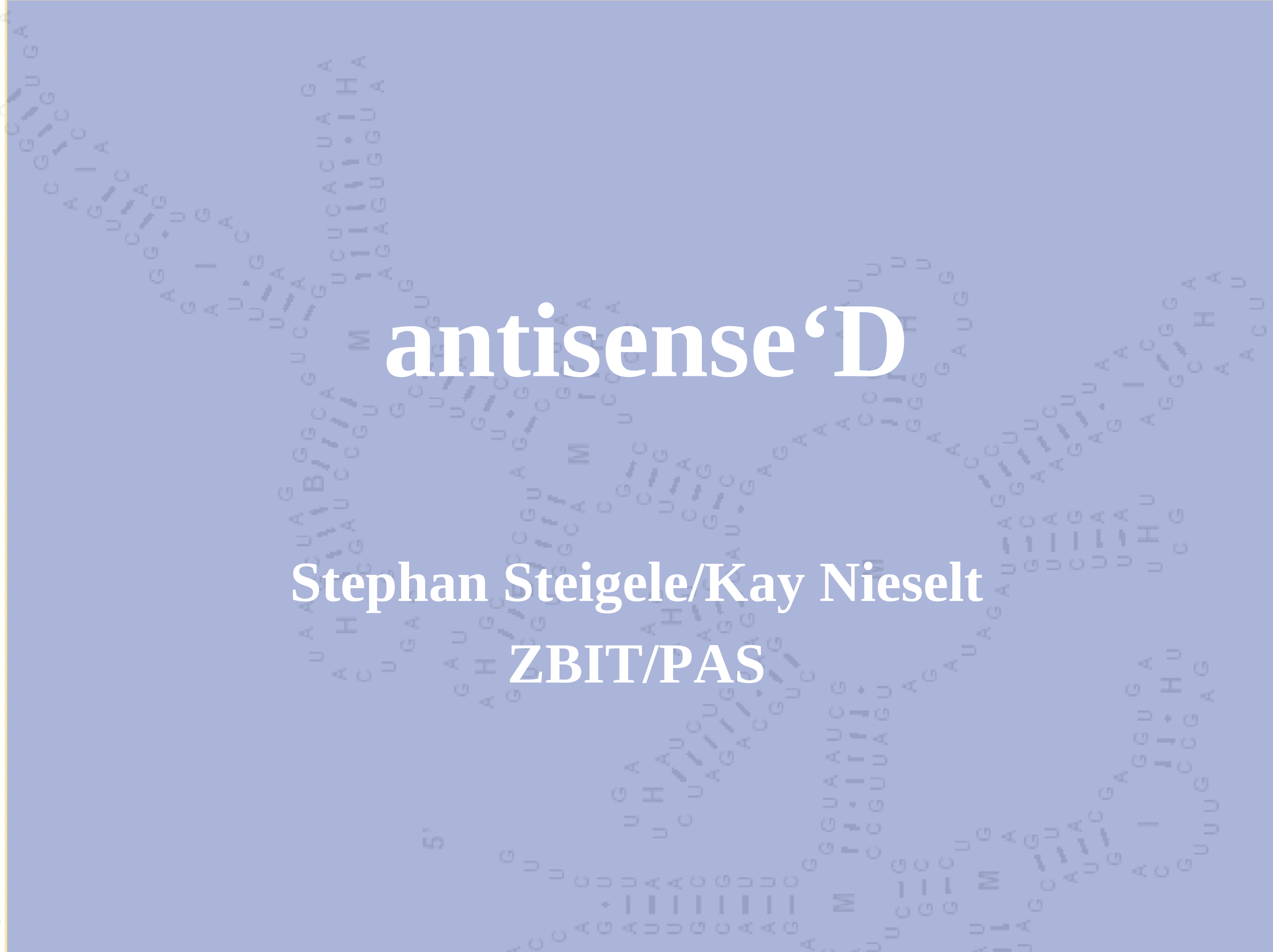


antisense‘D

Stephan Steigele/Kay Nieselt

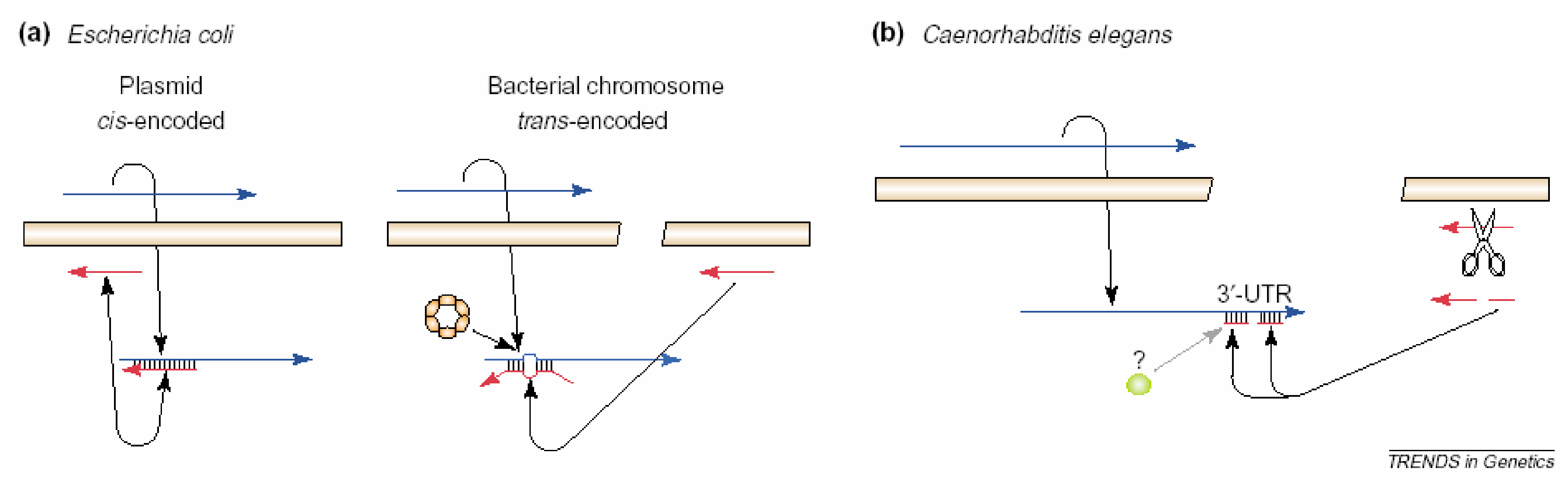
ZBIT/PAS

antisense principle

many VERY different biological activities relate to this principle ::

- processes like splicing, RNA editing (DRADA), RNA transport, translation, genomic imprinting and developmental regulation
- often mediated by small RNAs :: snRNAs, gRNAs, snoRNA, stRNA
- even tRNA / mRNA interaction is a complementary (antisense) process

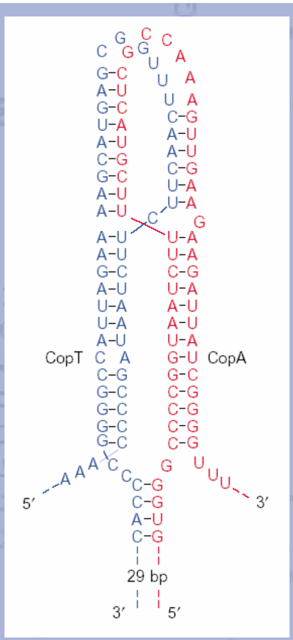
some established examples



a.) CopT / CopA cis-example

b.) OxyS / fhlA trans-example

c.) let-7 / lin-41 example for MIR



examples of known eukaryotic genes „nuclear events“

- thyroid hormone receptor, c-erbA
- genomic loci harbours two sense and one antisense transcript
- Rev-ErbAalpha **inhibits splicing** of inactive receptor alpha2. In addition it encodes a nuclear hormone receptor-related protein

- basic fibroblast growth factor, bFGF
- genomic loci harbours sense and antisense transcripts conserved in all vertebrata
- protein levels from sense transcript are shown to be inversely proportional to the level of antisense mRNA
- -> probably regulation through **mRNA polyadenylation**

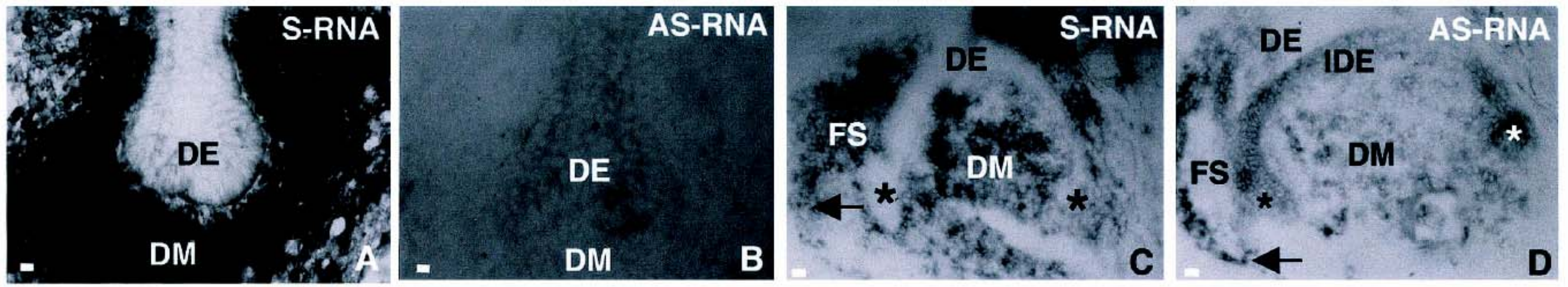
contd.

