

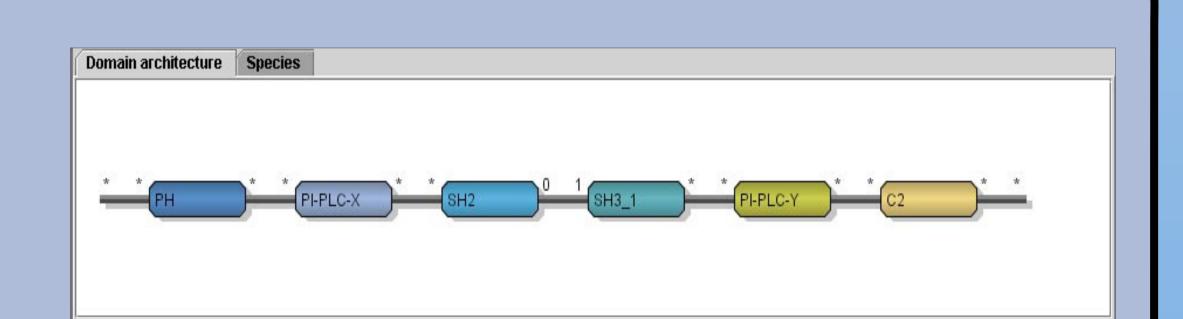
Volker Hollich and Erik L. L. Sonnhammer

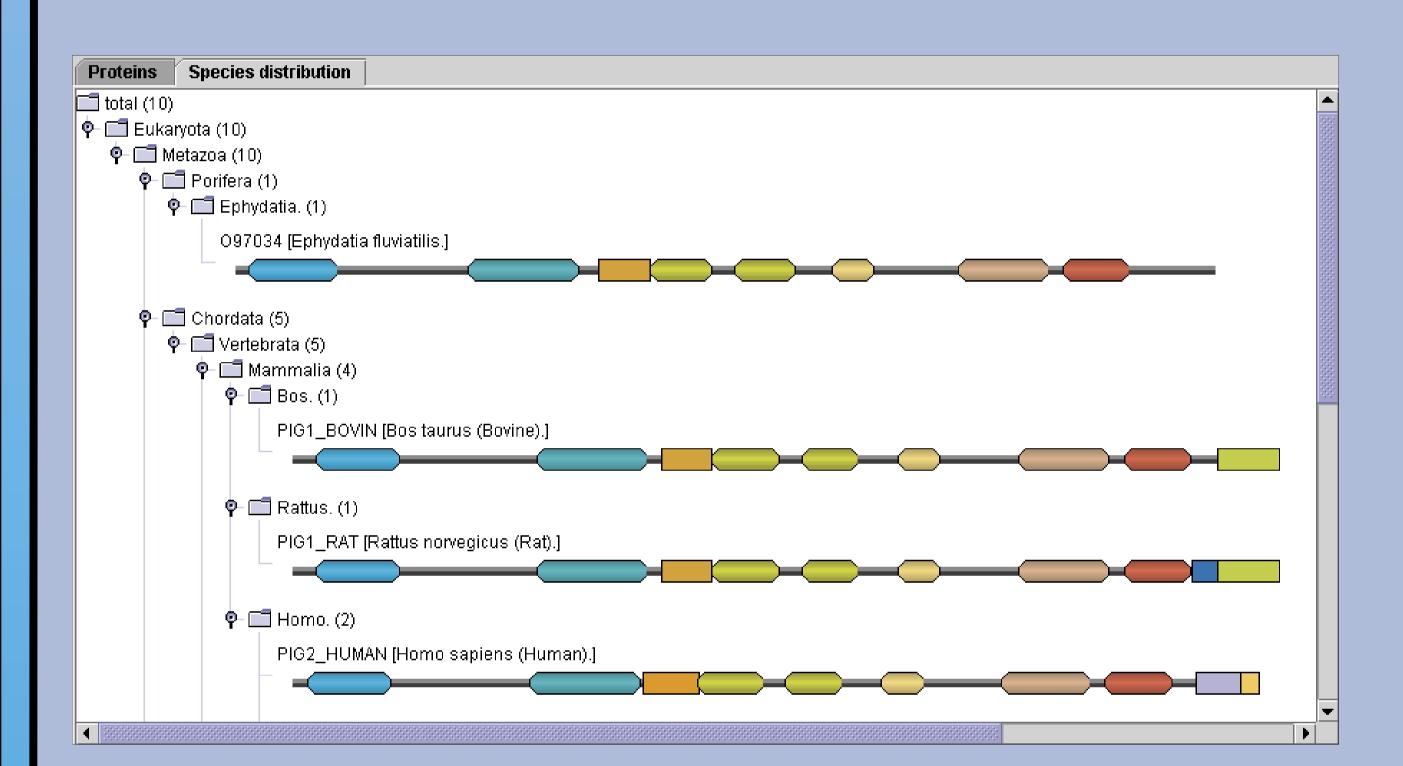
Center for Genomics and Bioinformatics Karolinska Institutet, Stockholm

PfamAlyzer paves way for an augmented understanding of protein domain architectures. It utilises the wealth of protein domain families present in the Pfam database and enables easy-to-use access to complex domain architectures as well as providing means for the study of their evolution.

