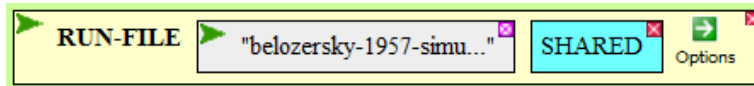


## Molecular Biology Through Discovery

### Problem Set 7: RNA

1. What does the data of Belozersky and Spirin (1957) mean? If you can calculate using only genomic information the substance of their Tables 1 and 2, you might get a step closer to understanding.

- Go to the [BioBIKE Portal](#)
- Enter the PhAnToMe/BioBIKE public site (second choice: ViroBIKE).
- Log in as usual (or enter non-registered).
- Once into BioBIKE, go to the INPUT/OUTPUT menu and click RUN-FILE.
- Fill in the *file-name* box with "belozersky-1957-simulation.bike" (in quotes).
- Mouse over the Options icon and click SHARED. You should get this:



- Execute the now completed function by double-clicking RUN-FILE.

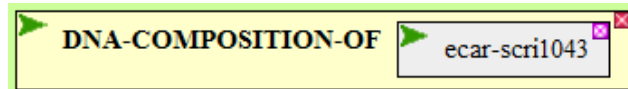
In a few seconds, you should get a message that loading completed successfully. As a result, you will have two new functions on your FUNCTIONS button.

- 1a. Reproduce a portion of Table 1 from Belozersky and Spirin (1957). To do this, mouse over the FUNCTIONS button and click DNA-COMPOSITION-OF. In the *organism* box type in the nickname for any one of the following organisms (these are both present in Table 1 and have genome sequences known by BioBIKE):

Organism name	Nickname
Clostridium-perfringens-str-13	Cper-13
Escherichia-coli-K12	EcoK12
Shigella-dysenteriae-M131649	Sdys-M131649
Erwinia-carotovora-subsp-atroseptica-SCRI1043	Ecar-Scri1043
Corynebacterium-diphtheriae-NCTC-13129	Cdip-13129
Brucella-abortus-biovar-1-str-9-941	Babo-9-941
Mycobacterium-tuberculosis-CDC1551	Mtub-1551

(In ViroBIKE, only Cper-13 will work)

You should see something like the following (depending on which nickname you chose)



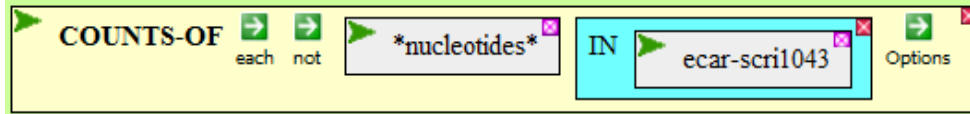
Execute the function. Try it again with a different organism.

How do your numbers relate to the numbers of Belozersky and Spirin (1957)?

- 1b. What do those numbers mean? Recall *What is a Gene?* You've done this problem before.

- From the green ALL menu, bring down the COUNTS-OF function.
- Click the *query* box to open it.
- Mouse over the blue DATA menu and click \*nucleotides\*.
- Mouse over the Options icon and click IN

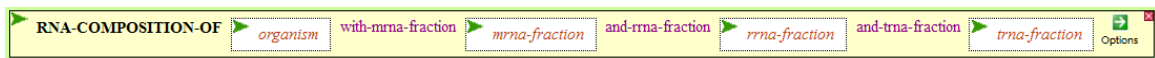
- Open the *value* box of IN and enter a nickname of an organism. You should see:



- Execute the function. You should see four numbers in the Result pane at the bottom of the screen.

What do these numbers mean? How could you get them yourself by hand (if you had a *lot* of time)? How do they relate to the numbers of **3a**? If you want to play more, you might recall the LENGTH-OF function, which will tell you how many total nucleotides are in a given genome (or just add up the four numbers from the previous function).

- 1c. Now on to Table 2, a listing of RNA base composition. As before, you'll calculate this... wait a second! What does RNA composition mean? How could you calculate it from just the genome sequence? What would you need to know?
- 1d. Let's see what the function can do for you. Mouse over the FUNCTIONS button and click RNA-COMPOSITION-OF, getting:



The function demands four things from you. In the *organism* box, enter your favorite organism nickname from the table on the previous page. The other three boxes ask you for the fractions of RNA for mRNA, rRNA, and tRNA (which should add up to 1). How do you know what they are? You don't, but be thankful the function asked you for the fraction of *classes* of RNA and not the fraction of every gene in the organism!

