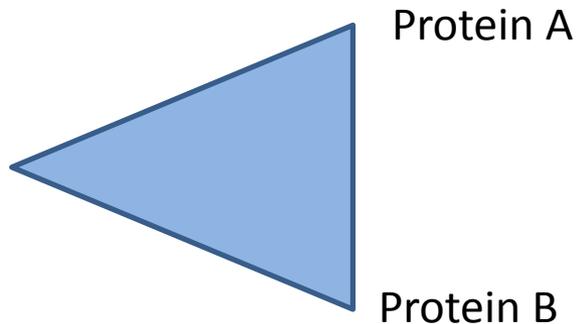


“Classic ISS”

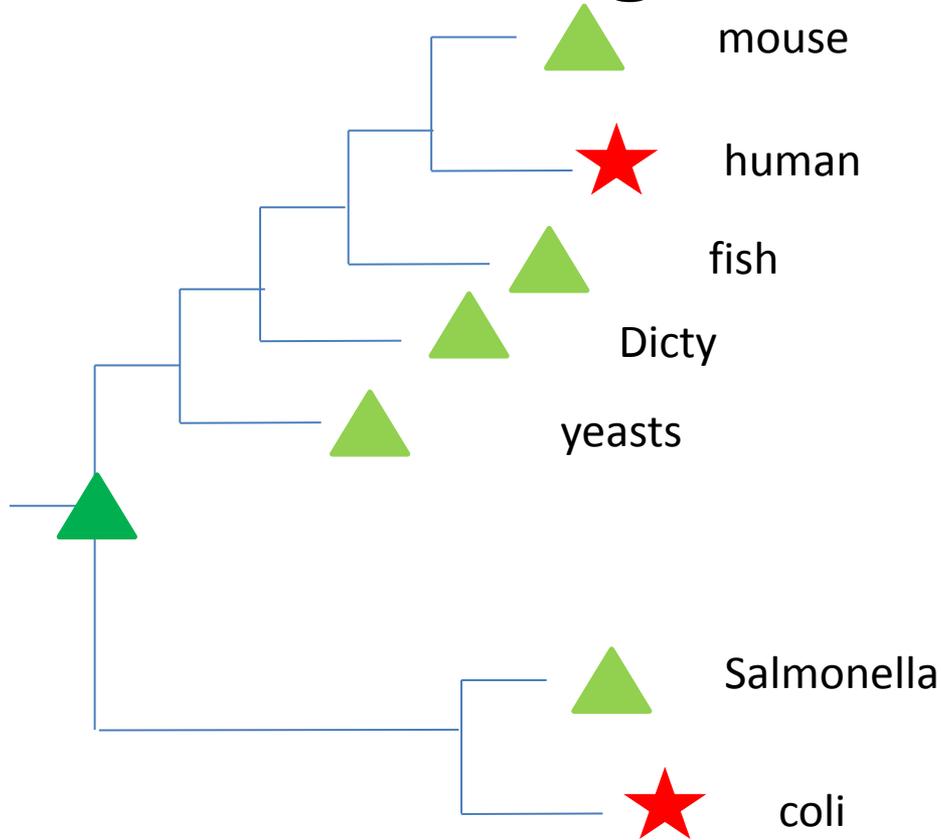
- Inference from sequence similarity
- A class of database search algorithm (e.g. BLAST) has become a metaphor
 - Implies “genes have similar functions because they have similar sequences”
 - Function is usually determined using pairwise comparison

Triangulation

- Inferences are based on EXP annotation from at least two sequences
- All sequences that fall 'between' those two sequences phylogenetically get annotated



