

Fauchon *et al.*<sup>2</sup> could also be interpreted to suggest that these other stresses do not require such a drastic increase in glutathione levels for cell survival, pointing to a pivotal role for glutathione in mediating resistance to cadmium.

How general is this response? Does the cadmium response occur in other yeast, and more importantly in higher eukaryotes? Fauchon *et al.*<sup>2</sup> suggest that the cadmium response could have evolved as a strategy for protecting against toxic levels of heavy metals accumulated by the plants that yeast live on. They go on to point out that other plant-living yeasts, such as *Schizosaccharomyces pombe* and *Pichia stipitis*, possess low-sulfur-con-

taining pyruvate decarboxylase isoenzymes, whereas non-plant-living yeasts such as *Candida albicans* do not. With the availability of various microbial and other genome sequences, it should be possible to determine how general this response is. In particular, it would be interesting to focus on plant pathogens. Equally intriguing is the finding that regulation of the cadmium response by Met4p is not the whole story. It turns out that many proteins, particularly those that are less abundant in the cell, are cadmium-inducible in a Met4p-independent manner. It will be especially interesting to examine in detail the factors required for the cadmium-mediated

induction of these proteins. As with all good studies, the report by Fauchon *et al.*<sup>2</sup> has provided us with many questions and avenues for future research. □

1. Avery, S.V. *Adv. App. Microbiol.* **49**, 111–142 (2001).
2. Fauchon, M. *et al. Mol. Cell* **9**, 713–723 (2002).
3. Westwater, J., McLaren, N.F., Dormer, U.H. & Jamieson, D.J. *Yeast* **19**, 233–239 (2002).
4. Jamieson, D.J. *Yeast* **14**, 1511–1527.
5. Wemmie, J.A., Szczypka, M.S., Thiele, D.J. & Moye-Rowley, W.S. *J. Biol. Chem.* **269**, 32592–32597 (1994).
6. Vido, K. *et al. J. Biol. Chem.* **276**, 8469–8474 (2001).
7. Meister, A. & Anderson, M.E. *Annu. Rev. Biochem.* **52**, 711–760 (1983).
8. Stephen, D.W. & Jamieson, D.J. *Mol. Microbiol.* **23**, 203–210 (1997).
9. Dormer, U.H. *et al. J. Biol. Chem.* **275**, 32611–32616 (2000).
10. Elskens, M.T., Jaspers, C.J. & Penninckx, M.J. *J. Gen. Microbiol.* **137**, 637–644 (1991).
11. Grant, C.M., MacIver, F.H. & Dawes, I.W. *Curr. Genet.* **29**, 511–515 (1996).



## Reversing Babel with GO

The Gene Ontology project allows biologists to share knowledge; a new study demonstrates that GO terms can aid in the identification of candidate 'disease' genes.

An untruth sometimes said about scientists is that they lack artistic creativity. As any reader of *Nature Genetics* can attest, scientific papers are rife with creative language describing experimental data, natural phenomena and interpretation. So it's likely that a not a few biologists grumbled when the Gene Ontology (GO) consortium announced its efforts to design a common language for describing the functions of genes across organisms; a tool, it was claimed, that would serve to unify biology<sup>1</sup>. Some two years later, GO is beginning to realize its lofty goal and, moreover, it is being applied in unanticipated ways. The tool described by Carolina Perez-Iratxeta and colleagues<sup>2</sup> on page 316 is one such example.

An ontology defines a controlled, consistent vocabulary to describe concepts and relationships, thereby enabling knowledge-sharing<sup>3</sup>. The GO consortium was inspired by the recognition of a bottleneck in the transfer of information between those studying different model organisms, owing to the absence of a shared vocabulary<sup>4</sup>. To circumvent this problem, they commenced development of three ontologies applicable to all eukaryotes: the biological process in which the gene product participates, the molecular function of the gene product and the cellular component within which the gene product acts. Although GO

*And the Lord said,  
"Behold, they are one people, and they  
have all one language; and this is only the beginning  
of what they will do; and nothing that they propose  
to do will now be impossible for them. Come, let us  
go down, and there confuse their language, that they  
may not understand one another's speech."  
Genesis 11*

terms are consistent, they are not complete, thereby allowing a dynamic vocabulary that evolves within the constraints of the ontology. The consortium initially included FlyBase, Mouse Genome Informatics and *Saccharomyces* Genome Database, but has subsequently grown to include the Arabidopsis Information Resource, WormBase, PomBase, the Rat Genome Database and DictyBase among others, which are now united by the use of a single shared vocabulary<sup>4</sup>.

Perez-Iratxeta and colleagues<sup>2</sup> report an approach in which they produce a score that links the functional annotation of proteins described using GO terms with the description of an inherited disease using medical subject heading (MeSH; the National Library of Science's controlled vocabulary). Linking this score to information from RefSeq yielded a list of most likely candidate genes for 455 mapped diseases of unknown genetic defect. As a blind test, the authors looked at 100 genes for

which disease-causing mutations had already been identified. In 55 of the cases, the disease-related gene was identified. As the authors point out, the data-mining system is highly dependent on the information that is being mined. Thus, continued improvements in GO and its increased integration in databases will enhance the authors' data-mining system (available on the Genes2Disease website).

The efforts of the GO consortium are helping to lead us beyond a Babel-like period and unify the field. And the tool reported by Perez-Iratxeta and colleagues<sup>2</sup> suggests that speaking the same language will have unexpected benefits. □

—David Gresham

1. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
2. Perez-Iratxeta, C., Bork, P. & Andrade, M.A. *Nature Genet.* **31**, 316–319 (2002).
3. Gruber, T.R. *Knowl. Acq.* **5**, 199–220 (1993).
4. Ashburner, M. *et al. Genome Res.* **11**, 1425–1433 (2001).

