Supporting Information

Pueyo *et al.* 10.1073/pnas.0804093105

SI Text

***Periplaneta americana* (*Delta* GenBank Accession Number FJ222590).** MRWTQQTRVQGAVVVVILAALQQVCCSGVFELRLKSFINDYGKDSVGQCCSGTPSPGTKACSGPCRTRFRVCLKHYQAKIDTTSPCTFGDVITPVLGENSVHLVGSAQHDGFANPIRFPFDFTWPGTFSLIVEALHDNNNATSRSGGETLITRLTTQRWLDVGEDWTEDEHRSSHSVMRYEYRVTCDAHYYGAGCANLCRPRDDNFGHYKCSATGDRVCLSGWQGDYCTKPRCLPGCDEQHGHCNQPNECLCHSGWKGRLCDECERYPGCLHGTCQKPWDCLCDEGWGGLFCNQDLNYCTNHKPCRHGGTCFNTGQGSYTCSCPPGYTGTDCERELDDCAHHPCLNGGTCKDNGTSSYRCECPKGWHGPHCETSAQTCEDQPCRHGGTCSDTAQGYTCKCPPGFNGTDCEHQIDECAPSPCKNNASCVDRINGFECICPTGFIGDRCETNVDDCQGNPCLNGGTCVDMVNQFRCQCVPGYVGPLCQSKVDYCQTKPCANGGECTYLLNDYKCTCRPGFTGKDCSIDADECSSSPCKNGGTCSNRVNSFQCQCPQGYHGKTCSDEATVVVPNQDASSAGAYLATPGNISRHVLEQSDDDAGLSTEQVVVIATLSTIVPIVVLVAAVVVMCMKRKRKREQQRADEEARMQNERNAVHSSMAKRGGGGTVGGDTHMIKNTWGKCINNVLAAGDDGGGGGGGGSVANIVSVSGDLSNASDLCYPKQQQQSQQVLDNSGPVYTLQRTRSHKQLNTDVHRGSQSNRTSTALLASKLDKDFDNLCPSNTQQQQRTPSSNAADKRISVLSVDSSLCNASDPTLVKRCVDKDSNSILSAAAGSGPTSSVFVIDEHFHHQEGLLATEV

***Periplaneta americana* (*Notch* GenBank Accession Number FJ222591).** MSARLRCCLLTMDYYVVFLILATCWIPSVTGFVSCSPSPCKNGGTCVSTPRGESYCNCTAMHVGEYCQYMNPCHTGPGPRCQNGGSCNVRSSPTSSPSFWCTCPIGYSASLCEIPVANSCDSDPCLNGGTCTLRSLDSYSCACAPGYTGQHCELQDHCASMPCRNGAECESLGDTYECTCAPGFVGSSCSEDIVECQSEPCVHGTCYNTHGSYKCVCHPGYTGQNCENKYIPCDPSPCLNDGTCRQVDALSYECQCPIGYRGKNCEENIDDCPGNLCQNGATCVDDINRYSCLCPPTYMGELCEQDVDECGQRPSVCQNGATCTNSIGGFPCICVNGWAGADCSVNIDDCAGAACFNGATCIDRVGSFYCRCTPGKTGLLCHLDDACTSNPCHADAICDTSPINGSYTCSCASGYKGVDCSEDIDECEQGSPCEHDGICVNTPGSFACNCTQGFTGPRCETNVNECES

***Periplaneta americana* (*Hairy* GenBank Accession Number FJ222592).** MVTGGGTVSPAGVAGPGVIATNPSACVASATPTPGPDSAGLPPQRRTNENRRSNKPIMEKRRRARINNCLNELRTLILDAMKKDPARHSKLEKADILEMTVKHLENLQRQQVAMSAATDPSVLNKFRAGFTECAGEVGRFPGLESPVRKRLLQHLANCLNGTTTAAPSGTGVPQPSPQDSSPAPPQHQTAVQVHILPSSTTTTVDNHQTPTAVSASSPPNGIFFTTGPNGAGLQLVPTRLPNGDIALVLPSSGSGTAQIFRQAATITTASPASSPAPSSSSSSSSPLPMLIPIPTRTSSTASASSSTSSSCASTSTSPVAFDRLAVSSPPQTSTVVNHPACRDMATSPAAYHLPVSPQSSHHGNGGYEVMDQHPPLRPYSPPMQKPLALVVKKETAPEVPVEEKPWRPW

