Evolution *in silico* of genes with multiple regulatory modules

Alexander Spirov
Computer Science and CEWIT
SUNY Stony Brook
The Sechenov Institute of Evolutionary Physiology & Biochemistry,
St.-Petersburg, Russia
The Content of the Talk

• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)

• Our Approach: Evolution in silico

• Test Cases, the *hunchback* & *even-skipped* Genes
  – Representing Gene Regulation
  – Constraints on in silico Evolution

• Modeling of *hb* and *eve* evolution
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain

• Comparison with the Known Evolutionary Biology of segmentation genes
Genes with Multiple CRMs

There is an extensive literature in vertebrate and invertebrate systems, of separable, apparently redundant enhancers (or semi-redundant enhancers, or enhancers capable of driving similar or overlapping patterns) within a single gene (reviewed in [Barolo, 2011]).

New evidence indicates that the embryonic patterns of all of the Drosophila segmentation genes are encoded by multiple elements.
Early fruit fly embryo

Strips 3/7, 2, 4/6, 1/5

CRM binding sites

even-skipped gene, pair-rule class
• Gap genes are defined by the effect of a mutation in that gene, which causes the loss of contiguous body segments, resembling a gap in the normal body plan.

• Pair-rule genes are defined by the effect of a mutation in that gene, which causes the loss of the normal developmental pattern in alternating segments.
• **Evolution of Genes with Multiple CRMs**
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)

• **Our Approach: Evolution in silico**

• **Test Cases, the *hunchback* & *even-skipped* Genes**
  – Representing Gene Regulation
  – Constraints on *in silico* Evolution

• **Modeling of *hb* and *eve* evolution**
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain

• **Comparison with the Known Evolutionary Biology of segmentation genes**
Evolution of segmentation: pair-rule genes

In **long germ insects** (e.g. *Drosophila*) all segments are specified (via segmentation pattern) almost simultaneously within the blastoderm (i.e. prior to gastrulation). In **short germ insects** (e.g., grasshoppers) only segments of the head are specified in the blastoderm, whereas the remaining segments of the thorax and abdomen form progressively from a posterior growth zone after gastrulation.
Mechanisms of segmentation

3 CRMs – 3 domains

Simultaneously
Blastoderm (syncytium) stage

Clock & wave-front model

Progressively
Growth zone stage
Evolution from short- to long-germ embryogenesis

Evolution

1 CRMs – 1 domain

CRM

domain

Evolution

2 CRMs – 2 domains

Evolution

3 CRMs – 3 domains
• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)
• Our Approach: Evolution in silico
• Test Cases, the hunchback & even-skipped Genes
  – Representing Gene Regulation
  – Constraints on in silico Evolution
• Modeling of hb and eve evolution
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain
• Comparison with the Known Evolutionary Biology of segmentation genes
Evolution of Genes with Multiple CRMs: CRM-Domain Correspondence

A simple working hypothesis might be that there is a one-to-one correspondence between a CRM and a domain:

*one CRM, one domain.*

Some classic molecular genetic studies of *pair-rule* genes do show such one-to-one correspondence.
Pair-rule genes:
Single- vs. multiple-stripe CRMs

**One CRM, one domain**

The single-stripe element

**One CRM, multiple domains**

The 7-stripe elements of run, ftz and odd contain significant input from the maternal and gap genes, just like the stripe-specific elements.

**One CRM, two domains**

The dual-stripe element

Schroeder MD, Greer C, Gaul U., Development. 2011 ;138:3067-78
Evolution of Genes with Multiple CRMs: Co-linearity Principle (CRM & Domain Order of Appearance)

• Many segmentation genes have no evident correlation between domain appearance and CRM appearance.

• For some cases, such as the HOM-C and HOX gene clusters, the correlation is very striking, suggestive of a co-linearity principle, in which the order of expression domains reflects the order of the CRMs on the DNA.
Co-linearity Principle (CRM & Domain Order of Appearance)
Evolution constrained by Co-linearity Principle
Constraints on evolution of segmentation

• CRM, domain correspondence
• Co-linearity Principle
One CRM, one domain correspondence

Evolution of segmentation
Control (?): parallel search

Evolution of segmentation

Without explicit intermediate cases
One CRM, multiple domains

Evolution of segmentation
Co-linearity Principle

Evolution of segmentation
• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)
• **Our Approach: Evolution *in silico***
• Test Cases, the *hunchback* & *even-skipped* Genes
  – Representing Gene Regulation
  – Constraints on *in silico* Evolution
• Modeling of *hb* and *eve* evolution
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain
• Comparison with the Known Evolutionary Biology of segmentation genes
Our Approach: Evolution *in silico*

