

## Problem Set 3, Problem 4

COMPSCI 260

Work Date: 10/10/2018

Extension Due Date: Sun 10/14/2018 5pm

**Jihong Tang**

NetID: jt290

Due Date: Fri 10/12/2018 5pm

## Introduction to the problem

Problem 4 from problem set 3 is the application of the read-matching program, which is based on the code we have written in problem 3.

For question 4a, we are asked to briefly discuss the ten microbes we will focus on in the following questions.

For question 4b, we are asked to write python program to find the microbe prevalences for each patient given the patients and microbe genome fasta files.

For question 4c, we need to take a deeper thinking into the results coming from question 4b. We should check if some viruses will bring changes in our case to find out the reason of the illness.

## Problem 4a

I searched information about all ten microbes on the internet and using some papers to get the following results and the reference informations are cited in the last part of this file.

1. *Bacteroides ovatus*

*Bacteroides ovatus* is a commensal intestinal microbe in gut, which is the predominant species. In addition, it acts as a predominant cause for a systemic antibody response. *Bacteroides ovatus* is an anaerobic, non-spore-forming, nonmotile, and gram-negative rod.

2. *Bacteroides thetaiotaomicron*

*Bacteroides thetaiotaomicron* is one of the most common bacteria found in human gut flora and is also an opportunistic pathogen. It is considered commensal. Its complement of enzymes for hydrolysis of glycosidic bonds is among the largest known in prokaryotes, and it is thought to be capable of hydrolyzing most glycosidic bonds in biological polysaccharides.

3. *Bifidobacterium longum*

*Bifidobacterium longum* is a Gram-positive, catalase-negative, rod-shaped bacterium present in the human gastrointestinal tract and one of the 32 species that belong to the genus *Bifidobacterium*. *B. longum* is considered to be a scavenger,

possessing multiple catabolic pathways to use a large variety of nutrients to increase its competitiveness among the gut microbiota.

4. *Eubacterium rectale*

*Eubacterium rectale* has been isolated as normal flora from feces, rumen, and periodontal tissue. *Eubacterium* spp. are thought to play a beneficial role in maintaining the normal ecology of the large intestine, in part by producing chemicals like butyric acid which act to inhibit the growth of other bacteria. These organisms are occasionally isolated from wounds and abscesses and may be an opportunistic pathogen. This genus has also been isolated from sewage and soil.

5. *Lactobacillus acidophilus*

*Lactobacillus acidophilus* is a species of gram positive bacteria in the genus *Lactobacillus*. *L. acidophilus* is a homofermentative, microaerophilic species, fermenting sugars into lactic acid. *L. acidophilus* occurs naturally in the human and animal gastrointestinal tract and mouth. Some strains of *L. acidophilus* may be considered to have probiotic characteristics.

6. *Peptoniphilus timonensis*

*Peptoniphilus timonensis* is a Gram-positive, anaerobic, indole-positive coccus that was isolated from the stool of a healthy Senegalese patient as part of a culturomics study aiming at cultivating individually all species within human feces.

7. *Prevotella copri*

*Prevotella copri* is a gram-negative bacteria. The cells are anaerobic and do not form spores. The cells grow best at 37 degree Celsius and are rod shaped. *Prevotella copri* is an intestinal bacteria, and its anaerobic qualities allows it to grow successfully in the human gut and intestines. Recent studies have shown a correlation between *Prevotella copri* and the development of rheumatoid arthritis.

8. *Roseburia intestinalis*

*Roseburia intestinalis* is a saccharolytic, butyrate-producing bacterium first isolated from human faeces. It is anaerobic, gram-positive, slightly curved rod-shaped and motile by means of multiple subterminal flagella.

9. *Ruminococcus bromii*

*Ruminococcus* is a genus of bacteria in the class Clostridia. They are anaerobic, Gram-positive gut microbes. *R. bromii* has a pivotal role in fermentation of RS3 in the human large intestine, and that variation in the occurrence of this species and its close relatives may be a primary cause of variable energy recovery from this important component of the diet. Also, *R. bromii* possesses an exceptional ability to colonize and degrade starch particles.

