ORI-GENE: A Tool for Gene Classification Based on Evolutional Tree

Hideaki Mizuno [1]
hmizuno@rtc.riken.go.jp
Yoshimasa Tanaka [1]
ytanaka@sakura.cc.tsukuba.ac.jp
Yoshihiro Ugawa [2]
ugawa@disc.dna.affrc.go.jp
Akinori Sarai [3]
sarai@rtc.riken.go.jp

[1] Institute of Biological Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-0006, Japan
[2] DNA Information & Stock center, Natl. Inst. of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-0856, Japan
[3] Tsukuba Life Science Center, The Institute of Physical & Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

As a result of progressing large-scale sequencing, a huge number of genes have been newly identified. Therefore, it became more important to predict function of these genes from its sequence. Generally, for this end, homology search by computational tools such as BLAST [1] and FASTA [2] is used as the most effective method. However, if no homologous sequences are found, or if the functions of the homologous sequences are not well-characterized, no functional information can be derived from the result.

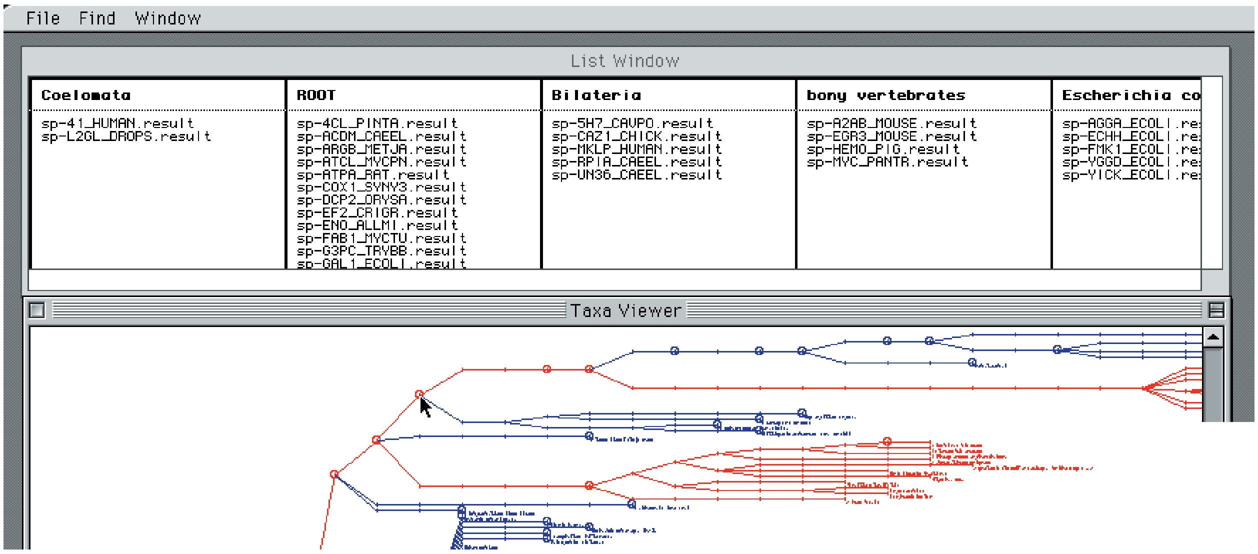
In this report, we offer new informatic approach, which is capable of extracting evolutional information from such a result in the following way.

1. First, regard homology search as a table that shows existence of similar sequences between different organisms.
2. Then classify genes by assigning the table to a given phylogenetic tree [3] and predicting its origin.

This could reveal genes that had appeared at evolutionary important periods; organisms transformed from unicellular to multicellular, from invertebrate to vertebrate and so on. It is likely that such genes have the corresponding function, and actually, the trial of sample data set of known proteins have reflected evolutional functions of genes. This suggests validity of this approach and its classification will be effective for functional prediction.

ORI-GENE is the first automated tool for this informatic approach with a graphical interface shown in Fig. 1. Using this tool, users can analyze many genes easily with point and click and visualize its result. In addition, since ORI-GENE handles all the sequences and phylogenetic relations among all species registered in GenBank as the source data, any organism's gene can be analyzed. ORI-GENE will be useful, especially for genome project, to classify a lot of newly identified genes from evolutional view and to predict functions of those, after the completion of sequencing.

Current version of ORI-GENE runs on Macintosh, and we are now working on WWW version.



Coelomata	ROOT	Bilateria	bony vertebrates	Escherichia co
sp-41_HUMAN.result	sp-4CL_PINTA.result	sp-5H7_CAVPO.result	sp-A2AB_MOUSE.result	sp-AGGA_ECOLI.re
sp-L2GL_DROPS.result	sp-ACDM_CAEEL.result	sp-CAZ1_CHICK.result	sp-EGR3_MOUSE.result	sp-ECHH_ECOLI.re
	sp-ARGB_METJA.result	sp-MKLP_HUMAN.result	sp-HEMO_PIG.result	sp-FMK1_ECOLI.re
	sp-ATCL_MYCPN.result	sp-RPIA_CAEEL.result	sp-MYC_PANTR.result	sp-YGGD_ECOLI.re
	sp-ATPA_RAT.result	sp-UN36_CAEEL.result		sp-YICK_ECOLI.re
	sp-COX1_SYNY3.result			
	sp-DCP2_ORYSA.result			
	sp-EF2_CRIGR.result			
	sp-ENO_ALLMI.result			
	sp-FAB1_MYCTU.result			
	sp-G3PC_TRYBB.result			
	sp-GAL1_ECOLI.result			