Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate

France Denoeud, Simon Henriet, Sutada Mungpakdee, Jean-Marc Aury, Corinne Da Silva, Henner Brinkmann, Jana Mikhaleva, Lisbeth Charlotte Olsen, Claire Jubin, Cristian Cañestro, et al.

PLASTICITY OF ANIMAL GENOME ARCHITECTURE UNMASKED BY RAPID EVOLUTION OF A PELAGIC TUNICATE

France Denoeud[1,2,3], Simon Henriet[4], Sutada Mungpakdee[4]*, Jean-Marc Aury[1,2,3]*, Corinne Da Silva[1,2,3]*, Henner Brinkmann[5], Jana Mikhaleva[4], Lisbeth Charlotte Olsen[4], Claire Jubin[1,2,3], Cristian Cañestro[6], Jean-Marie Bouquet[4], Gemma Danks[4,7], Julie Poulain[1,2,3], Coen Campsteijn[4], Marcin Adamski[4], Ismael Cross[8], Fekadu Yadetie[4], Matthieu Muffato[9], Alexandra Louis[9], Stephen Butcher[10], Georgia Tsagkogeorga[11], Anke Konrad[22],Sarabdeep Singh[12], Marit Flo Jensen[4], Evelyne Huynh Cong[4], Helen Eikeseth-Otteraa[4], Benjamin Noel[1,2,3], Véronique Anthouard[1,2,3], Betina M. Porcel[1,2,3], Rym Kachouri-Lafond[13], Atsuo Nishino[14], Matteo Ugolini[4], Pascal Chourrout[15], Hiroki Nishida[14], Rein Aasland[16], Snehalata Huzurbazar[12], Eric Westhof[13], Frédéric Delsuc[11], Hans Lehrach[17], Richard Reinhardt[17], Jean Weissenbach[1,2,3], Scott W. Roy[18], François Artiguenave[1,2,3], John H. Postlethwait[6], J. Robert Manak[10], Eric M. Thompson[4,19], Olivier Jaillon[1,2,3], Louis Du Pasquier[20], Pierre Boudinot[21], David A. Liberles[22], Jean-Nicolas Volff[23], Hervé Philippe[5], Boris Lenhard[4,7], Hugues Roest Crollius[9], Patrick Wincker[1,2,3†] & Daniel Chourrout[4†]

*equal contributions, [†]corresponding authors

[1]CEA, IG, Genoscope, Evry, France. [2] CNRS, UMR 8030, Evry, France. [3]Université d'Evry. Evry, France. [4]Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, Norway. [5]Département de Biochimie, Université de Montréal, Montréal, Canada. [6]Institute of Neuroscience, University of Oregon, Eugene, USA. [7]Bergen Center for Computational Science, University of Bergen, Bergen, Norway. [8]Laboratorio de Genética, Universidad de Cádiz, Cádiz, Spain. [9]Dyogen Lab, CNRS UMR8541, Ecole Normale Supérieure, Paris, France. [10]Department of Biology, University of Iowa, Iowa City, Iowa, USA. [11]Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Evolution, UMR5554-CNRS, Université Montpellier II, Montpellier, France. [12]Department of Statistics, University of Wyoming, Laramie, USA. [13]Institut de Biologie Cellulaire et Moléculaire du CNRS, Université de Strasbourg, Strasbourg, France. [14]Department of Biological Sciences, Osaka University, Osaka, Japan. [15]Centre Hospitalier d'Albi, Albi, France. [16]Department of Molecular Biology, University of Bergen, Bergen, Norway. [17]Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [18]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA. [19]Department of Biology, University of Bergen, Bergen, Norway. [20]Institute of Zoology and Evolutionary Biology, University of Basel, Basel, Switzerland. [21]INRA, Virologie et Immunologie Moléculaires, Jouy-en-Josas, France. [22]Department of Molecular Biology, University of Wyoming, Laramie, USA. [23]Institut de Génomique Fonctionnelle de Lyon, UMR5242 CNRS/INRA/UCBL/ENS, Ecole Normale Supérieure de Lyon, Lyon, France.

Correspondence should be addressed to :

Daniel Chourrout, Sars International Centre for Marine Molecular Biology, University of Bergen, Thormoehlensgate 55, N-5008 Bergen, Norway, Tel: +47 55 58 43 13, Fax: +47 55 58 43 05, email: Daniel.Chourrout@sars.uib.no

Patrick Wincker, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France, Tel: +33 1 60 87 25 68, Fax: +33 1 60 87 25 14, email: pwincker@genoscope.cns.fr

Genomes of animals as different as sponge and human show conservation of global architecture. Here we show that multiple genomic features including transposon diversity, developmental gene repertoire, physical gene order and intron-exon organization are shattered in the tunicate *Oikopleura*, belonging to the sister group of vertebrates and retaining chordate morphology. Ancestral architecture of animal genomes can be deeply modified and may therefore be largely non-adaptive. This rapidly evolving animal lineage thus offers unique perspectives on the level of genome plasticity. It also illuminates issues as fundamental as the gain of new introns.

