



A Segmentation Clock with Two-Segment Periodicity in Insects

Andres F. Sarrazin *et al.*

Science **336**, 338 (2012);

DOI: 10.1126/science.1218256

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 19, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/336/6079/338.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2012/03/07/science.1218256.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/336/6079/338.full.html#related>

This article **cites 27 articles**, 15 of which can be accessed free:

<http://www.sciencemag.org/content/336/6079/338.full.html#ref-list-1>

This article has been **cited by 1** articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/336/6079/338.full.html#related-urls>

This article appears in the following **subject collections**:

Development

<http://www.sciencemag.org/cgi/collection/development>

They are formed in response to the onset of new cold instabilities. They develop and a reorganization of the flow takes place to progressively produce a more triangular distribution. Variations in the shape of the distribution with time are consistent with reconstructions for the past 150 My (12, 14, 15). The shape of the distribution may have evolved from flat-like, when Pangea was barely splitting, to a skewed distribution after the birth of new ridges (15), ultimately transforming to the present-day triangular shape with dispersed continents.

Like the shape of the area-age distribution, the rate of production of new ocean floor (younger than 10 My) in the mantle convection models varies with time (Fig. 3). Fluctuations are moderate—32% and 21% of the mean value for the supercontinent and six-continent cases, respectively—but they can reach 100% at times, doubling or halving the production of new ocean floor. The strongest variations occur on a time scale of 500 million years, which corresponds to the time scale of flow reorganization through the onset of new plate boundaries. The peaks of production are generally correlated with the generation of new plate boundaries and peaks in heat flow (like the configuration in Fig. 2F). The fluctuations are stronger with one continent than with six continents. Many small continents make the flow adopt a smaller wavelength, so that a change in plate organization has a smaller contribution to the total (Fig. 4). The smaller wavelength imposes a higher time-averaged heat flow than for

the supercontinent case (20). The magnitudes and time scales of the computed variations of the production of new ocean floor are comparable to those extracted from seafloor spreading reconstructions (12, 13).

Our models provide a fundamental basis for realistically simulating Earth's mantle convection. Although they have relatively low Rayleigh numbers and simplified parameters for the interior of the mantle, they show that plate-like behavior and the presence of continents are the two necessary ingredients to build a model in which young seafloor is subducted like on Earth. Continents constrain the location and geometry of the downwellings that cool Earth's mantle. When subduction is confined at an ocean-continent boundary, convection forces the subduction of very young seafloor. Such a situation is favored by continental growth and dispersal. The distribution of seafloor age is a primary observation that should be used as a diagnostic when simulating Earth's mantle, predicting the long-term cooling of Earth, the fluctuations of sea level caused by tectonics (21) that ultimately condition climate change on geological time scales.

References and Notes

1. B. Parsons, *J. Geophys. Res.* **87**, (B1), 289 (1982).
2. C. Stein, S. Stein, *Nature* **359**, 123 (1992).
3. S. Labrosse, C. Jaupart, *Earth Planet. Sci. Lett.* **260**, 465 (2007).
4. S. Gaffin, *Am. J. Sci.* **287**, 596 (1987).
5. L. Husson, C. Conrad, *Geophys. Res. Lett.* **33**, L18303 (2006).
6. R. Berner, Z. Kothavala, *Am. J. Sci.* **301**, 182 (2001).