- p53
- AS-RNA is complementary to exon/intron 1
- inhibits the transport of sense RNA to cytoplasm

- Antisense Expression of Homeobox gene
- genomic loci of Hoxd-3 from mouse contains one antisense transcript (Doxh-3), which encodes a protein of 116aa
- „… suggests that Doxh-3 may be switched on when it is necessary to target Hoxd-3 sense transcripts for rapid degradation“

MSX1 + MSX1-AS

- MSX1 is a key regulator of tooth and craniofacial development
- the presence of an MSX1 antisense transcript is experimentally verified



antisense‘D ORFs

- modern genetic code evolved from a prototypic triplet code of general form RNY
- due to the complementary of sense / anti-sense strands there is always a potential to deduce an ORF from complementary strands

5‘-RNY RNY RNY RNY-3‘
3‘-YNR YNR YNR YNR-5‘

G+C content and length of antisense ORFs

- **the length of antisense ORFs relates to the G+C content of host genomes**

- **CTA and TTA (Leu) and TCA (Ser) are Stop – Codons (TAG, TAA, TGA) on the antisense strand**

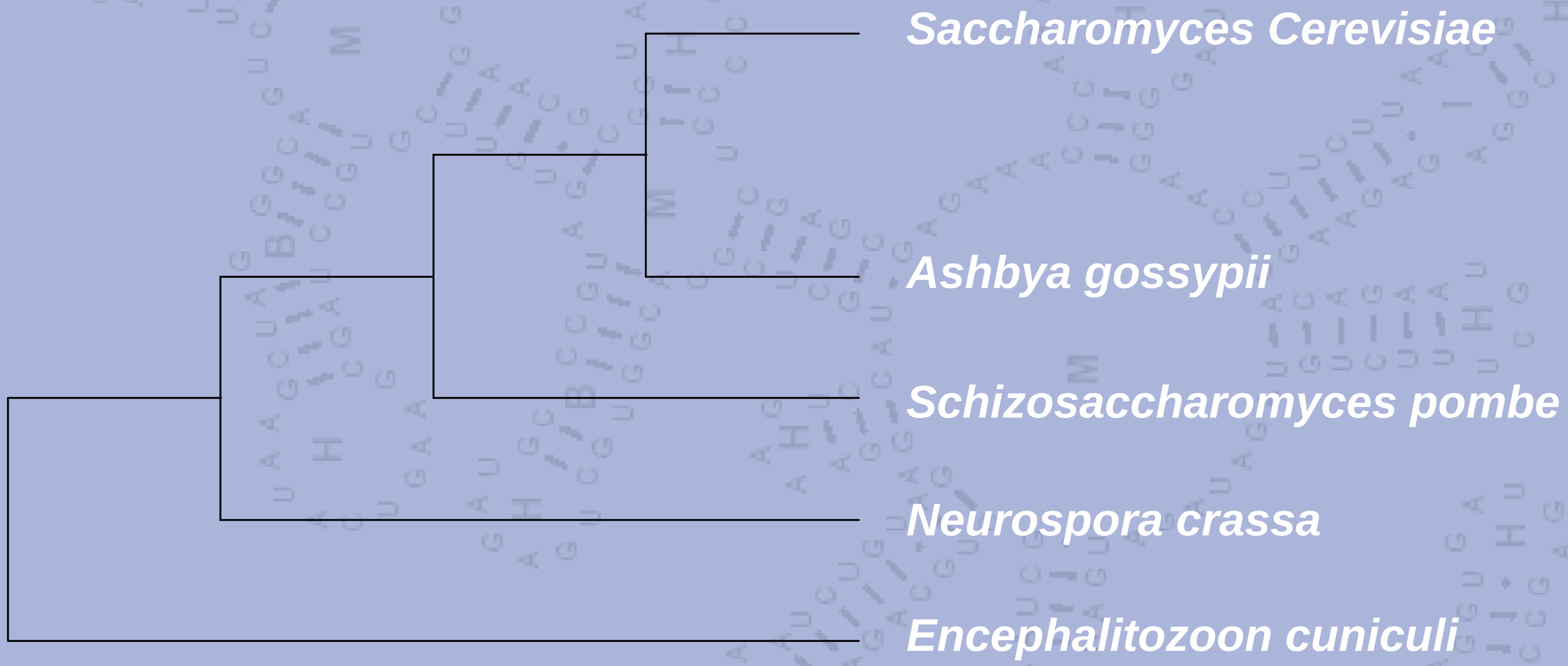
- **to form long antisense ORFs, alternative codon usage is necessary**

many papers covering Natural Antisense Transcripts (NATs)

- Lehner et al. 2002 (Trends in Genetics)
- Shendure et al. 2002 (Genome Biology)
- Yelin et al. 2003 (Nature Biotechnology)
- Chen et al. 2004 (Nucleic Acid Research)
- Kiyosawa et al. 2003 (Genome Research)

- only **one** author (Lehner et al.) suggests a method to detect both cis- and trans-NATs
- LACK of comparative analysis (15.2.05 appeared one in Genome Research cmp. human-fugu)

detection of fungi Antisense Transcripts



Saccharomyces Cerevisiae

Ashbya gossypii

Schizosaccharomyces pombe

Neurospora crassa

Encephalitozoon cuniculi

genomic features

- *S. cerevisiae* has lost argonaute, RNase III homolog to plant carpel (Dicer) and 5 other important RNAi related proteins

VS.

- in *S. pombe* are substantial functional needs for Dicer shown etc. : they are required for chromosome segregation and gene silencing, heterochromatin formation and centromere silencing

- *N.crassa* has even two antisense systems :: quelling & RNAi

contd.

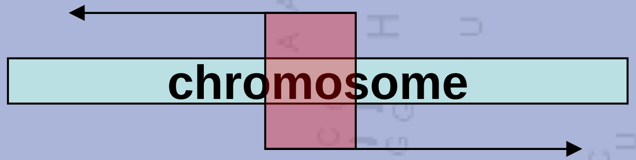
- 4% introns in *S. cerevisae* cmp. to 43% in *S. pombe*

- *A. gossypii* and *N. crassa* lack ANY repetitive element cmp. to yeast with hundreds of repetitive elements
- *N. crassa* likes to RIP (repeat induced point mutation) -> permits any duplicated sequence >80% ident.

