



Сравнительно-геномный анализ метаболизма лактозы у бактерий

Александра Еремина, Мария Тутукина, Анна Казнадзей



Введение

Лактоза – распространенный в природе дисахарид, состоящий из глюкозы и галактозы. Хорошо описаны несколько бактериальных метаболических путей деградации лактозы. Различные бактерии способны питаться лактозой, используя для этого разные наборы ферментов.

Известно, что гены, относящиеся к определенному метаболическому пути, часто встречаются рядом в бактериальном геноме, составляя кассеты. Анализ таких кассет позволяет исследовать эволюционные особенности разных генов, обнаруживать устойчивые сочетания и выявлять случаи горизонтального переноса.

Цели

- ▶ проанализировать все кассеты, имеющие отношение к лактозному метаболизму
- ▶ выявить консервативные сочетания генов
- ▶ проверить, разделяются ли кассеты на лактозные (транспорт и гидролиз) и галактозные (дальнейшие реакции)
- ▶ найти случаи горизонтального переноса консервативных кассет генов

Гипотеза 1:

Продуктами первого этапа расщепления лактозы являются глюкоза и галактоза – распространенные продукты питания бактерий, часто находящиеся в среде в свободном виде, а также являющиеся продуктами распада других углеводов и их производных. Это позволяет предположить, что гены, ответственные за гидролиз и за дальнейшие реакции метаболического пути, не обязательно будут наблюдаться в составе одних и тех же кассет.

Обсуждение

Более ¾ изученных бактерий способны питаться лактозой. Об этом свидетельствует наличие генов соответствующих гидролаз в их геномах. Многие бактерии используют для метаболизма лактозы несколько путей одновременно. Большинство бактерий, питающихся лактозой, не имеют длинных консервативных кассет, содержащих соответствующие гены. Гены, ответственные за гидролиз лактозы, чаще всего лежат отдельно от генов, кодирующих другие ферменты лактозного пути (более 85% случаев). Они также встречаются рядом с генами галактокиназ, но лишь в 5% случаев.

Было обнаружено только одна длинная консервативная кассета генов одного из самых распространенных метаболических путей деградации лактозы, известного для бактерий родов Bacillales и Lactobacillales. Интересно, что гены данной кассеты наблюдались также у множества бактерий Enterobacteriales, но все они располагались не друг рядом с другом. Наблюдался случай горизонтального переноса этой кассеты в бактерию *Leptotrichia buccalis*, также обитающую в организме человека. Найдена ошибка аннотации распространенного гена, ранее определенного, как фактор транскрипции из семейства DeoR. С помощью анализа структуры доменов соответствующего белка выяснено, что фермент является изомеразой.

Гипотеза 2:

Мы предполагаем, что имеющаяся у *E.coli* и других энтеробактерий консервативная кассета генов, участие которых в метаболизме сульфоглюкозы описано в статье Denger et al., также участвует в деградации лактозы или галактозы.

В пользу этой гипотезы свидетельствуют:

- ▶ полное соответствие основных функций ферментов кассеты функциям ферментов вышеописанного метаболического пути лактозы у бацилл.
- ▶ совпадение одного из ферментов (альдозазы) в обоих путях
- ▶ специфичность эписимеразы, присутствующей в найденном пути и в известном пути деградации галактозы.
- ▶ наличие в промоторной области сайта связывания фактора CRP*
- ▶ способность некоторых генов обоих путей сочетаться в составе других кассет

*Исследование генов методом филогенетического футпринтинга показало перекрытие областей сайта связывания транскрипционного фактора CRP и промотора. По-видимому, фактор CRP работает как репрессор, подавляя экспрессию генов этой кассеты в случае высокой концентрации глюкозы в клетке. Мы предполагаем, что на его работу не влияет сульфоглюкоза. Эта гипотеза может быть проверена в прямом эксперименте.

