

## Introduction to Bioinformatics 2024-2025

### Exercise 1 (M. Reczko):

(Adapted from: <https://web.archive.org/web/20150425010121/http://www.ableweb.org/volumes/vol-28/v28reprint.php?ch=8>)

In a hypothetical scenario many people in a city suddenly come down with a serious illness. All the victims have in common is that they were all in a downtown pedestrian mall at a certain time five days before. Could terrorists have released a cloud of viruses or bacteria from a vehicle downwind of the mall? You work for the Centers for Disease Control and Prevention, and you have to find out.

A sample of non-human DNA (bacterial or viral) has been isolated from the victims. Identify the DNA sample as well as you can. Some of the DNA molecules are very short, and have been partially degraded. You will notice that the sequence is sprinkled with Ns, “N” stands for “nucleotide” and means that the nucleotide at that position could not be determined.

Some judgment is called for as you interpret your results. First, everyone has bacteria and viruses in his or her body, and sometimes they can cause disease. However, we are looking for exotic pathogens with bioterrorism potential (e.g., anthrax or smallpox rather than the common cold). Even AIDS, although it is deadly, would not work as a bioterror weapon because the disease develops too slowly and the virus is too hard to disseminate. For the purposes of this exercise, we will not consider a pathogen a

bioterror agent unless it is listed as a potential agent on the Centers for Disease Control and Prevention Web site at <https://emergency.cdc.gov/agent/agentlist.asp>.

Second, organisms that are evolutionarily related have similar DNA, which might lead you to sound a false alarm. For example, say you find the following when you do a BLAST search on a certain DNA sample:

Sequences producing significant alignments:	Score (Bits)	E Value
<a href="#">gi 40012 emb X02369.1 BSORIC</a> Bacillus subtilis oriC region	<a href="#">5967</a>	0.0
<a href="#">gi 32468687 emb Z99104.2 BSUB0001</a> Bacillus subtilis complete ...	<a href="#">5967</a>	0.0
<a href="#">gi 467326 dbj D26185.1 BAC180K</a> B. subtilis DNA, 180 kilobase reg	<a href="#">5967</a>	0.0
<a href="#">gi 39877 emb X12778.1 BSDNAA</a> Bacillus subtilis dnaA gene 5'-regi	<a href="#">846</a>	0.0
<a href="#">gi 56160984 gb CP000002.2 </a> Bacillus licheniformis ATCC 14580, co	<a href="#">690</a>	0.0
<a href="#">gi 52346357 gb AE017333.1 </a> Bacillus licheniformis DSM 13, comple	<a href="#">690</a>	0.0
<a href="#">gi 39878 emb X12779.1 BSDNAAN</a> Bacillus subtilis genes for dnaA (	<a href="#">587</a>	8e-164
<a href="#">gi 39893 emb X17013.1 BSDPD</a> Bacillus subtilis lys gene for di...	<a href="#">525</a>	2e-145
<a href="#">gi 51973633 gb CP000001.1 </a> Bacillus cereus E33L, complete genome	<a href="#">337</a>	1e-88
<a href="#">gi 49328240 gb AE017355.1 </a> Bacillus thuringiensis serovar kon...	<a href="#">329</a>	3e-86
<a href="#">gi 50082967 gb AE017334.2 </a> Bacillus anthracis str. 'Ames Ancesto	<a href="#">329</a>	3e-86
<a href="#">gi 49176966 gb AE017225.1 </a> Bacillus anthracis str. Sterne, compl	<a href="#">329</a>	3e-86

*Bacillus subtilis* is a harmless and very common soil bacterium. It is closely related to *Bacillus anthracis*. *Bacillus anthracis* causes anthrax, and is a dangerous bioterror weapon. Note from the similarity score (second column from the right) that *Bacillus subtilis* DNA is far more similar to the sample than *Bacillus anthracis* DNA is. Unless one of your samples gives a stronger indication of *Bacillus anthracis* than this, the mention of *B. anthracis* in the output is probably just due to genetic similarities between it and *B. subtilis*.

#### 1. Analyze the samples

>outbreak14

```
GCCGAGTTAGTCTTGTGCTNACGGAAGCTTATTGTATGAGTANTGATTTGAAAGAGCTANANTTAAAA  
AATCACTAATNAATNTAAGAGCGGACTTAACNAGCGTAAAAGCTTACTAATTAATTGTCAGTTA
```

GCTCGTTCAGGTAATGGTTCCTANCGGNCAATGCAGGAAGAGTTCTACCTGGAAGTGANAGACCGC  
TGGCGGTGACAACACACTACGTCAAAATAAGA

>outbreak15

TAGTCTTGTGCTNACGGAACTTATTTATGAGGTACCCACCGANTCTGAAAACCGCTAATANAGCACT  
TAAAAATAAGAGCAGAATGGGATTTAAGGATAG

separately using both megablast and blastn at

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=&LINK\\_LOC=blasttab](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab) and to determine if there is any evidence of bioterror agents. Use the general nucleotide collection (nr/nt). Report any differences between the 2 algorithms.

2. Check the CDC Web site at <https://emergency.cdc.gov/agent/agentlist.asp>.

to see if the CDC considers any found organism to be a potential weapon. If you've found a bioterror agent, research it on the CDC site so you can describe its effects on humans.

3. The health effects of many pathogenic bacteria are briefly described on the NCBI Genomes Web site at <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>. Click on a species name to see its information. It also might be helpful to do a general Google search.

SEND SOLUTIONS (for M.Reczko exercises) ONLY TO:

[mareczko@di.uoa.gr](mailto:mareczko@di.uoa.gr)