

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

## A dictionary for genomes

ArticleInfo		
ArticleID	:	3754
ArticleDOI	:	10.1186/gb-spotlight-20000830-03
ArticleCitationID	:	spotlight-20000830-03
ArticleSequenceNumber	:	191
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	2
ArticleHistory	:	RegistrationDate : 2000-08-30 OnlineDate : 2000-08-30
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	
ArticleContext	:	130591111

With sequence information in hand, the search for regulatory sites in promoters can be done by computers rather than cloning. But the primary tools for analysis, [multiple-alignment algorithms](#), can only handle a small amount of sequence data. In the August 29 [Proceedings of the National Academy of Sciences](#), Bussemaker *et al.* introduce an alternative algorithm that they dub 'MobyDick' (*Proc Nat Acad Sci USA* 2000, **97**: 10096-10100). MobyDick treats DNA sequence as text in which allthewordshavebeenruntogether. It attempts to build a dictionary of 'words' by first finding over-represented pairs of letters. Letter frequency is used to determine the probability that the pairs exist thanks to chance, and this helps determine how larger fragments continue to be built. Bussemaker *et al.* test their algorithm on a space-less version of the first ten chapters of the novel *Moby Dick*, then attack a list of all of the upstream regions in the yeast genome. For yeast, approximately 500 dictionary entries fall above a plausible significance level, including 114 of the 443 [experimentally confirmed](#) sites, and good matches to approximately half of the motifs found in previous analyses of [co-regulated genes](#), the [cell cycle](#), and [sporulation](#).

## References

1. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment.
2. Proceedings of the National Academy of Sciences, [<http://www.pnas.org/>]
3. The Promoter Database of *Saccharomyces cerevisiae*, [<http://cgsigma.cshl.org/jian>]
4. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies.
5. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.
6. The transcriptional program of sporulation in budding yeast.