7. N. Flament, N. Coltice, P. Rey, *Earth Planet. Sci. Lett.* **275**, 326 (2008).
8. J. Sclater, C. Jaupart, D. Gelson, *Rev. Geophys.* **18**, 269 (1980).
9. J.-P. Cogné, E. Humler, *Earth Planet. Sci. Lett.* **227**, 427 (2004).
10. D. Rowley, *Geol. Soc. Am. Bull.* **114**, 927 (2002).
11. R. Demicco, *Geology* **32**, 485 (2004).
12. T. Becker, C. Conrad, B. Buffett, R. Müller, *Earth Planet. Sci. Lett.* **278**, 233 (2009).
13. C. Conrad, C. Lithgow-Bertelloni, *Geology* **35**, 29 (2007).
14. R. Müller, M. Sdrolias, C. Gaina, W. Roest, *Geochim. Geophys. Geosyst.* **9**, Q04006 (2008).
15. M. Seton, C. Gaina, R. Müller, C. Heine, *Geology* **37**, 687 (2009).
16. C. Conrad, B. Hager, *J. Geophys. Res.* **104**, 17551 (1999).
17. C. Lee, A. Lenardic, C. Cooper, F. Niu, A. Levander, *Earth Planet. Sci. Lett.* **230**, 379 (2005).
18. See supplementary materials on Science Online.
19. T. Yanagisawa, Y. Hamano, *Geophys. Res. Lett.* **26**, 791 (1999).
20. C. Grigné, S. Labrosse, P. Tackley, *J. Geophys. Res.* **110**, B03409 (2005).
21. R. D. Müller, M. Sdrolias, C. Gaina, B. Steinberger, C. Heine, *Science* **319**, 1357 (2008).

Acknowledgments: We thank R. D. Müller and T. W. Becker for fruitful reviews. Supported by Institut Universitaire de France and ANR grant DynBMO (Dynamique de l'océan de magma basal) ANR-08-JCJC-0084-01 (N.C. and S.L.) and by Crystal2Plate, a FP-7 funded Marie Curie action under grant agreement PITN-GA-2008-215353 (T.R.). Supercomputing resources were provided by ETH and the Swiss Supercomputer Centre (CSCS).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6079/335/DC1
Materials and Methods
Data File and Codes
References (22–27)

13 January 2012; accepted 21 March 2012
10.1126/science.1219120

A Segmentation Clock with Two-Segment Periodicity in Insects

Andres F. Sarrazin,*† Andrew D. Peel,† Michalis Averof‡

Vertebrate segmentation relies on a mechanism characterized by oscillating gene expression. Whether this mechanism is used by other segmented animals has been controversial. Rigorous proof of cyclic expression during arthropod segmentation has been lacking. We find that the segmentation gene *odd-skipped* (*Tc-odd*) oscillates with a two-segment periodicity in the beetle *Tribolium castaneum*. By bisecting embryos and culturing the two halves over different time intervals, we demonstrate that *Tc-odd* cycles with a period of about 95 minutes at 30°C. Using live imaging and cell tracking in green fluorescent protein-expressing embryos, we can exclude that cell movements explain this dynamic expression. Our results show that molecular oscillators represent a common feature of segmentation in divergent animals and help reconcile the contrasting paradigms of insect and vertebrate segmentation.

Many animals generate body segments sequentially from a posterior region known as the growth zone (1, 2). Whether there are common mechanisms underlying this process of segmentation in disparate segmented animals, such as vertebrates, annelids, and arthropods, has been

intensely debated (3–8). A role for molecular oscillators in segmentation was initially proposed on theoretical grounds by Cooke and Zeeman (9). Their “clock and wavefront” model explained how the temporal periodicity of a clock could be translated into a repetitive spatial pattern during sequential segmentation. Subsequent studies showed that oscillating patterns of gene expression sweeping through the growth zone play a key role in vertebrate segmentation (10–13).

A number of studies have indicated that an equivalent segmentation clock may operate in the presegmental zone of arthropods (3–5, 14, 15). These studies revealed changing patterns of gene

expression in the presegmental zone of an insect, a centipede, and a spider. They also showed that disrupting *Notch* signaling, which is an important element of the vertebrate segmentation clock (13, 16), leads to defects in segmentation in some of these species. These results have been interpreted by some researchers as evidence that a common mechanism for segmentation was inherited by vertebrates and arthropods from a segmented common ancestor.

However, several doubts remain regarding this interpretation. First, *Notch* signaling is known to be involved in many other developmental processes, such as specification of the presegmental zone. These diverse functions may provide alternative explanations for segmentation defects (8, 17, 18). Second, cycling expression patterns have been inferred from in situ hybridization stainings on fixed embryos and comparison of similarly staged embryos. However, embryo-to-embryo variation and difficulty in accurately staging embryos (relative to the speed of segment formation) limits the reliability of this approach. Moreover, it has not yet been proven that these dynamic expression patterns reflect intracellular changes in gene expression, rather than cell movements. Thus, there is no rigorous demonstration that cyclic waves of expression are sweeping through the growth zone of arthropods. We use embryo culture and live imaging in the insect *Tribolium castaneum* to address these issues.

Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology Hellas (FORTH), Nikolaou Plastira 100, GR-70013 Heraklio, Crete, Greece.

*Present address: Instituto de Química, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile.

†These authors contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: averof@imbb.forth.gr

We focus on the expression of the pair-rule ortholog *odd-skipped* (*Tc-odd*) during early germ-band elongation. *Tc-odd* is expressed in primary “pair-rule” stripes, corresponding to alternate segments, emerging from the growth zone (19); secondary segmental stripes appear later. Knockdown of *Tc-odd* arrests segmentation (19). We noticed that *Tc-odd* expression in the growth zone and in the overlying amnion varies among similarly staged embryos (Fig. 1). In particular, embryos that have the same number of primary *Tc-odd* stripes in the segmented part of the embryo show differences in the extent and strength of *Tc-odd* expression within the posterior growth zone (e.g., compare panels A' to F' in Fig. 1). These differences are also evident when *Tc-odd* expression is compared to the expression of the *Tribolium* ortholog of *even-skipped* (*Tc-eve*) (fig. S3). *Tc-odd* expression appears dynamic both in the ectoderm, the mesoderm, and the amnion (fig. S4). Taking the extent of amnion closure, the emergence of the secondary (segmental) *Tc-odd* stripes, and the overall shape of the germband as reference points (Fig. 1, A to F), we placed embryos in a putative temporal sequence and subdivided the period that leads to the production of each new *Tc-odd* stripe into three phases, named I, II, and III (see supporting online material and Fig. 1). By comparing these stainings to time-lapse recordings of live embryos (movie S1), using amnion closure as a reference, we roughly estimate that a new primary *Tc-odd* stripe is generated within 40 to 110 min at 30°C.

To address the criticism that different expression patterns may reflect variation among embryos, rather than temporal changes, we need to observe gene expression in the same embryo during successive time points. We achieved this using the approach taken by Palmeirim *et al.* (10) to prove oscillations in the vertebrate presomitic zone. First, we developed a method for *Tribolium* embryo culture, which allows the developing germband (embryo proper) to be dissected away from the surrounding yolk and serosa and to be grown in cell culture medium (Fig. 2). Germbands taken during early segmental stages (phase 3.III) continue to form segments for up to 5 hours in culture, at normal speed, with no overt morphological defects (Fig. 2C). This enabled us to bisect embryos in culture and to sample the expression pattern of *Tc-odd* at different time points in each half-embryo: Immediately after dissection, we cut individual germbands along the midline into halves (Fig. 2B); one half was fixed immediately after bisection ($t = 0$), whereas the other was cultured over different time intervals (15 to 105 min) before fixation. Both halves were then simultaneously subjected to *in situ* hybridization for *Tc-odd*. By comparing expression in these half-embryos, we demonstrate that *Tc-odd* expression in the growth zone changes from high to low levels, and back to high levels again, during production of the third and fourth primary stripes (Fig. 2, D and E). These experiments allow us to define more precisely the

cycling period of *Tc-odd*, which we estimate to be ~95 min at 30°C (fig. S5).

Next, we asked whether the observed changes in *Tc-odd* expression might be explained by cell movements in the growth zone, rather than changes in expression levels within cells. To directly address this, we established live imaging in *Tribolium* embryos and dissected germbands. We generated a transgenic line expressing nuclear-localized green fluorescent protein (GFP) driven by a *Tribolium* ubiquitous promoter (see supporting online material) and developed methods for imaging live *Tribolium* embryos under a spinning-disc confocal microscope. Whole embryos could be kept alive for more than 5 hours, with no apparent defects in morphogenesis, allowing us to visualize morphogenetic processes such as embryo condensation, amnion formation, germband elongation, and segmentation (Fig. 3A and movies S1 and S2). Dissected germbands were viable under a coverslip and could be imaged for 2 to 3 hours, allowing us to track the movements of individual nuclei in the growth zone and elongating germband (movie S3). Although only a small number of nuclei can be tracked reliably over a period of 1 hour at 30°C, this is sufficient to provide a map of cell movements in the early *Tribolium* germband (Fig. 3B). Compared with *Tc-odd* expression in corresponding stages, this map shows that cell movements could not account for the changes in *Tc-odd* expression that we observe in the growth zone