Domain query

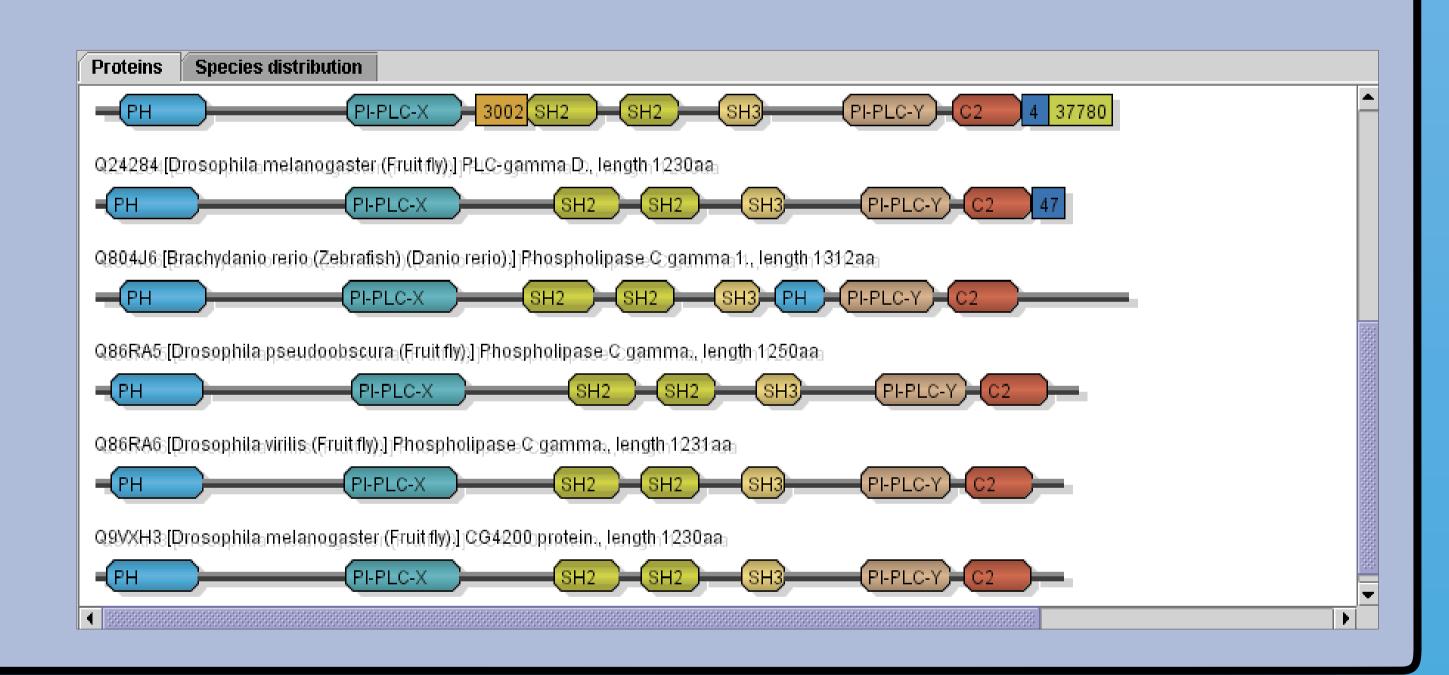
PfamAlyzer enables comprehensive study of domain architectures. Using a powerful, user-friendly graphical query language, arbitrary domains may be freely combined on the screen. PfamAlyzer uses Drag & Drop functionality as an intuitive approach for query formulation. Specific domains can be excluded from the search, in addition tolerated gaps between domains may be specified as domains or amino acids.





The manually curated Pfam-A domains are displayed as ovals, whereas automatically curated Pfam-B domains are shown as rectangles. Additional information for Pfam-A domains is available through context menus as well as links to other resources.

Optionally, the search can be limited to taxonomic groups, e.g., chordata. The proteins of SwissProt/TrEMBL with predicted domain arcitecture according to Pfam are queried for the selected architecture. The outcome is presented either in a list-fashion or as a species distribution. Here, the proteins are shown as leaves on the species tree. Subtrees may be in-/excluded from visualisation, which simplifies analysis of larges result sets.



Homology seach

Proteins | Species distribution

PfamAlyzer provides innovative homology searching. So far, the identification of homologues has been conducted by direct sequence comparison. This is appropriate if sequences are not very distant. However, for distantly related proteins the search for homologues is difficult.

ormone recep

ormone recep

Domain querying allows a new tool for finding such homologues.

