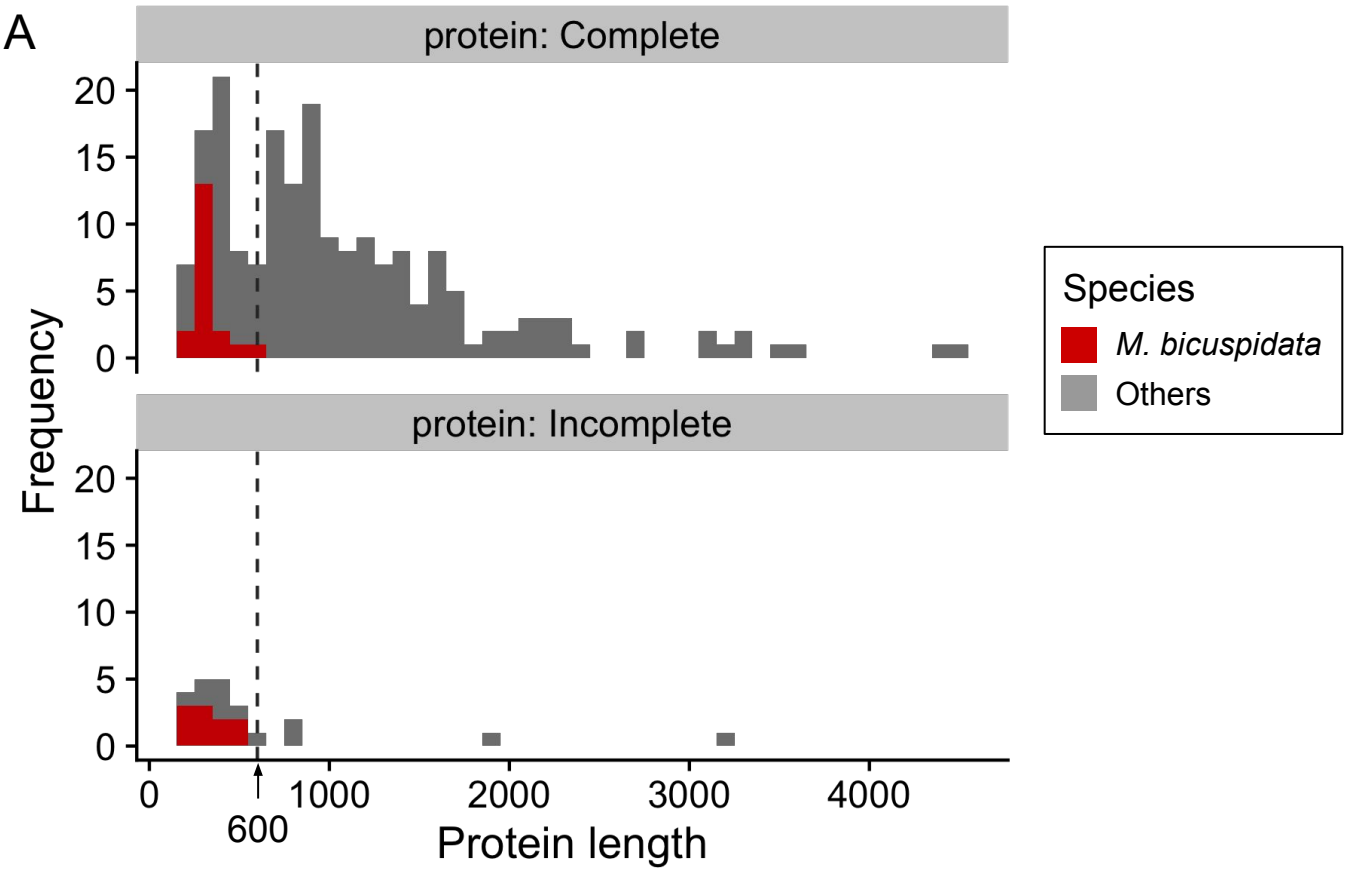
A



B

length	sp-gpi-	sp-GPI+	SP+gpi-	SP+GPI+	total
>600	3	14	9	112	138
251-600	12	5	27	4	48
0-250	3	0	3	1	7

Figure S1. Hil family protein length distribution and grouping by signal peptide (SP) and GPI-anchor signal presence. (A) Histogram showing the distribution of protein lengths for Hil family proteins from 32 yeast species. Top: protein sequence records were labeled as complete or "NA"; bottom: proteins labeled as incomplete (no-right, no-left, no-ends). Most of the short sequences (<600 aa, dashed vertical line) came from the species *M. bicuspidata* (red) (B) Summary of the number of Hil family proteins predicted to have a signal peptide (SP+) and GPI-anchor signal (GPI+), grouped by protein length. Proteins labeled as incomplete were excluded from this table.

