DNA to Darwin Case Study

A question of taste

Version 1.2

Anne Fischer
Formerly of the Max Planck Institute for Evolutionary Anthropology, Leipzig

Dean Madden
NCBE, University of Reading
Part 1: Introduction

The five basic tastes

Taste is one of our five senses. Humans are able to detect five basic tastes: sweet, sour, salty, bitter and umami (the savoury taste associated with meat and soy sauce). These tastes are detected when molecules bind to receptors in the taste buds, which are sensory organs on the surface of the tongue and mouth. This allows us to detect and distinguish between nutritious food and poisons, which is an important ability, affecting our well-being and even our survival.

You may have seen diagrams showing that different regions of the tongue are sensitive to different tastes, but these are now known to be wrong. Most parts of the tongue are sensitive to all the different tastes, because each taste bud contains receptors for all of the five basic tastes. Taste buds are also found on the roof of the mouth and in the throat.

Taste receptor families

There are several families of taste receptors. The best-understood are those that detect bitter substances (made of T2R proteins, encoded by TAS2R genes) and those that detect sweet and savoury substances (made of T1R proteins, encoded by TAS1R genes). There are other receptors for other tastes.
Bitter taste receptors

Bitter tastes are unpleasant and, when strong, nauseating. Bitter foods include unsweetened cocoa, citrus peel, olives and quinine. Bitter taste receptors are believed to play an important role in helping us to avoid potentially harmful and toxic substances and are thus crucial for survival. There are many sorts of bitter taste receptors, each of which is sensitive to a different bitter substance.

The structure of bitter taste receptors

Bitter taste receptors are transmembrane receptors. The receptors are found in the outer membrane of the cells and they cross the membrane, passing from outside the cell to the inside. Each receptor is made of three parts or domains:

- extracellular domains, which bind to flavour molecules;
- transmembrane domains, which anchor the protein in the cell membrane (all bitter taste receptors have seven of these); and
- intracellular domains, which interact with other proteins to convey the signal to the nerves that carry the impulse to the brain.

Image: Diagram of a transmembrane receptor that detects bitter-tasting compounds.

Sweet taste receptors

Before its discovery in the 1990s, some people doubted whether a 'sweetness receptor' really existed. In 2001, experiments with laboratory mice showed that mice with different versions of the gene TAS1R3 prefer sweet foods to different extents. Subsequent research has shown that the T1R3 protein forms a complex with a related protein, called T1R2, to form a two-part sweetness receptor in mammals.
Sweetness receptors in humans respond to a range of sugars, including sucrose, fructose and glucose. They can be ‘fooled’ by other substances including certain amino acids and sweeteners such as saccharin, aspartame and acesulfame-K. The ability to taste sweet things may differ significantly between species. Even amongst closely-related animals such as primates the ability to taste sweetness varies. For example, New World monkeys do not find aspartame sweet, while Old World monkeys and apes (including humans) all do. Cats have lost the ability to taste sweet substances entirely.

**Umami receptors**

The umami (savoury) taste receptor is similar to the sweetness taste receptor, except that the T1R2 protein is replaced by the T1R1 protein. The umami receptor can detect almost all of the 20 amino acids that are found in nature, so it’s therefore essentially a protein-detecting receptor.

**Questions**

a. Mammals can sense five main tastes: sweet, sour, bitter, salt and umami. What do you think the evolutionary (survival) advantage of each of these abilities might be?

b. Copy and complete the following summary table. Some parts have been filled in for you.

<table>
<thead>
<tr>
<th>TASTE</th>
<th>GENES INVOLVED</th>
<th>PROTEINS ENCODED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umami</td>
<td>TAS1R1 + ?</td>
<td>? + T1R3</td>
</tr>
<tr>
<td>Sweet</td>
<td>? + TAS1R3</td>
<td>? + ?</td>
</tr>
<tr>
<td>Bitter</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

**Naming genes**

For convenience, genes are given shortened names called gene symbols. These are often similar to the names of the proteins they encode. Although the conventions vary, gene symbols are usually written in italic letters.
T1R sequence analysis

The aim of this exercise is to study the similarities between the amino acid sequences of T1R (sweetness and umami) taste receptor proteins. This will allow you to investigate how new taste receptors evolve.

The data provided are the amino acid sequences for 17 taste receptors plus a distantly-related receptor protein to be used as an 'outgroup'. The sequences come from two fish (pufferfish and fugu), a bird (chicken), a marsupial (opossum) and two mammals (dog and human). The protein sequence used for the outgroup is from a small fish called a Medaka (the Japanese killifish). There are three sorts of taste receptors: T1R1, T1R2 and T1R3. The data file is called: T1Rproteins.geneious

The analysis will be performed using a programme called Geneious. First, you will need to align the amino acid sequences. From the alignment, you will use the software to generate a phylogenetic tree. This will show which proteins are more closely related to one another.