Tunicates, viewed as the closest living relatives of vertebrates, were probably simplified from more complex chordate ancestors (1). Larvacean tunicates represent the second most abundant component of marine zooplankton and filter small particles by their gelatinous house. *Oikopleura dioica* is the most cosmopolitan larvacean, has a very short life cycle (4 days at 20°C) and can be reared in the laboratory for hundreds of generations (2). Unique among tunicates, it has separate sexes. We sequenced its genome with high-coverage shotgun reads (14X) using males resulting from 11 successive full-sib matings (3). Two distinct haplotypes were retained despite inbreeding. Their comparison yielded a high estimate of population mutation rate ($\theta=4N_e\mu=0.234$) consistent with large effective population size (N_e) and/or high mutation rate per generation (μ) (4). Sequence comparisons among populations from Eastern Pacific and Eastern Atlantic and within the latter revealed low dN/dS values consistent with strong purifying selection, as expected for large populations (4). Phylogenetic analysis of over 1400 orthologous genes demonstrated that *Oikopleura* is, at the protein level, the fastest evolving animal for which a complete genome is known, despite strong purifying selection (5). Mitochondrial genes heavily modified by oligo-dT insertions also evolved impressively fast (6). Key components of DNA repair (especially in the non-homologous end-joining pathway) were not detected in the genome (7). Coincident rapid evolution of nuclear and mitochondrial genomes may also reflect a highly mutagenic context at the ocean surface.

At 70 megabases with 18,020 predicted genes, the *Oikopleura* genome is unusually compact. Introns are very small (peak at 47 bp, 2.4% > 1kb), as are intergenic spaces, partly due to numerous operons (8). Genes outside operons are also densely packed (53% intergenic distances < 1kb). Even compared with other compact genomes (9), the density of transposable elements (TEs) is low. Most pan-animal TE superfamilies are absent in *Oikopleura*, and only two novel clades of retrotransposons (10) have diversified (11). A massive purge of ancient TEs can be invoked, but TEs currently present in the genome show multiple signs of activity

(11**)**. The low copy number of each element and the uneven genome distribution of main TE clades suggest tight control of their proliferation **(**Fig.1A; (11)**)**.

Two exceptions to global compaction are particularly interesting, as they illustrate where excessive reduction could be harmful. First, a small population of *Oikopleura* genes have relatively large introns and intergenic spaces (Fig.1B). It is enriched for developmentally regulated transcription factor genes, that are long in other genomes due to an abundance of regulatory elements (12). Regulatory element sequences can be highly conserved, though rarely across phyla, but *Oikopleura* homologs of vertebrate conserved elements were not detected (13). Surprisingly, comparison of genes encoding developmental transcription factors from Atlantic and Pacific *Oikopleura* revealed short segments of higher sequence conservation in non-coding regions than in exons, suggestive of a rich regulatory content (Fig.1C; (13)). Interestingly, in a revolution of massive intron loss (see below), *Oikopleura* retained large introns more often than small introns, and the ratio of ancestral to new introns is highest in developmental transcription factor genes (13). Second, Mendelian analysis showed that sex in *Oikopleura* is genetically determined (14) and the genome sequence revealed X and Y chromosomes (Fig.1A). Seven genes on the Y chromosome , all expressed in the testis during spermatogenesis, have giant introns (Fig.1D). Their size probably grew with the non-recombining Y chromosome region, flaunting global compaction.

Oikopleura has a rather common number of introns per gene (4.1), but the turnover of its introns has been extraordinarily high: of 5589 introns mapped by inter-species protein alignments, 76% had positions unique to *Oikopleura* (new introns), 17% were at ancestral positions (old introns) and 7% could not be classified (15). Non-canonical introns, mostly GA-AG and with a very specific acceptor site, are unusually frequent (12%) (Fig2A; (15)). They show several peculiarities, including preferential insertion in phase 1, which is compatible with the current codon usage, as would be expected for the newest introns (16). Indeed, new introns are more often non-canonical than old introns (15). Since *Oikopleura* lacks the minor spliceosome, has only one type of each spliceosomal component (17), we propose that a single and permissive major spliceosome is utilized, with U1snRNP and U2AF able to recognize donor and acceptor sites (18). While cDNA sequence information suggests an efficient splicing for the vast majority of introns, a permissive spliceosome could favour intron gains by correctly splicing out new introns. The pattern of intron loss in *Oikopleura* is consistent with homologous recombination of reverse transcribed mRNA (19) (15). Among hypothetical mechanisms of intron gain, we provide evidence for the insertion of transposon-like elements, and more remarkably for reverse splicing, a reaction in which spliced out

introns can be ectopically reinserted into transcripts (20). We identified 32 compelling candidate introns for transposon insertion (Fig.2B), those matching repetitive elements containing terminal repeats at almost all nucleotides, with exons excluded (15). These introns were usually hemizygous in genotyped individuals, but one individual was homozygous and displayed spliced transcripts (15). We also identified four pairs of nearly identical introns (NII) with no or very weak similarity in flanking exons (Fig.2C), which represent the first reported candidates for reverse splicing (21). All animals with NII were homozygous and displayed spliced transcripts (15). Strikingly, introns of each pair of NII were found within the same gene or the same operon, suggesting intron propagation within their pre-mRNA. Many new introns of *Oikopleura* might have been propagated like these four NII before their sequences diverged, since new introns tend to be adjacent in their host gene (15). Competing mechanisms remain possible. First, introns could be reverse spliced into the genome itself, as can be group II introns (*12*). Some and possibly many new introns of *Oikopleura* could originate by repair of double strand breaks (DSB), as proposed for new introns in *Daphnia* (*13*). However, for the four mentioned intron pairs, a repair after DSB would not readily explain the systematic co-localization of homologous introns in the same transcription unit. No feature in the sequences of those introns in pairs and their surrounding brings particular support for this mechanism (15)."