Материалы и методы

- 300 тыс. генов углеводного метаболизма, собранных в 55 тыс. кассет.
- Metacyc – база данных метаболических путей
- База данных IMG JGI (COG, Pfam)
- Enzyme classification database – описание функций ферментов
- Clastal Omega – выравнивание нуклеотидных последовательностей
- Phylogenetic footprinting – поиск сайтов регуляции транскрипции
- PHYRE2 – построение трехмерной структуры белков, поиск доменов
- iTOL – построение филогенетических деревьев бактерий
- набор инструментов для сравнительно-геномного анализа кассет генов

Результаты

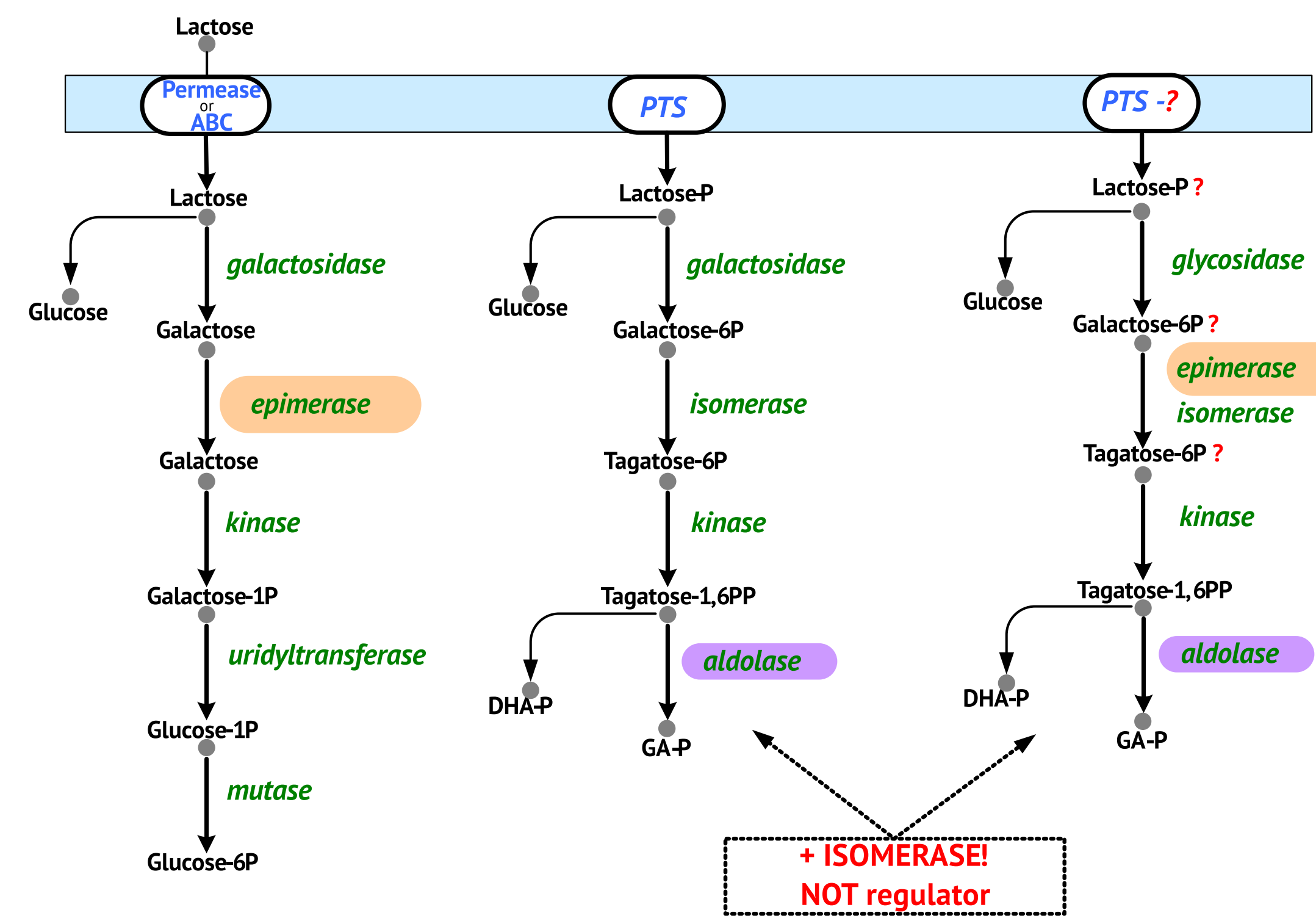


Рис. 1. Два распространенных пути деградации лактозы у бактерий и один предсказанный. В гипотетическом пути альдозаза совпадает с альдозазой из лактозного пути Bacillales, а