Relative abundance of microbes

After processing the sequencing data, we obtain the relative abundance of all microbes found in each sample. The microbes are identified at the genus level. Only the ten most abundant microbes are listed in the chart; the others are grouped together into an “Other” category. There is also an “Unclassified” category that includes sequences that were not identified. (Perhaps you have discovered a new species! – or it could just be an error.) Note that these charts only show the relative abundance of each microbe, they do not quantify the absolute amount of each microbe present.

Results are presented in the format below. The sample name includes the type of sample tested (RC, raw cheese and RM, raw milk in this case) and which replicate (1 and 5 in this case). The relative abundances are also represented in pie charts.
There is likely to be a wide diversity of microbes found in each sample. Common microbes such as pseudomonas and lactococcus are often found in many samples. Students are encouraged to use reference tools to independently search for information about the individual microbes found in their samples. (GC = goat cheese; M = Oakhurst milk; HM = Hood milk)
Comparing the microbes found in each sample

In order to compare the microbes in two different samples, we can use a Venn diagram. This Venn diagram only considers the top ten most abundant microbes in the sample, as reported by the pie charts. This diagram ignores the relative amount of each microbe, and only considers whether a particular microbe is among the top ten most abundant in the sample. Any microbes found in BOTH samples are recorded in the overlapping region of the Venn diagram.

Sample 1: RC1
- Lactococcus
- Brevibacterium
- Brachybacterium
- Corynebacterium
- Staphylococcus
- Facklamia
- Alkalibacterium

Sample 2: RM5
- Pseudomonas
- Janthinobacterium
- Gemella
- Acinetobacter
- Yersinia
- Veillonella
- Haemophilus
- Rothia
- Prevotella

For example, we will compare RC1 and RM5 – two results from different dairy products but the same process, not pasteurization or raw. The two samples can be considered as two sets, each containing ten elements, with some redundancy between the two sets. The overlapping region of the Venn diagram represents the INTERSECTION of the two sets.

In the example above comparing RC1 and RM5, there is 1 microbe found in both samples.

The entire non-redundant listing of microbes found in either set represents the UNION of the two sets.

In the example above comparing RC1 and RM5, there are 19 microbes present in either sample.

We can quantify the similarity between samples using Jaccard Similarity, which is calculated by dividing number of classifications found in both samples (intersection) by the total number of classifications found in either sample (union).
J = \frac{\text{(Microbes found in BOTH samples)}}{\text{(Microbes found in EITHER sample)}} \times 100\% \\
In the example above comparing RC1 and M5, the Jaccard similarity is \sim 5.3\% \\
J = \left(\frac{1}{19}\right) \times 100\% = \sim 5.3\% \\
This calculation can be performed for multiple samples. For example, we will compare M3 and HM2 – two results from the same dairy product but different companies, Oakhurst and Hood.

As before, we can calculate the Jaccard similarity, which turns out to be \sim 33\%.

At this point, we can try to answer the question concerning the microbiome relationship between different dairy products.

<table>
<thead>
<tr>
<th>Comparison #</th>
<th>Which Samples?</th>
<th>Jaccard Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RC1 (raw cheese) versus RM5 (raw milk)</td>
<td>5.3%</td>
</tr>
<tr>
<td>2</td>
<td>M3 (Oakhurst pasteurized milk) versus HM2 (Hood pasteurized milk)</td>
<td>33%</td>
</tr>
</tbody>
</table>

Based on these two comparisons, we may conclude that milk and cheese have different influences on the composition of microbiomes. This conclusion may be valid, but it is drawn from a relatively small amount of data. Our conclusion would be stronger if we considered data from comparisons between samples. (In practice, this requires comparisons between many milks and cheeses tested in the classroom.)
<table>
<thead>
<tr>
<th>Comparison #</th>
<th>Which Samples?</th>
<th>Jaccard Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RC1 (raw cheese) versus RM5 (raw milk)</td>
<td>5.3%</td>
</tr>
<tr>
<td>2</td>
<td>M3 (Oakhurst pasteurized milk) versus HM2 (Hood pasteurized milk)</td>
<td>33%</td>
</tr>
<tr>
<td>3</td>
<td>SM1 (smoked mozzarella) versus GC1 (goat cheese)</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>SM1 (smoked mozzarella) versus M6 (Oakhurst pasteurized milk)</td>
<td>5.3%</td>
</tr>
<tr>
<td>5</td>
<td>SC1 (sharp cheddar) versus RC2 (raw cheese)</td>
<td>11%</td>
</tr>
<tr>
<td>6</td>
<td>RM5 (raw milk) versus SC1 (sharp cheddar)</td>
<td>5.3%</td>
</tr>
</tbody>
</table>

As we increase the number of comparisons that we consider, the data increasingly supports students’ theories. This outcome is not guaranteed, and students can draw their own conclusions based on the comparisons they perform.