We explored the *Oikopleura* genome for genes involved in either development or immunity. Many conserved immunity genes failed detection, supporting a minimized immune system consistent with the short *Oikopleura* life history (Tab.1; (24)). While frequent gene losses may have affected families of developmental genes, we were most intrigued by an unusually large number of lineage-specific duplicates, thus far reported for homeobox genes only (25): 87 amplifications accounting for 266 current genes (26), versus 40 amplifications in *Ciona* giving 106 current genes (27). A survival analysis of early duplicates in the genome showed that duplicates are initially lost very rapidly with less relaxed selection than in mammalian genomes (28). In contrast, those that survive beyond 0.02 dS units are relatively more likely to be retained (Fig. 3A; (29)). To understand how older developmental gene duplicates are utilized, we focused on homeobox genes. Strikingly, broad expression signals were detected in the larval trunk epithelium for genes of most amplified groups (16 in 20), but rarely for other groups (1 in 19) (Fig.3B; (30)), likely reflecting roles in patterning of the house-building epithelium (31), a crucial novelty of larvaceans. A preferential retention of duplicates for developmental genes has occurred in vertebrates following whole genome duplications. Their massive retention in *Oikopleura* is exceptional among invertebrates. In

addition to exceptional recruitment for processes like house production through neofunctionalization, another hypothesis may take into consideration the general size reduction of *Oikopleura* genes. Developmental genes that can be very large in other invertebrates should rarely yield intact copies after the local rearrangements that generate duplications, due to greater likelihood of truncation for large genes (32). Other mechanisms may preserve developmental gene duplicates in *Oikopleura*.

Finally, we compared synteny relationships in *Oikopleura* and several invertebrates to ancestral chordate linkage groups (33) (Fig.3C; (34)). Amphioxus, *Ciona, Caenorhabditis* and sea anemone showed many cases of conserved chromosomal synteny. *Oikopleura* orthologs showed no such conservation. We also measured local synteny conservation between the same species and human (34). Amphioxus, *Ciona* and *Caenorhabditis* and sea anemone (to a much lower degree) displayed several-fold better conserved neighbourhoods than expected by chance. *Oikopleura* showed a local gene order indistinguishable from random for distances smaller than 30 genes, and a modest level of conserved synteny at larger distances (34).

We show that multiple genome organization features, conserved across metazoans including other tunicates and non-bilaterians, have dramatically changed in the *Oikopleura* lineage. Despite an unprecedented genome revolution, the *Oikopleura* lineage preserved essential morphological features, even maintaining the chordate body plan to the adult stage, unlike other tunicates. Evolution in this lineage was rapid and probably took place in a context favouring purifying selection against mildly deleterious features. Our results strengthen the view that global similarities of genome architecture from sponge to human (33,35-37) are not essential for the preservation of ancestral morphologies, as is widely believed (38-40).

REFERENCES

1. F. Delsuc, H. Brinkmann, D. Chourrout, H. Philippe, *Nature* **439,** 965-968 (2006).
2. J. M. Bouquet *et al., J. Plankton Res.* **31,** 359-370 (2009).
3. SOM:text§1,fig.S1-S2,tab.S1-S3.
4. SOM:text§2.
5. SOM:text§3,fig.S3,tab.S4-S6.
6. SOM:text§4,fig.S4-S6.
7. SOM:text§5,fig.S7,tab.S7.
8. SOM:text§6,fig.S8,tab.S8.
9. T. H. Eickbush, A. V. Furano, *Curr. Opin. Genet. Dev.* **12,** 669-674 (2002).
10. J. N. Volff, H. Lehrach, R. Reinhardt, D. Chourrout, *Mol. Biol. Evol.* **21,** 2022-2033 (2004).
11. SOM:text§7,fig.S9-S16.

12. A. Woolfe *et al., PLoS Biol.* **3,** e7 (2005).
13. SOM:text§8,fig.S17-S19.
14. SOM:text§9,fig.S20,tab.S10.
15. SOM:text§10,fig.S21-S31,tab.S11-S20.
16. H. D. Nguyen, M. Yoshihama, N. Kenmochi, *BMC Evol. Biol.* **6**, 69 (2006).
17. M. Marz, T. Kirsten, P. F. Stadler, *J. Mol. Evol.* (2008).
18. D. A. Zorio, T. Blumenthal, *Nature* **402**, 835-838 (1999).
19. T. Mourier, D. C. Jeffares, Eukaryotic *Science* **300**, 1393 (2003).
20. S. W. Roy, W. Gilbert, *Nat. Rev. Genet.* **7**, 211-221 (2006).
21. S. W. Roy, M. Irimia, *Trends Genet.* **25**, 67-73 (2009).
22. B. Cousineau *et al., Cell* **94,** 451-462 (1998).
23. W. Li *et al., Science* **326,** 1260-1262 (2009).
24. SOM:text§11,tab.S21.
25. R. B. Edvardsen *et al., Curr. Biol.* **15,** R12-3 (2005).
26. SOM:text§12,tab.S22.
27. Y. Satou, N. Satoh. *et al., Dev. Genes Evol.* **213,** 211-318 (2003).
28. T. Hughes, D. A. Liberles, *J. Mol. Evol.* **65,** 574-588 (2007).
29. SOM:text§13,fig.S32-S34,tab.S23.
30. SOM:text§14,fig.S35,tab.S24.
31. E. M. Thompson *et al., Dev. Biol.* **238,** 260-273. (2001).
32. V. Katju, M. Lynch, *Genetics* **165**, 1793–1803 (2003).
33. N. H. Putnam *et al., Nature* **453,**1064-1071 (2008).
34. SOM:text§15,fig.S36-S38,tab.S25.
35. N. H. Putnam *et al., Science* **317,** 86-94 (2007).
36. 1. M. Srivastava *et al., Nature* **466**, 720-726 (2010).
37. M. Srivastava *et al., Nature* **454**, 955-960 (2008).
38. M. Lynch, J. S. Conery, *Science* **302,** 1401-1404 (2003).
39. M. Lynch, *Mol. Biol. Evol.* **23,** 450-468 (2006).
40. M. Lynch, *Proc. Natl. Acad. Sci. USA*. **104,** 8597-8604 (2007).

