

Массовый поиск консервативных регуляторных структур, содержащих Т-боксы: результаты счета

Л.А. Леонтьев, В.А. Любецкий

Институт проблем передачи информации РАН,
127994, Россия, Москва, Большой Каретный переулок, 19,
e-mail: maravan@yandex.ru, lyubetsk@iitp.ru

Поступила в редколлегию 15.12.2005

Аннотация—Приводятся результаты поиска консервативных структур регуляции транскрипции с участием Т-боксов, найденные оригинальным алгоритмом авторов, у многих Грам положительных бактерий. Эти результаты значительно расширяют данные из известной базы Rfam, содержащей структуры РНК. Соответствующие множественные выравнивания приведены в другой публикации авторов.

1. ВВЕДЕНИЕ

Транскрипция многих генов бактерий регулируется Т-боксами — высоко консервативными альтернативными РНКовыми структурами, одна из конформаций которых стабилизируется незагруженной транспортной РНК. Детали механизма можно найти в [1]. Продолжая результаты, приведенные в [2], нами предсказано много новых регуляторных структур с Т-боксами, не содержащихся в базе Rfam. Наш алгоритм учитывает различные элементы РНК структуры, что позволяет снизить число перепредсказаний. Соответствующие множественные выравнивания приведены в [3].

2. МАТЕРИАЛЫ И РЕЗУЛЬТАТЫ

Геномы бактерий получены из базы данных GenBank (NCBI). Набор последовательностей образован 5'-нетранслируемыми областями перед размеченными генами. Краткое описание применяемого алгоритма приведено в [2]. Соответствующая программа разработана одним из авторов (Л. Леонтьевым). Из результатов нашего счета следует, что только для указанных ниже родов характерна Т-блочная регуляция транскрипции многих генов из одного генома. Типичные представители этих родов указаны в следующей таблице 1. Также наш счет показывает, что в ряде других геномов такая регуляция имеет место для отдельных генов, см. ниже.

Результаты работы алгоритма и их сопоставление с данными из базы Rfam приведены в таблице 2. Наибольшее число Т-боксов предсказано для фирмикутов. Большое число транскрипционных Т-боксов у *Symbiobacterium thermophilum* согласуется с гипотезой, что эта бактерия принадлежит к бациллам, а не актинобактериям.

Кроме того, найдены новые Т-боксы у протеобактерий и зеленых бактерий [2]. Хотя многие Т-боксы регулируют экспрессию аминоксил-тРНК синтетаз, Т-боксы часто встречаются в лидерных областях генов синтеза аминокислот. В частности, триптофановый Т-блочек перед генами антранилат синтазы присутствует во многих видах, что позволяет предположить возникновение такой регуляции синтеза триптофана на ранних этапах эволюции. Кроме того, известен случай Т-блочковой регуляции белка TRAP.

С другой стороны, поиск белок-ДНК регуляции генов, перед которыми не обнаружен Т-бокс, не привели к обнаружению предполагаемого белка-регулятора с близким филогенетическим профилем. Это свидетельствует в пользу предположения, что гены фирмикутов, потерявшие Т-боксовую регуляцию, которая была у их общего предка, сформировали новые пути регуляции независимо друг от друга на более поздних этапах эволюции.

Работа частично поддержана грантом ISTC (2766).

BC	<i>Bacillus cereus</i>	NC006274(CP000001)
BH	<i>Bacillus halodurans</i>	NC002570(BA000004)
BS	<i>Bacillus subtilis</i>	NC000964(AL009126)
BT	<i>Bacillus thuringiensis</i>	NC005957(AE017355)
CP	<i>Clostridium perfringens</i>	NC003366(BA000016)
EF	<i>Enterococcus faecalis</i>	NC004668(AE016830)
LL	<i>Lactococcus lactis</i>	NC002662(AE005176)
LP	<i>Lactobacillus plantarum</i>	NC004567(AL935263)
LI	<i>Listeria innocua</i>	NC003212(AL592022)
MM	<i>Mycoplasma mycoides</i>	NC005364(BX293980)
OI	<i>Oceanobacillus iheyensis</i>	NC004193(BA000028)
SA	<i>Staphylococcus aureus</i>	NC002952(BX571856)
StT	<i>Streptococcus thermophilus</i>	NC006449(CP000024)
SyT	<i>Symbiobacterium thermophilum</i>	NC006177(AP006840)
TT	<i>Thermoanaerobacter tengcongensis</i>	NC003869(AE008691)

Таблица 1. Геномы, для которых характерна регуляция с участием Т-боксов. В первой колонке содержатся обозначения видов.

