Sizes of the *Escherichia coli* and Human Genomes

8

Introduction

DNA is stored in cells in the form of chromosomes and plasmids. The amount of DNA required to store the information necessary for making even a simple organism such as a bacterial cell is very large. One of the wonders of biology is that cells are able to store and access the great lengths of DNA needed to encode their hereditary information.

The bacterium *Escherichia coli* is estimated to have about 2,000 genes in its genome. The average bacterial gene is considered to contain 1,200 base pairs; allowing for noncoding regions, the DNA of *E. coli* is estimated to be about 4 million base pairs long. A typical plasmid is about 3,000 base pairs long and encodes just a few genes.

What are the physical sizes of these DNA molecules? The *E. coli* chromosome consists of one large circular DNA molecule that if stretched out would be approximately 10^{-3} m (1 mm) long. By comparison, the *E. coli* cell is only 1×10^{-6} to 2×10^{-6} m long. The *E. coli* DNA molecule is thus 1,000 times longer than the cell! Even a lowly plasmid of 3,000 base pairs would be 10^{-6} m long, or approximately the length of the cell, if it were linear. Yet the chromosome of *E. coli* constitutes only 2 to 3% of the cell's weight and occupies only 10% of its volume.

DNA occupies such a small fraction of the cell's volume (considering the enormous length of the DNA molecule) because it is an extremely slender molecule. It is capable of a high degree of folding and coiling, an essential feature for packing it into the cell. Although the degree of folding required to fit the DNA of *E. coli* into the bacterium is impressive, the folding necessary for packaging DNA into a human cell is even more remarkable.

The human cell is approximately 2×10^{-5} m in diameter. The human genome is estimated to consist of about 3 billion base pairs. If the DNA of a single human cell were stretched out, it would be about 2 m long, or 100,000 times longer than the cell! Yet all this DNA not only fits into the cell but is also accessible to the cell's enzymes for information transfer and replication.

This activity uses models that are 10,000 times life size to demonstrate the relationship between the sizes of an *E. coli* cell, its chromosome, its plasmid, and a single gene. The $\times$ 10,000 *E. coli* is represented by a 2-cm gelatin capsule, the $\times$ 10,000 *E. coli* chromosome is represented by a 10-m length of thread, the plasmid is represented by 10 mm of thread, and a gene is represented by 4 mm of thread.

Materials

Obtain from your teacher a letter-size envelope containing the following items.

- A 2-cm (20-mm) gelatin capsule
- 10 m of thread
- An index card with 4 mm of thread labeled "Average length of bacterial gene"
- An index card with 10 mm of thread labeled "Length of a typical bacterial plasmid"

At the direction of your teacher, form groups of four for the next steps.

1. Remove the gelatin capsule from the envelope. It represents a single *E. coli* bacterium that has been enlarged 10,000 times.

2. Remove the two index cards from the envelope. The lengths of thread or string on the cards represent the length (but *not* the diameter) of an *E. coli* gene and plasmid magnified 10,000 times.

3. Remove the thread from the envelope, and stretch it out. The thread represents the bacterial chromosome magnified 10,000 times.

4. Let two people make a circle with the thread. The *E. coli* chromosome is circular and is attached to the cell membrane.

5. A third person can now hold up the index card with the $\times$10,000 bacterial gene next to the DNA loop. The average bacterial gene contains about 1,200 base pairs. Remember that a gene is composed of all the segments of DNA that instruct the cell to make a single protein, whether those segments are continuous or not.

6. With the bacterial chromosome and bacterial gene models still in view, a fourth person can hold up the index card with the $\times$ 10,000 bacterial plasmid. Plasmids carry one or a few genes necessary for their own replication and stability and often carry genes that give the bacterium important characteristics such as antibiotic resistance. You can see that the plasmid is tiny in comparison to the chromosome.

7. Now that you have compared the sizes of the chromosomes, plasmids, and genes, try to reconstruct the bacterium by inserting the "chromosome" into the capsule. It isn't easy! The real *E. coli* chromosome occupies about 10% of the cell volume.

Questions

1. How could two 10-m lengths of thread represent the *E. coli* chromosome more accurately?

2. How many bacterial genes would fit on your DNA circle (formed in step 4)?

3. Is the thread that you tried to stuff in the capsule too thick to represent the DNA's actual thickness? What is the reason for your answer? (Hint: What percentage of "bacterial cell" volume does the thread occupy in your model, and what is the actual volume that DNA occupies in *E. coli?*)

4. In this activity, how many meters of thread did it take to represent the *E. coli* genome? If the human genome is 1,000 times longer than the *E. coli* genome, how many meters would it take to represent the human genome? How many miles of thread would that be?

Mathematical Calculations

- The distance between DNA base pairs is 3.4×10^{-10} m.
- The *E. coli* chromosome contains about 3×10^6 base pairs.
- The average *E. coli* gene contains 1,200 base pairs.
- A typical plasmid contains about 3,000 base pairs.

Using the information given above, calculate the following.

5. How long (in meters) is the *E. coli* chromosome?

6. How long (in meters) is the average *E. coli* gene?

7. What is the circumference (in meters) of a typical *E. coli* plasmid?

8. If *E. coli* were magnified 10,000 times, how long (in meters) would its chromosome be?

9. If *E. coli* were magnified 10,000 times, how long would its average gene be?

10. If *E. coli* were magnified 10,000 times, how long would a typical plasmid be?

The human genome can be related to a length of railroad track. The railroad ties represent the base pairs, and the rails represent the sugar-phosphate backbone of the DNA molecule. The railroad ties are 2 ft apart.

11. The human genome contains 3×10^9 base pairs. How many miles of track will it take to represent the human genome?

12. The circumference of the Earth is 24,000 miles. How many times would the railroad track representing the human genome wrap around the Earth at the equator?

(continued on next page)

13. Another way to represent the size of the human genome is to relate the base pairs to characters on a page in a book. Calculate how many of these books it would take to represent the human genome in the following manner.
    - Choose a page in your text that is mostly print.
    - Count the number of characters on five randomly selected lines. Find the average number of characters per line (C). Record C.
    - Count the number of lines on the page (L). Record L. Calculate the average number of characters (N) on a page by multiplying C times L. Record N.
    - Calculate the number of characters in your text (T) by multiplying N by the number of pages in your text. Record T.
    - To determine how many books like your text it would take to represent the human genome if every character represented a base pair, divide the number of base pairs in the human genome (3×10^9) by the number of characters in your text (T). How many books would be required?