

Московский Государственный Университет им. М.В.Ломоносова
Факультет Биоинженерии и Биоинформатики

**Исследование обобщенного FruR (Cra)-регулона гамма-протеобактерий
методами сравнительной геномики**

Курсовая работа студентки 2-го курса
Закирзяновой В. В.

Научный руководитель: Равчеев Д. А.
Старший научный руководитель: Гельфанд М. С.

Москва, 2006

Введение.

Белок FruR из *Escherichia coli* представляет собой фактор транскрипции, регулирующий экспрессию генов, задействованных в метаболизме углеводов и энергетическом обмене. Первоначально данный белок был открыт как репрессор *fruBKA* оперона, ответственного за импорт и фосфорилирование фруктозы, за что и был назван FruR. Впоследствии же белок получил название Cra (Catabolite Repressor/Activator), которое более согласуется с его функциями.

Белок Cra принадлежит к белковому семейству LacI-GalR. Все белки данного семейства имеют N-концевые ДНК-связывающие домены, несущие характерный для факторов транскрипции мотив спираль-поворот-спираль. С-концевые домены белков данного семейства являются лиганд-связывающими и гомологичны лиганд-связывающим доменам у периплазмических рецепторов сахаров. С-концевые домены большинства белков LacI-GalR семейства связывают углеводы и их производные (Saier, 1996).

Экспериментально установлено, что влияние белка Cra на экспрессию регулируемых генов не зависит от цАМФ, что отличает данный белок от другого регулятора катаболизма сахаров, Crp (Saier, 1996). Физиологические свойства *cra*-мутантов и наблюдаемая гомология С-концевого домена с рецепторами сахаров указывает на то, что взаимодействие Cra с ДНК может контролироваться какими-либо производными сахаров. Действительно, впоследствии было показано, что взаимодействие Cra с микромолярными концентрациями фруктозо-1-фосфата или миллимолярными концентрациями фруктозо-1,6-бифосфата препятствуют связыванию данного белка с ДНК (Milton and Saier, 1996).

В ряде экспериментальных исследований было показано (Saier, 1996), что Cra репрессирует синтез белков фосфо-трансферазных систем, ферментов брожения, гликолиза и пути Энтнера-Дудорова, и активирует синтез ферментов, осуществляющих окисление субстратов, таких как ферменты глюконеогенеза и цикла трикарбоновых кислот.

На основе 11 экспериментально определенных операторов Cra был определен консенсус для его сайта связывания (Рис. 1). Cra-сигнал представляет собой несовершенный палиндром, в котором левая половина значительно более консервативна, чем правая. Высокая степень вырождения, обнаруженная в правой половине сайта, указывает на то, что Cra асимметрично взаимодействует с двумя полусайтами оператора (Negre et al., 1996). Было показано, что в тех случаях, когда сайты связывания располагаются перед промоторами регулируемого оперона, белок Cra действует как

активатор. Если же сайт связывания перекрывается с промотором или располагается за промотором, то экспрессия оперона репрессируется Cra.

Было обнаружено, что Cra может связываться с регуляторной областью оперона *fruBKA* посредством двух операторов, которые расположены между промотором оперона и сайтом инициации трансляции гена *fruB*. Для других регулируемых оперонов показано связывание Cra только с одним сайтом.

Потенциальные Cra-связывающие сайты с достаточно высокой степенью симметрии были найдены перед 15 оперонами, при этом, последовательности Cra-сайтов для активируемых и репрессируемых генов не различаются (Saier et al, 1996).

К настоящему времени к Cra-регулону *E.coli* относятся следующие опероны:

1) активируемые Cra: *ppsA*, *pckA*, *aceBA*, *icdA*, *fdp* и *cydAB* (Saier et al., 1996);

2) репрессируемые Cra: *edd-eda*, *mtlADR*, *ptsHI-crr*, *pfkA*, *pykF*, *epd-pgk*, *fruBKA* (Saier et al, 1996), *nirBCD-cysG* (Tyson et al, 1997) и *adhE* (Mikulskis et al., 1997).

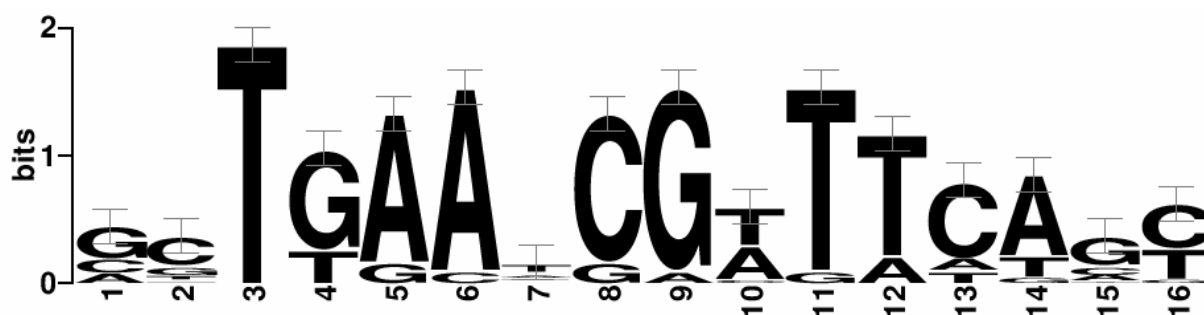


Рис. 1. Диаграмма Лого для сайтов связывания Cra. По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Цели и задачи.

Целью настоящей работы является исследование эволюции обобщенного Cra-регулона гамма-протеобактерий.

Для выполнения данной работы были поставлены следующие задачи:

- Поиск ортологов Cra в геномах гамма-протеобактерий. Данный поиск необходим для того, чтобы определить, в каких геномах будет проводиться дальнейшее исследование.
- Поиск ортологов для генов, входящих в Cra-регулон *E. coli*, и сайтов связывания Cra перед ними.
- Поиск потенциальных новых членов регулона, в частности, анализ таксон-специфичной регуляции.
- Описание эволюции обобщенного регулона.

Материалы и методы.

Исследованные геномы.

В данной работе было рассмотрено 29 геномов гамма-протеобактерий, принадлежащих к шести порядкам. Полные последовательности геномов *Escherichia coli* K12 (Blattner et al., 1997), *Salmonella enterica* serovar Typhi Ty2 (Deng et al., 2003), *Salmonella typhimurium* LT2 (McClelland et al., 2001), *Yersinia pestis* KIM (Deng et al., 2002), *Yersinia pseudotuberculosis* IP 32953 (Chain et al., 2004), *Pectobacterium carotovora* subsp. *atroseptica* SCRI1043 (Bell et al., 2004), *Photobacterium luminescens* subsp. *laumondii* TTO1, *Pasteurella multocida* Pm70 (May et al., 2001), *Haemophilus influenzae* KW20 (Fleischmann et al., 1995), *Haemophilus ducreyi* 35000HP, *Photobacterium profundum* SS9, *Vibrio vulnificus* CMCP6 KW20, *Vibrio parahaemolyticus* RIMD 2210633 (Makino et al., 2003), *Vibrio cholerae* O1 (Heidelberg et al., 2000), *Vibrio fischeri* ES114 (Ruby et al., 2005), *Shewanella oneidensis* MR-1 (Heidelberg et al., 2002), *Pseudomonas putida* KT2440 (Nelson et al., 2002), *Pseudomonas aeruginosa* PAO1 (Stover et al., 2000), *Pseudomonas syringae* pv. *tomato* str. DC3000 (Buell et al., 2003) и *Psychrobacter arcticum* 273-4 были взяты из базы данных GenBank (Benson et al., 1999). Последовательности геномов *Yersinia enterocolitica*, *Serratia marcescens* Db11 и *Pseudomonas fluorescens* SBW25 были взяты с интернет-сайта Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/>). Последовательность генома *Actinobacillus*

actinomycetemcomitans НК1651 была взята с интернет-сайта University of Oklahoma's Advanced Center for Genome Technology (<http://www.genome.ou.edu/>). Последовательность генома *Azotobacter vinelandii* AvOP была взята с интернет-сайта DOE Joint Genome Institute (<http://www.jgi.doe.gov>). Последовательность генома *Pectobacterium chrysanthemi* str. 3937 была взята с интернет-сайта The Institute For Genomic Research (<http://www.tigr.org>).

Сравнительно-геномный подход для исследования обобщенного регулона.

Для определения состава обобщенного Cга-регулона использовался метод проверки соответствия. Согласно данному методу, ген считается принадлежащим обобщенному регулону, если перед ним в нескольких организмах обнаруживается потенциальный сайт связывания. В противном случае, скорее всего, сайт является перепредсказанным и данный ген не рассматривается (Gelfand et al, 2000).

Для поиска потенциальных сайтов в геномных последовательностях использовался метод позиционных весовых матриц. В этом методе на основании множественного выравнивания известных сайтов связывания рассчитывается вес каждого нуклеотида во всех позициях. При этом учитывается разная значимость различных позиций сайта, а также возможность и вероятность замены одного нуклеотида на другой в каждой позиции (Gelfand et al, 2000).

Програмное обеспечение.

Поиск ортологов генов и потенциальных сайтов связывания регулятора производился программой GenomeExplorer (Миронов и др., 2000).

Для построения множественных выравниваний и филогенетических деревьев использовалась программа ClustalX (Thompson et al., 1997).

Для визуализации деревьев использовалась программа GeneMaster (Миронов А.А., неопубл.).

Для построения диаграммы Logo для сайтов связывания Cга использовалась программа WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

Результаты.

1. Поиск ортологов Cra

В геномах организмов из группы гамма-протеобактерий был произведен поиск ортологов для белка Cra из генома *Escherichia coli*. Всего было исследовано 28 геномов, в 19 из них были обнаружены ортологи Cra (Табл.1).

После этого было построено множественное выравнивание аминокислотных последовательностей и филогенетическое дерево (Рис.2) ортологов Cra.

2. Поиск ортологов членов Cra-регулона

В настоящее время регуляция за счет белка Cra показана для 29 генов *E.coli*, объединенных в 15 оперонов (см. Введение). Для всех этих генов был произведен поиск ортологов и потенциальных Cra-сайтов перед последними. Оперон считался потенциально регулируемым, если вес потенциального сайта связывания Cra был не меньше 3,75. Поиск проводился во всех геномах, где были найдены ортологи Cra (Табл.2).

3. Новые члены регулона

Как уже говорилось ранее, белок Cra регулирует в *E.coli* экспрессию ряда генов, кодирующих, ряд ферментов центрального метаболизма. Как видно из Рис.3, Cra-зависимая регуляция была показана не для полных метаболических путей, а лишь для отдельных реакций. Поэтому вполне логичным представляется предположение о том, что гены, кодирующие другие ферменты для этих путей, будут также находиться под Cra-регуляцией.

С целью проверки этой гипотезы был проведен поиск потенциальных сайтов связывания Cra перед генами *pgi*, *glpX*, *talA*, *talB*, *fbaB*, *fbaA*, *tpiA*, *gapA*, *gpmA*, *gpmB*, *gpmC*, *eno*, *pykA*, *pdhR-aceEF-lpdA*, *gltA*, *acnA*, *acnB*, *sucAB*, *sucCD*, *sdhCDAB*, *fumA*, *fumB*, *fumC*, *mdh* в *E.coli*, а также перед их ортологами в других геномах.

В случае, если потенциальные сайты были обнаружены более чем в трех геномах, ген считался принадлежащим к Cra-регулону.

Таким образом была предсказана регуляция четырех оперонов: *tpiA*, *gapA*, *gpmA*, *pdhR-aceEF-lpdA* (Табл.3).

Были рассмотрены выравнивания потенциальных регуляторных областей генов *tpiA*, *gapA*, *gpmA*, *pdhR* для изучения расположения в них Cra-связывающих сайтов и промоторов *E.coli*. По положению сайта относительно промотора в *E.coli* была сделана попытка определения типа регуляции гена.

Табл. 1. Исследованные геномы и ортологи *Cra*. Названия ортологов *Cra* приводятся в соответствии с идентификаторами в геномах.

