Exploring gene expression in *Yersinia pestis*

In 541-542, a plague swept in and killed 40% of the population in Europe. Almost 100 years later, a similar plague, infamously referred to as “The Black Death” resulted in the deaths of an estimated 75 to 200 million people, or 30% of the European population, with the epidemic peaking in Europe in the years 1346–53[1]. In 1894, Alexandre Yersin, a Swiss/French physician and bacteriologist from the Pasteur Institute, discovered the bacterium that caused plague, and the organism was named *Yersinia pestis* in his honor in 1967[2]. This bacterium is small, rod-shaped, and is able to infect both fleas and mammals. The bacterium must survive in both hosts to infect humans, as in most cases it is transmitted by the bite of an infected flea.

Fleas have a body temperature of 26°C, while the human body temperature is 37°C. Scientists wanted to explore what genes *Y. pestis* turned on (called up-regulated) and turned off (called down-regulated) at these two temperatures. They looked specifically at a strain of *Y. pestis* that they had made mutant at one critical gene, called Lpp, which was essential for survival in the mammal host[3].

We will examine this experiment and, using PATRIC tools, we will explore the up- and down-regulation of genes at the two different temperatures.

I. Exploring an experiment

1. Login to the PATRIC website so that you can use your workspace in the downstream analysis.

2. On the PATRIC homepage (patricbrc.org), open the Organisms tab at the top of the page.

3. When the tab opens to reveal the box listing the names of pathogens, click on Yersinia.
4. This will take you to the landing page for *Yersinia*, which summarizes all the information that PATRIC has about the genus, including the number of genomes, experiments associated with it, publications on it, and tools that can analyze the available data sorted at that taxonomic level.

5. Find the tab across the top that is labeled “Transcriptomics” and click on it.

6. This will take you to the landing page for the transcriptomics experiments that are available through the PATRIC resource. The left side has a filter to help people find specific experiments, and on the right is a table that summarizes the experiments available.
7. Find the experiment with the title “Gene expression profiles of wild-type Yersinia pestis CO92 and its Braun lipoprotein mutant at flea and human temp” and click the check box in front of it.

8. Next, in the header across the top of the table in the category called “View” click on the “Gene List” tool.

9. This will take you to the Transcriptomics Gene list for the experiment that you chose. There is a filter on the right side, and a table that will show the results of the filter on the left. When the page opens up and before any actions are taken, all genes available in the experiment are listed in the table.
10. To find the experimental conditions and see the two different temperatures, you will first need to expand the filter box by grabbing the right edge with your cursor and then increasing the width of the box.

11. Then you will have to grab the right edge of the column header called “Title” and use your cursor to increase the width of that box. Then you will be able to see which experimental condition was done at 26°C, and which at 37°C.

12. To see which genes are up-regulated at 26°C and down-regulated at 37°C, click the up arrow on the 26C row, and the down arrow on the row for 37C.

13. You will want to see which genes are significantly expressed. To do this, look at the bottom of the filter, and where is says “Filter by |Log Ratio|:” there is an arrow. Click the arrow and you will see a box with numbers in it. Choose 1, and this will give you a fold-change of two (log2). Then click “Filter”.
14. This will filter the table to show that there are 38 genes that meet those criteria. These genes are up-regulated at 26°C (the temperature of the flea host) and down regulated at 37°C (the temperature of the human host).

15. To see all the genes, you will need to scroll down to the bottom of the table and resize it to 38 and hit return.

16. To visualize these genes, PATRIC provides a heatmap view. To see the genes, click on the “Heatmap” tab.

17. You will notice that there are two columns. One is varying degrees of red, and the other is varying degrees of green. The red color signifies up-regulation, so this is the 26°C column. Green is down-regulation (37°C).
18. To get more details on specific genes, you can mouse a single cell of interest. I moused over a light green box in the 37C. When you do that, you can see details on that gene in the blue band at the top of the heatmap. In this example, that gene is the “Outer membrane protein C precursor” that is strongly down-regulated with a log ratio of -4.228.

II. Discovering if genes of interest have any functional significance.

1. From the previous example, return to the table by clicking on the Table tab.
2. In the column titled “Genome Name”, click on the box to the left of it select all the genes.

3. We need to save these features to a workspace. Click on the “Add Feature” and folder.

4. This will open up a pop-up window that allows you to save the group.

5. Select the “Create New Group” option.
6. Name the group and click “Save to Workspace”. Now that data is saved and you can use a number of tools to explore it.

7. This data is now safely in the Workspace, but we can also examine it here from this table. To look at the data with the Pathway tool, you first need to select all the features in this table again.