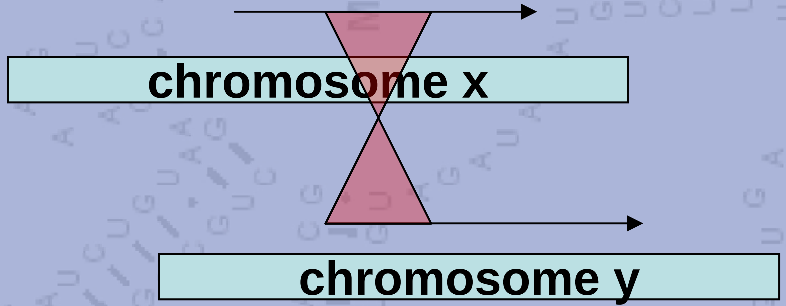
- GENERALLY LONG EVOLUTIONARY DISTANCE (eg. *S. pombe* and *S. cerevisiae* up to one billion years)

classification of NATs

Type-A NAT



Type-B NAT



brief protocol

- all predicted ORFs from fungi
- perform a pairwise BLAST using WU-BLASTN with an modified substitution matrix to handle G-U base pairing

- <u>**TYPE-A**</u>
- directly calculated from genomic coordinates

- <u>**TYPE-B**</u>
- analyze all high-scoring-pairs (HSPs) produced by one subject sequence
- Blastanalyzer builds plausible models
- longest consistent ordering is reported as the antisense overlap region of putative NAT partners

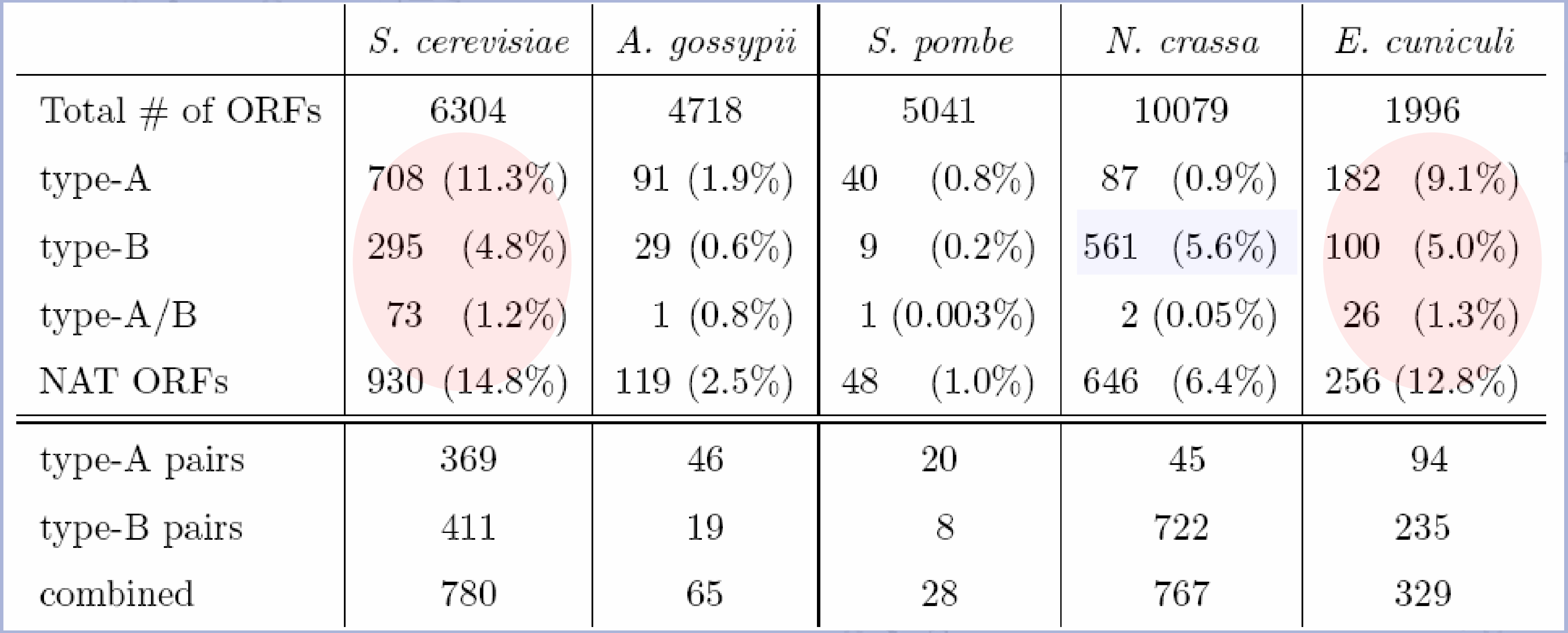
contd.

- define overlap regions
  (5‘-overlaps, 3‘-overlaps, inside-overlaps)
- define longest complementary Stretch (LCS)

- determine orthologs with pairwise FastA on translations of CDS
- define groups of reciprocally best-hits as orthologous groups

- cross-annotate gene_associations of GeneOntology
- annotate NATs with GO_slim terms

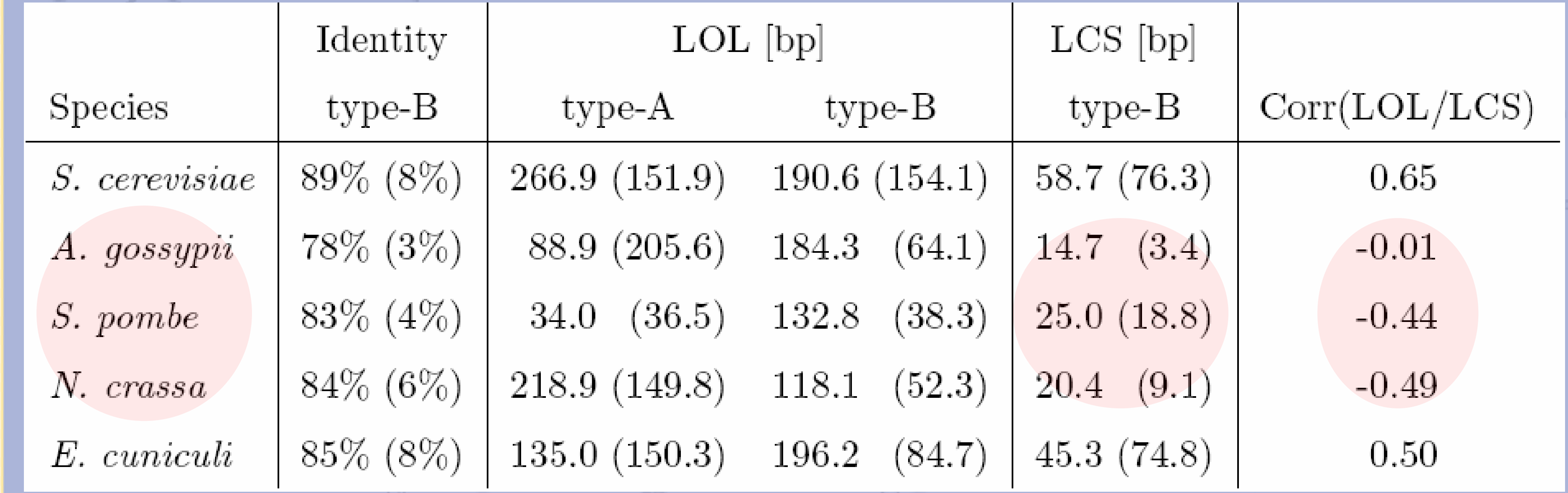
results :: overview



	S. cerevisiae	*A. gossypii*	*S. pombe*	*N. crassa*	*E. cuniculi*
Total # of ORFs	6304	4718	5041	10079	1996
type-A	708 (11.3%)	91 (1.9%)	40 (0.8%)	87 (0.9%)	182 (9.1%)
type-B	295 (4.8%)	29 (0.6%)	9 (0.2%)	561 (5.6%)	100 (5.0%)
type-A/B	73 (1.2%)	1 (0.8%)	1 (0.003%)	2 (0.05%)	26 (1.3%)
NAT ORFs	930 (14.8%)	119 (2.5%)	48 (1.0%)	646 (6.4%)	256 (12.8%)
type-A pairs	369	46	20	45	94
type-B pairs	411	19	8	722	235
combined	780	65	28	767	329

