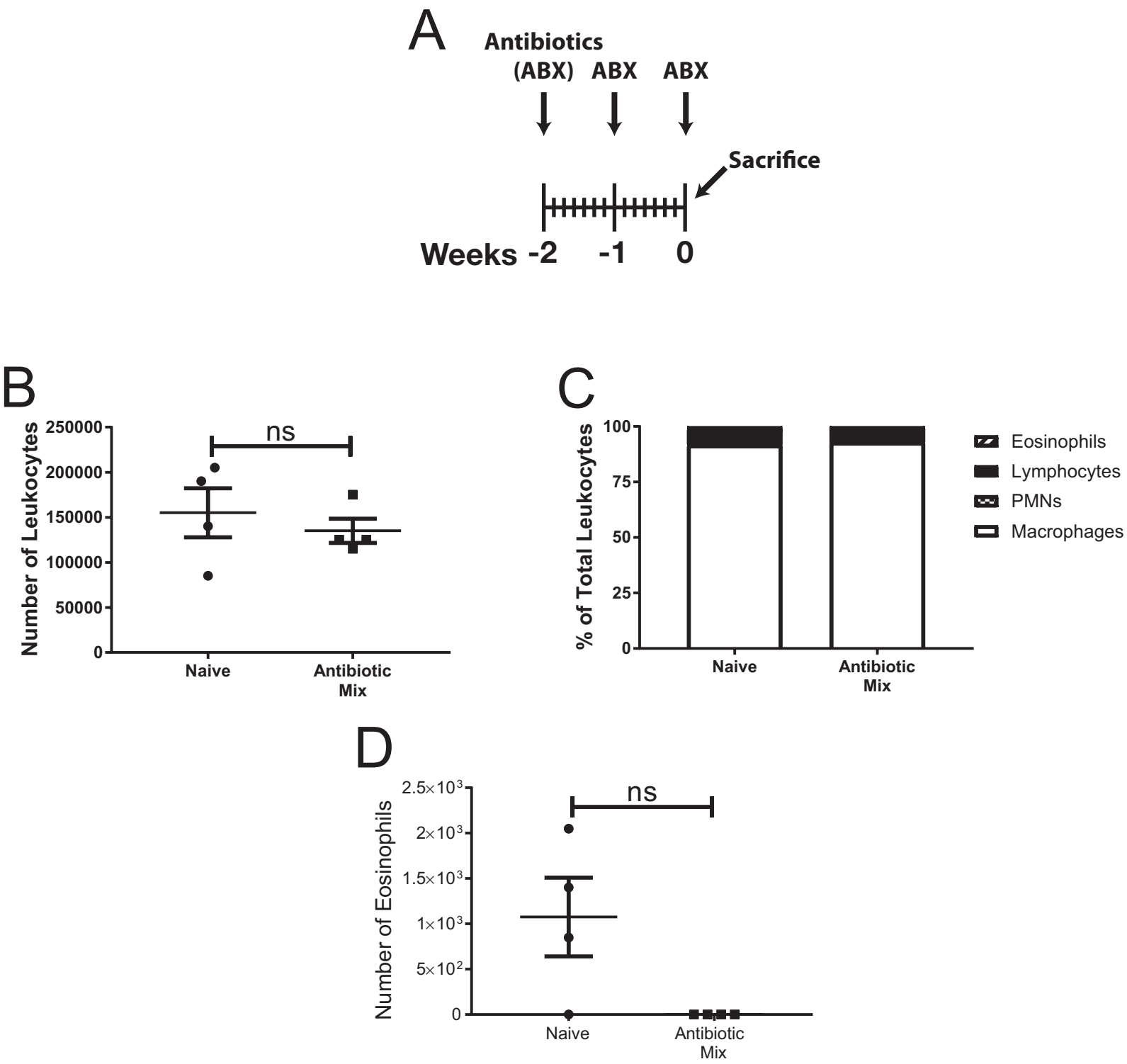
```
# Supplemental File S1, supplemental_file_1_qiime_analysis.sh: QIIME analysis
shell script used to process 16S rRNA reads.

