Supplementary material.

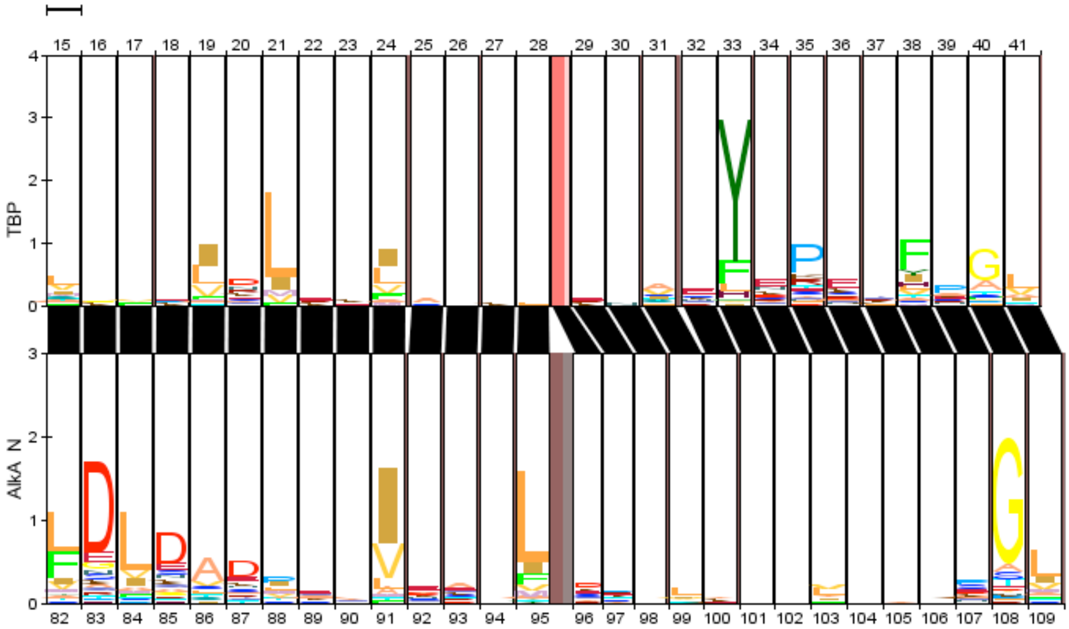

Figure S1. Pairwise HMM-logo showing relationships between Pfam families (AlkA_N and TBP). The x-axis displays information content. Combined letter height gives total information content, while the relative height of a letter shows its emission probability; letters are sorted in descending order of emission probability. Insert states are shown as pink columns, omitted states in grey, and aligned states across the profiles are connected by solid black lines. Column width varies, showing the relative contribution of a particular column to the protein family. Numbering refers to the position within each (top and bottom) HMM. Logos were generated using LogoMat-P (Schuster-Bockler and Bateman 2005). For a more detailed description of this representation, see: http://www.sanger.ac.uk/resources/software/logomat-p/

