05-05-25

Tomé mi archivo fasta (Delta.fasta) y lo traté de analiza con el servicio web de BLASTn, sin embargo, la secuencia es demasiado pesada (sobre los 5mb) y supera por mucho el máximo permitido.

Generé un script de python que fragmenta el fasta original en secuencias de 5120 bases, se generaron 1074 fragmentos. Usé 15 fragmentos aleatorios para hacer blastn contra ellos.

**Puesto 1:** >Delta\_fragmento\_319

**Puesto 2:** >Delta\_fragmento\_899

**Puesto 3:** >Delta\_fragmento\_102

**Puesto 4:** >Delta\_fragmento\_109

**Puesto 5:** >Delta\_fragmento\_172

**Puesto 6:** >Delta\_fragmento\_242

**Puesto 7:** >Delta\_fragmento\_829

**Puesto 8:** >Delta\_fragmento\_926

**Puesto 9:** >Delta\_fragmento\_482

**Puesto 10:** >Delta\_fragmento\_94

**Puesto 11:** >Delta\_fragmento\_735

**Puesto 12:** >Delta\_fragmento\_754

**Puesto 13:** >Delta\_fragmento\_179

**Puesto 14:** >Delta\_fragmento\_549

**Puesto 15:** >Delta\_fragmento\_226

El resultado del blastn arrojó que todos los fragmentos tienen scores máximos, con e-values tendientes a 0, coberturas del query del 100% y prácticamente 100% de identidad con *Blautia pseudococcoides.*

**Próximos pasos:** Hacer predicción de genes y proteínas con Prodigal.

07-05-25

Realicé la predicción de genes y proteínas con Prokka debido a que es más específica que Prodigal (se puede indicar el género y especie a predecir). El comando usado fue:

prokka /home/sastorga/proyecto\_final\_bioinfo1/Delta.fasta --outdir /home/sastorga/proyecto\_final\_bioinfo1/prokka\_out --genus Blautia --species pseudococcoides --prefix blautia --cpus 4

Resumen del resultado:

organism: Blautia pseudococcoides strain

contigs: 1

bases: 5498201

CDS: 5135

rRNA: 10

repeat\_region: 2

tRNA: 68

tmRNA: 1

Total de filas (ignorando encabezado si lo tiene)

total=$(tail -n +2 tabla.txt | wc -l)

Filas con valor en segunda columna

con\_valor=$(awk -F'\t' 'NF >= 2 && $2 != ""' tabla.txt | tail -n +2 | wc -l)

echo "Con datos: $con\_valor de un total de $total filas" Con datos: 1512 de un total de 5216 filas

08-05-25

Usé eggNOG-mapper para hacer la anotación funcional. El comando en particular fue:

emapper.py --cpu 8 -i /home/sastorga/proyecto\_final\_bioinfo1/prokka\_out/blautia.faa --itype proteins -o blautia --output\_dir /home/sastorga/proyecto\_final\_bioinfo1/emapper\_out --tax\_scope 57251

16-05-25

Puedo usar barrnap para apoyar el taxa identificado con BLAST

[GitHub - tseemann/barrnap: :microscope: Bacterial ribosomal RNA predictor](https://github.com/tseemann/barrnap)

18-05-25

El comando usado para barrnap fue:

barrnap --kingdom bac --threads 4 --outseq /home/sastorga/proyecto\_final\_bioinfo1/barrnap/hits\_rrna.fasta < /home/sastorga/proyecto\_final\_bioinfo1/Delta.fasta > /home/sastorga/proyecto\_final\_bioinfo1/barrnap/hits\_rrna.gff

El resultado fue:

##gff-version 3

Delta barrnap:0.9 rRNA 696985 698514 0 + . Name=16S\_rRNA;product=16S ribosomal RNA

Delta barrnap:0.9 rRNA 699219 702096 0 + . Name=23S\_rRNA;product=23S ribosomal RNA

Delta barrnap:0.9 rRNA 2884192 2887069 0 - . Name=23S\_rRNA;product=23S ribosomal RNA

Delta barrnap:0.9 rRNA 2887774 2889303 0 - . Name=16S\_rRNA;product=16S ribosomal RNA

Delta barrnap:0.9 rRNA 2898734 2901611 0 - . Name=23S\_rRNA;product=23S ribosomal RNA

Delta barrnap:0.9 rRNA 2902389 2903918 0 - . Name=16S\_rRNA;product=16S ribosomal RNA

Delta barrnap:0.9 rRNA 4021053 4022582 0 + . Name=16S\_rRNA;product=16S ribosomal RNA

Delta barrnap:0.9 rRNA 4023287 4026164 0 + . Name=23S\_rRNA;product=23S ribosomal RNA

Delta barrnap:0.9 rRNA 4830343 4831872 0 + . Name=16S\_rRNA;product=16S ribosomal RNA

Delta barrnap:0.9 rRNA 4832525 4835402 0 + . Name=23S\_rRNA;product=23S ribosomal RNA

Las secuencias de estos hits fueron blasteadas contra la base de datos nr.

