Для поиска метаболических путей с участием alkaline phosphatase, была использована программа gapseq <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02295-1> (2021).

Для построения моделей метаболической сети в масштабе генома gapseq использует модифицированную базу данных биохимии ModelSEED; добавлена часть реакций, и часть удалена; база данных автоматически обновляется.

Предсказание путей, транспортеров и комплексов основано на базе данных последовательностей белков, полученных из UniProt и TCDB.

Во всех исследуемых штаммах *Cobetia* был обнаружен метаболический путь с участием alkaline phosphatase – |NADPHOS-DEPHOS-PWY|. В базе ModelSEED обнаружен единственный ему соответствующий путь для *E. coli* [https://biocyc-curation.ai.sri.com/E COLI/NEW-IMAGE?type=PATHWAY&object=NADPHOS-DEPHOS-PWY&EXP-ONLY=NIL](https://biocyc-curation.ai.sri.com/ECOLI/NEW-IMAGE?type=PATHWAY&object=NADPHOS-DEPHOS-PWY&EXP-ONLY=NIL)

Найденная с помощью gapseq реакция - RXN-5822 в базе данных Biocyc -

[https://biocyc-curation.ai.sri.com/E COLI/NEW-IMAGE?object=RXN-5822&redirect=T](https://biocyc-curation.ai.sri.com/ECOLI/NEW-IMAGE?object=RXN-5822&redirect=T)

c участием [phoA](https://biocyc-curation.ai.sri.com/gene?orgid=ECOLI&id=EG10727) (*E. coli*) <https://www.uniprot.org/uniprotkb/P00634/entry>

|  |  |  |
| --- | --- | --- |
| X04586 (**EMBL** | **GenBank** | **DDBJ**) | CAA28257.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| M13345 (**EMBL** | **GenBank** | **DDBJ**) | AAA83893.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| M29664 (**EMBL** | **GenBank** | **DDBJ**) | AAA24364.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| M29665 (**EMBL** | **GenBank** | **DDBJ**) | AAA24365.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| U73857 (**EMBL** | **GenBank** | **DDBJ**) | AAB18107.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| U00096 (**EMBL** | **GenBank** | **DDBJ**) | AAC73486.2 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| AP009048 (**EMBL** | **GenBank** | **DDBJ**) | BAE76164.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| M33536 (**EMBL** | **GenBank** | **DDBJ**) | AAA24372.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| J01659 (**EMBL** | **GenBank** | **DDBJ**) | AAA24359.2 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| J01660 (**EMBL** | **GenBank** | **DDBJ**) | AAA24360.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| J01661 (**EMBL** | **GenBank** | **DDBJ**) | AAA24361.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |
| J05005 (**EMBL** | **GenBank** | **DDBJ**) | AAA24362.1 (**EMBL** | **GenBank** | **DDBJ**) | Genomic DNA |

Проведено мульти выравнивания белков из таблицы за исключением коротких последовательностей с KGA01942 с помощью Clustal, файлы: E. coli\_and\_KGA01942.txt, Mview \_ E. coli\_and\_KGA01942.pdf, Clustal Omega \_ E. coli\_and\_KGA01942.pdf

