Functional misassignment of genes

To the editor:

As we move into the postsequencing phase of many genome projects, a major focus for the next few years will be to accurately annotate the functions of genes on a genomic scale. A powerful approach for predicting the exact biochemical function of a new protein is to use bioinformatics to find its characterized orthologs that typically retain the same function in the course of evolution. Currently, computer-based programs (e.g., BLAST, PSI-BLAST, COG) have been developed that enable searches for orthologs in databases. However, these similarity-based analyses are problematic and can generate large amounts of grossly misannotated sequences.

Using isocitrate and isopropylmalate dehydrogenases family as a model, with insight into how distinct functions of orthologs and paralogs are conferred and evolved, we believe it is feasible to confidently identify orthologs. Our strategy is based on a recent finding from protein engineering studies that indicates substitutions of only a few amino acid residues in these enzymes are sufficient to exchange substrate and coenzyme specificities. Hence, a few major specificity determinants can serve as reliable markers for determining orthologous or paralogous relationships. The approach has effectively corrected similarity-based functional misassignments of some 30 or so family members. To illustrate our approach, we present two case studies below:

The sequence (Gene number: Aq1512) from the *Aquifex aeolicus* genome sequencing project is predicted to encode a NADP-isocitrate dehydrogenase (NADP-IDH, EC 1.1.1.42) based on its high sequence identity with bacterial NADP-IDH[1]. Structure-based sequence alignment shows that all of the substrate binding and catalytic residues identified in *Escherichia coli* NADP-IDH are conserved in this protein, including Ser113 and Asn115 which are the major determinants of specificity towards isocitrate[2]. In contrast, the key residues Lys344 and Tyr345 interacting with NADP in the *E.coli* NADP-IDH are replaced by Asp and Ile, as seen in NAD-dependent isopropylmalate dehydrogenase (EC 1.1.1.85)[3,4]. This observation allows us to correct the function of the pro-tein as NAD-isocitrate dehydrogenase (NAD-IDH, 1.1.1.41). We subcloned the coding region of the genomic DNA and expressed the enzyme in an *idh*- strain of *E. coli*. As expected, the enzyme is NAD-dependent. Calculated as the ratio of k_{cat}/K_M, this thermophilic enzyme favors NAD over NADP by a factor of 86.

As a second example, two *Arabidopsis* cDNA clones, NAD-IDH1 and NAD-IDH2 (U81993 and U81994)[5], have been identified by homology searches from EST database. It has been suggested that a single-subunit form of *Arabidopsis* enzyme may exist and that these two clones may represent isozymes. With a careful examination of the active site residues in these two sequences, it is obvious that the Mg^{2+} binding residues equivalent to Arg129, Asp129 and Asp311 found in *E. coli* NADP-IDH are all missing[3,4]. Hence, these gene products should correspond to different regulatory subunits of NAD-IDH, neither of which can form an active enzyme. This is consistent with the observation that the cDNA fails to complement yeast NAD-IDH mutants[5]. It is also noteworthy that the physiologically active form of tobacco NAD-IDH is composed of two regulatory subunits and one catalytic subunit[6].

Our studies demonstrate that correct biochemical function of new genes can be assigned with certainty—an important first step in characterizing their roles in various cellular processes. With the progress of genome-wide efforts to determine representative three-dimensional structures for all protein families, we believe our approach could become more powerful and broadly applicable. Extension of similar studies for other protein families would be much needed in order to take full advantages of an enormous wealth of biological information coming out of the EST and genome projects.

Soon-Seog Jeong and Ridong Chen
Department of Biochemistry, College of Medicine,
University of Saskatchewan
Saskatoon, SK, S7N 5E5, Canada
(ridong.chen@monsanto.com).

1. Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. & Swanson,R.V. *Nature* **392**, 353–358 (1998).
2. Chen, R., Grobler, J.A., Hurley, J.H. & Dean, A.M. *Protein Sci* **5**, 287–295 (1996).
3. Chen, R., Greer, A. & Dean, A.M. *Proc. Natl. Acad. Sci.* **92**, 11666–11670 (1995).
4. Chen, R., Greer, A. & Dean, A.M. *Proc. Natl. Acad. Sci.* **93**, 12171–12176 (1996).
5. Behal, R.H. & Oliver, D. *Plant Mol Biol*. **36,** 691–698 (1998).
6. Lancien, M., Gadal, P. & Hodges, M. **16**, 325–333 (1998).

Arabidopsis thaliana—a new crop?

To the editor:

In an influential paper in 1989, Meyerowitz[1] described *Arabidopsis thaliana* as a "useful weed" and indeed he was right. This weed, endogenous to temperate climes, has been the corner stone of a transformation in plant biology over the past decade and has been widely employed as a model for development and differentiation in plants. In addition, data about its genes and the isolation of cloned cDNAs have facilitated the improvement of crop plants.

Though not widely known, for several years, *Arabidopsis* has been a model for wood and fiber production in trees[2–4]. Repeated clipping of inflorescence stems of the weed induces several common genotypes (and probably many others) to produce much more secondary wood than usual in all plant parts. The increased secondary xylem production of *Arabidopsis* is sufficient not only for developmental studies[2,3] but also for the cloning of xylem-specific genes[4].

Turning weeds into crops is an ancient practice[5]. Now that sequencing of the *Arabisopsis* genome is complete[6], it should be possible to turn *Arabidopsis* from a useful weed into a useful crop. Why ignore its tremendous potential for genetic manipulation when many other crops are much less easy to transform or unresponsive to tissue culture protocols?

For many purposes, when large biomass is not required, this small plant is sufficient. We can make it sweet (with sugar or with sugar substitutions), sour or hot, enrich it with antioxidants, vitamins, pigments, amino acids, or fats, and use it to express many other medicines and natural products.

We already eat or process quite a number of the Brassicaceae, and *Arabidopsis* could join these crops. Turning a model plant into a crop is a test and challenge for modern biology. It is time to start *Arabidopsis* agriculture.

Simcha Lev-Yadun,
Department of Biology,
University of Haifa
at Oranim,
Tivon 36006,
Israel
(Levyadun@research.haifa.ac.il).

1. Meyerowitz E.M. *Cell* **56,** 263–269 (1989).
2. Lev-Yadun, S. *J. Exp. Bot.* **45,** 1845–1849 (1994).
3. Lev-Yadun, S. *Ann. Bot.* **80,** 125–129 (1997).
4. Zhao, C. et al. *Plant Physiol.* **123,** 1185–1196 (2000).
5. Zohary, D. & Hopf, M. *Domestication of plants in the Old World*. (Clarendon Press, Oxford, 2000).
6. The *Arabidopsis* genome initiative. *Nature* **408,** 796–815 (2000).