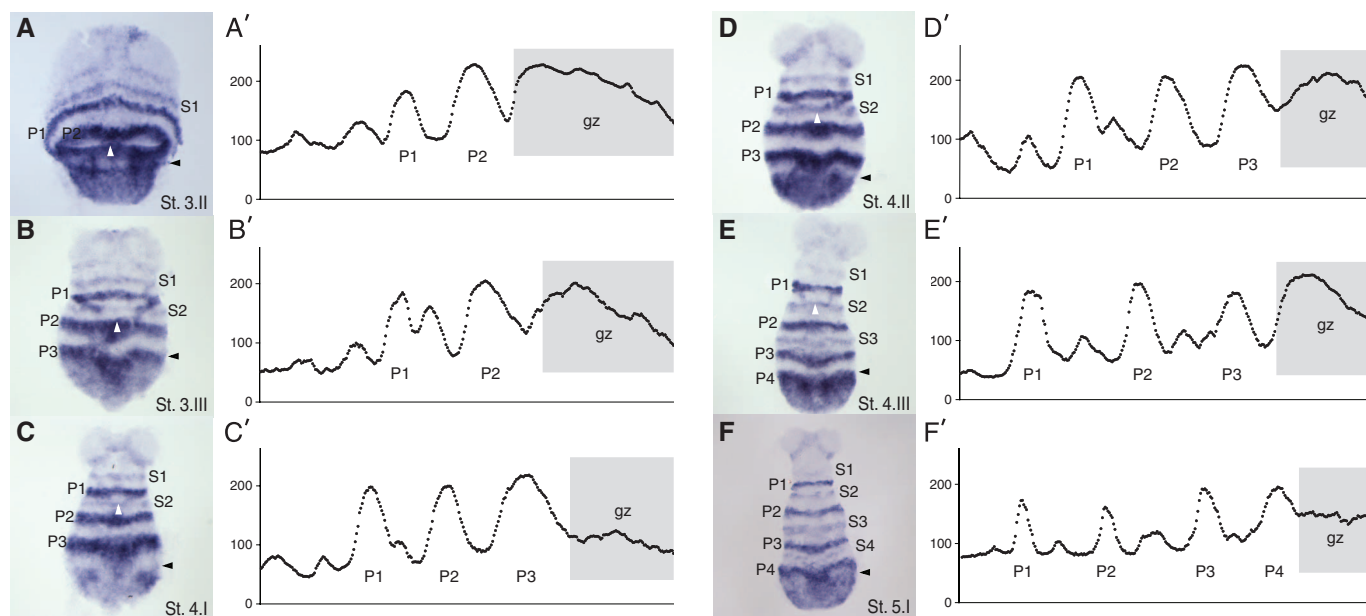


Fig. 1. Expression of *Tc-odd* during early germband elongation. (A to F) *In situ* hybridization for *Tc-odd* in elongating germbands at stages 3.II to 5.I, in ventral view (for staging, see supporting online material). The growth zone is located posterior to the black arrowheads. White arrowheads mark the posterior edge of the amniotic fold. (A' to F') Intensity profiles of *Tc-odd* expression along the anterior-posterior axis of germbands. The intensity of staining was measured on the lateral part of the germbands shown in (A) to (F), encompassing the ectoderm and the amnion (for explanations, see fig. S1). Note the changing profile of *Tc-odd* expression within the posterior growth zone (gray area) during the emergence of new primary *Tc-odd* stripes

from that zone. During phase I of stripe formation [(C) and (F)], the growth zone has low levels of *Tc-odd*, except for two lateral spots of expression that mark the initiation of a new cycle of *Tc-odd* expression. At phase II [(A) and (D)], *Tc-odd* expression extends through the growth zone, such that a posterior domain of high *Tc-odd* expression is established. At phase III [(B) and (E)], *Tc-odd* expression in the posterior-most part of the growth zone is reduced to low levels, whereas cells at the anterior of the growth zone still express high levels, establishing a broad primary stripe. In a typical collection of embryos, each phase is found in roughly equal numbers. P1 to P4, primary stripes; S1 to S4, intercalating secondary stripes; gz, growth zone.