эпимераза с эпимеразой из пути деградации галактозы Лелуара. Отмеченная желтым изомеразой ранее считалась регулятором транскрипции и присутствует в обоих указанных кассетах.

Fig. 1 Two most common bacterial lactose degradation pathways and one predicted pathway. In the hypothetical pathway aldolase matches aldolase from classic Bacillales lactose pathway, and epimerase matches epimerase from Leloir galactose degradation pathway. Isomerase marked yellow was previously considered a transcription regulator and is present in both respective gene cassettes.

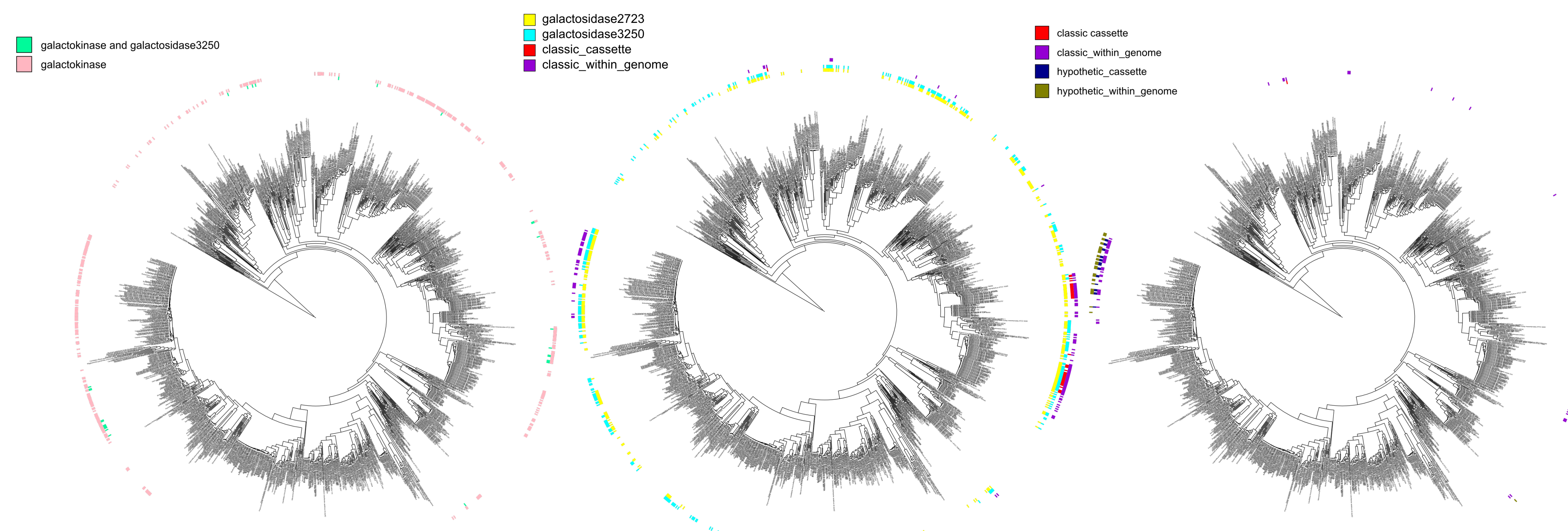


Рис. 5. Распространенность стандартной и гипотетической кассеты генов деградации лактозы относительно распространенности генов соответствующих путей, не лежащих внутри кассет