How triangulation works



EXP annotations

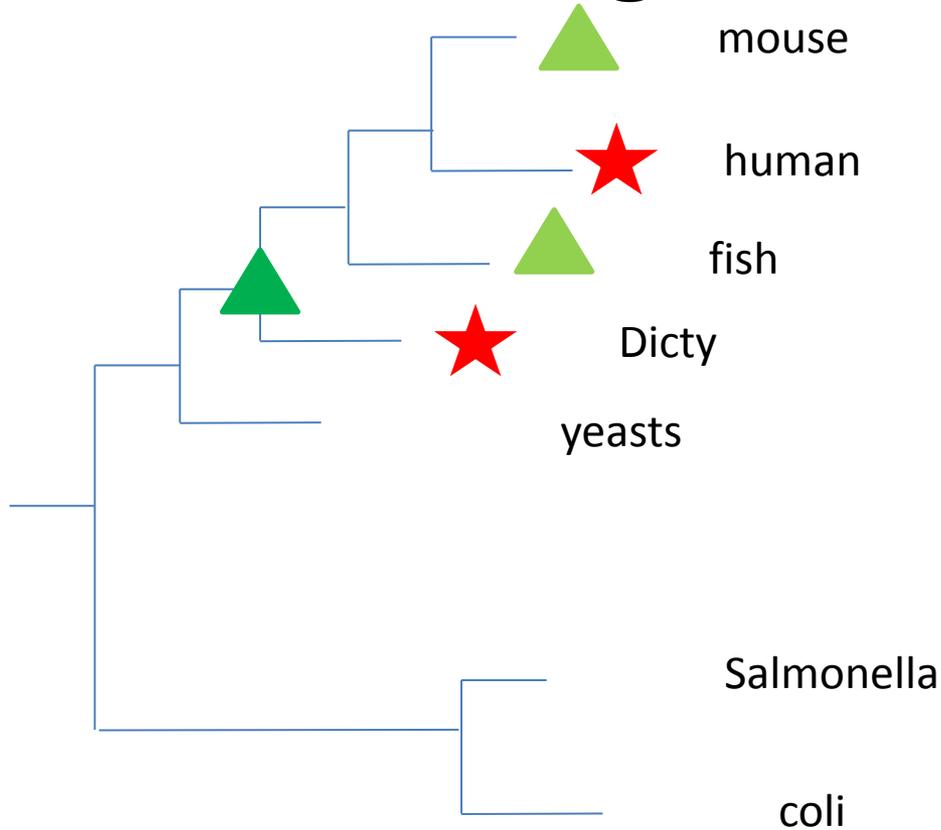


Annotated ancestor



Inferred annotations

How triangulation works



EXP annotations

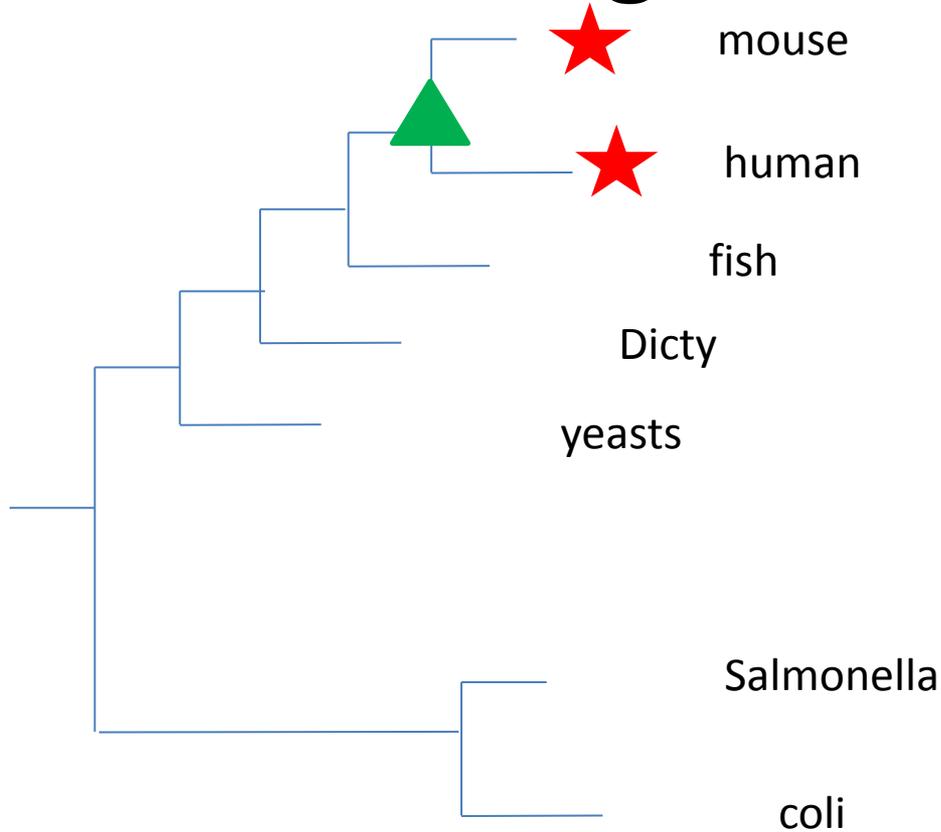


Annotated ancestor



Inferred annotations

How triangulation works



EXP annotations

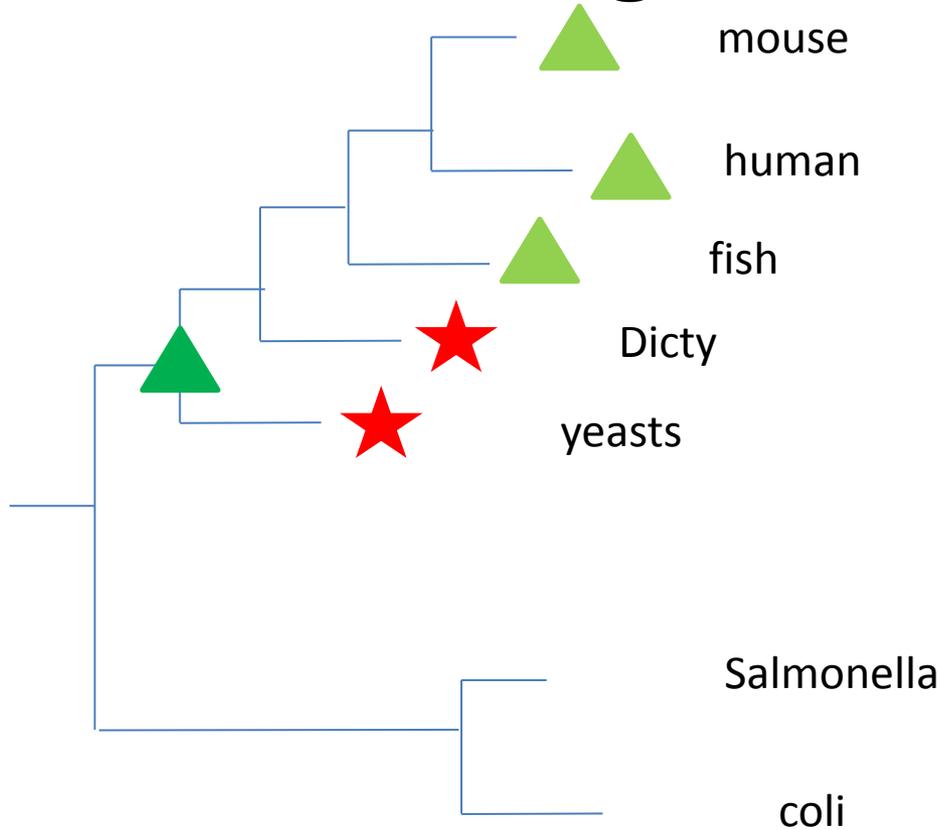


Annotated ancestor



Inferred annotations

