molecular
systems
biology

Deriving structure from evolution: metazoan segmentation

Paul François[1,3,*], Vincent Hakim[2,3] and Eric D Siggia[1,3]

[1] Center for Studies in Physics and Biology, The Rockefeller University, New York, NY, USA and [2] Laboratoire de Physique Statistique, CNRS UMR 8550, Ecole Normale Supérieure, Paris Cedex, France
[3] Order of authors is alphabetical.
* Corresponding author. Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA. Tel.: + 1 212 327 8835; Fax: + 1 212 327 8544; E-mail: pfrancois@rockefeller.edu

Segmentation is a common feature of disparate clades of metazoans, and its evolution is a central problem of evolutionary developmental biology. We evolved *in silico* regulatory networks by a mutation/selection process that just rewards the number of segment boundaries. For segmentation controlled by a static gradient, as in long-germ band insects, a cascade of adjacent repressors reminiscent of gap genes evolves. For sequential segmentation controlled by a moving gradient, similar to vertebrate somitogenesis, we invariably observe a very constrained evolutionary path or funnel. The evolved state is a cell autonomous 'clock and wavefront' model, with the new attribute of a separate bistable system driven by an autonomous clock. Early stages in the evolution of both modes of segmentation are functionally similar, and simulations suggest a possible path for their interconversion. Our computation illustrates how complex traits can evolve by the incremental addition of new functions on top of pre-existing traits.
Molecular Systems Biology 18 December 2007; doi:10.1038/msb4100192
Subject Categories: simulation and data analysis; development
Keywords: evolution; modeling; oscillations; segmentation; somitogenesis

Introduction

Although only a minority of the around 30 bilaterally symmetric animal phyla display a segmental anterior–posterior body plan, examples can be found in each of the three major clades of bilaterians (Tautz, 2004). Thus the question, did segmentation appear once in some putative 'urbilaterian' and was then lost from most phyla or did it evolve several times? The great variety of embryonic patterning mechanisms found among insects, annelids, and vertebrates argues for the later alternative. The appearance of similar genes in the segmentation networks of diverse phyla (Peel *et al*, 2005) raised the question of whether this was independent recruitment or common ancestry. In this study, we simulate the evolution of transcriptional regulatory networks subject to a fitness function favoring segments. We find the repeated emergence of networks sharing a common structure that resembles what is observed in nature.

The segmented bodies of insects and the somites of vertebrates are the two best characterized segmentation networks. Insects divide into long-germ band species such as

Drosophila that create all segments in parallel, and short-germ species such as *Tribolium* that segment the head and thorax first and pattern the abdomen later, together with the elongation of the posterior growth zone (Liu and Kaufman, 2005b). Segmentation is under the control of morphogen gradients, for example, the anterior activator *bicoid* (*bcd*) in *Drosophila* that is established maternally. Segments of long-germ band insects are carved from broad domains of activation by short-range repression due to gap genes (Rivera-Pomar and Jäckle, 1996). The genes responsible for these patterns can change, most famously *bcd* is not present in other long-germ band insects such as mosquito, though its functions of anterior activation and translational repression of *caudal* (*cad*) mRNA are subsumed by other genes (Liu and Kaufman, 2005b).

The segmented structures of the vertebrate body, such as the vertebrae and skeletal muscles arise from the somites, blocks of paraxial mesoderm cells partitioned into anteroposterior domains by segmentation of the pre-somitic mesoderm (PSM). Somites are produced from the posterior growth zone or tail bud (Pourquié, 2003), through the interaction of time-periodic gene expression or a 'clock' with a morphogen wavefront that

moves along with posterior growth, as insightfully proposed 30 years ago by Cooke and Zeeman (1976). Candidates for the morphogen include signaling molecules such as FGF8 (Dubrulle and Pourquié, 2004) and Wnt3a (Aulehla *et al*, 2003). While many cycling genes are known, including a homolog to the *Drosophila* pair-rule gene *hairy* (Dequéant *et al*, 2006), much less is known about the segmentation clock than other genetic oscillators such as the circadian clock. The transition between the oscillating posterior PSM and the more anterior segmented part is also still under investigation (Morimoto and Saga, 2005).

We have previously designed an evolutionary algorithm (François and Hakim, 2004) that generates networks of interacting genes and proteins that implement a predefined function in a single cell. The simulations successfully produced circuits very similar to circadian clocks when subject to selection that favored periodic behavior. We evolved segmentation *in silico* using a suitably adapted version of this procedure to investigate its possible origin and mechanisms. In the following, we show that for a static morphogen gradient, segmentation networks evolve toward cascades of repressors. When the gradient becomes dynamic, they evolve toward molecular networks implementing a 'clock and wavefront' mechanism in a new way.

Evolution *in silico*

The evolutionary procedure is similar to the one described in François and Hakim (2004). The strategy is to evolve genetic networks by repeated rounds of selection, growth, and mutation.

The algorithm evolves a population of transcriptional regulatory networks. We represent the binding of a transcription factor A to the regulatory region of gene encoding a protein P by a Michaelis function, with an equilibrium binding constant, A_P and a Hill coefficient n to represent cooperative effects. A gene is described by a maximum protein production rate, S_P, a time delay, τ_P, (to represent all the steps between transcription and the appearance of new protein (Lewis, 2003; Monk, 2003)) and a decay rate, δ_P. The rate of protein production defined by the regulatory interactions in Figure 1 is described mathematically as follows:

$$\frac{dP}{dt} = \left(\max\left[\frac{A_1^{n_1}}{A_1^{n_1} + (A_{1P})^{n_1}} , \frac{A_2^{n_2}}{A_2^{n_2} + (A_{2P})^{n_2}} \right] \frac{S_P}{1 + (R/R_P)^{n_3}} \right)$$

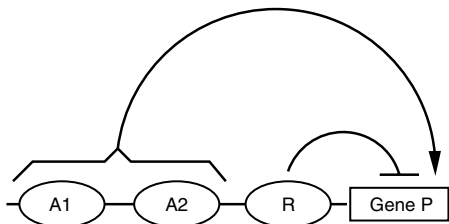$$(t - \tau_P) - \delta_P P \qquad (1)$$

Figure 1 Transcriptional regulation of a prototypical gene P. In the example shown, the expression of gene *P* is activated by proteins *A1* and *A2* and repressed by protein *R*. The rate of production of the corresponding protein is described mathematically by Equation (1).

