Phage Gene Annotation

Overview

- All annotators should establish a single login name, providing full name, e-mail address, and institution (spelled out)
- Each phage gene will be annotated by at least two annotators, in sequence (**Pass 1** and **Pass 2**). But anyone is free to annotate any gene at any time.
- Genes will be annotated with respect to legitimacy, start codon, function, transcriptional unit, and DNA features.
- All annotations should be entered on the appropriate gene's annotation page in PhAnToMe/BioBIKE.
- The group as a whole is responsible for a phage's annotation. If someone in the group takes off, the group is responsible for finding a way to get the missing annotation accomplished.

Pass 1: Initial annotation

1. Genes for which you're responsible will be dispersed throughout the genome. See the **Genome Analysis** web page, and click on your phage.

2. All annotations should be entered into the gene's annotation page, which can be accessed through the INFORMATION-ABOUT-GENE/S function in PhAnToMe/BioBIKE.

3. Consider the **start codon** of the gene.

   a. Use GeneMark and other tools to seek alternative start codons.

   b. Identify a likely ribosome binding site if possible.

   c. Consider overlapping with the upstream gene and the size of the intergenic sequence.

   d. Use SEQUENCE-SIMILAR-TO/Blast to determine which start codon leads to the greatest conservation of amino acids with proteins of other phages and bacteria. Problem Set 7, Question 5, provides an example of this approach.

   e. When you've reached your decision, enter it in the From or To field as appropriate (From for forward genes, To for backwards genes). Even if you choose not to change the value, enter the evidence behind your decision, including the sequence and coordinates of the ribosome-binding site, if you could identify one.

4. Consider the **function** of the encoded protein.

   a. Use SEQUENCE-SIMILAR-TO/Blast to collect candidate functions. Consider both the quality and the extent of similarities.

   b. Use DOMAINS-OF to collect candidate functions.

   c. When you've reached your decision, enter it in the Main Annotation field. Even if you choose not to change the functional annotation, enter the evidence behind your decision.

5. Consider the **structural features** of the encoded protein
    a. Use DOMAINS-OF and possibly other means to identify possible membrane spans and signal sequences.
    b. If you find features worth annotating, enter them in the appropriate field.
6. Consider the **transcriptional unit** of the gene
    a. Examine the context of the gene to predict whether it is part of a multigene operon.
    b. If you believe it to be so, enter in the Operon Structure field the first and last gene of the proposed transcriptional unit, along with the evidence behind your decision.
7. Consider **DNA features** near the gene
    a. Look for plausible promoter sequences, transcriptional terminator sequences, repeated sequences, and long palindromes.
    b. If you find anything plausible, enter it in the Regulation field, along with the evidence behind your decision.

Pass 2: Broader analysis and confirmation

1. Genes for which you're responsible will be in one contiguous segment. See the **Genome Analysis** web page, and click on your phage.
2. Examine your entire segment for alternative genes that were not found during the automated annotation.
    a. GeneMark may be helpful (try both the heuristic option and GeneMark.hmm-p, using a Mycobacterium as the reference species).
    b. SEQUENCE-SIMILAR-TO using the Translated-DNA-vs-Protein option (BlastX) may also be illuminating.
    c. Alternative genes will often conflict with genes present in the automated annotation. Use comparisons with other phages to decide which gene is the more likely to be functional, Functional genes are more likely to be conserved than artifactual genes.
    d. If you decide that an alternative gene should become part of the phage annotation, or if you believe that a gene in the present annotation should be removed, please let me know. At present there is no way for you to add or subtract genes from a genome.
3. Examine the first pass annotation of each gene in your segment. Make changes you believe are warranted. Even if none are required, resave the existing annotation so that your name is associated with it.