



# Сладкие Парочки: Любовь и Равнодушие

## Sweet Couples: Love and Indifference

### Anastasia Zakharova, Anna Kaznadzey

#### Abstract

Carbohydrate metabolism genes are often located together on bacterial chromosomes forming combinations that here will be called cassettes. Many cassettes contain several genes of similar functions, which often can't be explained by respective metabolic pathway characteristics. Our goal was to analyze similar function genes co-localization tendencies in comparison with random distribution pairwise and study how such "sweet couples" behave in different genomes, to find out if generally they become closer or farther away on the chromosomes with time. We chose glycosidases and glycosyltransferases as subject functions. Glycosyltransferase and glycosidase gene families generally do not tend to be co-localized with other carbohydrate metabolism gene families, but almost all of them prefer to be co-localized with members of the same family. Numerous examples of pair formation and breakdown were observed. At that, co-localized ancestor pairs are generally younger compared to unpaired ancestors. We suggest that the tendency to co-localization is higher than the breakdown tendency.

#### Введение

Гены углеводного метаболизма часто располагаются рядом на бактериальных хромосомах, формируя комбинации, которые здесь мы будем называть кассетами. Многие кассеты содержат одновременно несколько генов похожих функций, что часто нельзя объяснить с точки зрения особенностей соответствующих метаболических путей. Нашей задачей было попарно проанализировать тенденцию генов сходных функций к ко-локализации в сравнении со случайным распределением, а также исследовать как такие "сладкие парочки" ведут себя в различных геномах и понять, сходятся они с течением времени или расходятся. В качестве объекта исследования мы выбрали гликозидазы и гликозилтрансферазы.

#### Материалы:

##### Programs:

developed *ad hoc* scripts on Python  
pruneTree (for tree parsing) by Zoya Chervontseva  
treehmmweb (for ancestral conditions/events probability calculation) by Nadya Bykova