Todas las secuencias arrojaron resultados perfectos contra *Blautia pseudococcoides.*

Para apoyar esto también se buscó contra la base de datos SILVA usando la secuencia 16S predicha por barrnap. El resultado no arrojó nada.

04-06-25

Debo recopilar información bibliográfica de la especie para definir en que genes enfocarme.

\*KOALA y KAAS no entregan necesariamente la misma información que eggNOG por lo que no es redundante usar esas herramientas

11-06-25

KOALA: 44% del proteoma de entrada fue anotado ([resultado](https://www.kegg.jp/tool-bin/blastkoala_result?id=1f9e2042de8bf64ce1016fb0e96b17987dcfe1d7&passwd=wBmumm&type=blastkoala#graph_link)), las 3 categorías funcionales más frecuentes son: 333 proteínas (~15% total anotado) corresponden a procesos celulares y señalización; 325 (~14%) a procesamiento de información ambiental y 263 (~12%) a procesamiento de información genética. Si solo nos restringimos a pathways el top 3 corresponde a metabolismo de carbohidratos (236; ~10%), metabolismo de aminoácidos (124; ~5%) y biosíntesis de glicanos (79; 3,5%). Además 75 proteínas anotadas (3%) no fueron clasificadas.

KAAS: de 5135 proteínas, 2060 (40,12%) fueron anotadas ([KEGG Automatic Annotation Server](https://www.genome.jp/kaas-bin/kaas_main?mode=result&id=1749064954&key=fMfTvFTp)).

Al comparar KOALA y KAAS en KEGG-mapper, en KOALA se predicen 209 vías metabólicas y 39 módulos completos, mientras que en KAAS se predicen 236 vías metabólicas y también 39 módulos completos, sin embargo no son los mismos módulos. Por ejemplo, KAAS predice el módulo de resistencia a vancomicina completa pero KOALA no lo predice. Al tomar una proteína del módulo (CKJNFLKG\_01121) para contrastar con otra herramienta como InterproScan se predice como una proteína kinasa sensora de dos componentes, mientras que al analizar su arquitectura de dominios con CDD se haya un hit contra kdpD, un dominio kinasa sensor de histidina. Se sabe que VanS es un sensor de histidina kinasa transmembrana clave en la resistencia a vancomicina. Con la información que poseo solo soy capaz de decir que si es muy probable que se trate de un sistema de dos componentes que sensa en la membrana un elemento externo para regular la expresión de genes pero no se puede afirmar que este elemento sea vancomicina necesariamente.

Ahora analizando la salida de eggNOG-mapper hay 2608 proteínas a las que se les asignó un KO. En KEGGmapper se puede notar que el mayor grupo de proteínas asignadas corresponde (fuera de metabolic pathways)) a la categoría ‘biosíntesis de metabolitos secundarios’ (148)

[El resultado en kegg mapper para eggNOGmapper\_annotations se obtiene con el archivo [emapper\_kegg\_mapper.txt](https://uccl0-my.sharepoint.com/:t:/r/personal/sastorj_uc_cl/Documents/Semestre_1_doc/proyecto_final_bioinfo/emapper_out/emapper_kegg_mapper.txt?csf=1&web=1&e=DNnCzd)]

Obeservando aquellos módulos más completos dados por eggNOG-mapper obtengo que la glicolisis está completa, la biosíntesis de pirimidinas (glutamina + PRPP > UMP) también, en la **síntesis de cobalamina** faltan 3 pasos en la parte anaerobia (cobyrinate a,c-diamide) y está completa en su segunda parte hasta cobalamina (comparar con KOALA y KAAS), puede ser importante porque es la vitamina B12 que no es producida por animales, la **vía Shikimate está completa**, esta vía permite formar folatos y aminoácidos aromáticos y no es hallada en mamíferos (IMPORTANTE). La síntesis de tiamina también se halla completa. La biosíntesis de purinas está casi completa (PRPP + glutamina > IMP; falta 1). La **biosíntesis de lisina** está casi completa para la vía succinyl-DAP (falta uno; comparar con KAAS, KOALA), está completa para la vía DAP aminotransferase y también completa para la vía DAP dehydrogenase, este aminoácido también es esencial por lo que es importante. La **biosíntesis de isoleucina** también está completa (vía threonine => 2-oxobutanoate => isoleucine), este es un aminoácido esencial. La biosíntesis de metionina desde aspartato está casi completa (falta 1), este también es un aminoácido esencial

Debido a que [*B. pseudococcoides*](https://pmc.ncbi.nlm.nih.gov/articles/PMC8107234/) es una bacteria residente (o no) de la microbiota intestinal es interesante buscar aquellas vías que puedan estar relacionadas con la síntesis de metabolitos secundarios que puedan tener beneficios para el organismo.