**Data processing flow**

**# S1\_initial\_processing.**  (Rm-off-by-one for all >99, remove if not in inoc, and make summary files)

##### loop 1 - open and process RAW files for INOC, and make merged list of inoc barcodes

Read: ‘sample\_names.txt'

– contains at least two columns corresponding to output name and raw input name for each sample

raw\_data /", name, ".codes

Write: filtered\_data / name .countsCln

filtered\_data / name .codesCln

filtered\_data / name \_plus.codesCln

threshold\_lists / morethan2prcnt\_", name, ".codesCln.

R\_output / All\_inoc\_samples.codes *(inoc codes with counts in each inoc replicate. Used in S4\_diversity\_assess.)*

R\_output / Inoc\_barcode\_list.txt. *(list of just barcodes detected in all inoc replicates. Used in S2\_Venn.)*

##### loop 2 - open and process RAW files for LN filtering off-by-ones and filter in inoc

Read: raw\_data /", name, ".codes

Write: filtered\_data / ", name, ".countsCln

filtered\_data / ", name, ".codesCln

filtered\_data / ", name, "\_plus.codesCln

filtered\_data / off-by-one /", name, ".codesCln. *(codes after first filter, and before second inoc filter – for Fig S5)*

threshold\_lists / morethan2prcnt\_", name, ".codesCln *(only codes >2% of reads – for Fig 3)*

R-output / Table\_S1\_Sample\_Summary\_Stats.txt. *(summary data statistics)*

##### Assemble counts data for each sample

Read: filtered\_data /", name, ".countsCln

Write: R\_output / All\_LN\_quadrants.counts *(counts data for all LN quadrant samples – for Fig S6)*

R\_output / All\_inoc\_samples.counts *(counts data for all inoc samples – for Fig S6)*

R\_output / Ranked\_Barcode\_counts\_Inoc\_trimmed.txt *(ranked counts for inoc samples – for Fig S6)*

R\_output / Ranked\_Barcode\_counts\_LN\_trimmed.txt *(ranked counts for LN samples – for Fig S6)*

R\_output / All\_LN\_quadrants.codes *(inoc filtered codes w/ counts in LN fragments – for diversity assessment)*

R\_output / All\_PT\_LN\_quadrants.codes *(inoc filtered codes in PT samples w/ counts - for diversity assessment)*

R\_output / All\_PT\_LN\_samples\_without\_inoc\_filter.codes *(includes codes not in the inoculum – for S5 - venn)*

*Used .counts data and Ranked\_bardoce\_counts data for Figure S6 (plotted in prism).*

*Used .codes data in S4\_diversity\_assessment or S2\_venn\_diagrams*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**#S1b\_Plotting\_off-by-one\_codes**

(find all off-by-one from top code in a manually selected sample)

Read: raw\_data /", name, ".codes. *(output from Multicodes.pl)*

‘sample\_names.txt' (contains at least two columns corresponding to each sample output and raw input name)

Write: Off\_by\_ones/", name, "\_topcode.errors

*plotted in prism (Fig S1)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**#S1c\_replicate samples**

Read: R\_output / Inoc\_barcode\_list.txt. *(from S1 part 1)*

raw\_data / replicates\_raw / ", samp1, ".codes". *(output from Multicodes.pl)*

Write: replicates / filtered / ", name, "\_1.countsCln

replicates / filtered / ", name, "\_2.countsCln

replicates / filtered / ", name, "\_1.codesCln

replicates / filtered / ", name, "\_2.codesCln

replicates / pairs / ", name, ".codesCln

*From the paired replicate code lists, plotted replicates in prism (Fig S4)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**#S2\_venn\_diagrams\_of\_barcodes**

Read filtered\_data / Inoc\_x\_( A-C) .codesCln. *(from S1, for inoc replicate overlap)*

R\_output / Inoc\_barcode\_list.txt. *(from S1 part 1)*

abortus\_pool.txt. *(pool file from mapping library)*

R\_output / All\_PT\_LN\_samples\_without\_inoc\_filter.codes " *(from S1 part 3)*

*Reads in code lists. Convert counts to T. Full join lists and replace NA with F.*

*Use euler function in eulerr to calculate and plot overlap in a) inoculum replicates, and b) between inoc, mapped codes and codes detected in parotid LNs*