ВИД	BC	BH	BS	BT	CP	EF	LL	LP	LI	MM	OI	SA	StT	SyT	TT
<i>alaS</i>	R+	LR+	LR+	R+	LR+	L+	L+	L+	L+	+	L+	LR+	L+	LR+	+
<i>alsS</i>	-	+	+	LR+	-	-	-	-	+	-	-	+	-	-	-
<i>argS</i>	L+	+	+	L+	L+	LR+	+	L R+	+	+	+	+	+	+	+
<i>aroA</i>	L+	+	+	L+	+	+	+	+	+	-	-	+	+	-	+
<i>aroF</i>	L+	+	+	LR+	-	-	+	+	+	-	+	-	+	-	-
<i>asnA</i>	L+	-	-	L+	LR+	-	-	L+	-	+	-	-	+	-	-
<i>asnS</i>	L+	+	+	L+	+	+	+	+	-	+	+	+	+	+	-
<i>aspS</i>	L+	+	+	L+	+	+	+	+	+	+	+	+	+	+	+
<i>azr</i>	+	-	-	R+	-	-	-	-	-	-	-	-	-	-	-
<i>brnQ</i>	L R+	+	+	LR+	LR+	+	+	LR+	-	-	-	-	+	-	-
<i>cysE</i>	L+	-	L+	L+	+	L+	+	+	L+	-	L+	+	+	-	L+
<i>cysK</i>	+	+	+	+	R+	+	+	+	+	-	+	+	-	-	+
<i>dxr</i>	R+	+	+	R+	-	-	-	-	-	-	-	-	-	+	+
<i>entA</i>	+	-	-	+	R+	-	-	-	-	-	-	-	-	-	-
<i>gcsH</i>	-	-	-	-	-	-	-	L+	-	-	-	-	-	-	-
<i>glnP</i>	+	+	+	+	-	-	+	L+	-	-	-	-	-	-	-
<i>gltX</i>	+	+	+	+	L+	+	+	+	+	+	-	+	+	+	-
<i>glyQ</i>	-	LR+	L+	-	-	L+	-	L+	LR+	-	L+	-	L+	-	-
<i>glyS</i>	L+	+	+	L+	R+	+	+	+	+	+	+	L+	+	+	-
<i>hisS</i>	L+	LR+	LR+	L+	+	+	+	L+	LR+	+	L+	L+	+	+	R+
<i>hisX</i>	-	-	-	-	-	-	-	L+	-	-	-	-	-	-	-
<i>hom</i>	LR+	+	+	L R+	-	+	+	+	+	-	-	-	+	-	-
<i>ileS</i>	LR+	LR+	LR+	L+	+	LR+	L+	L+	L+	LR+	L+	LR+	L+	L+	LR+
<i>ilvB</i>	+	-	L R+	+	-	-	+	-	+	-	+	+	+	-	+
<i>ilvH</i>	+	+	-	+	-	-	-	-	-	-	-	-	-	-	R+
<i>leuS</i>	L+	L R+	L+	L+	L+	L+	+	L+	LR+	+	LR+	L+	+	L+	L+
<i>lysS</i>	L+	+	+	LR+	+	+	+	+	+	+	-	+	+	L+	-
<i>metA</i>	+	+	+	+	-	-	+	LR+	-	-	-	-	+	-	-
<i>metB</i>	-	+	+	-	+	-	+	L+	-	-	-	+	+	-	-
<i>metE</i>	+	+	+	+	-	-	+	L+	-	-	-	+	+	-	-
<i>metG</i>	+	-	-	+	-	+	-	-	-	-	-	-	-	-	L+
<i>metS</i>	+	L+	-	+	+	-	+	+	+	+	-	+	+	+	-
<i>nha</i>	+	+	+	+	-	+	-	LR+	-	-	-	-	+	-	+
<i>nhaC</i>	LR+	+	+	LR+	-	LR+	-	-	-	-	-	-	-	-	+
<i>osmC</i>	-	-	-	-	-	-	R+	-	-	-	-	-	-	-	-
<i>pah</i>	L+	-	-	L+	-	-	-	-	-	-	-	-	-	-	-
<i>panE</i>	+	-	-	+	-	-	+	L+	-	-	-	-	+	-	-
<i>pheS</i>	L+	R+	R+	L+	L+	L+	L+	LR+	LR+	+	-	L+	+	+	L+