ACKNOWLEDGEMENTS

This article is dedicated to Hans Prydz and Kaare Rommetveit for crucial contributions to the Sars Centre establishment. The Sars Centre budget, the FUGE Programme of Norwegian Research Council, Genoscope, IOS-0719577 and DBI-0743374 supported the research. GENBANK/EMBL sequence accession numbers are CABV01000001-CABV01005917, CABW01000001-CABW01006678, FN653015-FN654274, FN654275-FN658470, FP700189-FP710243, FP710258-FP791398, FP791400-FP884219.

FIGURE AND TABLE LEGENDS

Figure 1. Genome compaction features. (A) Chromosome regions assembled with physical links and genetic markers**.** The location of TEs is indicated with horizontal lines (left sides: DNA transposons; right sides: short lines for LTR-retrotransposons and long lines for LINEs). **(B)** Distribution of gene models over 10% abundance classes of intron size and upstream intergenic distance for 8812 non-operon genes (left); 189 developmentally regulated genes,

mainly transcription factors (right). **(C)** Conserved elements revealed in *Oikopleura* inter-ocean alignments: density of conserved blocks (top) gene annotation (middle) and perfectly conserved elements >100 bp on grey line (blue=Norway vs. Northwest America; red=Norway vs. Japan). **(D)** Expression of a giant Y gene observed by RT-PCR and *in situ* hybridization.

Figure 2. Intron gain scenarios and candidate introns. (A) Main intron logos **(B)** Transposon insertion: duplicated insertion sites (framed in blue) allow MITE-like insertions to be spliced out exactly (red: exons, black: introns). **(C)** Reverse splicing: 4 pairs of homologous introns (black) and their immediate exonic environments (red).

Figure 3. Changes affecting the gene complement. (A) Early gene duplicates. Lower: Histogram of binned recent duplicate pairs; mixture model (discrete distribution plus truncated Weibull distributions) accommodating heterogeneous birth/death processes is fitted. Inset: Nonsynonymous substitution accumulation declines with ongoing synonymous substitution. **(B)** Expression of amplified homeobox gene groups in the trunk epithelium of larvae (arrows). **(C)** Loss of ancestral gene order. Positions of orthologous genes in a given metazoan genome (Y-axis) compared to ancestral chordate linkage groups (CLGs, X-axis). The width of CLGs corresponds to the number of orthologs in species. Amphioxus and sea anemone genome segments represent the largest 25 assembled scaffolds, while *Ciona*, nematode and *Oikopleura* segments are chromosomes.

Table 1. Minimal immune system predicted from the *Oikopleura* genome. Numbers of genes or domains in families encoding potential immunity factors.

	D.m.	*S.p.*	*O.d.*	*C.i.*	*B.f.*	*L.f.*	*H.s.*
sensors							
TLR	9	222	**1**	3	48	21	10
NLR	0	203	**0**	20	92	140-220	20
SRCR	14	218	**1**	81	270	287	81
PGRP	15	5	**4**	6	>20	ND	6
RIG-I-like helicases	0	12	**0**	ND	7	ND	3
C-type lectins	32	104	**31**	120	1215	ND	81
IgSF-ITIM	>3	ND	**5**	>6	>5	>3	>50
adaptors							
MyD88-like (DEATH-TIR)	1	4	**0**	1	4	ND	1
SARM1-like, TIRAP-like, TICAM2-like	1	15	**0**	>2	12	ND	3
potential effector							
PLA2	8	65	**128**	7	>7	ND	11

D.m.= Drosophila melanogaster, S.p.= Strongylocentrotus purpuratus, O.d.= Oikopleura dioica, C.i.= Ciona intestinalis, B.f.= Branchiostoma floridae, L.f.= Lampetra fluviatilis, H.s.=Homo sapiens