8. Once the genes are selected, you can use the Pathway Summary tool that you can see at the top of the table under the Tools heading. Click on that.

9. This will return a table that shows you the list of pathways that the genes you selected are involved in.
10. Clicking on any of the pathway names in the table will take you to a KEGG map that shows how many the genes you selected acted as enzymes are involved in that pathway.

III. Exploring specific genes that are significantly expressed in an experiment

1. In the heatmap view above we saw that a gene called “Outer membrane protein C precursor” was strongly down-regulated at 37°C. Return to the workspace by clicking on Workspace at the top of the page.

2. We’re going to return to the group created earlier, so double-click on the name Feature Groups as you did before.

3. Once again, double-click on the name you selected for your group. I called mine “Yersinia pestis up-regulated at 26C”.
4. Scroll down until you see the gene called “Outer membrane protein C precursor” and click on it once to select it.

5. When you click on it, the description panel to the right shows information related to the gene. The informative band that separates the table from the description panel also shows tools that you can use with your selection.
6. Mouse over the icons in the band until you see one that says “View Feature” (it’s the eyeball). Click on that.

7. This will take you to the landing page for that gene where all the information available for it in PATRIC is summarized, including its different gene identifiers, tools and resources that can be used to examine this gene, and any publications that might have been written about it.

8. Along the top panel, click on the tab called “Transcriptomics.”

9. This takes you to all the experimental evidence for these gene that is available in the PATRIC website. The page is broken down to show the expression pattern for the gene across all the available experiments, a list of those experiments, and a circular diagram that can display different experimental conditions. Above the diagrams you will see some tools that allow you to filter the experiments.
10. As in the example above, we are going to filter the experiments displayed by the level of expression. Next to |Log Ratio|: there is an arrow. Click the arrow and you will see a box with numbers in it. Choose 1, and this will give you a fold-change of two (log2).

11. This will set the log ratio to 1, but to see the results you have to click the “Filter” button.

12. This will filter the table and the graphs to show those results. To see a visual representation of the experimental conditions where this gene had this fold-change, click on the “Experimental Condition” button above the bar graph.

13. This will show you that the majority of the experimental conditions where this gene was significantly up- or down-regulated dealt with temperature.
14. You can also see that in the table that has been filtered to match the log ratio selection. You can also see the log ratio values. Note that many of them are negative, as we saw in the initial experiment we were looking at.

<table>
<thead>
<tr>
<th>Title</th>
<th>PubMed</th>
<th>Accession</th>
<th>Strain</th>
<th>Gene Mod.</th>
<th>Experimental Condition</th>
<th>Time Point</th>
<th>Avg Intensity</th>
<th>Log Ratio</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yersinia enterica, OmpF mutant strain (sacBEF4932), 30</td>
<td>139552</td>
<td>GSE92274</td>
<td>ompF mutant vs wild type</td>
<td>30 mins</td>
<td>temperature</td>
<td>0</td>
<td>-0.157</td>
<td>-0.279</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, growth phase, THM in</td>
<td></td>
<td></td>
<td>growth phase</td>
<td></td>
<td>temperature</td>
<td>0</td>
<td>-1.979</td>
<td>-2.194</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, grown at 37°C / Yersinia enterica</td>
<td>139552</td>
<td>GSE92274</td>
<td>temperature</td>
<td>2 hrs</td>
<td>temperature</td>
<td>0</td>
<td>-2.613</td>
<td>-2.322</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, small-mutant, pH autolysis 20 min /</td>
<td>139552</td>
<td>GSE92274</td>
<td>ompF mutant vs wild type</td>
<td>26 mins</td>
<td>temperature</td>
<td>0</td>
<td>-1.46</td>
<td>-3.619</td>
<td></td>
</tr>
<tr>
<td>small-mutant (sacBEF4932), LT</td>
<td></td>
<td></td>
<td>ompF mutant vs wild type</td>
<td>26 mins</td>
<td>temperature</td>
<td>0</td>
<td>-4.173</td>
<td>-6.363</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, 15 degree C / Yersinia enterica</td>
<td>139552</td>
<td>GSE92274</td>
<td>temperature</td>
<td>13 hrs</td>
<td>temperature</td>
<td>0</td>
<td>1.573</td>
<td>1.396</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, ORF (mutant strain of 30°C -)</td>
<td>200729</td>
<td>GSE10840</td>
<td>ompF mutant vs wild type</td>
<td>13 hrs</td>
<td>temperature</td>
<td>0</td>
<td>1.096</td>
<td>1.190</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, Experimental phase, THM / Yersinia p.</td>
<td>139552</td>
<td>GSE92274</td>
<td>growth phase</td>
<td></td>
<td>temperature</td>
<td>0</td>
<td>-1.873</td>
<td>-2.155</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, phoP mutant, pH autolysis 20 min /</td>
<td>139552</td>
<td>GSE92274</td>
<td>ompF mutant vs wild type</td>
<td>26 mins</td>
<td>temperature</td>
<td>0</td>
<td>-2.114</td>
<td>-1.8</td>
<td></td>
</tr>
<tr>
<td>Yersinia enterica, 15 degree C / Yersinia enterica</td>
<td>139552</td>
<td>GSE92274</td>
<td>temperature</td>
<td>13 hrs</td>
<td>temperature</td>
<td>0</td>
<td>-3.063</td>
<td>-3.216</td>
<td></td>
</tr>
<tr>
<td>COX13 Total RNA extracted after 1 hour at 37 degrees</td>
<td>224754</td>
<td>GSE36034</td>
<td>temperature</td>
<td>1 hr</td>
<td>temperature</td>
<td>0</td>
<td>-1.873</td>
<td>-2.846</td>
<td></td>
</tr>
<tr>
<td>PBF13 Total RNA extracted after 2 hours at 37 degrees</td>
<td>224754</td>
<td>GSE36034</td>
<td>temperature</td>
<td>2 hrs</td>
<td>temperature</td>
<td>0</td>
<td>-1.998</td>
<td>-2.35</td>
<td></td>
</tr>
<tr>
<td>PBF13 Total RNA extracted after 6 hours at 30 degrees</td>
<td>224754</td>
<td>GSE36034</td>
<td>temperature</td>
<td>4 hrs</td>
<td>temperature</td>
<td>0</td>
<td>-2.368</td>
<td>-3.053</td>
<td></td>
</tr>
<tr>
<td>PBF13 Total RNA extracted after 2 hours at 37 degrees</td>
<td>224754</td>
<td>GSE36034</td>
<td>temperature</td>
<td>2 hrs</td>
<td>temperature</td>
<td>0</td>
<td>-2.171</td>
<td>-4.669</td>
<td></td>
</tr>
</tbody>
</table>