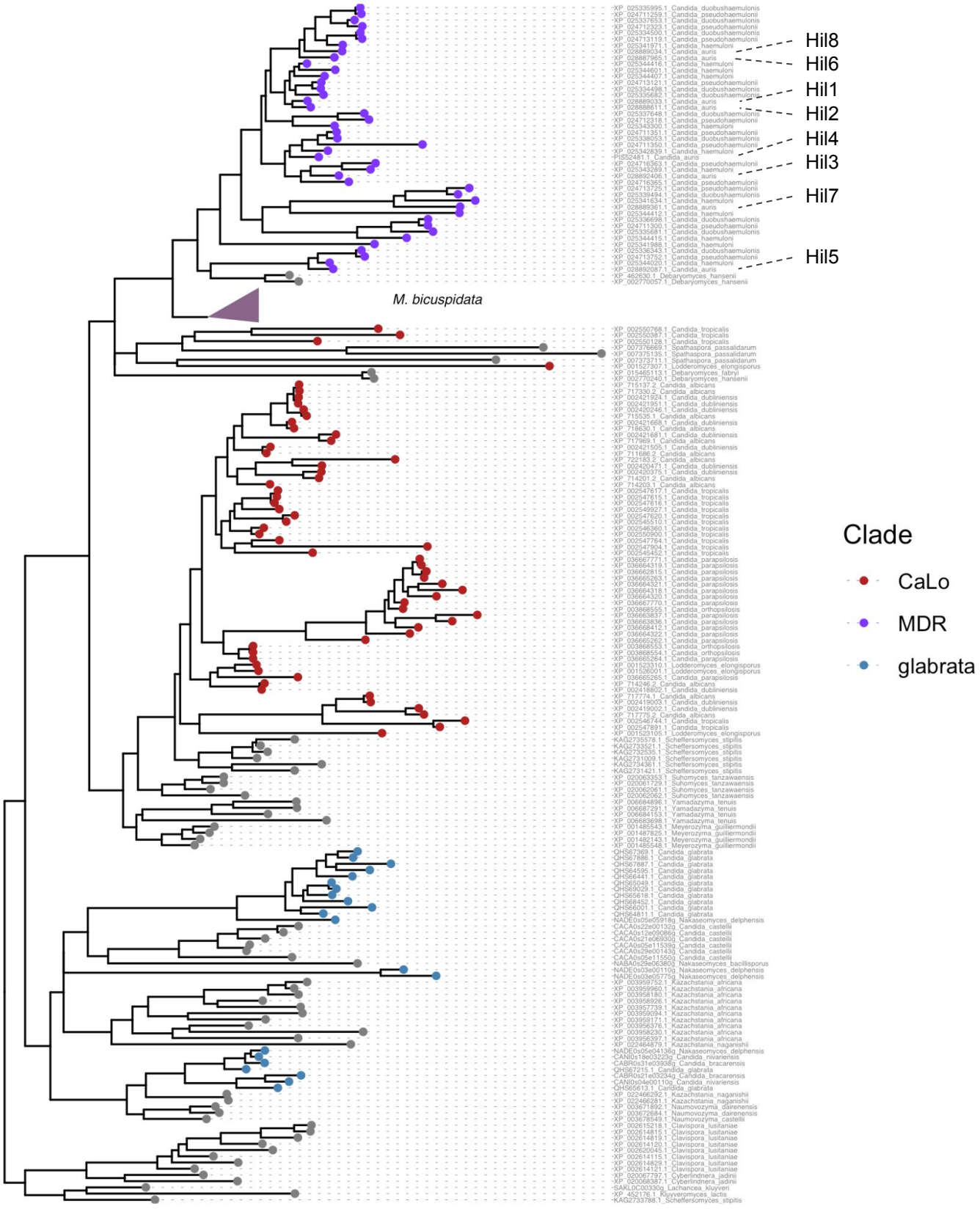


Figure S2. Maximum likelihood tree for the Hil family genes. This tree is identical to the one shown in Fig 1C, except that it is shown in a rectangular format with the sequence names in the form of refseqID_species_name. We identified and named the *C. auris* homologs as Hil1-8 to be consistent with the latter figures. Note that all 29 *M. bicuspidata* homologs form a single clade, which is collapsed for the ease of viewing. Their sequence IDs can be found in Table S1.