When multiple activators are present as in Figure 1, their combined output is defined by the maximum of their activities. Repressors are combined multiplicatively. This is analogous to using an OR between activators and an AND between repressors in discrete systems. Gene regulation is vastly more complex than our simple model, yet if one were to measure protein production in response to changing levels of activators and repressors, the trends would be captured by our model.

Each network is evaluated in an 'embryo', which consists of a linear array of cells (typically 100) sharing the same genetic network. The embryo is one-dimensional: there is only one cell for each anteroposterior position x along the embryo. Cells are independent and cannot communicate with each other. A protein G provides positional information to these cells. We impose the dynamics of this 'effector protein', which we assume to reflect the effect of a morphogen gradient. For simplicity, we call G the morphogen in the following. The only difference between cells in the same embryo is through their exposure to G, which is a prescribed function of x and the same in all embryos. A reporter gene E, whose product defines the segments, is introduced into the system. Selective pressure acts on the final profile of this protein.

Integration of the gene network equations in each embryo then allows us to rank the genetic networks according to a predetermined scoring or fitness function that counts the number of boundaries in the profile between low and high states of E (see below).

A hundred networks are followed in parallel. The first generation network for all the simulations is the morphogen G and uncoupled from it the reporter E. At each step of the algorithm, after the fitness function is computed, the 50 best networks are retained; then each is copied and mutated. Mutations can create or destroy genes, add or remove transcriptional regulatory linkages (e.g. G activates E), and change any parameter. After the mutation step, the entire process is iterated. A 'generation' is one iteration of this selection/growth/mutation process, and corresponds to many generations in a real organism since we are only concerned with mutations in the one network under study. An overview of the algorithm is provided in Figure 2.

The fitness function is a crucial ingredient in our evolution of segmentation. Darwin originally proposed that a complete eye could have evolved from a simple sensor if gradual changes increased the ability to derive information about the environment from light and were inheritable (Darwin, 1859; Nilsson and Pelger, 1994). This argument potentially applies to any complex trait; if there is a sequence of small steps each increasing the fitness that lead to the desired trait, then evolution can be rapid. It thus appears reasonable to assume that segmentation arose from the potential benefit of more specialization and modularity in the body plan.

To test this scenario in our computations, segmentation was simply defined as the spatial alternation between high and low concentrations of the test protein E. We then chose as a fitness function the number of boundaries (transitions in either direction between low and high) or 'steps' in the concentration of E. This mathematically represents the general evolutionary pressure that an increase in the number of segments is beneficial without making arbitrary assumptions such as insisting on predefined levels of E or rewarding only seven

segments in the case of *Drosophila*. These more specific functions not only contravene the notion of genericity and gradualism, they also greatly slow the evolution, which reduces to a random search through a vast space with no clues indicating proximity to the desired state (see Figures S1–S3 in Supplement).

The second crucial ingredient in our evolutionary computation is the choice of the dynamics of the morphogen *G*. We explore two possibilities. In a first introductory set of simulations, we take *G* to be a static gradient as in the fly (Figure 3). For the more complicated case of sequential segmentation, we take a uniformly translating (sweeping)
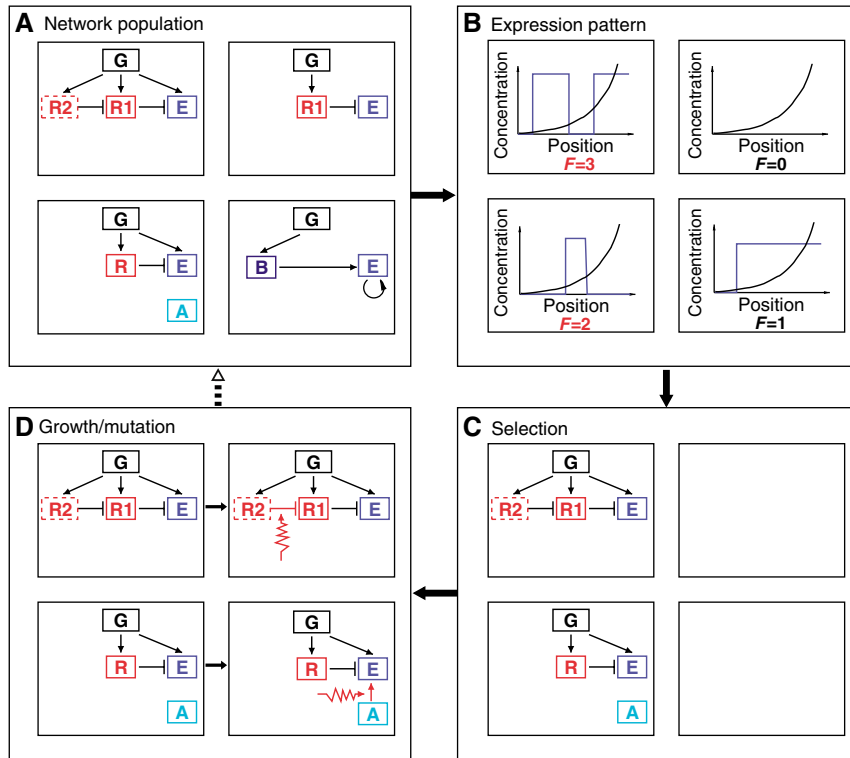

