

CHEN, Xin

Assistant Professor

Division of Mathematical Sciences
School of Physical and Mathematical Sciences
B.S. in Mathematics, Peking University
Ph.D. in Mathematics, Peking University

Major Research Interest: **Computational Biology and Bioinformatics**
Other Interests: **Data Compression and Algorithm Design**

Email: chenxin@ntu.edu.sg

<http://www.ntu.edu.sg/home/chenxin/>

Tel: (65) 6790 3731



Orthology assignment via genome rearrangement

Orthologous genes are typically evolutionary and functional counterparts in different species. Many existing methods that assign orthologs based on the similarity between DNA/protein sequences may make erroneous assignments when sequence similarity does not clearly delineate the evolutionary relationship among genes of the same families. We proposed a novel approach for assigning orthology by taking into account both local base mutations and global rearrangement events. Specifically, a most parsimonious rearrangement scenario that transforms one genome to the other should suggest orthology relationships in a straightforward way. Following this approach, we implemented a high-throughput system for assigning orthologs on a genome scale, called SOAR, and identified potential orthologs between human, mouse, and rat.

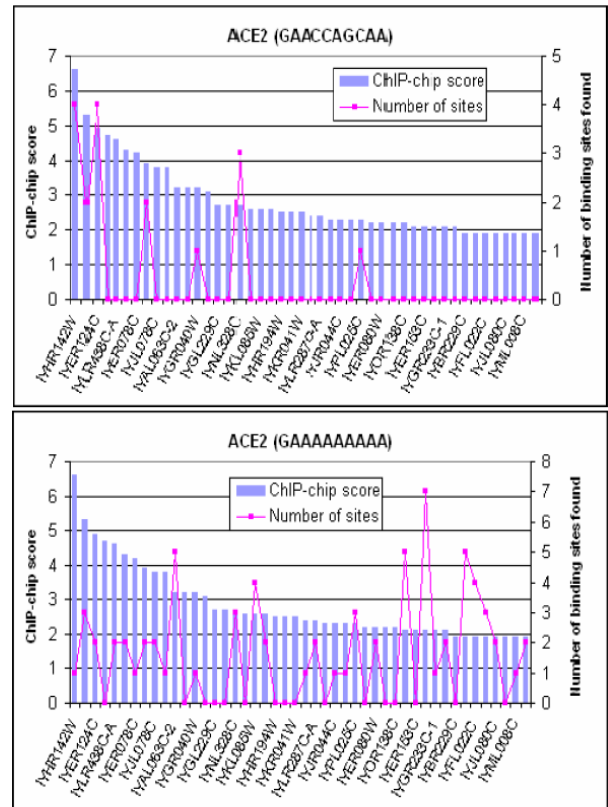
An improved Gibbs sampling method for motif finding

The discovery of motifs in DNA sequences remains a fundamental and challenging problem in computational molecular biology and regulatory genomics, although a large number of computational methods have been proposed in the past decade. Among these methods, the Gibbs sampling strategy has shown great promise and is routinely used for finding regulatory motif elements in the promoter regions of co-expressed genes. We present an enhancement to the Gibbs sampling method when the expression data of the concerned genes is given. A sequence weighting scheme is proposed by explicitly taking gene expression variation into account in Gibbs sampling to find position weight matrices with high specificity. We have incorporated the new approach into the popular motif finding program AlignACE, and our modified program is called W-AlignACE. Large-scale tests demonstrate that W-AlignACE is an effective tool for discovering TF binding sites from gene expression or ChIP-chip data and, in particular, has the ability to find very weak motifs.

Selected Recent Publications:

X. Chen, J. Zheng, Z. Fu, P. Nan, Y. Zhang, S. Lonardi, and T. Jiang. Assignment of orthologous genes via genome rearrangement. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 4, pp. 302-315, 2005.

Z. Fu, **X. Chen**, V. Vacic, P. Nan, Y. Zhong, and T. Jiang. A parsimony approach to genome-wide ortholog assignment. *RECOMB'06*, 578-594 (Joint First Author).



Distributions of ChIP-chip scores and occurrences of putative TF binding sites. The top figure depicts the distribution for a spurious motif ranked the first by AlignACE, and the bottom figure corresponds to the true STE12 motif, which is ranked the first by our program W-AlignACE. We can see that the correct motif occurs more frequently in promoter sequences with higher ChIP-chip scores. However, occurrences of a spurious motif are not expected to have any position correlation with ChIP-chip scores.

X. Chen, T. Jiang. An improved Gibbs sampling method for motif discovery via sequence weighting. *CSB'2006*, pp. 239-247.

X. Chen, L. Guo, Z. Fan, and T. Jiang. Learning position weight matrices from sequence and expression data. *In submission*, 2007.