

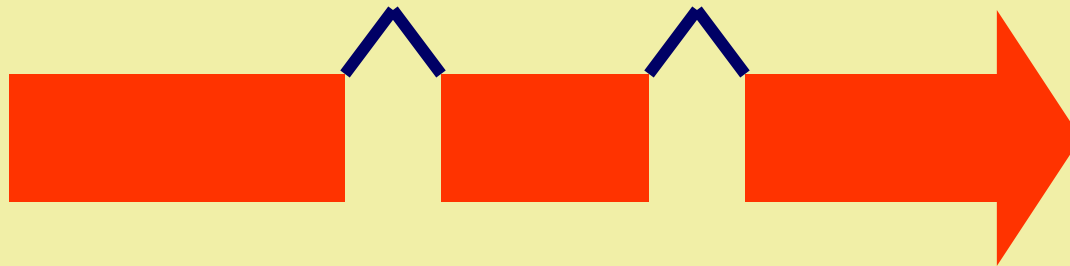
Origins of recently gained introns in *Caenorhabditis*



Avril Coghlan and Kenneth H. Wolfe
Department of Genetics,
Trinity College Dublin, Ireland

Rate of intron gain and loss

- Over the long term, losses and gains occurred at high rates in eukaryotes
- Over the short term, losses are more frequent than gains