Геном	Условные обозначения геномов	Ген белка-регулятора
Enterobacteriales		
<i>Escherichia coli</i> K-12	EC	<i>fruR</i>
<i>Serratia marcescens</i> Db11	SM	<i>020_0139</i>
<i>Salmonella typhi</i> Ty2	ST	<i>fruR</i>
<i>Salmonella typhimurium</i> LT2	STM	<i>fruR</i>
<i>Yersinia enterocolitica</i> 8081	YE	<i>001_0554</i>
<i>Yersinia pestis</i> KIM	YP	<i>fruR</i>
<i>Yersinia pseudotuberculosis</i> IP 32953	YPS	<i>fruR</i>
<i>Photorhabdus luminescens subsp. laumondii</i> TTO1	PHL	<i>plu3664</i>
<i>Pectobacterium chrysanthemi</i>	PCH	<i>089_0019</i>
<i>Pectobacterium carotovorum subsp. atrosepticum</i>	PCA	<i>001_3381</i>
Pasteurellales		
<i>Pasteurella multocida</i> PM70	PM	<i>fruR</i>
<i>Haemophilus influenzae</i> KW20	HI	отсутствует
<i>Haemophilus ducreyi</i> 35000HP	HD	отсутствует
<i>Actinobacillus actinomycetemcomitans</i> HK1651	AA	<i>001_1052</i>
Vibrionales		
<i>Vibrio cholerae</i> N16961	VC	<i>VCA0519</i>
<i>Vibrio fischeri</i> ES114	VF	#3670
<i>Vibrio parahaemolyticus</i> O3:K6	VP	<i>VPA0814</i>
<i>Vibrio vulnificus</i> CMCP6	VV	<i>VV20201</i>
<i>Photobacterium profundum</i> SS9	PBP	#1571
Pseudomonadales		
<i>Pseudomonas aeruginosa</i> PAO1	PA	<i>fruR</i>
<i>Pseudomonas putida</i> KT2440	PP	<i>fruR</i>
<i>Pseudomonas fluorescens</i> SBW25	PF	<i>013_0174</i>
<i>Pseudomonas syringae pv. tomato</i> DC3000	PS	<i>PSPTO0953</i>
<i>Azotobacter vinelandii</i> AvOP	AV	отсутствует
<i>Psychrobacter</i> 273-4	PSY	отсутствует
Alteromonadales		
<i>Shewanella oneidensis</i> MR-1	SO	отсутствует
Xanthomonadales		
<i>Xanthomonas axonopodis pv. citri</i> 306	XA	отсутствует
<i>Xanthomonas campestris pv. campestris</i> ATCC33913	XC	отсутствует
<i>Xylella fastidiosa Temecula1</i>	XF	отсутствует

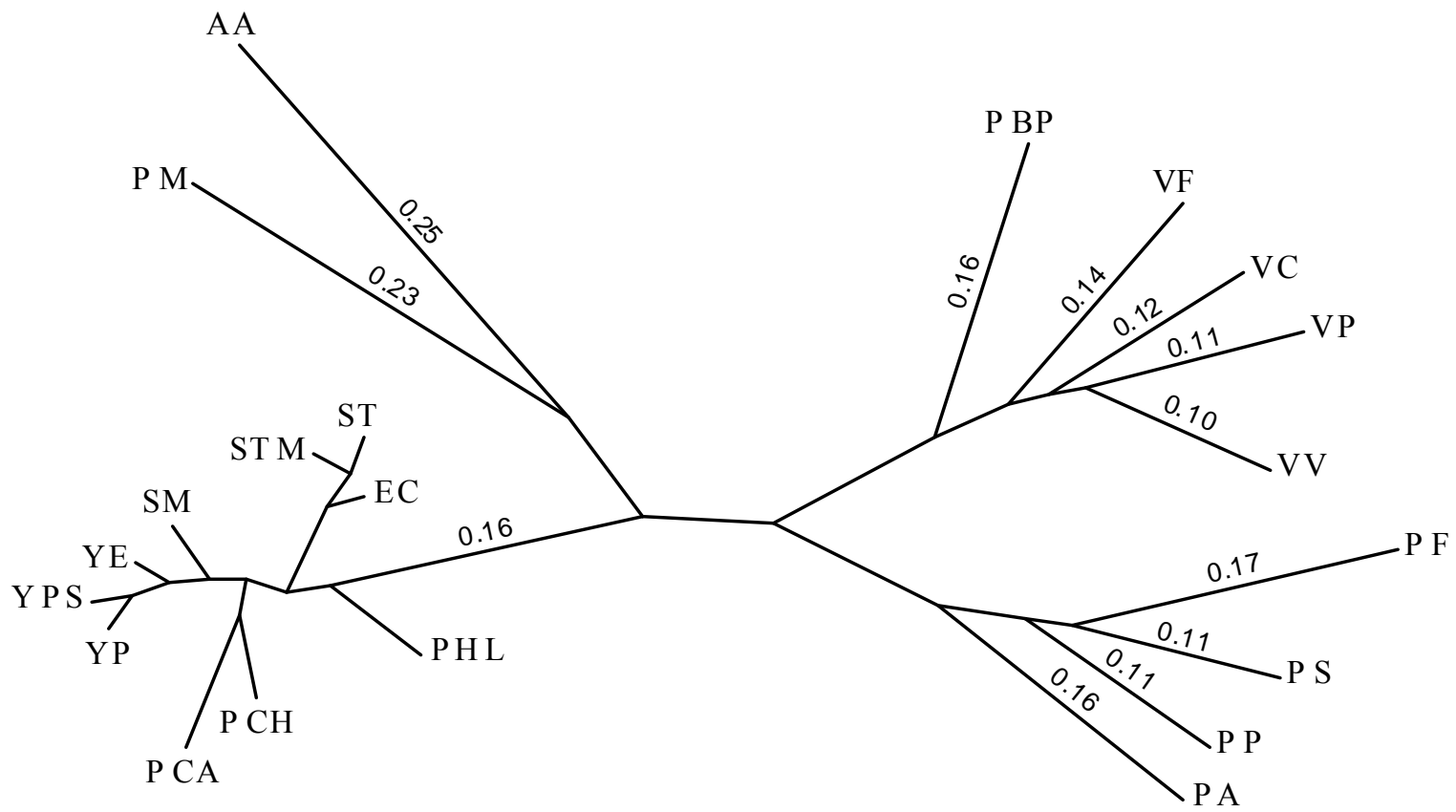


Рис. 2. Филогенетическое дерево для белков Sca из гамма-протеобактерий.

Условные обозначения – как в Табл. 1.

Табл. 2. Ортологи генов, для которых была экспериментально показана регуляция генов в *E.coli*.

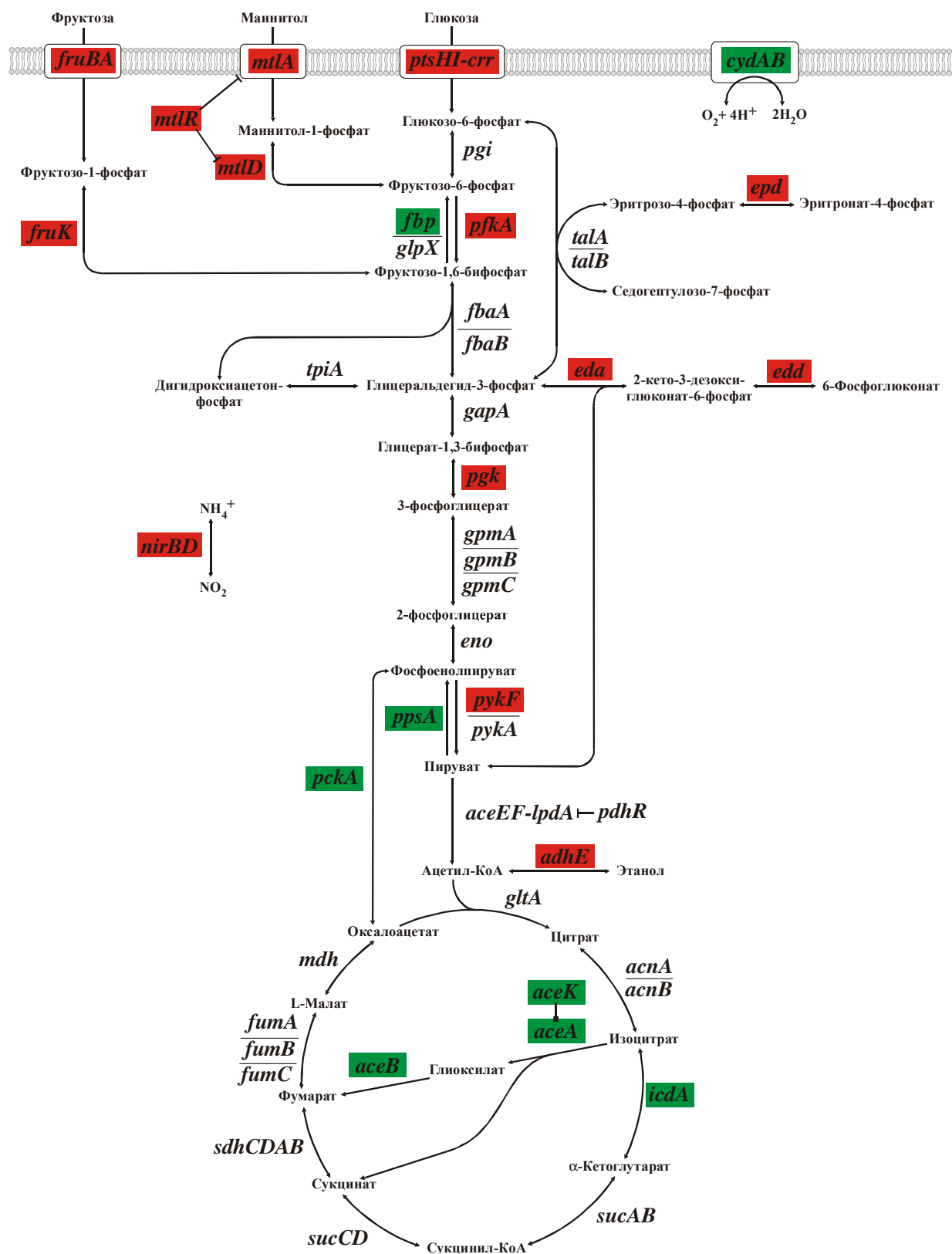
Условные обозначения: те же, что и в таблице 1, полужирным шрифтом обозначены гены, начинающие оперон,

«+» – перед геном расположен потенциальный сайт связывания Cга; «*» – ген находится в регулируемом опероне (но не начинает его);

«-» – экспрессия гена не регулируется Cга; «0» – ортолога для гена не найдено.

Gene	EC	ST	STM	PCA	PCH	YP	YE	YPS	SM	PHL	PM	AA	PBP	VF	VC	VP	VV	PA	PP	PF	PS
fruB	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+
<i>fruK</i>	*	*	*	*	*	*	*	*	*	+	-	-	*	*	*	*	*	*	*	*	*
<i>fruA</i>	*	*	*	*	*	*	*	*	*	*	-	-	*	*	*	*	*	*	*	*	*
epd	+	+	+	+	+	+	+	+	+	+	0	0	-	-	-	-	-	-	-	-	-
<i>pgk</i>	*	*	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	-
edd	+	+	+	+	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-
<i>eda</i>	*	*	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
nirB	+	+	+	-	-	+	+	+	+	0	0	0	-	-	0	-	-	-	-	-	-
<i>nirD</i>	*	*	*	-	-	*	*	*	*	0	0	0	-	-	0	-	-	-	-	0	-
<i>nirC</i>	*	*	*	0	0	*	*	*	*	0	0	0	-	-	0	-	-	0	0	0	0
<i>cysG</i>	*	*	*	-	-	*	*	-	*	0	0	0	-	-	-	-	-	-	0	-	-
mtlA	+	+	+	-	-	-	-	-	-	0	-	0	-	0	-	-	-	0	0	0	0
<i>mtlD</i>	*	*	*	-	-	-	-	-	-	0	-	0	-	0	-	-	-	0	0	0	0
<i>mtlR</i>	*	*	*	-	-	-	-	-	-	0	-	0	-	0	-	-	-	0	0	0	0
ptsH	+	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-	0	0	-	0
<i>ptsI</i>	*	*	*	-	-	*	*	*	*	-	-	-	-	-	-	-	-	0	0	0	0
<i>crr</i>	*	*	*	-	-	*	*	*	*	-	-	-	-	-	-	-	-	0	0	-	0
aceB	+	+	+	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0
<i>aceA</i>	*	*	*	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	-	-	-
<i>aceK</i>	*	*	*	-	-	-	-	-	-	-	0	0	-	0	0	0	0	-	-	-	0
cydA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	0
<i>cydB</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	0
pykF	+	+	+	+	+	+	+	+	+	+	0	0	-	-	-	-	-	0	0	-	0
pfkA	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	0	0	0	0
adhE	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	0	0	0	0
icdA	+	+	+	+	+	+	+	+	+	+	-	0	-	0	0	0	0	-	-	-	0
fbp	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ppsA	+	+	+	+	+	+	+	+	+	+	0	0	-	-	-	-	-	-	-	-	-
pckA	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Рис. 3. Экспериментальные данные о Cra-регулоне *Esherichia coli*.



