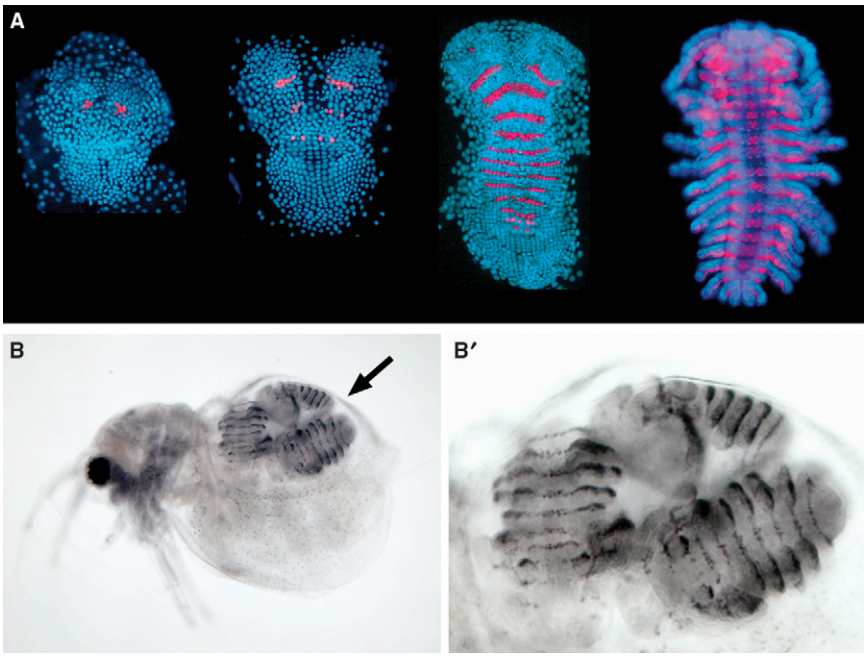

Figure 3. An example of sequential segmentation frequently observed in crustacean development.

(A) Series of sequentially older embryos of the amphipod *Parhyale hawaiensis* (anterior is oriented up). All the nuclei of the animal are shown in blue, while stripes of expression of the Engrailed protein (in red) mark the boundaries between segments. As the embryo develops, segments are added from the head to the tail. (B) Here the developing embryos (arrow) of the branchiopod *Moinodaphnia* are shown inside the brood pouch of the mother (Engrailed stripes are black). (B') Higher magnification of the embryos within the brood pouch.

a boom in a wild population, although with some unintended consequences: As a way to feed the Soviet Union's growing northern population, king crabs native to the North Pacific were introduced into the Barents Sea in the 1960s. Now thriving, this population has been welcomed by the fishing industry, as these crabs are a profitable catch. Unfortunately, these crabs are also ravenous omnivores and have few natural predators in these waters. Their feeding activities strip bare huge swaths of seabed, depleting capelin and cod stocks by consumption of the fishes' eggs and competition for food. The king crabs are steadily making their way south towards European waters and despite short-term benefits damage to fisheries in the long term seems likely. Other crustacean species have also become invasive, including Chinese mitten crabs, which have spread through the ballast water of ships, and the crayfish *Procambarus clarkii*, a native of the US Gulf Coast, that has displaced many other crayfish from their habitats. The recent finding of a parthenogenetic crayfish strain in the aquarium trade also poses a potential threat if it makes its way into the wild.

Crustaceans are clearly a remarkable group of organisms with a long evolutionary history and remarkable adaptability. Given their primarily aquatic habitats, however, they are not as well studied as their terrestrial arthropod relatives, but our knowledge of their diversity, behavior, development, and physiology is growing rapidly.

Further reading

Averof, M., and Patel, N.H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682–686.

Browne W.E., Price, A.L., Gerberding, M., and Patel, N.H. (2005). Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. Genesis *42*, 124–149.