How triangulation works



EXP annotations

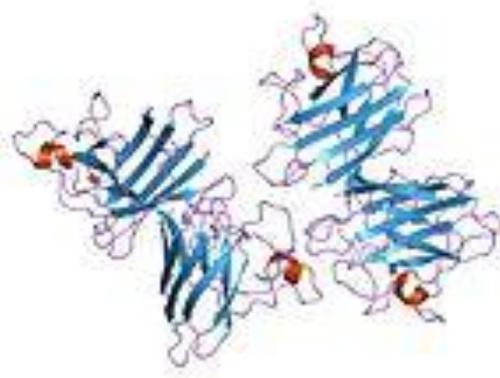


Annotated ancestor

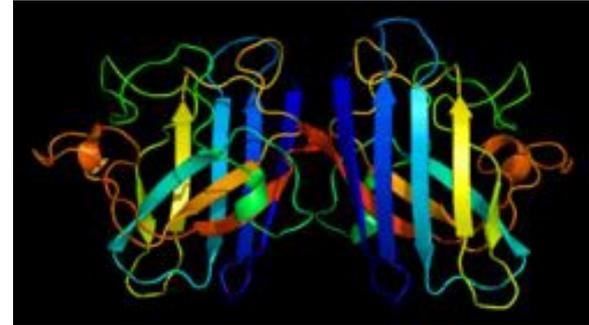


Inferred annotations

Example: Superoxide dismutase



Yeast SOD1



Human SOD1

- SOD requires copper as a co-factor. In Eukaryotes there is a copper chaperone (CCS) that delivers the Cu to SOD.
- There are SODs in all kingdoms of life.