#### 10. *Vibrio cholerae*

*Vibrio cholerae* is a Gram-negative, comma-shaped bacterium. The bacterium's natural habitat is brackish or saltwater. Some strains of *V. cholerae* cause the disease cholera. *V. cholerae* is a facultative anaerobe[1] and has a flagellum at one cell pole as well as pili. *V. cholerae* can undergo respiratory and fermentative metabolism.

## Problem 4b

### Program consideration

Before writing the program to achieve the task to work out the microbe prevalences, I consider several details in the program.

Firstly, in order to save the space, I do not store the position of the mapped reads.

Secondly, I considered the situation when the reverse-complement steps need to be done. From my perspective, I prefer reporting the position in a easy way, therefore, I do the reverse-complement step for each read in order to get the right position index in the genome sequence. I understand that it may cause a little more time since the total number of reads length is larger than the number of genome length. I tested this two methods in my code, and it seems that the difference is only 3 to 4 seconds. Therefore, I still choose to use the easy coordinate method. That is an easy start for question 4c as well.

### Program results

The code solution of the program I mentioned can be found in **infection\_investigator.py**, the detailed comments and the outputs can also be seen in the code file.

From my code, I got the following results about the different microbes' prevalence in different patients.

- Patient 1
  - The estimated prevalence of microbe *Bacteroides ovatus* is 19.51%.
  - The estimated prevalence of microbe *Bacteroides thetaiotaomicron* is 16.41%.
  - The estimated prevalence of microbe *Bifidobacterium longum* is 4.75%.
  - The estimated prevalence of microbe *Eubacterium rectale* is 3.00%.
  - The estimated prevalence of microbe *Lactobacillus acidophilus* is 6.65%.
  - The estimated prevalence of microbe *Peptoniphilus timonensis* is 13.15%.
  - The estimated prevalence of microbe *Prevotella copri* is 8.75%.
  - The estimated prevalence of microbe *Roseburia intestinalis* is 14.46%.
  - The estimated prevalence of microbe *Ruminococcus bromii* is 8.15%.

- The estimated prevalence of microbe *Vibrio cholerae* is 5.17%.
- Patient 2
  - The estimated prevalence of microbe *Bacteroides ovatus* is 21.45%.
  - The estimated prevalence of microbe *Bacteroides thetaiotaomicron* is 15.45%.
  - The estimated prevalence of microbe *Bifidobacterium longum* is 6.10%.
  - The estimated prevalence of microbe *Eubacterium rectale* is 4.35%.
  - The estimated prevalence of microbe *Lactobacillus acidophilus* is 5.85%.
  - The estimated prevalence of microbe *Peptoniphilus timonensis* is 13.15%.
  - The estimated prevalence of microbe *Prevotella copri* is 9.85%.
  - The estimated prevalence of microbe *Roseburia intestinalis* is 13.00%.
  - The estimated prevalence of microbe *Ruminococcus bromii* is 7.80%.
  - The estimated prevalence of microbe *Vibrio cholerae* is 0.05%.
- Patient 3
  - The estimated prevalence of microbe *Bacteroides ovatus* is 6.50%.
  - The estimated prevalence of microbe *Bacteroides thetaiotaomicron* is 3.75%.
  - The estimated prevalence of microbe *Bifidobacterium longum* is 5.15%.
  - The estimated prevalence of microbe *Eubacterium rectale* is 4.05%.
  - The estimated prevalence of microbe *Lactobacillus acidophilus* is 1.30%.
  - The estimated prevalence of microbe *Peptoniphilus timonensis* is 7.65%.
  - The estimated prevalence of microbe *Prevotella copri* is 12.95%.
  - The estimated prevalence of microbe *Roseburia intestinalis* is 4.35%.
  - The estimated prevalence of microbe *Ruminococcus bromii* is 4.65%.
  - The estimated prevalence of microbe *Vibrio cholerae* is 49.65%.

## Result analysis

The total time to run my code including the code for problem 4c is about 67 seconds. I believe it is acceptable for the large amount of data.

As we can see from the program results, patient 3 has a marked relevant prevalence for *Vibrio cholerae*. 49.65% is definitely a large proportion. Therefore, as a world-leading gastroenterologic scientist, I believe that the microbe *Vibrio cholerae* accounts for patient 3's complaint, and patient 3 may catch the disease cholera. In addition, the cure treatment should focus on the method to deal with the *Vibrio cholerae*.