Recent intron gains

- mosquito TPI gene



- fly XDH gene



- midge globin genes



- plant catalase gene



- worm chemoreceptor genes



Intron gain in nematodes

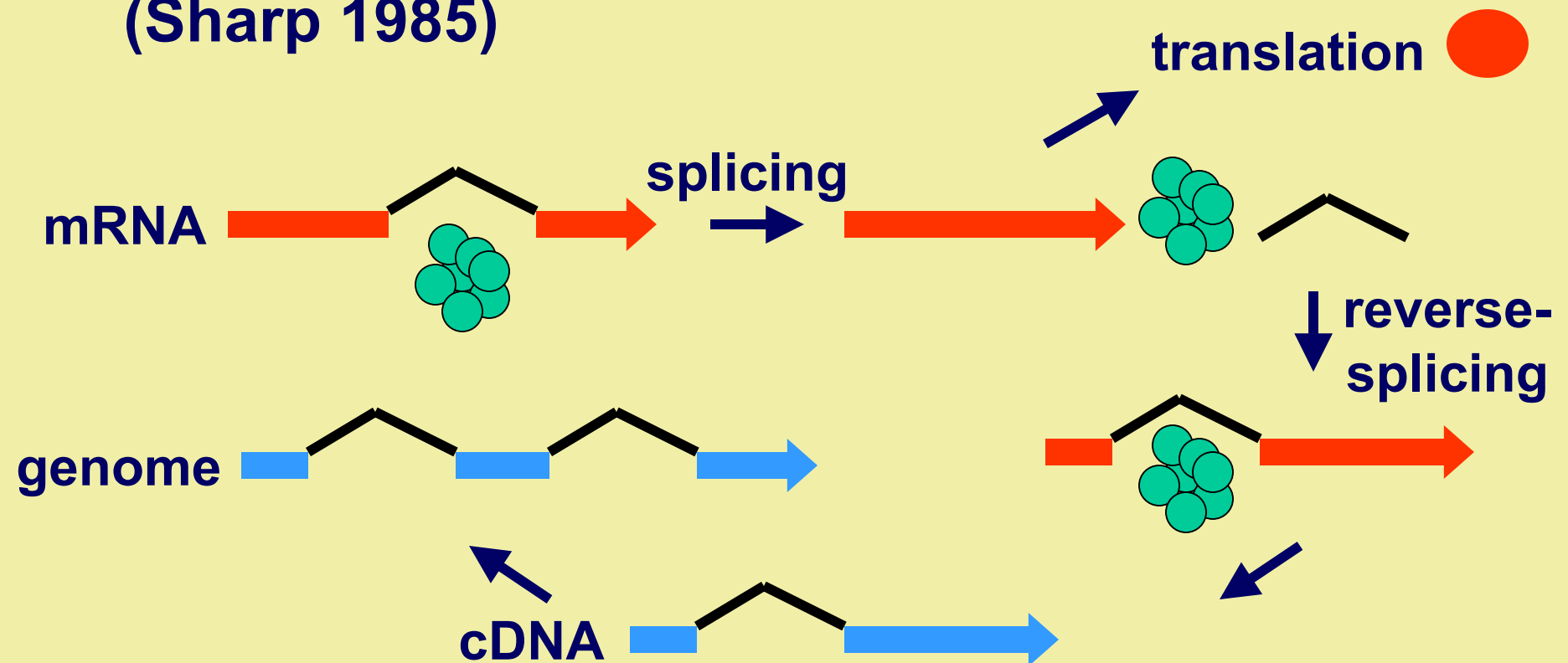


- 4400 *C. elegans*- and 2200 *C. briggsae*-specific introns
- 0.005 gains/losses per gene per My in nematodes

Mechanism of intron gain

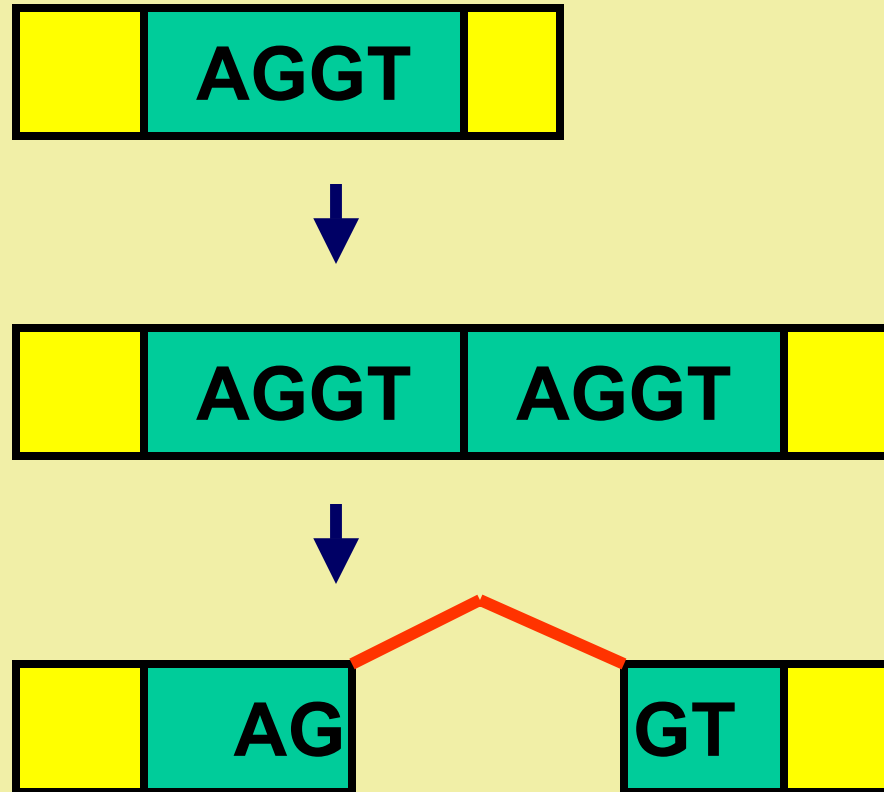
? Insertion of a transposon (Crick 1979)

? Reverse-splicing of a pre-existing intron (Sharp 1985)