Fig. 5. Distribution of standard and hypothetical cassette of lactose degradation genes compared to non-cassette distribution of the same genes

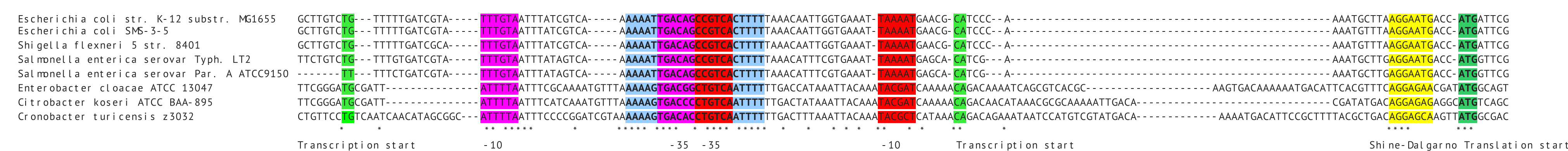


Рис 8. Множественное выравнивание межгенных областей исследуемой кассеты Enterobacteriales

Голубым отмечен сайт связывания фактора транскрипции CRP. Видно, что он перекрывается с промотором, поэтому мы предполагаем, что при связывании фактора происходит репрессия транскрипции.

Fig 8. Multiple alignment of intergenic regions of studied Enterobacteriales cassette

Transcription factor CRP binding site is marked blue, it is overlapping with promoter region, thus we suggest that when factor binds to the site it prevents transcription.

Results

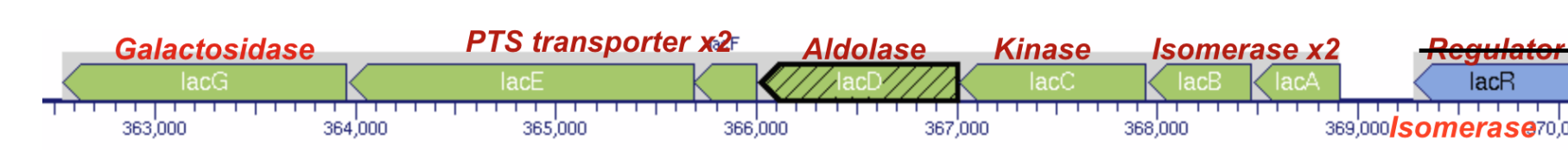


Рис. 2 Гены пути деградации лактозы (Streptococcus)

Fig. 2 Lactose degradation pathway genes (Streptococcus)

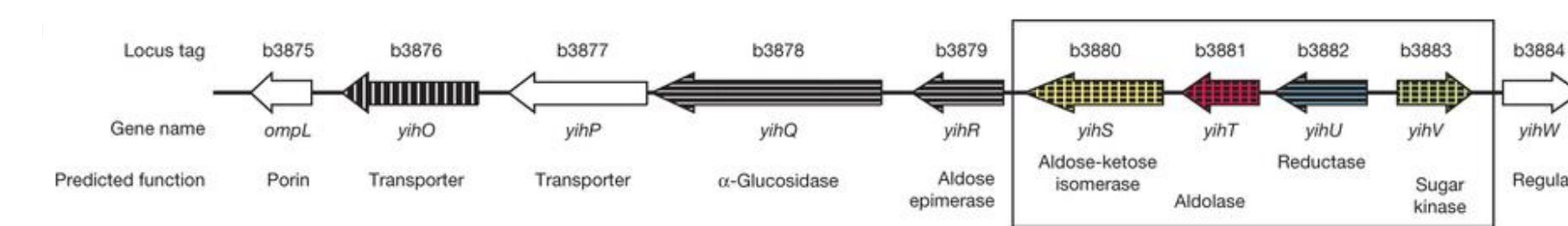


Рис. 3 Гены пути деградации сульфодезокси-глюкозы, также, возможно, способные деградировать лактозу (E. coli)