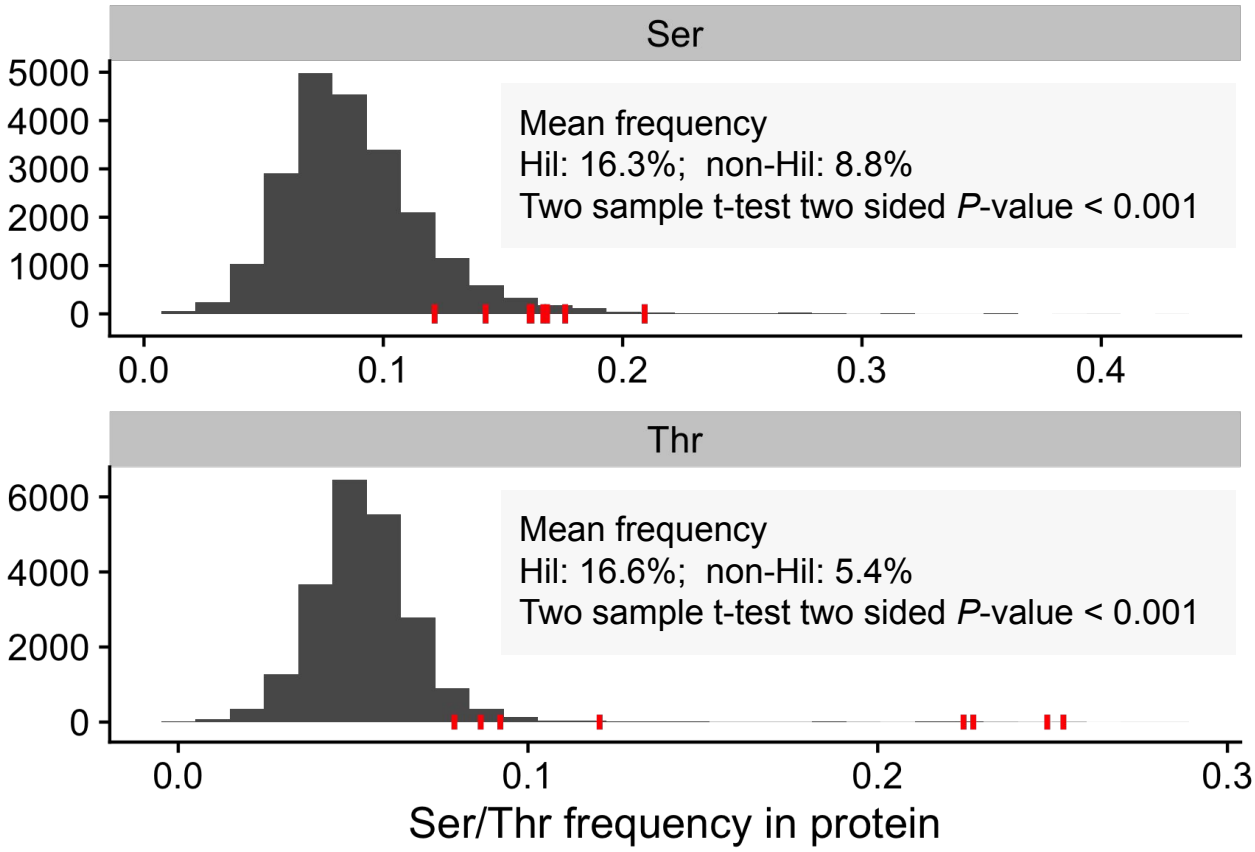


Figure S3. Comparison of the Ser/Thr frequencies in *C. auris* Hil family members with all protein-coding genes in *C. auris*. B8441 strain genome is used for this analysis. The frequency of Ser or Thr residues as a percent of the entire protein length is plotted as a histogram for all protein-coding genes. Red ticks indicate the eight *HIL* genes. A Student's t-test was used to assess the significance of the difference in Ser/Thr frequencies between the Hil family proteins vs the rest of the proteome.