Human SOD1 vs human CCS

Score = 137 bits (344), **Expect = 2e-37**, **Method:** Compositional matrix adju

Identities = 73/146 (50%), Positives = 94/146 (64%), Gaps = 4/146 (3%)

```
Query 5   AVCVLKGDGPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVHEFGDNTAGCTSAGPH 64
          AV +L G G VQG++ F Q      + + G+I GL  GLHG HVH++GD T  C S G H
Sbjct 89   AVAILGGPGTVQGVVRFLLQLTPERCL-IEGTIDGLEPGLHGLHVHQQYGDLTNNCNSCGNH 147

Query 65  FNPLSRKHGGPKDEERHVGDLGNVTADKDGVADVSIEDSVISLSGDHCIIGRTLTVVHEKA 124
          FNP      HGGP+D +RH GDLGNV AD DG A  +ED  + +      +IGR+L++ E
Sbjct 148  FNPDGASHGGPQDSDRHRGDLGNVRADADGRAIFRMEDEQLKVWD---VIGRSLIIDEGE 204

Query 125 DDLGKGGNEESTKKTGNAGSRLACGVI 150
          DDLG+GG+ S  TGN+G RLACG+I
Sbjct 205  DDLGRGGHPLSKITGNSGERLACGII 230
```

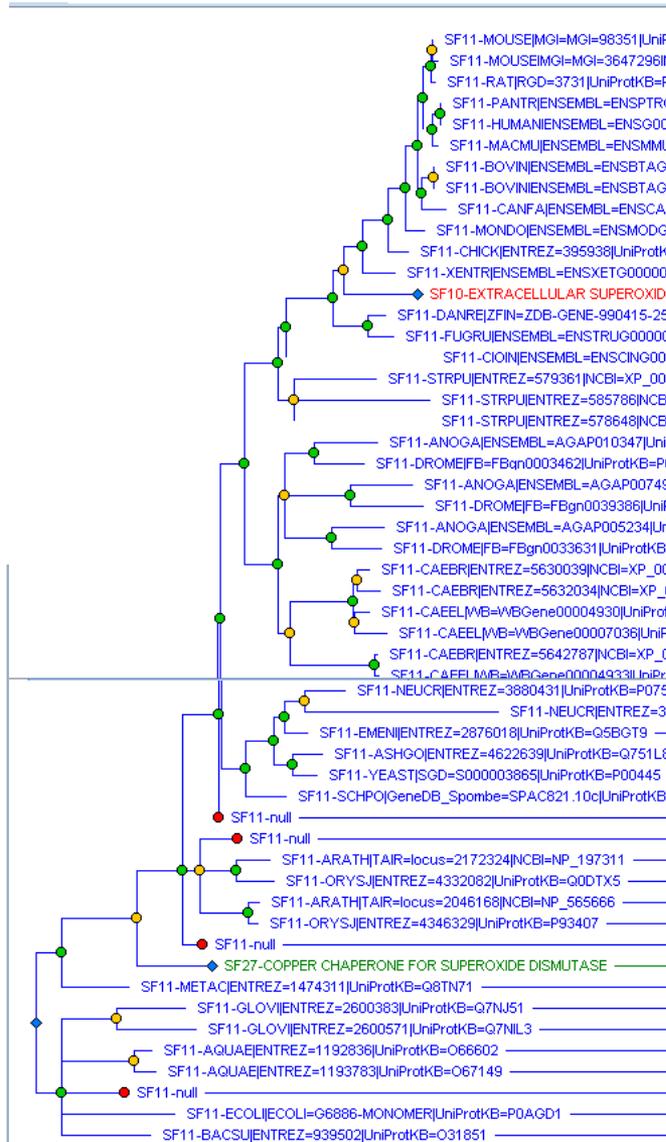
Human SOD1 versus E. coli SodC

```
>lcl|14001 sp|P0AGD1|SODC_ECOLI Superoxide dismutase [Cu-Zn] OS=Escherichia coli (strain K12) GN=sodC PE=1 SV=1  
Length=173
```

Score = 49.3 bits (116), **Expect = 3e-11**, Method: Compositional matrix adjust.