**Phylogenetic Tree:** allBacteriaTree from microbesonline

**Random distributions:** tables obtained with help of Pavel Shelyakin

**Carbohydrate metabolism genes cassette database**

**Nucleotide search:** nsmscan tool

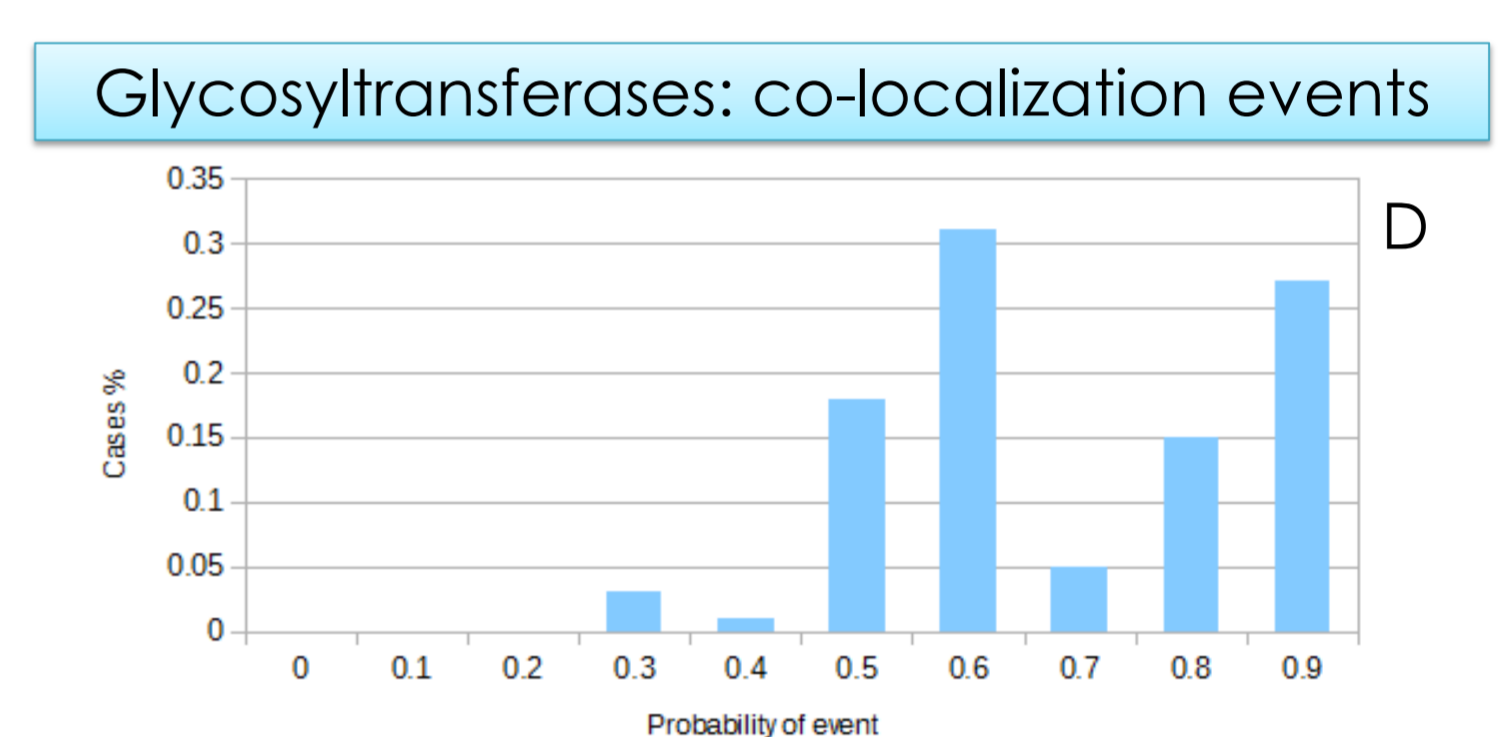