Figure_1

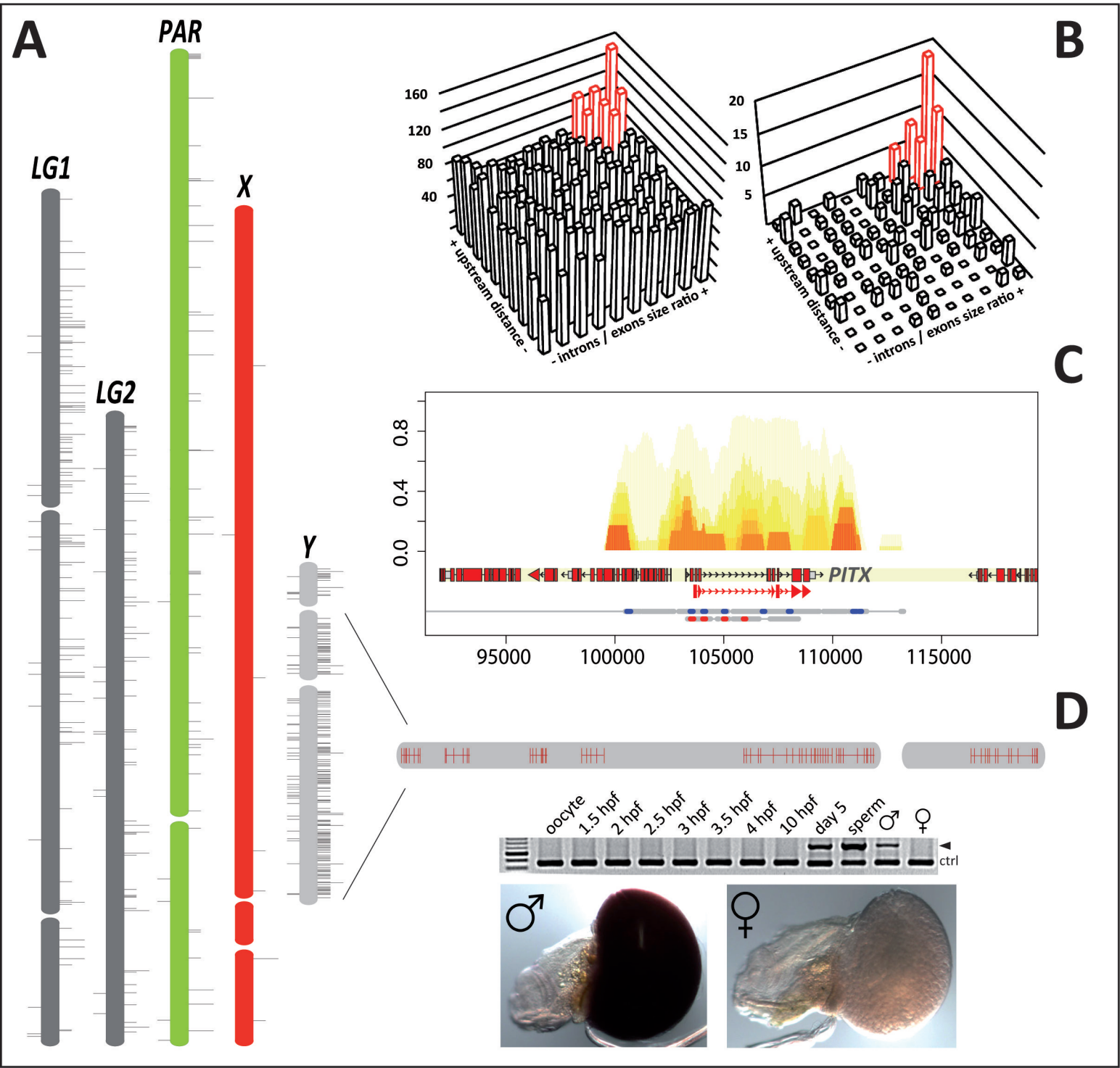
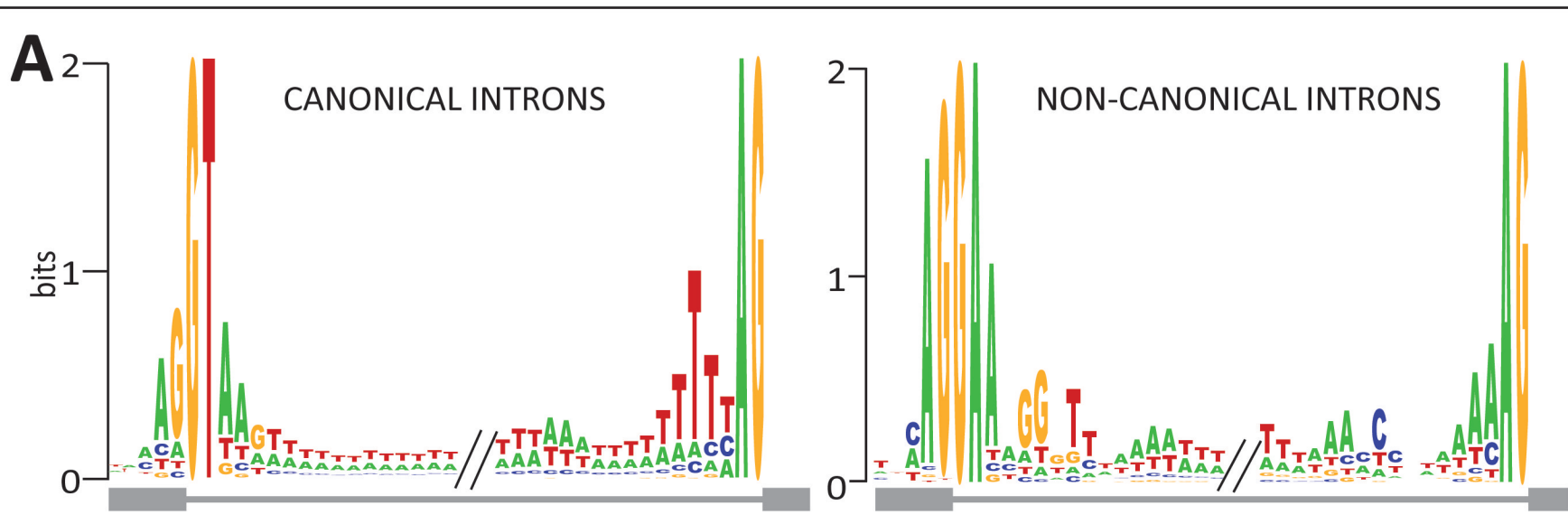
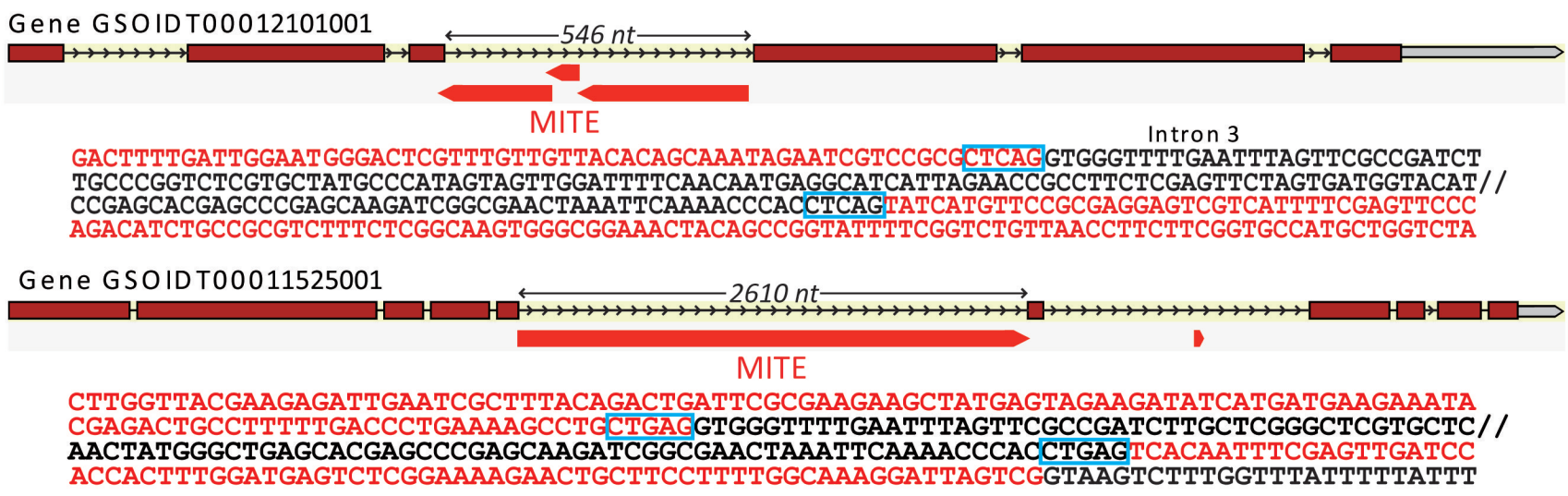