Brusca, R.C., and Brusca, G.J. (2003). Invertebrates (Sunderland, MA: Sinauer Associates, Inc.)

Duffy, J.E. (1996). Eusociality in a coral-reef shrimp. Nature *381*, 512–514.

Eads, B.D., Andrews, J., and Colbourne, J.K. (2008). Ecological genomics in Daphnia: stress responses and environmental sex determination. Heredity *100*, 184–190.

Schram, F.R. (1986). Crustacea (New York, NY: Oxford University Press).

Stott, R. (2003). Darwin and the Barnacle: The Story of One Tiny Creature and History's Most Spectacular Scientific Breakthrough. (New York, NY: W.W. Norton & Company, Inc.)

Department of Integrative Biology and Department of Molecular Cell Biology
University of California, Berkeley,
CA 94530-3140, USA.
*Email: nipam@uclink.berkeley.edu

Correspondences

Widespread recycling of processed cDNAs in dinoflagellates

Claudio H. Slamovits and Patrick J. Keeling

Dinoflagellates are ubiquitous algae with extraordinary nuclear genomes that are among the largest known [1–3]. Dinoflagellate nuclear genes are also exceptional in that an invariant 22 bp trans-spliced leader (SL) caps most or all mRNAs [4,5]. We re-examined cDNAs from diverse dinoflagellates, and show that about 25% of them include additional, sometimes multiple, relict SL sequences in tandem. All of these relict SLs are truncated after nucleotide 7 of the canonical SL, corresponding to an AG dinucleotide. Genomic sequences confirm that these additional SLs are genome-encoded, and that the relict SL provides the AG acceptor site for subsequent trans-splicing. Altogether, this shows that large numbers of genes in dinoflagellate genomes have been cycled through an mRNA intermediate at least once in recent evolutionary history, an event potentially encouraged by the conservation of splice acceptor sites within the SL. This dynamic cycling between the genome and mRNA predicts that introns will be rare in dinoflagellate genomes since they would be purged with each cycle, and those that do remain are most likely to have been inserted relatively recently.

An identical 22 bp leader is found on the 5' end of most or all mRNAs in all known dinoflagellates, making the trans-splicing mechanism that adds this leader both ancient and pan-genomic. We aligned the 5' end of approximately 500 full-length dinoflagellate cDNAs and found the conserved 22 bp SL sequence in all cases, but also found evidence for a second, relict SL immediately downstream of the first one in over 100 cDNAs (Figure 1A, and Figures S1 and S2 in Supplemental Data, published with this article online). A third SL was also present in six