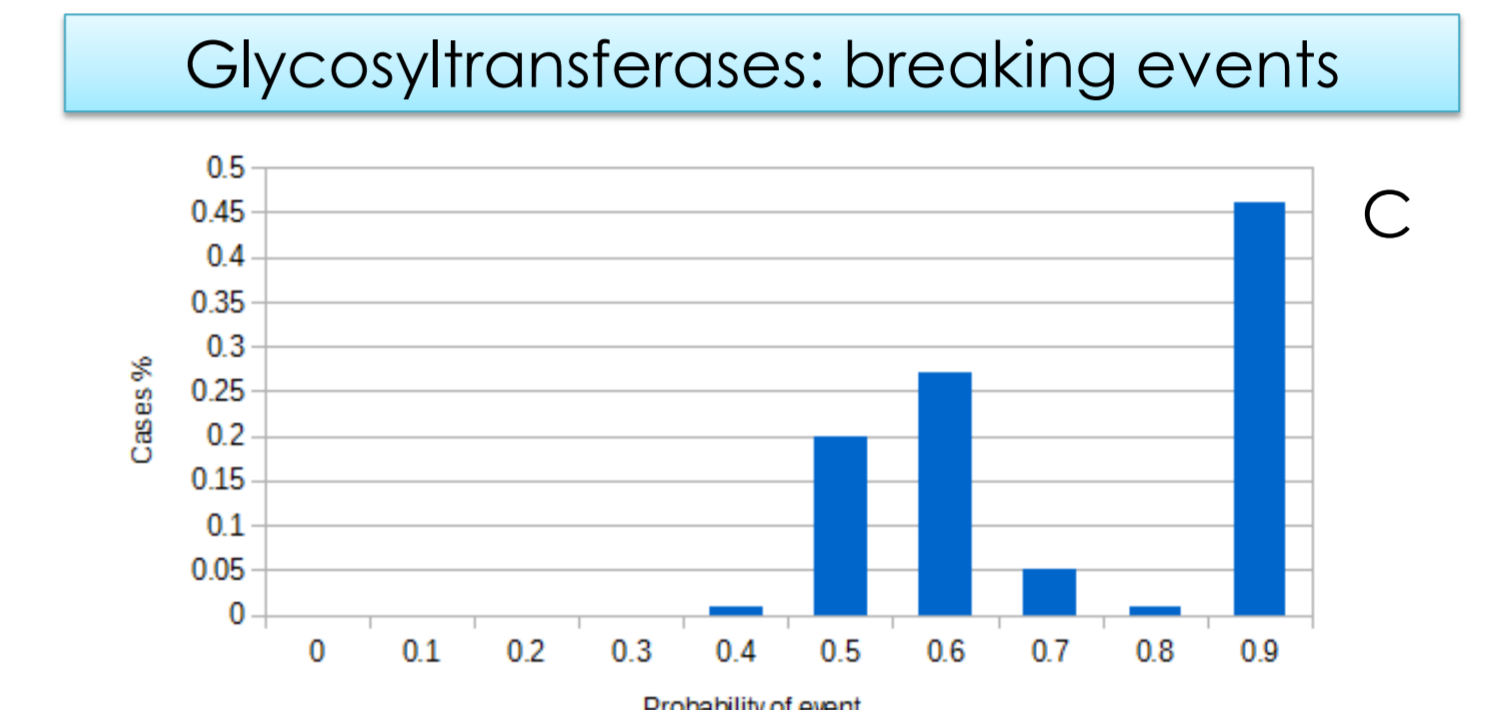
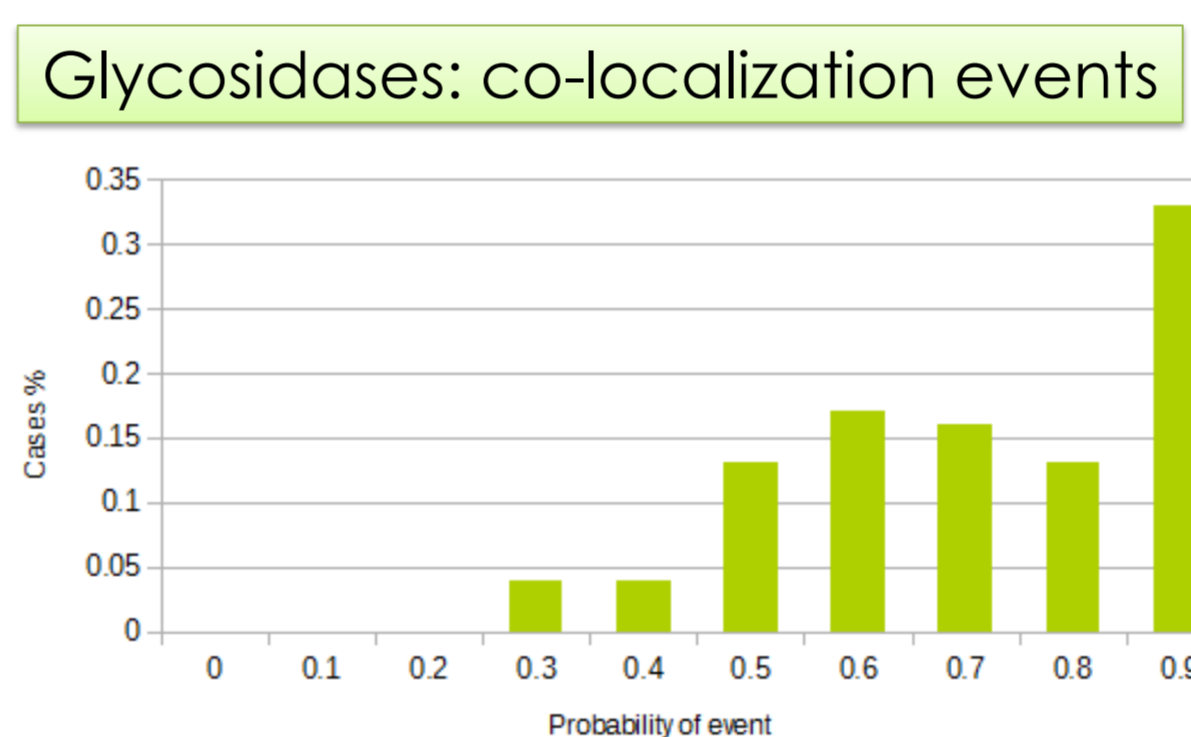
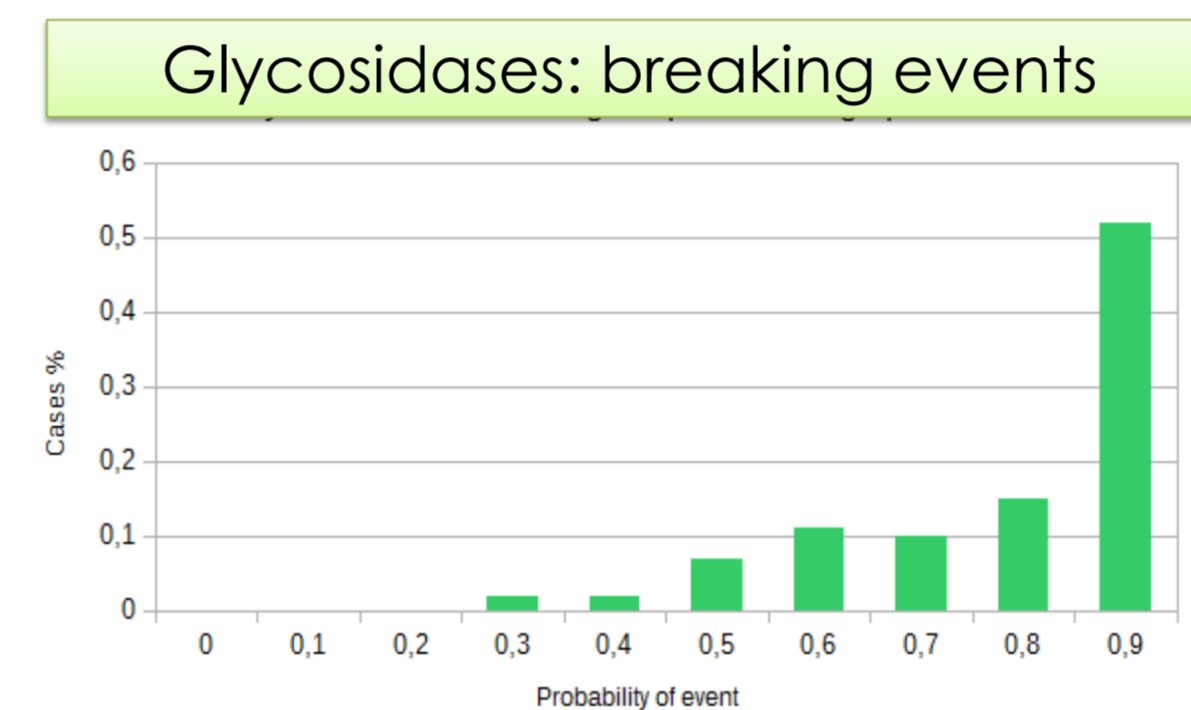
##### Методы:

Значимость ко-локализации семейств генов была установлена путем сравнения со случайным распределением и вычислением P-value. Поиск родственных генов в бактериальных геномах осуществлялся с помощью программы nsmscan (условия: более 60% сходства более 30% длины гена). Данные о событиях слияния и разбиения кассет и о среднем возрасте соответствующих предков были получены с помощью анализа вероятности событий перехода между состояниями и вероятности состояний узлов на филогенетических деревьях бактерий, содержащих исследуемые пары генов. Для этой работы был разработан ряд программ на Python.

## Results

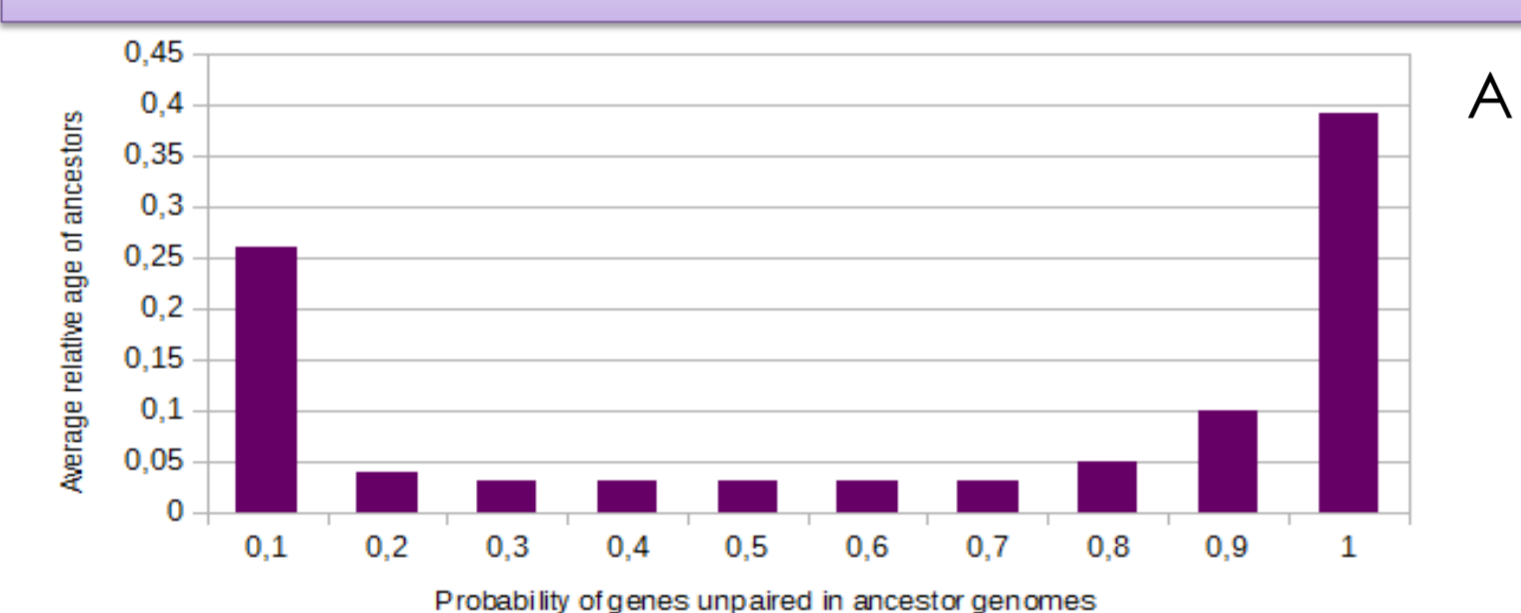
Function	COG	Gene number in cassette	Cassette number	Average random cassette number	Standard deviation	P-value	Significance
glycosyltransferase	COG2236	1	31	66.2937	5.7657818472	1.00E-004	No
glycosyltransferase	COG0438	1	1881	2771.0125	37.7215766339	1.00E-004	No
glycosyltransferase	COG0463	1	1434	1666.8884	29.0232207971	1.00E-004	No
glycosyltransferase	COG1819	1	114	196.0891	9.6110645191	1.00E-004	No
glycosyltransferase	COG1442	1	50	79.0429	5.9514586103	1.00E-004	No
glycosyltransferase	COG1215	1	689	714.7328	18.3589488849	0.078192181	No
glycosyltransferase	COG0380	1	139	101.1349	6.8356493466	1.00E-004	Yes
glycosyltransferase	COG0297	1	201	166.3851	8.6211135006	1.00E-004	Yes
glycosyltransferase	COG3754	1	60	43.651	4.4204523524	0.00019998	Yes
glycosyltransferase	COG1216	1	755	696.064	18.3942193093	0.00089991	Yes
glycosyltransferase	COG2236	2	1	0.0254	0.15797101	0.00019998	Yes
glycosyltransferase	COG0438	2	842	185.3919	12.6410013207	1.00E-004	Yes
glycosyltransferase	COG0463	2	235	63.9089	7.5006000287	1.00E-004	Yes
glycosyltransferase	COG0297	2	4	0.3514	0.5878078257	1.00E-004	Yes
glycosyltransferase	COG1819	2	21	1.5887	1.2417456704	1.00E-004	Yes
glycosyltransferase	COG3754	2	8	0.1411	0.3762855166	1.00E-004	Yes
glycosyltransferase	COG1442	2	13	0.9066	0.9202589894	1.00E-004	Yes
glycosyltransferase	COG1215	2	36	10.8688	3.2181029443	1.00E-004	Yes
glycosyltransferase	COG1216	2	129	13.0837	3.5360846017	1.00E-004	Yes

Glycosyltransferase and glycosidase genes are highly abundant among carbohydrate metabolism cassettes. After comparison with the random distribution and selecting cases passing a strong P-value filter, we see that 60% of glycosyltransferase families and 84% of glycosidase families do not have a preference towards having specific carbohydrate metabolism neighbors when they are located within cassettes as single representatives; however 84% glycosidase families and 79% glycosyltransferase families have a strong preference towards having a same-family neighbor within the same cassette.