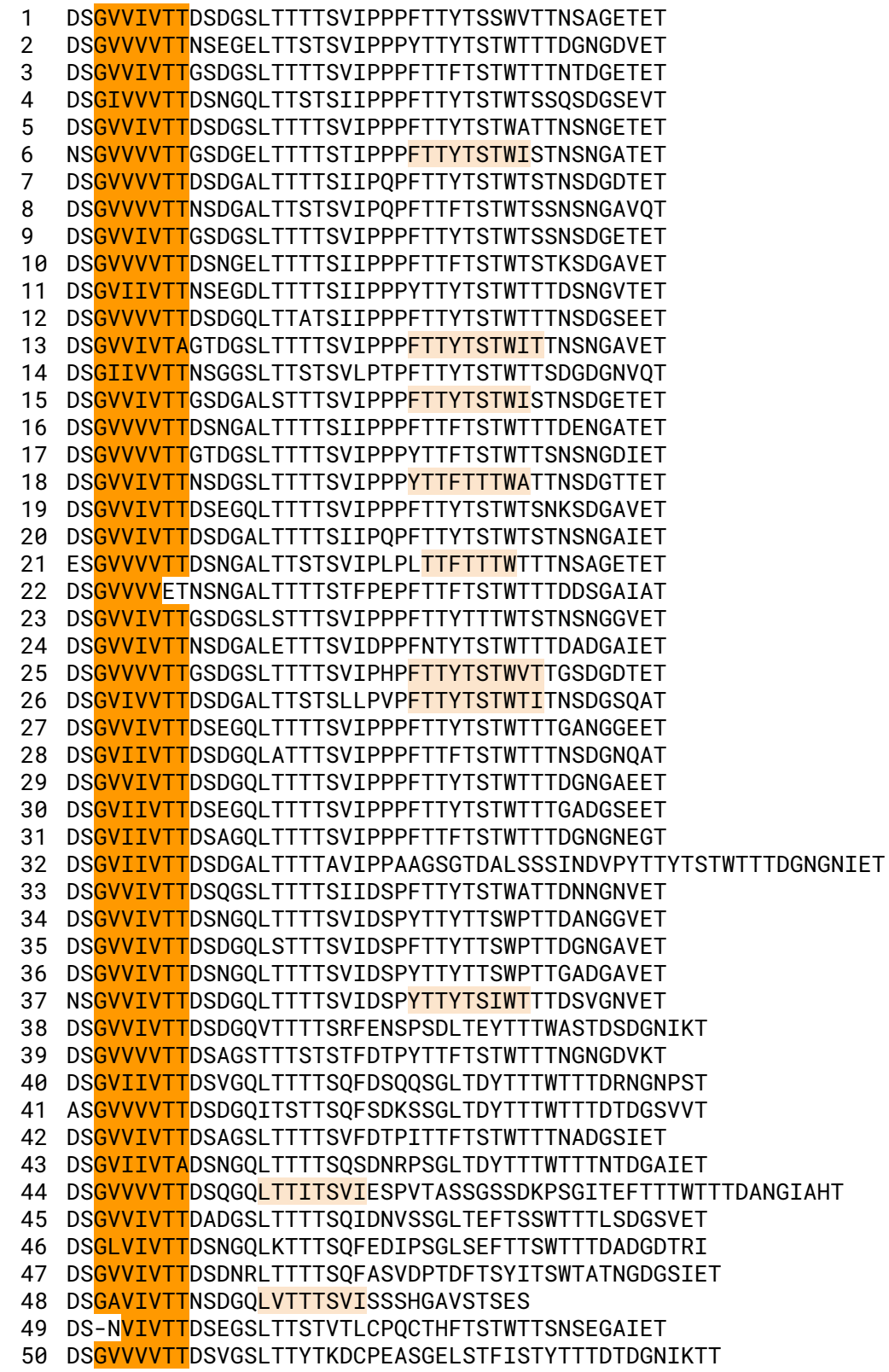


Figure S4. Tandem repeats in the *C. auris* Hil1 central domain. The majority of the 50 tandem repeat copies have a conserved 44 aa period. Dark and light orange highlights show sequences predicted by TANGO to have strong (>90%) or moderate (30-90%) β-aggregation potentials.

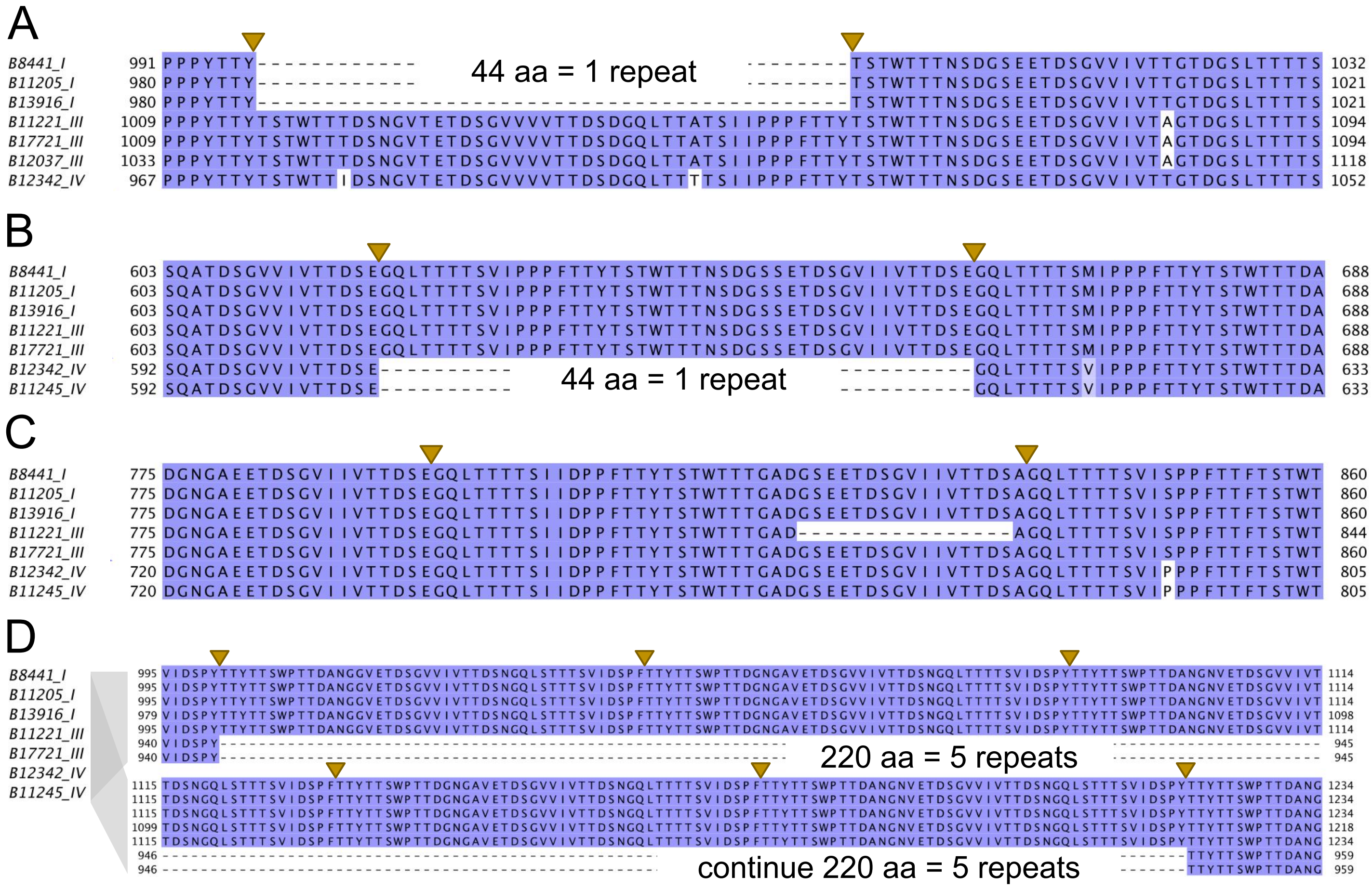


Figure S5. Examples of tandem repeat copy number variation in Hil1-Hil4 among the *C. auris* strains. (A) A 44 aa indel in Hil1 removes exactly one repeat in all three Clade I strain orthologs. (B) A similar indel polymorphism of exactly one repeat length in Hil2 affecting the Clade IV strains. (C) An indel polymorphism in Hil2 that affects one Clade III strain and spans 16 aa, not a full repeat, but includes a predicted strong β-aggregation prone sequence "GVIIVTT". (D) An indel polymorphism in Hil2 that spans 220 aa or five full repeats affecting the Clade IV strains. Similar patterns were observed in Hil3 and Hil4.