in human approx. 3-5 % of all protein coding genes in type-A relations

results :: classification

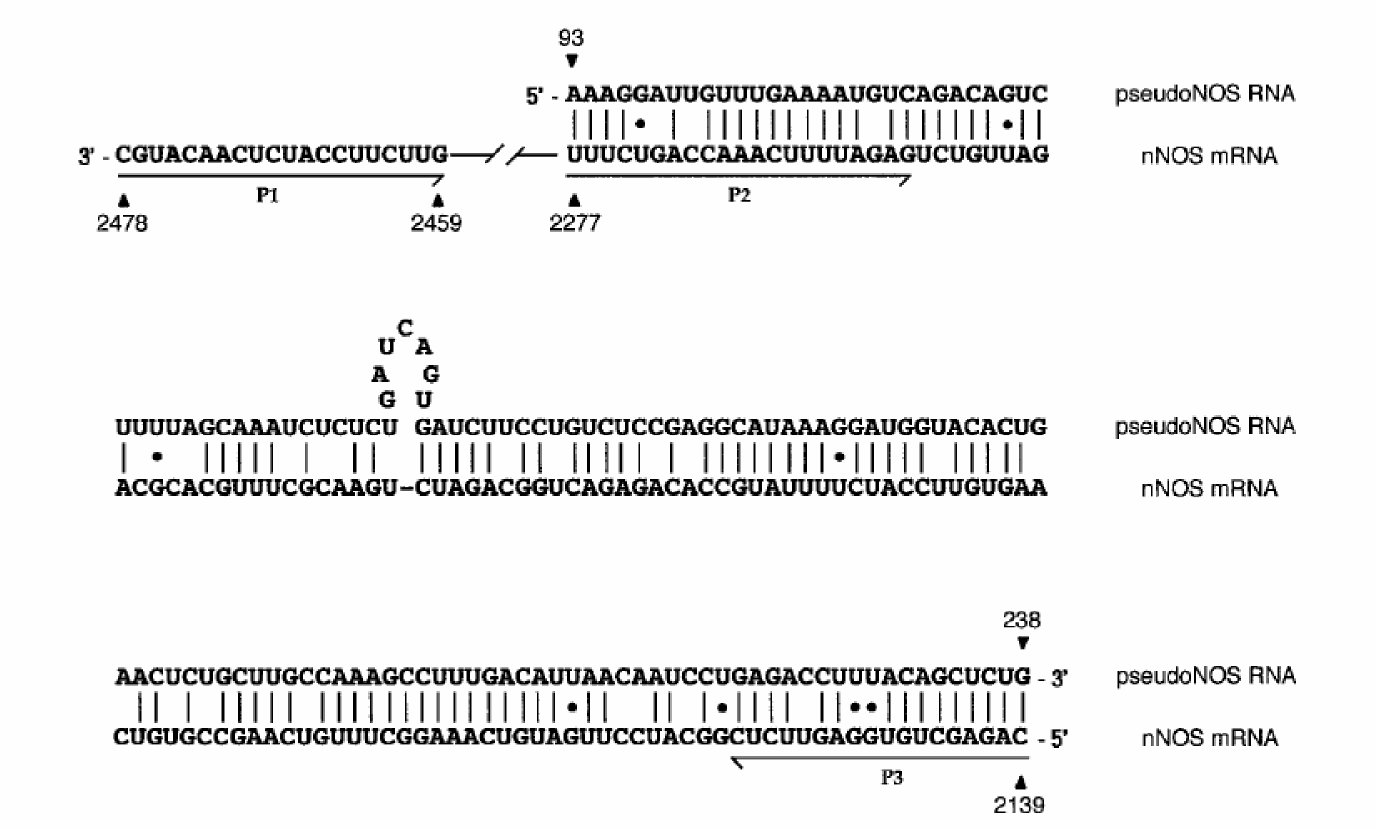


	Identity	LOL [bp]		LCS [bp]	
Species	type-B	type-A	type-B	type-B	Corr(LOL/LCS)
S. cerevisiae	89% (8%)	266.9 (151.9)	190.6 (154.1)	58.7 (76.3)	0.65
A. gossypii	78% (3%)	88.9 (205.6)	184.3 (64.1)	14.7 (3.4)	-0.01
S. pombe	83% (4%)	34.0 (36.5)	132.8 (38.3)	25.0 (18.8)	-0.44
N. crassa	84% (6%)	218.9 (149.8)	118.1 (52.3)	20.4 (9.1)	-0.49
E. cuniculi	85% (8%)	135.0 (150.3)	196.2 (84.7)	45.3 (74.8)	0.50

open question :: are these differences reflected by the fact, that

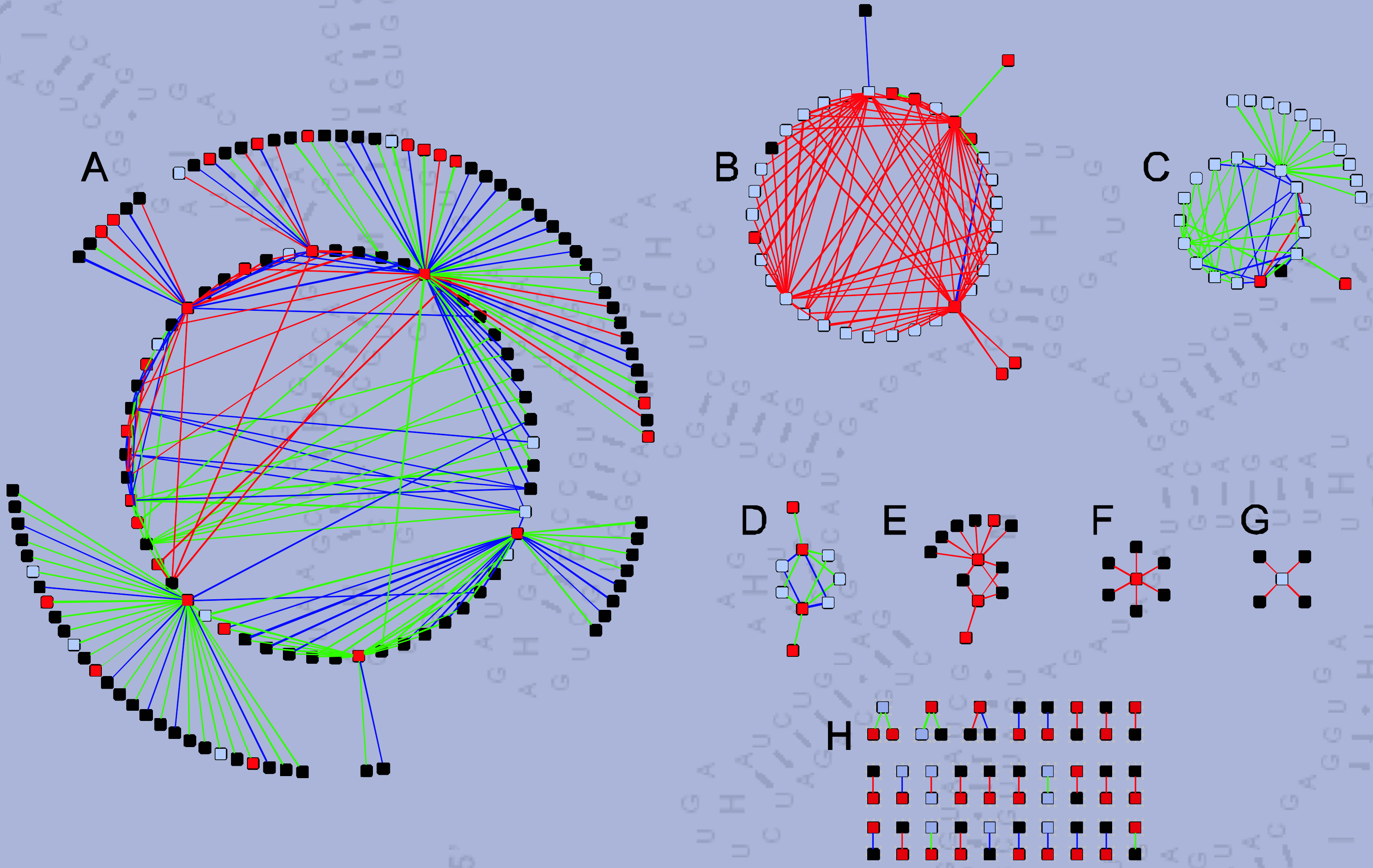
- ***S. pombe*** **and** ***N. crassa*** **use massively RNAi**
- ***N. crassa*** **and** ***A. gosypi*** **lack any repetitive element,**
- ***N. crassa*** **has RIP ????**