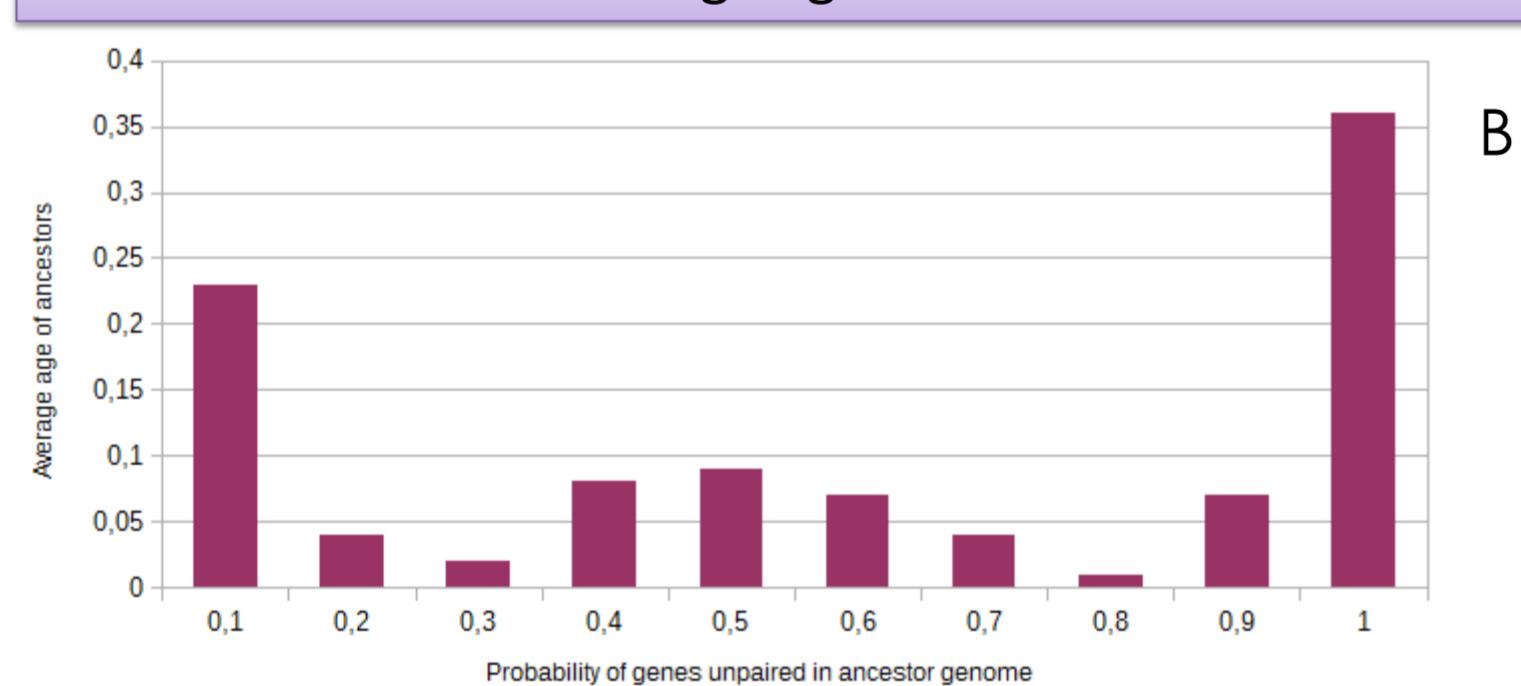


Events of pair co-localization and breaking apart are shown both for glycosidase (A, B) and glycosyltransferase (C, D) genes.