Мульти выравнивания показало, что последовательности phoA *E. coli* идентичны между собой, но существенно отличаются от phoA *Cobetia amphilecti* (KGA01942) ([Per. Ident](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&ADV_VIEW=yes&ADV_VIEW=on&ALIGNMENTS=100&ALIGNMENT_VIEW=Pairwise&BLAST_SPEC=blast2seq&CONFIG_DESCR=ClustMemNbr,ClustComn,Ds,Sc,Ms,Ts,Cov,Eval,Idnt,AccLen,Acc&DATABASE_SORT=0&DESCRIPTIONS=100&DYNAMIC_FORMAT=on&FIRST_QUERY_NUM=0&FORMAT_OBJECT=Alignment&FORMAT_PAGE_TARGET=&FORMAT_TYPE=HTML&GET_SEQUENCE=yes&I_THRESH=&LINE_LENGTH=60&MASK_CHAR=2&MASK_COLOR=1&NEW_VIEW=yes&NUM_OVERVIEW=100&PAGE=Proteins&QUERY_INDEX=0&QUERY_NUMBER=0&RESULTS_PAGE_TARGET=&RID=BRYW1UD6114&SHOW_LINKOUT=yes&SHOW_OVERVIEW=yes&STEP_NUMBER=&WORD_SIZE=3&ADV_VIEW=on&DISPLAY_SORT=3&HSP_SORT=3) 32.21%, [Cover](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&ADV_VIEW=yes&ADV_VIEW=on&ALIGNMENTS=100&ALIGNMENT_VIEW=Pairwise&BLAST_SPEC=blast2seq&CONFIG_DESCR=ClustMemNbr,ClustComn,Ds,Sc,Ms,Ts,Cov,Eval,Idnt,AccLen,Acc&DATABASE_SORT=0&DESCRIPTIONS=100&DYNAMIC_FORMAT=on&FIRST_QUERY_NUM=0&FORMAT_OBJECT=Alignment&FORMAT_PAGE_TARGET=&FORMAT_TYPE=HTML&GET_SEQUENCE=yes&I_THRESH=&LINE_LENGTH=60&MASK_CHAR=2&MASK_COLOR=1&NEW_VIEW=yes&NUM_OVERVIEW=100&PAGE=Proteins&QUERY_INDEX=0&QUERY_NUMBER=0&RESULTS_PAGE_TARGET=&RID=BRYW1UD6114&SHOW_LINKOUT=yes&SHOW_OVERVIEW=yes&STEP_NUMBER=&WORD_SIZE=3&ADV_VIEW=on&DISPLAY_SORT=4&HSP_SORT=0) 61%)

Дальнейший анализ результатов gapseq показал, что gapseq ассоциирует с путем |NADPHOS-DEPHOS-PWY| не phoA (alkaline phosphatase), а ген **surE - нуклеотидаза, проявляющая фосфатазную активность в отношении нуклеозид-5'-монофосфатов.**

**Дальнейшие поиски поэтому проведены самостоятельно с помощью clusltal и blast против искомого фермента** phoA Cobetia (KGA01942).

С помощью скриптов R, bash проанализировали gtf всех исследуемых штаммов *Cobetia*.

С помощью igv browser в gtf ряда штаммов был обнаружен безымянный ген с продуктом alkaline phosphatase, находящийся рядом с геном metG (+/- 2 позиции).

Поэтому решили выделить регион, центрировав по metG. По результатам анализа создана таблица ALP\_blast.xlsx (лист ALP neighbors).

В каждой ячейке указан ген и продукт в формате «gene|product», например: «metG |methionine--tRNA ligase». Ген KGA01942 в таблице присутствует как «NA |alkaline phosphatase», потому что название гена в карточке не обозначено. В таблице подкрашено только ближайшее окружение metG, однако можно заметить сходство соседей на протяженном участке для различных штаммов.

В результатах, приведенных в таблице ALP\_blast.xlsx (лист ALP neighbors) необходимо отметить следующие особенности:

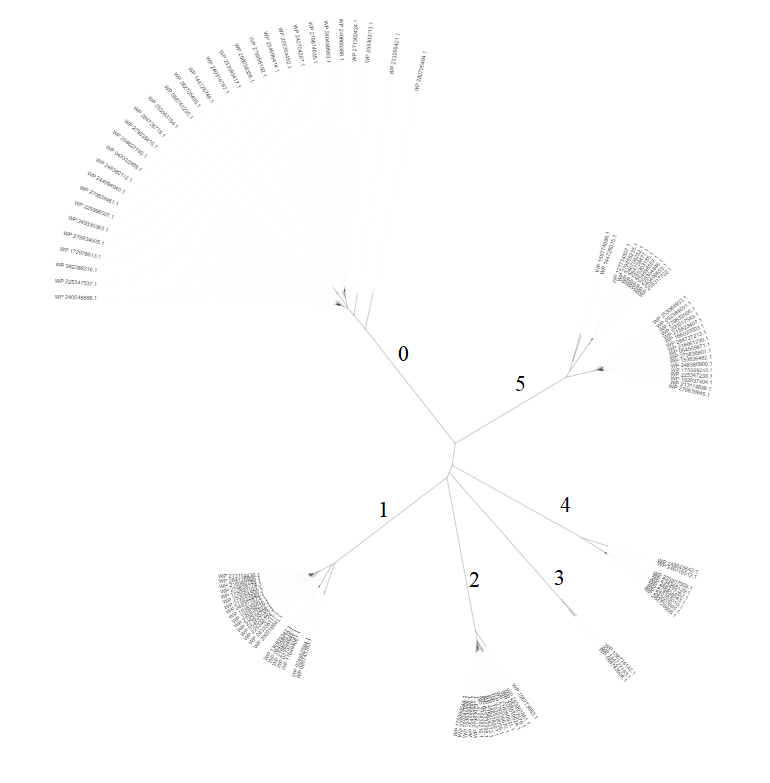
* У штаммов Cobetia\_sp\_UCD-24C\_GCF\_001306765.1, Cobetia\_crustatorum\_JO1T\_GCF\_000591415.1 KGA01942 лежит не по соседству metG, нужно привести регион расположения гомолога.
* У штаммов Cobetia\_sp\_UCD-24C\_GCF\_001306765.1, Cobetia\_crustatorum\_JO1T\_GCF\_000591415.1 KGA01942, Cobetia\_marina\_NBRC\_15607\_GCF\_006540105.1, Cobetia\_sp\_10Alg\_146\_GCF\_029846385.1 окружение metG сильно отличается.

