

# 동역학 시스템의 진화적 학습에 의한 초파리 발생과정 모델링

이제근<sup>01,2</sup> 남진우<sup>1,2</sup> 정제근<sup>1,2</sup> 장병탁<sup>1,2,3</sup>

서울대학교 생물정보학 협동과정<sup>1</sup>

서울대학교 바이오정보기술연구센터<sup>2</sup>

서울대학교 컴퓨터공학부<sup>3</sup>

{jkrhee<sup>0</sup>, jwnam, jgjoung, btzhang}@bi.snu.ac.kr

## Modeling Embryonic Development in Drosophila by Evolutionary Learning of Dynamical System

Je-Keun Rhee<sup>01,2</sup> Jin-Wu Nam<sup>1,2</sup> Je-Gun Joung<sup>1,2</sup> Byoung-Tak Zhang<sup>1,2,3</sup>

Graduate Program in Bioinformatics, Seoul National University<sup>1</sup>

Center for Bioinformation Technology (CBIT), Seoul National University<sup>2</sup>

School of Computer Science and Engineering, Seoul National University<sup>3</sup>

### 요 약

초파리 초기 발생과정은 gap 유전자, pair-rule 유전자, polarity 유전자의 세 가지 유전자 그룹에 의해서 조직화 된다. Gap 유전자들에 의해 pair-rules 유전자들의 발현이 조절되며, 이들에 의해 결국 polarity 유전자들의 발현을 조절함으로써, 정확한 위치에서 각 기관의 형성을 유도한다. 특히 분열 14단계에서는 pair-rule 유전자 중의 하나인 eve 유전자의 발현이 조절되는데, eve 유전자는 배아의 분할의 줄무늬를 형성시키는 유전자에 해당된다. 본 논문에서는 eve 유전자의 발현조절자인 hunchback, giant, kruppel, bicoid의 gap 유전자들로 구성된 조절 네트워크를 S-system을 이용하여 모델링하였다. 이를 통해 각 유전자들의 발현 데이터로부터 파라미터들을 진화 연산을 통해 예측하고, 각 유전자들의 발현에 대한 시뮬레이션 결과를 보여준다. 예측된 결과와 실제 데이터의 비교는 전체적으로 패턴이 서로 유사함을 보여주고 있다.

### 1. 서 론

초파리의 초기 발생은 다른 고등동물의 발생과 구별된다. 가장 큰 특징 중 하나는 발생시 분할(segmentation)을 이룬다는 것이다. 각 분할은 성체의 기관 위치를 정확히 유도하게 되며, 머리, 가슴, 배의 큰 단위 뿐 아니라, 다리, 날개 등의 위치까지 정확히 분화시키게 된다. 초파리의 초기 발생을 담당하는 gap 유전자들은 자신들의 발현을 유도하고, 또한 pair-rule 유전자들의 발현을 유도함으로써 배아(embryo)의 분할을 형성하게 된다[1]. 이러한 gap 유전자는 Kruppel (Kr), bicoid (bcd), hunchback (hb), giant (gt), knirps (kni)로 구성되며, pair-rule 유전자들의 전사조절인자(transcription factor)로써 작용한다[2]. pair-rule 유전자들은 even-skipped 유전자 (eve), odd-skipped 유전자 (odd)등으로 구성된다. 여기서 eve 유전자는 홀수 번째 띠에서 색을 발현시키며, odd 유전자는 짝수 번째 띠에서 색을 발현시키게 된다[3]. 띠 발현은 배아상에 gap 유전자들의 분포에 따라 pair-rule 유전자의 상류지역(upstream)의 유전자 발현을 촉진하는 영역(enhancer)에 작용하여 특정 위치에서의 pair-rule 유전자의 발현을 유도함으로써 이루어진다. 결국 eve 유전자는 홀수 번째 7개의 띠를, odd 유

전자는 짝수 번째 7개의 띠를 형성하게 된다(그림 1).

최근 시스템 생물학(systems biology)에서는 boolean 네트워크에서부터 임의형태의 구체화된 미분방정식까지 다양한 방법으로 네트워크의 모델링이 시도되었는데, 이러한 기법들 중 S-system은 동적인 역학 관계를 잘 표현할 수 있는 풍부한 구조를 제공하며, 지금까지 몇몇 S-system에 의한 네트워크 분석은 비교적 성공적인 결과들을 보여주고 있다[4, 5].