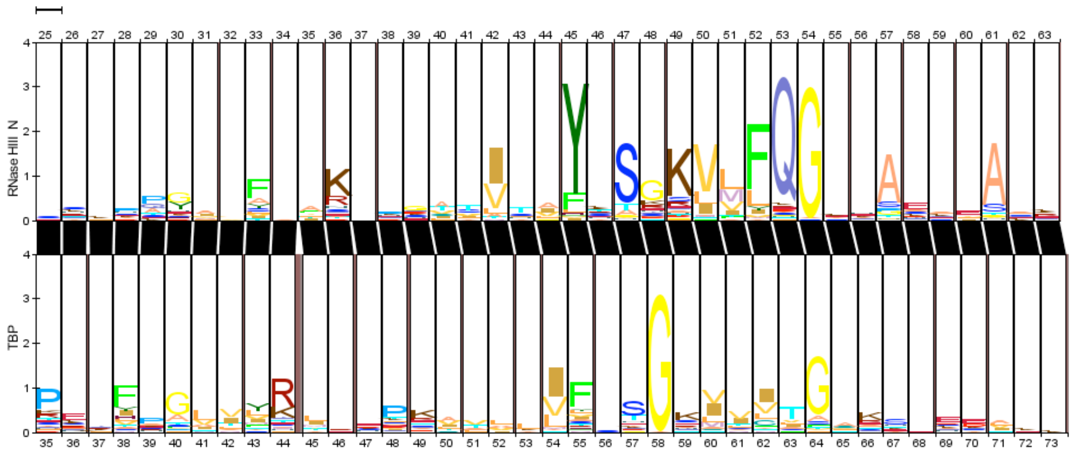

Figure S2. Pairwise HMM-logo showing relationships between Pfam families (TBP and the N-terminus of DUF3378, which corresponds to RNase HIII). The x-axis displays information content. Combined letter height gives total information content, while the relative height of a letter shows its emission probability; letters are sorted in descending order of emission probability. Insert states are shown as pink columns, omitted states in grey, and aligned states across the profiles are connected by solid black lines. Column width varies, showing the relative contribution of a particular column to the protein family. Numbering refers to the position within each (top and bottom) HMM. Logos were generated using LogoMat-P (Schuster-Bockler and Bateman 2005). For a more detailed description of this representation, see: http://www.sanger.ac.uk/resources/software/logomat-p/

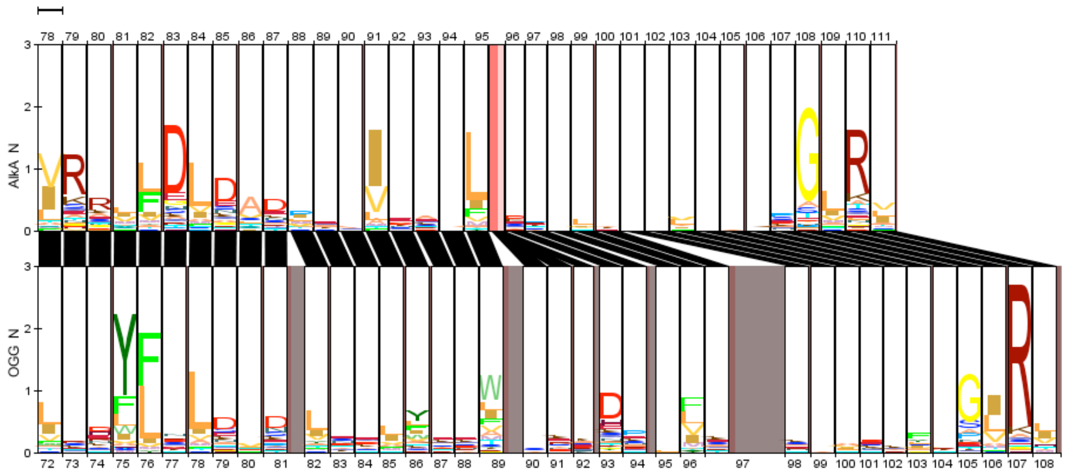

Figure S3. Pairwise HMM-logo showing relationships between Pfam families (OGG_N and AlkA_N). The x-axis displays information content. Combined letter height gives total information content, while the relative height of a letter shows its emission probability; letters are sorted in descending order of emission probability. Insert states are shown as pink columns, omitted states in grey, and aligned states across the profiles are connected by solid black lines. Column width varies, showing the relative contribution of a particular column to the protein family. Numbering refers to the position within each (top and bottom) HMM. Logos were generated using LogoMat-P (Schuster-Bockler and Bateman 2005). For a more detailed description of this representation, see: http://www.sanger.ac.uk/resources/software/logomat-p/