Figure 2 Overview of one generation of the evolution algorithm. At each generation, the algorithm evolves the network collection (**A**). First, in each embryo, we solve for the expression pattern of a reporter protein *E* in the presence of a morphogen *G* as a function of position (**B**). The fitness, F, is defined as the number of jumps between high and low values of *E*. Half the networks of lowest fitness are discarded (**C**) and replaced by mutated copies of the fittest ones (**D**). This produces the starting network collection for the next generation (dashed arrow).
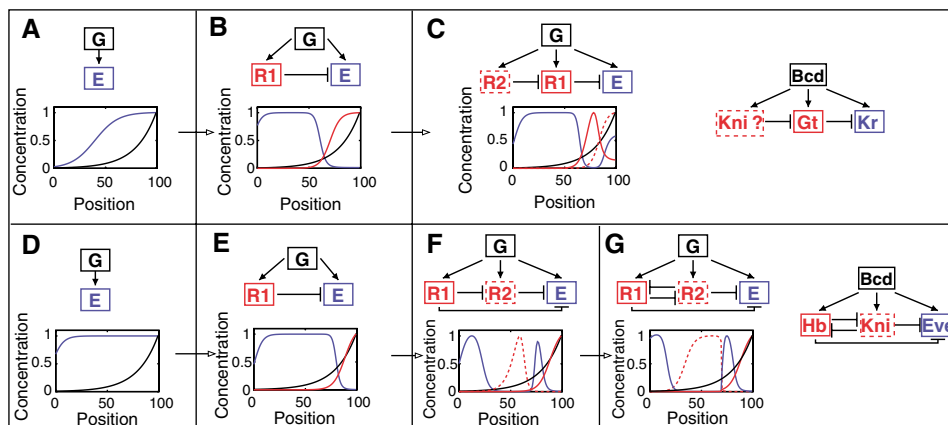

Figure 3 Evolution of two segmentation networks in a static morphogen gradient. Two different evolutionary pathways are displayed (**A**–**C**, **D**–**G**). Successive stages run from left to right and show both the network and the spatial profile of the proteins. Note that the first two stages are common to both evolutionary trajectories. The morphogen *G* is depicted in black, the protein *E* defining the segments is in blue, and the repressors *R1* and *R2* are in red (dashed lines represent the last to be added). Concentrations have been normalized by their maximum value for plotting purposes. See the text for details.

exponential step (Figure 4B–E) decreasing from a constant concentration to a concentration that is effectively 0 (i.e. smaller than relevant binding affinities). This mimics formation of a signaling gradient upon exit from a growth zone, similar to what happens in the PSM during somitogenesis. Therefore, it is not necessary to model growth of the tail bud, and for computational convenience, we just extend the (constant) morphogen over all the posterior cells. The number of boundaries in E is counted after the morphogen front has run down the row of cells and is close to zero everywhere.

Results

Static gradient: *in silico* evolution produces cascades of repressors

To test the relevance of the method on a simplified example, we first considered the *in silico* evolution of segmentation under the control of a static morphogen gradient. Two typical examples of simulations are displayed in Figure 3A–C and D–G with the major stages of their evolutionary pathway.

Both evolutionary pathways begin in the same way: after few generations, the gene E is put under the control of the graded morphogen, and is transcriptionally activated when the morphogen G exceeds some level (Figure 3A and D).

A few generations later, the next event creates R_1, a repressor of E, only active in the region where the morphogen is highest. As a result, this creates a second boundary (E from high to low) in the region of R_1 activity (Figure 3B and E).

The evolutionary processes then fork for the later generations. In the first case, a second repressor R_2 is created, which represses R_1 in regions of yet higher concentration of the morphogen. This creates a new region of E activity, where R_2 lifts R_1 repression (Figure 3C), so that E is expressed in two broad domains.

In the second case, a second repressor R_2 is created by neutral evolution, at an intermediate concentration of the morphogen G, and is repressed by R_1 in regions of higher concentration of G. Then, R_2 represses E, which increases the fitness by splitting the broad stripe of Figure 3E into two smaller stripes, delimited by the regions of high repressor (Figure 3F). Later in the evolution, mutual repression between repressors R_2 and R_1 is selected to sharpen the boundary of one stripe (Figure 3G and Supplementary Figure S4).

Each stage in this evolutionary process creates new boundaries by repressing or derepressing E. In the example of Figure 3C, the evolutionary pathway creates a cascade of repressors. Cascades were often found by the algorithm, since cross-regulating repressors create adjacent regions of expression of E. Such networks, based on a cascade of repressors, are suggestive of some part of the actual *Drosophila* gap-gene regulation network as discussed below.

Dynamic gradient: spontaneous evolution of a 'clock and wavefront' mechanism

Since the algorithm succeeded in creating some simple networks controlling localized expression of genes in a static gradient, we allowed the gradient to sweep to create a dynamic morphogenetic profile.
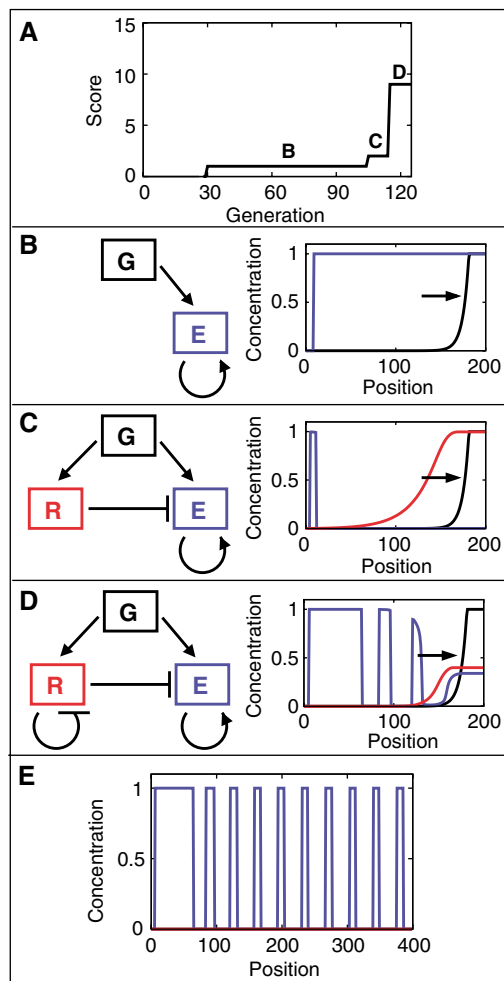