one example of a functional Type-B NAT :: snail nNOS

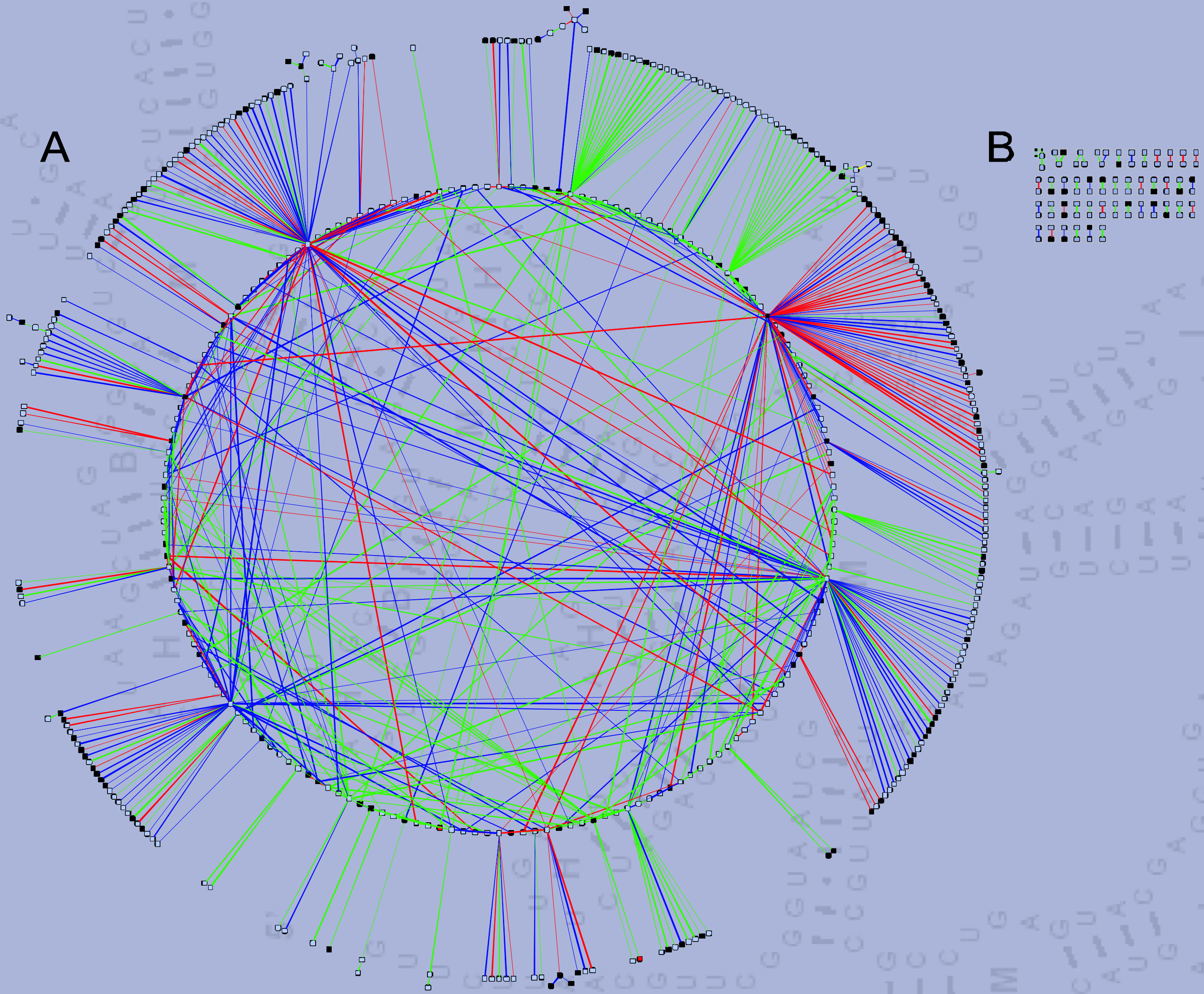


here :: translational control

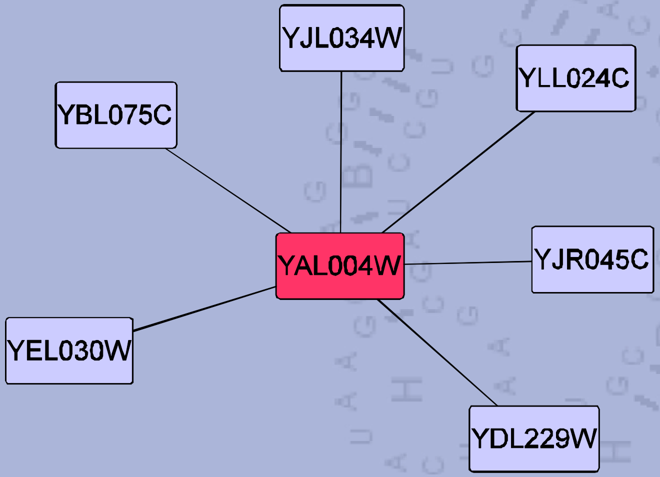
antisense‘D networks : *S. cerevisiae*



Antisense‘d Networks : *N. crassa*

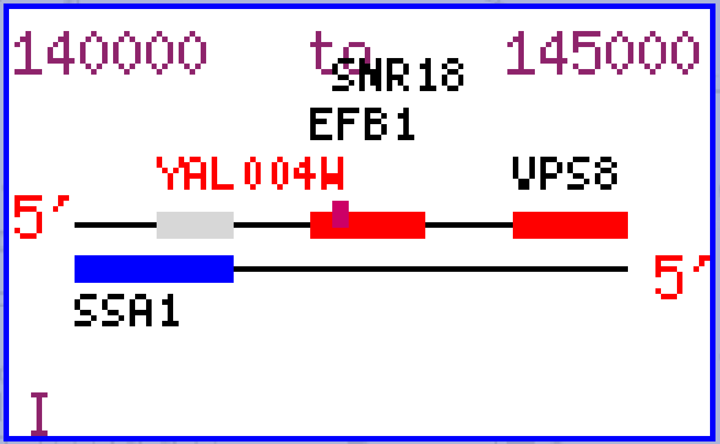


founders ??

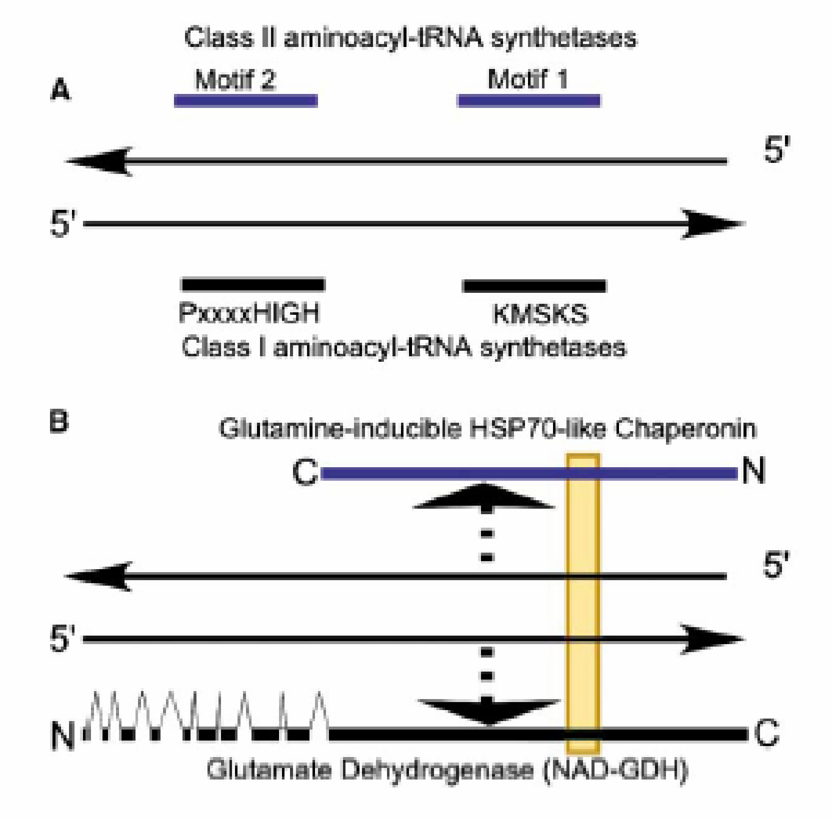


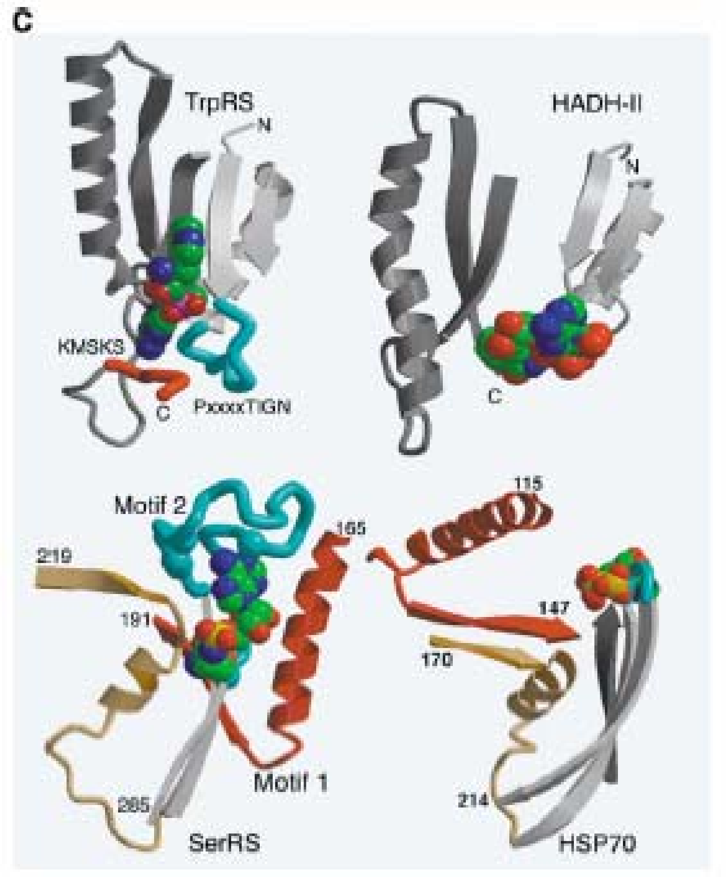
- conserved example of sense / antisense transcription

