

Identify conserved domain(s)
that define the family
using multiple sequence alignments



Perform multiple tblastn searches
with individual domains
including all divergent domains
using high cut-off of 10.0



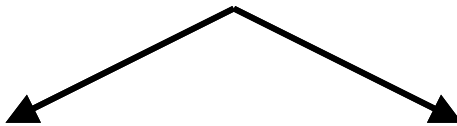
Pool hits into one data set



Assemble into contigs



Analyse each hit using blastx



Delete false positives

Annotate positives

- Full or partial domains
- Unique identifier



Catalogue minimum number of genes