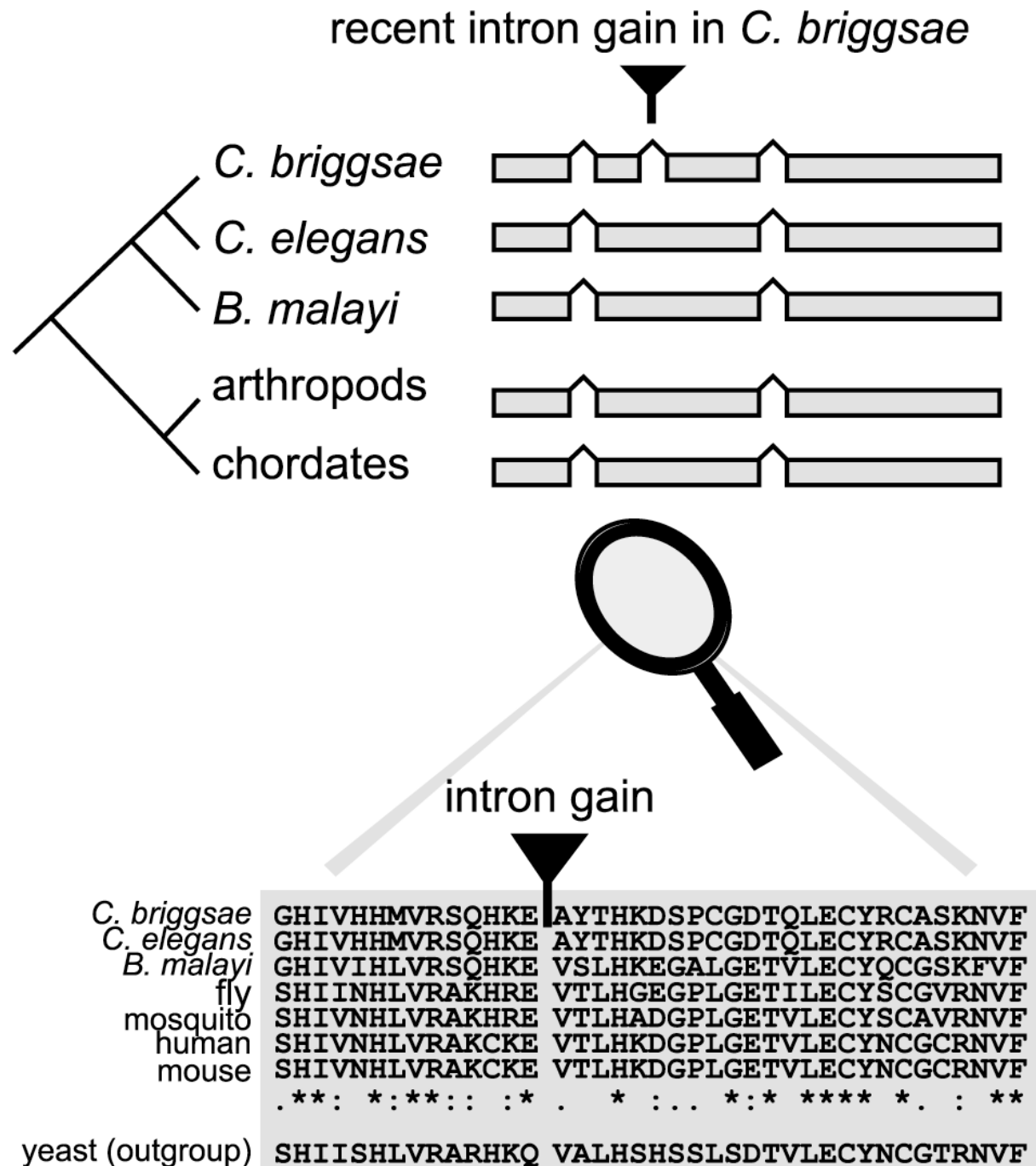


Mechanism of intron gain

? Tandem duplication of an exon fragment containing AGGT (Rogers 1989)



Method



Exon splice site consensus



-2 -1

+1 +2

... A G

G T...

C.e. control 56 60

30 38

81 → C.e. novel 79 84

54 46

C.b. control 55 59

30 37

41 → C.b. novel 71 88

54 39

Germline expression



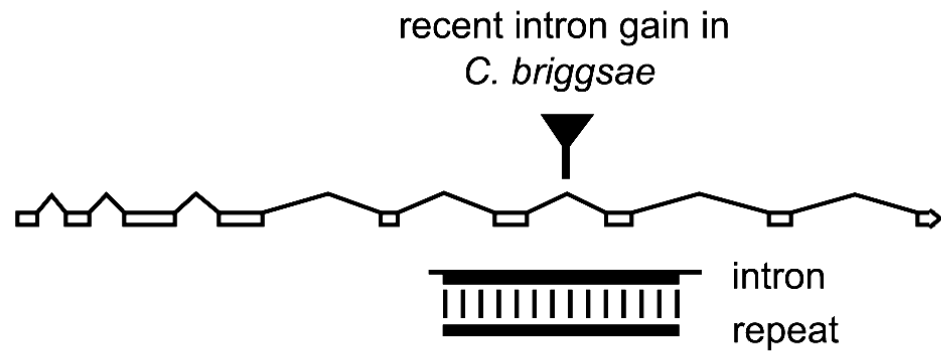
- 63% of *C. elegans* genes that gained introns are expressed in the germline, compared to 42% of control genes ($P = 0.001$)