#### Glycosidases: correlation of ancestor condition with average age of ancestors



#### Glycosyltransferases: correlation of ancestor condition with average age of ancestors

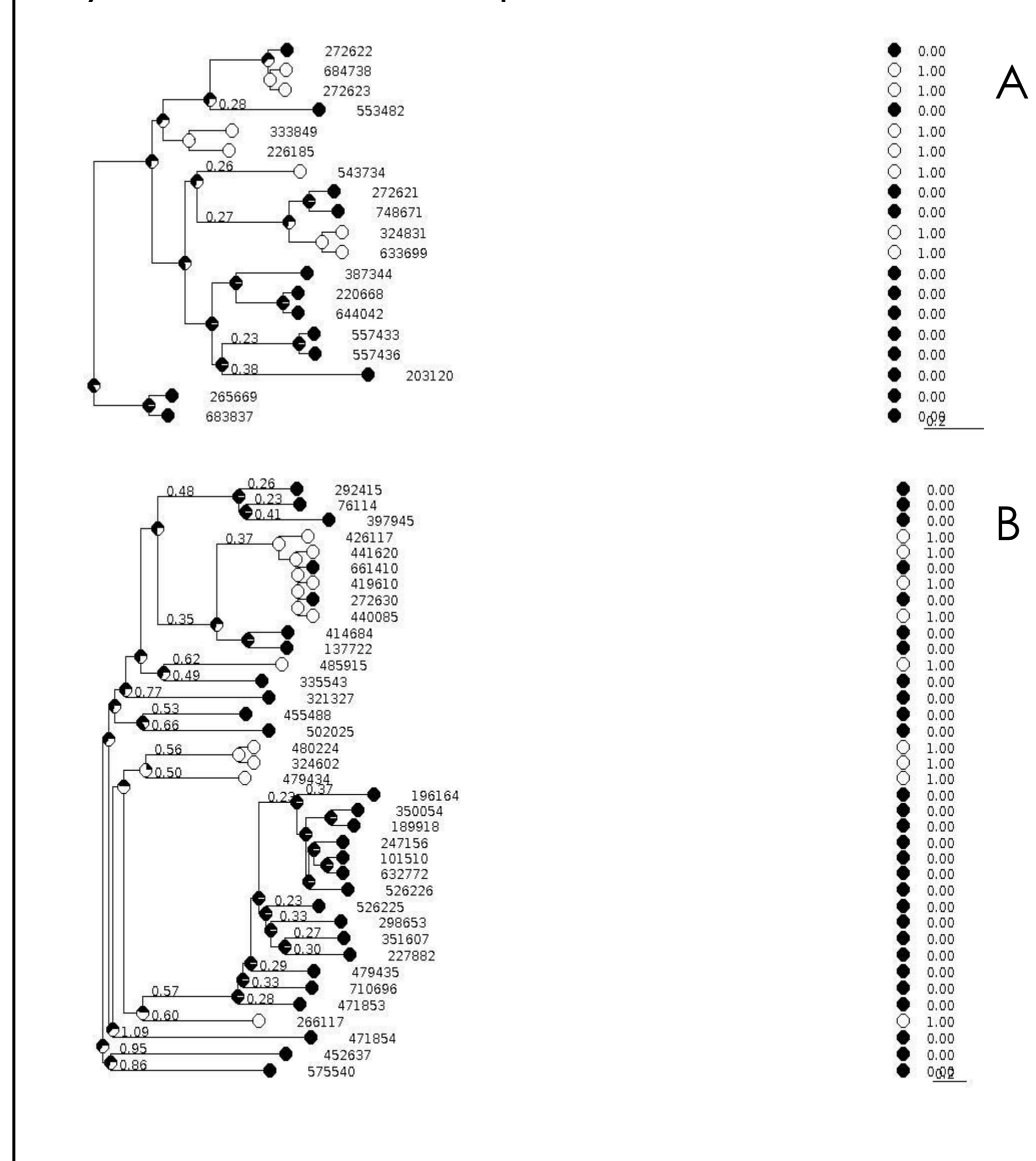


Ancestors with co-localized genes are generally younger, and the same pattern can be seen both for glycosidases (A) and glycosyltransferases (B).