Cloning *Pa-Dl Pa-N*, and *Pa-h* Genes. *Pa-Dl* primers; Exon4 degenerate primers: (Dl3F-GCNGAYTGGACNGAGGAYGAG; Dl3R-GCAGTAGTCNCCYTTCCANCCN; Dl4R-GTAGTGNCCGAAGTTGTCGTCNC). Exon6 degenerate primers: (Dl5F-AGGGNTGGGNGGNYTNTT; Dl1R-GGANGTGTANAGNCCYTGNCCNGT; Dl2R-TGTTGAAGGANGTNCCNCCG).

Exon4 gene-specific primers: (Dle4F-CTCGTCGCACTCCGTGATGC; Dle4R-CGTAGTAGTGTGCGTCGCACG). Exon6 gene specific primers: (Dle6F-TAACCAGGACCTCAATTACTGCACG; Dle6R-TGTCTGCACGGCTTATGATTCG).

Pa-N primers: *Pa-N* gene specific primers: (N1R-GGTCCTGTGAAGCCCTGTGT; N2R-AGGTGTAGGAGCCGTTGATAGG).

(Notch7stoll; Notch8stoll; Notch7restoll; Notch8restoll) (13) *Pa-h* genomic degenerated primers: (hF-WSBAMHAARCCBATYATG; hF2-AARCCBATYATGGARAAA; hR-TCTTTYTTBRTBGCATCYA; hR2-TTBRTBCATCYARRATVAG).

Pa-h cDNA primers: (outer primers he2F1-AAGCGTCGAGCTCGAATCAATA and he2R1-TGTTCTCAGCTCATTAAGACAG, nested primers he2F2-CTGTCTTAATGAGCTGAGAACA and he2R2-TTATTGATTCGAGCTCGACGC).

Sequence Analysis. The protein domains in the sequenced fragments of *P. americana* Notch, Delta, and Hairy were compared to other metazoan Notch, Delta, and Hairy homologues using publicly available protein domain databases and domain prediction algorithms. The predicted structure of the majority of Notch, Delta and Hairy homologues was taken from the Pfam database (http://www.sanger.ac.uk/). These domain predictions are calculated by cross-referencing results from a series of independent databases: PfamA (14), which is a database of well-characterized protein domains; SMART (Simple Modular Architecture Research Tool) (15) which uses a similar database to PfamA with a different domain prediction algorithm; SignalP (16), which tests the sequence for putative signal peptides; and TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) which tests the sequence for putative transmembrane helices. Where Notch and Delta homologues did not have Uniprot identifiers, predicted amino acid sequences were checked against each of these databases sequentially and results were collated. A cut-off value of 1,500 units was used with the TMpred algorithm. Where predictions from different databases overlapped, databases were given precedence in the order in which they are listed earlier. Accession numbers are given later.

For phylogenetic analyses, sequences (other than for *P. americana*) were aligned by eye using SE-AL (http://evolve.zoo.ox.ac.uk). Regions for which homology among sites could not confidently be inferred were removed. This procedure resulted in an alignment of 403 aa for Delta (from the 3′ end of the signal peptide up to and including the fifth EGF repeat), and an alignment of 414 aa for Notch (covering EGF repeats 6 to 12). Alignments are available on request. Maximum Likelihood phylogenetic trees were calculated using PhyML (17). Trees were calculated from the amino acid sequence using the WAG (18) amino acid replacement matrix with four categories of gamma-distributed rates across sites and the proportion of invariant sites estimated from the data. 500 non-parametric maximum likelihood bootstraps were performed on each tree to assess the confidence associated with each node. Bayesian analyses were performed using MrBayes version 3.1.2. Trees were calculated using an amino acid alignment and an identical model to that used for the maximum likelihood analysis (the WAG amino acid replacement matrix with four categories of gamma-distributed rates across sites, and proportion of invariant sites estimated

from the data). Starting trees were random trees. Bayesian analyses were run for 1,000,000 generations, with the first 250,000 generations discarded as burnin. Following burnin, trees were sampled every 20 generations, providing a posterior sample of 37,500 trees. Analyses were performed using one cold and one heated chain to improve the ability of the algorithm to find the global optimum. Convergence was assessed by running two identical analyses in parallel and examining the average SD of split frequencies. This diagnostic method gives an indication of the similarity of the posterior sample of the trees in the two replicate runs. Runs were considered to have converged when the SD of split frequencies decreased to less than 0.01, which occurred well before 1,000,000 generations in all analyses.

