TEACHING THE

GENOME GENERATION

Sequence Comparison with ACE

The Jackson Laboratory
Leading the search for tomorrow's cures

NIH National Institutes of Health
Turning Discovery Into Health

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
SUPPORTED BY THE NATIONAL INSTITUTES OF HEALTH

Sequence Comparison with *ACE*

Did you know that humans are, on average, 99.9% genetically identical? Only 0.1% of all our DNA bases are different, but those differences are what influence our traits and help make us each who we are.

The angiotensin I converting enzyme (*ACE*) gene codes for the protein angiotensin-converting enzyme (ACE), which functions as a protease that cuts other proteins. ACE plays a central role in the system that controls blood pressure by regulating the volume of fluids in the body. Variants in the human *ACE* gene are associated with differences in athletic endurance performance. In this activity, you'll compare ACE DNA sequences from different individuals and different organisms.

When comparing DNA sequences, **percent identity** provides a measure of how similar two sequences are. The formula for percent identity uses the total number of nucleotide **positions** in the sequence comparison and the number of nucleotide positions that are different, or **divergent**, between the sequences:

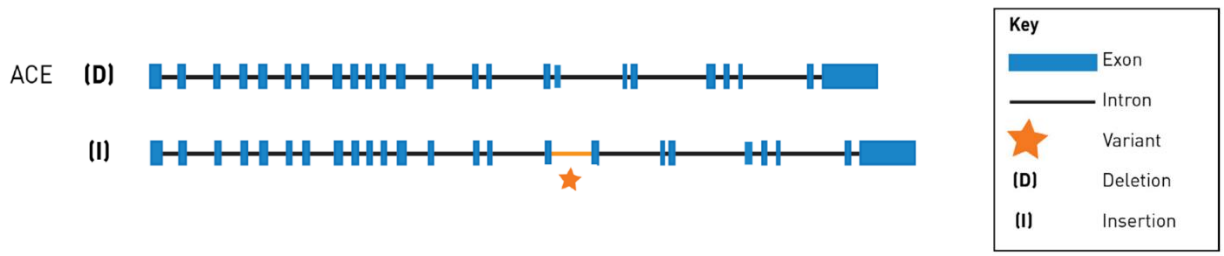
$$Percent\ Identity = \frac{\#\ positions - \#\ divergent\ positions}{\#\ positions}\ x\ 100\%$$

Part 1. Compare to a Reference Sequence

One common type of sequence comparison is comparing an individual's DNA sequence to a reference sequence. A **reference sequence** is a DNA sequence that is assumed by scientists to be a representative example of the genetic material of a specific species. Reference sequences are typically created by combining the DNA sequences of multiple individuals from the same species.

Comparing an individual person's DNA to a reference sequence allows us to identify variants, or differences, between that person's DNA sequence and the reference.

Below are box-line diagrams representing the two common alleles of the *ACE* gene. Box-line diagrams are a common visual representation of a gene structure where boxes indicate the parts of the gene that code for proteins (also called exons), and the black lines are the regions in between called introns.



The insertion (**I**) allele of the *ACE* gene has an insertion of 287 base pairs within intron 16. The deletion (**D**) allele of the of the *ACE* gene does not contain the 287 base pair insertion.

The entire human *ACE* gene sequence, including the 287 base pair insertion, is 21,597 nucleotides.

1. You want to compare an individual person's DNA sequence for the *ACE* gene to the human reference sequence for *ACE* to identify which *ACE* gene variant that person carries. We'll call this person **Jean**.

   Jean's *ACE* DNA sequence contains the insertion, while the reference sequence contains the deletion. What is the percent identity of this gene for this comparison? You can assume that all of the nucleotides are the same between the two sequences besides the 287 nucleotide insertion. Round your answer to two decimal places.

Sometimes researchers know which gene a DNA sequence is from, but other times they don't. Comparing sequences and calculating percent identity can help them figure out which gene their sequence is from. Since humans are 99.9% genetically identical to each other on average, a very high percent identity provides more confidence that the sequence is from a given gene.

2. If you didn't know which gene **Jean**'s DNA sequence was from, would you feel confident that this sequence was from *ACE* given the percent identity you calculated in question 1? Why or why not?

3. What if the percent identity was 95%? 75%? Justify your answer.

Part 2. Compare Across Species

Imagine that you are a researcher studying athletic performance, and you want to learn more about the *ACE* gene. You plan to use a model organism to conduct some experiments to better understand the impact of the *ACE* gene on physical endurance. Model organisms, such as mice and fruit flies, are often used as a representation of human biology because they are easier to study in controlled environments and share much of the same physiology as humans.

