

Для анализа были выбраны виды:

*Candidatus Methanoplasma termitum*

*Cuniculiplasma divulgatum*

*Ferroplasma acidarmanus fer1*

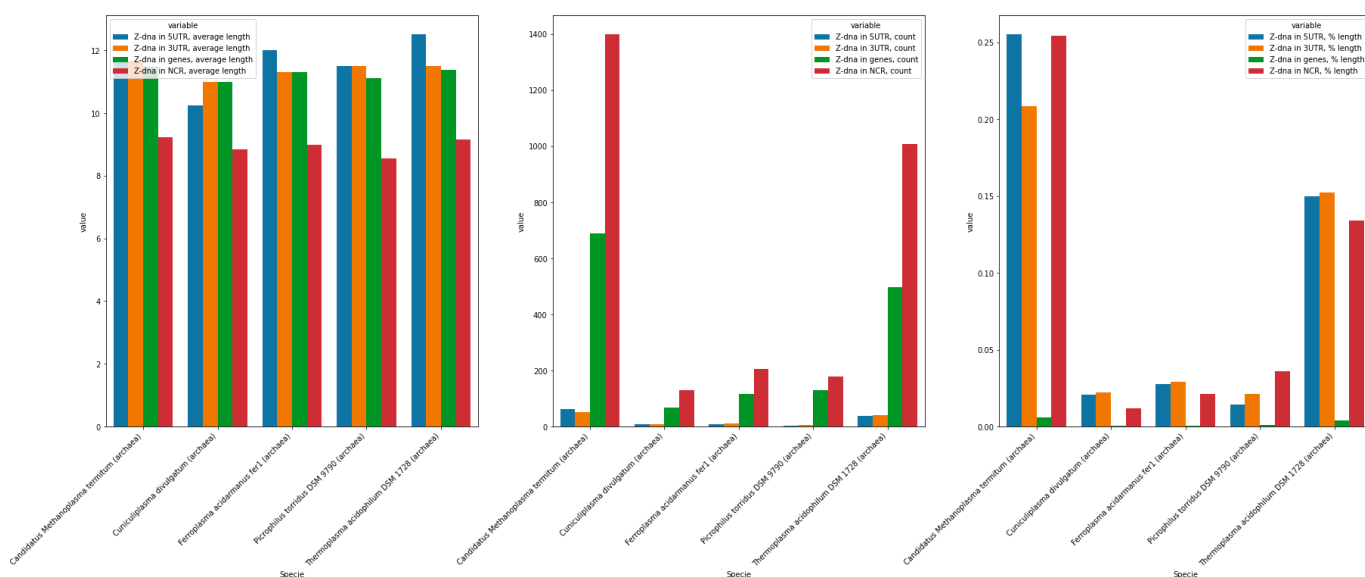
*Picrophilus torridus* DSM 9790

*Thermoplasma acidophilum* DSM 1728

Средствами edirect ыли скачаны геномы, аннотации и белок-кодирующие последовательности.

С помощью zhunt3 проведен поиск и анализ положения z-формы ДНК.

Было найдено множество предполагаемых участков, однако у большинства score был довольно низок. В анализе рассматривались только участки со скором, превосходящим 500. Подробные данные по количеству и длине участков представлены в файле `result_table.tsv`. По количеству участков что-то сказать сложно – все-таки очень не равна длина категорий. Однако если взглянуть на график ниже, то можно заметить, что средняя длина участков внутри не кодирующих последовательностей меньше всего, при этом везде, что косвенно указывает на случайность (однако для более точной проверки нужны более тонкие подходы). Количество участков, однако, наибольшее в межгенниках, а потом в генах, что, скорее всего, объясняется высокой суммарной длиной. А вот доля, занимаемая z-формой в кодирующих последовательностях низка – возможно, она мешает транскрипции. А вот в UTR ее наоборот очень много – возможно, она несет некоторые функции.



О консервативности что-то сказать довольно сложно – даже в очень консервативных кластерах стабильно z-форма не обнаруживалась, что заставляет нас усомниться в ее необходимости во всех типах просмотров. Однако ее высокое количество в UTR косвенно может указывать на ее функциональность – для проверки нужен более тонкий инструментарий, чем тот, которым я располагаю. Один из примеров эволюционной (не)консервативности ниже. Еще больше в репозитории