What values for the three fractions enable you to reproduce Table 2 from Belozersky and Spirin? (Try a few) Why?

2. Consider the two scenarios in the [companion](#) to Belozersky and Spirin (1957), bearing in mind your experience with Question 1.
  - 2a. Relate Scenario 1 (word length survey) to the results in the article.
  - 2b. Relate Scenario 2 (automobile structure) to the results in the article.
3. Consider the results of Brenner et al (1961), presuming alternate realities.
  - 3a. Redraw Fig. 2, presuming that ribosomes freely dissociate into their component proteins and then reassociate to form ribosomes again, prior to centrifugation.
  - 3b. Redraw Fig. 6, presuming that their Model I is correct.
  - 3c. Redraw Fig. 7, presuming that their Model II is correct. Redraw it presuming that Model I is correct.
4. What were the two critical observations (i.e. experiment + evidence) of Astrachan and Volkin (1958)? How did the experiments of Brenner *et al* show a way out of the dilemma posed by their work?

5. Below is the genome sequence of bacteriophage lambda in the region of the *cI* and *cro* genes.

```
5' - CTTTTTTGTGCTCATACGTTAAATCTATCACCGCAAGGGATAAAATATCTAACACCGTGC GTGTTG - 3'
3' - GAAAAAACACGAGTATGCAATTTAGATAGTGGCGTTCCTATTTATAGATTGTGGCAGCACAAC - 5'

5' - ACTATTTTACCTCTGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATGGAACAACGCATA - 3'
3' - TGATAAAATGGAGACCGCCACTATTACCAACGTACATGATTCCTCCAACATACCTTGTGCGTAT - 5'
```

5a. Copy the sequence shown above and paste it into your favorite word processor. Highlight or otherwise mark the beginnings of the two genes, the -35 and -10 regions of the promoters, and the Repressor/Cro binding sites.

5b. Extract all instances of the promoters and align them (single stranded, 5'→3'). Highlight or otherwise mark deviations from the consensus -35 and -10 regions.

5c. Extract all instances of the Repressor/Cro binding sites and align them (single stranded, 5'→3'). Use the alignment to guess at what might be a consensus binding site. Do any of the binding sites contain palindromic sequences?

6. See if you can identify regulatory elements without much prior knowledge. Is the repressor of Lactobacillus Phage A2 regulated in a manner similar to the lambda repressor? Go to ViroBIKE and bring up the SEQUENCE-OF *lact-a2* (*lact-a2* is the nickname of Lactobacillus phage A2). Now go to the *cI* repressor – *cro* region.... (Hmmm, that might be a problem).

6a. In ViroBIKE look for the repressor gene by going to the alphabetized list of functions (the ALL button) and clicking GENE/S-DESCRIBED-BY. Then fill in the *query* box with "repressor", choose the IN option and give its *value* as *lact-a2*, then execute the function. A suggested gene should pop up. Is that the one you want? If not, then try the same thing with "cro" in the *query* box.

6b. Copy the full name of the desired gene (beginning *NC\_004112.gene...*) and paste it into the **Go to** box of the sequence viewer. Locate the *cro* gene. If Lact-A2 is like Lambda, then where should the *cI* repressor be? Find it. Note that you can use the **Go to** box to specify the first nucleotide you want to appear on the screen.

6c. Make the display double-stranded by clicking the **Double-stranded?** check box and clicking **Go**. Then copy the double-stranded intergenic region between *cI* and *cro* and paste it into your favorite word processor.

6d. Look for sequences that might be -35 and -10 regions of promoters. Highlight them.

6e. Look for sequences that might be Repressor/Cro binding sites. This is not easy. Repressor binding sites are notorious for changing from organism to organism. Do you see any sequences that appear more than once on the two strands? Do you see any sequences that contain palindromes?

---

### References

- Astrachan L, Volkin E, (1958). Properties of ribonucleic acid turnover in T2-infected *Escherichia coli*. *Biochim Biophys Acta* 29:536-544.
- Belozersky, A. N. & Spirin, A. S. A correlation between the compositions of deoxyribonucleic and ribonucleic acids. *Nature* **182**, 111-112 (1958).
- Brenner S, Jacob F, Meselson M (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576-581.