All ccr genes were identified using blastp against all protein sequences and extracted from the cds fasta files. Only complete coding sequences were included.

Nucleotide sequences from a total of 18184 ccr genes were extracted and combined with 135 IWG reference sequences. These 19319 sequences were represented by 1363 unique sequences which were aligned using MUSCLE.

Definition of ccr types

A complete-linkage hierarchical clustering was performed based on pairwise dissimilarity between sequences in the MUSCLE alignment.