#!/bin/bash

#PBS -l nodes=1:ppn=40
#PBS -N name_of_job
#PBS -e /path/to/working/directory/
#PBS -o /path/to/working/directory/
#PBS -V

# qiime_analysis.sh
#
#
# Created by Alexander J. Adami on 2016-02-09.
# Last revised: 2017-02-01
# This script is designed to work with QIIME 1.9.1 in a cluster environment
using torque
# Adapted in part from script provided by Kendra Maas of the University of
Connecticut Microbial Analysis, Resources, and Services (MARS) Facility and
from scripts referenced in Nelson et al 2015 PLOS ONE DOI:
10.1371/journal.pone.0094249 - PubMed
https://www.ncbi.nlm.nih.gov/pubmed/24722003


# Used to track runtime
START=`date +%s`

# Number of processors
# Modify to suit your environment
torque_procs=40

# References
# Modify to point to your reference files
reference_seqs=/path/to/greengenes/rep_set/97_otus.fasta
reference_tree=/path/to/greengenes/trees/97_otus.tree
reference_tax=/path/to/greengenes/taxonomy/97_otu_taxonomy.txt

# Script variables
map_file=/path/to/working/directory/map_file.txt
preprocess_directory=/path/to/working/directory/preprocess
work_directory=/path/to/working/directory
merged_reads_out=/path/to/working/directory/merged_reads_out
split_libs_out=/path/to/working/directory/split_libs_out
chimeric_seqs_out=/path/to/working/directory/usearch_chimeras_out
otus_out=/path/to/working/directory/otus_out

START=`date +%s`

# Move into main analysis directory
cd $work_directory

# Clear the preprocessing directory of any content and copy original files
into it for work
# Remove any intermediate directories created by future steps
# This is useful only if you had a previous failed run and wish to retry it
```

```
# It will clear all existing work from previous runs in the working
directory!
# Note that it expects source data in the orig_files directory
rm -rf *.fna
rm -rf usearch_*
rm -rf otus_out/
rm -rf $preprocess_directory/*
cp -rf $work_directory/orig_files/* $preprocess_directory/

# Move into the preprocessing directory
cd $preprocess_directory

# Loop through data folders and move files, renaming them by sample
# NOTE: This step will need to be modified for data downloaded from the SRA,
as the fastq files are no longer in Data/Intensities/BaseCalls
mkdir $preprocess_directory/fastq
for i in */Data/Intensities/BaseCalls/*.gz; do mv $i "fastq""/"${i%%-
*}"."`basename $i`; done

# Merge all paired ends by sample
mkdir qiime
multiple_join_paired_ends.py -i . -o $preprocess_directory/qiime/

# Remove any fastq files that failed to join
mkdir nonjoin
find qiime/ -name "fastqjoin.un*" -print -exec mv {} nonjoin/ \;
find qiime/ -size "0" -print -exec mv {} nonjoin/ \;

# Merge all demultiplexed libraries by sampleID using filenames, retaining
only QScores above 19
multiple_split_libraries_fastq.py -i $preprocess_directory/qiime/ -o
$preprocess_directory/qiime/ --demultiplexing_method sampleid_by_file --
include_input_dir_path -p $work_directory/qiime_parameters.txt

# Remove extra BaseSpace info from sequence names in seqs.fna
cp $preprocess_directory/qiime/seqs.fna
$preprocess_directory/qiime/demultiplexed_with_basespace_info_seqs.fna
sed 's/_S.*_L001.*join.fastq//g'
$preprocess_directory/qiime/demultiplexed_with_basespace_info_seqs.fna >
$preprocess_directory/qiime/basespace_cleaned_seqs.fna

# Count sequences: what did we get?
echo "Sequences after demultiplexing\n"
count_seqs.py -i $preprocess_directory/qiime/basespace_cleaned_seqs.fna

# Move into main analysis directory
cd /path/to/working/directory/

# Length filter sequences to between 245 and 260 bp
# This command
# 1) checks to see if the current line begins with a fasta sequence
identifier, which always starts with >.  In other words, says if the
beginning of the line (^) is NOT (!) followed by >
# 2) stops all activity and moves to the next line
# 3) if it passes 1 and 2, grabs the current line and copies it into the
variable seq
```

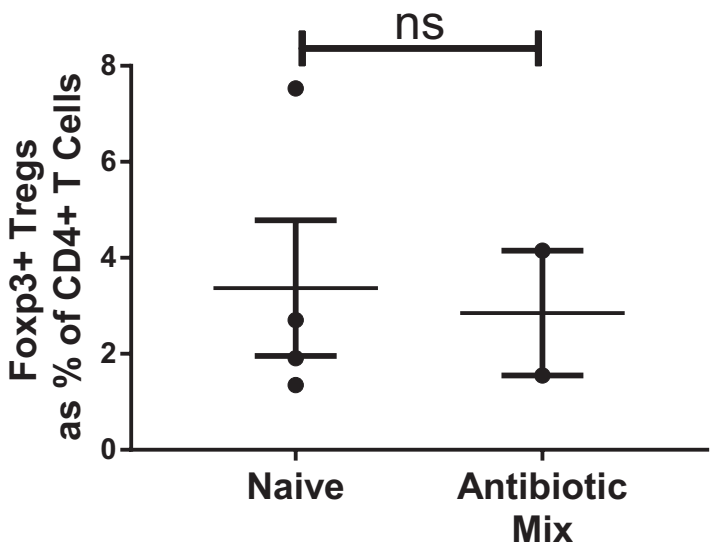
```
# 4) asks if the length of seq is between 245 and 260 bp inclusive, and
finally 5) prints a newline and then the contents of seq if the line met the
length conditions.