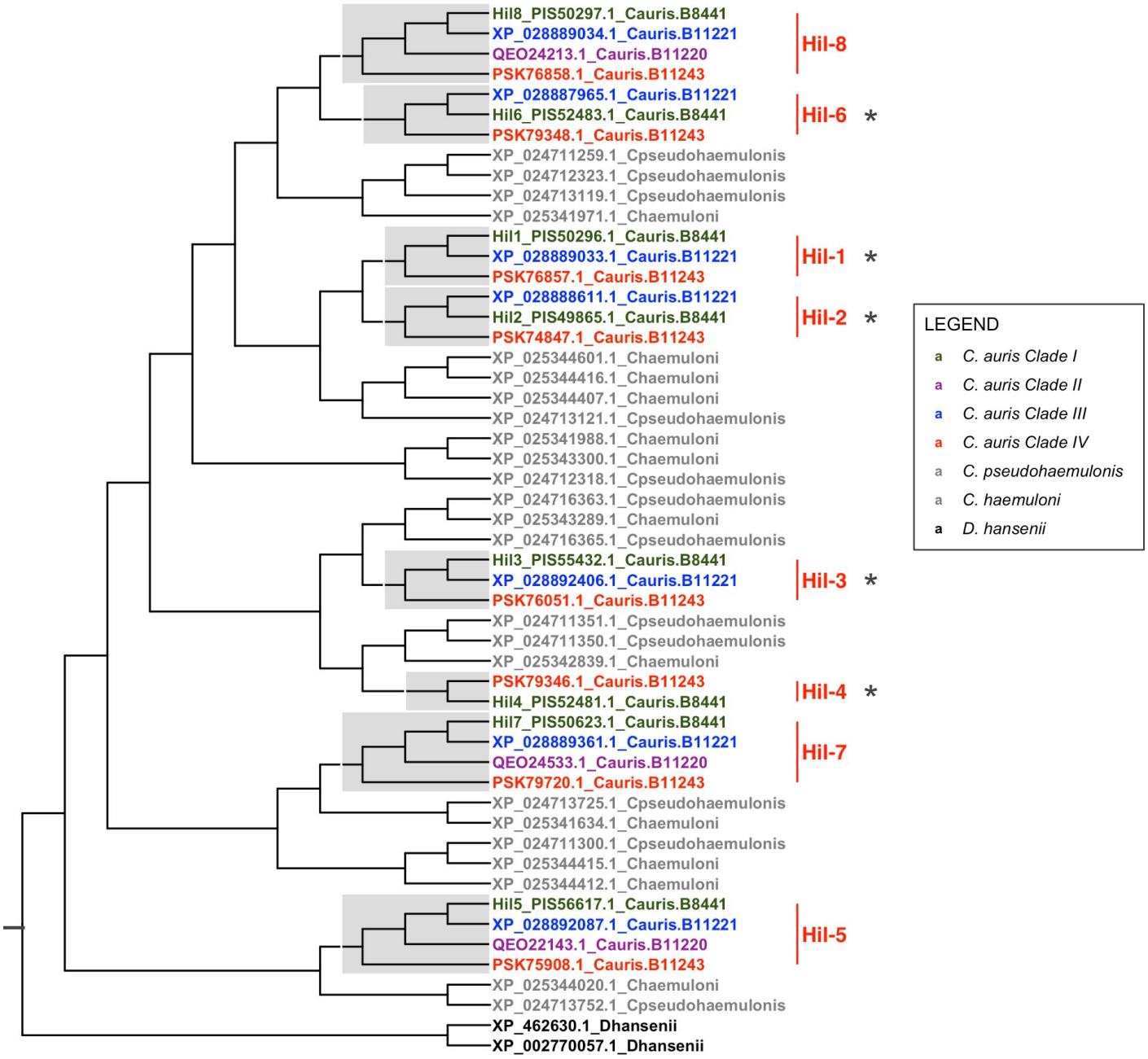


Figure S6. Reconciled Hil family gene tree based on the Hyphal_reg_CWP domain alignment in the four clades of *C. auris* strains and two closely related species. The tree is rooted by the two homologs from the outgroup *D. hansenii*. The gene tree was corrected with the species/strain tree (see Materials and Methods). *HIL* genes lost in *C. auris* Clade II strains are labeled with an asterisk next to the Hil1-8 group labels.

