

Materials and Methods

We re-classified the subfamilies of EF-hand proteins¹⁻³ using our updated database. We added new sequences from the Uniprot database⁴ by selecting entries with the keyword “EF-hand” in the domain section for five species – *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens* – and by checking the similarities of these EF-hands. Our database now contains 905 entries.

Several new subfamilies have been identified. All by all comparisons within and between closely related subfamilies were performed with FASTA⁵. The top one hundred hits for each entry were selected. Network (sif and eda) files were created as the node for each entry and the edge for similarity or identity between entries was created by using a script of perl (<https://www.perl.org/>). The network was visualized and analyzed by Cytoscape 2.8.2⁶. MCL clustering was performed with clusterMaker (v1.11), a Cytoscape plug-in⁸ in the Markov Clustering Algorithm^{11,12}.

This classification is the basis for determining the relationships among these subfamilies and their modes of action; it also permits us to infer their functional relationships and their evolution(s). Molecular evolutionary analyses were made with the PHYLIP package¹⁰ (seqboot, protdist, neighbor, and consense) (<http://evolution.genetics.washington.edu/phylip/>). Dendrograms (trees) were visualized and analyzed with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

We selected homologs of CAM, CDC, CANB, VIS, CALP and SORC from 20 species. We identified a set of bi-directional best hits for each entry of these subfamilies by searching with FASTA the top hit against each sequence database for these 20 species. We classified the hits by analyzing with Cytoscape the network created with the FASTA search.

Further, the relationships among these subfamilies was analyzed based on domain level alignment. Multiple sequence alignments were performed with clustalw¹¹. Sequence alignments were edited with Seaview¹² or Jalview¹³. RAXML was used to obtain the best maximum likelihood tree or to compare tree topology^{14,15}.

References

1. Kawasaki H, Kretsinger RH (1994) Calcium-binding proteins. 1: EF-hands. *Protein Profile* 1:343–517.
2. Kawasaki H, Nakayama S, Kretsinger RH (1998) Classification and evolution of EF-hand proteins. *Biometals* 11:277–295.
3. Nakayama S, Kawasaki H, Kretsinger R Evolution of EF-Hand Proteins. In: Carafoli PDE, Krebs P-DDJ, editors. *Calcium Homeostasis. Topics in Biological Inorganic Chemistry*. Springer Berlin Heidelberg; 2000. pp. 29–58. Available from: http://link.springer.com/chapter/10.1007/978-3-642-58306-3_2
4. Consortium TU (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204–D212.
5. Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* 85:2444–2448.
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13:2498–2504.
7. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, Bader GD, Ferrin TE (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* 12:436.
8. Enright AJ, Dongen SV, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–1584.
9. Dongen SV (2000) Graph Clustering by Flow Simulation.
10. Felsenstein J PHYLIP (Phylogeny Inference Package). 2005.
11. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.

12. Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* 27:221–224.
13. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* 20:426–427.
14. Rokas A Phylogenetic Analysis of Protein Sequence Data Using the Randomized Accelerated Maximum Likelihood (RAXML) Program. In: *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.; 2001. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/0471142727.mb1911s96/abstract>
15. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.