

Additional File 15:

```
Conf: 979998898779999999999789979999658565444035896799954677466899
Pred: CEEEEEECCCCHHHHHHHHHHHHCCCEEEEEECCHHHHHHHCCCCCEEEEEECCHHHHHH
  AA: MSILVIGGTGTLGRQIVRQAIDEGYQVKCMVRDFRRGAFLKEWGAELIYGDLSTIPL
      10         20         30         40         50         60

Conf: 96378576541102578753103456888877888864355643776245566799981
Pred: HHCCCCEEEEECCECCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCEEEEECCCCCCCCCCC
  AA: ALKGVTTVIDSSTIRSTSSYTAETIDWRGKLALLEASKIVGLKKFISFGVLNASANSSIP
      70         80         90        100        110        120

Conf: 245566999999872544231354101366326555444327853426899743834499
Pred: HHHHHHHHHHHHHHHCCCCCECCCCCCCCCHHHHHHHHHHHCCCCCEEECCCCCCCCCEEHH
  AA: LMDLKLKIEEKITTSGLNITIFQCSGFFQGLISQYALPILENETIWVQNAAPVAYLDTQ
      130        140        150        160        170        180

Conf: 9999999997095657958997189855899999999997899877988989999999
Pred: HHHHHHHHHHHHCCCCCCCCEEEEECCECCCCCHHHHHHHHHHHHHCCCCCEEECHHHHHHHHH
  AA: DAAKAVVNALNKSSYDNKIVSLIGEFWASNEIIELCERLCGKRANISYIPFLAFSLRR
      190        200        210        220        230        240

Conf: 8742454330899999888754787004760210467786159999999999999986
Pred: HHHHCCCCCHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHH
  AA: FFRLFEFTWNIADRLQFGEVVDNSKTIVKPNNEVDWPNGRLSLESYLQEYFSKILKKLRE
      250        260        270        280        290        300

Conf: 176520111166439
Pred: CCCCCHHHHCCCCC
  AA: TNYEKTQKSNDISFL
      310
```

Supplementary Figure 7: Secondary structure prediction of ycf39 sequence using Psipred (Jones 1999). Top row or 'Conf' shows confidence in prediction (9 being highest and 0 being lowest). Middle row or 'Pred' is the secondary structure predicted, 'H' for helix, 'E' for strand and 'C' for coil/loop. Third row, 'AA' is the amino acid identity at each position in ycf39 sequence.