

Supporting Text

Methods

Sequence data and annotations were downloaded from GenBank. Homology searches were performed with BLAST (1) and WU-BLAST (<http://blast.wustl.edu>). Shared genes were defined by using a BLASTP value cutoff of 10^{-6} and share similarity for at least 50% of the length of their protein products. Functional schemes in clusters of orthologous groups (2) were utilized for protein classification. Functional comparisons at a protein domain level were based on INTERPRO (3) and GENE ONTOLOGY (GO) (4). Maps from KEGG were used for metabolic reconstruction (5).

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410.
2. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000) *Nucleic Acids Res.* **28**, 33-36.
3. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M. D., *et al.* (2001) *Nucleic Acids Res.* **29**, 37-40.
4. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* (2000) *Nat. Genet.* **25**, 25-29.
5. Kanehisa, M., Goto, S., Kawashima, S. & Nakaya, A. (2002) *Nucleic Acids Res.* **30**, 42-46.