- Initial population
- Population of models
  - Evaluation of models
  - Mutation of models
  - Reproduction of models / recombination
  - Selection of models
- Elimination of weakest

Overview of evolutionary computation approach
Gene expression model: string representation

Symbolically, CRM level (in BS):

***B K B G G B K B H G B K***...***N H H/N N H H N H K H H H***

Element 1

Element 2

Symbolically, in (octal) numbers:

0 0 0 0 1 7 4 4 0 1 7 2 4 0 1 0 0 0...0 0 0 3 2 2 3 3 2 2 3 2 1 2 2 2 0 0 0

Element 1

Element 2
PDE (reaction-diffusion) model

At each in silico evolution generation, candidate strings are used to solve a reaction-diffusion model of \(hb\) gene expression. Expression of the gene, \(C \equiv [Hb]\), under control of a given CRM is quantitatively described by the following reaction-diffusion equation:

\[
\frac{dC}{dt} = D \frac{\partial^2 C}{\partial x^2} + R\sigma \left( \sum_{i=1}^{n} S_i - h \right) - \lambda C,
\]

where \(S_i\) is the strength of the \(i\)-th activator BS, \(n\) is the number of the activator BSs in a given CRM, \(D\) is a diffusion coefficient, \(h\) represents regulatory input from ubiquitous factors, and \(\lambda\) is a decay coefficient.

The strength \(S_i\) is a sum of three terms: the local concentration \(A_i\), the short-range co-activation term, and the short-range repression (quenching) term:

\[
S_i = A_i + \alpha_i \left( \sum_{k=1}^{m} A_k \right) - \sum_{j=1}^{l} R_j,
\]

where \(A_k\) is local concentration of \(k\)-th activator, \(R_j\) is local concentration of \(j\)-th repressor, \(\alpha_i\) is the co-activation coefficient, \(m\) is the amount of the neighbor activator BSs and \(l\) is the amount of the neighbor repressor BSs. \(\sigma(x)\) is a sigmoid regulation-expression function.
The 3-step algorithm to sum the activation strengths for a given activator BS, taking into account both repression via quenching and co-activation from neighboring BSs. We assume that both repression and co-activation are short-range, limited to three neighboring sites.

1) Local BS strengths are tallied;
2) neighboring activation is added (co-activation);
3) neighboring repression is added (quenching).

1) \( \text{Strength}_{a(i)} \approx A_i \).

2) \( \text{Strength}_{a(i)} \approx A_i + \alpha_i (A_{i-3} + A_{i-2} + A_{i+1} + A_{i+3}) \).

3) \( \text{Strength}_{a(i)} = A_i + \alpha_i (A_{i-3} + A_{i-2} + A_{i+1} + A_{i+3}) - (R_{i-1} + R_{i+2}) \).

If \( \text{Strength}_{a(i)} < 0 \) then \( \text{Strength}_{a(i)} = 0 \).
How the model works

3) Strength $a(i) = A_i + \alpha_i(A_{i-3} + A_{i-2} + A_{i+1} + A_{i+3}) - (R_{i-1} + R_{i+2})$.

\[
\frac{dC}{dt} = D \frac{\partial^2 C}{\partial x^2} + R \sigma \left( \sum_{i=1}^{n} S_i - h \right) - \lambda C
\]
Evolution of models,
But not the set of parameters (solutions)

\[
\frac{dC}{dt} = D \frac{\partial^2 C}{\partial x^2} + R\sigma \left( \sum_{i=1}^{n} S_i - h \right) - \lambda C,
\]

One-CRM case

\[
\frac{dC}{dt} = D \frac{\partial^2 C}{\partial x^2} + R\sigma \left( \sum_{i=1}^{n} S_i^1 - h \right) + \sigma \left( \sum_{j=1}^{n} S_j^2 - h \right) + \sigma \left( \sum_{k=1}^{n} S_k^3 - h \right) - \lambda C,
\]

Three-CRM case
• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)
• Our Approach: Evolution in silico
• Test Cases, the hunchback & even-skipped Genes
  – Representing Gene Regulation
  – Constraints on in silico Evolution
• Modeling of hb and eve evolution
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain
• Comparison with the Known Evolutionary Biology of segmentation genes
**hunchback gene (3 CRMs)**

- **P2** proximal
- **P1** oogenesis element
- distal
- shadow distal

**Intensity, a.u.**

![Graph showing intensity variations with labeled regions](image-url)
even-skipped gene (5 CRMs)
Evolution of population of the models