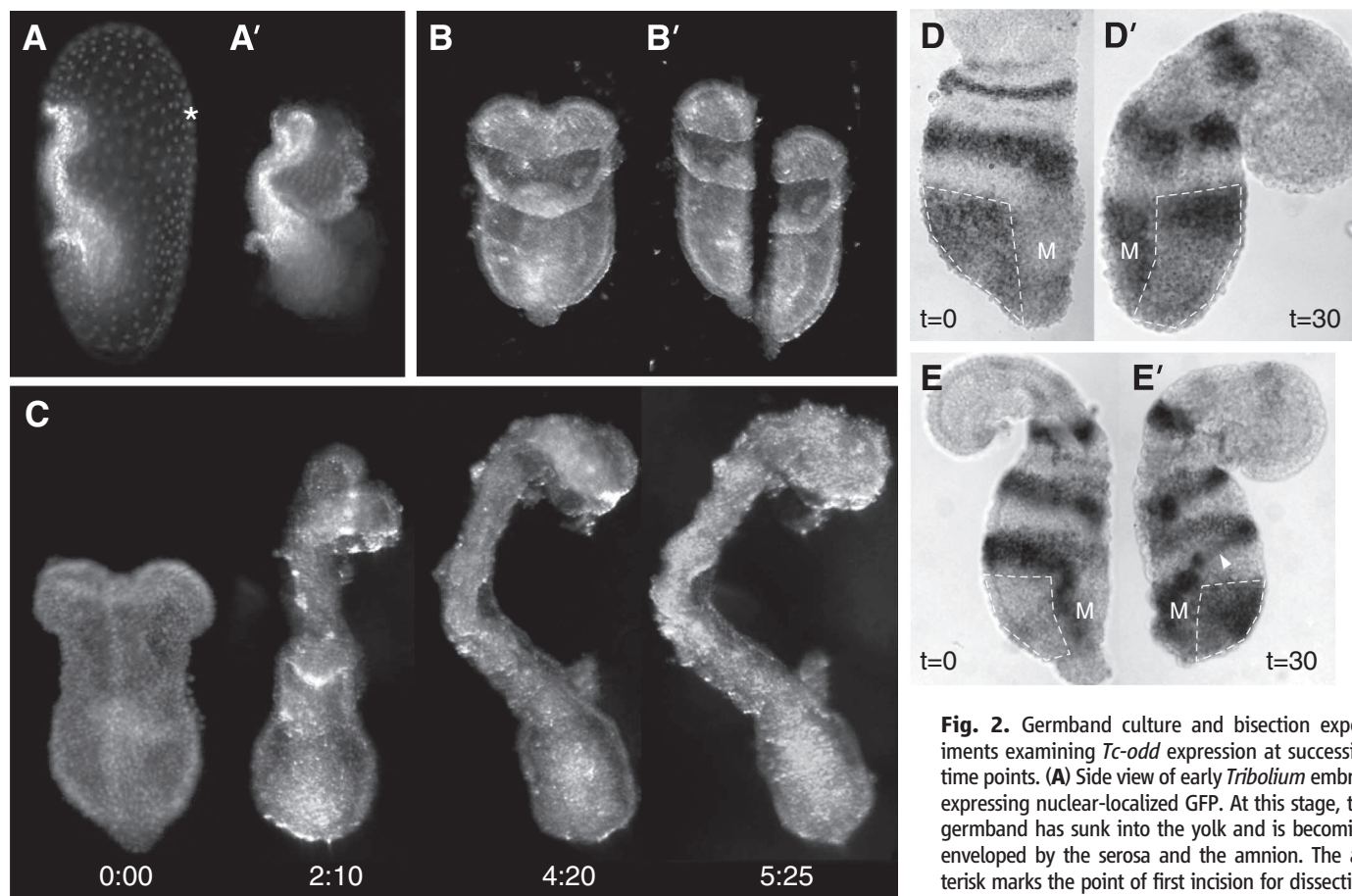
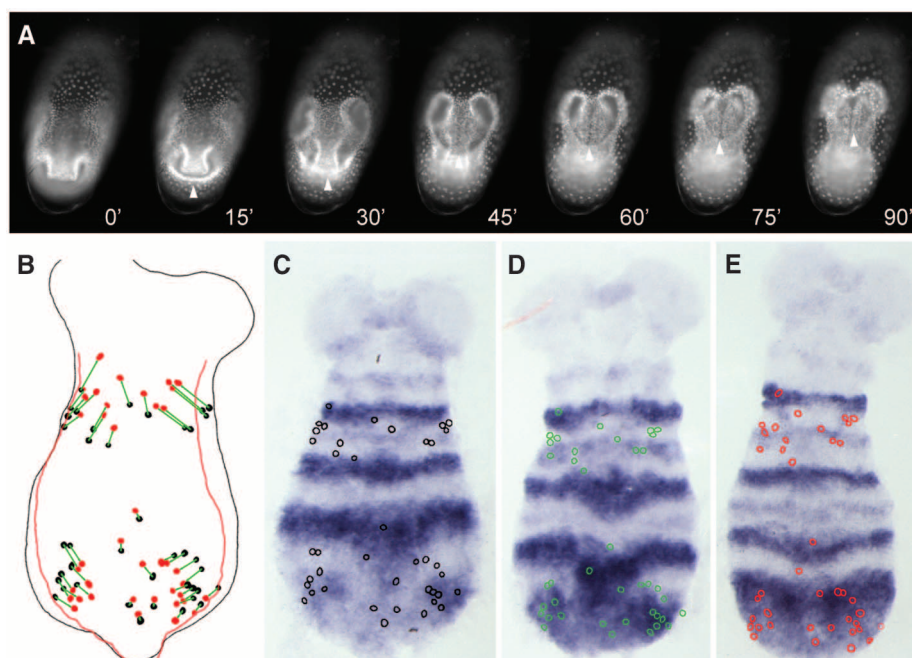