A
PAR
LG1
LG2
X
Y
B
160
120
80
40
+ upstream distance -
- introns / exons size ratio +
20
15
10
5
+ upstream distance -
- introns / exons size ratio +
C
0.8
0.4
0.0
PITX
95000
100000
105000
110000
115000
D
oocyte
1.5 hpf
2 hpf
2.5 hpf
3 hpf
3.5 hpf
4 hpf
10 hpf
day 5
sperm
♂
♀
ctrl
♂
♀

Figure_2


A
CANONICAL INTRONS
NON-CANONICAL INTRONS
bits
2
1
0


B
HOMOLOGY WITH TRANSPOSON-LIKE REPEATS
Gene GSOIDT00012101001
546 nt
MITE
Intron 3
GACTTTTGATTGGAATGGGACTCGTTTGTTGTTACACAGCAAATAGAATCGTCCGCGCTCAGGTGGGTTTTGAATTTAGTTCGCCGATCT
TGCCCGGTCTCGTGCTATGCCCATAGTAGTTGGATTTTCAACAATGAGGCATCATTAGAACCGCCTTCTCGAGTTCTAGTGATGGTACAT//
CCGAGCACGAGCCCGAGCAAGATCGGCGAACTAAATTCAAAACCCACCTCAGTATCATGTTCCGCGAGGAGTCGTCATTTTCGAGTTCCC
AGACATCTGCCGCGTCTTTCTCGGCAAGTGGGCGGAAACTACAGCCGGTATTTTCGGTCTGTTAACCTTCTTCGGTGCCATGCTGGTCTA
Gene GSOIDT00011525001
2610 nt
MITE
CTTGGTTACGAAGAGATTGAATCGCTTTACAGACTGATTCGCGAAGAAGCTATGAGTAGAAGATATCATGATGAAGAAATA
CGAGACTGCCTTTTTGACCCTGAAAAGCCTGCTGAGGTGGGTTTTGAATTTAGTTCGCCGATCTTGCTCGGGCTCGTGCTC//
AACTATGGGCTGAGCACGAGCCCGAGCAAGATCGGCGAACTAAATTCAAAACCCACCTGAGTCACAATTTCGAGTTGATCC
ACCACTTTGGATGAGTCTCGGAAAAGAACTGCTTCCTTTTGGCAAAGGATTAGTCGGTAAGTCTTTGGTTTATTTTTATTT

