The diagram above shows a portion of the multiple sequence alignment for the gene that codes for the hemagglutinin protein in 11 influenza viruses. The full gene has about 1,000 nucleotides. Samples are identified in the left column by the country or city of origin and the year the virus was isolated from a patient. In this alignment, a period (.) means that the nucleotide is the same as the most common nucleotide in the sample. Nucleotides that remain the same in all samples have a darker gray background. Nucleotides that vary across the samples have a lighter background. Use the alignment to answer the questions on page 2.

HKG = Hong Kong; ENG = England; VIC = Victoria, Australia; TXS = Texas, USA; PHP = Philippines; SIG = Singapore; PAR = Paris, France; MAD = Madrid, Spain; JOH = Johannesburg, South Africa; AUK = Auckland, New Zealand; FIN = Finland
Questions

1. Calculate the number of changes per nucleotide in this 100-nucleotide sequence by using the following formula:

\[
\frac{\text{number of nucleotides that have at least one change}}{\text{total number of nucleotides}}
\]

2. The viruses in this study were collected over a span of 35 years. Calculate the number of changes per nucleotide per year by using the following formula:

\[
\frac{\text{answer from Step 1}}{\text{total number of years}}
\]

3. In Lesson 3, you investigated a portion of the sequence of a gene called Irf6 that is involved in the development of the head and face. You compared the sequence of this gene in many different species. Thirty nucleotides from the Irf6 sequence are shown below:

\[\text{TGGGCCAC}c\text{AGCCAGGGCTT}t\text{AGcCGgACT}\]

The lowercase (gray) letters show nucleotides that differ among some of the species. The uppercase (black) letters did not change. The amount of time represented in this comparison is 1,009 million years (or 1.009 billion years). Use the same formulas you used in Steps 1 and 2 to calculate the expected number of changes per nucleotide per year in this sequence.

4. Compare the rate of change per nucleotide per year for the hemagglutinin gene in influenza with the rate for the Irf6 gene. Do this by dividing the rate for the hemagglutinin region by the rate for the Irf6 gene. The number you calculate will show how many times faster one region changes than the other.

5. The number of changes in the hemagglutinin gene for six samples compared with the sample from Finland in 2003 is as follows:
   - Hong Kong, 1968 = 140 changes
   - Victoria, Australia, 1975 = 128 changes
   - Philippines, 1982 = 95 changes
   - Singapore, 1989 = 74 changes
   - Madrid, Spain, 1993 = 71 changes
   - Auckland, New Zealand, 1997 = 34 changes

   a. How do the number of changes to the sequence relate to time?

   b. What do you think this means?