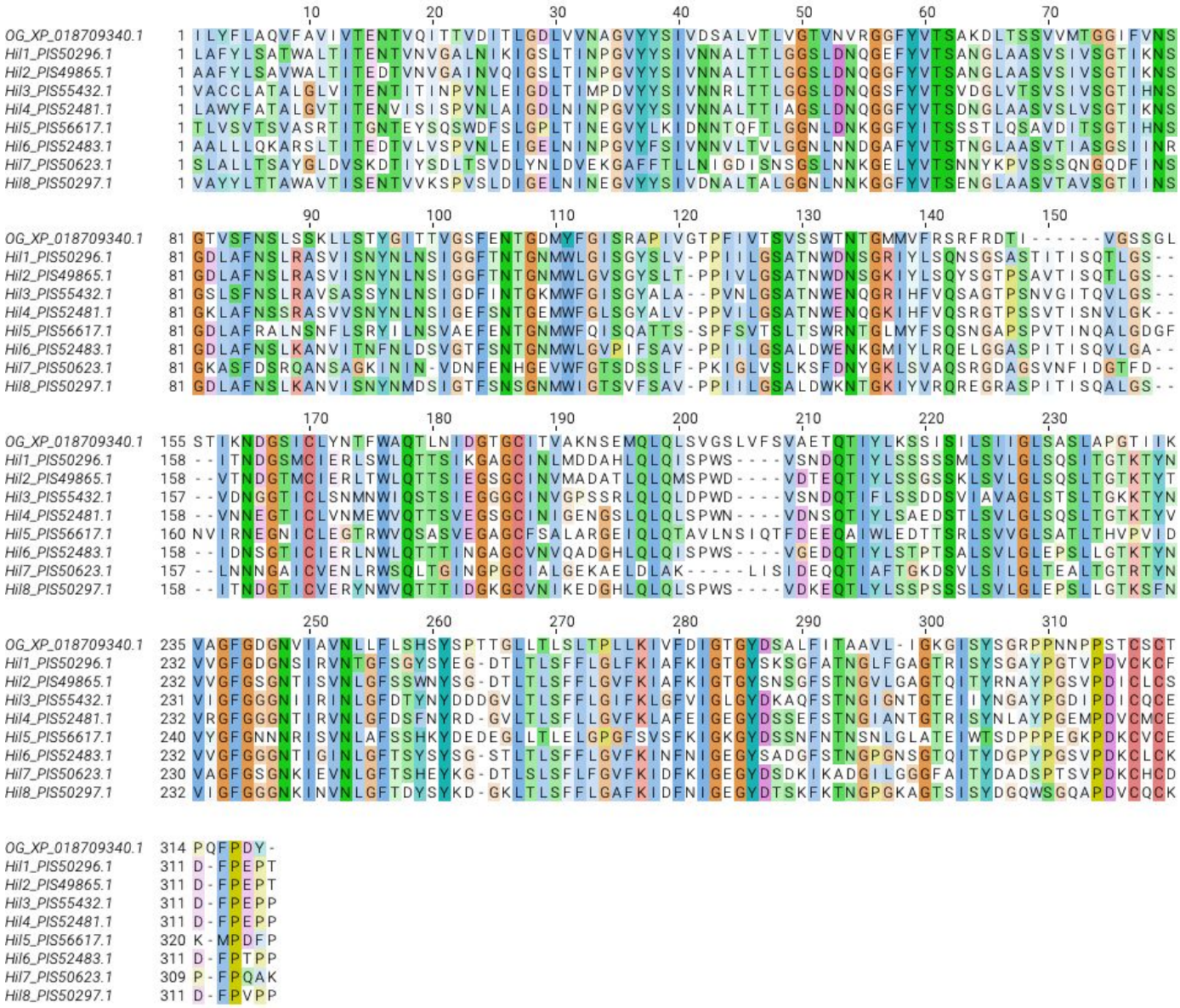


Figure S7. Multiple sequence alignment of *C. auris* Hil1-8 Hyphal_reg_CWP domain. The domain regions were identified using HMMScan against the Pfam-A database. The sequences were aligned with clustalo v1.2.4 and the result visualized in Jalview v2.11.1.4 with the ClustalW color scheme. OG = outgroup from *M. bicuspidata*

