Regulatory Sequence Analysis

Transcription factor databases

Jacques.van.Helden@ulb.ac.be

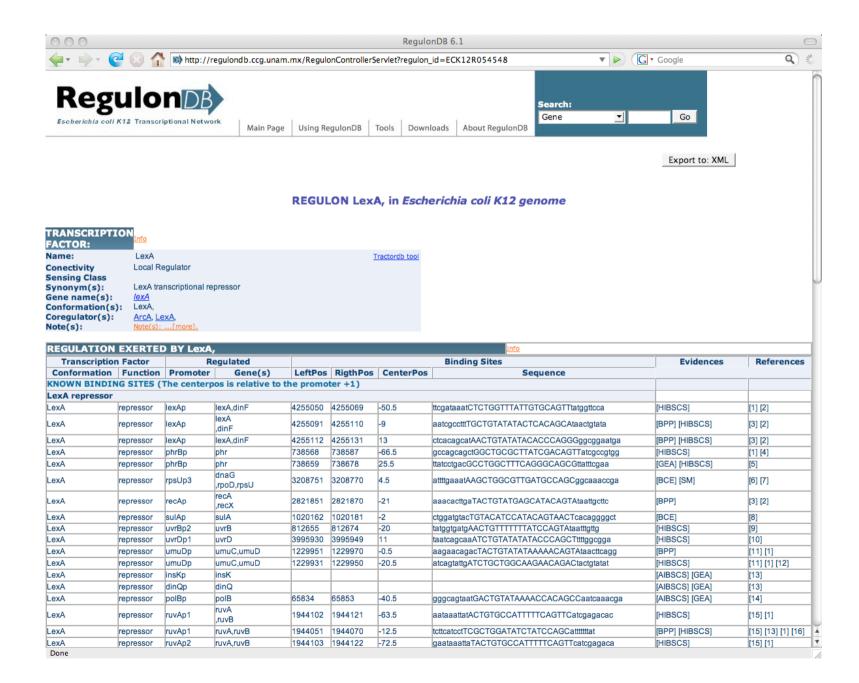
Université Libre de Bruxelles, Belgique Laboratoire de Bioinformatique des Génomes et des Réseaux (BiGRe) http://www.bigre.ulb.ac.be/

RegulonDB Transcriptional regulation in Escherichia coli

- RegulonDB Web site
 - http://regulondb.ccg.unam.mx/
- Model organism: Escherichia coli
- Data content
 - Transcription factors
 - Transcription factor binding sites (TFBS)
 - Position-specific scoring matrices (PSSM)
 - Promoters
 - Operons
- Collaboration with EcoCyc
 - EcoCyc is the reference database about metabolism in Escherichia coli
 - RegulonDB is integrated in the EcoCyc database



Example of regulon in RegulonDB



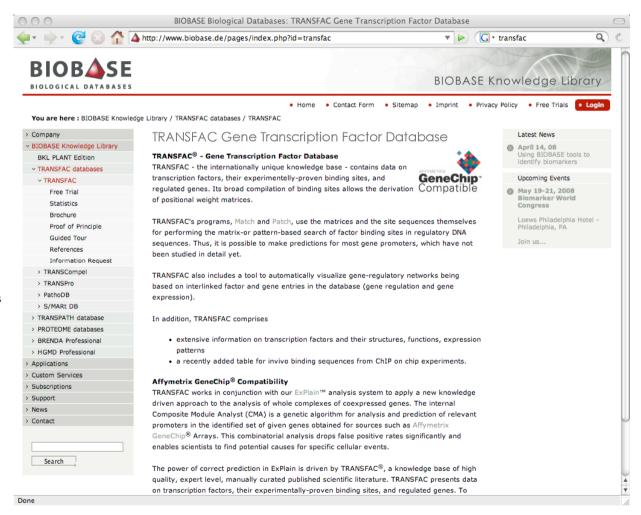
PSSM in RegulonDB

- RegulonDB contains a collection of PSSM built by aligning annotated binding sites.
 - http://regulondb.ccg.unam.mx/data/Matrix_AlignmentSet.txt
- This collection can be used to scan genomes and predict new TFBS.