Fig. 3 Sulphodexoxyglucose degradation pathway genes predicted to be able to also degrade lactose (E. coli)

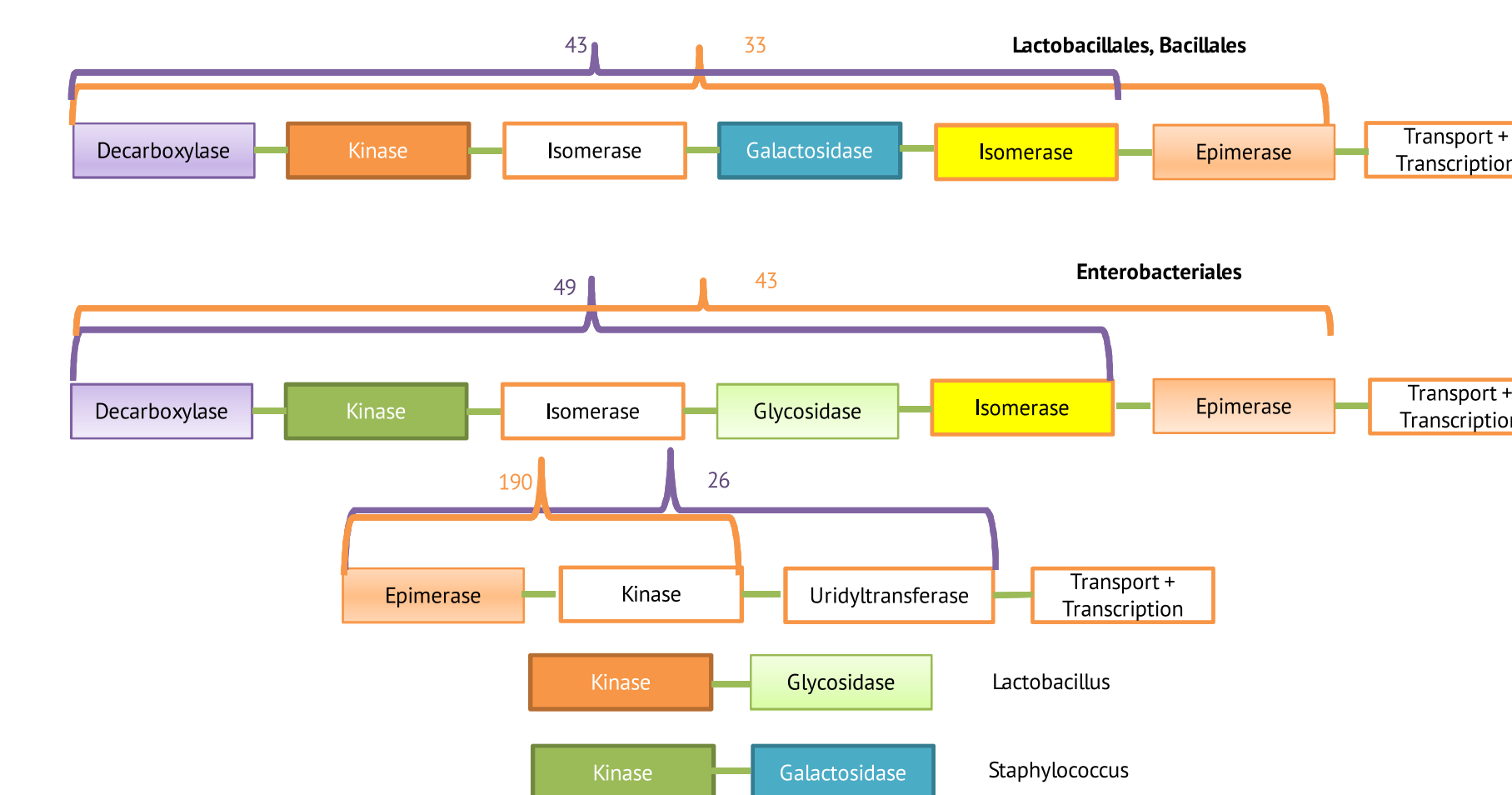


Рис. 4 Комбинации генов в исследуемых кассетах

Fig. 4 Gene combinations in studied cassettes

Introduction

Lactose is a common disaccharide that consists of glucose and galactose monomers. There are a few well-described bacterial metabolic pathways of lactose degradation. Different bacteria can consume lactose using different sets of genes.

It is a known fact that genes from the same metabolic pathway are often located together on the bacterial chromosome. Analysis of such cassettes allows to study evolution, stable combinations of such genes and reveal horizontal transfer cases.

Aims

- ▶ Analyze all gene cassettes related to lactose metabolism
- ▶ Obtain conservative combinations
- ▶ Check whether cassettes divide into lactose (transport and hydrolysis of disaccharide) and galactose (further reactions). See hypothesis 1.
- ▶ Find cases of horizontal transfer of conservative cassettes.

Hypothesis 1:

On the first stage of lactose degradation lactose is hydrolysed, and glucose and galactose are obtained. Both of them are common nutrients and could be transported into cell separately or obtained after degradation of different kinds of carbohydrates and their derivatives. Thus we suggest that genes providing hydrolysis and genes providing further reactions concerning galactose could be often located on different cassettes.

Discussion

More than ¾ of known bacteria seem to have ability to consume lactose. We make this conclusion based on the fact that they possess galactosidase genes. Many bacteria use several pathways of lactose degradation simultaneously. Most bacteria that consume lactose do not contain long conservative cassettes of respective genes. Our studies show that galactosidase genes are usually located apart from other genes coding other enzymes from lactose degradation pathways (more than 85% cases). Sometimes they are located near galactokinases, but only in 5% cases.