Looking at the results for patient 1 and patient 2, all microbe prevalences are similar to each other in different patients respectively except *Vibrio cholerae*. Therefore, something must be connected! I believe this is a good start for the problem 4c.

## Problem 4c

I found the prevalence of *Vibrio cholerae* in patient 1 is 100 times larger than that in patient 2. Therefore, in problem 4c, I will focus on the patient 1 and patient 3.

The code solution of the procedure can be found in **infection\_investigator.py** as shown in function **find\_count** and **find\_zeros**, the detailed comments and the outputs can also be seen in the code file.

The output of function **find\_zeros** for patient 3 after the addressing of edge effect is that no continuous zeros in the reference genome, which means that patient 3's genomic reads distributed all across the reference genome.

The output of function **find\_zeros** for patient 1 is 5015 for start position and 5079 for end position in 0-indexed. In addition, the two position index will be 5016 and 50800 in 1-indexed position. The string I extracted from the reference genome is:

*TACCGGCCAGGTGCAACTTTTCAAGTAGAAGT*

*ACCAGGTAGTCAACATATAGATTACAAACCTTT*

For the blast result, I have attached the result from the website behind. It seems that the relationship between the virus and the bacteria exist since I have the **Vibrio virus CTXphi, complete genome** result and **Vibrio cholerae strain PM19 phage CTX genomic sequence** result.

I have learned that the role of phages in the analysis of disease is very important and we cannot ignore such things. Also, I would tell my patient 1 to pay attention on the disease cholera due to the microbe *Vibrio cholerae* and recommend some treatments to the patient. That's all for my work!

## Reference

1. [https://en.wikipedia.org/wiki/Bacteroides\\_thetaiotaomicron](https://en.wikipedia.org/wiki/Bacteroides_thetaiotaomicron)
2. [https://en.wikipedia.org/wiki/Bifidobacterium\\_longum](https://en.wikipedia.org/wiki/Bifidobacterium_longum)
3. <https://www.ncbi.nlm.nih.gov/genome/?term=Eubacterium%20rectale>
4. [https://en.wikipedia.org/wiki/Lactobacillus\\_acidophilus](https://en.wikipedia.org/wiki/Lactobacillus_acidophilus)
5. Mishra AK, Lagier J-C, Robert C, Raoult D, Fournier P-E. Non contiguous-finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. Standards in Genomic Sciences. 2012;7(1):1-11. doi:10.4056/sigs.2956294.
6. [https://microbewiki.kenyon.edu/index.php/Prevotella\\_copri](https://microbewiki.kenyon.edu/index.php/Prevotella_copri)
7. [https://en.wikipedia.org/wiki/Roseburia\\_intestinalis](https://en.wikipedia.org/wiki/Roseburia_intestinalis)

8. <https://en.wikipedia.org/wiki/Ruminococcus>
9. Ze X, Duncan SH, Louis P, Flint HJ. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. The ISME Journal. 2012;6(8):1535-1543. doi:10.1038/ismej.2012.4.
10. [https://en.wikipedia.org/wiki/Vibrio\\_cholerae](https://en.wikipedia.org/wiki/Vibrio_cholerae)