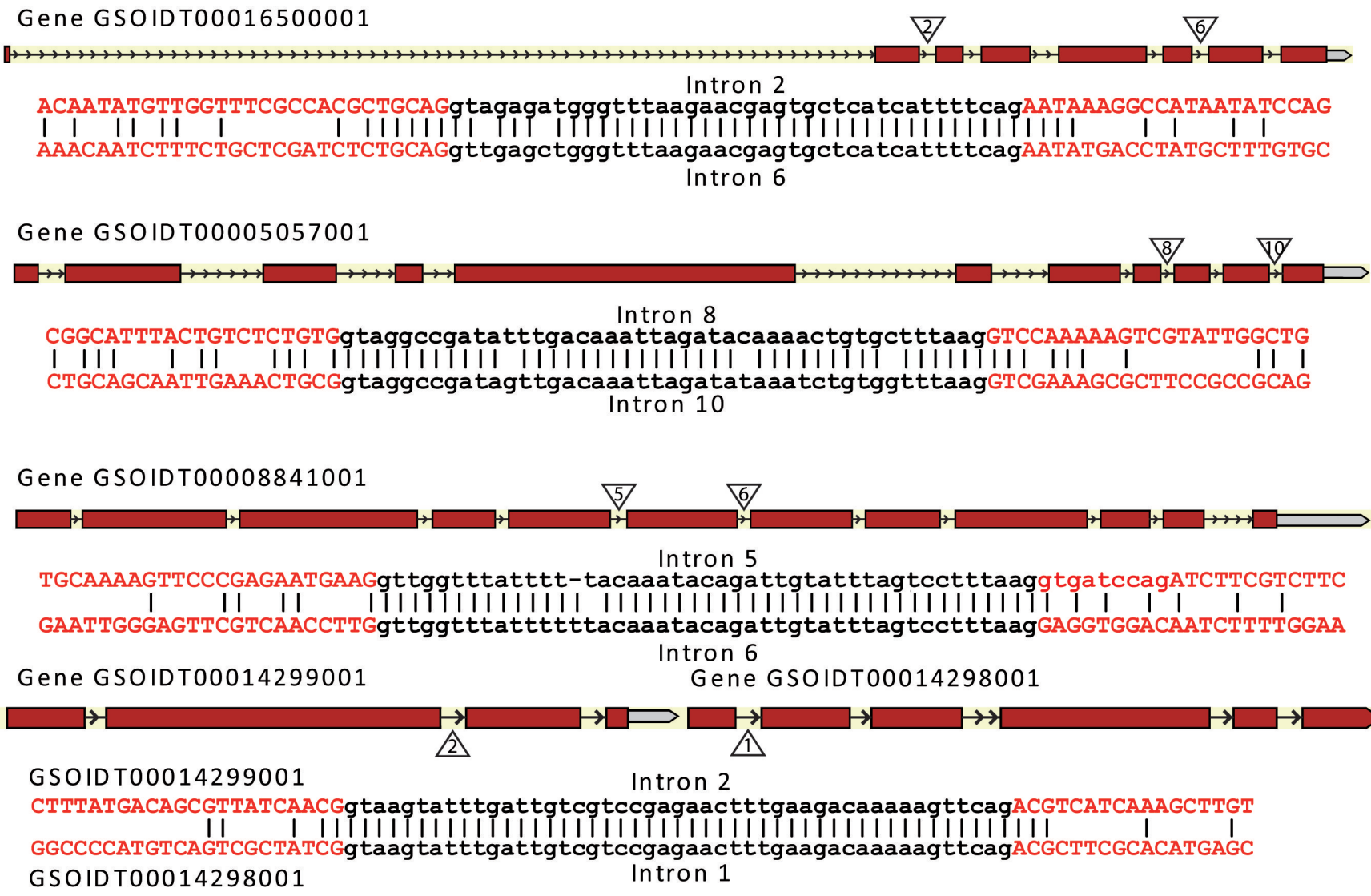
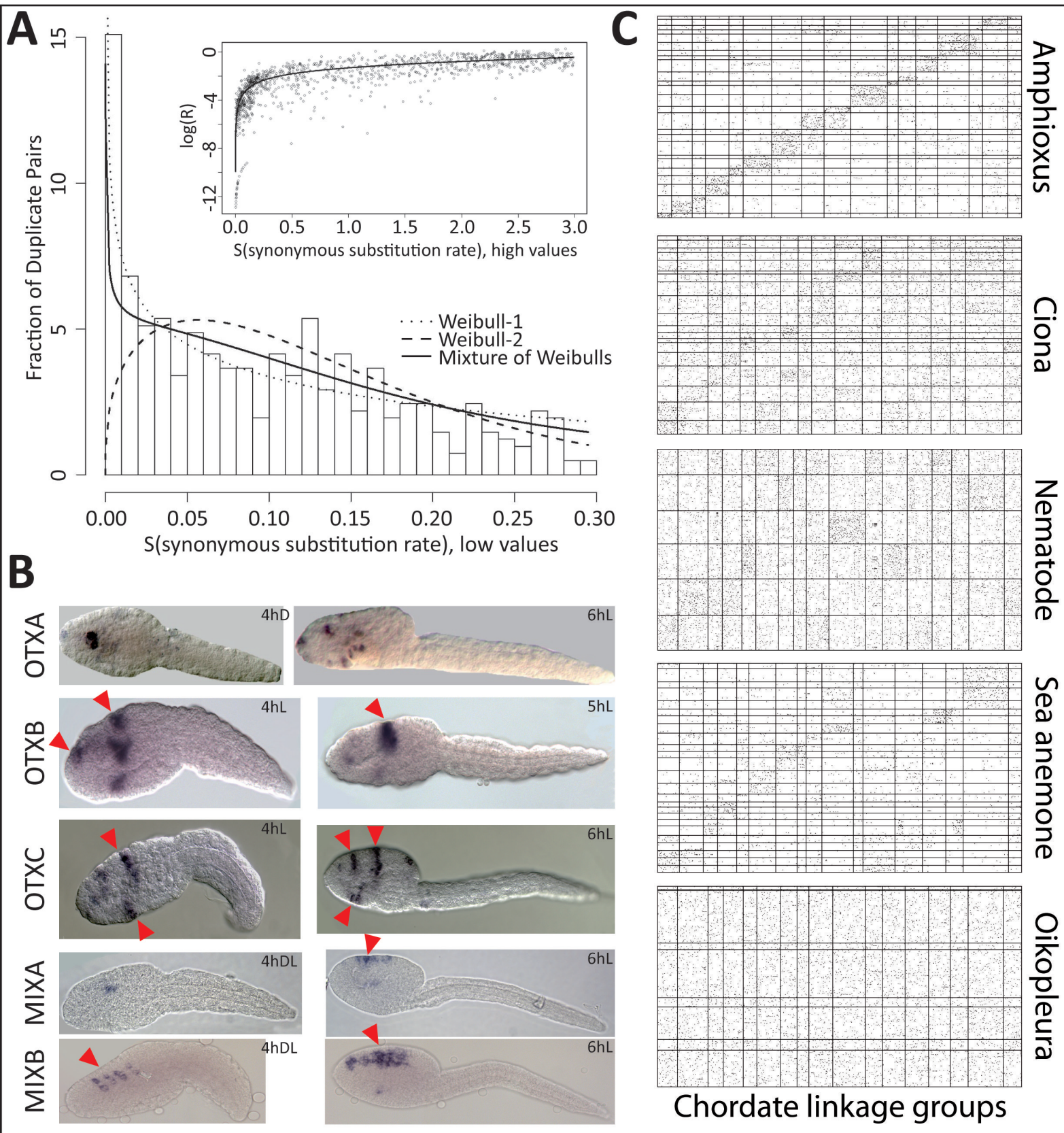