<i>plpA</i>	-	-	-	-	-	-	L+	-	-	-	-	-	-	-	-
<i>proB</i>	L+	-	LR+	L+	+	+	+	+	+	-	+	-	+	-	+
<i>proC</i>	R+	-	-	R+	+	-	+	+	-	-	-	-	+	-	+
<i>proI</i>	-	-	L+	-	-	-	-	-	-	-	-	-	-	-	-
<i>proS</i>	+	+	+	+	L+	+	+	+	+	+	-	+	+	+	+
<i>pta</i>	+	+	+	+	+	+	R+	+	+	+	-	+	+	-	+
<i>serS</i>	L+	-	L+	L+	L+	L+	+	LR+	+	+	-	LR+	+	+	L+
<i>thrS</i>	L+	L+	L+	L+	L+	LR+	L+	L+	L+	L+	LR+	L+	L+	L+	+
<i>thrZ</i>	-	-	LR+	-	-	-	-	-	-	-	-	-	-	-	-
<i>trpE</i>	LR+	+	+	LR+	-	-	L+	L+	LR+	-	-	-	+	-	+
<i>trpS</i>	LR+	-	L+	LR+	LR+	+	+	LR+	LR+	+	LR+	+	L+	L+	+
<i>tyrS</i>	L+	-	L+	L+	+	LR-	+	L+	LR+	+	L+	R+	+	L+	+
<i>tyrZ</i>	-	L+	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>valS</i>	LR+	LR+	LR+	LR+	L+	L+	+	LR+	LR+	+	L+	LR+	+	+	L+
<i>ycgJ</i>	-	+	+	-	L+	-	+	-	-	-	-	-	-	-	-
<i>yczA</i>	-	-	LR+	-	-	-	-	-	-	-	-	-	-	-	-
<i>yvbW</i>	-	-	L+	-	-	-	-	-	-	-	-	-	-	-	-

Таблица 2. Результаты нашего алгоритма по массовому поиску регуляторных структур с Т-боксами. Плюс означает наличие гомолога у гена, указанного в строке, в геноме бактерии, указанной в столбце, а минус означает отсутствие такового. Буква R отмечает Т-боксы, содержащиеся в базе Rfam, буквы LR отмечают Т-боксы, предсказания и нашим алгоритмом, а буква L — Т-боксы, впервые предсказанные нашим алгоритмом.

СПИСОК ЛИТЕРАТУРЫ

1. Grundy F.J., Henkin T.M. The T box and S box transcription termination control systems. *Front Biosci*, 2003, 8, d20 – 31.
2. Леонтьев Л.А., Любецкий В.А., Селиверстов А.В., Алгоритм массового поиска опосредованной Т-боксом транскрипционной регуляции у бактерий. *Молекулярная биология*, 2005, 39, No 6, 1076 – 1078.
3. Леонтьев Л.А., Селиверстов А.В., Любецкий В.А. Регуляция транскрипции с участием Т-боксов, Труды конференции РАН “Проблемы управления и моделирования в сложных системах, VII Международная конференция”, 2005, Самара, 222 – 224.