1. Double click on the document named T1Rproteins.geneious. This will start the Geneious software and load the file of protein sequence data into the programme. 
   Hint: if a box appears over the Geneious start-up screen, saying that your trial of the 'Pro' version has ended, click on 'Use Geneious Basic'.

2. The 18 amino acid sequences will now open in Geneious:

   ![Geneious software interface]

   The names of the sequences and the organisms they come from are shown here. The amino acid sequences are in this central window.
3. Use the magnifying glass buttons to zoom in on the sequence data and check that the sequences are in fact proteins made of amino acids (the single-letter amino acid codes are used here):

Asp  D  Aspartic acid  Ala  A  Alanine
Glu  E  Glutamic acid  Gly  G  Glycine
Arg  R  Arginine  Val  V  Valine
Lys  K  Lysine  Leu  L  Leucine
His  H  Histidine  Ile  I  Isoleucine
Asn  N  Asparagine  Pro  P  Proline
Gln  Q  Glutamine  Phe  F  Phenylalanine
Ser  S  Serine  Met  M  Methionine
Thr  T  Threonine  Trp  W  Tryptophan
Tyr  Y  Tyrosine  Cys  C  Cysteine

4. Select all 18 sequences at the same time, by clicking on the file name in the top window so that it is highlighted. Next, click on the Alignment button:
5. A box will appear, asking you to choose a method of alignment. Only one method is possible with the basic Geneious software, so select **Geneious Alignment** then click **OK**. The alignment will take a few minutes to complete (slightly longer on a slow computer):

![Alignment window](attachment:image.png)

6. When the alignment process has finished, select the aligned protein sequences (‘Alignment of 18 sequences’) in the top window.

The aligned protein sequences will then appear in the central Geneious window:

![Geneious window](attachment:image.png)

*The aligned amino acid sequences are in this central window.*
7. You can zoom in and see how dashes have been inserted in the protein sequences to create the alignment:

![Alignment View](image)

8. Next, with the aligned sequences still selected in the top window, click on the Tree button to create a phylogeny (evolutionary tree).

A box will appear, offering some options for the tree building. Select the values shown below and click OK.

![Tree button](image)

**IMPORTANT**

You must select ‘Jukes-Cantor’ as the Genetic Distance Model, ‘Neighbor-Joining’ as the Tree build Method and ‘Medaka V2R’ as the Outgroup.

**Technical note**

The Jukes-Cantor distance model assumes that all amino acid substitutions (mutations) happen at the same rate (1 in 20 or 5%). Other mathematical models assume that different amino acids mutate at different rates.

The Neighbor-Joining method is a quick and popular mathematical model for calculating genetic distances and drawing trees. Other methods will produce slightly different results (and take longer to do it).

9. A tree will be produced in the lower central window. Re-size the other windows so that you can study the tree. The software will group similar sequences (that is, similar proteins) closer together on the tree.
Note

Depending upon the options you select in the ‘Formatting’ panel, your tree may not look exactly like the one shown.

The relationships between genes on the tree should be identical, however.

Questions

Before you attempt to analyse the tree and answer the questions below, you might find it helpful to obtain a printed copy of the tree diagram.

c. There should be three clusters of proteins in your tree, plus the outgroup. Which genes are in each cluster?

d. All the proteins in the tree (except for the outgroup) are very similar to one another (you can check this by looking back at their amino acid sequences). Evolutionary biologists infer that this is because the proteins are descended from a common ancestor. Where in your tree would the common ancestor of all the taste receptor proteins be?

e. Which proteins are more closely-related to one another: T1R1 and T1R2; T1R1 and T1R3; or T1R2 and T1R3? How can you tell?

f. Which taste receptor evolved first, and which came next? Which is the most recent receptor to evolve?

g. From your tree, which sense of taste came first: the ability to taste sweet things or savoury flavours? How can you tell?

h. Copy and complete the table on the right, showing which genes each animal has (a couple have been done for you). Which animal in the tree probably cannot taste sugar? Explain the reason for your answer.

i. Why might your answer to Question h, above, be wrong?
The giant panda’s diet

In January 2010, the genome of the giant panda (Ailuropoda melanoleuca) was published.

Most of a panda’s diet is bamboo which, like all plant material, contains cellulose. The researchers didn’t find any digestive cellulase genes in the panda’s genome, however, indicating that the animal probably has to rely upon gut bacteria to digest the cellulose in the bamboo. Unlike a cow or other ruminant, however, the panda lacks a digestive system that is suited to this purpose. It doesn’t have a rumen, for example, and its teeth are like those of meat-eating bears.