→ it is unlikely that introns originate by partial exon duplication

Mechanism of intron gain

- ✘ Tandem duplication of an exon fragment containing AGGT**
- ? Reverse-splicing of a pre-existing intron**
- ? Insertion of a transposon**

Repeat elements in introns



```

R          533 TTTTCTATTTCTTTTTTTTTCAATTTTTCATGTCGTTAGAATTG
I          ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
R          GTGAGGTTTTCGGAATTTTTTCGATT---TTTTTTTTTGAAAATTTCAACT-TTTCATGCCG
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTACTCATGGAATTCCCCTCATTTTGATACCCATATTGCCTTATATTTCCAAATTTTATT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTAGAATTGTTACCCATGCGATTCCCGTCACTTTGATACC-CATATTGCTCTATATTAGC
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          CTGAATTTGAGAATTCTGAATTCAGAAGACGTCAAAAATCGAATCTTCTCGAATCGGGTC
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          AAATTTTCGAAAATTTCTGAATTCAGATGACGTCAAAAATCCAATATTTCTCGAACATAGGC
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          AATATTGGTATCAAAATGACCGGAAAACGAGGGAATTGTTTTAACGAAGTGAAATTTG
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          AATATGGGTATCAAAATGACTGGAAAATTACAGAGAATAGTTCGGATGAAGTGAAATTTG
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          CTTAAAATGTTTCTGGAGTCCCAAAATCCGAAAAATTGAAAAA-TTGAAAAATATTTTT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          CTGAAAATGATTTCTGAAGTTACAAAATTTGAAAAATTCAAAACCTGGA---ATTTTT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTTTGAATTTTTCGAAATTTTTTTTCTATTTTCAATT-TTCAACTTTTCATGTCGTTA
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTTTGAATTTTTCGAAATTTTTTTTCTATTTTCCCTTTCCCTTCAACTTTTCTGTCGTTA
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          GAATTGTTACCCATTGAATTTCCCCTCATTTTGATACCCATATTGCCTATATTTCCAAATT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          GAATTGTTCCCGTGAAGTTTCCGTCATTTTGATACCCATATTGCCATGTTTCCAAATT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTGTGAATTTCTGAATTCAGAAGACGTCAAAAATCGAATCTTCCGAATCGGGTCAATATT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          TTGTGAATTTCTGAATTCAGAAGACGTCAAAAATCGAATCTTCTCGAATCGAGTCAAGATT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          GGTATCAAAATGACCGGAAAACGAGGAAATCGTTTTAACGAAATGAAATTTGCTTAAA
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          GGTATCAAAATGACCGGAAAACACAGAAAATTGATAGAACGCAACGAAATTTGCTGAAA
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          ATGATTGTGGAGTCTGAAAATTTCAAAAAAC 1
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
R          ATGGTCTGGAGATTGAAAATTTCAAAAAATCTGAAAAGAATTTTTTTTAAAGTTAAAGTT
I          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
I          TTTTCAAG 597