*Export plots as .eps and make decorative modifications in Illustrator. (Figure S5)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**#S3\_LN\_code\_counts** (groups samples by lymph node and merges counts in each LN)

Read: filtered\_data /", name, ".codesCln" *(from S1)*

Write: LN\_lists/ LN1L.txt

*(for each LN; includes BC, count in each quad, #quads hit, total, mean, median and fractional counts)*

Write: R\_output / CodeCounts\_by\_lymphnode.codes *(all barcodes, with total count in each LN, used in diversity anal.)*

R\_output / Table\_S1b\_LN\_Summary\_Stats.txt *(for each LN, nCodes, nCnt1, fracCnt1; for each cow, nCodes tot and nCodes shared R-L)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**#S4\_Diversity\_assessments**

Setup: Open lists of codes for different types of samples. Merge all lists. Replace NA with 0. Transpose the matrix.

Part A: Calculate standard diversity metrics for figure 5

Part B: generates dissimilarity metrics between samples

Read code lists with counts for inoc, parotid LN pieces, all LN pieces, totLN summaries:

R\_output / All\_inoc\_samples.codes *(inoc samples generated with S1 part 1)*

R\_output / All\_LN\_quadrants.codes *(PT quads AND retro samples generated with S1)*

R\_output / CodeCounts\_by\_lymphnode.codes *(summary of each LN generated with S3)*

R\_output / Barcode\_count\_by\_sample\_groups.txt *(merged list of all samples generated by this script with NA=0)*

Write R\_output / Barcode\_count\_by\_sample\_groups.txt *(merged list of all samples with NA=0)*

R-output / DiversityMetrics.txt *(for figure 2)*

*DiversityMetrics were plotted in Prism (Figure 2)*

*Part B generates plot of Bray-curtis and Jaccard dissimilarity matrix*

*Plots were exported as .eps and modified in Illustrator (Figure 4)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**# S5\_heatmap\_of\_codes>2prcnt**

Read ‘sample\_names.txt'

threshold\_lists / morethan2prcnt\_", name, ".codesCln *(only codes >2% of reads, from S1)*

filtered\_data/", samp, "\_plus.codesCln *(to collect counts from all samples, generated by S1)*

Write R\_output / Morethan2prcnt\_AllOccurances.codes  *(all codes >2% of reads in any sample)*

R\_output / Morethan2prcnt\_uniquecodes.txt. *(list of unique codes >2% of reads in any sample)*

R\_output / BC\_2prcnt\_gathered\_fracCounts.txt *(fractional cnts from all samp of 160 codes >2% in any sample)*

R\_output / BC\_2prcnt\_gathered\_Log10fracCounts.txt  *(log transformed values from table above)*

*\*\*here I open* ‘BC\_2prcnt\_gathered\_Log10fracCounts.txt’ *in excel, manually sort, add a column with BC number and resave as* “R\_output / BC\_2prcnt\_gathered\_Log10fracCounts\_Sorted.txt” and read in this file to make heat map.

*Then generate heatmap from the Sorted file. Export as .eps and modify in Illustrator for Figure 3.*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**# S6\_Probability analysis**

Read: filtered\_data / Inoc\_x\_(A-C).codesCln.  *(inoculum codes with nCounts, from S1)*

"LN\_lists / ", LN, ".txt"  *(summary files with all BC in each LN, from S3)*

"LN\_lists / LN1L\_distribution.txt".  *(generated by this code – frac detected in each inoc bin)*

Write: R\_output / Inoc\_ranked\_summary.txt. *(all filtered inoc codes, with total counts, frac counts, and log10(frac))*

"LN\_lists / ", LN, "\_distribution.txt". *(for each LN, fraction detected in each inoculum bin)*

R\_output / FracNotInLN\_ByAbundanceBin.txt *(summary of frac detected/lost in each bin, for all LNs, w mean)*

R\_output / Prob\_0\_theoretical.txt *(p of loss with different # of trials and different P of success)*

*Used data from the last two files to generate figure 5.*

*Logistic fit (part 4)*

Read: *R\_output/FracNotInLN\_ByAbundanceBin.txt (generated above)*

Write: *R\_output/DoseResponseFit.txt (curve corresponding to parameters from dose response fit)*

*Plotted DoseResponseFit.txt with the Mean +/- SD from FracNotInLN\_ByAbundanceBin.txt for Sigure S7.*