Accession Numbers. Accession numbers for protein domain prediction and phylogenetic analyses are as follows: *Pa-Dl*, *Periplaneta americana* Delta (reported here, GenBank accession number JF222590); *Dm-Dl*, *Drosophila melanogaster* Delta (Uniprot accession: P10041); *Ag-Dl*, *Anopheles gambiae* Delta (predicted protein sequence from GenBank accession XM_319454); *Tc-Dl*, *Tribolium castaneum* Delta (predicted sequence from GenBank accession XM_964994); *Am-Dl*, *Apis mellifera* Delta (predicted protein sequence from GenBank accession XM_393831, minus the first 247 aa, which appear to be mis-annotated); *Ph-Dl*, *Parhyale hawaiensis* Delta (predicted protein sequence from GenBank accession number DQ917570); *At-Dl*, *Achaearanea tepidariorum* Delta (predicted protein sequence from GenBank accession AB287420); *Cs-Dl1*, *Cupiennius salei* Delta1 (Uniprot accession: Q81498); *Cs-Dl2*, *Cupiennius salei* Delta2 (Uniprot accesssion: Q81497); *Hs-Dl1*, *Homo sapiens* Delta 1 (Uniprot accession: O00548). *Pa-N*, *Periplaneta americana* Notch (reported here, GenBank accession number FJ222591); *Dm-N*, *Drosophila melanogaster* Notch (Uniprot accession: P10041); *Lc-N*, *Lucilia cuprina* Notch (Uniprot accession: Q25253); *Aa-N*, *Aedes aegypti* Notch (predicted protein sequence from GenBank accession CH477646); *Ph-N*, *Parhyale hawaiensis* Notch (predicted protein sequence from GenBank accessionDQ917572); *Gm-N*, *Glomeris marginata* Notch (Uniprot accession: Q869J7); Lf-*N*, *Lithobuis forficatus* Notch (Unprot accession: Q68QF3); *Cs-N*, *Cupiennius salei* Notch (Uniprot accession: Q8I498); Bm-*N*, *Boophilus microplus* Notch (Uniprot accession: Q8I6X6); Hs-*N*, *Homo sapiens* Notch1 (Uniprot accession: P46531). *Pa-H*, *Periplaneta americana* Hairy (reported here, accession number pending); *Dm-H*, *Drosophila melanogaster* Hairy (predicted protein sequence from GenBank accession NM_079253); *Aa-H*, *Anopheles aegypti* Hairy (predicted protein sequence from GenBank accession CH477820, join(655775.655885,656659.656754,656853.657677)); *Tc-H*, *Tribolium castaneum* Hairy (predicted protein sequence from GenBank accession XM_966842); *Am-H*, *Apis mellifera* Hairy (predicted protein sequence from GenBank accession XM_393948); *Cs-H*, *Cupiennius salei* Hairy (predicted protein sequence from GenBank accession AJ252154); *At-H*, *Achaearanea tepidariorum* Hairy (predicted protein sequence from GenBank accession AB125743); and *Hs-HES1*, *Homo sapiens* Hairy (predicted protein sequence from GenBank accession NM_005524).