```

Homology outside repeats



```

old  GTAAGTTTTACATTGAAAGTATTTTCATAAAAAGTTGAGGGAGAGA-TTCCAGACACTTTGA
new  ----G-TGCGTTT-----CTAATTGTTAAAATTT---TGAGAAAATTGGCTAAATTTTGG
      * * * * *
old  GCCTAAAATCATAAAAAACGGTTGAAAATGAAGAAAGTTATGAAGATTTGAAACATTTCA
new  TTCAGAAATCATCAAAATCGGTTGAAAATTTGCCTATTTATGAGGGTTTGAATTTTGGGA
      * * * * *
old  TAAAGTCGTAACTTGGTCCATTTTCACCCGATTCTCTTGATTTTGAGCTCGAAATGACTC
new  AAAGGT-GTAAGTTGGCTAATTTTCAATCGATTTTATATTTTTTGGCTAAAATG----
      * * * * *
old  GAAAAAATTTTTCTCTGGTTTTGTGAAATATCTTTTTTGAATTTTGTGAACAAAT-CCA
new  TATGAAATCCTTGAATATTCGGATGAAAATATTTCAAACCTTTCCGGAAAAAATTCCA
      * * * * *
old  ATCTTTTAAAGCCAAAAACCATGAAAA-TCGTTGAAGATTAAGAAAGTTATGAACGAAA
new  GACATTTTAAGTAAAAAACATTAAAAATGGTTAAAAATGAATCAAATTATGGA-GTTT
      * * * * *
old  CTCATAACTTGGTCCATTTTCGAACCGATTTCTATGTTTTTTGGCTCAAAAATGTCTGAAA
new  TCCCATAACTTTCTTCATTTTCATCCGATTCCTATGATTCCTAGCTTGAAGTGTCTGGAA
      * * * * *
old  ATGTTTCTTAAACACTTAAAATGCATAAAATATTTAAAATTCAAAAATCCAGACATTTT
new  AT-TTTC-----CAATTTTCAATGA----TCATAGCTATGAAATTTCCAGACGTTTT
      * * * * *
old  ACGCCTGAAATCATGAATATAGGATGAAAATCAGCCAAGTTATGGT-----CAAATCCC
new  GAACCAAAAATATTGAAATCGGTGGAAGATGAACCAAGTTATGGTACTTTGACCATAAC
      * * * * *

```

●
●
●

```

old  CA----AAAAATCCAGAAATTTTAAGCCAAAAACCACAAAATCGGTTGGCAAAGTAAT
new  TTGGAAAAACCTTCCAGAAATTTTAATCCAAAAATCAGGAAAATCGATGGAAAATGAACC
      * * * * *
old  ATTTTATTGAAATTTTTCAAAAAATCCAAAGTTCATTTTTTACAAAAATCCACCTTATCT
new  AAGTTATGGAATTTTTATCCGAATTCACGATTTTATAGATTAAAATGATCCAA-----
      * * * * *
old  CAGCTCATTTGACTCAAATTTTATGGATTTTAGGTTTTTACTCAGAAAAACAAGGCC
new  -----
old  AACAAAATGAGTATAATCATGACATTCTTTTTTTGAAAATTATTTTTCCAAAAATCTGAA
new  -----TTTCTAA-----
old  ATTTCTTGTTTTTTCAG
new  -----TTTTCAG
      * * * * *

```

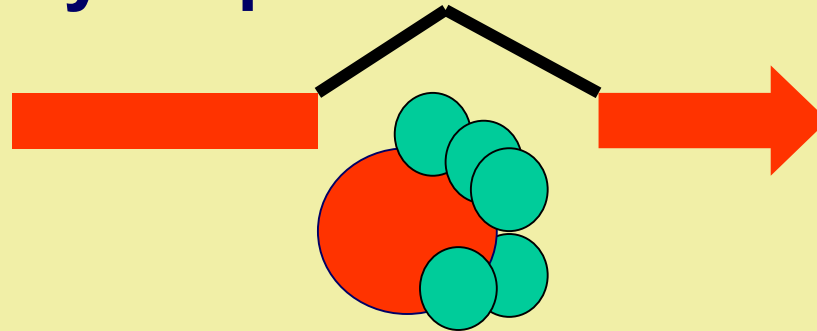
Functions of genes that gained introns

Several genes with novel introns code for proteins involved in splicing/surveillance:

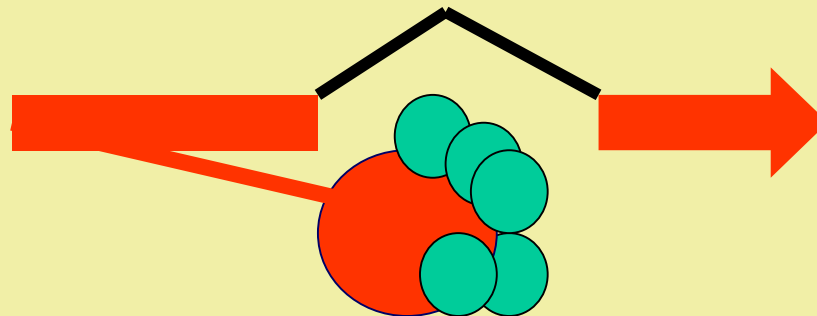
- smg-2***
- homolog of yeast Dbp2**
- homolog of yeast Cdc40**
- homologs of yeast Hsh155, Prp6, Prp19**
- homologs of yeast Imd2 and Ssa1**
- homolog of yeast Dis3**
- homolog of human *CPSF5***