C
PAIRS OF HOMOLOGOUS INTRONS SUPPORTNG REVERSE SPLICING
Gene GSOIDT00016500001
2
6
Intron 2
ACAATATGTTGGTTTCGCCACGCTGCAGgtagagatgggtttaagaacgagtgctcatcatttcagAATAAAGGCCATAATATCCAG
AAACAATCTTTCTGCTCGATCTCTGCAGgttgagctgggtttaagaacgagtgctcatcatttcagAATATGACCTATGCTTTGTGC
Intron 6
Gene GSOIDT00005057001
8
10
Intron 8
CGGCATTTACTGTCTCTGTGgtaggccgatatttgacaaattagatacaaaactgtgctttaagGTCCAAAAAGTCGTATTGGCTG
CTGCAGCAATTGAAACTGCGgtaggccgatagttgacaaattagatataaatctgtggtttaagGTCGAAAGCGCTTCCGCCGCAG
Intron 10
Gene GSOIDT00008841001
5
6
Intron 5
TGCAAAAGTTCCCGAGAATGAAGgttggtttatttt-tacaaatacagattgtatttagtcctttaaggtgatccagATCTTCGTCTTC
GAATTGGGAGTTCGTCAACCTTGgttggtttatttttttacaaatacagattgtatttagtcctttaagGAGGTGGACAATCTTTTGGAA
Intron 6
Gene GSOIDT00014299001
Gene GSOIDT00014298001
2
1
GSOIDT00014299001
Intron 2
CTTTATGACAGCGTTATCAACGgtaagtatttgattgtcgtccgagaactttgaagacaaaaagttcagACGTCATCAAAGCTTGT
GGCCCCATGTCAGTCGCTATCGgtaagtatttgattgtcgtccgagaactttgaagacaaaaagttcagACGCTTCGCACATGAGC
GSOIDT00014298001
Intron 1

Figure_3


A
Fraction of Duplicate Pairs
0
5
10
15
log(R)
0
-4
-8
-12
0.0
0.5
1.0
1.5
2.0
2.5
3.0
S(synonymous substitution rate), high values
Weibull-1
Weibull-2
Mixture of Weibulls
0.00
0.05
0.10
0.15
0.20
0.25
0.30
S(synonymous substitution rate), low values
B
OTXA
4hD
6hL
OTXB
4hL
5hL
OTXC
4hL
6hL
MIXA
4hDL
6hL
MIXB
4hDL
6hL
C
Amphioxus
Ciona
Nematode
Sea anemone
Oikopleura
Chordate linkage groups