The giant panda is in fact a bear, and is classified with other bears in the order Carnivora*, despite the fact that it is predominantly vegetarian — a herbivore. If the giant panda has been designed, it would appear to have been very badly designed.

In the panda genome, the TAS1R2 and TAS1R3 genes are intact, meaning that the animal can taste sweet substances. The TAS1R1 gene has gained mutations, however, and become a pseudogene, so pandas probably cannot taste umami (the flavour of meaty and protein-rich foods). This may explain in part their preference for bamboo rather than meat.

<table>
<thead>
<tr>
<th>Human</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panda</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Dog</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Structure of the umami (protein taste) receptor TAS1R1 gene in humans, giant pandas and dogs. The gene is made up of six exons (coding sequences) broken up by five introns (non-coding sequences). In the giant panda, two of the exons (shown in red) have undergone mutations. The third exon has a two base-pair insertion, and the sixth exon has a four base-pair deletion.

Discussion questions

a. Which do you think came first: the giant panda’s bamboo diet or the loss of its umami taste receptor?
b. If the giant panda’s ability to taste protein-rich foods was restored, what effect might this have on the animal’s diet, ecology and survival chances?

* the so-called red panda, which was once thought to be closely related to the giant panda, is now known to be only a distant relative, which is so distinctive it now has its own family, Ailuridae.
Part 2: Introduction

New and ‘fossil’ genes

A key mechanism in evolution which leads to the emergence of new genes is gene duplication. There are several ways in which genes can be duplicated: all of them are complex and it remains unclear exactly how the process happens. About 40% of the genes in the human genome have arisen in this way, however.

When a gene is duplicated the DNA sequences of each copy are at first identical. With time, each copy accumulates changes (mutations) and the two copies gradually become different. Nevertheless, even after a long time, parts of the genes will be identical and this similarity can be used to identify genes that have been duplicated and are part of the same ‘multigene family’.

Immediately after duplication, the two genes both carry out the same function. One copy is not really needed however: it is effectively redundant. Over several generations the new copy of the gene can experience one of several different fates. Sometimes the new copy is beneficial: as it acquires mutations it may gain a new function and become a ‘new’ gene. It can also share part of its function with the original gene. But often, gene redundancy is not advantageous. Many duplicated genes will accumulate harmful mutations and cease to function. These non-functional copies often remain like ‘fossils’ in the genome; the technical term for them is pseudogenes.

Taste receptors: multigene families

Taste receptors, such as those you have studied in the previous exercise, belong to multigene families.

In the human genome, 25 genes encoding bitter taste receptors (called TAS2Rs) have been identified. In addition, ten related pseudogenes have been found. As a comparison, the mouse genome contains 33 bitter taste genes but only three pseudogenes. The presence of multiple taste receptor genes that are obviously related shows that several gene duplication events have taken place.

Different animals live in very different environments and their diets are dramatically different. We would therefore expect to see differences in the number of active taste receptors. In adaptation to diverse environments, some receptors might have been duplicated in one species and not in another. A comparison of the protein sequences of taste receptors from species allows us to see how this gene family has evolved. We can also work out the order in which the duplication and the inactivation (pseudogene) events have taken place.
Bitter taste receptors in primates

Previously, we have said that new genes can arise by duplication of existing genes followed by mutation. After a gene is duplicated, one of the copies will not be essential to the organism. It will be under less selection pressure, and can accumulate mutations, allowing it to evolve a new function, or, the copy gene can be inactivated by the accumulation of mutations disrupting the gene (such as ‘Stop’ mutations). This creates a pseudogene (a non-functional, ‘fossil’ gene).

Thus, if we compare the number of mutations in the DNA sequence of a gene and its copy we might expect to find more mutations in the copy.

Harmful mutations

Consider this DNA sequence:

DNA: ATG CCA TAC GCA CCA
Amino acid: Met Pro Tyr Ala Pro

If the red C mutates to an A, a ‘Stop’ mutation arises and the protein sequence is cut short. The protein cannot work because it is incomplete:

DNA: ATG CCA TAA GCA TGA
Amino acid: Met Pro Stop

Depending on where the ‘Stop’ mutation arises in the protein, part of the protein’s function might remain. For example, if the mutation occurs near the end of the amino acid chain in the protein, the function may remain, although maybe with reduced efficiency.