1. Stollewerk A, Schoppmeier M, Damen WGM (2003) Involvement of Notch and Delta genes in spider segmentation. *Nature* 423:863–865.
2. Peel A, Akam M (2003) Evolution of segmentation: Rolling back the clock. *Curr Biol* 13:708–710.
3. Campuzano S, Modolell J (1992) Patterning of the Drosophila nervous system: the achaete-scute gene complex. *Trends Genet* 8:202–208.
4. Adam J, *et al.* (1998) Cell fate choices and the expression of Notch, Delta and Serrate homologues in the chick inner ear: parallels with Drosophila sense-organ development. *Development* 125:4645–4654.
5. Couso JP,Bishop SA, Martinez Arias A (1994) The wingless signalling pathway and the patterning of the wing margin in Drosophila. *Development* 120:621–636.
6. Jacobs DK, *et al.* (2000) Molluscan engrailed expression, serial organization, and shell evolution. *Evol Dev* 2:340–347.
7. de Celis JF, Bray S (1997) Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the Drosophila wing. *Development* 124:3241–3251.
8. Couso JP, Knust E, Martinez Arias A (1995) Serrate and wingless cooperate to induce vestigial gene expression and wing formation in Drosophila. *Curr Biol* 5:1437–1448.
9. Kim J, Irvine KD, Carroll S (1995) Cell interactions and inductive signals at the dorsal/ventral boundary of the developing Drosophila wing. *Cell* 82:795–802.
10. Jiang YJ, *et al.* (2000) Notch signalling and the synchronization of the somite segmentation clock. *Nature* 408:475–479.
11. Pourquie O (2003) The segmentation clock: converting embryonic time into spatial pattern. *Science* 301:328–330.
12. Ciudad L, Piulachs MD, Belles X (2006) Systemic RNAi of the cockroach vitellogenin receptor results in a phenotype similar to that of the Drosophila yolkless mutant. *FEBS J* 273:325–335.
13. Stollewerk A (2002) Recruitment of cell groups through Delta/Notch signalling during spider neurogenesis. *Development* 129:5339–5348.
14. Finn RD, *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34:D247–D251.
15. Letunic I, *et al.* (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 34:D257–D260.
16. Bendtsen JD, *et al.* (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
17. Guindon S, *et al.* (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33:W557–W559.
18. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699.

