Wen-Hsiung Li

James Watson Professor, Department of Ecology and Evolution, University of Chicago

2003 Balzan Prize for Genetics and Evolution

Wen-Hsiung Li has made seminal contributions to the field of evolutionary molecular genetics. He has developed widely used methods for inferring phylogenetic relationships and has made important discoveries about the rate of genetic change in different groups of animals.

Institution Administering Funds: The University of Chicago

Adviser for the Balzan General Prize Committee: John Krebs

Evolution of Gene Regulation and Regulatory Modules in Yeast

The development and the physiology of an organism are controlled by genes. For this purpose a gene must be turned on or off at the right time and under the right conditions, and when it is on, the level of its expression must be appropriate; otherwise, the organism can become sick or even die. The turn-on and -off and the level of expression of a gene are called gene regulation. Thus, one can imagine that evolutionary change in gene regulation (in short, regulatory evolution) might be important for the morphological or physiological differences between organisms. However, although this idea has existed since the 1960s, the subject is still not well studied because of experimental difficulties. Recent advances in molecular biology and genomics have allowed fruitful investigations of this subject. These advances notwithstanding, it is still not simple to study higher organisms. He has therefore chosen the budding yeast as the model organism for this purpose because its genetics and molecular biology are well understood and it is experimentally much easier to manipulate than are any higher organisms.

The purpose of the project was to study how the regulation of yeast genes have evolved over time. Also, instead of looking at one gene at a time, the aim was to look at a group of genes that are subject to the same or similar regulation at the same time. Such a group of genes is called a regulatory module.

Researchers:
Y.-W. Chang
F.-G. R. Liu
Elizabeth Marland
Anuphap Prachumwat
H.-M. Sung

Publications (and major results):
1. Marland, E., A. Prachumwat, N. Maltsev, Z. Gu and W.-H. Li. (2004) *Higher gene duplicabilities for metabolic proteins than for non-metabolic proteins in yeast and E. coli*. J. Mol. Evol. 59:806-814.
Gene duplication produces an extra copy that may be free to evolve in function. Therefore, gene duplication is the primary source of genetic novelties. Because in the yeast and the bacterium *E. coli*, the expression level of genes whose products (usually enzymes and proteins) are required for metabolism may often be high, it would be advantageous for them to have extra copies, that is, to have duplicate genes, we examined whether this view is supported by DNA sequence data of the yeast and *E. coli*. We indeed found strong support for this view. In fact, we found that metabolic proteins tend to have higher gene duplicability than non-metabolic proteins. Moreover, a detailed analysis of metabolic pathways in these two organisms revealed that genes in the central metabolic pathways and the catabolic pathways have, on average, higher gene duplicability than do other genes.

2. Prachumwat, A. and W.-H. Li. (2006) *Protein Function, Connectivity, and Duplicability in Yeast*. Mol. Biol. Evol. 23:30-39.
Protein-protein interaction networks have evolved mainly through connectivity rewiring and gene duplication. However, how protein function influences these processes and how a network grows in time have not been well studied. Using protein-protein interaction data and genomic data from the budding yeast it was first examined whether there is a correlation between the age and connectivity of yeast proteins. A steady increase in connectivity with protein age was observed for yeast proteins except for those that could be traced back to bacteria. Second, it was investigated whether protein connectivity and duplicability vary with gene function. Results showed a higher average gene duplicability for proteins interacting with external environments than for proteins localized within intracellular compartments. For example, proteins that function in the cell periphery (mainly transporters) show a high duplicability but are lowly connected. Conversely, proteins that function within the nucleus (e.g.

transcription, RNA and DNA metabolisms, and ribosome biogenesis and assembly) are highly connected but have a low duplicability. Finally, a negative correlation between protein connectivity and duplicability was demonstrated.

3. Chang, Y.-W., F.-G. R. Liu, N. Yu, H.-M. Sung, P. Yang, D. Wang, C.-J. Huang, M.-C. Shih, and W.-H. Li (2008) *Roles of* cis- *and* trans-*changes in the regulatory evolution of genes in the gluconeogenic pathway in yeast*. Mol. Biol. Evol. 25: 1863-1875.