Figure 4 Stages in the evolution of sequential segmentation for a morphogen gradient moving to the right. (**A**) Evolution of the best network fitness as a function of the number of generations. (**B**–**D**) The letters correspond to the network topologies and protein profiles in the three subsequent panels following the conventions in Figure 3. (**E**) Final profile produced by network of (D) for twice as long an embryo showing regular spacing of the stripes. See the text for details and Supplementary Information for other examples.

An example of a network with the corresponding evolutionary pathway is displayed in Figure 4.

The first stage in the evolution of the sequential segmentation simulation (Figure 4) is always activation of E at some intermediate level G_E of the initial gradient and then positive feedback of E onto itself so that high values are self sustaining while low values decay to zero. This generates one step after the morphogen gradient runs over the cells and G equals 0 everywhere (Figure 4B). Bistability is the first functionality to evolve.

The next event is to create a repressor, R, of E that itself is induced only for morphogen levels G greater than G_E. This can occur either by placing a pre-existing repressor of E under the control of G or adding repressor function to some other gene controlled by G.

Several conditions should be met for R to create a second boundary of E expression without destroying the first one. First, to preserve the boundary from low to high E expression:

- *R* should turn on later than *E* while the morphogen is sweeping past, so that *E* has time to autoactivate before being repressed. Typically *R* turns on for higher value of *G* than *E*, but the opposite could be true if *R* turns on much more slowly than *E*.
- the high value of *E* in the bistable system should be stable (persist) for low values of *R*. Weak induction of *R* by the exponential tail of *G* should not destroy the stable high value of *E*.

The creation of a second boundary, from high to low *E* concentration, requires that full expression of *R* should repress *E* expression strongly enough to prevent its autoactivation, thereby creating a limited domain where *E* concentration is high.

When these conditions are satisfied, *R* drives *E* to zero well ahead of the morphogen front, while in the morphogen-decaying tail, an island of autoactivated *E* remains (Figure 4C and Supplementary Figure S5).

Subsequent evolution creates, however, a more spectacular improvement (Figure 4D) and a large fitness increase. In addition to repressing *E*, the protein *R* becomes a transcriptional repressor of its own gene. This negative feedback loop produces temporal oscillations in *R* so long as *G* is large enough to induce *R*. The oscillations in *R* produce islands of *E* for the same reasons as before. The oscillation phase is encoded in a binary way and rendered permanent by the bistable dynamics of *E* as *G* decreases to 0. Figure 5 illustrates this mechanism at the level of individual cells. As a consequence of this process (and with the exception of the very first segments), all segments are of the same size as can be seen in Figure 4E: segment size is simply given by the product of the velocity of the front times the period of the clock. Bistability further implies sharp boundaries between domain of high and low *E* expression. There are only two possible concentrations for *E* at steady state and therefore *E* concentration profile jumps abruptly from one cell to the next.

After several dozen evolution experiments for a variety of mutation parameters, the only networks that produced more than one or two stripes always displayed a clock coupled to a bistable system (see several examples in Supplementary Figures S6–S11). However, some variability in the detailed sequence of events was observed, most often in the final stage (*R* represses itself) when sometimes multiple repressors would appear, as in Figure 6 (evolutionary pathway is displayed in Supplementary Figure S8). Each repressor creates a single additional step in *E*, until finally an oscillator (similar in principle to the engineered 'repressilator'; Elowitz and Leibler, 2000) is created for a big increase in fitness. An interesting property of this repressilator network is that delays are not required for it to work efficiently and to produce many stripes (even though delays are needed along the evolutionary pathway leading to it).

To investigate the generality of these conclusions and the ubiquity of the path leading to the clock and wavefront model, we did 20 additional simulations with identical parameters and stopped them after a fixed number of generations (Table I). Five of the simulations consisted of a clock coupled to a bistable system. In one, the oscillations were damped, and in network 17 a repressilator-like clock evolved (as illustrated in
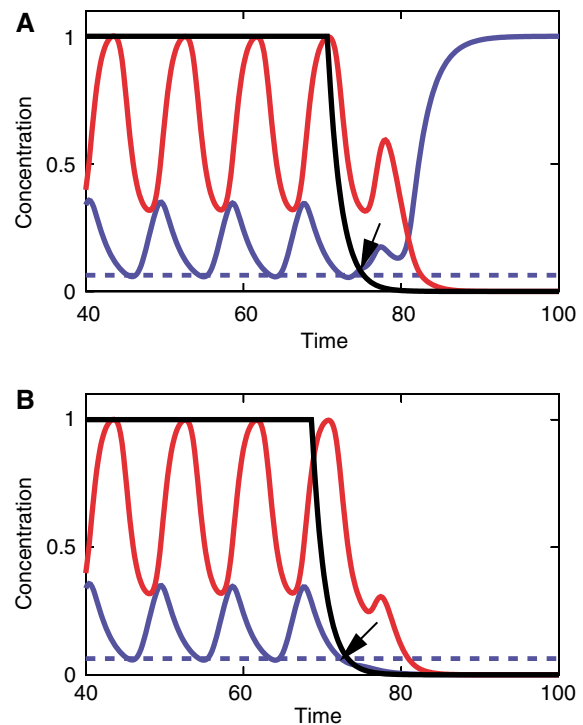