Transcription Factor Name Total of uniq binding sites																					
																					Ma+
Mat A	12	0	0	0	1	12	1	12	6	10	7	13	4	12	0	23	0	1	12	6	11
C		22	0	0		3		2		5		2		7		0			2		3
G		0				3					0			2		0		1	3		1
T	3	_	23	0	14				15			5				0	0	13	6	13	8
•	3	-	23	Ū	17	3	13	,	13	Ü		3	12	2	Ū	Ū	Ü	13	·	15	Ü
Ali	gnme	nt	5	Score	9																
ACTGTATAAAACCACAGCCAA						12	.05														
GCTGCGCTTATCGACAGTTAT						8.	48														
CCTGGCTTTCAGGGCAGCGTT						7.	51														
ACTGTTTTTTTATCCAGTATA						16	.18														
ATTGGCTGTTTATACAGTATT					12	12.01															
CCTGTTAATCCATACAGCAAC					10	10.7															
ACTGTACATCCATACAGTAAC					14	14.66															
TCTGCTGGCAAGAACAGACTA					3.	36															
ACTGTATATAAAAACAGTATA					17	17.23															
GCTGGATATCTATCCAGCATT					15	15.55															
GCTGGATATCTATCCAGCATT					15	.55															
ACTGTGCCATTTTTCAGTTCA						8.	61														
ACTGTGCCATTTTTCAGTTCA						8.	61														
ACTGTATATAAAACCAGTTTA						16	.16														
ACTGTACACAATAACAGTAAT						12	.47														
ACTGTATGAGCATACAGTATA						14	.73														
GCTGGCGTTGATGCCAGCGGC						4.	27														
ACTGTTTATTATACAGTAAA						16	.67														
TCTGTATATATACCCAGCTTT						14.73															
TCTGGTTTATTGTGCAGTTTA						9.	97														
GCTGTATATACTCACAGCATA						15.05															
ACTGTATATACACCCAGGGGG						9.28															
CCT	GAAT	GAAT	ATAC	AGTA	TT		12	. 9													

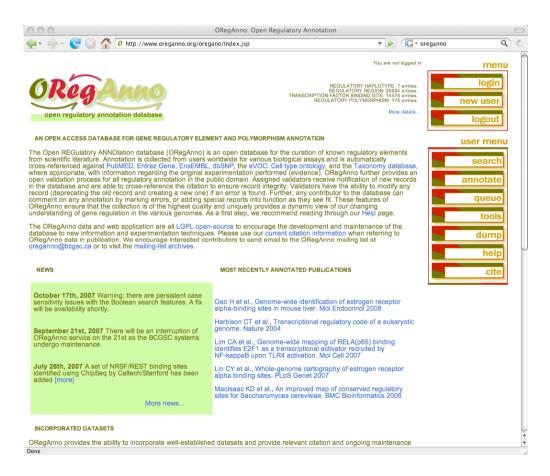
TRANSFAC - Gene transcription factor database

- Organisms
 - Eukaryotes
 - Particular emphasis on mammals (specially human, mouse, rat)
- Distribution
 - The public version is not updated anymore
 - Commercial version (TRANSFAC PRO)
 - Distributed by BioBaseTM
 - http://www.biobase.de/
- Data content
 - Transcription factors
 - Binding sites
 - Fy idences
 - Publications
 - Position-specific scoring matrices
 - Pattern matching tools (patch, match)



ORegAnno

- All organisms (with specific focus on metazoan)
- Web site
 - http://www.oreganno.org/oregano/Index.jsp
 - Also available from the UCSC genome browser
 - http://genome.ucsc.edu/
- Community-based annotation (Jamboree)
- Data content
 - Transcription factor binding sites
 - Mapping on the genomes



Other databases

- YeasTract http://www.yeastract.com/
 - Yeast-specific database. Factors, binding sites and motifs + tools.
- JASPAR http://jaspar.cgb.ki.se/
 - Essentially PSSMs for vertebrates
- FlyReg http://www.flyreg.org/
 - Drosophila DNase I Footprint Database
- PlantCARE http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
 - Plant Cis-Acting Regulatory Elements