Initial population

Population of models → Evaluation of models → Reproduction of models / recombination → Selection of models → Elimination of weakest

Mutation of models
• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)
• Our Approach: Evolution \textit{in silico}
• Test Cases, the \textit{hunchback} & \textit{even-skipped} Genes
  – Representing Gene Regulation
  – Constraints on \textit{in silico} Evolution
• \textbf{Modeling of \textit{hb} and \textit{eve} evolution}
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain
• Comparison with the Known Evolutionary Biology of segmentation genes
Parallel search (control) is the most effective

In the case, the complete hb pattern is used for fitting. CRMs are free to evolve, but the number of CRMs is constrained (1, 2 or 3). There is no requirement for CRM-domain correspondence: solutions are allowed in which only one (of the three) CRMs controls formation of all domains; in which CRMs are one-to-one with domains; or in which the CRMs share control of domains.
Constrained evolution: 
*One CRM, one domain correspondence*

It can be roughly a hundred times easier to find one CRM governing formation of all three domains of the *hb* pattern, than to find three separate CRMs independently controlling separate *hb* domains (one CRM – one domain).

This suggests that genes which show multiple domain control by single CRMs may have evolved quite quickly.
Typical solutions have only one CRM (sometimes two) controlling the expression pattern. In only a few percent of the solutions are all three CRMs involved in the patterning.

Multiple and redundant control

A solution of the *hb* gene problem with all three CRMs participating in patterning the anterior domains.

A) CRM-domain diagram.
B) Solution of the *hb* gene problem for each of three CRMs, with redundancies outlined by the dashed boxes.
It takes substantially more time to evolve a CRM for each domain

• In general, there is abundant evidence that evolution of autonomous CRMs is responsible for many cases of morphological evolution.

• The computational approach outlined here will help to understand the correspondence between CRM evolution and domain appearance (morphological effect).

• As shown here, different cases of the CRM-domain dependence lead to different evolutionary costs, and help to understand how a number of regulatory motifs have arisen in evolution, and what their particular advantages might be.
Co-linearity Principle is Computationally Expensive
Evolution in silico under the co-linearity constrain

• This is the colinear case, where the one-to-one CRM-domain correspondence includes both order of domain appearance and CRM order on the DNA.

• The evolutionary computations with this scenario are quite slow, but are also quite reproducible (with a small standard deviation).

• Biological examples of this are not ubiquitous, but the cases which do display this, such as the HOX cluster, are quite important and famous.
• Evolution of Genes with Multiple CRMs
  – CRM-Domain Correspondence
  – Co-linearity Principle (CRM & Domain Order of Appearance)
• Our Approach: Evolution in silico
• Test Cases, the hunchback & even-skipped Genes
  – Representing Gene Regulation
  – Constraints on in silico Evolution
• Modeling of hb and eve evolution
  – Single CRMs tend to control multiple domains when there are no constraints on CRM-domain correspondence
  – It takes substantially more time to evolve a CRM for each domain
• Comparison with the Known Evolutionary Biology of segmentation genes
Single- vs. multiple-stripe CRMs: pair-rule genes (5 genes)

One CRM, one domain (10)

The single-stripe element

One CRM, two domains (12)

The dual-stripe element

One CRM, multiple domains (>3)

Schroeder MD, Greer C, Gaul U., 2011

The 7-stripe elements of run, ftz and odd contain significant input from the maternal and gap genes, just like the stripe-specific elements.

Pair-rule genes tend to have all three classes of CRMs
The colinear case: Drosophila HOX genes

iab-8 cis-regulatory elements, which direct expression in parasegment (ps) 13, are located both upstream and downstream of the Abd-B transcription unit. The other iab regulatory domains that control Abd-B expression are also indicated (iab-5, iab-6, and iab-7). These latter domains regulate expression in ps10, ps11 and ps12, respectively. It has been proposed that the 3¢ Abd-B cis-regulatory DNA contains a series of insulators, which are indicated by the red ovals.

Zhou and Levine, 1999
Evolutionary colinearity

- Biological examples of this are not ubiquitous, but the cases which do display this, such as the HOX cluster, are quite important and famous.
- Biological examples such as the HOX cluster are extremely well conserved through evolution, comparing between species.
- It is possible that the reproducibility of this search is associated with the stability and conservation of these regulatory structures.
Acknowledgements

• David M. Holloway, Math Dept, British Columbia Institute of Technology, Canada

• This work supported by the National Institutes of Health, 2-R01 GM072022.