Figure 5 Binary encoding of the phase of the oscillation by bistability. Creation of high (**A**) and low (**B**) values of the segmentation marker *E* in two cells by the coupled effects of oscillatory and bistable dynamics. The network corresponds to Figure 4D and the colors and scalings are identical. While the morphogen *G* is high, *E* is high and oscillates in response to the clock variable *R*. As time passes, *G* decreases and at a given moment (black arrow), it can no longer significantly activate gene *E*. The cell fate is determined by the concentration of *E* at this particular moment relative to a threshold E^0 (shown by a dashed line). E^0 is the (unstable) fixed point (for $G=R=0$) that separates protein concentrations $E>E^0$ converging to the high state of *E* expression, from smaller values that end in the low state of *E* expression. In (A), *E* is high enough at the arrowed time so that *G* and *R* can disappear while leaving $E>E^0$. In (B), the concentration of *E* at the arrowed time is under the threshold E^0.
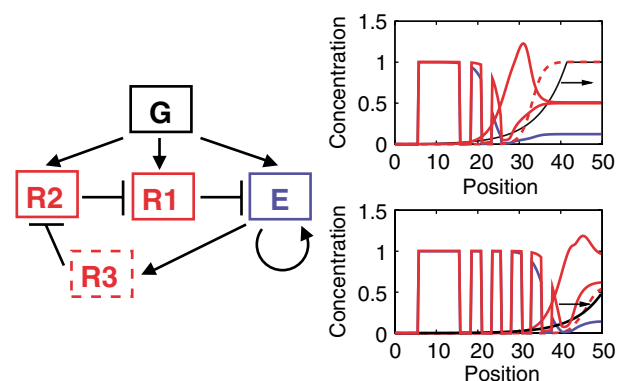

Figure 6 An alternative pathway from repression to oscillations. This network evolved from the network in Figure 4C and replaces the network in Figure 4D. A triplet of repressors creates oscillations by a mechanism similar to the synthetic network created in Elowitz and Leibler (2000). Two protein expression profiles are shown on the right for the same network at different times with the same conventions as in Figure 4.

Supplementary Figure S12), slightly different from the one displayed in Figure 4. Interestingly, all networks evolved at least a bistable system, and several others one or more

repressors, for example, network 7 had a cascade of repressors and two others stopped at the stage displayed in Figure 2C.

When we increased the rate of topology-changing mutations relative to those that changed parameters, the evolutionary path remained the same but the fraction of high fitness 'clock and bistable' systems increased. Finally, we chose random mutation rates and repeated the evolution using the same dynamic morphogen and the same fitness function. Again we found only networks with topologies coupling clock and bistable systems or ones on the path leading to it, as detailed in Table II. In this sense, there is a common path or fitness funnel that all systems traverse as their fitness increases.

Table I Outcome of a typical run of 20 random evolutionary simulations with the same evolutionary parameters

Index	Fitness	Function and topology
0	30	Clock + bistable system, Figure 4D
1	1	Bistable system, Figure 4B
2	1	Bistable system, Figure 4B
3	6	Damped clock + bistable system, Figure 4D
4	44	Clock + bistable system, Figure 4D
5	1	Bistable system, Figure 4B
6	1	Bistable system, Figure 4B
7	7	Cascade of two repressors + bistable system, Supplementary Figure S12
8	1	Bistable system, Figure 4B
9	1	Bistable system, Figure 4B
10	2	Repressor + bistable system, Figure 4C
11	1	Bistable system, Figure 4B
12	1	Bistable system, Figure 4B
13	35	Clock + bistable system, Figure 4D
14	1	Bistable system, Figure 4B
15	1	Bistable system, Figure 4B
16	2	Repressor + bistable system, Figure 4C
17	6	Clock ('repressilator') + bistable system, Supplementary Figure S12
18	1	Bistable system, Figure 4B
19	1	Bistable system, Figure 4B

Fitness was the number of boundaries, with a term to prevent apparition of traveling wave, as explained in the text and the Supplement. Here, relative probability of mutating the parameters of the network was six times the probability of changing the topology of the network. Each evolutionary simulation was stopped after 400 generations.

Discussion

The evolved cascades of repressors resemble some modules of *Drosophila* segmentation network

Both terminal networks in Figure 3 have approximate analogs in selected modules of the *Drosophila* system with E a pair-rule gene such as *even-skipped (eve)*, and the various repressors acting as gap genes. From Figure 3C, *knirps* (*kni*) and *giant* (*gt*) are known to be complementary and mutually repress in the anterior part of the embryo (Kraut and Levine, 1991a,). *gt* defines the anterior border of the central *Krüppel* (*Kr*) domain (Kraut and Levine, 1991b) and possibly splits it off from the weak and dynamic *Kr* anterior domain (e.g. Figure 2 of Wu *et al*, 1998).

The situation in Figure 3G is loosely like *eve* stripes 3–7 or 4–6 in that one module creates two stripes with two repressors, one central *kni* and the other *hunchback* (*hb*) defining the outer boundaries of the pair of stripes (Clyde *et al*, 2003). The analogy is incomplete in that the *eve* stripes are created from a uniform activator, while we have a spacial gradient and the simulation simply uses the limit of the activator domain in place of *hb* to define the anterior boundary of the anterior stripe. Mutual repression seen between R1 (*hb*) and R2 (*kni*) occurs in the evolved system when a small term favoring sharp boundaries is added to the fitness.

The algorithm is therefore able to generate small modules with topology close to gap genes interaction generating localized expression domains. The complete *Drosophila* patterning system involves both anterior- and posterior-anchored morphogens, hence is much more complex than what we have simulated. *Drosophila* segmentation network is also a highly derived and hierarchical system (Peel *et al*, 2005), presumably from more primitive (short germ) insects, so that evolution of a complete long-germ band segmentation system may require steps passing through a dynamic mode of segmentation (see below). The networks presented here only suggest possible building blocks for formation of gap or segmentation gene domains (or anterior patterning in short-germ band insects).

