Algorithm S1: Recognition of genetic codes and alternating genetic codes in phage genomes and annotation coding regions.

Input: FASTA file
Output: GTF/GFF2 formatted gene annotations that account for the genetic code

1. Run MetaGeneMark with models of protein-coding region with genetic code models 11, 4, 15, and 101
2. If not predict alternating genetic codes
   a. Assign genetic code to input sequences based on model of protein-coding regions with highest coding potential and provide corresponding gene annotation

   Else
   a. Apply sliding window approach (5000 bp)
      i. Determine model with highest coding potential (log-odds scores) in each window;
      ii. Assign window labels accordingly
   b. If a model has the highest potential in all windows
      i. This model's predictions are used as gene annotations in the whole genome

      Else
      i. Identify the two models that have most frequently the highest coding potential in a window (A & B)
      ii. Segment genome into genomic blocks
         1. Set longest consecutive sequence of windows, which have the same window label and have not been updated, as seed block; Set corresponding label as reference
         2. Extend seed block into both direction until *n* consecutive mismatches are observed
         3. Update labels of extended sequence of windows according to reference
         4. Repeat 1-3 until all window labels have been updated
         5. Define tentative block boundaries as the center of intergenic region separating the two blocks
         6. If the PES switches within one window of predicted block boundary, update prediction to center of intergenic region separating the genes between which the PES switches
      iii. Compile set of predicted protein coding regions
         1. For all predictions made by models A & B
            a. If model A predicts a long gene and model B many short genes on the same PES, and coding potential of model A ≥ model B
               1. Keep prediction of model A; Drop predictions of model B; Assign gene label A
            b. Elif model A predicts a long gene and model B many short genes on the same PES, and coding potential of model A < model B
               i. Drop prediction of model A; Keep predictions of model B Assign gene label B
            c. Elif predictions of model A & B are identical
               i. Keep either prediction Assign gene label C
            d. Elif (predictions of model A and B share stop coordinates) or (share start coordinates and stop coordinates differ by less than *t* bp)
               i. Keep both predictions and annotate as isoforms Assign gene label C
            e. Else
               i. Keep prediction Assign gene label A or B accordingly