# Note that this will fail to generate the expected output if your input
FASTA file has sequences that are not represented on a single line!
awk '!/^>/ { next } { getline seq } (length(seq) >= 245 && length(seq) <=
260) { print $0 "\n" seq }'
$preprocess_directory/qiime/basespace_cleaned_seqs.fna >
length_filtered_seqs.fna

# Count sequences: what did we get?
echo "Sequences after length filtering\n"
count_seqs.py -i length_filtered_seqs.fna

# Check for chimeric sequences using usearch61
identify_chimeric_seqs.py -i length_filtered_seqs.fna -m usearch61 -o
$chimeric_seqs_out/ -r $reference_seqs
filter_fasta.py -f length_filtered_seqs.fna -o seqs_chimeras_filtered.fna -s
$chimeric_seqs_out/chimeras.txt -n

# Count sequences: what did we get?
echo "Sequences after chimera checking\n"
count_seqs.py -i seqs_chimeras_filtered.fna

# Pick OTUs using usearch61 in an open reference manner (i.e. reference based
and then de novo on unmatched sequences)
pick_open_reference_otus.py -o $otus_out/ -i seqs_chimeras_filtered.fna -r
$reference_seqs -m usearch61 -a -O $torque_procs

# Summarize OTU table information (i.e. how many sequences do we have for
each sample)
biom summarize-table -i $otus_out/otu_table_mc2_w_tax_no_pynast_failures.biom
-o $otus_out/otu_table_mc2_w_tax_no_pynast_failures_summary.txt

# Calculate alpha and beta diversity metrics (metrics specified in parameters
file), rarefying to 10000 reads
# Classification parameters should be specified based on your map file (e.g.
treatment, timepoint)
core_diversity_analyses.py -o $otus_out/diversity/ -i
$otus_out/otu_table_mc2_w_tax_no_pynast_failures.biom -m $map_file -t
$otus_out/rep_set.tre -p $work_directory/qiime_parameters.txt -e 10000 -c
"ClassificationParameter1,ClassificationParameter2" -a -O $torque_procs

# Move into main analysis directory
cd $otus_out/diversity/

# Make 2D plots, coloring by classification parameters
make_2d_plots.py -i $otus_out/diversity/2dplot_unweightedunifrac -m $map_file
-b 'ClassificationParameter1,ClassificationParameter2' --
ellipsoid_opacity=0.5 --ellipsoid_method=IQR -o
$otus_out/diversity/2dplots_out_unweighted_unifrac

echo ''
END=`date +%s`
RUNTIME=$(( END - START ))
echo "This job took $RUNTIME seconds to complete"
echo "Began "
```

```
date -d @$START
echo "\n"
echo "Ended "
date -d @$END
exit
```

```
# Supplemental File S2, supplemental_file_2_qiime_parameters.txt: QIIME
parameters file used to specify QIIME settings in the analysis file.

split_libraries_fastq:phred_quality_threshold 19
alpha_diversity:metrics PD_whole_tree,chao1,observed_otus,shannon,simpson
beta_diversity:metrics
bray_curtis,binary_ochiai,euclidean,unweighted_unifrac,weighted_unifrac
```



Supplemental Figure S1. Animals treated with antibiotics do not exhibit signs of allergic airways inflammation in the absence of HDM exposure. Experimental schematic is shown in (A). At sacrifice, broncho-alveolar lavage (BAL) was performed, total leukocytes counted (B), cellular differentials determined (C), and total eosinophils determined (D). No significant differences were noted for any measured parameters between naive animals not exposed to antibiotics or HDM and animals exposed to antibiotics but no HDM. Values in (B) and (D) represent mean ± the SEM while data in (C) represent the mean. n=4 per group. ns, no significant difference, $p > 0.05$.



Supplemental Figure S2. HLN Treg proportions are highly variable and similar between animals treated with antibiotics and animals not treated with antibiotics in the abscence of HDM exposure. Experimental schematic can be found in Supplemental Figure 1A. At sacrifice, lung draining (hilar) lymph nodes (HLNs) were processed and cells isolated and stained for flow cytometric assessment of CD4+Foxp3+ regulatory T cells (Tregs). No significant difference was noted between naive animals not exposed to antibiotics or HDM and animals exposed to antibiotics but no HDM, although conclusions are limited by sample number and variability. Values in represent mean ± the SEM. n=4 for naive and n=2 for Antibiotic Mix. ns, no significant difference, $p > 0.05$.