Cra-зависимая регуляция транскрипции:

- репрессия
 - активация
- } экспериментальные данные

Регуляция другими белками:

- аллостерическая репрессия
- репрессия на уровне транскрипции

Табл. 3. Гены, для которых была предсказана регуляция Стг.

Условные обозначения: те же, что и в табл. 2.

Gene	EC	ST	STM	PCA	PCH	YP	YE	YPS	SM	PHL	PM	AA	PBP	VF	VC	VP	VV	PA	PP	PF	PS
<i>tpiA</i>	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
<i>gapA</i>	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	0	0	0	0
<i>gpmA</i>	+	+	+	+	+	+	+	+	+	+	-	-	0	0	0	0	0	0	0	0	0
<i>pdhR</i>	+	+	+	+	+	+	+	+	-	-	0	0	-	-	-	-	-	-	-	-	-
<i>aceE</i>	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>aceF</i>	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>lpdA</i>	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>crp</i>	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
<i>nuoA</i>	+	+	+	+	0	+	+	+	+	+	0	0	0	0	0	0	0	-	-	-	-
<i>nuoB-N</i>	+	+	+	+	0	+	+	+	+	+	0	0	0	0	0	0	0	-	-	-	-
<i>manX</i>	+	+	+	-	+	+	+	+	-	-	-	-	0	0	0	0	0	0	0	0	0
<i>manY</i>	*	*	*	-	*	*	*	*	-	-	-	-	0	0	0	0	0	0	0	0	0
<i>manZ</i>	*	*	*	-	*	*	*	*	-	-	-	-	0	0	0	0	0	0	0	0	0
<i>copA</i>	+	+	+	-	-	+	+	+	+	-	0	-	-	-	-	-	-	0	0	0	0

В результате были сделаны следующие заключения:

tpiA: Сайт расположен в консервативной области, в *E.coli* располагается перед промотором (Рис. 4а). Поэтому в данном организме экспрессия оперона, скорее всего, активируется Сra. О других организмах этого нельзя сказать с такой же уверенностью, поскольку промотор неконсервативен.

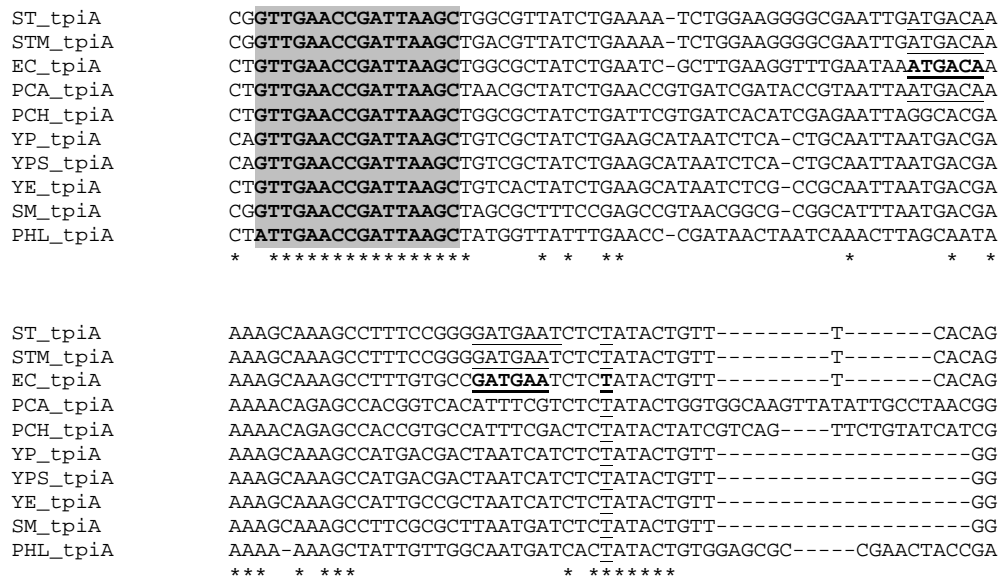


Рис. 4а. Выравнивание регуляторных областей ортологов гена *tpiA*.
Условные обозначения: серым цветом и полужирным шрифтом выделены потенциальные сайты связывания Сra, подчеркнуты промотеры (в *E.coli* выделен полужирным шрифтом); перед названием гена название генома (обозначения – как в Табл.1).

gapA: Промоторы оперона, равно как и потенциальные сайты связывания Сra, располагаются в высоко консервативной области (Рис. 4б). Тем не менее, невозможно определить, является ли ген репрессируемым или активируемым, поскольку сайт, с одной стороны, находится перед промотором, а с другой, расположен вплотную к -35-боксу.

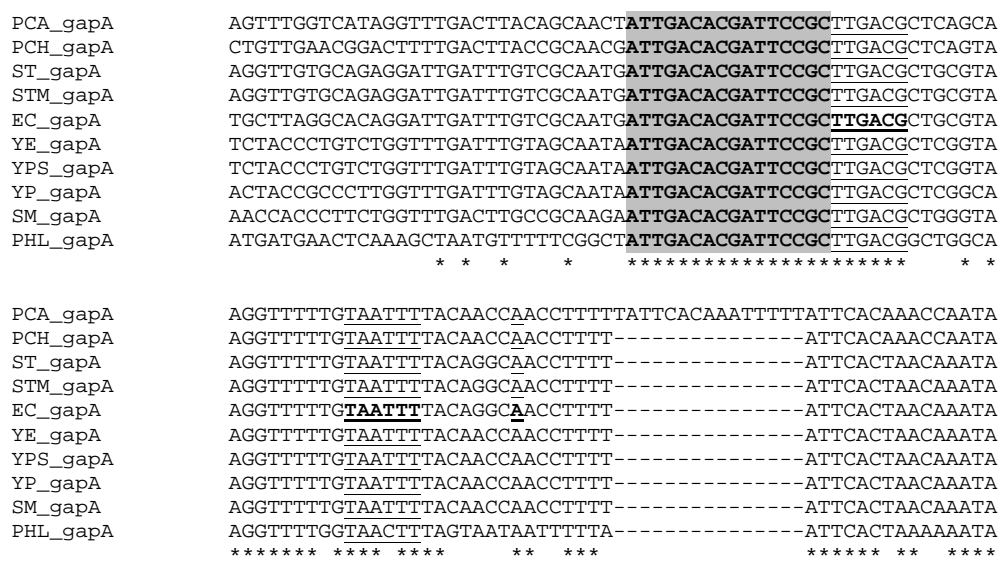


Рис. 4б. Выравнивание регуляторных областей ортологов гена *gapA*.

pdhR: Промотор и потенциальный сайт расположены в консервативных областях (Рис. 4г). Нахождение сайта перед промотором указывает на то, что экспрессия оперона активируется Cra. Кроме того, для бактерий из рода *Yersinia* было показано наличие еще одного сайта, расположенного за промотором. Данный факт указывает на возможность амбивалентной регуляции.

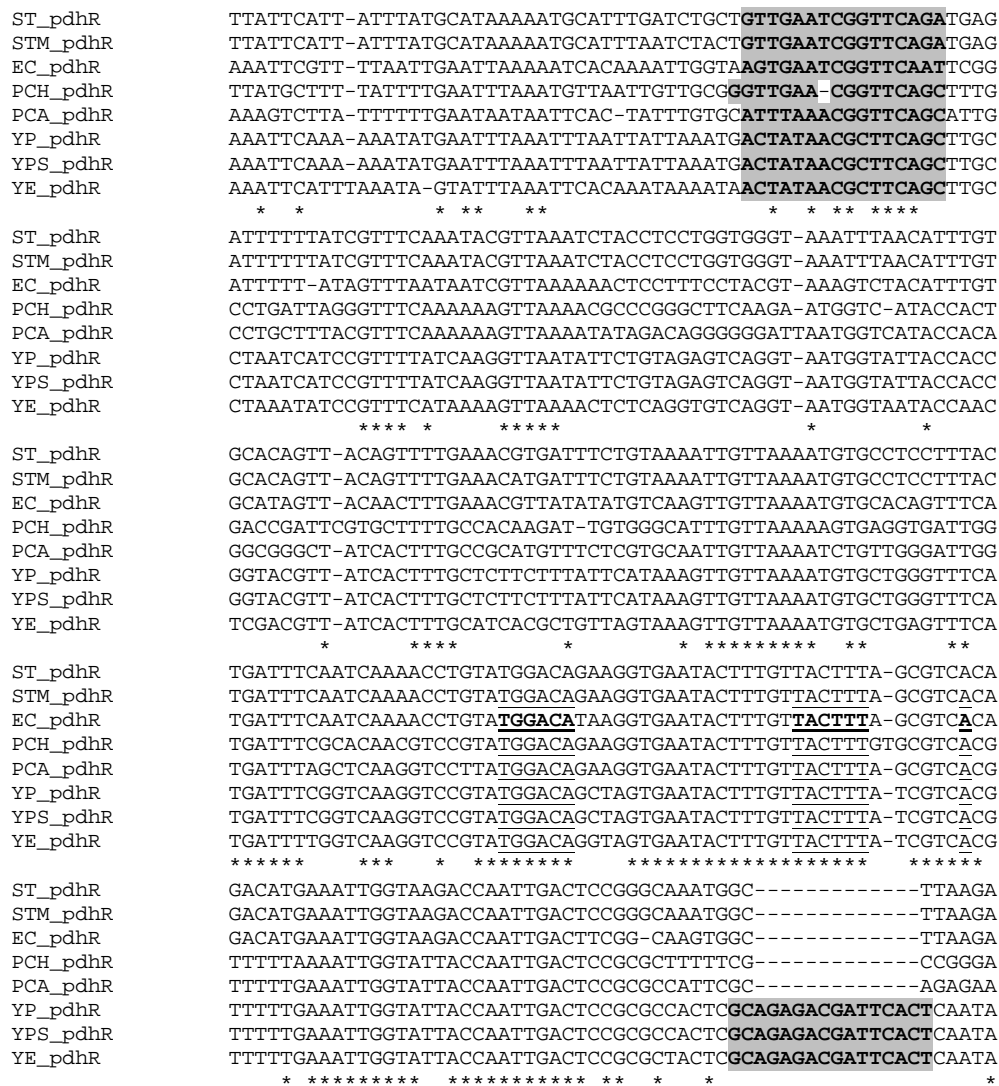


Рис. 4г. Выравнивание регуляторных областей ортологов гена *pdhR*.

gpmA: Для *E.coli* и *Salmonella* spp. сайт находится за экспериментально показанным промотором, из чего следует, что экспрессия оперона должна репрессироваться белком Cra (Рис. 4в). Для остальных белков генов определение типа регуляции невозможно, поскольку и сайт, и промотор неконсервативны.