Identities = 45/145 (31%), Positives = 67/145 (46%), Gaps = 18/145 (12%)

```
Query 16 QGIINFEQKESNGPVKVGWGSIKGLTEGLHGFHVHEFGD-----NTAGCTSAGPHFNP 67  
      Q I + E++ ++ +K L G HGFH+H G + SAG H +P  
Sbjct 36 QSIGSVTITETDKGLEFSPDLKALPPGEHGFHIHAKGSCQPATKDGKASAAESAGGHLDP 95  
  
Query 68 LSR-KHGGPKDEERHVGD LGNVTADKDGVA-DVSIEDSVISLSGDHCIIGRTL VVHEKAD 125  
      + KH GP + H+GDL + + DG A D I + SL I + L+VH D  
Sbjct 96 QNTGKHEGP-EGAGHLGDLPALVVNNDGKATDAVIAPRLKSLDE---IKDKALMVHVGGD 151  
  
Query 126 DLGKGGNEESTKTGNAGSRLACGVI 150  
      ++ +++ G G R ACGVI  
Sbjct 152 NM----SDQPKPLGGGGERYACGVI 172
```

| Protein Id | sf_name | definition | organism |
|--|----------------------------|--------------------------------|------------------|
| MOUSE[MG]=... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Mus musculus |
| MOUSE[MG]=... | SUPEROXIDE DISMUTASE [C... | PREDICTED: hypothetical pr... | Mus musculus |
| RAT[IRGD]=37... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Rattus norve... |
| PANTR[ENSE]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Pan troglodyt... |
| HUMAN[ENS]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Homo sapiens |
| MACMU[ENS]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Macaca mula... |
| BOVIN[ENSE]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Bos taurus |
| BOVIN[ENSE]... | SUPEROXIDE DISMUTASE [C... | Thymus HYPOCHOLESTERO... | Bos taurus |
| CANFA[ENS]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Canis familiaris |
| MONDO[ENS]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Monodelphis... |
| CHICK[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Gallus gallus |
| XENTR[ENSE]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Xenopus tro... |
| AN54 | EXTRACELLULAR SUPEROX... | | |
| DANRE[JZFN]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Danio rerio |
| FUGRU[ENSE]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Fugu rubripes |
| CIOIN[ENSEM]... | SUPEROXIDE DISMUTASE [C... | | Ciona intesti... |
| STRPU[JENTR]... | SUPEROXIDE DISMUTASE [C... | PREDICTED: similar to super... | Strongyloce... |
| STRPU[JENTR]... | SUPEROXIDE DISMUTASE [C... | PREDICTED: similar to super... | Strongyloce... |
| STRPU[JENTR]... | SUPEROXIDE DISMUTASE [C... | PREDICTED: similar to super... | Strongyloce... |
| ANOGA[ENS]... | SUPEROXIDE DISMUTASE [C... | SOD3B (Fragment) | Anopheles g... |
| DROME[IFB]=F... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Drosophila m... |
| ANOGA[ENS]... | SUPEROXIDE DISMUTASE [C... | | Anopheles g... |
| DROME[IFB]=F... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Drosophila m... |
| ANOGA[ENS]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Anopheles g... |
| DROME[IFB]=F... | SUPEROXIDE DISMUTASE [C... | RE42883p (Fragment) | Drosophila m... |
| CAEBR[ENTR]... | SUPEROXIDE DISMUTASE [C... | Hypothetical protein CBG11... | Caenorhabdi... |
| CAEBR[ENTR]... | SUPEROXIDE DISMUTASE [C... | Hypothetical protein CBG00... | Caenorhabdi... |
| CAEEL[WB]=... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Caenorhabdi... |
| CAEEL[WB]=... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Caenorhabdi... |
| CAEBR[ENTR]... | SUPEROXIDE DISMUTASE [C... | hypothetical protein CBG10... | Caenorhabdi... |
| CAEEL[WB]=... | SUPEROXIDE DISMUTASE [C... | Extracellular superoxide di... | Caenorhabdi... |
| NEUCR[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Neurospora ... |
| NEUCR[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Neurospora ... |
| EMEN[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Aspergillus n... |
| ASHGO[ENT]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Ashbya gos... |
| YEAST[SGD]=S000003865[UniProtKB=P00445] | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Saccharomy... |
| SCHPO[Gene]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Schizosacch... |
| AN139 | | | |
| AN157 | | | |
| ARATH[TAIR]... | SUPEROXIDE DISMUTASE [C... | CSD3 (COPPER/ZINC SUPE... | Arabidopsis ... |
| ORYSJ[ENTR]... | SUPEROXIDE DISMUTASE [C... | Os03g0219200 protein (Fra... | Oryza sativa |
| ARATH[TAIR]... | SUPEROXIDE DISMUTASE [C... | CSD2 (COPPER/ZINC SUPE... | Arabidopsis ... |
| ORYSJ[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Oryza sativa |
| AN146 | | | |
| AN3 | COPPER CHAPERONE FOR ... | | |
| METAC[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase | Methanosarc... |
| GLOVI[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Gloeobacter ... |
| GLOVI[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Gloeobacter ... |
| AQUAE[ENT]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Aquifex aeol... |
| AQUAE[ENT]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Z... | Aquifex aeol... |
| AN165 | | | |
| ECOLI[ECOLI]=G6886-MONOMER[UniProtKB=P0AGD1] | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase [Cu-Zn] | Escherichia ... |
| BACSU[ENTR]... | SUPEROXIDE DISMUTASE [C... | Superoxide dismutase-like p... | Bacillus subt... |