- first detected in *Achlya klebsiana* and *Drosophila auraria*
- YAL004W is described as having strong similarity to *A. klebsiana* NAD-GDH
- is FARs for family of HSP70 proteins

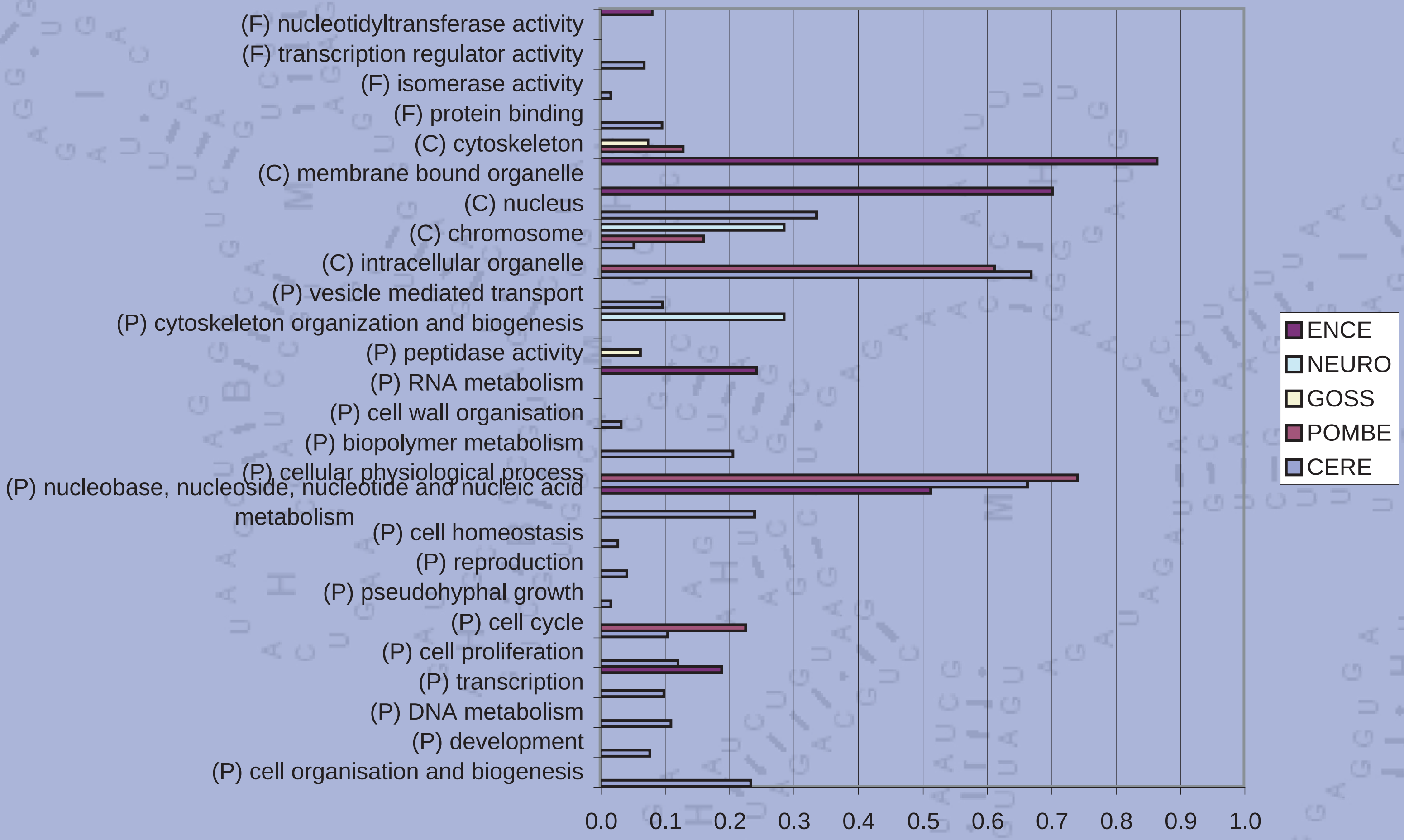


hints for evolution of aaRS-classes ??