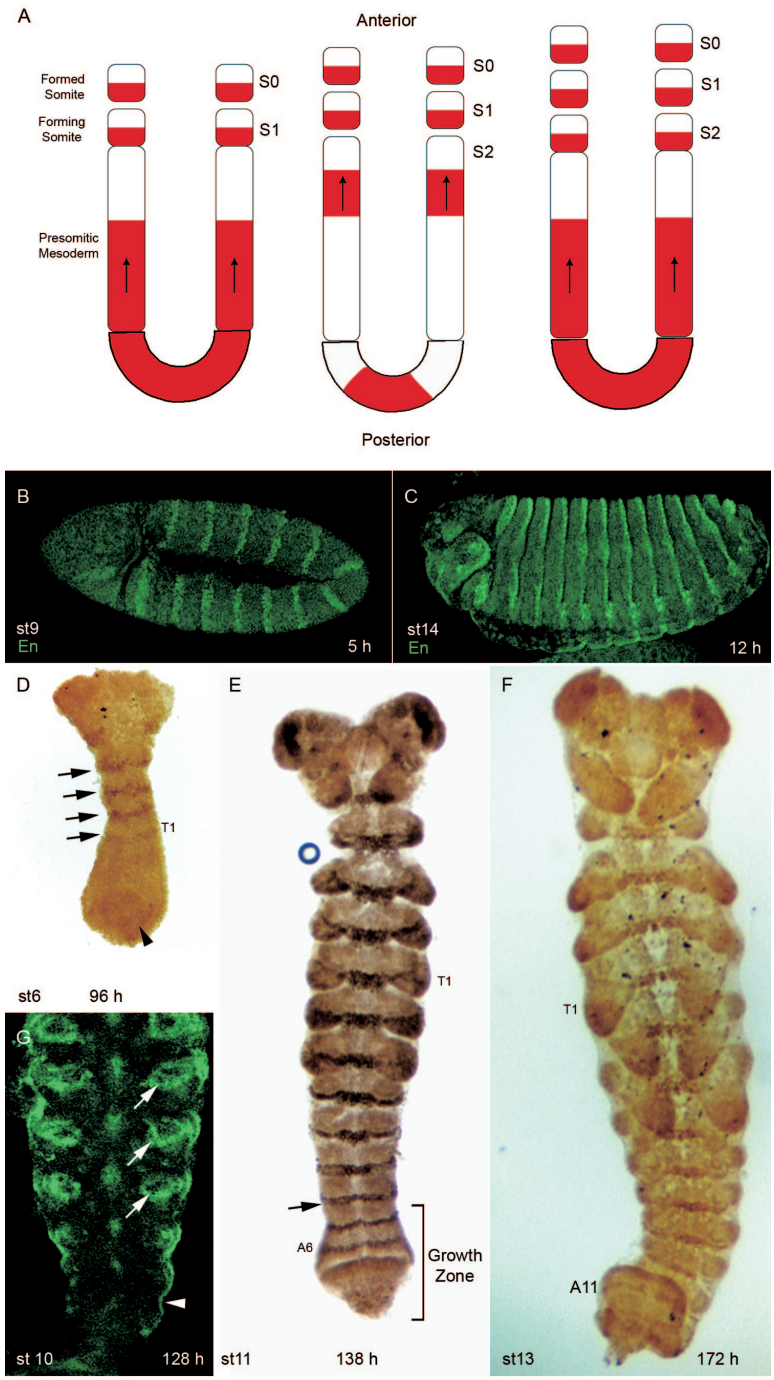

Fig. S1. Modes of Segmentation in vertebrate and insect embryos. (*A*) Diagram shows how segmentation proceeds in a vertebrate embryo. The most posterior part of the embryo consists of an undifferentiated and proliferative mesoderm called the presomitic mesoderm (PSM), followed by single segmented units of mesoderm called the somites (S). Somites are acquired sequentially from the posterior in a cyclical fashion. Members of the Notch pathway show coordinated and cyclic patterns of expression in the PSM and are required for somite formation. A stripe of expression of segmentation genes sweeps across the PSM. Every wave of expression gives rise to the formation of a new somite. (*B*,*C*) Segmentation of a long-germ band insect, *Drosophila*, takes place in a syncytial blastoderm. All segments form rapidly and simultaneously by a cascade of segmentation genes. Engrailed (En) expression revealed with an anti-En antibody appears simultaneously in all segments as observed in a stage 9 embryo (*B*). This segmental pattern of En is maintained until the end of development (*C*). (*D*–*F*) Segmentation of a short–germ band insect, *Periplaneta*, proceeds sequentially. (*D*) Shortly after the blastoderm stage, the stage 6 embryo has determined only the most anterior segments: cephalic labial, maxillar, and first thoracic segment as revealed by the stripes of En protein expression (arrows). As embryogenesis proceeds the embryo acquires the rest of the trunk segments sequentially from the posterior part of the embryo, the ''growth zone'' (arrowhead). (*E*) By stage 11, the sixth abdominal segment is formed as revealed by En expression and the abdominal segment four starts developing segmental furrows (arrow). The growth zone is denoted by a bracket. (*F*) All of the segments have been formed by stage 13. (*G*) *Periplaneta* somites (arrows), revealed by the F3.38 antibody, are composed of an outer epidermal layer and a inner cluster of mesodermal cells (arrow). The segregation of mesodermal cells follows the appearance of the segmental furrow (arrowhead).

