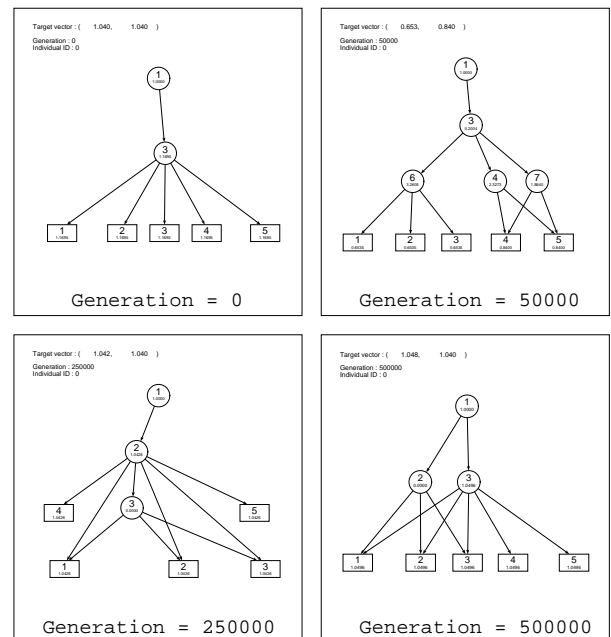


Fig. 5: Evolution of the genetic networks

Fig.5 represents the evolution of genetic networks. The circles represent genes, the rectangles phenotype components. The arrows the direction of the regulation of the gene(or phenotype). In each circle (rectangle) the gene (phenotype) number is displayed, under which the amplification coefficient (phenotype vector component) is. In this figure all the genes that have

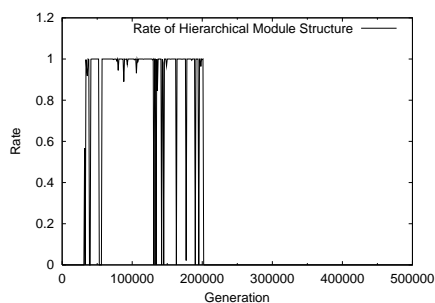


Fig. 6: The rate of the individuals which have the hierarchical modular structure

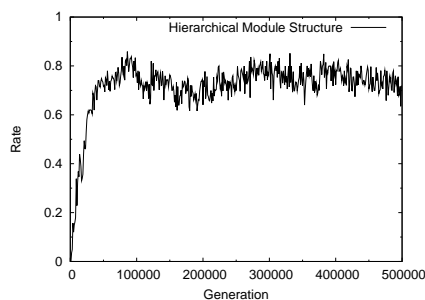


Fig. 7: The ensemble average of the rate of the individuals which have the hierarchical modular structure

no influence on phenotype are omitted, for simplicity. The individual with the highest fitness is displayed. Fig.6,7 are plots of the rate of individuals which have the hierarchical modular structure on genetic networks. Fig.6 is the result of the trial shown in Fig.5, and Fig.7 is an ensemble average of 50 trials.

## 7 Discussion

According to Fig.5, at generation 0 the fittest individual doesn't have hierarchical modular structure, but at generation 50,000 an individual with such structure appeared. However, the fittest in the generations 250,000 and 500,000 didn't show hierarchical modular structure. According to Fig.6, the genetic networks obtained the hierarchical modular structure at about the 40,000th generation. However, The structure collapsed after 200,000th generation.

In Fig.7, we can see that the genetic networks evolves to the hierarchical modular structure at the probability of about 70% at the 50,000th generation.

Our results indicates that the individual which have the genetic networks with the hierarchical modular structure are in an advantageous position in the environmental perturbation. If the genetic networks are without hierarchical modular structure, it needs simultaneous adaptive mutation in all genes in one functional module to follow the environmental perturbation. But such mutations are extremely rare.

## 8 Conclusion

We modeled the simplified genetic networks, defined the hierarchical modular structure, and simulated the evolution by GA on the environmental perturbation. The simulation results show that the environmental perturbation possibly gave rise to the evolution of the hierarchical modular structure.

## References

- [1] Akam, M. (1987) The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* 101:1-22
- [2] Ingham, P.W. (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* 335:25-34
- [3] Scott, M.P.; O'Farrell, P.H. (1986) Spatial programming of gene expression in early *Drosophila* embryogenesis. *Annu. Rev. Cell Biol.* 2:49-80
- [4] Erwin, D.H., Davidson, E.H. (2009) The evolution of hierarchical gene regulatory networks. *Nat Rev. Genet.* 10, 141-148
- [5] Erwin, D.H., Davidson, E.H. (2006) Response to comment on "Gene Regulatory Networks and the Evolution of Animal Body Plans". *Science* 761c, 313
- [6] He, J., Deem, M.W., Hierarchical evolution of animal body plans, *Dev. Biol.* (2009), doi: 10.1016/j.ydbio.2009.09.038
- [7] Barabási, A.L, Zoltán, N.O. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5,101-113
- [8] Somogyi R, Sniegowski CA (1996) Modeling the complexity of genetic networks: understanding multigene and pleiotropic regulation. *Complexity* 1: pp45-63
- [9] H. De Jong (2002) Modeling and Simulation of Genetic Regulatory Systems: A literature Review, *Journal of Computational Biology*, 9, pp67-103
- [10] H. De Jong, R. Lima (2005) Modeling the Dynamics of Genetic Regulatory Networks: Continuous and Discrete Approaches, *Lecture Notes in Physics*, 671, pp304-340
- [11] Savageau, M.A. (1976) Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology. Addison-Wesley, Reading, MA.
- [12] Holland, J. (1975, 1992) Adaptation in Natural and Artificial Systems, The University of Michigan, and MIT Press