Usually, harmful mutations are quickly ‘weeded out’ by natural selection: often such mutations are lethal and they quickly vanish. Even if a harmful mutation is not immediately lethal, any individuals that have such mutations may be less successful and leave fewer offspring and after several generations the harmful gene will be eliminated from the population. The technical term for this ‘weeding out’ is selective constraint. When a gene has been duplicated, however, such selective constraint is reduced or relaxed. This is because even if one copy of the gene gains harmful mutations, the other copy will still be able to function correctly.

In this exercise, you will study the changes in the DNA sequences of two sorts of taste receptor collected from seven different primate species: bonobo, chimpanzee, human, gorilla, orangutan, rhesus macaque and baboon.
Sequence analysis

1. Double click on the document named Bittertastereceptors. geneious. This will start the Geneious software and load the file of DNA sequence data into the programme. **Hint:** if a box appears over the Geneious start-up screen, saying that your trial of the 'Pro' version has ended, click on 'Use Geneious Basic'.

2. The 14 ready-aligned DNA sequences will now open in Geneious:

![Geneious software interface with aligned DNA sequences]

3. You can use the zoom buttons to enlarge the view of the sequences in the main Geneious window, and confirm that the sequences are in fact DNA.

![Geneious software interface with zoom buttons highlighted]

These dashes show where the sequences have been aligned.

If you select the WRAP button, the sequences will be arranged in several sets of rows, one above the other, with each set of sequences continuing in rows below where the last ones ended.
4. When you zoom in, you’ll see that the DNA data consists of four nucleotide bases: A (adenine), T (thymine), G (guanine) and C (cytosine). The numbers of the bases are shown at the top of each set of rows (see screenshot below).

You will also see that there are two sets of genes: one set of T2R16 genes and one set of $ps1$ genes. The two genes are thought to be copies of each other that have been produced by gene duplication.

Notice that the start or each sequence is ‘ATG’ — the ‘Start’ codon for DNA, and the end of each sequence is ‘TAG’ — a ‘Stop’ codon for DNA.

5. While you are zoomed in on the DNA sequences, quickly scan the data by eye. Compare the DNA sequences of the two genes from the seven different primates, first the T2R16 genes, then the $ps1$ genes.

6. Next, compare (again, just by eye) the sequences of the two different receptor genes from the same species. To help you do this, notice that you can move different sequences alongside each other by clicking and dragging on the sequence names on the left (see screenshot above).

Questions

a. By eye, does it look like the taste receptor gene sequences are similar in the seven primates?
b. How many bases make up the DNA sequences of each bitter taste receptor gene? (Hint: Click on a sequence and look at the statistics panel — see page 15.)
c. How would you calculate the probability that two genes shared their sequences by chance?
You can move sequences next to each other by clicking and dragging.

You can select different sequences at the same time by holding down the ⌘ or the % key (on a Mac) or the Ctrl key (on a PC) as you click.

The statistics panel on the bottom right of the Geneious screen allows you to compare data from two or more selected DNA sequences.

<table>
<thead>
<tr>
<th>Identity</th>
<th>humanT2R16</th>
<th>chimpanzeeT2R16</th>
<th>bonoboT2R16</th>
<th>gorillaT2R16</th>
<th>orangT2R16</th>
<th>rhesusT2R16</th>
<th>baboonT2R16</th>
<th>humanPs1</th>
<th>chimpanzePs1</th>
<th>bonoboPs1</th>
<th>gorillaPs1</th>
<th>orangPs1</th>
<th>rhesusPs1</th>
<th>baboonPs1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73</td>
<td>148</td>
<td>219</td>
<td>292</td>
<td>365</td>
<td>73</td>
<td>148</td>
<td>219</td>
<td>292</td>
<td>365</td>
<td>73</td>
<td>148</td>
<td>219</td>
<td>292</td>
</tr>
</tbody>
</table>

Note: to compare the DNA sequences by eye, you would have to zoom in from the view shown here.

d. By eye, can you tell whether the two receptor genes (T2R16 and ps1) are likely to have been derived from one another?
e. What tools are there in the Geneious software to help you compare gene sequences?
**Drawing an evolutionary tree**

In the previous exercise you produced a phylogenetic tree to investigate the relationships between various taste receptors. You can do the same for the two bitter taste receptor genes (T2R16 and ps1) from the seven primates.

This time, however, we will use DNA sequence data to create the tree.

1. With the DNA sequences selected in the top window, click on the **Tree** button at the top of the screen. A window with several options will appear. Ensure that the Genetic Distance Model is **Jukes-Cantor** and that the **Tree build Method** is **Neighbor-Joining**, then click **OK**.

![Tree button](image.png)

**Technical note**

The Jukes-Cantor distance model assumes that all nucleotide substitutions (mutations) happen at the same rate (1 in 4 or 25%). Other mathematical models assume that different bases mutate at different rates.