● Nodes (Sequences) ● Duplication ● (Collapsed)

← SOD activity by EXP: Fly, pombe, cerevisiae, A. thaliana, E. coli

CCS

What triangulation would do

1. Annotate entire tree to SOD activity (since one mammal and bacteria have it)
2. Fail to annotate Copper chaperone activity – since there is only one annotation

Could we come up with rules?

- Duplications are opportunities for divergence (see CCS), but not always (SOD1 versus SOD3)
- ‘Compara-type’ of inference process: whereby you chose a group of related species (mammals, animals, etc) among which you can transfer all annotations (see next slide)

Compara annotations for Human SOD1

| | |
|--|---|
| activation of MAPK activity | regulation of blood pressure |
| aging | regulation of multicellular organism growth |
| anti-apoptosis | relaxation of vascular smooth muscle |
| auditory receptor cell stereocilium organization | removal of superoxide radicals |
| cellular iron ion homeostasis | removal of superoxide radicals |
| chaperone binding | response to amphetamine |
| cytoplasm | response to axon injury |
| DNA fragmentation involved in apoptotic nuclear change | response to copper ion |
| double-strand break repair | response to drug |
| embryo implantation | response to ethanol |
| glutathione metabolic process | response to heat |
| heart contraction | response to hydrogen peroxide |
| hydrogen peroxide biosynthetic process | response to nutrient levels |
| hydrogen peroxide biosynthetic process | response to oxidative stress |
| locomotory behavior | response to reactive oxygen species |
| mitochondrion | response to superoxide |
| muscle cell homeostasis | retina homeostasis |
| myeloid cell homeostasis | sensory perception of sound |
| negative regulation of apoptosis | spermatogenesis |
| negative regulation of neuron apoptosis | superoxide anion generation |
| neurofilament cytoskeleton organization | superoxide dismutase activity |
| neuronal cell body | superoxide dismutase activity |
| nucleus | superoxide metabolic process |
| ovarian follicle development | superoxide metabolic process |
| peripheral nervous system myelin maintenance | transmission of nerve impulse |

“ISS”

- Inference from sequence similarity
- A class of database search algorithm (e.g. BLAST) has become a metaphor
 - Implies “genes have similar functions because they have similar sequences”
 - Function is usually determined using pairwise comparison

“ISS”

- More properly, ISS of function is inheritance!
 - “related genes have a common function because their common ancestor had that function, which was inherited by its descendants”
 - ISS is not just a statement about one gene. It is also making assertions about
 - The common ancestor
 - Inheritance of a “character” by
 - Both “pairwise similar” descendants
 - Other descendants