본 연구에서는 발생초기에서 gap 유전자와 pair-rule 유전자의 단백질 농도 데이터를 이용하여, 최종적으로 pair-rule 유전자의 발현 양상을 시뮬레이션 하고자한다. 이를 위해서 S-system과 진화 연산(evolutionary computation)을 적용하여, gap 유전자의 전사 조절 네트워크의 파라미터를 추정하게 되며, 최종적으로 eve 유전자의 발현 양상을 시뮬레이션 한다.

### 2. 실험 방법

#### 2-1. 실험 데이터 및 네트워크 구성

유전자 발현 데이터는 각 유전자에 해당하는 단백질의 항체(antibody)를 생산하여 면역 형광(immunofluorescence) 방법을 통해 발생하는 형광의 양을 통해 얻는다. 데이터

는 <http://flyex.ams.sunysb.edu/flyex/>에 각 분열시기별 (cleavage10~cleavage14-8, 총 12 time point), 각 위치별 (Anterior~Posterior, 약 2200 A-P point)로 정리되어 있다.

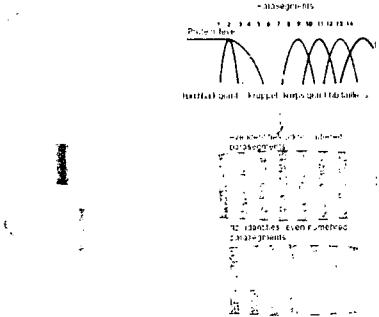


그림 1. gap 유전자와 pair-rule 유전자들의 농도 분포에 의한 배아 분할 모습

본 연구에서는 gap 유전자의 발현조절이 현저하게 나타나며 eve 유전자의 분할 현상이 나타나는, 14-1부터 14-8의 단계의 시간 단계 데이터를 사용하였다. 또한 A-P 축을 중심으로 100의 상대적인 값으로 표시하고, 각 단위에서 가장 큰 대표값을 잡아 그 단위 위치에서의 발현량으로 정하였다.

그림 2는 gap 유전자와 eve 유전자의 단순화된 유전자 조절 네트워크(regulatory network)를 보여주고 있다. 본 연구에서는 그림 2의 구조를 바탕으로 실험하였다.

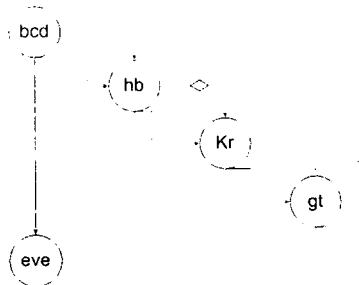


그림 2. 본 논문에서 사용된 유전자 조절 네트워크. 상각형 화살표는 활성화(activation), 원형 화살표는 억제(inhibition)를 나타낸다.

### 2-2. S-System을 이용한 모델링과 진화 연산을 통한 파라미터 예측

각 위치에서 시간에 따른 발현패턴들을 유전자 네트워크로 모델링하기 위하여 S-system을 이용한다. S-system은 다음과 같이 비선형 미분방정식(nonlinear differential equation)의 집합이다[4].

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^N X_j^{g_{ij}} - \beta_i \prod_{j=1}^N X_j^{h_{ij}} \quad (1)$$

여기서,  $X_i$ 는 상태 변수이고  $N$ 은 네트워크의 컴포넌트 수이다. 즉 본 실험에서는  $X_i$ 는 각 유전자의 발현 정도이고,  $N$ 은 네트워크를 구성하고 있는 유전자의 수를 의미하게 된다.  $\alpha_i$ 와  $\beta_i$ 는 속도 상수(rate constant)에 해당

하는 파라미터이며,  $g_{ij}$ 와  $h_{ij}$ 는 반응 차수(kinetic order)에 해당하는 지수 파라미터이다.

S-system을 통한 유전자 네트워크 구조의 분석은 파라미터 학습 과정으로 생각되어질 수 있다. 여기서는 진화 알고리즘을 이용하여 파라미터들  $\alpha_i$ ,  $\beta_i$ ,  $g_{ij}$ ,  $h_{ij}$ 를 학습한다. 진화 알고리즘에 있어서 적합도 함수는 식 (2)와 같은 목적함수를 최소화하는 것이다.

$$\tilde{f} = \sum_{i=1}^N \sum_{t=1}^T \left( \frac{X_{i,cal,t} - X_{i,exp,t}}{X_{i,exp,t}} \right)^2 \quad (2)$$