[BLAST®](#) » [blastn suite](#) » RID-W76PJCMY015

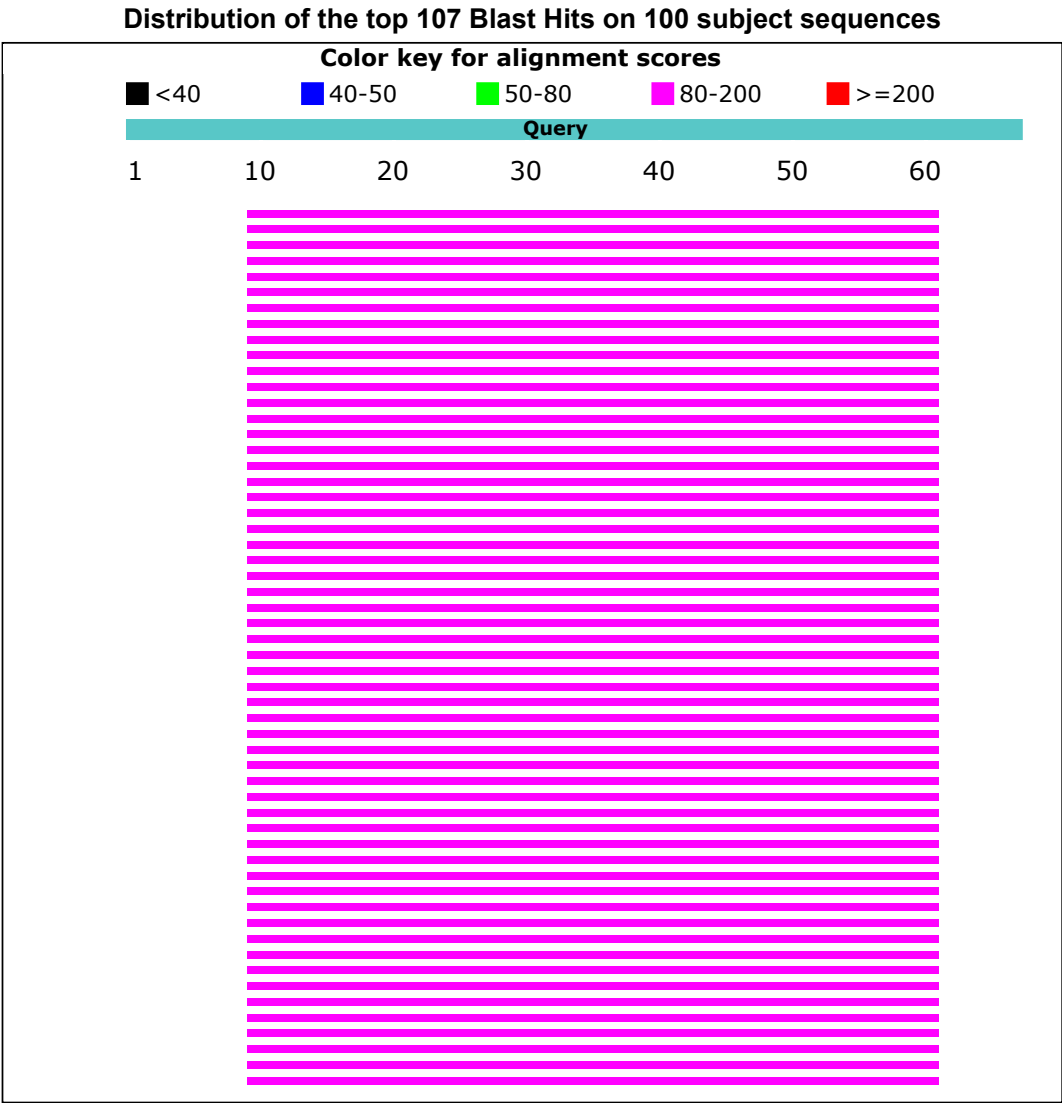
BLAST Results

[Questions/comments](#) \_

Job title: Nucleotide Sequence (65 letters)

<b>RID</b>	<a href="#">W76PJCMY015</a> (Expires on 10-16 03:36 am)	<b>Database Name</b>	nr
<b>Query ID</b>	Idl Query_153039	<b>Description</b>	Nucleotide collection (nt)
<b>Description</b>	None	<b>Program</b>	BLASTN 2.8.1+
<b>Molecule type</b>	nucleic acid		
<b>Query Length</b>	65		

[Graphic Summary](#).



## Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
Vibrio cholerae strain OG80 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356517.1</a>
Vibrio cholerae strain KA66 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356516.1</a>
Vibrio cholerae strain AB45 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356515.1</a>
Vibrio cholerae strain OS30 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356514.1</a>
Vibrio cholerae strain BA29 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356513.1</a>
Vibrio cholerae strain B24 CtxB (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MG356512.1</a>
Vibrio cholerae strain WB-0050 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320542.1</a>
Vibrio cholerae strain GJ-0071 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320541.1</a>
Vibrio cholerae strain CG-0005 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320516.1</a>
Vibrio cholerae strain WB-0036 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320515.1</a>
Vibrio cholerae strain RJ-0003 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320514.1</a>
Vibrio cholerae strain CH-0001 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320513.1</a>
Vibrio cholerae strain OR-0001 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320512.1</a>
Vibrio cholerae strain AS-0005 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320511.1</a>
Vibrio cholerae strain AP-0006 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320510.1</a>
Vibrio cholerae strain MP-0003 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320509.1</a>
Vibrio cholerae strain KA-0001 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320508.1</a>
Vibrio cholerae strain MH-0009 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320507.1</a>
Vibrio cholerae strain GJ-0003 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320506.1</a>



