

# Supporting Information

Menke et al. 10.1073/pnas.0909950107

## SI Text

The Dataset S1 file contains all 506 protein sequences that contain Pfam's YYY motif. Each sequence is matched against the structural motifs using random fields (SMURF) double seven-bladed propeller template, and the likelihood that the sequence contains a double seven-bladed propeller, together with the best-scoring alignment of the sequence to the template is shown. In SMURF output files, a greater-than sign (>) followed by a sequence's name marks the start of the results for each sequence. The raw score is the sum of the log probabilities, normalized by residue background frequencies. The *p* value is computed from this by fitting a Gaussian distribution to the raw scores of a non-

redundant negative set of protein data bank sequences. Following that is the alignment of the sequence to the SMURF Markov random field (MRF). The MRF is the first line, and uses all lowercase letters, indicating the residue most likely to occupy a given state, relative to its background frequency. The third line of the alignment, in all caps, is the sequence itself. Dashes (-) indicate gaps and insertions. In the line between the MRF and the sequence, each "|" indicates a position where a residue is matched to a position in the MRF. A "B" appears at each MRF position in a beta strand. Note that, because of how SMURF works, every beta strand position corresponds to a match state as well.

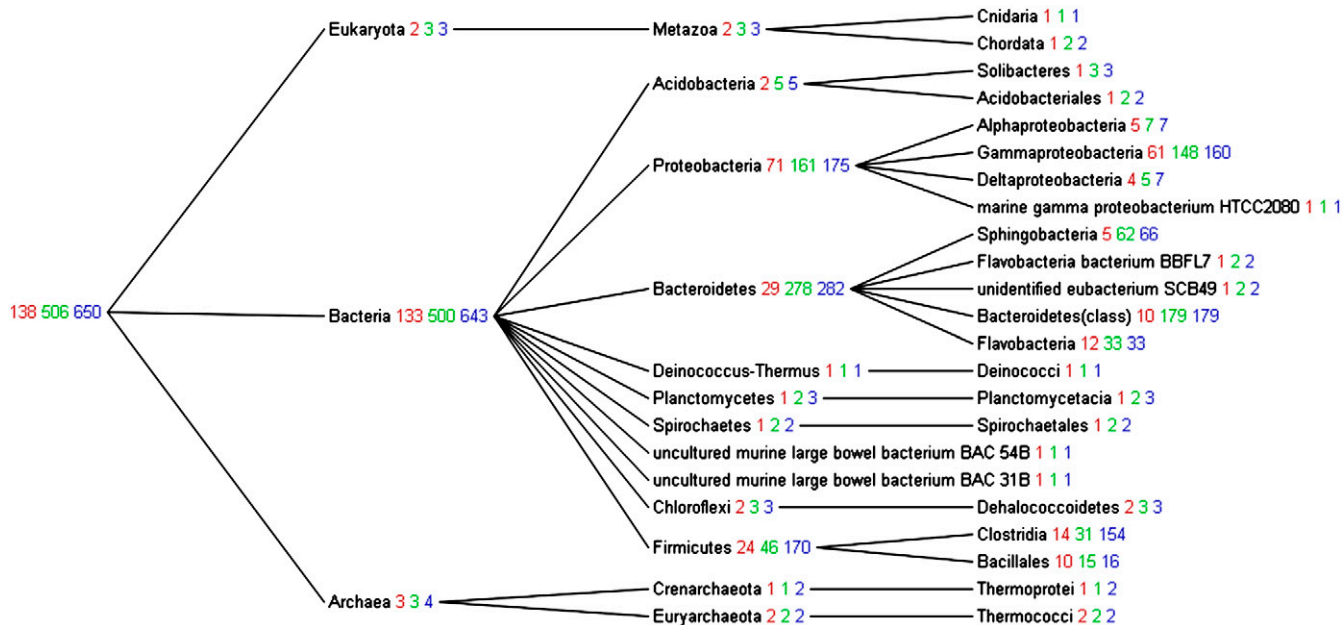


Fig. S1. Species distribution of the 506 protein sequences that contain Pfam's YYY motif at the top level. The red number gives the number of species, the green number is the number of sequences, and the blue number is the number of regions.

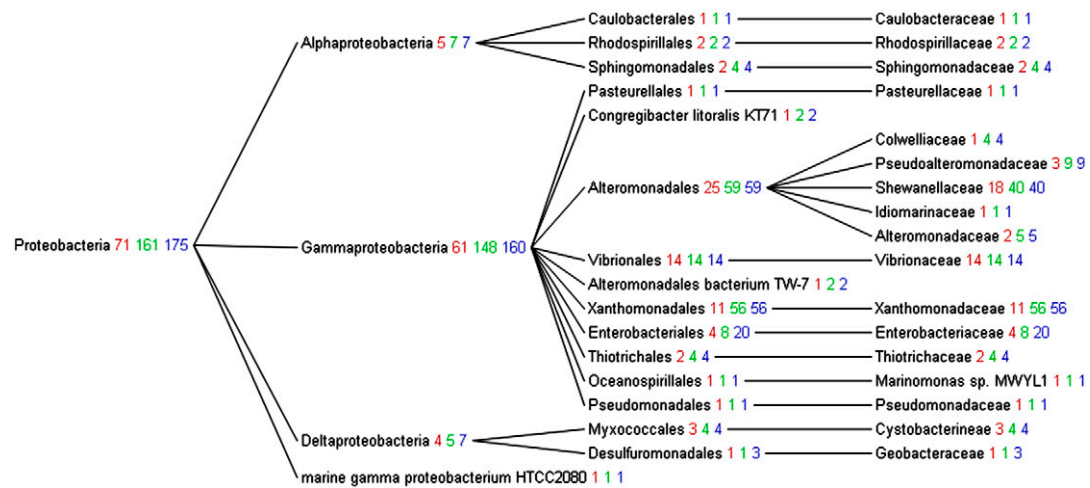


Fig. S2. Species distribution of the protein sequences that contain Pfam's YYY motif within Proteobacteria. The red number gives the number of species, the green number is the number of sequences, and the blue number is the number of regions.

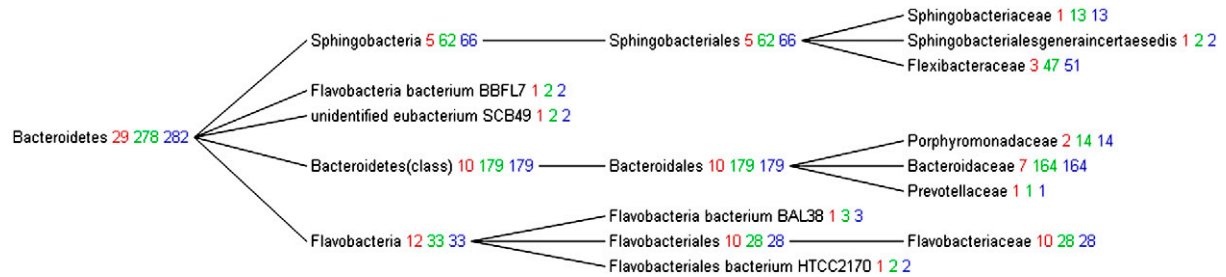


Fig. S3. Species distribution of the protein sequences that contain Pfam's YYY motif within Bacteroidetes. The red number gives the number of species, the green number is the number of sequences, and the blue number is the number of regions.

### Other Supporting Information Files

[Dataset S1 \(TXT\)](#)

[Dataset S2 \(TXT\)](#)

[Dataset S3 \(TXT\)](#)