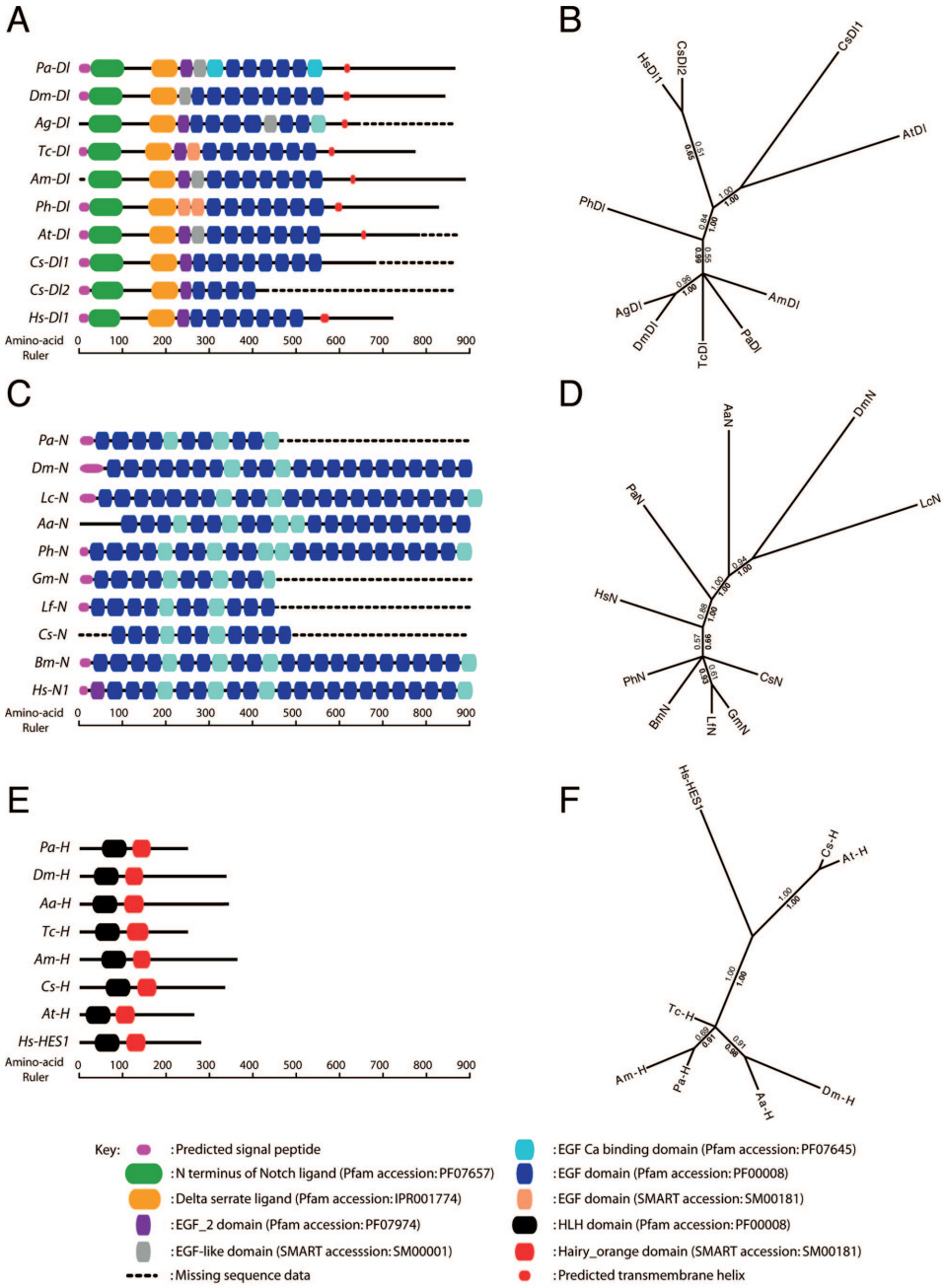

Fig. S2. Sequence analysis of metazoan Delta, Notch, and hairy homologues. (*A*,*C*,*E*) predicted protein domains of metazoan Delta, Notch, and hairy homologues, respectively. (*B*,*D*,*E*) The 50% majority rule: maximum likelihood phylogeny of metazoan Delta, Notch, and hairy homologues, respectively. In *B*, *D*, and *E*, the numbers in standard type above each branch represent the proportion of 500 maximum likelihood bootstrap replicates in which that split appeared, and the numbers in bold below each branch represent the Bayesian posterior probability of each split. Abbreviations: Pa, *P. americana*; Dm, *D. melanogaster*; Ag, *Anopheles gambiae*; Tc, *Tribolium castaneum*; Am, *Apis mellifera*; Ph, *Parhyale hawaiensis*; At, *Achaearanea tepidariorum*; Cs, *Cupiennius salei*; Hs, *Homo sapiens*; Lc, *Lucilia cuprina*; Aa, *Aedes aegypti*; Gm, *Glomeris marginata*; Lf, *Lithobius forficatus*; Bm, *Boophilus microplus*. (*A*) Pa-Dl consists of a domain that is common to many Notch ligands, followed by a DSL (Delta-Serrate-Lag2) domain and 9 EGF repeats. In the two insects (*P. americana* and *D. melanogaster*) for which the full coding sequence of Delta is known, the length of the protein is highly conserved (858 and 833 aa respectively), whereas in vertebrates the protein is somewhat shorter (723 aa). (*B*) The ML Delta tree exactly matches the expected (i.e., species) tree, and both internal nodes are well supported by bootstrap analysis. (*C*) The region of Pa-N that we have sequenced (amino acids 1–469) contains 11 EGF repeats, three of which have calcium-binding affinity, an organization highly conserved among the metazoa. (*D*) The Pa-N sequence appears in the extended position in the Notch ML tree, recovering the monophyly of Neoptera (*P. americana* and *D. melanogaster*). There is some disagreement of the Notch ML tree with the expected (i.e., species) tree, in that the Myriapoda are recovered as paraphyletic. However, bootstrap proportions indicate that these regions of the tree are not well supported, and that there is therefore conflicting phylogenetic signal in this region of the tree. (*E*) The protein encoded by *Pa-h* is also similar to those of other *hairy* genes in metazoans, and contains a basic helix-loop-helix domain related to its activity as a transcription factor. (*F*). The phylogenetic position of Pa-H in a ML tree is also compatible with a expected species tree.