Fig. 2. Germband culture and bisection experiments examining *Tc-odd* expression at successive time points. **(A)** Side view of early *Tribolium* embryo expressing nuclear-localized GFP. At this stage, the germband has sunk into the yolk and is becoming enveloped by the serosa and the amnion. The asterisk marks the point of first incision for dissecting the germband. **(A')** Same embryo, after the serosa and yolk have been partially removed. **(B)** Ventral view of fully dissected germband in culture. **(B')** Same embryo shown after bisection along the midline. **(C)** Successive pictures of a germband in culture. The germband was mounted under a coverslip and cultured for more than 5 hours at 30°C. **(D and D')** In situ hybridization for *Tc-odd* on two halves of the same embryo. One half was fixed immediately after dissection (**D**), whereas the other was cultured for 30 min at 30°C before fixation (**D'**). *Tc-odd* expression in each half-embryo is at phase 3.II and 3.III, respectively. **(E and E')** Same experiment, with half-embryos at phases 4.I and 4.II. The lateral part of the growth zone, encompassing the ectoderm and amnion, is outlined with a dashed line; the medial part, which also includes mesoderm, is marked M (see fig. S1). In panel **E'**, the oblique stripe of *Tc-odd* expression (white arrowhead) is in the amnion. Anterior is toward the top in all panels.

Fig. 3. Live imaging and tracking cell movements in *Tribolium* germbands. **(A)** Live imaging of *Tribolium* embryo expressing nuclear-localized GFP, during early stages of germband elongation. White arrowheads mark the posterior edge of the amniotic fold as it extends over the germband (compare to Fig. 1). The frames shown were taken at 15-min intervals at 30°C; the corresponding time-lapse movie is shown in movie S1. **(B)** Tracks of individual ectodermal cells over a 60-min period at 30°C; black circles mark the initial position of cells and red circles mark their final position. **(C to E)** Cell positions at 0, 30, and 60 min are superimposed on *Tc-odd* stained embryos at the corresponding developmental stages (4.I, 4.II, and 4.III) (see Fig. 1, C to E). Cell populations that initially express very low levels of *Tc-odd* subsequently lie in areas that express the gene at high levels, and then low levels. All panels show ventral views.



