

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Many yeasts win the vote

ArticleInfo		
ArticleID	:	4775
ArticleDOI	:	10.1186/gb-spotlight-20030515-01
ArticleCitationID	:	spotlight-20030515-01
ArticleSequenceNumber	:	127
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	2
ArticleHistory	:	RegistrationDate : 2003-5-15 OnlineDate : 2003-5-15
ArticleCopyright	:	BioMed Central Ltd2003
ArticleGrants	:	
ArticleContext	:	130594411

The [Human Genome Mapping Project](#) set out to unravel the secrets of the genes by determining the primary sequence of the human genome, but it has become clear that this information is insufficient. [Determination](#) of functional and coding sequences in a primary genome sequence depends on an *a priori* knowledge of gene function and on [statistics](#), and so the information obtained is incomplete and probabilistic. In the May 15 [Nature](#), Manolis Kellis and colleagues at the [Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research](#) develop and apply a general approach to determining regions of significance in primary sequence by whole genome comparison of several related species. They reasoned that evolution would conserve protein coding and regulatory elements and that comparison of more than two genomes would increase the signal:noise ratio by highlighting changes that were not due to chance (*Nature* 2003, **423**:241-254).

Kellis et al. compared the sequences of four related species of yeast, *Saccharomyces cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus* and employed a "voting system" to reach a conclusion on the validity of theoretical open reading frames (ORFs) and on the accuracy of the determination of proposed gene structures such as promoters, translation start and stop sites, and intron/exon boundaries. They propose to reduce the number of genes in the yeast gene catalogue by eliminating 503 invalid ORFs and to redefine gene structure assignments in at least 300 cases. They identified 188 genes that encode small proteins of <100 amino acids and many new genes and regulatory elements; they were also able to infer functions for more than half of their 42 newly discovered sequence motifs by categorizing the genes associated with them. In addition, they found evidence for rapid genome evolution at all of the telomeres.

"The analyses will produce a substantial revision in our knowledge of the yeast genome and provide strategic directions for how we might select other sequencing targets to advance understanding of the human genome," writes Steven Salzberg of [The Institute for Genomic Research](#) in an accompanying News and Views article. "This new study of yeast genomes makes it clear that comparative genome sequencing has tremendous analytical power," he concludes.

## References

1. Human Genome Project, [<http://www.genome.gov/>]
2. Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*.
3. Greedy mixture learning for multiple motif discovery in biological sequences.
4. *Nature*, [<http://www.nature.com>]
5. Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research, [<http://www-genome.wi.mit.edu/>]
6. The Institute for Genomic Research, [<http://www.tigr.org>]