PCA_gpmA GCAGCGGATCGCCCGCCGATGATAGCGGAATGGTTAGTGCCCTCTGCTGT-CAATGC
PCH_gpmA GCAACGTATCGCCGCGCCGATGGTCGCGGCATGGTCAGCGCCGCTGCTGT-CGATGC
ST_gpmA ATCAGTTTATTCGGTTCTAACGATATGCCGCTACAT--GCCGATGACGCAAG-CTTATC
STM_gpmA ATCAGTTTATTCGGTTCTAACGATATGCCGCTACAT--GCCGATGACGCAAG-CTTATC
EC_gpmA ATCAGTTTATTCGGTTCTAACGATATGCCGCTACAT--GCCGATGACGCAAG-CTTATC
SM_gpmA ACTGCCTGAAAAAATCGCCGCTGATTTCCCCCA-----CGTCATGA-----
YP_gpmA TATGGGGTATTTAATGCCATATATTGATA-----AGTTGCCATT--TGGCAGAATCG
YPS_gpmA TATGGGGTATTTAATGCCATATATTGATA-----AGTTGCCATT--TGGCAGAATCG
YE_gpmA CAAAAA-GGGACACTGGC--TCAGACGATT-----GGATGCAATT-----AAATCG
PHL_gpmA CATGGA-TGTGGAGTGGCCGCTCTGTGGGTTTCTGTTGGGTTCCATTTCTGCTCAATTTG

PCA_gpmA TGGTGATTCAGCGGTTTATCTATTATTGCAC-----AAGG--GAGATTCGGTTTGC
PCH_gpmA TGGTGATTCAGCACTGTTTTTGTGCTCCACCACAAT-AAGGTGGAGAAACAGACAAGC
ST_gpmA GGGCCTACGCGCCATTGCCCTGCTCAAACCTCC-----ACCACTTAAAGACAAAAATCGC
STM_gpmA GGGCCTACGCGCCATTGCCCTGCTCAAACCTCC-----ACCACTTAAAGACAAAAATCGC
EC_gpmA GGGCCTTATCTCATAT---TGTTCAAATCACC-----AGCAAACACCGACATATTGC
SM_gpmA GAGCGATTCGCGCGC-CGTTTCCCCCTCACA-----GCGCCAGCTGAACGATTGAG
YP_gpmA **ATTTCGGCT**TATTGGTCTTACGTTAAAGTAAAAAATTGG--CTATGGTTAAACACAATCGA
YPS_gpmA **ATTTCGGCT**TATTGGTCTTACGTTAAAGTAAAAAATTGG--CTATGGTTAAACACAATCGA
YE_gpmA **ATTTCGAC**CTCTGGCCTTACCTTATGATAAAAAAATTGG--CTATGGTTAAACACTATCGA
PHL_gpmA GTTCATTTGTTGTTCTTACTTATCCAATCAAGTTGGTGACTGCGTTATAAACGTTGTCAA

PCA_gpmA **TGAAACGGTAAAGC**CAGAGCGTGATTTACGCAATAAC-CTGCGGATA---AATCAGCG
PCH_gpmA **TGAAACGATAAAGC**---GCCTATTTTGTGCGCCATAATGACGGCATATA---AATCAGCG
ST_gpmA AACTCTGAGTTCACAAGAACCTTACACTGCGCCACTATT-----TTCGCTA
STM_gpmA AACTCTGAATTCACAAGAACCTTACACTGCGCCACTATT-----TTCGCTA
EC_gpmA AACTCAATATTCACAACAACCTTACACTGCGCCACTATT-----TTCGCTA
SM_gpmA **CTTTTTCCTTTCGCGCAGC-CCTTACTTTGCGGCAAGAAA-----TTGGCTA**
YP_gpmA TTGCTGTTTCGCGAGTATCAGCAATAATAATGATAATAAATTATCATTGAATGTGATTTAT
YPS_gpmA TTGCTGTTTCGCGAGTATCAGCAATAATAATGATAATAAATTATCATTGAATGTGATTTAT
YE_gpmA TTGCTGTTTCGCGAGTATCAGCAATAATAATGATAATAAGTTATCATTGAATATACCCTGAA
PHL_gpmA ACTTGGCCAGTATATATTAGCAATCTGCAAAAGATGCAAACTTAAACCAATAAAGTT

PCA_gpmA CTGCCATCGACTATCGATTGCGCGATAACCGTTCTCCCCCTATAATGAAATCAGTTATC
PCH_gpmA GGCCCGACGACTAGCGATTGCGCGATAACCGCTCTCCGGCCCTATAATGAGAGCAGTTATC
ST_gpmA TGGTTATGCGTAAGC-ATTGCGCGC-CAACACCTCGCGGCAATATAATGAGAATTATTATC
STM_gpmA TGGTTATGCGTAAGC-ATTGCGCGC-CAACACCTCGCGGCAATATAATGAGAATTATTATC
EC_gpmA TGGTTATGCGTAAGC-ATTGCGCGC-CAACACCTCGCGGCAATATAATGAGAATTATTATC
SM_gpmA TGGTTAGACACTATCGATTGCGCGAGCCCGCGCCGCGGCAATATAATGATTTCGTTATC
YP_gpmA TGGTTAGACACTATCGATTGCGCGAGCCCGCGCCGCGGCAATATAATGATTTCGTTATC
YPS_gpmA TGGTTAGACACTATCGATTGCGCGAGCCCGCGCCGCGGCAATATAATGATTTCGTTATC
YE_gpmA TGGTTAGACACTATCGATTGCGCGAGCCCGCGCCGCGGCAATATAATGATTTCGTTATC
PHL_gpmA TGGTTAGACACTATCGATTGCGCGAGCCCGCGCCGCGGCAATATAATGATTTCGTTATC

PCA_gpmA ACAGAAATCTTAGGATAGGAAGGAGTTAGCTATGGCTGTAACCTAAGCTAGTACTGGTAA
PCH_gpmA ATAGAAATCTTAGTATATGAAGGAGTTAAGCTATGGCTGTAACCTAAGCTAGTACTGGTGC
ST_gpmA ATTACAT---ATGATTTATAGGAGTGAGAGTATGGCTGTAACCTAAGCTAGTACTGGTGC
STM_gpmA ATTACAT---ATGATTTATAGGAGTGAGAGTATGGCTGTAACCTAAGCTAGTACTGGTGC
EC_gpmA ATTAATAA---GATGATTTGAGGAGTAAAGTATATGGCTGTAACCTAAGCTAGTACTGGTGC
SM_gpmA ATTGAAATGTCAGATTATTAAGGAGTTAAGCTATGGCTGTAACCTAAGCTAGTACTGGTGC
YP_gpmA ACTGGAGT--AAGTGATTTCGGCTGATTGAGCGCAGCCAACGCACATGCAGCTGGAAGTAT
YPS_gpmA ACTGGAGT--AAGTGATTTCGGCTGATTGAGCGCAGCCAACGCACATGCAGCTGGAAGTAT
YE_gpmA --TAAATT--AGTCATCTGAGGAGTTCGAACTATGGCAGTAACAAAGCTGGTTCGGTAC
PHL_gpmA ATTAAGTT--TGATTTTAAAGGAGTTAAGCTATGGCTGTCACCTAAGCTAGTACTGGTAA

PCA_gpmA GACACGGTGAGAGCCAGTGAACAACGAAAACCGCTTCACAGGCTGGTACGATGTTGATC
PCH_gpmA GACATGGTGAGAGCCAGTGAACAACGAAAACCGCTTCACCGGCTGGATGGACGTTGATC
ST_gpmA GTCACGGTGAAGTCAATGGAACAAAGAGAACCGTTTCACCGGCTGGTATGATGTTGACC
STM_gpmA GTCACGGTGAAGTCAATGGAACAAAGAGAACCGTTTCACCGGCTGGTATGATGTTGACC
EC_gpmA GTCATGGCGAAAAGTCAAGTGAACAACGAAAACCGTTTCACCGGCTGGTACGACGTTGATC
SM_gpmA GTCACGGCGAAAAGCCAGTGAACAACGAAAACCGCTTCACCGGCTGGTATGATGTTGATC
YP_gpmA GACGGCATATACCAAACCTAATTGGAGTCAACGTTA---TGGCAGTAACTAAGCT----
YPS_gpmA GACGGCATATACCAAACCTAATTGGAGTCAACGTTA---TGGCAGTAACTAAGCTAGTT
YE_gpmA GTCACGGCGAAAAGTCAATGGAACAACGAAAACCGCTTCACCGGCTGGTATGACGTTGAT
PHL_gpmA GACATGGCGAAAAGCAGTGAACAACGAAAACCGTTTCACCGGCTGGACTGACGCTCGCGT

Рис. 4в. Выравнивание регуляторных областей ортологов гена *gpmA*.

Были рассмотрены гены, перед которыми есть потенциальные сайты связывания Cra, в *E.coli*, организмах рода *Salmonella* и *Y.pestis*. После того, как были рассмотрены выравнивания потенциальных регуляторных областей ортологов генов *ybaR*, *hldD*, *rpoA*, *accA*, *manX*, *crp*, *rplF*, *nuoA*, *pyrI*, *rfaH* из всех геномов порядка Enterobacteriales, где есть потенциальные сайты связывания Cra, регуляция следующих генов была признана достоверной (Табл.3):

nuoABCEFGHIJKLMN: Потенциальный сайт консервативен отдельно для *E.coli* и рода *Salmonella* и отдельно для остальных организмов порядка Enterobacteriales (Рис. 4д). Функция генов *nuo*-оперона указывает на то, что его экспрессия может регулироваться Cra, поскольку в части реакций рассматриваемых метаболических путей NADH играет важную роль. Поскольку в *E.coli* и бактериях рода *Salmonella* промотор и сайт перекрываются, -35-бокс расположен перед сайтом, то экспрессия оперона, скорее всего, репрессируется Cra.

PHL_plu3089	AGTGGAAAGGCAGGTTTTCCTGAGATAATGATGTAGTAAA--AGCGAGTTATCGGTTAA
PCA_in_dna	CATGGGGGAACAGGCATTTTCTGTGAAACTGGTATAGTGACGCAGCAAACCTGCTTTCTG
YP_nuoA	AGTGGAAAGAACAGGATATTCCTGAGACAATGGTATAGTTAA--AGCGATCTATCTACGCG
YPS_nuoA	AGTGGAAAGAACAGGATATTCCTGAGACAATGGTATAGTTAA--AGCGATCTATCTACGCG
YE_001_1104	AGTGGAAAGAACCAGGTTATTCCTGAGACAATGGTATAGTTGG--TAAGAGCTATTCATGTT
SM_in_dna	AGTGGAAAGAACAGGGAATTCCTGAGACAATGGTATAGTAAA--ACTGAAATGTGCATGTT
STM_nuoA	GGAGGCATCTTCACAAATTCCTGTCAAAATAGGTGTGTGTTATTTCTTTTGCTTCACAAC
ST_nuoA	GGAGGCATCTTCACAAATTCCTGTCAAAATAGGTGTGTGTTATTTCTTTTGCTTCACAAC
EC_nuoA	GGAGGAATCGACACAATTCCTGTCAAAATAGACCCGATATTTTTCATTGCTTCACAAC
	* * * * * * * * * * * *
PHL_plu3089	ACTAGCT-TCCGCCACTGTT--TGCGCCAGTTAACTGACACGTTTTAGC--AGTGCGGTA
PCA_in_dna	TGTCGCTGTACTCTGGTTAA--CCGACGGAAGTACTGACTCGATTTAGCTACGGGTGATA
YP_nuoA	AGTTTCTATGTGTATGTTCTTGCTGAACAAATGACTGACTCGATTTAGCT-CGTCAGGGC
YPS_nuoA	AGTTTCTATGTGTATGTTCTTGCTGAACAAATGACTGACTCGATTTAGCT-CGTCAGGGC
YE_001_1104	GGTTTATATGTGTATGTTCTGACTTAACAAATTACTGACTCGATTTAGCT-TGCAGGTTCT
SM_in_dna	GGTTTATATCTATCTGTTTCTACACAAAAATTACTGACACGATTTAGCCTGTCCCTGGC
STM_nuoA	GGACACGATTCAACAACATAAAATCCCCCCTGGTCGAAGAGTCATTTTCT--CCAGGATG
ST_nuoA	GGACACGATTCAACAACATAAAATCCCCCCTGGTCGAAGAGTCATTTTCT--CCAGGATG
EC_nuoA	GGACACGATTCAACAACATCTAATTATCCTGGAGTCGTCGAAGGATCTGGG--GTGAAAGG
	* * * * * * * * * * * *

Рис. 4д Выравнивание апстримов ортологов гена *nuoA*.
-35-бокс слабый, невозможно найти по консенсусу.

manX: Сайт консервативен отдельно для *E.coli* и рода *Salmonella*, отдельно для рода *Yersinia*, также есть отдельный сайт для *P.luminescens* (Рис. 4е). Промотор консервативен только для *E.coli* и *Salmonella* spp., экспрессия оперона *manXYZ* для которых репрессируется Cra, как и для других систем транспорта сахаров, например, кодируемых оперонами *ptsHI-crr*, *mtlADR* и *fruBKA* (см. Введение).

