TEACHING THE GENOME GENERATION

Sequence Comparison with TAS2R38

Sequence Comparison with *TAS2R38*

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 Taste 2 Receptor Member 38 (*TAS2R38*) gene produces the Taste 2 Receptor Member 38 (TAS2R38) protein, which functions as a receptor to perceive a wide range of bitter compounds. Variants in the human *TAS2R38* gene are associated with differences in ability to taste specific bitter compounds. In this activity, you'll compare *TAS2R38* 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's DNA to a reference sequence allows us to identify variants, or differences, between the individual's DNA sequence and the reference.

In this example, we will compare an individual person's DNA sequencing data for the *TAS2R38* gene to the human reference sequence for *TAS2R38* to identify which *TAS2R38* gene variant that person carries. We'll call this person **Jean**.

The sequence comparison below looks at a portion of the *TAS2R38* gene for two sequences:
- The **Query** sequence is **Jean**'s *TAS2R38* sequence data.
- The **Subject** ("Sbjct") sequence is the human **reference** sequence for *TAS2R38.*

When comparing sequences, it can be helpful to record the source of each sequence. In the figure below, write in which sequence is from **Jean** and which is the **reference** sequence.

```
_____Query  1    GTGAATTTTTGGGATGTAGTGAAGAGGCAGCCACTGAGCAACAGTGATTGTGTGCTGCTG  60

_____Sbjct       ...............................G............................
```

1. How many nucleotides are different between the two sequences within this portion of the *TAS2R38* gene? What is/are the variant nucleotide(s) in the individual's DNA sequence and the reference DNA sequence?

2. In this comparison, there are 60 positions total. What is the percent identity for this comparison?

3. The entire human *TAS2R38* gene sequence is 1,002 nucleotides. What is the percent identity for the entire *TAS2R38* sequence? You can assume that all the other nucleotides are the same between the two sequences. 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.

4. If you didn't know which gene **Jean**'s DNA sequence was from, would you feel confident that this sequence was from *TAS2R38* given the percent identity you calculated in question 3? Why or why not?

5. What if the percent identity was 99%? 95%? 75%? Justify your answer.

Part 2. Compare Across Species

Imagine that you are a researcher studying taste perception, and you want to learn more about the *TAS2R38* gene. You plan to use a model organism to conduct some experiments to better understand the impact of the *CYP2C19* gene on the ability to taste bitter compounds. Model organisms, such as mice and fruit flies, are often used as a representation of human biology and behavior because they are easier to study in controlled environments and share much of the same physiology as humans.

Although you're interested in human taste perception, you plan to use a model organism. Model organisms, such as mice, are often used as a representation of human biology and behavior.

When you try to find the sequence for *TAS2R38* in mice, you are surprised to find that mice do not have a *TAS2R38* gene! Mice instead have a very similar gene called Taste 2 Receptor Member 138 (*TAS2R138)*. These two genes are part of the same gene family, or group of genes with similar functions.

You decide to compare a small section of the mouse reference sequence for *TAS2R138* (**Query**) to the human reference sequence for *TAS2R38* (**Sbjct**). 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    GTAAATGTTTGGGATGTGGTAAAAAAGCAGCCCTTGAACAACTGTGACATCGCACTGCTG  60

_____ Sbjct       ..G...T..........A..G..G.G....G.AC...G....A....TTGT.TG......
```

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?

Now, look at a different section of mouse *TAS2R138* (**Query**) and human *TAS2R38* (**Sbjct**).

```
_____ Query  1    -AGATTTCTTCAGATGCTTCTAGTTGTTCTTCTTCTCTCCTGCATCTGCACTGCCCTTTG  60

_____ Sbjct       A....C..-C........C..G.G.A..A.....TG.................T...C..
```

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

4. What is the percent identity for this comparison?

The percent identity for the comparison of the whole protein coding sequences of human *TAS2R38* and mouse *TAS2R138* is 76%. The percent identity of mouse and human genes is 85% on average, but it varies from 60% to 99% for individual genes.

5. Given a 76% identity for human *TAS2R38* and mouse *TAS2R138*, would you feel confident using mice as a model for your research on taste receptors? Justify your answer.

6. What is one other system or method you might be able to use to study human *TAS2R38* and taste perception?

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 DNA sequence to human *TAS2R38*: mice or chimpanzees? Justify your answer.

Check your hypothesis by comparing a portion of the chimpanzee *TAS2R38* reference sequence (**Query**) and the human *TAS2R38* reference sequence (**Sbjct**).

```
_____ Query  1     GTGAATTTTTGGGATGTAGTGAAGAGGCAGCCACTGAGCAACAGTGATTGTGTGCTGCTG  60

_____ Sbjct        ..............................G.............................
```

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 *TAS2R38* 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 *TAS2R138* or chimpanzee *TAS2R38* DNA sequence is more similar to the human *TAS2R38* 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. *TAS2R38* is in the same gene family as taste 2 receptor member 3 (*TAS2R3*). How similar are these two sequences?

Start by comparing a small section of the human *TAS2R3* reference sequence (**Query**) to the human *TAS2R38* 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 ***TAS2R3*** and which is from ***TAS2R38***.

```
_____  Query  1    TGGCTTGCCACCTGTCTTGGTGTCCTCTACTGCCTGAAAATCGCCAGTTTCTCTCACCCC  60

_____  Sbjct       ........TG....C..CA.CC.G..T......TCC..GC..AT.C...........A..
```

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?

Overall, human *TAS2R3* has an 41% identity with human *TAS2R38*.

3. Given this percent identity and what you know about the function of TAS2R38, what might you predict the function of TAS2R3 in the body to be?