Description	Max score	Total score	Query cover	E value	Ident	Accession
Vibrio cholerae strain DL-0001 cholera toxin subunit B (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">MH320505.1</a>
Vibrio cholerae V060002 DNA, complete genome	95.3	190	78%	1e-16	100%	<a href="#">AP018677.1</a>
Vibrio virus CTXphi, complete genome	95.3	95.3	78%	1e-16	100%	<a href="#">MF155889.1</a>
Vibrio cholerae strain 4295STDY6534248 genome assembly, chromosome: 1	95.3	190	78%	1e-16	100%	<a href="#">LT992492.1</a>
Vibrio cholerae strain 4295STDY6534232 genome assembly, chromosome: 1	95.3	190	78%	1e-16	100%	<a href="#">LT992488.1</a>
Vibrio cholerae strain 4295STDY6534216 genome assembly, chromosome: 1	95.3	190	78%	1e-16	100%	<a href="#">LT992486.1</a>
Vibrio cholerae O1 biovar El Tor strain HC1037 chromosome I, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">CP026647.1</a>
Vibrio cholerae strain R-18588 core encoded pilin (cep), OrfU (orfU), accessory cholera enterotoxin (ace), zonula occludens toxin (zot), cholera toxin A-subunit (ctxA), and cholera toxin B-subunit (ctxB) genes, ctxB1 allele, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KX584736.1</a>
Vibrio cholerae strain R-18368 core encoded pilin (cep), OrfU (orfU), accessory cholera enterotoxin (ace), zonula occludens toxin (zot), cholera toxin A-subunit (ctxA), and cholera toxin B-subunit (ctxB) genes, ctxB1 allele, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KX584735.1</a>
Vibrio cholerae strain R-18367 core encoded pilin (cep), OrfU (orfU), accessory cholera enterotoxin (ace), zonula occludens toxin (zot), cholera toxin A-subunit (ctxA), and cholera toxin B-subunit (ctxB) genes, ctxB1 allele, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KX584734.1</a>
Vibrio cholerae strain M2140 chromosome 1, complete sequence	95.3	190	78%	1e-16	100%	<a href="#">CP013315.1</a>
Vibrio cholerae strain E1320 chromosome 2, complete sequence	95.3	190	78%	1e-16	100%	<a href="#">CP013312.1</a>
Vibrio cholerae strain E506 chromosome 1, complete sequence	95.3	190	78%	1e-16	100%	<a href="#">CP013307.1</a>
Vibrio cholerae strain PTCC 1611 cholera toxin B subunit (ctxB) gene, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">KU646842.1</a>
Synthetic construct cholera toxin subunit B/tetanus toxin heavy chain fusion protein gene, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">KX022510.1</a>
Vibrio cholerae O1 biovar El Tor cholera toxin subunit A (ctxA) and cholera toxin subunit B (cxtB) genes, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KU215666.1</a>
Vibrio phage pre-CTX, partial genome	95.3	95.3	78%	1e-16	100%	<a href="#">KT779274.1</a>
Vibrio phage pre-CTX, partial genome	95.3	95.3	78%	1e-16	100%	<a href="#">KT779273.1</a>
Vibrio phage pre-CTX, partial genome	95.3	95.3	78%	1e-16	100%	<a href="#">KT779272.1</a>
Vibrio phage pre-CTX, partial genome	95.3	95.3	78%	1e-16	100%	<a href="#">KT779270.1</a>
Vibrio cholerae O1 strain DN74 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734657.1</a>
Vibrio cholerae O1 strain DN73 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734656.1</a>