Table II Statistics on the nature and the dynamics of the networks obtained in 100 simulations, after 400 generations

Topology	Void	Bistable		Figure 4C	Clock and wavefront			
		Figure 4B	Other		Figure 4D		Other	
					Sustained	Damped	Sustained	Damped
Number	17	63	2	3	9	4	1	1
Fitness (± s.d.)	0 ± 0	1 ± 0	1 ± 0	3.3 ± 1.1	32 ± 11	21 ± 10	30 ± 0	3 ± 0

To assess the influence of mutation rates, they were chosen at random in different simulations as follows. For each possible mutation, the mutation rate was chosen uniformly in $[r/5, r]$, the non-zero lower bound ensuring that each mutation remained possible. Here, r for creation of a gene or an interaction was 0.1, r for modification of one kinetics parameter was 1, while rate for removal of a gene or condition was enforced to be one half the rate for creation. This table indicates the number of networks of each topology and the average fitness for each type with the standard deviation. 'Void' networks are network with fitness 0 where there is no relevant connections between G and E. In general, these networks correspond to simulations where the probability of creation of nodes or links is too low for any network to be built. Bistable networks are networks of fitness 1; most of the networks share the topology displayed in Figure 4B, two other networks built a positive feedback loop via another gene activating E. Three networks where found to be similar to the network in Figure 4C; in two of them, parameter adjustment by evolution succeeded in producing two stripes. Finally, 15 networks display clock and wavefront dynamics; 13 have the same topologies of the network of Figure 4D (including four damped oscillators). The two alternate topologies add an additional gene but implement the same logical operator as in Figure 4D. Therefore, even though the mutation parameters are chosen at random, the statistics are close to the one displayed in Table I.

More uniform domains of expression could be obtained if a subsidiary term were added to our primary fitness function favoring that feature (see for an example, Supplementary Figure S13). Uniform stripes in the fly may be a vestige of its presumed short-germ band ancestor where uniform segments arise naturally from a clock and wavefront system. Simply counting boundaries as we do, leaves the spacing, shape, and intensity of the segmentation marker completely unregulated. In reality, selection will further play on these features while leaving the topology of the network unchanged.

Finally, note that the network topology displayed in Figure 3B (and also in Figure 4C) is a very general mechanism to create a limited zone of expression of a protein E at intermediate concentration of an effector G. There are several examples in contexts other than segmentation where one signal activates a protein over a threshold and shuts it down at higher concentration via activation of a repressor. For instance, in *Xenopus* development, the gradient of *Activin* induces *Brachyury* (Xbra) and then *Goosecoid* (Gsc), that in turn represses Xbra (Green, 2002).

Network evolution suggests new models of sequential segmentation

Somitogenesis has long been the subject of quantitative modeling (reviewed in Baker *et al*, 2006). The model suggested by the evolutionary algorithm is characterized by a combination of two features: (1) a cell autonomous clock dependent on the morphogen and (2) an independent bistable system driven both by the morphogen and clock, which we identify as the anterio-posterior somite marker. Most existing models focus on more specific aspects of somitogenesis. Lewis (2003) proposed a clock based on delayed negative feedback among genes in the *hairy/E(spl)* family and demonstrated its robustness against changes in transcription rate and mRNA levels. The evolved network extends this model and suggests that a bistable system may provide a discrete encoding of the phase and sidestep the more subtle task of remembering the continuous phase suggested in Kerszberg and Wolpert (2000). Evidence for some bistability in single cells comes from experiments where elimination of one of several delta paralogs is thought to desynchronize oscillations, and yields a salt and pepper pattern for an anterior somite marker (Jiang *et al*, 2000). We find partition of the anterior PSM into Notch1/Dll1 high and low activity domains suggestive of our bistable 'gene' E (Morimoto and Saga, 2005).

Another model (Baker *et al*, 2006) mathematically describes the behavior in the transition zone by imposing thresholds on the moving morphogen front. This is only part of the problem, and our evolved network by construction gives a full model transforming the prepattern into a static array of segments. Meinhardt (1986) was the first to suggest an explicit model of the complete process and proposed that diffusion or long-range inhibition between cells played an essential role in segmentation. Our model is situated in the opposite limit and is purely cell autonomous.

The sweeping of the morphogen step ties segmentation to the growth of new tail bud. In a static gradient, each segment requires a new inhibitory interaction as can be seen in Figure 3.

Therefore, it is hard to generate more than a few segments. In contrast, in a temporal gradient, the addition of a clock to a series of one or more repressors appears as the highly favored way to create many segments within our simulations.

Thus, it is tempting to speculate that in short-germ insects or in myriapods, a clock will be found to complement the moving morphogen gradient implicit in the elongating posterior growth zone (Peel *et al*, 2005). Recently, a gradient of *mkp3* has also been observed during chick limb bud growth (Pascoal *et al*, 2007a) and cyclic *hairy2* expression has been demonstrated (Pascoal *et al*, 2007b). Interestingly, the period of the limb bud formation clock is much longer than in the segmentation clock. These results suggest that similar clock and wavefront mechanisms, based on different molecular players, may apply to other sequential segmentation processes.

Simple as it is, our evolved model suggests biochemical interactions (such as the need of bistability to fix the pattern) that may be checked experimentally. It admits refinement, such as cell–cell communication, that may be necessary to keep the phase of the different oscillators synchronized and makes the somite boundary straight in the direction perpendicular to the morphogen gradient, and deals with the bilateral symmetry. Such elaborations need not destroy the core cell autonomous dynamics and the basic synchrony imposed on a field of cells by the morphogen will enhance the efficacy of cell–cell communication. Other variants can display additional features of the experiments that did not spontaneously evolve such as the traveling waves posterior to the front (Palmeirim *et al*, 1997; Dequéant *et al*, 2006). Such phase waves could be simulated if one allowed the period of the clock (mainly under control of the transcriptional delay in R in the network displayed on Figure 4) to vary along the oscillatory region, for instance under control of a secondary gradient. Further work is needed to explore more precisely the interplay among the clock, bistability and morphogen interactions in our system.