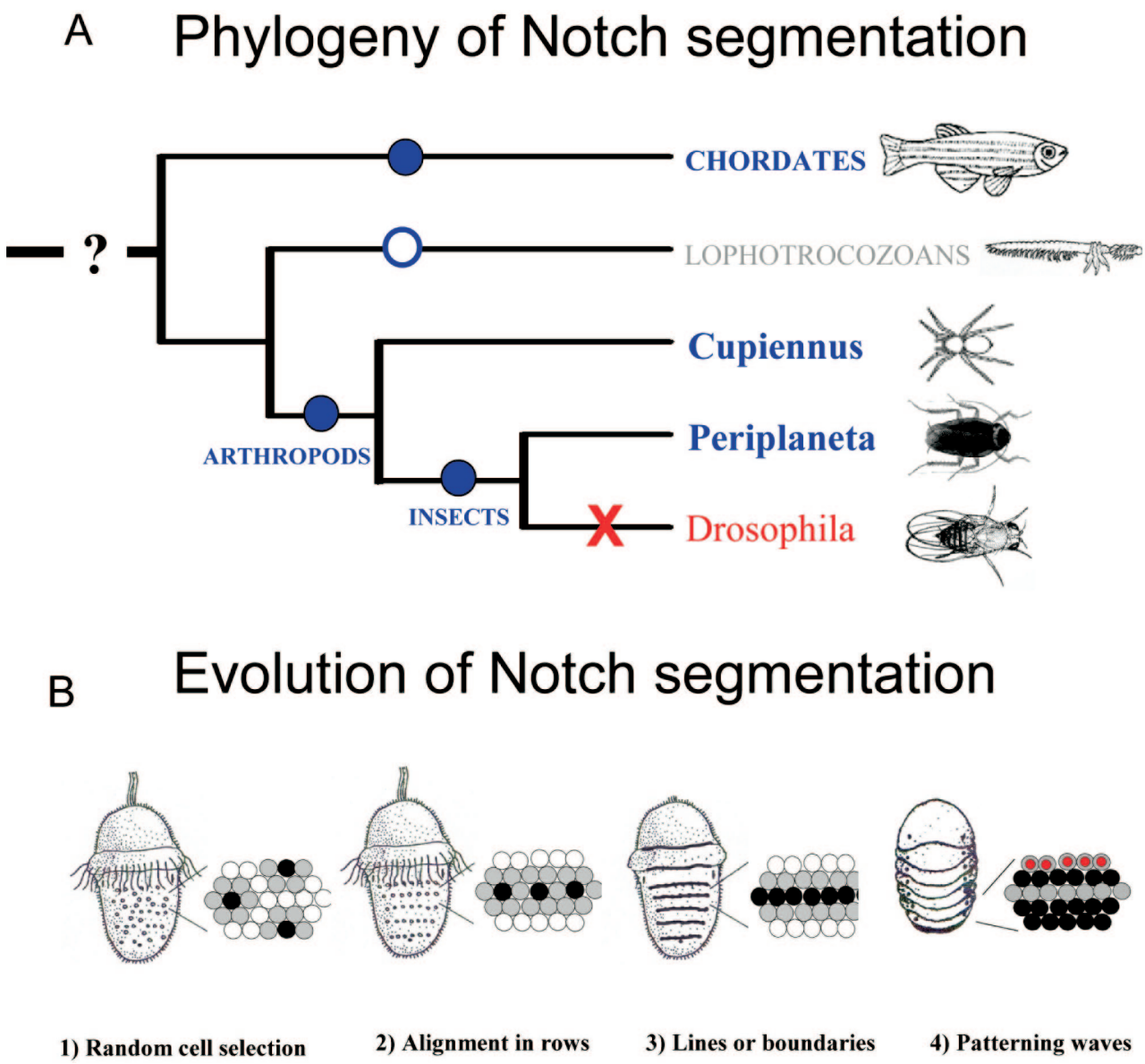

Fig. S3. Evolution of Notch-mediated segmentation. (A) Simplified phylogenetic tree showing segmented taxa with (blue type) and without (gray and red type) a reported role for Notch in segmentation. Ancestors with Notch segmentation are labeled as blue filled dots and others with metameric organization revealed by Engrailed expression as blue ringed circles. The presence of Notch-mediated segmentation in the insect *P. americana* (in the present study) and in the chelicerate *Cupiennus salei* (1) indicates that it is the ancestral condition for arthropods, but it has been lost in some insects such as *Drosophila,* possibly in correlation with the appearance of meroistic ovaries and gap segmentation (2). Notch-mediated segmentation is also present in vertebrates. This coincidence might be a result of Notch segmentation being inherited by both vertebrates and arthropods from a common bilaterian ancestor, or by Notch segmentation evolving independently in each lineage. The presence of Notch and Engrailed expression in annelids or other lophotrochozoans might suggest that the first possibility is the most parsimonious hypothesis. (*B*) Putative evolution of Notch signaling and metameric organization. First, there is segregation of sensory elements or other isolated pattern elements from clusters of competent cells at random, as in sensory bristles in Drosophilid flies and vertebrates (3, 4). Second, there is alignment of such competent clusters and selected elements in rows, as in bristles at the anterior wing margin in Drosophilids (5) or goblet cells in quitons (6). Third, there is patterning of lines or boundaries, as in the fly veins (7). The veins carry some neural and sensory cells but the intervening cells