Description	Max score	Total score	Query cover	E value	Ident	Accession
Vibrio cholerae O1 strain DN72 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734655.1</a>
Vibrio cholerae O1 strain DN71 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734654.1</a>
Vibrio cholerae O1 strain DN70 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734653.1</a>
Vibrio cholerae O1 strain DN69 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734652.1</a>
Vibrio cholerae O1 strain DN67 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734651.1</a>
Vibrio cholerae O1 strain DN65 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734650.1</a>
Vibrio cholerae O1 strain DN64 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734649.1</a>
Vibrio cholerae O1 strain DN63 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734648.1</a>
Vibrio cholerae O1 strain DN62 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734647.1</a>
Vibrio cholerae O1 strain DN60 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734646.1</a>
Vibrio cholerae O1 strain DN56 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734645.1</a>
Vibrio cholerae O1 strain DN55 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734644.1</a>
Vibrio cholerae O1 strain DN54 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734643.1</a>
Vibrio cholerae O1 strain DN52 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734642.1</a>
Vibrio cholerae O1 strain DN51 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734641.1</a>
Vibrio cholerae O1 strain DN50 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734640.1</a>
Vibrio cholerae O1 strain DN49 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734639.1</a>
Vibrio cholerae O1 strain DN47 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT734638.1</a>
Vibrio cholerae strain EDC-409 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278777.1</a>
Vibrio cholerae strain EDC-159 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278776.1</a>

Description	Max score	Total score	Query cover	E value	Ident	Accession
Vibrio cholerae strain EDC-038 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278775.1</a>
Vibrio cholerae strain DCC-203 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278774.1</a>
Vibrio cholerae strain EDC-506 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278773.1</a>
Vibrio cholerae strain EDC-344 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278772.1</a>
Vibrio cholerae strain EDC-200 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278771.1</a>
Vibrio cholerae strain EDC-025 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278770.1</a>
Vibrio cholerae strain EDC-583 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278769.1</a>
Vibrio cholerae strain DCS-121 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278768.1</a>
Vibrio cholerae strain DCS-041 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278767.1</a>
Vibrio cholerae strain DCC-006 cholera enterotoxin B-subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KT278766.1</a>
Vibrio cholerae O1 str. KW3 chromosome I, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">CP006947.1</a>
Vibrio cholerae strain 16150 RstR (rstR), RstA (rstA), RstB (rstB), core encoded pilin (cep), hypothetical protein, accessory cholera enterotoxin (ace), zonula occludens toxin (zot), cholera toxin subunit A (ctxA), and cholera toxin subunit B (ctxB) genes, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KR072210.1</a>
Vibrio cholerae strain TSY216 chromosome 1, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">CP007653.1</a>
Vibrio cholerae O1 biovar El Tor strain FJ147 chromosome I, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">CP009042.1</a>
Vibrio cholerae strain RND81 DNA binding protein gene, RstR (rstR), RstA (rstA), RstB (rstB), Cep (cep), hypothetical protein, Ace (ace), Zot (zot), CtxA (ctxA), CtxB (ctxB), RtxA (rtxA), RtxC (rtxC), Chp (chp), RtxB (rtxB), RtxD (rtxD), RtxE (rtxE), and SHK/RR genes, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KM401563.1</a>
Vibrio phage CTX strain 81, partial genome	95.3	95.3	78%	1e-16	100%	<a href="#">KM352500.1</a>
Vibrio cholerae strain 2012EL-2176 chromosome 1, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">CP007634.1</a>
Synthetic construct gene for recombinant T-cell epitopes of major Japanese cedar pollen allergens fused with cholera toxin B subunit, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">AB978369.1</a>