여기서  $X_{i,exp,t}$ 는  $i$  번째 유전자의 시간  $t$ 에 관찰된 발현량이고  $X_{i,cal,t}$ 는 미분방정식에 의해 얻어진 예측된  $i$  번째 유전자의 시간  $t$ 의 발현량이다. 그리고  $T$ 는 관찰된 샘플링 수이다. 이 목적함수는 각각의 시간에서 예측한 발현량과 실제 발현량의 차이를 계산한다. 하지만 여기서는 시간뿐만 아니라 공간까지 고려해야하기 때문에 이러한 목적함수는 재정의 되어야 한다.  $C$ 개의 공간에 대한 데이터가 주어졌을 때 최종적인 목적함수는 식 (3)과 같이 정의된다.

$$F = \frac{1}{C} \sum_{i=1}^C \tilde{f}_i \quad (3)$$

여기서 공간에 대한 위치수는 100이지만 S-system의 파라미터를 학습할 때 상당한 시간이 요구된다. 그래서 실제로 학습할 때는  $C$ 를 적은 수로 제한한다.

### 3. 실험 결과

#### 3-1. 파라미터 추정 결과

진화 연산을 이용한 파라미터 추정에서 파라미터  $\alpha_i$ 와  $\beta_i$ 의 초기값 범위는  $[0, 10]$ 으로,  $g_{ij}$ 와  $h_{ij}$ 는  $[-4, 4]$ 으로 설정하였다. 그리고 진화알고리즘에 있어서 개체 수는 10,000으로, 교차연산(crossover)의 확률은 0.8로 설정하였다. 진화알고리즘으로 학습되어 결정된 파라미터 값은 표 1과 같다.

	$X_1$	$X_2$	$X_3$	$X_4$
$\alpha$	8.87	0.35	1.76	8.60
$g_{i1}$	0.0	0.0	0.0	0.0
$g_{i2}$	-3.00	1.50	2.96	0.0
$g_{i3}$	0.20	0.0	0.0	5.54
$g_{i4}$	9.55	0.0	0.0	0.0
$\beta$	1.64	19.21	4.24	6.20
$h_{i1}$	3.77	0.0	0.0	0.0
$h_{i2}$	0.0	9.61	0.0	0.0
$h_{i3}$	0.0	0.0	1.43	0.0
$h_{i4}$	0.0	0.0	0.0	1.74

표 1. 진화 연산을 이용한 학습을 통해 계산된 S-system의 파라미터 추정값 ( $X_1$ : eve  $X_2$ : hb  $X_3$ : Kr  $X_4$ : gt  $X_5$ : bcd)

#### 3-2. 모델링 결과

그림 3은 실제 발현량의 데이터와 S-system을 통하여 예측된 그래프를 보여주고 있다. 예측된 결과에서는 8개의 시간 포인트와 36개 공간 포인트로 학습한 다음, 시간상 8 포인트, 공간상의 96포인트를 예측하였다.

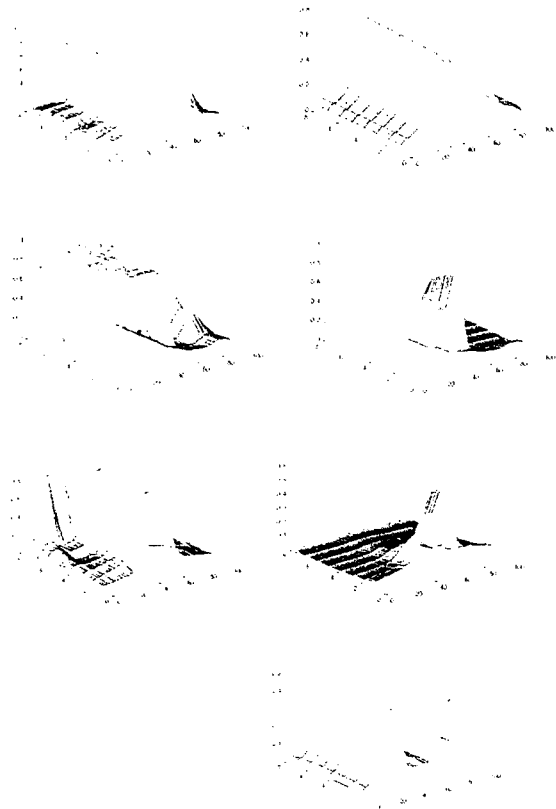


그림 3. 시뮬레이션 결과, 왼쪽의 그림은 실제 데이터에 대한 그래프이며, 오른쪽 그림은 S-system을 이용하여 예측된 결과이다. 위로부터 차례로 eve, hb, Kr, gt에 대한 결과이다. 각 축은 각각 공간, 시간, 발현량을 나타낸다.