(Fig. 3, C to E). These changes must be due to changes in mRNA levels within cells.

We find that cells in the early extending germband tend to become separated along the anterior-posterior axis and converge toward the ventral midline, which indicates that they undergo convergent extension. This is likely to be a major force driving germband elongation. In contrast, mitosis does not appear to contribute significantly to early germband elongation (see fig. S6).

Overall, our analysis provides compelling evidence for a segmentation clock in the growth zone of arthropods. By exploiting methods for embryo culture, transgenic markers, live imaging, and cell tracking in *Tribolium*, we are able to demonstrate that oscillating expression is due to temporal changes in expression levels, proof of which was missing in previous studies. The clock involves *Tc-odd*, a pair-rule gene known to be essential for *Tribolium* germband elongation and segmentation (19). An *odd*-related gene is also expressed dynamically with a two-segment periodicity in the growth zone of a centipede (4), raising the possibility that a widely conserved segmentation clock may exist in the arthropods. These results are consistent with the hypothesis that pair-rule genes were ancestrally part of a segmentation clock and subsequently evolved regulation by gap genes, which underlies *Drosophila* segmentation (20).

In a wider context, our results support the idea that a clock-based mechanism underlies segmen-

tation in animals as widely separated as arthropods and vertebrates. It will be interesting to discover whether this common feature reflects a common evolutionary origin of segmentation, or a design principle that was reinvented on separate occasions. In the latter case, the clock mechanism may have evolved independently but became integrated with a preexisting mechanism of posterior growth (1, 21, 22). Ultimately, this question might be resolved by comparing the gene regulatory networks underpinning the segmentation clock across phyla, as has already been attempted within the vertebrates (16). *Tribolium*, as an emerging model organism with an increasing array of genetic tools and resources (23), provides opportunities to investigate the arthropod clock mechanism by genetic means.

References and Notes

1. B. L. Martin, D. Kimelman, *Curr. Biol.* **19**, R215 (2009).
2. V. Wilson, I. Olivera-Martinez, K. G. Storey, *Development* **136**, 1591 (2009).
3. A. Stollewerk, M. Schoppmeier, W. G. Damen, *Nature* **423**, 863 (2003).
4. A. D. Chipman, W. Arthur, M. Akam, *Curr. Biol.* **14**, 1250 (2004).
5. J. I. Pueyo, R. Lanfear, J. P. Couso, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16614 (2008).
6. N. Dray *et al.*, *Science* **329**, 339 (2010).
7. A. D. Chipman, *Bioessays* **32**, 60 (2010).
8. F. Kainz, B. Ewen-Campen, M. Akam, C. G. Extavour, *Development* **138**, 5015 (2011).
9. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455 (1976).
10. I. Palmeirim, D. Henrique, D. Ish-Horowitz, O. Pourquié, *Cell* **91**, 639 (1997).

11. Y. Masamizu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1313 (2006).
12. M. L. Dequéant, O. Pourquié, *Nat. Rev. Genet.* **9**, 370 (2008).
13. A. C. Oates, L. G. Morelli, S. Ares, *Development* **139**, 625 (2012).
14. M. Schoppmeier, W. G. Damen, *Dev. Biol.* **280**, 211 (2005).
15. A. D. Chipman, M. Akam, *Dev. Biol.* **319**, 160 (2008).
16. A. J. Krol *et al.*, *Development* **138**, 2783 (2011).
17. H. Oda *et al.*, *Development* **134**, 2195 (2007).
18. T. Mito *et al.*, *Development* **138**, 3823 (2011).
19. C. P. Choe, S. C. Miller, S. J. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6560 (2006).
20. A. Peel, M. Akam, *Curr. Biol.* **13**, R708 (2003).
21. T. Copf, R. Schröder, M. Averof, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17711 (2004).
22. R. de Rosa, B. Prud'homme, G. Balavoine, *Evol. Dev.* **7**, 574 (2005).
23. S. J. Brown *et al.*, *Cold Spring Harbor Protocols* **2009**, pdb.emo126 (2009); <http://dx.doi.org/10.1101/pdb.emo126>.