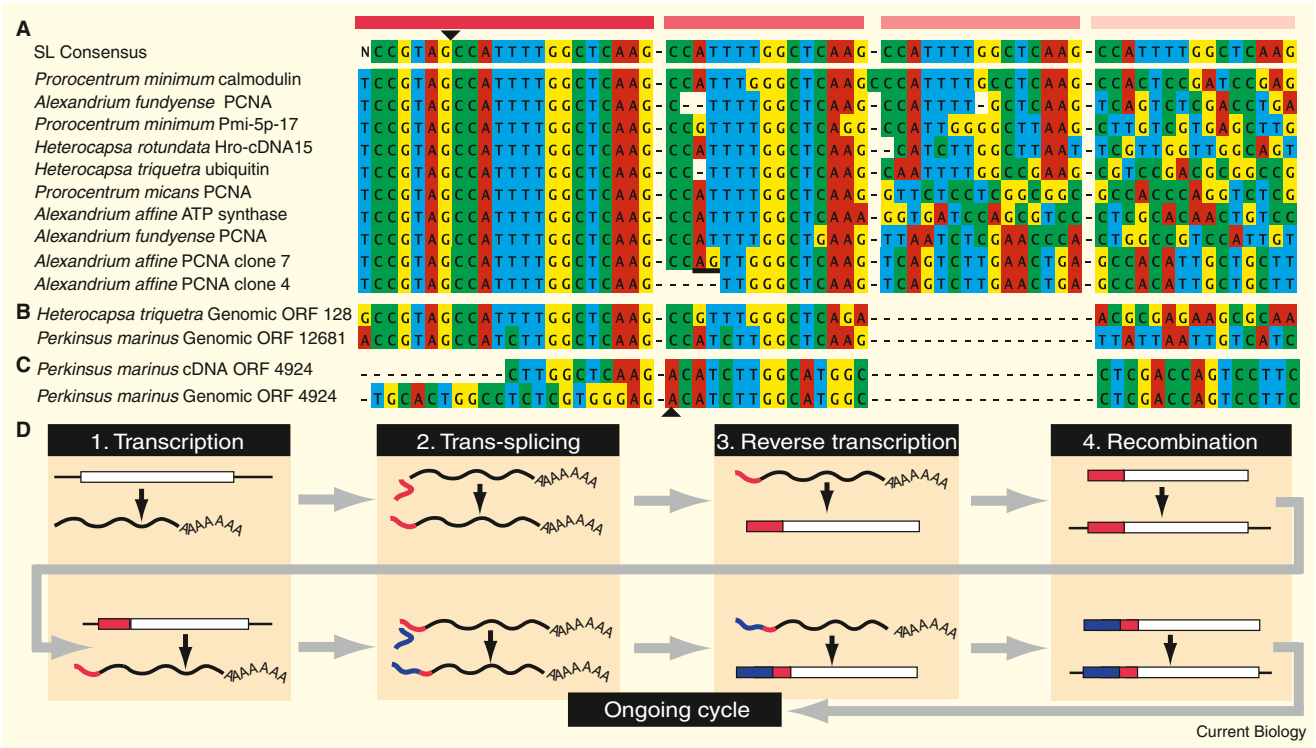

Figure 1. Dinoflagellate mRNAs contain multiple, tandem, relict SL sequences.

(A) Alignment of 5' ends of dinoflagellate cDNAs. The top line is a canonical SL repeated from nucleotides 8 to 22 for reference. The arrowhead indicates the AG splice-acceptor used in subsequent trans-splicing. Below this are examples where a canonical SL is followed by 1, 2, or potentially 3 (top two lines) relict SLs. The last two *A. affine* PCNA cDNAs show how both canonical and alternative AG-acceptor sites (underlined) have been used in mRNAs from one gene. (B) Examples of genome-encoded relict double SL sequences upstream of *H. triquetra* and *P. marinus* genes. (C) Alignment of homologous cDNA and genomic DNA sequences from *P. marinus*. The cDNA is 5' truncated. The canonical SL has been added to the splice-acceptor site of the genome-encoded relict SL at the arrowhead. (D) A cyclical process for progressive addition of SLs to dinoflagellate genes. Transcription (1) is followed by trans-splicing (2 – SL in red). Reverse transcription (3) and insertion of a processed cDNA into the genome by recombination results in a gene with a relict SL (4). Transcription of this gene (1 – lower) leads to the addition of a second SL (2 – SL in blue) at the AG dinucleotide at position 6–7 of the relict SL. Relict SL sequences decay over time, but new relict SLs are also continuously added, refreshing the pool of genes with recognizable SL relicts.

cDNAs, and in two of these there was limited sequence similarity to suggest a possible fourth SL (Figure 1A). One relict SL was 100% identical to the canonical SL sequence, but generally substitutions or small deletions distinguish them from the canonical sequence.