Our simulations and recent experiments suggest a possible evolutionary path between short- and long-germ insects. In the basal short-germ insect *Oncopeltus*, Liu and Kaufman (2005a) show that the prototype fly pair-rule gene *eve* actually feeds back and activates the anterior gap genes *hb* and *Kr*. The situation is very suggestive of the evolved network in Figure 6. The cascade of repressors is common to what we found for a static morphogen, and *eve* is one target of that cascade. The feedback of *eve* into this cascade (i.e. E activates R_3 in Figure 6) could create an oscillator and thus drive sequential segmentation. The early anterior gap-like expression of *eve* in *Drosophila* could be a vestige of this feedback interaction, which could be a key transition linking short- and long-germ band insects. A recent reassessment of the phylogeny among the major groups of Holometabolous insects (Savard *et al*, 2006) makes the Hymenoptera (which includes the long-germ band wasp *Nasonia* (Lynch *et al*, 2006)) an outgroup with respect to Diptera (which includes long-germ band insects) and Coleoptera (which includes short-germ species such as *Tribolium*). Thus interconversion between these two modes of development has occurred multiple times, and evolutionary simulations suggest a possible pathway for the evolution.

Incremental evolution produces complex networks from a generic fitness function

How can a numerical simulation performed in complete ignorance of the actual mutation rates hope to capture even a cartoon version of actual evolutionary pathways? The key assumption, we believe, is the incremental increase in fitness, and the contingency imposed on subsequent stages of evolution by the earlier ones.

Evolutionary simulations for segmentation under the control of static gradients spontaneously evolved toward cascades of repressors. New segment boundaries arose by the addition of new proteins to the cascade, which restrict the domain of expression of the downstream proteins. Evolution of sequential segmentation seems much more constrained. The first stage that couples bistability to a moving front of G is the only way to create one persistent step. A single cell sees a temporal pulse of G, and from that stimulus has to assume either the high or low state. This defines bistability.

Adding one or more repressors creates additional high/low E boundaries just as it did for a static gradient G. A negative feedback oscillator is then only one mutation away from the repressor itself. Note that repression accomplishes both the regulation of E and the oscillations, so one protein can perform both functions. In conformity with ideas about the evolution of biochemical pathways (Horowitz, 1945), somitogenesis in our simulation evolves by adding functionality upstream to an existing, less fit, process.

The strict serial construction of a network renders the process independent of rates of mutation. Numerically correct rates become important when there are several parallel routes to the same end; then the one realized depends on the rates evolution explores them. A computational reconstruction of evolution is most likely to succeed when real biological networks are built serially. This may only be true on the macroscale that we have modeled, while there may be many ways of making a bistable system and negative feedback oscillator and the one chosen depends on numbers.

Certain of our 'genes' such as the bistable E may be preexisting modules that are just recruited by the morphogen G. The composite nature of our 'genes' does not matter, evolution merely has to stitch pieces together. The relationships among the nodes of our network, such as bistability and negative feedback, could be realized no matter what the actual molecular interactions are.

The choice of fitness is crucial, and it is noteworthy that separate bistable and oscillatory modules evolved in response to a fitness function that only scored the number of boundaries in E. Both observations and theory support our choice of a nonspecific fitness function. Although somite number is a characteristic of an organism, it is notably disassociated from other taxonomic features (Richardson *et al*, 1998). Temperature shock can change somite number (Primmett *et al*, 1988), and local environment can shift the segment number of centipedes (Arthur and Kettle, 2001). Once the topology of the somitogenesis system is established, the number or spacing of somites can be varied with the clock period or front velocity as occurs between the backbone and tail of a mouse (Richardson *et al*, 1998). Our function describes a component of fitness common to all metazoans that favors segmentation. For a particular phylum, there will be additional terms that are hard to deduce *a priori* that favor a particular number and size of segments. Mathematical search is efficient when one can move continuously uphill and arrive at the optimum. Similarly, biological evolution is rapid when positive selection controls the succession of states realized. Our fitness function, since it is parameter free and rewards only 'topological' properties of the solution, provides just such a smooth bias to facilitate, in a natural way, the evolutionary search.

Robustness and evolution of genetic networks

Screening for parameter independence has been used to select among biological networks. The one which realizes the known transformation (taken from an extant organism) for the widest range of parameters is considered as the most plausible (von Dassow *et al*, 2000; Eldar *et al*, 2002).

This implicitly corresponds to a mode of evolution in big jumps, each one creating a complex network (with associated parameters); the first one that works is selected. However, searching through all possible networks takes much longer than building them incrementally. When evolution progresses by small fitness-increasing increments, purifying selection will control deleterious changes in the values of sensitive parameters (see also Galis *et al*, 2002) once there is a workable network topology.

For sequential segmentation, we found an almost unique evolutionary path from an unsegmented ancestor to a species embodying a novel version of the clock and wavefront model of somitogenesis. We term this a fitness funnel and note that it is an argument for convergent evolution. Conversely, the observation of networks with similar topology realized by non-homologous genes is evidence for convergent evolution.

The traditional strategy for modeling a biological system is to start with a network defined by genetics, obtain constants for the interactions (from diverse sources), and then hope it works. However, this strategy does not shed light on the invariant dynamical structure that a particular set of genes implement. This structure is important to understand since it can be implemented by different genes in different species. For instance, rather surprisingly, there appears to be only a small overlap when comparing the genes that oscillate during somitogenesis in the chick and mouse (O Pourquié, private communication). Evolving a network as we have done inverts this procedure. A logically complete dynamical model comes first, lacking all connections to the genes. These models are compact or condensed in comparison to the number of genes implicated in the processes. Thus if correct, multiple genes will impact each model variable, allowing for non-trivial predictions or at least a principled organization of the genes into a network.