eve 유전자의 경우 초파리의 배아에서 출수 번째 띠를 나타내는 유전자이다. 즉 발생이 진행되어 가면서 띠의 개수가 점점 늘어나게 되어 마지막 14-8번째에는 eve 유전자에 의해 7개의 봉우리가 나타나게 됨을 알 수 있다. eve 유전자의 발현 패턴 예측은 몇 개의 봉우리가 보여지고는 있지만 명확하게 구분되지는 않는다. 또한 시간의 변화에 따르는 eve 유전자의 발현 패턴 변화 또한 불명확해 보인다.

hunchback (hb) 유전자의 경우 배아의 전단부(anterior)에서 많은 발현량을 보인다. 또한 후단부(posterior) 쪽에서도 작은 봉우리를 계속적으로 보임을 알 수 있다. 시뮬레이션상의 hb 유전자는 실제 발현 패턴과 비교적 유사한 형태를 보인다. 전단부 쪽에 전체적으로 발현량이 많으며, 후단부 쪽에도 역시 발현량이 많은 부분이 나타난다. 하지만 실제와는 달리 극단적으로 발현량이 '0'에 가까운 부분도 나타나고 있다.

Kruppel (Kr) 유전자는 전체적으로 배아의 중앙 영역에서 많은 발현량을 보인다. 시뮬레이션 결과에서 Kr 유전자는 초기에는 실제와 유사하게 나타난다. 하지만 시간의 진행에 따라 발현량이 '0'으로 떨어진다.

giant (gt) 유전자의 발현 패턴을 Kr 유전자의 발현 패턴과 비교해보면 Kr 유전자가 많이 발현되는 위치에서 giant 유전자의 발현량이 낮다는 것을 알 수 있다. 따라서 giant 유전자는 Kr 유전자에 의해 발현이 억제된다는 것을 알 수 있다. 시뮬레이션 상에서 gt 유전자의 경우 전체적으로 실제와 유사한 경향성을 보인다. Kr 유전자가 많이 발현되는 중앙 부위에서는 발현량이 낮고, 그 주위로 높은 발현량을 보인다.

#### 4. 결 론

본 논문에서는 초파리 초기 발생에 관여하는 gap 유전자들과 eve 유전자의 단백질 발현 데이터를 이용하여 S-system에 기반하여 조절 네트워크를 모델링하고 이를 시뮬레이션 하였다.

S-system은 각 분자가 최종적으로 steady-state로 가는 것을 가정으로 하기 때문에, 변동(oscillation)이 있는 발현데이터의 모델링은 어려운 문제이다. 본 연구에서 다른 초파리 초기발생에 관여하는 유전자의 단백질 발현 데이터는 최종 발생단계가 아니기 때문에 steady state라고 볼 수 없다. 이러한 점은 S-system을 이용한 시뮬레이션에서 정확한 결과를 가져다주지 못했다. 또한, 초파리 초기 발생에 관여하는 유전자는 본 연구에서 다뤘던 유전자들 외에도 많은 유전자가 존재하기 때문에, 실제와 유사한 파라미터 예측과 시뮬레이션 결과를 얻기가 쉬운 일이 아님을 예상할 수 있다. 그럼에도 불구하고, gt, bcd, hb 등의 시뮬레이션 상에서의 발현 양상이 실제와 유사하게 나타나고 있음은 고무적이라 할 수 있다. 후속 연구로, 공간상의 주어진 데이터 포인트만큼의 S-system을 각기 학습하는 방법을 생각해 볼 수 있으며, 이때 C개의 S-system이 학습할 수 있게 된다. 또한, 각각의 규칙을 서로 다르게 생성해 줄 수 있는 방법도 생각해 볼 수 있을 것이다.

#### 감사의 글

이 논문은 과학기술부 국가지정연구실 사업(NRL)에 의하여 지원되었음

#### 참고문헌

- [1] S. Small, A. Blair, and M. Levine. Regulation of two pair-rule stripes by a single enhancer in the drosophila embryo. *Developmental Biology*, 175, 314-324, 1996.
- [2] D. Stanojevic, S. Small, and M. Levine. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science*, 254, 1385-1387, 1991.
- [3] J. Reinitz and D. H. Sharp. Mechanism of eve stripe formation. *Mechanisms of Development*, 49, 133-158, 1995.
- [4] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi; and M. Tomita. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, 19, 643-650, 2003.
- [5] D. Tominaga, N. Koga, and M. Okamoto. Efficient numerical optimization algorithm based on genetic algorithm for inverse problem. *In Proceedings of the Genetic and Evolutionary Computation Conference*, 251-258, 2000.