Aim 1: Isolate effect of urbanization on evolutionary forces in human microbiome

* Question: How has urbanization influenced evolutionary forces in the human microbiome
* Approach: infer demographic models of microbial species present in human oral and gut microbiome.
* Hypothesis: Urbanization has decreased microbial diversity in the gut microbiome and increased microbial diversity in the oral microbiome (based on preliminary results for gut and previous work for oral).

Aim 2: Identify host-symbiote evolutionary interactions between humans and gut microbiota

* Question: Does biologically interesting selection occur with respect to the human microbiome, and if so, *where*, i.e., can we apply a systematic scan to identify these sites under biologically interesting selection?
* Approach: Systematic scan to find biologically interesting selection and evolutionary interactions between human and microbial species.
* Hypothesis: I expect to find evidence of negative selection in the gut microbiome, as evidenced by the decrease in microbial diversity and the increased model fit when accounting for negative selection in inferring demographic models.

Aim 1: Differentiate within-host and between-host evolutionary effects on selection in human microbiome.

* Question: Do within-host and between-host evolutionary effects on selection in the human microbiome target different genes.
* Approach: Perhaps some form of 3D SFS, where we examine prevalence x frequency x host. Maybe some form of DFE analysis with and without the pseudoisolate / quasiphasing?
* Hypothesis: I would expect that within-host and between-host evolutionary effects are subject to different bottlenecks, and thus should target different aspects of the evolutionary landscape.

Aim 2: Quantify effect of urbanization on DFE of gut microbiota

* Question: Does urbanization produce a detectable shift in the DFE of gut microbiota?
* Approach: compare DFE of gut microbial species in urbanized vs. ruralized populations
* Hypothesis: I would expect that the DFE of gut microbial species in urbanized populations demonstrate a notable increase in negative selection, aligned with the observed demographic contraction.

**Notes from 2021\_08\_19:**

Aim 1: Infer demographic model of microbial species present in human oral and gut microbiome.

Aim 2: Isolate effect of urbanization on evolutionary forces in human microbiome.

* Explain hypothesis that urbanization effects things
* 1st aim is harder to argue
  + Insufficient as standalone aim – perhaps include as part of second aim, e.g., we need to form demographic models to test the hypothesis for urbanization
  + Perhaps merge aim 1 and 2 into one stronger aim
  + Make sure aims are *hypothesis-driven*
    - Easier to read for reviewers – clearly defines research question and its importance / feasibility
    - **Clearly define research question / goal**
  + Aim 2 question:
  + When constructing hypothesis, be “cartoonish” about expectations to clearly illustrate point
* If this ends up the point of the pre-print, than probably don’t use this for the NSF

Aim 1: Design statistically valid bioinformatics pipeline to apply traditional population genetics techniques to prokaryotic species.

Aim 2: Differentiate within-host and between-host evolutionary effects on selection in human microbiome.

* Idea is that currently we are phasing metagenomic data (using phased data as a stand-in for pseudo-isolates to build an SFS)
  + We are throwing away a lot of within-host data since we only consider the highest prevalence strain for each species per host
  + We could have a “2D” or “3D” analysis where we look at prevalence by frequency per host, e.g., one axis is prevalence, one axis is frequency, and one axis is the host
    - Novel way to represent data in an SFS-like approach
  + This one would require a lot of cartoons to explain / envision
* It would be important to highlight why we should care about the differences between within-host and between-host evolution.
* If we propose something more technical e.g., the “3D” sfs cube phased data thing, then we need to justify why we would want to use such a statistic or method
  + Also we need to figure out if this tells us anything? Can we even infer anything?
* Rephrase *and* rework aim 1 – don’t use term “pipeline”
  + Change to tool or software or just straight up change aim 1 to be less exploratory
  + Reword it as a question instead of an open-ended adventure

Aim 1: Apply systematic scan for biologically interesting selection in human gut microbiome.

Aim 2: Identify host-symbiote evolutionary interactions between humans and gut microbiota.

* Aim 2 could be exciting
  + Maybe another case where Aim 1 could be folded into Aim 2 to make a stronger singular aim, e.g.,: Systematic scan to find interesting selection and interactions between humans and gut microbiota.

Broadly, I had a pattern of aim 1: apply method, aim 2: interpret results. The problem with this is that aim 2 is dependent on aim 1, so if the method doesn’t work than rest in pepperonis. It’d be better to package method and result-based aims into one stronger aim that is *hypothesis-driven*.

Instead of aims, approach as pairs of questions.

Other ideas:

1. Could have aim about inferring demography using LD and an Approximate Bayesian Computing (ABC) framework.
2. Infer demography + infer selection
   1. Don’t need demographic model to infer selection (not sure)
   2. Do we have a contraction in population size?
   3. Do we observe selection present in urbanized populations?
      1. Did changes in diet create selective pressure on microbial species in gut microbiome?

Could also write about Kirk stuff (recessive deleterious mutations)