



Additional data file 3. Maximum-likelihood phylogenetic analysis of CDC7 proteins. Branching confidence was estimated based on 100 bootstraps. Identifiers are ensembl protein ids preceded by a species abbreviation. CDC7 is present in all genomes except *Tribolium*. The *Tribolium* protein is missing from the genome, the closest *Tribolium* protein is TC_1853 which is the ortholog of casein kinase CSNK2A1/CNSKA2. As *Anopheles* CDC7 is not annotated in ensembl its annotation is from a homology-based gene annotation pipeline (<http://cegg.unige.ch>). DBF4 is fast evolving and the alignment is not sufficiently long for a phylogenetic tree.