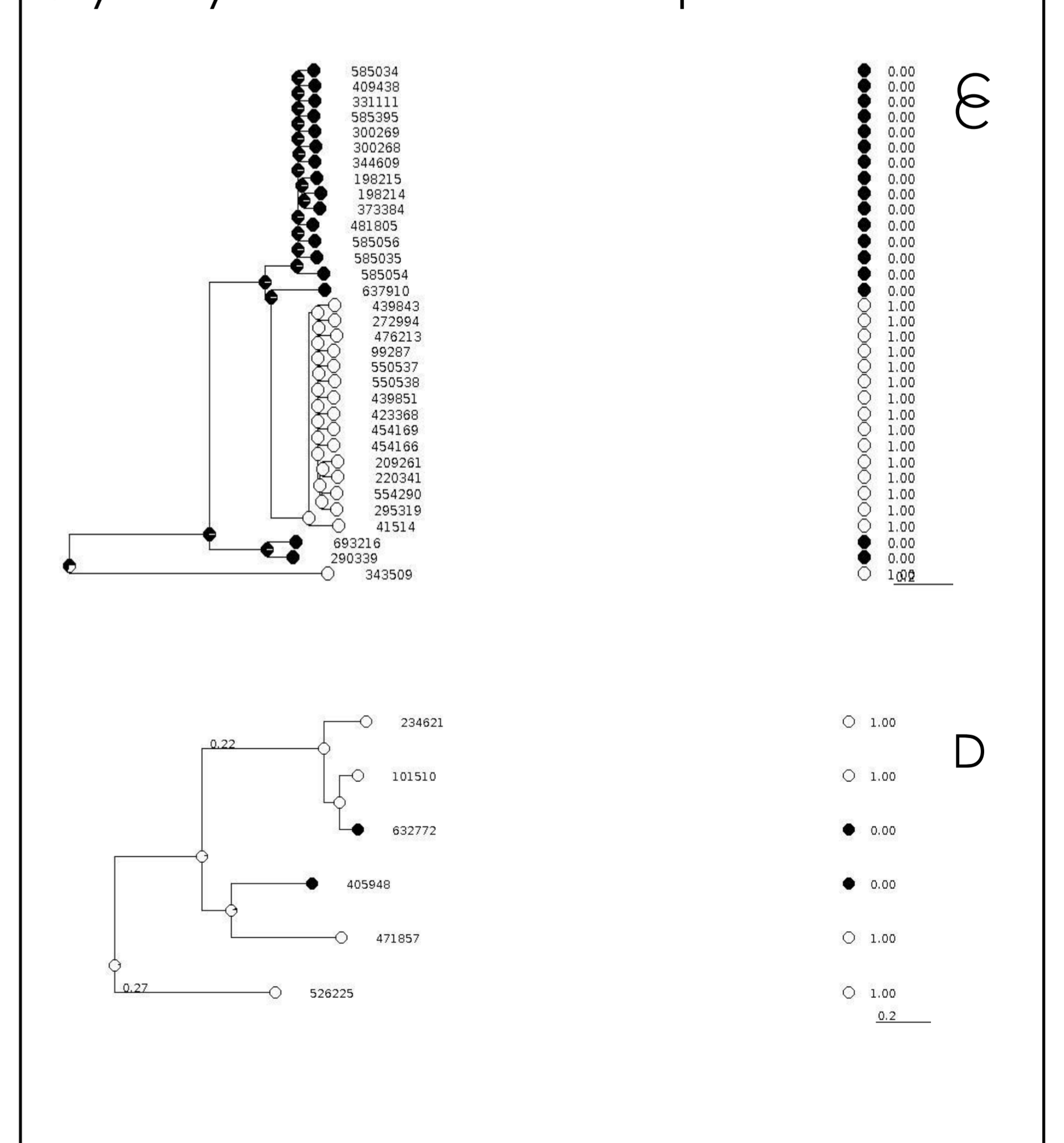
#### Выводы:

Большинство семейств генов гликозилтрансфераз и гликозидаз не имеют специфических предпочтений к ко-локализации с другими семействами генов углеводного метаболизма. Однако практически все они демонстрируют высокое предпочтение к ко-локализации генов, принадлежащих к одному и тому же семейству. Были обнаружены достоверные события как слияния, так и распада пар исследуемых генов на геномах бактерий с течением времени. С помощью реконструкции предков мы выяснили, что ко-локализованные предки - более "молодые", чем не ко-локализованные, что позволяет предположить наличие эволюционной тенденции, направленной на ко-локализацию генов как гликозилтрансфераз, так и гликозидаз.

#### Glycosidase example trees



#### Glycosyltransferase example trees



Examples of phylogenetic trees for glycosidase (A, B) and glycosyltransferase (C, D) gene pairs. [1] condition (white) stands for a co-localized pair, [0] condition (black) for a broken pair.

A. Four possible events of co-localization of a glycosidase pair

B. Four possible events of co-localization and at least one possible event of breakdown of a glycosidase pair

C. Two possible events of co-localization of a glycosyltransferase pair

D. Two possible events of breakdown of a glycosyltransferase pair