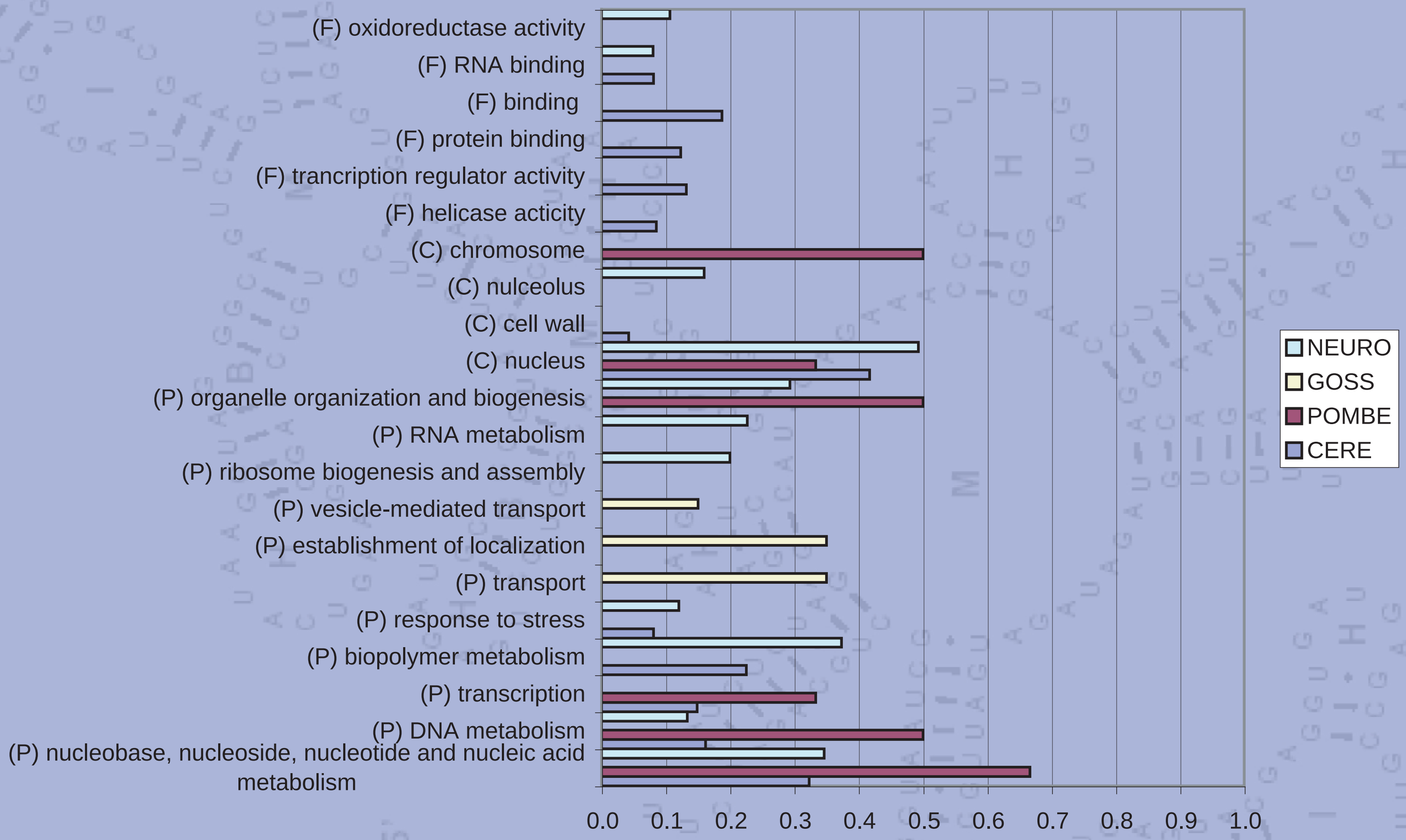




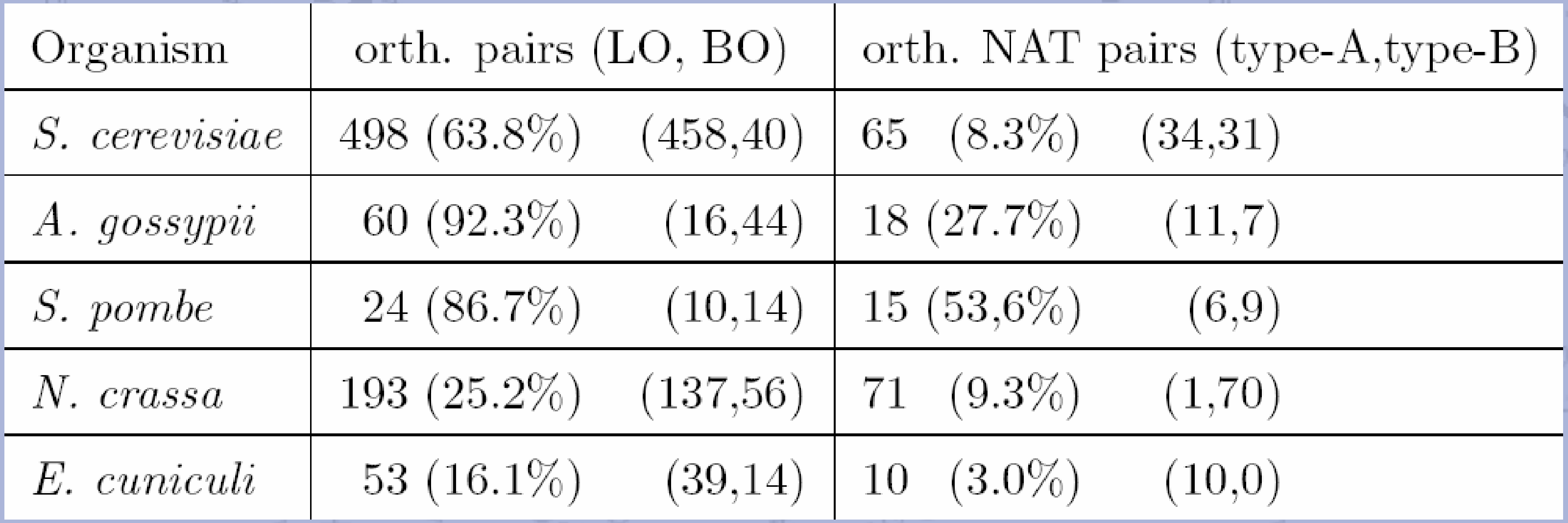
GO-annotation type-A NATs



GO-annotation type-B NATs



ORFaned NATs ??

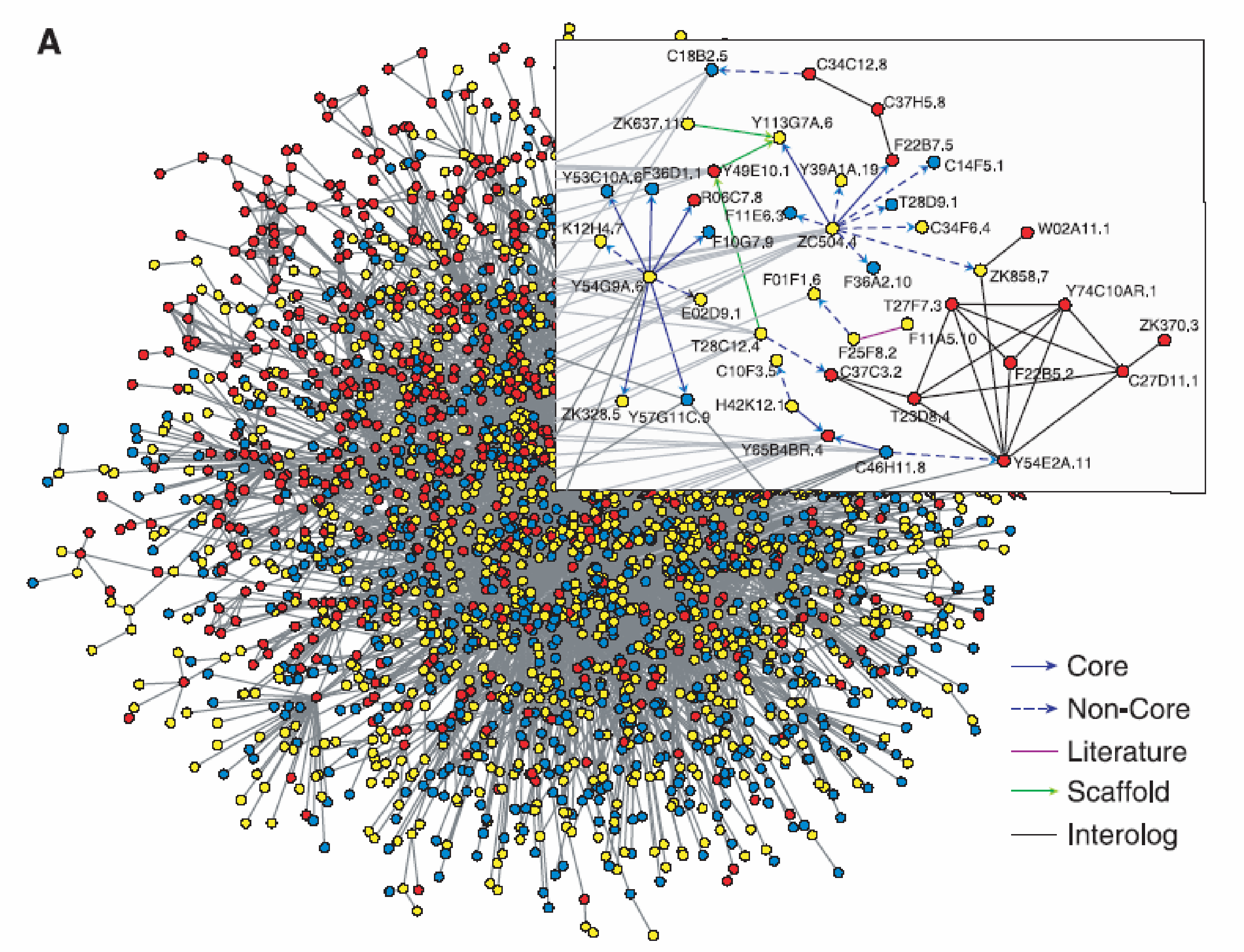


Organism	orth. pairs (LO, BO)	orth. NAT pairs (type-A,type-B)
S. cerevisiae	498 (63.8%) (458,40)	65 (8.3%) (34,31)
A. gossypii	60 (92.3%) (16,44)	18 (27.7%) (11,7)
S. pombe	24 (86.7%) (10,14)	15 (53,6%) (6,9)
N. crassa	193 (25.2%) (137,56)	71 (9.3%) (1,70)
E. cuniculi	53 (16.1%) (39,14)	10 (3.0%) (10,0)

