

Homework questions: (10 points each). **These questions will be graded; do them alone. The accompanying sheet has (practice) questions that you should do with a group.**

1) (10 points) Where is the following sequence from (what organism?) Does it encode a protein? If so, what is the likely function of that protein?

>MysterySequence08

```
TGTCTCACTTTTCTCTGTTGTCCTGCCTTTCAATTAAATAATCTTAATAAAAAGACACTGTATGCAGATCA
GTGAAATTTGACCTCCTGTCTTACACCATATACAAAAATCAGCTTAGAATTGATTAATGACTTATATGTC
```

This sequence comes from the rabbit, *Oryctolagus cuniculus*. To learn that, you needed to use blast against the wgs database. It does not encode a protein. That is apparent from the conceptual translation, which is open in none of the six frames and from the lack of similarity to annotated proteins in any organism (blastx or tblastx). This sequence is in the intron of eIF4A1, so you might have said something about eIF4A1.

ATP-dependent RNA helicase which is a subunit of the eIF4F complex involved in cap recognition and is required for mRNA binding to ribosome. In the current model of translation initiation, eIF4A unwinds RNA secondary structures in the 5' untranslated region of mRNAs which is necessary to allow efficient binding of the small ribosomal subunit, and subsequent scanning for the initiator codon.

2) (10 points) The protein product of the yeast gene SNP1 (NP_012203) is homologous to two human proteins, [NP_003080](#) and [NP_851030](#).

What Arabidopsis protein(s) are orthologous to NP_003080? **NP_190636**

What Arabidopsis protein(s) are orthologous to NP_851030? **NP_850395**

(AAB64330/NP_18165 is an alternatively spliced form from the same gene)

Consider the three simpler photosynthetic organisms, each with a completed genome sequence.

[Physcomitrella patens](http://www.cosmoss.org/) (<http://www.cosmoss.org/>) This genome can also be searched at NCBI (and there are refseq entries).

[Volvox carteri](http://genome.jgi-psf.org/Volcal/Volcal.home.html) (<http://genome.jgi-psf.org/Volcal/Volcal.home.html>) and

[Chlamydomonas reinhardtii](http://genome.jgi-psf.org/Chlre3/Chlre3.home.html) (<http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>).

Which genomes have one ortholog (as is true for *S. cerevisiae*)?

Which have two or more orthologs (and do they have one or more for [NP_003080](#) and one or more for [NP_851030](#))

?

Physcomitrella (a moss) has orthologs to **both** U1 70K (NP_190636) and U11 35K (NP_850395).

Phypa122893 (XP_001760398) and Phypa151189 (XP_001782496) are nearly identical and both are orthologous to U1 70K (NP_190636).

Phypa78228 (XP_001764761) is orthologous to U11 35K (NP_850395).

Volvox and *Chlamydomonas* have orthologs to U1 70K **only** (not to U11 35K).

They are [jgi|Volcal1|106420|estExt_fgenes4_pg.C_430081](#) and [jgi|Chlre3|159152|acegs_kg.scaffold_37000005 / XP_00167897](#).

Volvox is not on the NCBI site; you **had** to use the link above.

It might have been useful to use the algal protein as a query and "blast back" to see that these are orthologous to U1 70K but not U11 35K.

What is the significance of this difference? Is it likely that *Volvox* and *Chlamydomonas* have lost the U12 spliceosome (there is other evidence to support this). It was helpful to make use of known phylogenetic relationships (all of the photosynthetic species share a common ancestor).