We have found one long conservative cassette corresponding to a well-known lactose degradation pathway of Bacillales and Lactobacillales. Interestingly, most Enterobacteriales seem to possess all those genes, but they are scattered on the chromosome.

We have also observed a horizontal transfer case, where the whole cassette was transferred into *Leptotrichia buccalis*, which also live inside human organism. A few annotation mistakes were spotted, including a common gene from many bacteria genomes previously described as a transcriptional factor from DeoR family. After analyzing its domain structure we came to a conclusion that it is an isomerase.

Hypothesis 2:

We suggest that a conservative cassette found in most *E.coli* strains and several other Enterobacteriales involved in metabolism of sulphoquinovose (6-deoxy-6-sulphoglucose), described by Denger et al., can also participate in lactose or galactose degradation.

Arguments in favor of this hypothesis:

- ▶ Full compliance of basic functions of enzymes from this cassette with the functions of the mentioned above enzymes from the common metabolic pathway of lactose degradation in Bacilli.
- ▶ Matching of one of the enzymes (aldolase) in both pathways
- ▶ Specificity of the epimerase present in this cassette and in the cassettes of the well-known pathway of galactose degradation.
- ▶ Presence of the binding site of factor CRP in the promoter region*
- ▶ The ability of some genes from both pathways to be mixed in other kinds of cassettes

*Investigation of genes by phylogenetic footprinting revealed the overlap of CRP transcription factor binding site and promoter. Apparently factor CRP functions as a repressor suppressing gene expression cassette when the glucose concentration

Methods and Materials

- 300k genes of carbohydrate metabolism assembled into 55k cassettes
- Metacyc – metabolic pathway database
- IMG JGI (COG, Pfam)
- Enzyme classification database
- Clastal Omega – alignment of nucleotide sequences
- Phylogenetic footprinting – search tool for transcription regulation sites
- PHYRE2
- 3D protein structure study, search for domains
- iTOL – construction of bacterial phylogenetic trees