for most NAT pairs, only ONE of the two NATs has an an ortholog

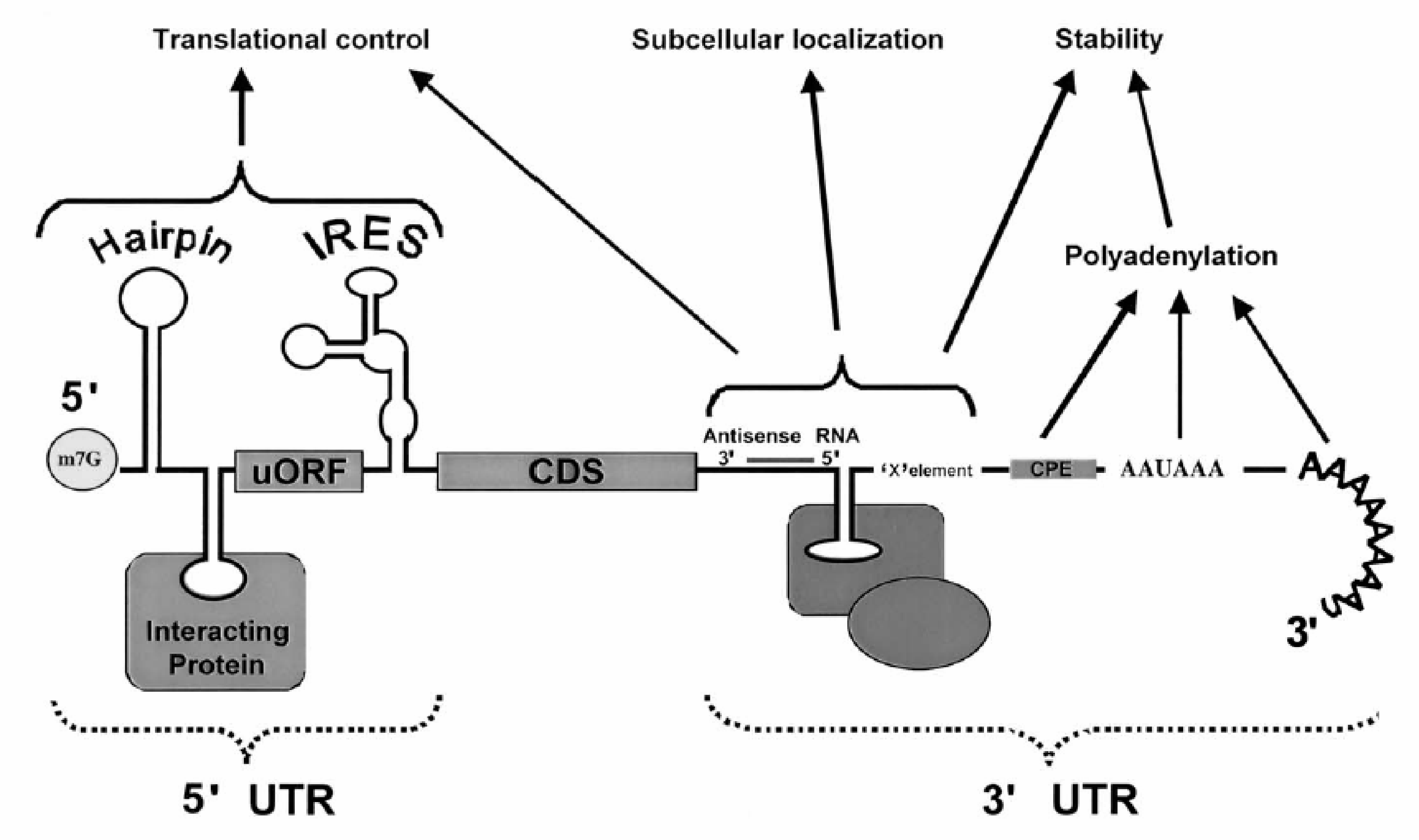
most antisense relations are unique in each species

ORFans everywhere



protein interaction network in C. elegans
> ORFan proteins are involved in any type of interaction

many possible targets for antisense functions



general existence of an orphaned antisense set ???

- use human antisense dataset of Chen et al. 2004

  approx. 2750 type-A NAT pairs

- search for antisense pairs with annotated proteins for both NATs
- search with this set for orthologs in animal species ::

  chimp, mouse, rat, chicken

  fugu, zebrafish

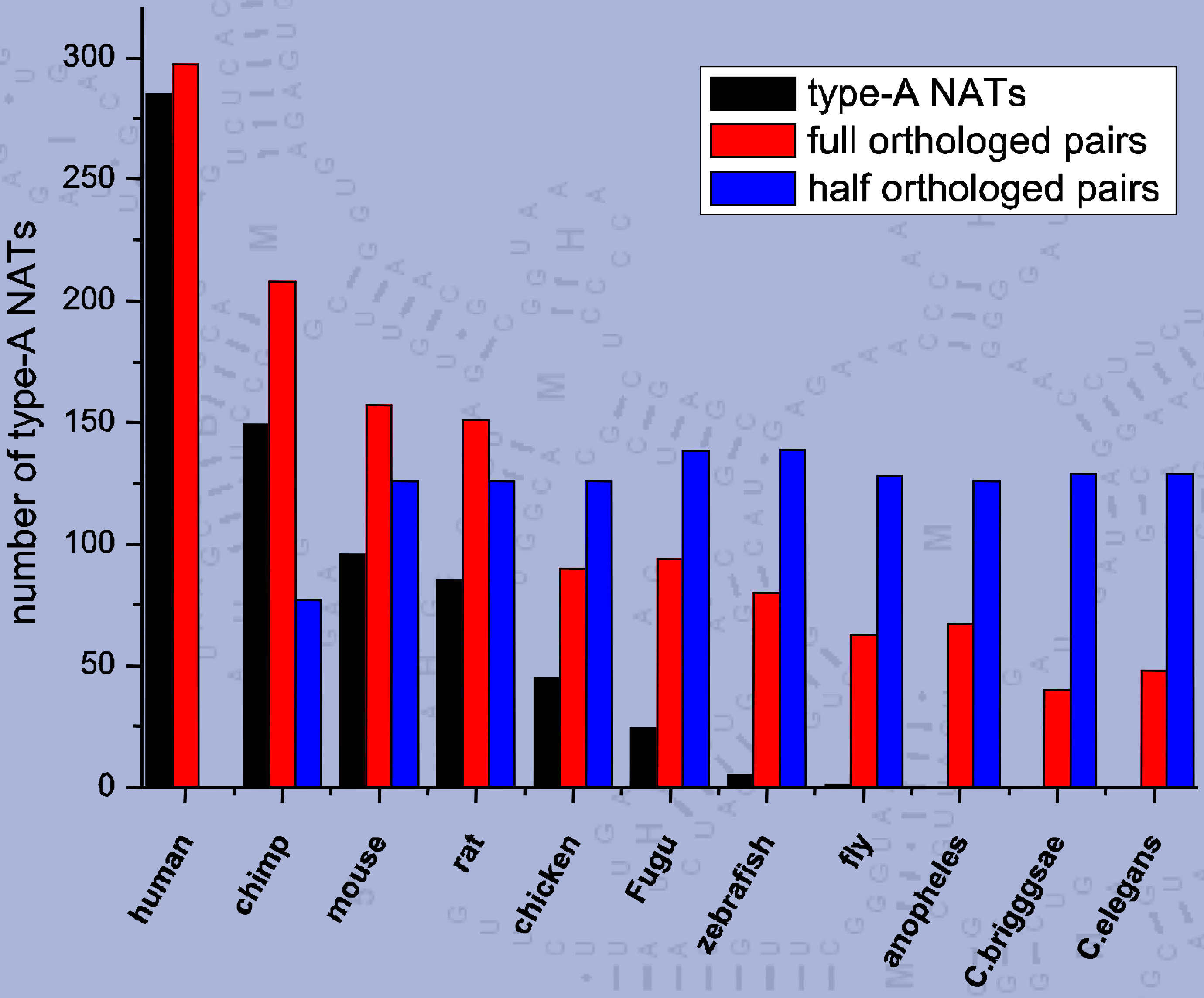
  fly, anopheles

  C. elegans, C. briggsae
- report how many orthologs exist and how many are in NAT relations

rough overview of preliminary results

Species	number of full pairs	number of type-A NATs	half orthologed pairs
human	297	285	
chimp	208	149	77
mouse	157	96	126
rat	151	85	126
chicken	90	45	126
Fugu	94	24	138
zebrafish	80	5	139
fly	63	1	128
anopheles	67	0	126
C.brigggsae	40		129
C.elegans	48	0	129

rough overview of preliminary results



Conclusion

- fungi species have specific sets of antisense‘d genes

  >their number possibly correlates with the sensitivity of an organism to sense dsRNA

  >annotation shows common functions of NATs between different species

- most type-B NATs have more than one antisense partner and appear in large networks
- many of the NATs are ORFans and are unique to its host

- many NATs are specific to human (at least ORFans)
- human NATs could only be traced back in vertrebrata
- species from the same „taxonlevel“ seem to preserve similar numbers of NATs

THX