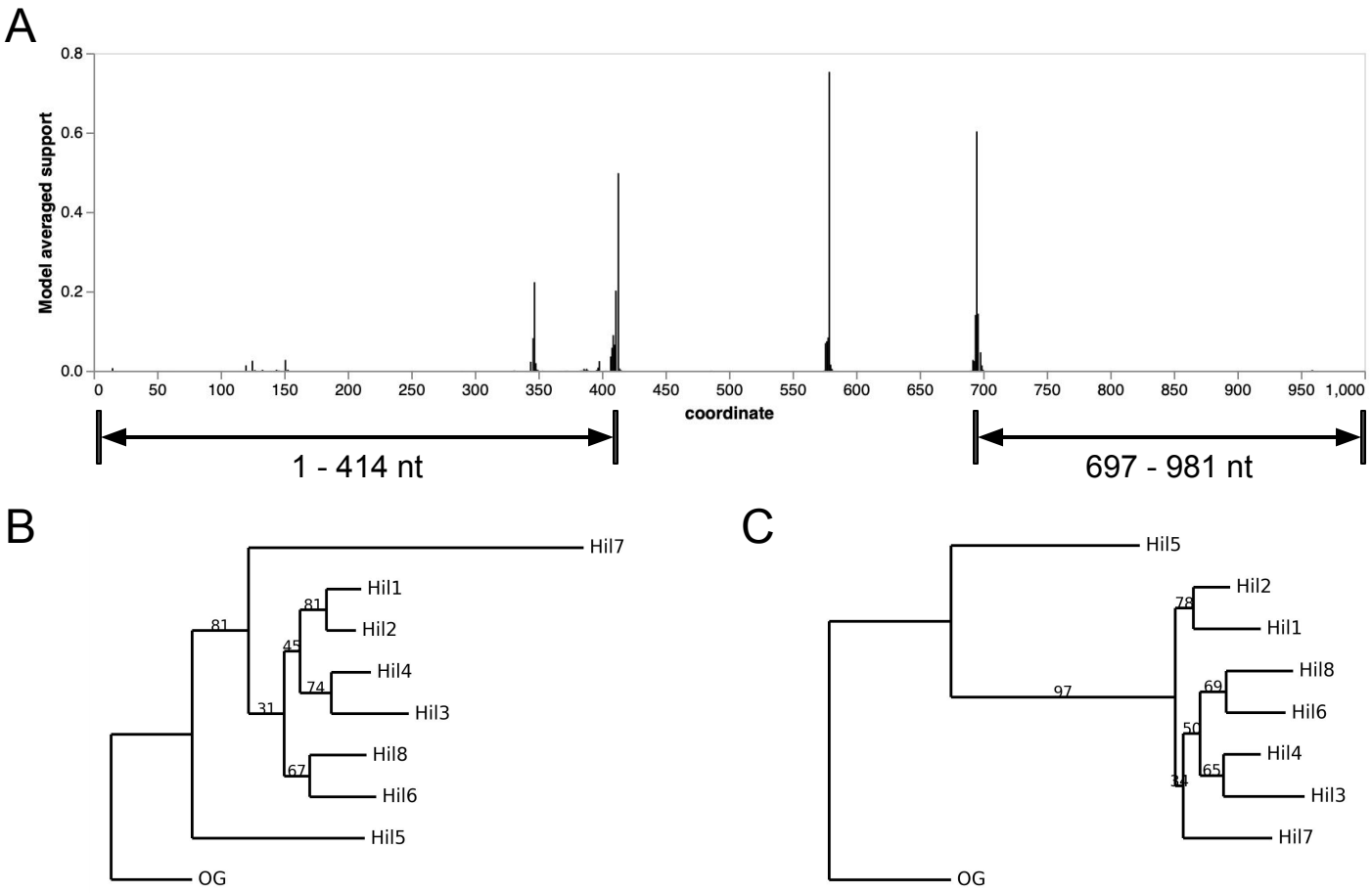


Figure S8. Detecting intra-domain recombination and identifying non-recombining partitions in the Hyphal_reg_CWP domain using GARD. (A) Model averaged support for breakpoint locations along the Hyphal_reg_CWP domain alignment for the eight Hil proteins in *C. auris* and an outgroup sequence from *M. bicuspidata* (protein ID: XP_018709340.1) to root the gene tree. Based on the GARD output, we chose the N- and C-terminal partitions for downstream analyses, i.e., coordinates 1-414 nt and 697-981 nt. (B) A maximum likelihood tree for partition 1-414 was constructed using RAxML-NG v1.1.0. Branch length is proportional to the amount of sequence divergence. OG stands for outgroup. Bootstrap support for internal splits are shown as a percentage and are based on 1000 replicates or until bootstrapping converges. (C) tree for 697-981nt, same format as in (B)