The yeast *Saccharomyces cerevisiae* proliferates rapidly in glucose-containing media. As glucose is getting depleted, yeast cells enter the transition from fermentative to non-fermentative metabolism, known as the diauxic shift, which is associated with major changes in gene expression. To understand the expression evolution of genes involved in the diauxic shift and in non-fermentative metabolism within species, a laboratory strain (BY), a wild strain (RM), and a clinical isolate (YJM) were used in this study. Data showed that the RM strain enters into the diauxic shift ~1 hour earlier than the BY strain with an earlier, higher induction of many key transcription factors (TFs) involved in the diauxic shift. Sequence data revealed sequence variations between BY and RM in both coding and promoter regions of the majority of these TFs. The key TF Cat8p, a zinc-finger cluster protein, is required for the expression of many genes in gluconeogenesis under non-fermentative growth and its derepression is mediated by deactivation of Mig1p. The kinetic study of CAT8 expression revealed that CAT8 induction corresponded to the timing of glucose depletion in both BY and RM and CAT8 was induced up to 50-90 folds in RM, whereas only 20-30 folds in BY. In order to decipher the relative importance of *cis*- and *trans*-variations in expression divergence in the gluconeogenic pathway during the diauxic shift. Studies on the expression levels of MIG1, CAT8, and their downstream target genes in the co-cultures and in the hybrid diploids of BY-RM, BY-YJM, and RM-YJM, and in strains with swapped promoters, were carried out. Data showed that the differences between BY and RM in the expression of MIG1, the upstream regulator of CAT8, were affected mainly by changes in *cis* elements, though also by changes in *trans*-acting factors, whereas those of CAT8 and its downstream target genes were predominantly affected by changes in *trans*-acting factors.

In addition to the evolution of yeast regulatory modules, Anuphap Prachumwat, a graduate student, had studied the origins of vertebrate genes and paper was published in “Genome Research”, the leading journal in genomics:

Prachumwat, A. and W.-H. Li. (2008) *Gene number expansion and contraction in vertebrate genomes with respect to invertebrate genomes*, "Genome Research" 18: 221-232.
Where did vertebrate genes come from? This question was addressed by analyzing eight completely-sequenced land vertebrate genomes and six completely-sequenced invertebrate genomes. Approximately 70% of the vertebrate genes can be found in the six invertebrate genomes with the standard homology search criteria (denoted as *V.MCL*), another ~6% can be found with relaxed search criteria, and an additional ~2% can be found in sequenced fungal and bacterial genomes. Thus, a substantial proportion of vertebrate genes (~22%) cannot be found in the non-vertebrate genomes studied (denoted as *Vonly*). Interestingly, genes in *Vonly* are predominantly singletons, while the majority of genes in the other three groups belong to gene families. The proteins of *Vonly* tend to evolve faster than those of *V.MCL*. Surprisingly, in many cases the family sizes in *V.MCL* are only as large as or even smaller than their counterparts in the invertebrates, contrary to the general perception of a larger family size in vertebrates. Interestingly, in comparison with the family size in invertebrates, vertebrate gene families involved in regulation, signal transduction, transcription, protein transport, and protein modification tend to be expanded, whereas those involved in metabolic processes tend to be contracted. Furthermore, for almost all of the functional categories with family-size expansion in vertebrates, the number of gene types (i.e. the number of singletons plus the number of gene families) tends to be overrepresented in *Vonly* but underrepresented in *V.MCL*. The study suggests that gene function is a major determinant of gene family size.

Note:
In pursuant to the intention of the Balzan research project, which is to cultivate a new generation of scholars, most of the researchers involved in the above studies were all graduate students or postdoctoral fellows. Anuphap Prachumwat, then a graduate student, has gone on to pursue postdoctoral research at the Genomics Research Center, Academia Sinica, Taiwan. Dr. Elizabeth Marland, then a postdoc, has later became a research scientist at Argonne National Laboratory, Illinois. Dr. Y.-W. Chang, then a postdoctoral fellow, has become an assistant professor at National Taiwan University Medical School, Taipei, Taiwan. Dr. F.-G. R. Liu, who was a postdoctoral fellow, has become an assistant professor at National Central University, Taiwan. Finally, Dr. H.-M. Sung, who was a postdoctoral fellow, has become an assistant professor at National Cheng-Kung University, Taiwan. Thus, most of the young scholars involved in the project have continued to pursue scientific research and are now faculty members at prestigious universities or research institutes.