GCR_SAISC [Saimiri sciureus (Common squirrel monkey):] Glucocorticoid receptor (GR):, length 778aa.

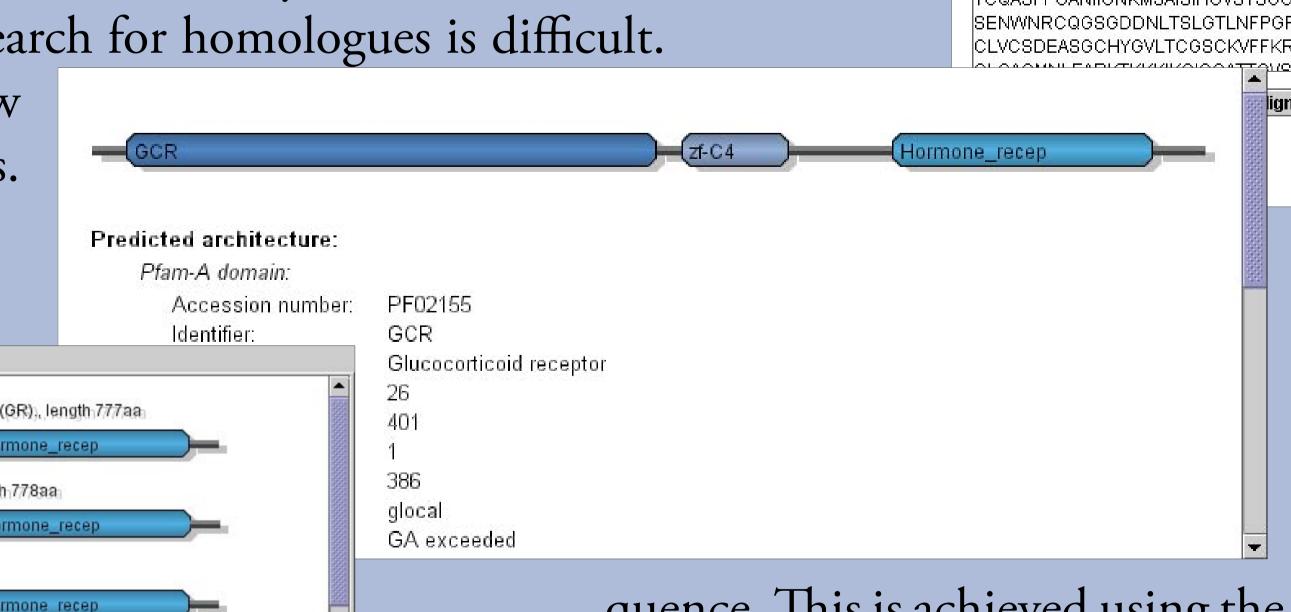
GCR_HUMAN [Homo sapiens (Human).] Glucocorticoid receptor (GR)., length 777aa

GCR_RAT [Rattus norvegicus (Rat).] Glucocorticoid receptor (GR)., length 795aa

GCR_MOUSE [Mus_musculus (Mouse).] Glucocorticoid receptor (GR)., length 783aa

GCR_CAVPO [Cavia porcellus (Guinea pig).] Glucocorticoid receptor (GR)., length 77/1 aa

GCR: ONCMY [Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).] Glucocorticoid receptor (GR)., length 758aa



MDSKESLTPGKEENPSSVLTQERGNVMDFCKILRGGATLKVSVSSTSLAAASQSDSKQQR
LLVDFPKGSVSNAQQPDLSKAVSLSMGLYMGETETKVMGNDLGFPQQGQISLSSGETDLQ
LLEESIANLNRSTSVPENPKSSASSSVSAAPKEKEFPKTHSDVSSEQQNLKGQTGSNGGN
VKLYTADQSTFDILQDLEFSSGSPGKETNQSPWKSDLLIDENCLLSPLAGEEDSFLLEGN
SNEDCKPLILPDTKPKIKDNGDLVLSSSSNVTLPQVKTEKEDFIELCTPGVIKQEKLSTV
YCQASFPGANIIGNKMSAISIHGVSTSGGQMYHYDMNTASLSQQQDQKPIFNVIPPIPVG
SENWNRCQGSGDDNLTSLGTLNFPGRTVFSNGYSSPSMRPDVSSPPSSSSTATTGPPPKL
CLVCSDEASGCHYGVLTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRYRK

PfamAlyzer introduces a domaincentric homology search based on Pfam domains. In a first step, Pfam domains are predicted from the se-

Pfam gathering threshold (GA)

○ **E-value** 1.0

quence. This is achieved using the hmmer package which is also applied for generating Pfam full alignments from curated seeds. Subsequently, SwissProt/TrEMBL is queried for the obtained domain architecture. This type of domain-centric homology analysis provides a means to identify relationships that are distant yet possess a conserved domain architecture.