If the topology of biological networks can be deduced by evolutionary simulations from a generic fitness function that gives rise to a fitness funnel, then evolution should be viewed more as a learning process than an optimization process. Networks that can be learned quickly are what we observe, even if they are not the global optimum of some fitness function specific to an organism and an environment.

Materials and methods

The evolutionary algorithm used in this study is similar to the one described in François and Hakim (2004). Further details on this are provided in the Supplementary information.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

References

Arthur W, Kettle C (2001) Geographic patterning of variation in segment number in geophilomorph centipedes: clines and speciation. *Evol Dev* **3**: 34–40

Aulehla A, Wehrle C, Brand-Saberi B, Kemler R, Gossler A, Kanzler B, Herrmann BG (2003) Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev Cell* **4**: 395–406

Baker RE, Schnell S, Maini PK (2006) A clock and wavefront mechanism for somite formation. *Dev Biol* **293**: 116–126

Clyde DE, Corado MSG, Wu X, Pare A, Papatsenko D, Small S (2003) A self-organizing system of repressor gradients establishes segmental complexity in *Drosophila*. *Nature* **426**: 849–853

Cooke J, Zeeman EC (1976) A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J Theor Biol* **58**: 455–476

Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, Chapter Difficulties on theory, pp 186–187 John Murray: London

Dequéant ML, Glynn E, Gaudenz K, Wahl M, Chen J, Mushegian A, Pourquié O (2006) A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* **314**: 1595–1598

Dubrulle J, Pourquié O (2004) fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature* **427**: 419–422

Eldar A, Dorfman R, Weiss D, Ashe H, Shilo BZ, Barkai N (2002) Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419**: 304–308

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335–338

François P, Hakim V (2004) Design of genetic networks with specified functions by evolution *in silico*. *Proc Natl Acad Sci USA* **101**: 580–585

Galis F, van Dooren TJM, Metz JAJ (2002) Conservation of the segmented germband stage: robustness or pleiotropy? *Trends Genet* **18**: 504–509

Green J (2002) Morphogen gradients, positional information, and *Xenopus*: interplay of theory and experiment. *Dev Dyn* **225**: 392–408

Horowitz NH (1945) On the evolution of biochemical syntheses. *Proc Natl Acad Sci USA* **31**: 153–157

Jiang YJ, Aerne BL, Smithers L, Haddon C, Ish-Horowicz D, Lewis J (2000) Notch signalling and the synchronization of the somite segmentation clock. *Nature* **408**: 475–479

Kerszberg M, Wolpert L (2000) A clock and trail model for somite formation, specialization and polarization. *J Theor Biol* **205**: 505–510

Kraut R, Levine M (1991a) Mutually repressive interactions between the gap genes giant and Kruppel define middle body regions of the *Drosophila* embryo. *Development* **111**: 611–621

Kraut R, Levine M (1991b) Spatial regulation of the gap gene giant during *Drosophila* development. *Development* **111**: 601–609

Lewis J (2003) Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr Biol* **13**: 1398–1408

Liu PZ, Kaufman TC (2005a) even-skipped is not a pair-rule gene but has segmental and gap-like functions in *Oncopeltus fasciatus*, an intermediate germband insect. *Development* **132**: 2081–2092

Liu PZ, Kaufman TC (2005b) Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evol Dev* **7**: 629–646

Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C (2006) Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature* **439**: 728–732

Meinhardt H (1986) Models of segmentation. In: Bellairs R, Edde DA, Lash JW (eds) *Somites in Developing Embryos*, pp 179–191, Plenum Press, New York and London

Monk NAM (2003) Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delays. *Curr Biol* **13**: 1409–1413

Morimoto M, Saga Y (2005) The Mesp2 transcription factor establishes borders by suppressing notch activity. *Nature* **435**: 354–359

Nilsson D, Pelger S (1994) A pessimistic estimate of the time required for an eye to evolve. *Proc Biol Sci* **256**: 53–58

Palmeirim I, Henrique D, Ish-Horowicz D, Pourquié O (1997) Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**: 639–648

Pascoal S, Andrade RP, Bajanca E, Palmeirim I (2007a) Progressive mRNA decay establishes an mkp3 expression gradient in the chick limb. *Biochem Biophys Res Comm* **352**: 153–157

Pascoal S, Carvalho CR, Rodriguez-León J, Delfini MC, Duprez D, Thorsteinsdóttir S, Palmeirim I (2007b) A molecular clock operates during chick autopod proximal–distal outgrowth. *J Mol Biol* **368**: 303–309

Peel AD, Chipman AD, Akam M (2005) Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Gen* **6**: 905–916

Pourquié O (2003) The segmentation clock: converting embryonic time into spatial pattern. *Science* **301**: 328–330

Primmett DR, Stern CD, Keynes RJ (1988) Heat shock causes repeated segmental anomalies in the chick embryo. *Development* **104**: 331–339

Richardson MK, Allen SP, Wright GM, Raynaud A, Hanken J (1998) Somite number and vertebrate evolution. *Development* **125**: 151–160

Rivera-Pomar R, Jäckle H (1996) From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet* **12**: 478–483

Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, Tettelin H, Lercher MJ (2006) Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res* **16**: 1334–1338

Tautz D (2004) Segmentation. *Dev Cell* **7**: 301–312

von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. *Nature* **406**: 188–192

Wu X, Vakani R, Small S (1998) Two distinct mechanisms for differential positioning of gene expression borders involving the *Drosophila* gap protein giant. *Development* **125**: 3765–3774