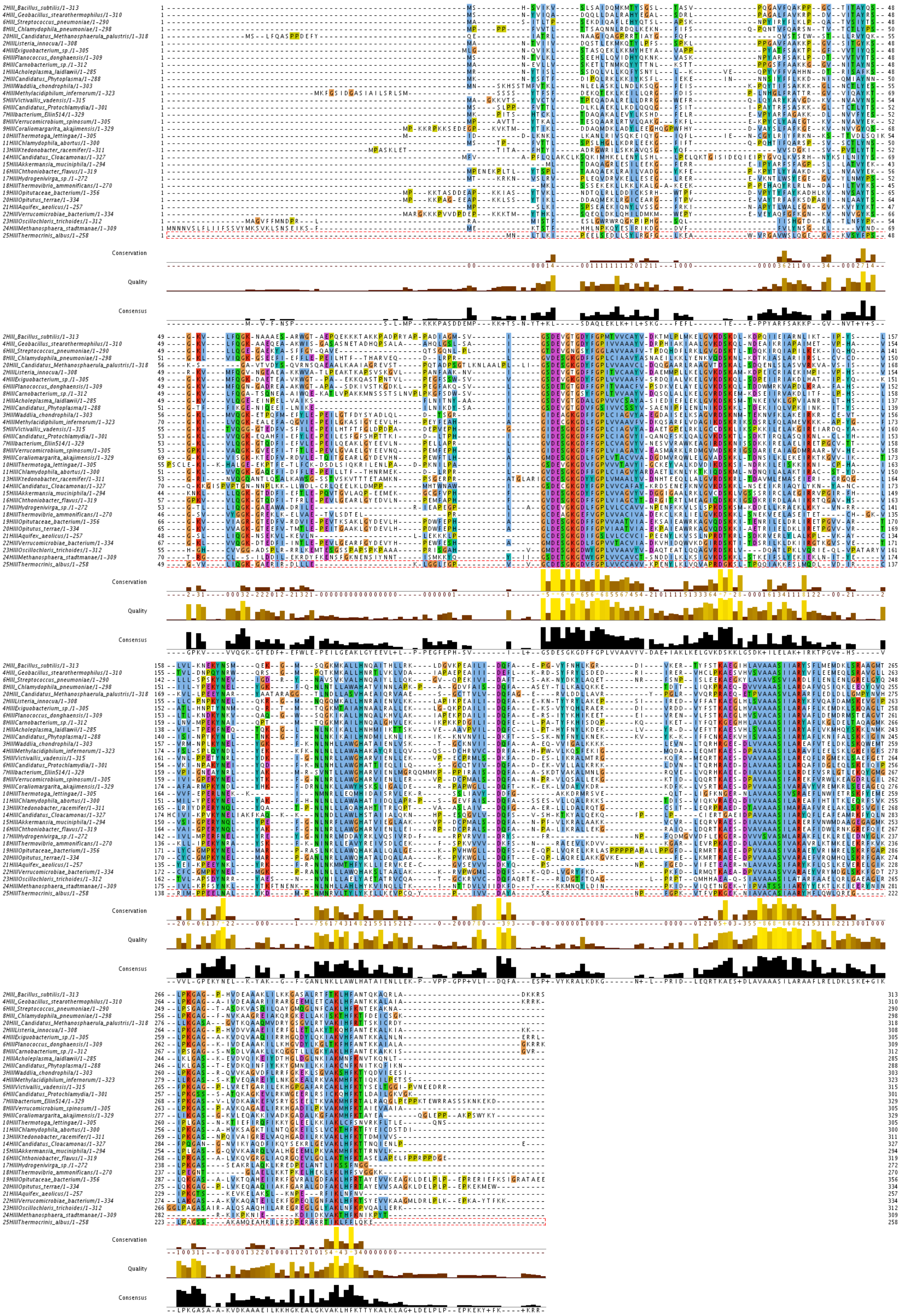

Figure S4. Multiple sequence alignment of of RNase HIII sequences, coloured by physicochemical property (as per the default ClustalX colour scheme). Conservation score of total alignment is for sites with fewer than 25% gaps, alignment quality is based on Blosum62 scores. The multiple alignment was made with Kalign v2.04 (Lassmann and Sonnhammer 2006) and figure was prepared with Jalview v. 2.07 (Waterhouse, et al. 2009). Sequence alignment files are included as part of supplementary material.

Lassmann T, Sonnhammer EL 2006. Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. Nucleic acids research 34: W596-599.
Schuster-Bockler B, Bateman A 2005. Visualizing profile-profile alignment: pairwise HMM logos. Bioinformatics (Oxford, England) 21: 2912-2913.
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191. doi: 10.1093/bioinformatics/btp033