Acknowledgments: We thank J. Schinko and G. Bucher for teaching us how to handle *Tribolium* and for help in generating the GFP-expressing transgenic line; J. P. Couso and S. Bishop for advice on embryo culture; D. Kosman, W. McGinnis, and C. Delidakis for advice on fluorescence in situ; and M. Strigini for critical comments on the manuscript. Our work was funded by the Marie Curie programs CELLIMAGE and ZOONET (FP6, European Union).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1218256/DC1
Materials and Methods
Figs. S1 to S6
Movies S1 to S3
References (24–28)

21 December 2011; accepted 27 February 2012
Published online 8 March 2012;
10.1126/science.1218256

Synthetic Genetic Polymers Capable of Heredity and Evolution

Vitor B. Pinheiro,¹ Alexander I. Taylor,¹ Christopher Cozens,¹ Mikhail Abramov,² Marleen Renders,^{2*} Su Zhang,³ John C. Chaput,³ Jesper Wengel,⁴ Sew-Yeu Peak-Chew,¹ Stephen H. McLaughlin,¹ Piet Herdewijn,² Philipp Holliger^{1†}

Genetic information storage and processing rely on just two polymers, DNA and RNA, yet whether their role reflects evolutionary history or fundamental functional constraints is currently unknown. With the use of polymerase evolution and design, we show that genetic information can be stored in and recovered from six alternative genetic polymers based on simple nucleic acid architectures not found in nature [xeno-nucleic acids (XNAs)]. We also select XNA aptamers, which bind their targets with high affinity and specificity, demonstrating that beyond heredity, specific XNAs have the capacity for Darwinian evolution and folding into defined structures. Thus, heredity and evolution, two hallmarks of life, are not limited to DNA and RNA but are likely to be emergent properties of polymers capable of information storage.

The nucleic acids DNA and RNA provide the molecular basis for all life through their unique ability to store and propagate

information. To better understand these singular properties and discover relevant parameters for the chemical basis of molecular information encoding, nucleic acid structure has been dissected by systematic variation of nucleobase, sugar, and backbone moieties (1–7).

These studies have revealed the profound influence of backbone, sugar, and base chemistry on nucleic acid properties and function. Crucially, only a small subset of chemistries allows information transfer through base pairing with DNA or RNA, a prerequisite for cross-talk with extant biology. However, base pairing alone cannot conclusively determine the capacity of a given chemistry to serve

as a genetic system, because hybridization need not preserve information content (8). A more thorough examination of candidate genetic polymers' potential for information storage, propagation, and evolution requires a system for replication that would allow a systematic exploration of the informational, evolutionary, and functional potential of synthetic genetic polymers and would open up applications ranging from biotechnology to materials science.

In principle, informational polymers can be synthesized and replicated chemically (9), with advances in the nonenzymatic polymerization of mononucleotides (10) and short oligomers (11, 12) enabling model selection experiments (13). Nevertheless, chemical polymerization remains relatively inefficient. On the other hand, enzymatic polymerization has been hindered by the stringent substrate selectivity of polymerases. Despite progress in understanding the determinants of polymerase substrate specificity and in engineering polymerases with expanded substrate spectra (7), most unnatural nucleotide analogs are poor polymerase substrates at full substitution, as both nucleotides for polymer synthesis and templates for reverse transcription. Notable exceptions are 2'OMe-DNA and α -L-threofuranosyl nucleic acid (TNA). 2'OMe-DNA is present in eukaryotic ribosomal RNAs, is well tolerated by natural reverse transcriptases (RTs), and has been shown to support heredity and evolution at near full substitution (14). TNA allowed polymer synthesis and evolution in a three-letter system (15) but only limited reverse transcription (16).

¹Medical Research Council (MRC) Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. ²Rega Institute, Katholieke Universiteit Leuven, Minderbroederstraat 10, B 3000, Leuven, Belgium. ³Center for Evolutionary Medicine and Informatics, The Biodesign Institute at Arizona State University, 1001 South McAllister Avenue, Tempe, AZ 85287–5301, USA. ⁴Nucleic Acid Center, Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark.

*Present address: Department of Chemistry, University of British Columbia, Vancouver V6T 1Z1, Canada.

†To whom correspondence should be addressed. E-mail: ph1@mrc-lmb.cam.ac.uk