С помощью скриптов R, bash проанализировали gtf всех исследуемых штаммов *Cobetia*.

Создали список генов, которые описаны в gtf как alkaline phosphatase, таблица ALP\_blast.xlsx.

По найденным protein\_id выполнили мульти выравнивание, построили филогенетическое дерево с помощью Clustal и [ITOL](file:///C:\Users\1\Desktop\vl2\ITOL), соответственно.

<https://itol.embl.de/tree/21216438178297041689965799>



На дереве можно видеть 6 ветвей. В ALP\_blast.xlsx обозначены в branch соответствующими номерами (0-5).

В gtf файле есть 3 вида описания alkaline phosphatase: “alkaline phosphatase”, “alkaline phosphatase D family protein”, “alkaline phosphatase family protein”.

Все гомологи KGA01942 - это branch 4, они в файле gtf отмечены как “alkaline phosphatase”. Белки с описанием “alkaline phosphatase D family protein” относятся к branch 0 (неклассифицированные) или к branch 5. Остальные – относятся к “alkaline phosphatase family protein”, которые попали в branch 0, 1, 2, 3, 5, но не 4. У всех исследуемых штаммов больше одного гена alkaline phosphatase разных типов (2-5).

Результаты этого исследования согласуются с нашим предыдущим анализом, где штаммы, содержащие гомологи KGA01942, это:

|  |  |
| --- | --- |
| Strain | Mismatch |
| Cobetia\_amphilecti\_B2M13\_GCF\_018860945.1 | 240F>Y, 308F>Y |
| Cobetia\_amphilecti\_NRIC\_0815T\_GCA\_030010415.1 | 240F>Y, 308F>Y, 372E>G, 399A>T |
| Cobetia\_crustatorum\_JO1T\_GCF\_000591415.1 | Down start +45p, отсутствует ключевая позиция 21D, Tblastn\_Identities 413/466 (89%) |
| Cobetia\_crustatorum\_SM1923\_GCF\_007786215.1 | Tblastn\_Identities 456/511 (89%) |
| Cobetia\_sp\_1AS1\_GCF\_029846435.1 | 240F>Y, 308F>Y |
| Cobetia\_sp\_1CM21F\_GCF\_023161745.1 | 240F>Y, 308F>Y |
| Cobetia\_sp\_AM6\_GCF\_009617955.1 | 240F>Y, 308F>Y, 419A>V, 422P>L |
| Cobetia\_sp\_UCD-24C\_GCF\_001306765.1 | 240F>Y, 308F>Y, 390G>D |

У всех исследуемых штаммов больше одного гена alkaline phosphatase разных типов (2-5).

Относительно вопроса: является ли KGA01942 маркерным геном *Cobetia amphilecti*, необходимо отметить, что штамм Cobetia\_amphilecti\_N-80\_GCF\_020217465.1 не содержит этого гомолога ни в геноме, ни в окрестности metG. Но штаммы, которые Вы предполагаете, что являются *Cobetia amphilecti* (см. в таблице) его содержат.