

Григорьева Мария

Примечание:

[ChIPMunk](#) выдавал ошибку на длину мотива 27, так как максимальное значение равно 22

На сайте бактериальной базы мотивов [RegPrecise](#) была выбрана группа [Micrococccineae](#).

Phylum/class	Taxonomic group	Reconstructed			Studied
		TF regulogs (regulons)	RNA regulogs (regulons)	Total regulogs (regulons)	Genomes
Actinobacteria	Bifidobacteriaceae	68 (299)	2 (18)	70 (317)	10
Actinobacteria	Corynebacteriaceae	59 (262)	13 (56)	72 (318)	8
Actinobacteria	Frankineae/Propionibacterineae/Pseudonocardiceae	29 (123)	0 (0)	29 (123)	9
Actinobacteria	Micrococccineae	40 (220)	0 (0)	40 (220)	14
Actinobacteria	Mycobacteriaceae	30 (154)	12 (80)	42 (234)	9
Actinobacteria	Nocardiceae	35 (114)	0 (0)	35 (114)	4

Далее были выбраны четыре регулятора:

Имя регулятора	Zur	PaaR	LexA	NrdR
Количество сайтов	35	31	50	23

Family	Type	Regulator	Arthroact	Arthroact	Arthroact	Beutenbier	Brachybact	Brevibacter	Clavibacter	Ianibacter	Ionisia dei	Kocuria rhi	Kvrococcus	Leifsonia x	Remibacter	Trophenavr	Function
ArgR	TF	ArgR	1	1	1	1	1	2		2	1	2	1		1		Arginine biosynthesis
ArsR	TF	SutR			1		1	1			1						Sulfur metabolism
CsoR	TF	CsoR	3	2	2	3	2	1		1		1	2				Copper resistance
DtxR	TF	MntR	3	1	2	1	1	1			1	1		1	1		Manganese homeostasis
FUR	TF	Zur	4	3	3	2	2	5	5	1	2	1	1	2	4		Zinc homeostasis
GntR/MocR	TF	Bcav 0044				1											Metabolite transport
	TF	Bcav 0410				1											Pyridoxine/pyridoxal homeostasis;
	TF	PdxQ	2	2	2					2							Metabolite transport
	TF	PdxR						1			1	1					Pyridoxine biosynthesis
	TF	PdxR2					2						2				Pyridoxine biosynthesis
	TF	PdxR2(2)								1							Pyridoxine/pyridoxal homeostasis
	TF	PdxR3	1	1	1	1		1	1	1							Pyridoxine/pyridoxal homeostasis
GntR/Others	TF	AAur 0050	2	2	3			1		2		1	1				Multidrug resistance; Multidrug efflux
	TF	AAur 2567	1	1		1	1	1			1	1	1		1		Aspartate utilization; Asparagine degradation
	TF	AnsR	2	1	2												Multidrug efflux
	TF	Arth 3541	2	4	4				4								N-acetylglucosamine utilization
	TF	Bcav 0107				1			1								Inositol utilization
	TF	DasR	5	2	5	2											Lactate utilization
	TF	IolR	7	6	5	4			4	2				2	4		Antibiotic resistance
	TF	LldR	2		2		2	4			1	2			2		Heat shock response
	TF	LndYR				1				1							Energy metabolism
HrcA	TF	HrcA	3	4	4	4	3	2	4	3	4	4	3	5	3		Leucine biosynthesis
HxlR	TF	QorR		2					1	2		1					Sugar utilization
IclR	TF	LtbR	3	3	3	3	2	3		3		3			3		Galactosides utilization
Lacl	TF	AAur 3503	1		1												Sugar utilization
	TF	AAur 4078	2			2	3				4						Galactosides utilization
	TF	Arth 2426		2	2	1	1			2							Sugar utilization
	TF	BgaR							2		2						Galactosides utilization
	TF	CelR	1	1		1					1						Galactosides utilization
	TF	FruR							2	3							Galactosides utilization
	TF	IolR	2	4													Galactosides utilization
	TF	MsmR	2		2		1										Galactosides utilization
	TF	RhaR	4	5	7	1	4										Rhamnose utilization
LexA	TF	LexA	5	5	5	4	5	4	4	3	4	3		4	4		SOS response
NrdR	TF	NrdR	2	2	2	2	2	1	1	1	1	2	2	1	3	1	Deoxyribonucleotide biosynthesis
NrtR	TF	NrtR	3	1	1	1	1					1					NAD biosynthesis
Rex	TF	Rex	5	5	5			3		3					4		Energy metabolism
TetR	TF	AmtR	2		2				1								Nitrogen metabolism
	TF	BioQ					3	2			2		2	2			Biotin biosynthesis
	TF	PaaR	5	7	7			4			5				3		Phenylacetic acid degradation