The Neighbor-Joining method is a quick and popular mathematical model for calculating genetic distances and drawing trees. Other methods will produce slightly different results (and take longer to do it).

![Geneious window](image.png)

2. A tree like the one below will appear in the main Geneious window:

![Geneious tree](image.png)

**Note**

Depending upon the options you select in the ‘Formatting’ panel, your tree may not look exactly like the one shown.

The relationships between genes on the tree should be identical, however.
Questions

f. How are the two taste receptor genes arranged in the tree?
g. What does this suggest about their evolutionary relationships?
h. Why do you think that the two parts of the evolutionary tree produced by the two genes show different relationships?

Which gene is under less selection pressure?

You will now try to discover which of the two genes (T2R16 or Ps1) is under less selection pressure is therefore evolving faster. To do this you will have to search for mutations in the genes.

A common sort of mutation is a STOP mutation, which appears in the DNA as ATC, ATT or ACT (corresponding to the RNA codons UAG, UAA and UGA). It would be tedious to search through all the DNA sequences for these mutations; fortunately the software can do this for you.

The first step is to translate the DNA sequences into their corresponding amino acid sequences.

3. Ensure that you select the DNA sequences you have imported in the top window, then click on Translate. A dialogue box will appear, asking you which version of the genetic code you wish to use:

4. Although you are going to use the ‘Standard’ genetic code, you may like to see the other, slightly different, versions that are available from the drop-down menu.
5. Once you have seen the different options available, choose the ‘Standard’ code and click OK. The protein sequence data should appear in the main Geneious window, and also in the upper window:

![Geneious window with protein sequence data]

6. If you scroll to the end of the amino acid sequences, you’ll notice that they all end with an asterisk: this represents a ‘Stop’ codon, signifying the end of the protein. You may also notice that the baboon psi protein is one amino acid shorter than the other sequences:

![Amino acid sequences with asterisks]

Consequently, so that the amino acid sequences may be properly compared, it is necessary to align them, in a similar way to that in which the amino acid sequences were aligned in the previous exercise.

7. Select the translated sequences in the upper window, then click on the ‘Alignment’ button:

![Select the translated sequences in the top window, then click the 'Alignment' button.]

Asterisks in the sequences represent ‘STOP’ codons.

The baboon psi protein is one amino acid shorter than the other psi sequences.
8. You can now identify disrupting mutations, that is, changes that produce a 'Stop' mutation. By looking through the sequences, identify which of the two copies was inactivated (remember, 'Stop' mutations are shown as asterisks (*) in the protein sequences).

9. Mark the positions of any 'Stop' mutations on a copy of this species tree:

Another way of looking for mutations is to compare pairs of genes, and to see how much they differ from one another.

You will now compare each gene with the human version to see which one has more mutations.

10. Return to the DNA sequences of the receptors by clicking on the appropriate line in the top window:

The DNA sequences will now appear in the main Geneious window.
11. In the main Geneious window, select the human T2R16 gene by clicking on the sequence name on the left. Hold down the button or the key (on a Mac) or the key (on a PC) as you click and simultaneously select the chimpanzee T2R16 gene (you may wish to click and drag to move the two gene sequences so that they are next to each other — see page 15):

12. While the two genes are selected, look at the Statistics panel on the lower right of the screen. This tells you how long the DNA sequence is (945 bases) and how many sites (bases) are shared by the human and the chimp (Identical sites = 941).

13. Write down the number of shared bases ('Identical sites') in the table below and work out the number of differences (the first one has been done for you). Also record the percentage differences in another table.

<table>
<thead>
<tr>
<th>Gene pairs</th>
<th>Receptor T2R16</th>
<th>Receptor ps1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shared</td>
<td>Different</td>
</tr>
<tr>
<td>Human-Chimpanzee</td>
<td>941</td>
<td>4</td>
</tr>
<tr>
<td>Human-Bonobo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-Gorilla</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-Orangutan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-Rhesus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-Baboon</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Further reading


Note

*These papers are available free-of-charge online.*

Additional references are listed in the teacher’s notes.

---

14. Repeat this process for each pair of T2R16 genes in turn, completing the tables as you go.

15. Do the same for the pairs of ps1 genes, again filling in the tables as you go.

**Questions**

i. Which gene (T2R16 or ps1) has accumulated the greatest number of mutations?

j. Which gene (T2R16 or Ps1) therefore seems to be under less selection pressure?

k. If you were generating an evolutionary tree based on the data from one of these taste receptor genes alone, which one do you think would be the best gene to choose? Explain the reasons for your suggestion.