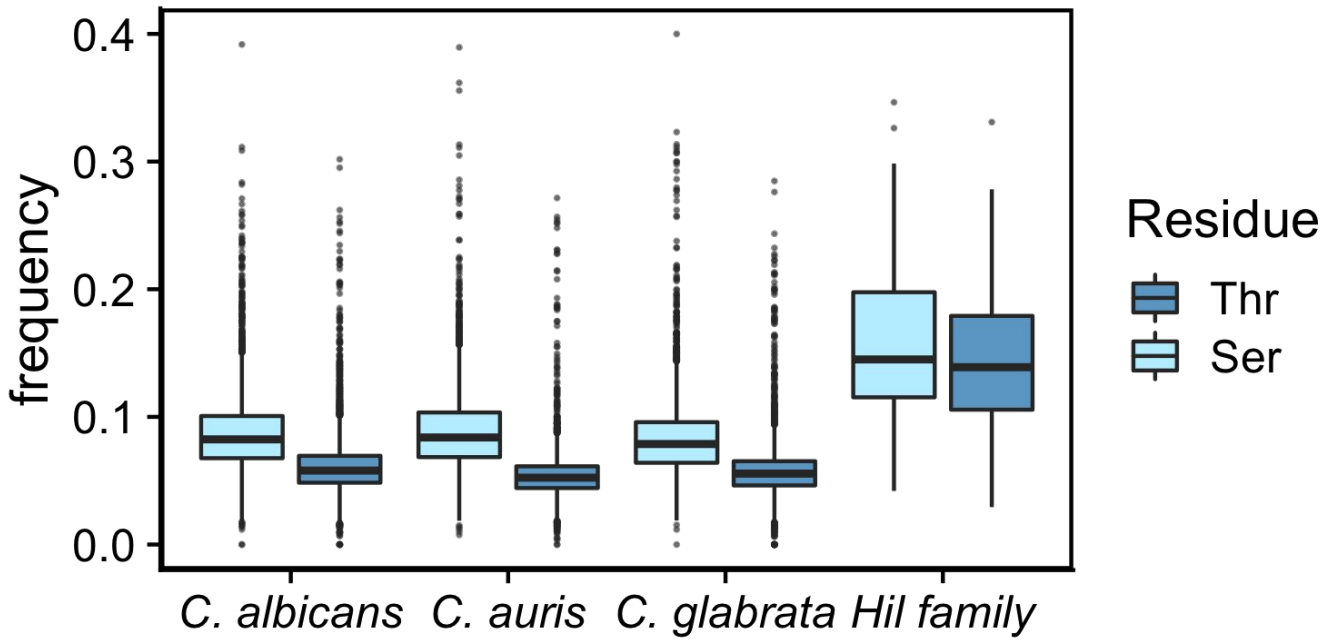


Figure S9. Yeast Hil family proteins have on average higher Ser/Thr frequencies than the rest of the proteome. Proteome-wide distribution of Thr/Ser frequencies per protein from three species, compared with the yeast Hil family proteins (*M. bicuspidata* homologs were excluded because a large number of them were incomplete). The boxes represent the interquartile range (IQR), the middle thick line the median, the whiskers the 1.5 x IQR and the dots outliers outside that range.

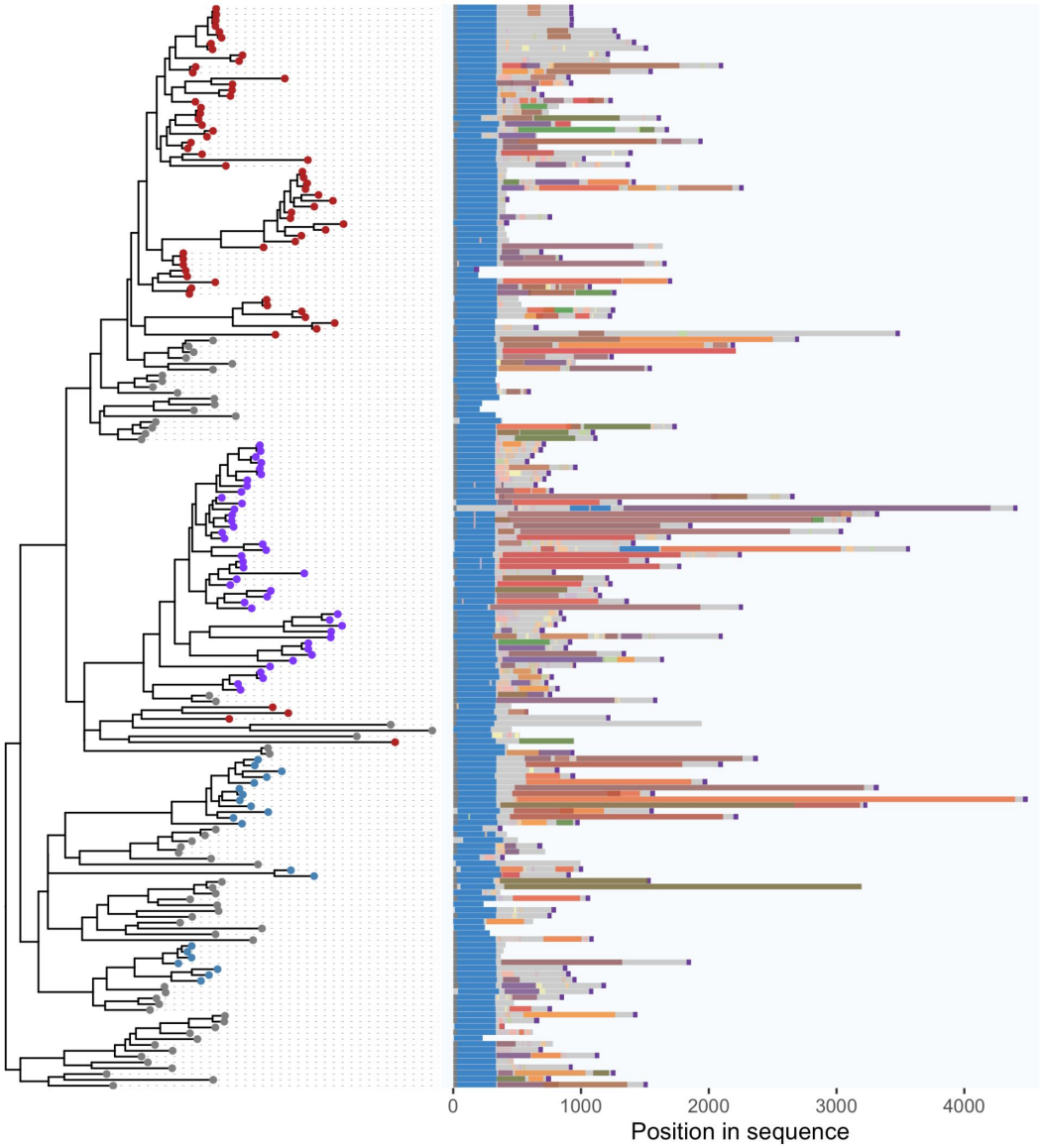


Figure S10. Domain schematic for the Yeast Hil family showing rapidly evolving tandem repeat sequences in the central domain of the proteins. Same as Fig. 6A except that tandem repeats belonging to different sequence clusters as determined by XSTREAM are shown in different colors.