crp: Сайт и промотор консервативны; экспрессия оперона активируется Cra, поскольку сайт расположен перед промотором (Рис. 4ж). Для рода *Yersinia* показано присутствие еще одного потенциального сайта, расположенного за промотором, результатом чего может явиться амбивалентная регуляция.

```

ST_manX      ATTACAG-----ATCTCCATCACACAA----AATATTTTTTTTCGATA--TCTAAAATA
STM_manX     ATTACGG-----ATCTCCATCACACAA----AATATTTTTTTTCGATA--TCTAAAATA
EC_manX      ATTACGG-----ATCTTCATCACATAA-----AATAATTTTTTTTCGATA--TCTAAAATA
YP_manX      GGTACTG-----AGAACGGTTGCCTGA-----ATCGATTCAATT--AATA--GTAGATGCA
YPS_manX     GGTACTG-----AGAACGGTTGCCTGA-----ATCGATTCAATT--AATA--GTAGATGCA
YE_001_1444 CACATTG-----CGA-TATGTACCTGA-----ATCGATTCAATT--GGCA--TTGCGAATA
SM_011_0780 CGT-TCG-----TGAAGTATCAGCACC---AGGCATT---T-GAAA--CTGCGTGAG
PCH_111_0001 AGGCTGGCCTGAATCGATTCATAACCCGGCGTGATGAGTGTGCTAAACGCTTTTTGTGCC
PCA_001_2120 GATTTC-----TCGCCTGACAACCTGG---AGGAACACAGTGAATG---TCATGA-
PHL_plu2697  GGGATGA-----TTCATCAAATTTATTGCAGTAACCTGTGTAAT-AATCCTGCGTTTATA

```

* * *

```

ST_manX      ATTCACGAAAATCATGCGTTTTTTCGTTGTAACCCTCAT---CTGAATCGATTCGCTTGCG
STM_manX     ATTCACGAAAATCATGCGTTTTTTCGTTGTAACCCTCAT---CTGAATCGATTCGCTTGCG
EC_manX      AATCGGAAACGCAGGGGTTTTTGGTTGTAGCCCTTAT---CTGAATCGATTCGATTGTG
YP_manX      AATCGTCAGCGACAAGCGAATATTGCGACTATCATAATAGATTCAAGTAGCTCAACAGTA
YPS_manX     AATCGTCAGCGACAAGCGAATATTGCGACTATCATAATAGATTCAAGTAGCTCAACAGTA
YE_001_1444 AATCATCAGCAGTACGCTGATATTGTGACTATCATAATAGATTCAATATGTACCGAAGTG
SM_011_0780 AAATGTTAGGCGTGCGTTTTCTTTTTTT-CTACCATAATGTATCTGAACCG--CGGAAGTG
PCH_111_0001 ATCTATCGGGCAAGAAACCGTATTACAGCCATACTGAATATCGCTCACGG-ATAAACGTC
PCA_001_2120 ATCCAC--GGTAAAATTTTTCTTTGCGCCATACT-AATACCGCG-ACGG-TTAACATTC
PHL_plu2697  CCCTATGGGTTTCAAGATGG----ATCGCGACGGCAAGGGAGCGAATCCCCGGGAGTATA

```

*

Рис. 4е Выравнивание регуляторных областей ортологов гена *manX*.

```

PCA_001_3585  GCTGAAGCGAGACAACACAAGACACATCCTTAAAGCTGTGTTA-AAACGAAAATA-GAAT
PCH_014_0010  GCTGAAGCGAGACAACACAAGACACTACACTCAACCTGTGTTACAAAACGAAGGCGTAAAC
SM_015_0228  GCTGAAGCGAGACAACACGAGACACCCGCCCTTACTCTGTGCTACAAAACAGAGCCG-CAAT
YP_crp        GCTGAAGCGATACAACACAAGACAGTAGCCTCATCCTGTGCTACAAAACAGAATCGAAAGT
YPS_crp       GCTGAAGCGATACAACACAAGACAGTAGCCTCATCCTGTGCTACAAAACAGAATCGAAAGT
YE_001_3158  GCTGAAGCGATACAACACAAGACACTAGCCTCATCCTGTGCTACAAAACAGAACCAGAAAGT
PHL_plu0395  GCTGAACAATAACAAGACAAGACAGAAACCCCTTTCTGTGTTAAAAACAGAAGTG-GGCT
ST_crp        GCTGAAGCGAGACACCAGGAGACACACGGCGAAAGCTATGCTA-AAACAGACAAGATGCT
STM_crp       GCTGAAGCGAGACACCAGGAGACACACGGCGAAAGCTATGCTA-AAACAGACAAGATGCT
EC_crp        GCTGAAGCGAGACACCAGGAGACACAAGCGAAAGCTATGCTA-AAACAGTCAGGATGCT
***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

```

PCA_001_3585  AGACSTTTTCTGGACA---AACCC--AGAAACGAAGAAGATGATGACGTAGCCAGAAATA
PCH_014_0010  CGGTGTGCGTTGAGCCTCTGCCTG--ACCAGCCGCCGGCTTACTGGCGTCATCCGATATA
SM_015_0228  AAATCTTTCCGGGCACTGAATTTTCAGGGAACGCCAGAGAAAGGTCCTTTACC GGTA
YP_crp        GATGCTGT---AGCAATTAATTAAT-----AGAGTAGAGATCAATACTCAGTGAAGGGTT
YPS_crp       GATGCTGT---AGCAATTAATTAAT-----AGAGTAGAGATCAATACTCAGTGAAGGGTT
YE_001_3158  ATTTGTTTTT---AGAATTAATTAATAG-----GGTAGCTTCTCAAT-ACTCAGTGAAGGGTT
PHL_plu0395  AAGTATAGGTAGGTTAT--GATCA-----ATACTTAAGGTTCAAATTTTCAGATA
ST_crp        ACAGTAATAC-----ATTG-----ACGTACTGCATGTATGCAGAGGA
STM_crp       ACAGTAATAC-----ATTG-----ACGTACTGCATGTATGCAGAGGA
EC_crp        ACAGTAATAC-----ATTG-----ATGTAAGTGCATGTATGCAAAGGA

```

```

PCA_001_3585  CCC--AGCCA--CGTTTACGGTGCGGCAAAG----CAGCAACC-GAAAAACCAAGG-GTA
PCH_014_0010  CCCTCAATAATTCGATTGCAGGAAAGCCAACGCGCCTGCAGCCTGAAGTATGACGG-TTA
SM_015_0228  TGTT-AACAGAATGTAGCTTGCAGCGCTTAC-----GGCAAGTTAATATGCTAGAG-CGA
YP_crp        TCT-----ATTCGGT-----ACGAACCTTGG-CTA
YPS_crp       TCT-----ATTCGGT-----ACGAACCTTGG-CTA
YE_001_3158  TCT-----ATTCGGT-----ACGAACCTTGG-CTA
PHL_plu0395  ACC-----TATATAAAC-CCT
ST_crp        CAT-----CACATTACAGGCTA
STM_crp       CAT-----CACATTACAGGCTA
EC_crp        CGT-----CACATTACCG--TG

```

Рис. 4ж Выравнивание регуляторных областей ортологов гена *crp*.

Также возможна регуляция оперона *copA*, поскольку сайт расположен в консервативной области (Рис. 4з); функция белка (Cu-транспортующая АТФаза) также говорит о том, что экспрессия оперона возможно регулируется Cra. Промотор неконсервативен, в *E.coli* частично перекрывается с сайтом; поскольку -35-бокс расположен перед сайтом, экспрессия гена репрессируется Cra.

```

PCH_116_0104   CG-CTGATATTCA-----TGATTC CCCCTTGACCCCTCCCTTGCTGGAAGGTTACGC
PCA_001_1061   CA-CTGATATTCA-----TGATTTCC CCCTTGACCCCTCCCTTGCTAGAAGGTTACCT
SM_025_0281   CG-СТААТАТТСА-----ТТТCCCTCCCTTGACCTTACCCTTGCTGGAAGGTTAAGC
ST_ybaR        CG-СТААТАТТСА-----ТАТТТCC CCCTTGACCTTACCCTTGCTGGAAGGTTAACC
STM_copA        CG-СТААТАТТСА-----ТАТТТCC CCCTTGACCTTACCCTTGCTGGAAGGTTAACC
YP_y1093        CG-СТААТГТТСА-----ТТATTC CCCTTGACCTTACCCTTGCTGGAAGGTTAACC
YPS_ATCU        CG-СТААТГТТСА-----ТТATTC CCCTTGACCTTACCCTTGCTGGAAGGTTAACC
YE_001_2446     CG-СТААТАТТСА-----ТТCTTCC CCCTTGACCTTCCCTTGCTGTAAGGTTAACC
PHL_plu3824     CA-СТААТГТТСА-----ТТТТТCC CCCTTGACCTTACCCTTGAGGGAAGGTTAACC
EC_ybaR         GGACTTTTACCCCGCTGGTTТААТТТТ TTGACCTTCCCTTGCTGGAAGGTTAACC
                   **  *  *                *  *   **** *  **** *  ****