Для каждого из выбранных регуляторов была найдена длина мотива:

Regulog Zur - Micrococcineae

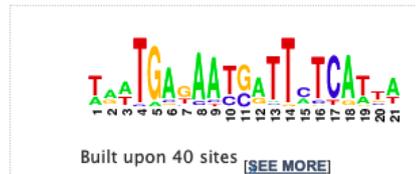
Properties

Regulator type: Transcription factor
 Regulator family: FUR
 Regulation mode: repressor
 Biological process: Zinc homeostasis
 Effector: Zinc ion, (Zn²⁺)
 Phylum: Actinobacteria

Visualization:



Allows to visualize regulog content in the context of metabolic pathways



Regulog PaaR - Micrococcineae

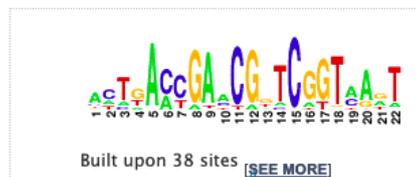
Properties

Regulator type: Transcription factor
 Regulator family: TetR
 Regulation mode: repressor
 Biological process: Phenylacetic acid degradation
 Effector: Phenylacetyl-CoA
 Phylum: Actinobacteria

Visualization:



Allows to visualize regulog content in the context of metabolic pathways



Regulog LexA - Micrococcineae

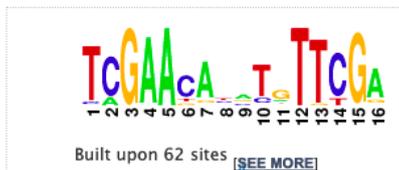
Properties

Regulator type: Transcription factor
 Regulator family: LexA
 Regulation mode: repressor
 Biological process: SOS response
 Effector: DNA damage
 Phylum: Actinobacteria

Visualization:



Allows to visualize regulog content in the context of metabolic pathways



Regulog NrdR - Micrococcineae

Properties

Regulator type: Transcription factor
 Regulator family: NrdR
 Regulation mode: repressor
 Biological process: Deoxyribonucleotide biosynthesis
 Effector: Deoxyribonucleotides
 Phylum: Actinobacteria

Visualization:



Allows to visualize regulog content in the context of metabolic pathways



С помощью инструмента [ChIPMunk](#) я искала мотив найденной длины во всех известных мне последовательностях. Полученные данные были записаны в таблицу ниже.

Затем был выполнен поиск мотивов в хромосоме 14 (привожу список командных строк в системе linux):

```
java -jar sarus.jar chr4.fa aaaaaaajib_motif.pwm 11.74123 > log_jib.txt
wc log_jib.txt
43 127 1283 log_jib.txt
```

```
java -jar sarus.jar chr4.fa aaaaaaajia_motif.pwm 9.86845 > log_jia.txt
wc log_jia.txt
332 994 9956 log_jia.txt
```

```
java -jar sarus.jar chr4.fa aaaaaaajhz_motif.pwm 5.51548 > log_jhz.txt
wc log_jhz.txt
77608 232822 2284742 log_jhz.txt
```

```
java -jar sarus.jar chr4.fa aaaaaaajhv_motif.pwm 9.18885 > log_jhv.txt
wc log_jhv.txt
2624 7870 77892 log_jhv.txt
```

Первая цифра (-l значение) поля вывода — число находок + 1.

Результаты в таблице:

TF name	motif length	ChIPMunk ticket	KDIC	Threshold	Number of hits
<u>Zur</u>	21	aaaaaaajhv	0.5356	9.18885	2623
<u>PaaR</u>	22	aaaaaaajib	0.63977	11.74123	42
<u>LexA</u>	16	aaaaaaajia	0.68866	9.86845	331
<u>NrdR</u>	16	aaaaaaajhz	0.58907	5.51548	77687

Значение информационного содержания (KDIC) равное ~ 0.5 считается нормальным (значение KDIC должно входить в промежуток $[0,5:1]$). Чем выше значение KDIC, тем лучше, следовательно, количество находок должно быть меньше.

В целом, эта корреляция выполняется (особенно в работах у моих однокурсников, но мне, похоже, не очень повезло). Наименьшее число находок получается при $KDIC = 0.63977$ (второе по убыванию значение), а наибольшее — при $KDIC = 0.58907$ (второе по возрастанию). Интересно, что если выстроить все значения в порядке возрастания, то каждое последующее будет отличаться от предыдущего на порядок.

TF name	KDIC*100000	Number of hits
<u>Zur</u>	53560	2623
<u>PaaR</u>	63977	42
<u>LexA</u>	68866	331
<u>NrdR</u>	58907	77687