The most conspicuous difference between canonical and relict SLs is that the latter are invariably truncated and, with rare exceptions, they are truncated at exactly the same position: after nucleotide 7 of the canonical SL sequence (arrowhead in Figure 1A). Interestingly, nucleotides 6 and 7 correspond to an AG dinucleotide, which is the acceptor site for spliceosomal introns [6], leading to the obvious possibility that the upstream SL is spliced onto a pre-existing SL sequence. Further support for this idea comes from the single, unambiguous exception: a mutation in the relict SL from the *Alexandrium affine* proliferating cell nuclear antigen (PCNA) gene has created a novel AG dinucleotide at positions 10 and 11. In one mRNA transcribed from this gene, the SL has been trans-spliced at this acceptor site, whereas in a second mRNA the SL is trans-spliced at the standard position (underlined in Figure 1A).

Tandemly duplicated SL sequences could arise by trans-splicing one onto another following transcription, but the situation appears to be more complicated. Because the SL sequence is strictly conserved, sequential trans-splicing onto a single mRNA would lead to identical tandem repeats. In contrast, the sequence conservation of relict SL sequences ranges from 100% to barely detectable (Figures 1A, S1, and S2), so they cannot have been added sequentially to the same mRNA. Moreover, in cases where there are two relict SLs, the downstream relict is more divergent, suggesting it is older. Lastly, a 5' fragment of only 7 bp is likely to be too short to support the trans-splicing reaction, suggesting that the precursor mRNA must extend beyond the start of the relict SL.

These characteristics are all consistent with the progressive accumulation of relict SLs on genes in the genome itself. Direct support for this comes from genomic data from *Heterocapsa triquetra.* In a genome sequence survey, we identified one ORF with two tandemly duplicated SL sequences 66 bp upstream of the start codon. The first repeat was complete and 100% identical to the canonical SL sequence, whereas the downstream copy had three substitutions and was truncated after nucleotide 7 (Figure 1B). Similarly, from the *Perkinsus marinus* genome project, we identified several genes with one or two relict SL sequences upstream of genes (e.g., Figure 1B).

We also identified three cases in which data from both a gene and its transcript could be compared. For example (Figure 1C), an ORF with a divergent but recognizable relict SL 336 bp upstream of the start codon is also represented in the *P. marinus* expressed sequence tag (EST) data, and the EST encodes a partially truncated SL identical to the *P. marinus* consensus that has been trans-spliced at the expected position within the genome-encoded relict SL (arrowhead, Figure 1C).

Altogether, these data show that a significant proportion of genes in dinoflagellate genomes have recently been recycled from processed mRNA. This chain of events is shown in Figure 1D. Briefly, transcripts capped with SL sequences by trans-splicing are reverse-transcribed and re-integrated into the genome. Upon expression of these newly integrated genes, a second SL is added to the 5' end of the now-relict SL at the splice-acceptor site. The integration of processed cDNAs into the genome cannot involve homologous replacement of the locus from which it was transcribed because this would lead to the loss of the cDNA ends, including the relict SL. Instead, processed cDNAs are more likely to be integrated by non-homologous recombination at some other site. In this scenario, only the new copy of the gene would encode the relict SL, so how does this copy become fixed in the genome? Given sufficient time (e.g., the entire evolutionary history of dinoflagellates), even random loss would inevitably lead to the accumulation of many relict SL sequences, because the process is a ratchet. Losing the original copy of the gene leads to the retention of the processed cDNA, which is irreversible (except for the loss of detectability by accumulated mutations as discussed below). Losing the recycled cDNA merely returns the gene to its original state, which is reversible by any subsequent recycling that leads to the replacement of the original copy, making the process inevitable given sufficient time.

The appearance of multiple relict SLs on a single gene and the range of conservation of relict SL sequences attests to this being an ongoing process. The absence of detectable relict SLs on most genes may reflect some bias for certain genes to be recycled (e.g., highly expressed genes should be recycled more frequently because more mRNA is available); however, it is also possible that most genes have been recycled at some point, but that their relict SL sequences have mutated beyond recognition. Indeed, if the relict SLs are not under strong selection (other than the splice acceptor site), then the cycle may be operating at a very high frequency and it is possible that all dinoflagellate genes have been recycled in the past. The frequency of mRNA recycling may also be underestimated because trans-splicing could completely excise relict SL sequences in two ways. First, when two relict SLs are present, trans-splicing to the downstream copy would remove the upstream one entirely. Second, the last 2 nucleotides of the canonical SL sequence are AG (Figure 1A), and trans-splicing to this acceptor site would also excise the entire relict SL (indeed, this may be the case with another *A. affine* PCNA gene: Figure S3). Unfortunately, analysis of cDNAs will reveal neither of these events unequivocally.