You discover that mice also have an *ACE* gene. How similar is the mouse *ACE* gene to human *ACE*?

Start by looking at a small section of the coding region of the *ACE* gene. The sequence comparison below compares a portion of the ACE gene for two sequences:

- The **Query** sequence is the **mouse** reference sequence for *ACE.*
- The **Subject** ("Sbjct") sequence is the **human** reference for *ACE*.

When comparing across species, it can be helpful to record which sequence corresponds to which organism. In the figure below, write in which sequence is **mouse** and which is **human**.

```
_______ Query  1    TTGTATGAGTCCATTTGGCAGAACTTTACTGACTCAAAGCTGCGAAGGATCATCGGATCT  60
_______ Sbjct       C.......AC.G..C...........C..G...C.GC.......C............G..
```

1. How many nucleotides are different between the two sequences? What types of differences are they?

2. In this comparison, there are 60 positions total. What is the percent identity of this section of the *ACE* gene?

Now, look at a different section of the *ACE* gene and see how it compares between mouse (**Query**) and human (**Sbjct**).

```
_______ Query  1    GGCTCTACAACATCCGTAACCATCACAGCCTCCGCCGGCCCCACCGTGGGCCCCAGTTTG  60
_______ Sbjct       ......T..G......---...C.G........A.....A.TC..AC...........C.
```

3. How many nucleotides are different between the two sequences? What types of differences are they?

4. What is the percent identity of this section of the *ACE* gene?

The percent identity for the comparison of the whole protein coding sequences of human *ACE* and mouse *ACE* is 83%. The percent identity of mouse and human genes is 85% on average, but it varies from 60% to 99% for individual genes.

5. Given an 83% identity for *ACE* and knowing that mice are often used as a model organism for human biology, would you feel confident using mice as a model to study the impact of *ACE* on endurance performance?

6. What other information might you want to know about the mouse *ACE* gene or protein to help you make your decision?

Comparing DNA sequences across species can also provide information about how species evolved over time. In general, the greater the percent identity of DNA sequences between two species, the more recently they have shared a common ancestor. It takes time for genetic differences to accumulate, so organisms with fewer genetic differences are typically more closely related.

7. Which species would you expect to have a more similar *ACE* DNA sequence to humans: mice or chimpanzees? Justify your answer.

Check your hypothesis by comparing a portion of the chimpanzee *ACE* reference sequence (**Query**) and the human *ACE* reference sequence (**Sbjct**).

```
_______Query  1     CTGTATGAACCGGTCTGGCAGAACTTCACGGACCCGCAGCTGCGCAGGATCATCGGAGCT  60

_______Sbjct        ............A...............................................
```

8. How many nucleotides are different between the two sequences? What types of differences are they?

9. What is the percent identity of this section of the *ACE* gene?

10. Compare your percent identity calculations for questions 2 and 4 (mouse) with your calculation for question 9 (chimpanzee). Do they support your hypothesis about whether the mouse or chimpanzee *ACE* DNA sequence is more similar to the human sequence? Explain your reasoning.

11. If not, why do you think that might be?

Part 3. Compare Within Species

Another type of sequence comparison is comparing the DNA sequences of different genes within the same species. Comparing different genes within the same species can help scientists identify gene families.

Gene families are groups of genes with similar functions. Comparing sequences helps identify gene families because a gene's sequence determines its associated protein's structure, which determines protein function. This is especially useful in organisms where a full genome sequence is not known. By comparing new gene sequences to known genes, scientists can determine if the new gene serves a similar function to a known gene.

Comparing genes within species can also provide us information about evolution. Sometimes, as species evolve, genes get duplicated. Over time, these gene duplicates accumulate changes and become different enough that they serve different, yet related, functions.

Let's look at an example. *ACE* is in the same gene family as Angiotensin Converting Enzyme 2 (*ACE2*). How similar are these two sequences?

Start by comparing a small section of the human *ACE2* reference sequence (**Query**) to the human *ACE* reference sequence (**Sbjct**). When comparing sequences, it can be helpful to record the source of each sequence. In the figure below, write in which sequence is from ***ACE2*** and which is from ***ACE***.

```
_______ Query  1   CCAATTCCAGTTTCAAGAAGCACTTTGTCAAGCAGCTAAACATGAAGGCCCTCTGCACAA  60
_______ Sbjct      ...G........C..C..G.....G..C..G......GGC..CACG.....C........
```

1. How many nucleotides are different between the two sequences? What types of differences are they?

2. What is the percent identity for this comparison?

3. Given this percent identity and what you know about the function of ACE, what might you predict the function of ACE2 in the body to be?