Reverse-splicing in genes involved in splicing

? autoregulatory loops



? incorporation in the spliceosome of a protein still joined to its mRNA

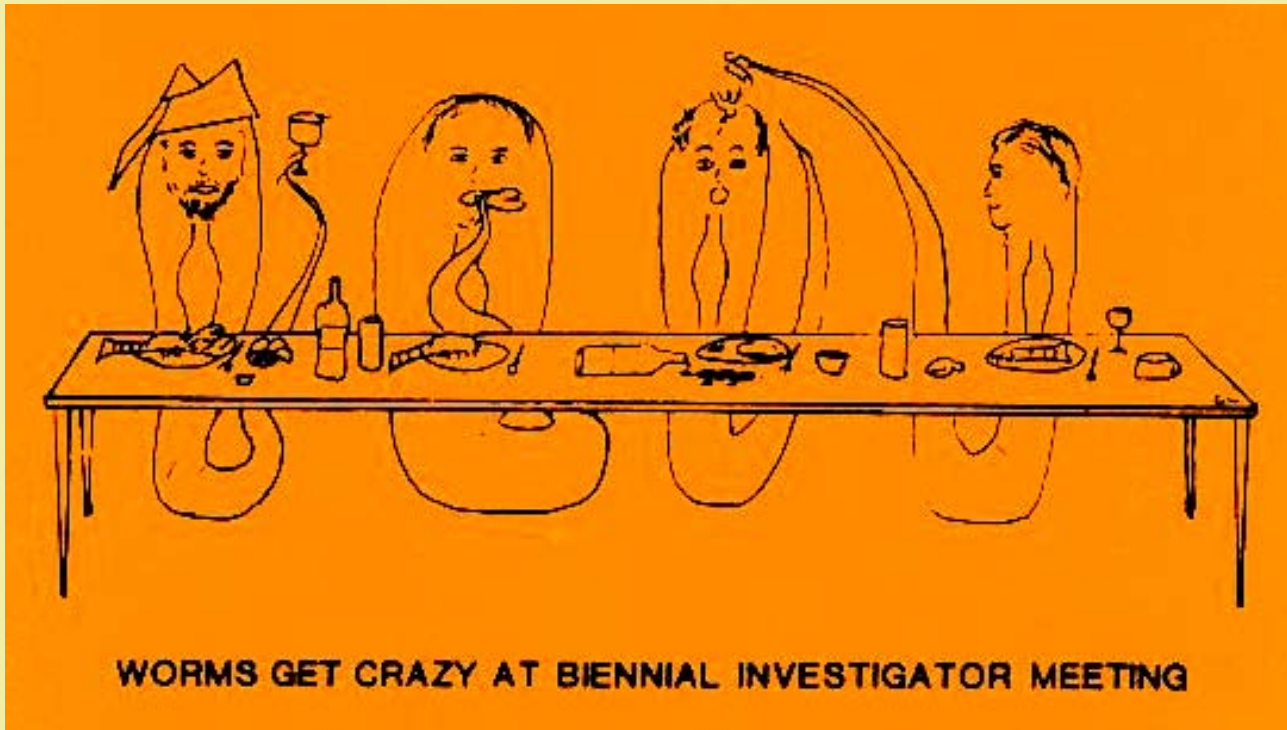


Conclusions



- some gains are due to reverse-splicing
- introns gained by reverse-splicing are copies of introns from the same gene
- reverse-splicing is especially frequent in genes involved in splicing

Acknowledgements



- Science Foundation Ireland for funding
- TIGR for use of *B. malayi* sequence
- Richard Durbin and Lincoln Stein for use of *C. briggsae* sequence