

## Identification of conserved C2H2 zinc-finger gene families in the Bilateria

### ABSTRACT

On the basis of current metazoan phylogeny, these 39 groups represent the minimum complement of C2H2 zinc-finger genes present in the genome of the bilaterian common ancestor.

### INTRODUCTION

Background Model organisms such as the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans* are commonly used to investigate gene function. Frequently, genes with similar sequence can be identified in the human genome, allowing prediction of human gene function by extrapolation from *Drosophila* and/or *C. elegans*. Implicit in such extrapolations is that the genes being compared are orthologous, that is, they derive from the same ancestral gene in the common ancestor of the model organism and humans. Correct identification of such relationships is therefore essential if extrapolation of function is to be fully exploited. In one form, such identifications typically utilize database comparisons with algorithms such as BLAST, with the highest-scoring sequences inferred to be orthologs. Additional criteria can then be applied to confirm orthologous relationships, including checking that orthologs have similar domain structures, and ensuring that no sequence from a more distantly related taxon is more closely related to one proposed ortholog than to another. In more complex analyses, molecular phylogenetic reconstruction of gene family history is employed. Such reconstructions help distinguish speciation from gene duplication, thereby revealing orthologous and paralogous relationships. With the near-completion of the human, *C. elegans* and *Drosophila* genome sequences, it is becoming possible to extend the identification of such relationships to analyses of large, complex gene superfamilies in the Metazoa. Such an exercise essentially reconstructs the minimum gene complement, for a particular superfamily, that would have been present in the last common ancestor of these three taxa and, given their phylogenetic relationship, gives insight into the genome complexity of the bilaterian common ancestor. Here we present an analysis of the C2H2 zinc finger (C2H2 ZNF) genes: a superfamily that, with over 600 members in humans, contains 1-2% of all human genes. C2H2 ZNF genes primarily encode DNA- and chromatin-binding transcription factors, and include familiar and well-studied developmental genes such as Krox-20, snail, Gli, Krüppel and hunchback, as well as numerous genes whose function is yet to be established. By defining orthologous relationships within this superfamily, we aim to reconstruct the minimum complement of C2H2 ZNFs present in the bilaterian common ancestor.

### CONCLUSION

Conclusions We have conclusively identified 39 families of C2H2 ZNF genes by comparing *Drosophila* and *C. elegans* sequences with human sequences. Of these, 17 have not been previously defined, and we propose that 38 represent definitive groups of orthologous genes, each deriving from a single gene in the common ancestor of these three organisms. Therefore, on the basis of current metazoan phylogeny, a member of each of these groups was primitively present in all triploblast bilaterian taxa, and they represent the minimum C2H2 ZNF complement in the bilaterian common ancestor.