Description	Max score	Total score	Query cover	E value	Ident	Accession
Vibrio cholerae strain DN24 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KJ730047.1</a>
Vibrio cholerae strain DN19 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KJ730043.1</a>
Vibrio cholerae strain DN18 cholera toxin B subunit (ctxB) gene, complete cds	95.3	95.3	78%	1e-16	100%	<a href="#">KJ730042.1</a>
Vibrio cholerae strain PM19 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540278.1</a>
Vibrio cholerae strain PM18 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540277.1</a>
Vibrio cholerae strain PM17 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540276.1</a>
Vibrio cholerae strain PM15 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540274.1</a>
Vibrio cholerae strain IB5230 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540264.1</a>
Vibrio cholerae strain IB4540 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540263.1</a>
Vibrio cholerae strain IB4405 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540262.1</a>
Vibrio cholerae strain IB4247 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540261.1</a>
Vibrio cholerae strain IB4563 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540260.1</a>
Vibrio cholerae strain IB5419 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540259.1</a>
Vibrio cholerae strain IB4712 phage CTX genomic sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540258.1</a>
Vibrio phage CTX plasmid pCTX-6 Kan, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540272.1</a>
Vibrio phage CTX plasmid pCTX-5 Kan, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540271.1</a>
Vibrio phage CTX plasmid pCTX-3 Kan, complete sequence	95.3	95.3	78%	1e-16	100%	<a href="#">KJ540270.1</a>
Vibrio phage CTX, complete genome	95.3	95.3	78%	1e-16	100%	<a href="#">KJ619459.1</a>
Vibrio cholerae strain SC2008 CtxB (ctxB) gene, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">KJ437653.1</a>
Vibrio cholerae strain GX2006022 cholera toxin B subunit (ctxB) gene, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">KC754366.1</a>
Vibrio cholerae strain BJ823 cholera toxin B subunit (ctxB) gene, partial cds	95.3	95.3	78%	1e-16	100%	<a href="#">KC754362.1</a>

## Alignments

Vibrio cholerae strain OG80 CtxB (ctxB) gene, complete cds

Sequence ID: **MG356517.1** Length: 375 Number of Matches: 1

Range 1: 196 to 246

Score	Expect	Identities	Gaps	Strand	Frame
95.3 bits(51)	1e-16()	51/51(100%)	0/51(0%)	Plus/Plus	

Features:

Query	10	GGTGC AACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATT CACAA	60
Sbjct	196	GGTGC AACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATT CACAA	246

Vibrio cholerae strain KA66 CtxB (ctxB) gene, complete cds

Sequence ID: **MG356516.1** Length: 375 Number of Matches: 1

Range 1: 196 to 246

Score	Expect	Identities	Gaps	Strand	Frame
95.3 bits(51)	1e-16()	51/51(100%)	0/51(0%)	Plus/Plus	

Features:

```

Query 10  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 60
Sbjct 196  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 246

```

Vibrio cholerae strain AB45 CtxB (ctxB) gene, complete cds

Sequence ID: **MG356515.1** Length: 375 Number of Matches: 1

Range 1: 196 to 246

Score	Expect	Identities	Gaps	Strand	Frame
95.3 bits(51)	1e-16()	51/51(100%)	0/51(0%)	Plus/Plus	

Features:

```

Query 10  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 60
Sbjct 196  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 246

```

Vibrio cholerae strain OS30 CtxB (ctxB) gene, complete cds

Sequence ID: **MG356514.1** Length: 375 Number of Matches: 1

Range 1: 196 to 246

Score	Expect	Identities	Gaps	Strand	Frame
95.3 bits(51)	1e-16()	51/51(100%)	0/51(0%)	Plus/Plus	

Features:

```

Query 10  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 60
Sbjct 196  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 246

```

Vibrio cholerae strain BA29 CtxB (ctxB) gene, complete cds

Sequence ID: **MG356513.1** Length: 375 Number of Matches: 1

Range 1: 196 to 246

Score	Expect	Identities	Gaps	Strand	Frame
95.3 bits(51)	1e-16()	51/51(100%)	0/51(0%)	Plus/Plus	

Features:

```

Query 10  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 60
Sbjct 196  GGTGCAACTTTTCAAGTAGAAGTACCAGGTAGTCAACATATAGATTACAA 246

```

BLAST is a registered trademark of the National Library of Medicine



[Support center Mailing list](#)

[YouTube](#)



[National Library Of Medicine](#)



- [National Institutes Of Health](#)



- [U.S. Department of Health & Human Services](#)



- [USA.gov](#)

## **NCBI**

*National Center for Biotechnology Information*. *U.S. National Library of Medicine* 8600 Rockville Pike, Bethesda MD, 20894 USA  
[Policies and Guidelines](#) | [Contact](#)