Genes that cycle continuously between DNA and RNA and a genome that readily absorbs processed cDNAs and puts them back into service as genes are two more unusual features of the already strange genomes of dinoflagellates [2,3]. The constant recycling of mRNA may be favoured by the expansive nature of dinoflagellate genomes, although the *P. marinus* genome lacks many of the most extreme characteristics and still appears to recycle mRNA. The conservation of two AG splice-acceptor sites in the canonical SL sequence might also have favoured this process. Attaching two potentially viable splice-acceptor sites to the 5' end of all mRNAs pre-equips recycled genes for subsequent capping by trans-splicing. Assuming that the addition of the SL at the mRNA level is essential, the SL sequence itself favours its own presence at the 5' ends of genes. Another major implication for the genome is that with every pass through the DNA–RNA cycle, the gene will be purged of introns, leading to two predictions. First, introns should be relatively rare in dinoflagellate genomes in general. Second, those genes with the most highly conserved relict SLs (i.e., the most recently recycled genes) should have especially few introns, and whatever introns they do contain are predicted to have been inserted recently. Indeed, any intron in any gene with a relict SL has probably been inserted after the gene was recycled from mRNA. Spliceosomal intron insertions and mobility has been an elusive phenomenon [6,7], and recently recycled dinoflagellate genes may help illuminate the origin of new introns. Lastly, it should be pointed out that the presence of trans-splicing does not necessarily drive the genome to recycle mRNAs [8,9] — SLs are only an indicator of recycling in dinoflagellates. It is possible such recycling takes place at comparable frequencies in other genomes where no such obvious indicator is available.

Supplemental data

Supplemental data are available at http://www.current-biology.com/cgi/content/full/18/13/R550/DC1

Acknowledgments

This work was supported by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. We thank J. Saldarriaga, M. Gray, and M. McEwan. The *P. marinus* genome project (www.tigr.org/tdb/e2k1/pmg/) is a joint TIGR-COMB initiative.

References

1. Keeling, P.J., and Slamovits, C.H. (2005). Causes and effects of nuclear genome reduction. Curr. Opin. Genet. Dev. *15*, 601–608.
2. Rae, P.M. (1976). Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrrophyta (dinoflagellates). Science *194*, 1062–1064.
3. Wargo, M.J., and Rizzo, P.J. (2001). Exception to eukaryotic rules. Science *294*, 2477.
4. Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T., and Lin, S. (2007). Spliced leader RNA trans-splicing in dinoflagellates. Proc. Natl. Acad. Sci. USA *104*, 4618–4123.
5. Lidie, K.B., and van Dolah, F.M. (2007). Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. J. Eukaryot. Microbiol. *54*, 427–435.
6. Roy, S.W., and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. Nat. Rev. Genet. *7*, 211–221.
7. Coghlan, A., and Wolfe, K.H. (2004). Origins of recently gained introns in *Caenorhabditis*. Proc. Natl. Acad. Sci. USA *101*, 11362–11367.
8. Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., *et al*., (2002). A global analysis of Caenorhabditis elegans operons. Nature *417*, 851–854.
9. Simpson, A.G., Stevens, J.R., and Lukes, J. (2006). The evolution and diversity of kinetoplastid flagellates. Trends Parasitol. *22*, 168–174.

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver V6T 1Z4, BC, Canada.
E-mail: pkeeling@interchange.ubc.ca