The kallisto quant outputs in the .TABULAR format for each sample were used as the inputs for tximport. Transcript-to-gene mapping was done using a table of transcript-to-gene data provided by the Kallisto developers, available on their Github. All other tximport settings were kept at defaults for Kallisto-generated data, and the output was consolidated into gene-level data. tximport generated tables of both counts and abundances by table for all genes covered by the reference genome used in Kallisto. Only the abundance data was used for downstream analysis. Non-DEGs according to DESeq2 were filtered out of the dataset.



R's hclust function was used to perform hierarchical clustering following a tutorial developed by Dr. Hugo Tavares at the Sainsbury Laboratory, Cambridge University. First, data was scaled using R’s scale() function, which converts the original data in a column of a dataframe into z-scores, such that the column has a mean of 0 and a standard deviation of 1. This scaling function is a pre-requisite for clustering using R’s hierarchical clustering function.

Next, a Euclidean distance matrix was generated using dist() and the distance matrix was used as an input for hclust(), which clustered the data using the complete-linkage agglomerative method, where each element begins as an individual cluster and then are sequentially clustered until the entire dataset is consolidated. Clustering was performed along both the *sample* and *gene* axes.

The heatmap with dendrograms to show clustering was generated using the heatmap.2 function in R’s gplots package.