PCH_116_0104   TTCAGGTAGTGTGAGAACTAATCAAAATCACGCTTGC GGTCAATGTCTTTGACGTT-CG
PCA_001_1061   TAC-----TCAATAACTAGTAGAAATCAGATGGCCGGTCAACGTCTTTCACCTC-CG
SM_025_0281   TCC-----ATTCTTAGTGAACACAGGTGAGGCGGTCAAC--CTTTAAACAC-CG
ST_ybaR        ТТТА-----ТCACAGTGAGTCACTACGATТАСТАATGTCAATG-AATTGATGAC-CA
STM_copA        ТТТА-----ТCACAGTGAGTCACTACGATТАСТАAGGTCATG-AATTGATGAC-CA
YP_y1093        TCT-----ATGGTTAGTGAAAATAGCCATGCCAGTCAAGCGGAATAAAAAGC-CG
YPS_ATCU        TCT-----ATGGTTAGTGAAAATAGCCATGCCAGTCAAGCGGAATAAAAAGC-CG
YE_001_2446     TCT-----ATAATTAGTGAAAAGGAAGGATGACAGTCAAGCGGAATAAAAACC-TG
PHL_plu3824     TCCCAACAAGTCAGAAACTCATTGAACTGTGTAGATATTTCATCACCAATTATCACACA
EC_ybaR        ТТТА-----ТCACAAGCCAGTCAAAACTGTCTTAAAGG--AGTG--TTTATGTCA
                   *                               *          *  *
  
```

Рис. 4з Выравнивание регуляторных областей ортологов гена *copA*.

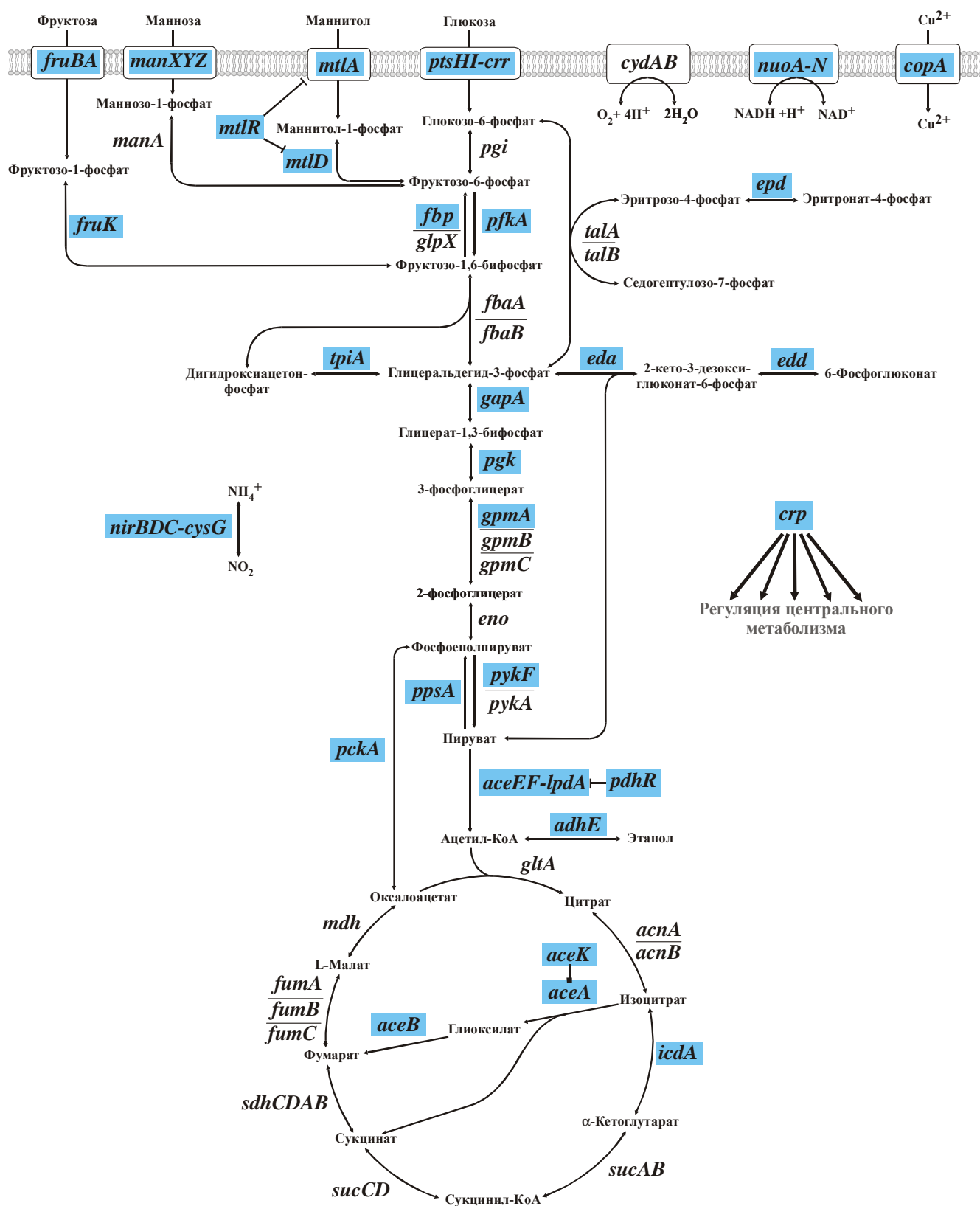
4. Эволюция обоженного Cra-регулона

Как уже было сказано ранее, в *E.coli* экспериментально показана регуляция для 15 оперонов (Рис.3), из которых 9 репрессируются, а 6 активируются Cra (см. Введение). Еще для восьми генов была предсказана регуляция; из них три активируются, причем для двух из них (*pdhR* и *cnp*) возможна амбивалентная регуляция, четыре репрессируются Cra (в двух из них – *nuoA* и *copA* – происходит перекрывание сайта и промотора), а тип регуляции гена *gapA* определить невозможно (см. «Новые члены регулона»). Было принято решение исключить оперон *cydAB* из Cra-регулона, что объясняется двумя причинами. Во-первых, экспериментальные данные по его регуляции недостоверны (Ramsaier et al., 1996), а во-вторых, сайт перед геном имеет вес ниже порогового значения (3,49) и обнаружен только в *E.coli*.

У обоих видов рода *Salmonella* состав регулона совпадает с *E.coli* (Рис.5а), за исключением отсутствия сайтов перед опероном *cydAB*, который, как было сказано ранее, был исключен из состава регулона.

У остальных организмов, принадлежащих группе Enterobacteriales, – более дальних родственников *E.coli* – наблюдается редукция Cra-регулона.

Рис. 5а. Структура регулона в геномах различных гамма-протеобактерий: *Escherichia coli*, *Salmonella typhi*, *S.typhimurium*.



FruR-зависимая регуляция транскрипции:

- предсказанная регуляция

У *P.chrysanthemi* и *P.carotovorum* (Рис.5б) пропадает регуляция оперонов *mtlADR*, *ptsHI-crr*, *pckA*, *aceBAK*, *nirBD*, *fbp*, *copA*; у *P.carotovorum* также не регулируется *manX*. Оперон *edd-eda* распадается, при этом ген *edd* в *P.chrysanthemi* полностью отсутствует. Также в обоих организмах исчезают гены *nirC* и *nuoA*.

У организмов рода *Yersinia* и у *S.marcescens* (Рис.5в) исчезает регуляция оперонов *edd-eda*, *mtlADR*, *pckA*, *aceBAK*, *fbp* (за исключением *Y.enterocolitica*, у которой перед данным геном обнаружен потенциальный сайт). Также отсутствует сайт связывания Cга перед опероном *pdhR-aceEF-lpdA* и *manXYZ* у *S.marcescens*.

У *P.luminescens* (Рис.5г) не регулируются опероны *ptsHI-crr*, *pckA*, *aceBAK*, *fbp*, *edd-eda*, *pdhR-aceEF-lpdA*, *manX*, *copA* и полностью отсутствуют *edd*, *nirBDC-cysG*, *mtlADR*. Наблюдается оперонная перестройка: *fruBKA*-оперон распадается на опероны *fruBA* и *fruK*. Такое разделение соответствует функциям закодированных генами белков: FruB и FruA составляют единый мембранный комплекс, тогда как FruK является цитоплазматическим ферментом. Но, несмотря на то, что *fruK* транскрибируется отдельно от оперона *fruBA*, оба оперона регулируются Cга.

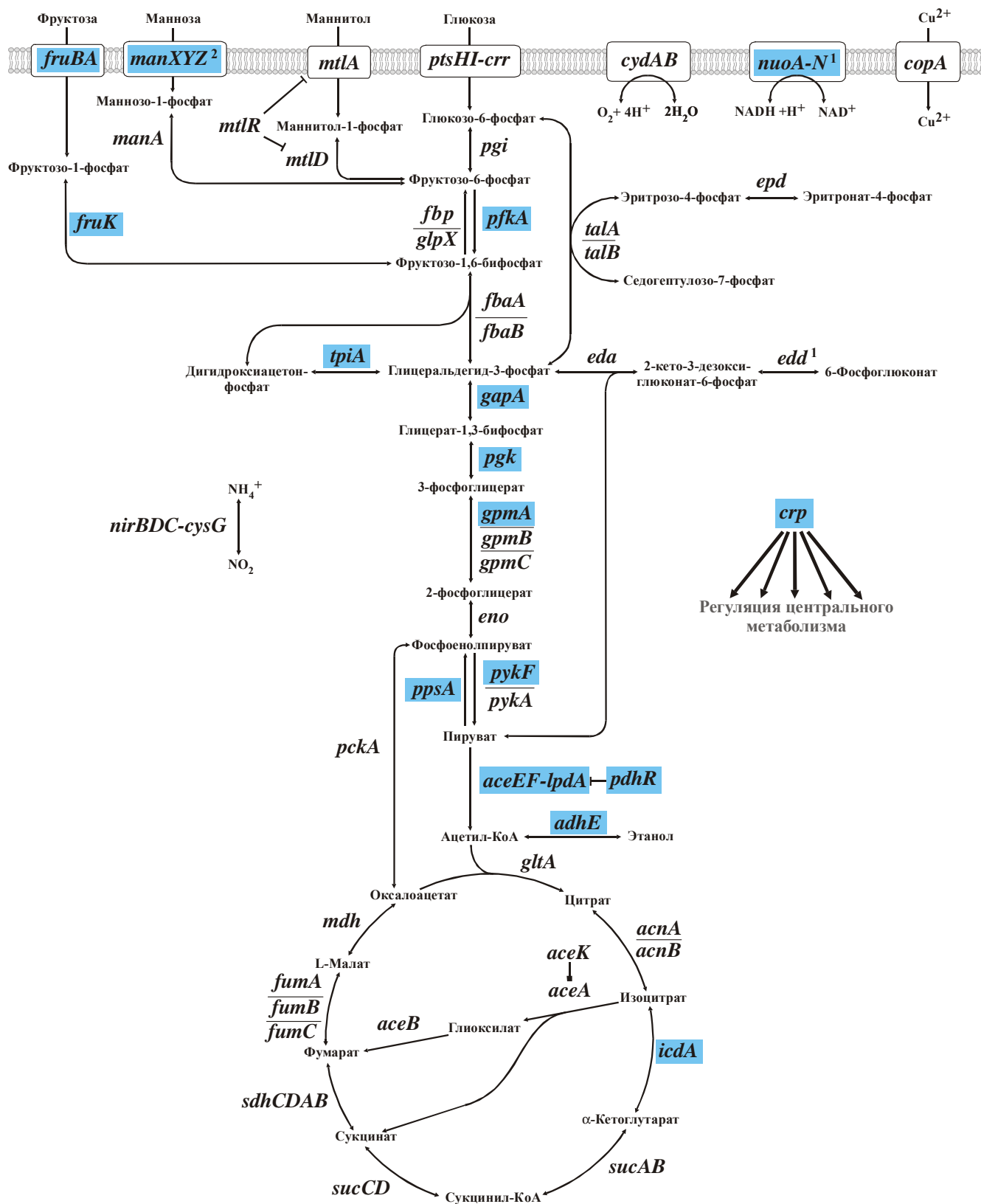
Несмотря на то, что, судя по филогенетическому дереву, порядок Pasteurellales находится в более близкородственных отношениях с порядком Enterobacteriales, чем другие исследуемые семейства, регулон у порядка Pasteurellales не сохраняется. Кроме того, не найдены ортологи для ряда генов (Табл. 2,3).

У порядка Vibrionales сохраняется регуляция только *fruBKA* оперона (Рис.5д).

У бактерий группы Pseudomonadales (Рис.5е) также присутствует только *fruBKA*-оперон, причем у него была обнаружена доменная перестройка. Так, первый ген фруктозного оперона *fruBKA* кодирует, по всей видимости, составной белок: N-концевая часть данного белка гомологична FruB, тогда как C-концевая часть – PtsI. Скорее всего, в ходе эволюции произошло встраивание гена *ptsI* внутрь *fru*-оперона и слияние рамок считывания *fruB* и *ptsI*.

Рис. 56. Структура регулона в геномах различных гамма-протеобактерий:

Pectobacterium carotovorum, *Pectobacterium chrysantemi*.