are not selected out, thus whole lines of cells are selected and can take a different fate. In turn, these cells can signal to their neighbors, as in the dorsal–ventral wing boundary of *Drosophila* (8, 9). Finally, patterning waves are seen as in the segmentation clock. Once cells in patterning boundaries gain the ability to signal to adjacent cells, they recruit more cells and fates to the boundary. If one of the signals passed on is Notch pathway cyclic activation, a run-away patterning wave is created. This can generate a series of metameric boundaries along the main body axis (10, 11).

Table S1. Penetrance of the N and h RNAi phenotypes

RNAi	No. of embryos	WT	Segmentation defects	Nonspecific
Notch	60	6 (10)	46 (76.67)	8 (13.33)
Hairy	73	43 (58.90)	16 (21.92)	14 (19.18)

Values in parentheses are percentages. Quantification of the penetrance of the RNAi phenotypes was carried out by scoring the number of embryos showing segmentation and nonspecific defects, and WT phenotype. In Notch-RNAi experiments, ≈76% embryos showed segmentation defects. This percentage coincides with number of embryos showing specific phenotypes treated with RNAi in a related cockroach species [Ciudad L, Piulachs MD, Belles X (2006) Systemic RNAi of the cockroach vitellogenin receptor results in a phenotype similar to that of the Drosophila yolkless mutant. *FEBS J* 273:325–335]. In addition, near total loss of Notch expression was observed in N-RNAi treated embryos using a *Pa-N*-specific probe. In hairy-RNAi experiments, the penetrance and expressivity of the phenotypes are not as strong as those produced by N-RNAi. This hypomorphy correlates with the fact that h-RNAi does not produce a total loss of the *Pa-h* transcripts even after increasing the amount of injected RNAi.