15. To see what other genes have a similar expression pattern, scroll to the top of the page and select the Correlated Genes tab.

16. The page that returns will show you what genes have a similar expression pattern to the gene we chose, Outer membrane protein C precursor.
17. Notice that the default parameters are set to a correlation cutoff of 0.4, with positive correlation.

![](Correlation_settings.png)

18. Click on the arrow next to positive and select “negative,” then hit filter.

![](Filter_selection.png)

19. This will show you all the genes that are also negatively correlated with this gene.

![](Gene_list.png)

20. At the bottom of the page, resize the table to show all 146 of the negatively correlated genes.

21. At the top of the table, click on the column header called “RefSeq Locus Tag”.

![](RefSeq_sort.png)

22. This will sort the results showing the genes in the order they appear in the genome. Look at the results...talk about **AWESOME!** All the members of a Type III Secretion System are correlated, and have similar correlation values. Certain pathogenic bacteria use these structures to inject proteins into the cell that helps the bacteria infect the host.
Just for fun...here’s a picture of what they think the Type III secretion system looks like and how the proteins line up.

And here’s a transmission electron microscope image of one a Type III secretion system from *Salmonella*. You can clearly see the needle-like apparatus that they use to inject proteins into the host cell.
Assignment: Answer the following questions using the PATRIC website.

1. Choose one of the Type III Secretion System genes and get to the landing page for it. Open the Correlation Tab.
   a. Look at the genes that are positively correlated with the gene you chose. Adjust the correlation cut-off (located above the table) to narrow the number of genes you see.
   b. Which genes are the most strongly correlated?
2. Go to your workspace and select other genes from the group you created Part II, Number 6. Examine the positively and negatively correlated genes for several of them. Are there any pathways that these correlated genes share?
3. Follow the same protocol above, but in this case select the 37°C condition as the one that is up-regulated, and the 26°C condition as the one that is down-regulated. Set the log ratio to 1 and filter. Then follow all the steps listed above, and chose one specific gene to drill down on.
   a. How many genes are significantly up-regulated at that specific log-ratio when you choose those conditions.
   b. How many pathways do the genes appear in?
   c. What gene(s) did you choose to examine?
      i. What log ratio did you filter on to examine this gene, and what experimental conditions surfaced when you changed that?
      ii. When you did the positive and negative correlation, did you see any genes of interest (Hint: The flagellar genes looked interesting to me)?
   d. Bonus question: We haven’t discussed this, but look at the tabs across the gene landing page. How would you look at the gene you chose and its gene neighborhood (the genes around it)? Which tab did you choose and what did you see?

References