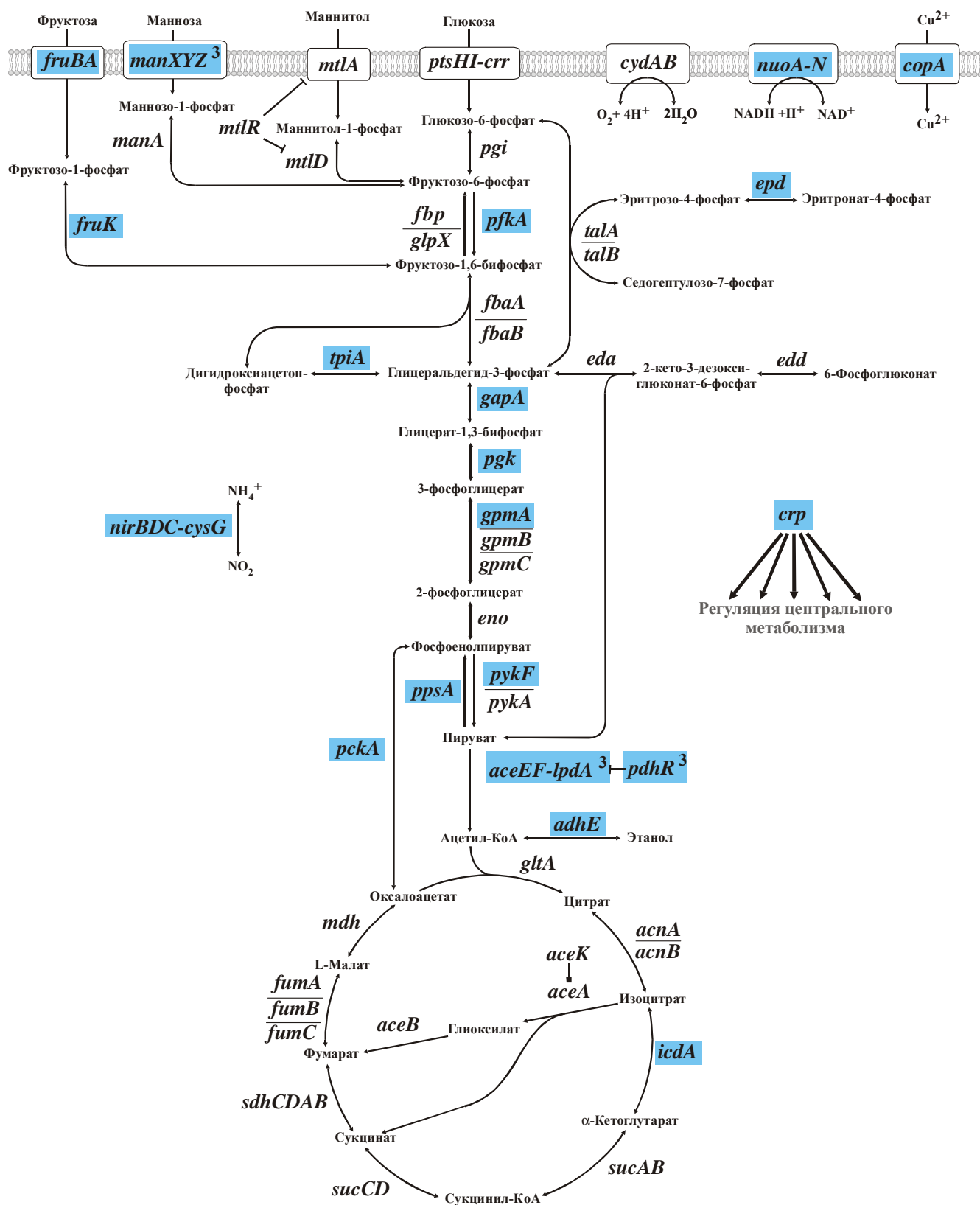


¹ ген отсутствует в *Pectobacterium chrysantemi*

² сайт перед геном отсутствует в *Pectobacterium carotovorum*

Рис. 5в. Структура регулона в геномах различных гамма-протеобактерий:

Yersinia pestis, *Y.pseudotuberculosis*, *Y.enterocolitica*, *Serratia marcescens*.



³ сайт перед геном отсутствует в *Serratia marcescens*

Рис. 5г. Структура регулона в геномах различных гамма-протеобактерий:

Photorhabdus luminiscens.

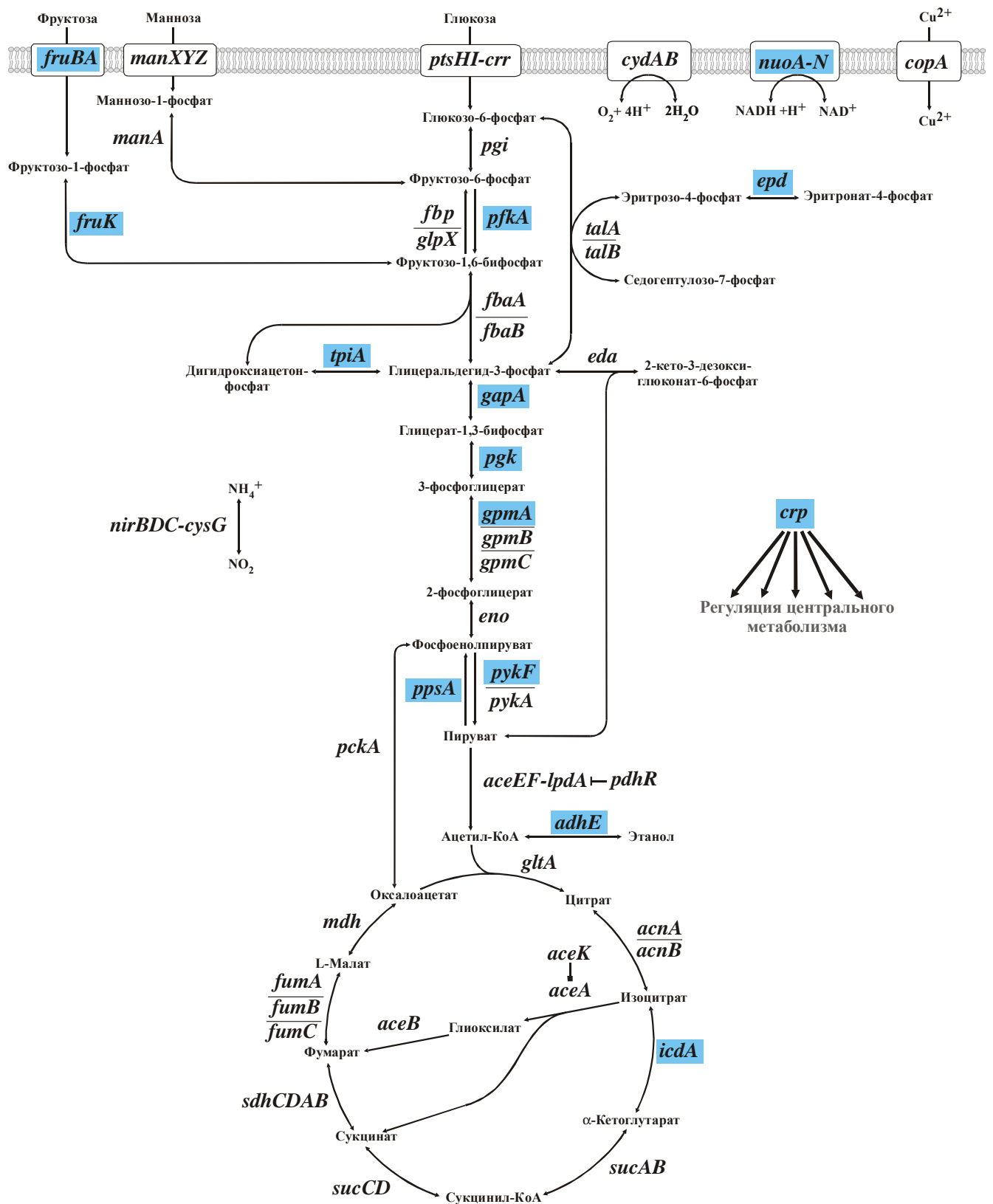
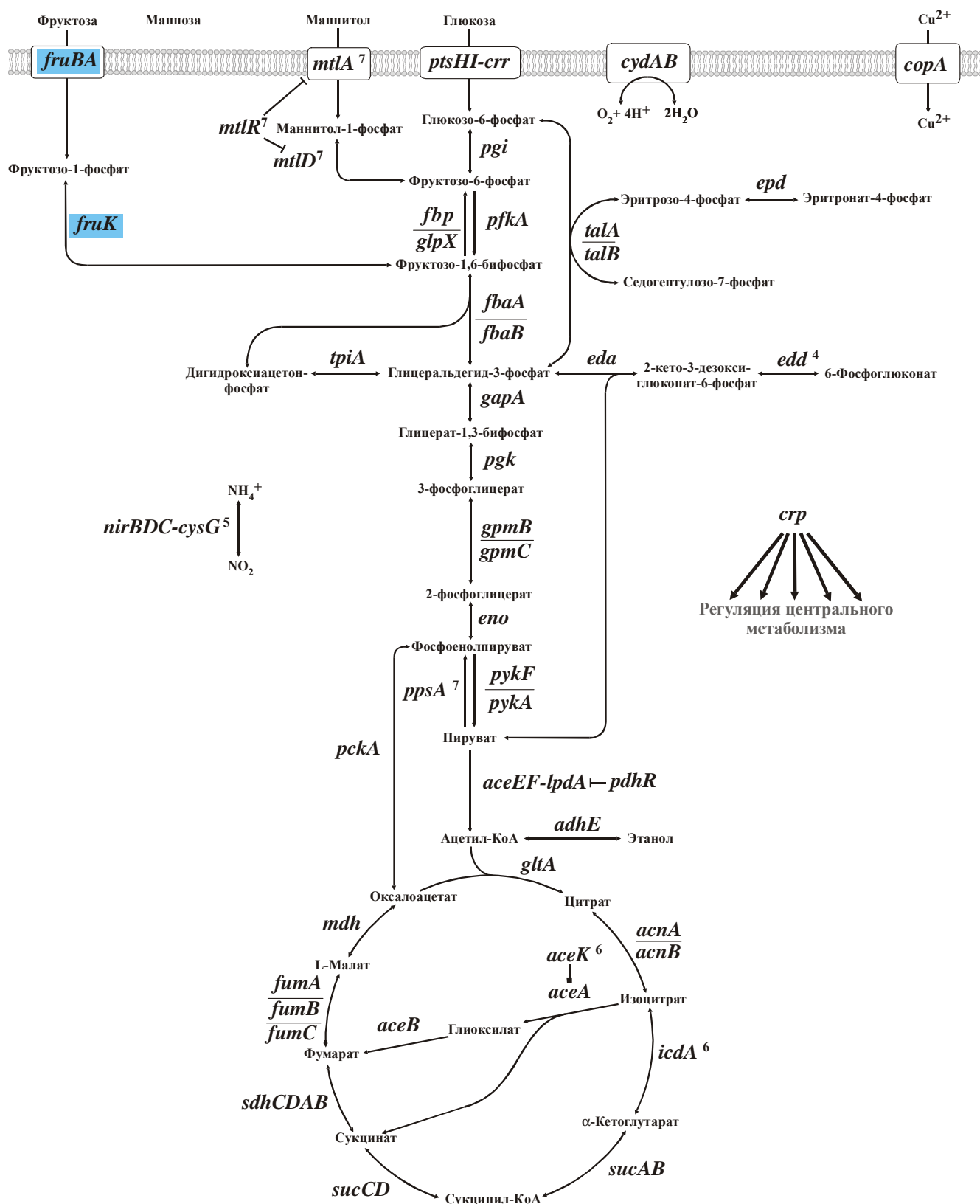


Рис. 5д. Структура регулона в геномах различных гамма-протеобактерий:

Vibrio vulnificus, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*, *Photobacterium profundum*.



⁴ ген отсутствует в *Photobacterium profundum*

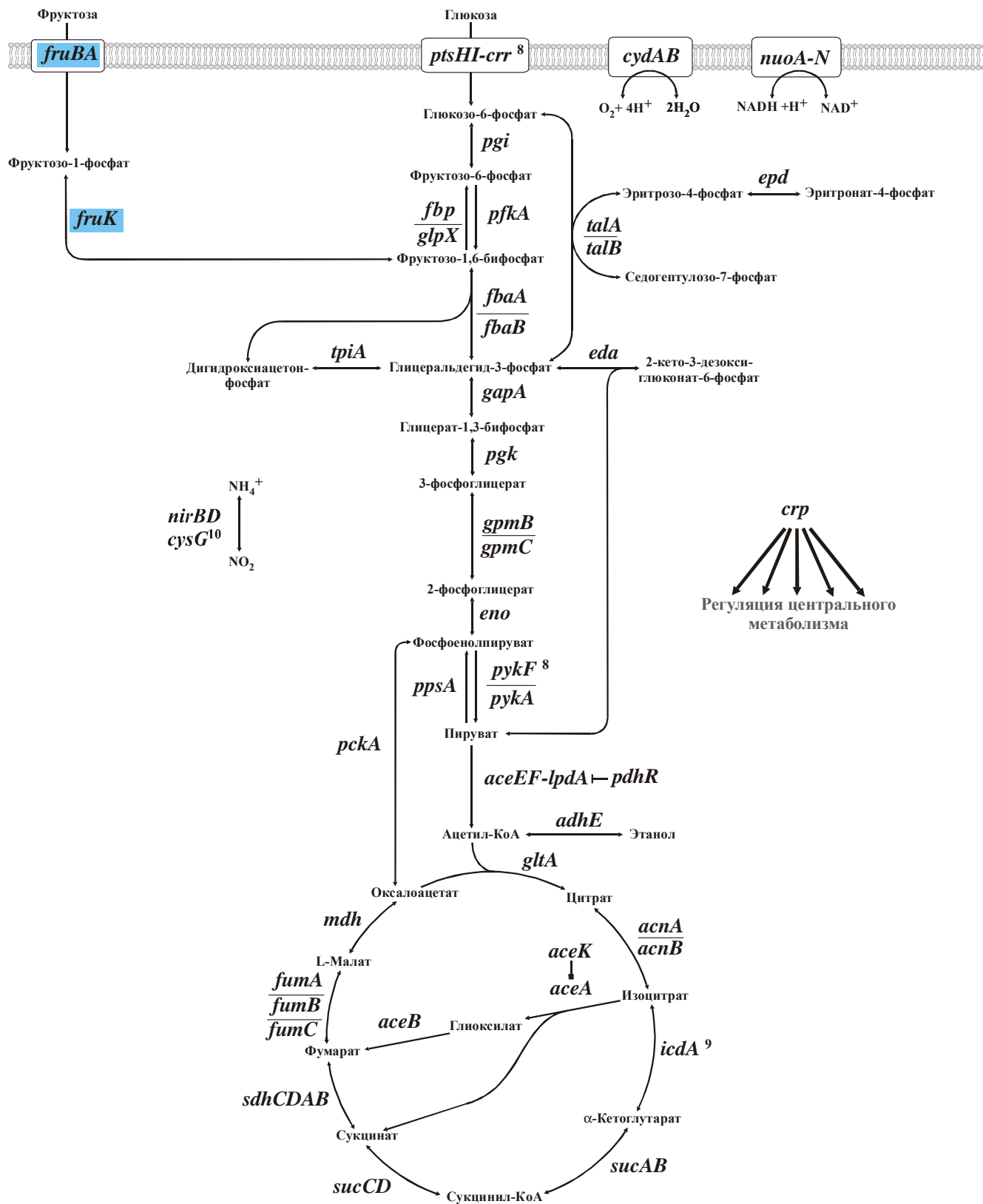
⁵ ген отсутствует в *Vibrio cholerae*

⁶ ген отсутствует в *Vibrio vulnificus*, *V. parahaemolyticus*, *V. cholerae* и *V. fischeri*

⁷ ген отсутствует в *Vibrio fischeri*

Рис. 5е. Структура регулона в геномах различных гамма-протеобактерий:

Pseudomonas fluorescens, *P. aeruginosa*, *P. syringae*, *P. putida*.



⁸ ген отсутствует в *Pseudomonas aeruginosa*, *P. syringae* и *P. putida*

⁹ ген отсутствует в *Pseudomonas syringae*

¹⁰ ген отсутствует в *Pseudomonas putida*

Выводы.

На основании проведенного исследования были сделаны следующие выводы (Рис.6):

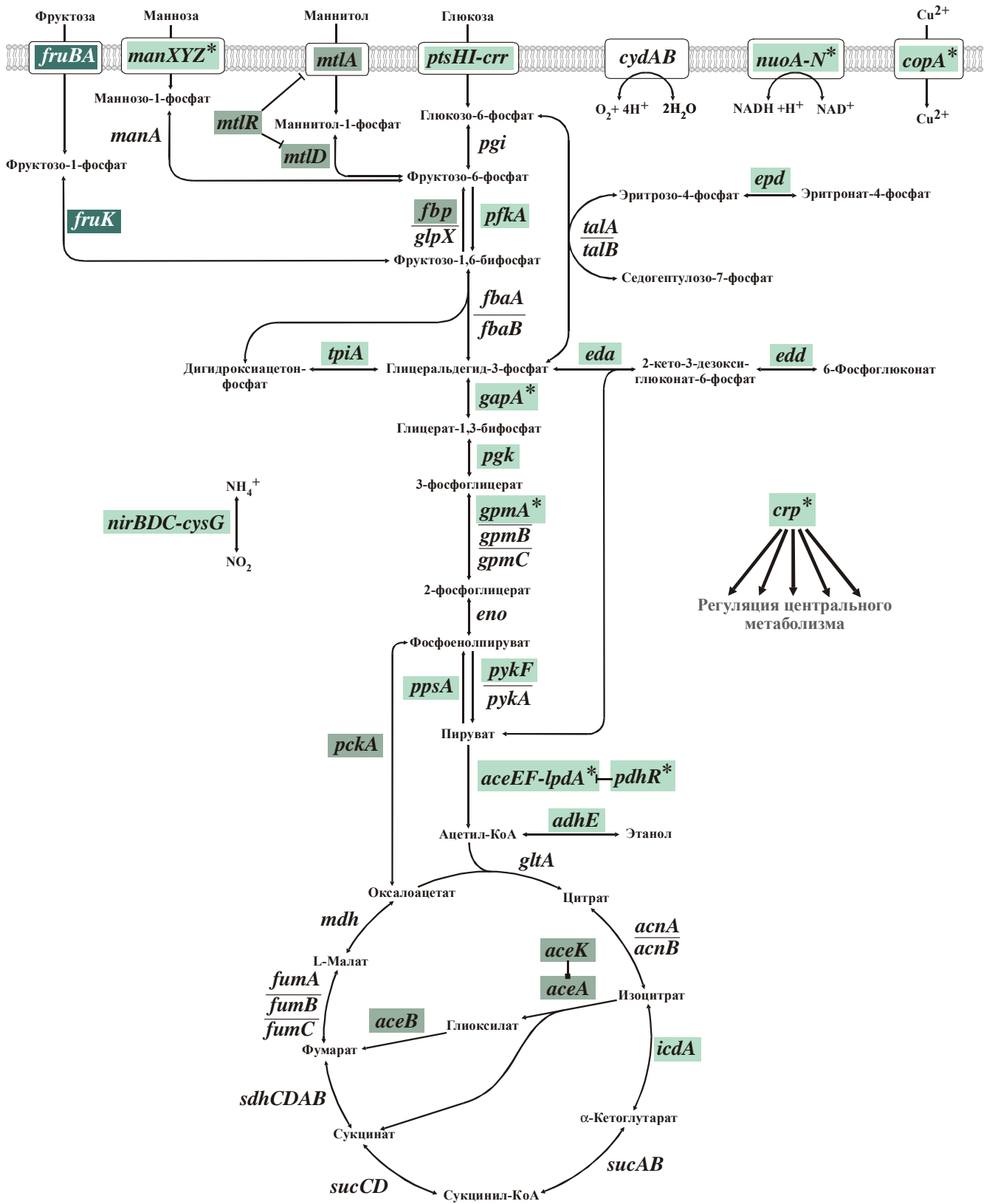
1. Наиболее консервативной является регуляция оперона *fruBKA*, отвечающего за систему транспорта и фосфорилирования фруктозы, – сайты связывания перед ним обнаружены в абсолютном большинстве случаев. Таким образом, наиболее вероятно, что Cra первоначально возник как узкоспецифичный регулятор фруктозо-фосфотрансферазной системы.

2. У бактерий порядка Pasteurellales наблюдается полное отсутствие регулона.

3. В геномах многих бактерий из порядка Enterobacteriales Cra регулирует гены большинства ферментов гликолиза/глюконеогенеза, а также глобальный регулятор *crp*. Достаточно консервативна регуляция генов для цикла Кребса (*icdA*), ассимиляторной нитрит-редуктазы (*nirBD*), NADH дегидрогеназы (*nuoA-N*), системы транспорта маннозы (*manXYZ*) и Су-транспортирующей АТФазы (*copA*).

4. Для узкой группы энтеробактерий (*E.coli*, *Salmonella sp.*) характерна также регуляция генов фосфотрансферазной системы (*mtlADR*), глиоксилатного пути (*aceBAK*) и пути Энтнера-Дудорова (*edd-eda*).

Рис. 6. Эволюция регулона в геномах различных гамма-протеобактерий.



Dark Green Box - регуляция сохраняется во всех геномах, за исключением Pasteurellales

Light Green Box - регуляция сохраняется в большинстве геномов Enerobacteriales

Medium Green Box - регуляция сохраняется лишь в геномах *Escherichia coli* и *Salmonella* spp.

* - регуляция предсказана в данной работе

Список литературы:

- Bell K.S., Sebaihia M., Pritchard L., Holden M.T., Hyman L.J., Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J., Mungall K., Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H., Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S., Salmond G.P., Birch P.R., Parkhill J., Toth I.K. (2004) **Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors.** *Proc. Natl. Acad. Sci. U. S. A.* **101 (30)**: 11105-11110.
- Benson D.A., Boguski M.S., Lipman D.J., Ostell J., Ouellette B.F., Rapp B.A., Wheeler D.L. (1999) **GenBank.** *Nucleic Acids Res.* **27 (1)**: 12-17.
- Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y. (1997) **The complete genome sequence of *Escherichia coli* K-12.** *Science.* **277 (5331)**: 1453-1474.
- Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T., Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H., Nelson W.C., Davidsen T., Zafar N., Zhou L., Liu J., Yuan Q., Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T., Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L., White O., Fraser C.M., Collmer A. (2003) **The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000.** *Proc. Natl. Acad. Sci. U. S. A.* **100 (18)**: 10181-10186.
- Chain P.S., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L., Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C., Simonet M., Chenal-Francois V., Souza B., Dacheux D., Elliott J.M., Derbise A., Hauser L.J., Garcia E. (2004) **Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*.** *Proc. Natl. Acad. Sci. U. S. A.* **101 (38)**: 13826-13831.
- Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., Perry R.D. (2002) **Genome sequence of *Yersinia pestis* KIM.** *J. Bacteriol.* **184 (16)**: 4601-4611.
- Deng W., Liou S.R., Plunkett G., Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R. (2003) **Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.** *J. Bacteriol.* **185 (7)**: 2330-2337.
- Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C. (1995) **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science.* **269 (5223)**: 496-512.
- Gelfand M.S., Koonin E.V., and Mironov A.A. (2000) **Prediction of transcription regulatory sites in Archaea by a comparative genomic approach.** *Nucleic Acids Research.* **28 (3)**: 695-705.

Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M. (2000) **DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae***. *Nature*. **406 (6795)**: 477-483.

Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M. (2002) **Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis***. *Nat. Biotechnol.* **20 (11)**: 1118-1123.

Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y.K.S.Y.T., Honda T., Shinagawa H., Hattori M., Iida T. (2003) **Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae***. *Lancet*. **361 (9359)**: 743-749.

May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V. (2001) **Complete genomic sequence of *Pasteurella multocida*, Pm70**. *Proc. Natl. Acad. Sci. U. S. A.* **98 (6)**: 3460-3465.

McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K. (2001) **Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2**. *Nature*. **413 (6858)**: 852-856.

Mikulskis A., Aristarkhov A., and Lin E.C.C.(1997) **Regulation of Expression of the Ethanol Dehydrogenase Gene (*adhE*) in *Escherichia coli* by Catabolite Repressor Activator Protein Cra**. *Journal of Bacteriology* **179 (22)**: 7129-7134.

Negre D., Bonod-Bidaud C., Geourjon C., Deleage G., Cozzone A.J., and Cortay J.-C. (1996) **Definition of a consensus DNA-binding site for the *Escherichia coli* pleiotropic regulatory protein, FruR**. *Molecular Microbiology* **21 (2)**: 257-266.

Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H., Martins dos Santos V.A., Fouts D.E., Gill S.R., Pop M., Holmes M., Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J., Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I., Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A., Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S., Kiewitz C., Eisen J.A., Timmis K.N., Dusterhoft A., Tumbler B., Fraser C.M. (2002) **Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440**. *Environ. Microbiol.* **4 (12)**: 799-808.

Ramseier T.M., Chien S.Y., and Saier M.H. Jr. (1996) **Cooperative Interaction Between Cra and Fnr in the Regulation of the *cydAB* Operon of *Escherichia coli***. *Current Microbiology* **33 (4)** 270-274

Ruby E.G., Urbanowski M., Campbell J., Dunn A., Faini M., Gunsalus R., Lostroh P., Lupp C. , McCann J., Millikan D., Schaefer A., Stabb E., Stevens A., Visick K., Whistler C., Greenberg E.P. (2005) **Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners**. *Proc. Natl. Acad. Sci. U. S. A.* **102 (8)**: 3004-3009.

Saier M.H. Jr., (1996) **Cyclic AMP-independent catabolite repression in bacteria.** *FEMS Microbiology Letters* **138**: 97-103.

Saier M.H. Jr., and Ramseier T.M. (1996) **The Catabolite Repressor/Activator (Cra) Protein of Enteric Bacteria.** *Journal of Bacteriology* **178 (12)**: 3411-3417.

Saier M.H. Jr., Ramseier T.M., and Reizer J. (1996) **Regulation of carbon utilization.** In Neidhart F.C. (eds.) *Escherichia coli and Salmonella*. Cellular and Molecular Biology, ASM Press, Washington, pp. 1325-1333.

Stover C.K., Pham X.Q., Erwin A.L., Mizoguchi S.D., Warrener P., Hickey M.J., Brinkman F.S., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R., Smith K., Spencer D., Wong G.K., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E., Lory S., Olson M.V. (2000) **Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.** *Nature*. **406 (6799)**: 959-964.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., and Higgins D.G. (1997) **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids. Res.* **25 (24)**: 4876-4882.

Tyson K., Busby S., and Cole J. (1997) **Catabolite regulation of two *Escherichia coli* operons encoding nitrite reductases: role of the Cra protein.** *Arch Microbiol.* **168 (3)**: 240-244.

Миронов А.А., Винокурова Н.П., и Гельфанд М.С. (2000) **Програмное обеспечение анализа бактериальных геномов.** *Мол. Биол.* **34 (2)**: 253-262.