(Updated 20 Aug 2013)

What is a Gene?

A BioBIKE Tour

IV. What determines the end of a protein-encoding gene?

Stuffed with insights as to how a gene begins, you must be raring to figure out how genes end as well. Don't let me stop you!

Consider your successful strategy to characterize the beginning of a gene. First, you isolated the beginning of one gene and then generalized the method to work on all genes. Then you examined the beginning sequences and identified a pattern. Then you tested different hypotheses concerning the pattern. At each step, if you were wise, you tried to test your tools and your ideas against reality. When working with large numbers of entities, it is easy to make a small mistake that has a large effect.

I'll offer a few variations on the tricks you already know. Perhaps they can make your life easier.

35. Start as you did before, with the sequence of *Avar*. What is the end of Ava0001?

    35a. What is the end of Ava0001?

36. Now extract the last few nucleotides of the gene, using SEQUENCE-OF. The number system at the end of genes is similar to that at the beginning. Using the end of the sequence of Ava0001 as an example:

-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15
T	A	A	A	A	G	T	T	C	G	A	C	G	T	T	**T**	**T**	**C**

    The FROM and TO options of SEQUENCE-OF start counting from the beginning of a gene. The FROM-END and TO-END options start counting from the end.

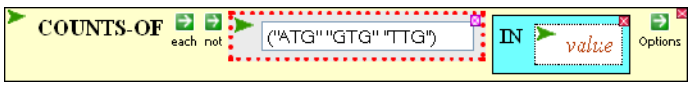
    36a. Use SEQUENCE-OF and the FROM-END and TO-END, display the end of Ava0001.

    36b. Display the ends of all the genes of *Avar*.

    36c. Any pattern?

37. Now count the different versions of the end. You used COUNT-OF before. Here's a possibly useful variant (of course you should use not what's shown but whatever sequences are pertinent):

    37a. Use this function to count what you want to count in the genes (or perhaps coding genes) of *Avar*.



Take it from there.

Supplemental Problems

P8. What are the most common ends of genes in *Avar*? Why?

P9. Do other cyanobacteria use the same ends with the same frequency?

P10. Are there any interesting downstream features, say in the first 50 nucleotides?