**Slide 1:** The Horizontal Gene Transfer Project – a Community Science Project (CSP)

**Slide 2**: The work you have done so far, working through the previous slide decks and exercises gives you tools to investigate genomes of bacteria. You now know what data are available, where to find these data, and some basic ways to analyze the data.

In order to give you a specific project to join, we have developed what we call the Horizontal Gene Transfer Project or the Community Science Project (CSP). Much of what you’ve done so far is to replicate the findings of a paper we did with some of the undergraduates in our initial Microbial Genomics class – finding that orthologs of the gene encoding the Putative Replication Initiation Protein (PRIP) from Chlamydia phage was found in all sequenced isolates of *Chlamydophila pneumoniae* (at least as of 2014 when this article was published).

Anne Rosenwald, Bradley Murray, Theodore Toth, Ramana Madupu, Alexandra Kyrillos, and Gaurav Arora. Evidence for horizontal gene transfer between *Chalmydophila* *pneumoniae* and Chlamydia phage (2014) Bacteriophage doi: 10.4161/21597073.2014.965076

**Slide 3:** We’ve extended this work in several ways. The next few slides are from a second paper we published on horizontal gene transfer between phages and bacteria, this time examining *Helicobacter pylori* genomes and a *Helicobacter* phage called phiHP33.

Brad Murray was an undergraduate in our initial Microbial Genomics class and was interested in *Helicobacter*. For his senior thesis project, he wanted to look at HGT between phiHP33 and *Helicobacter*. He began the work, then Alex picked it up and did a much more extensive job.

Alexandra Kyrillos, Bradley Murray, Gaurav Arora, and Anne Rosenwald. phiHP33 Orthologous Genes in *Helicobacter pylori* (2016) *Helicobacter* doi: 10.1111/hel.12282

**Slide 4:** *Helicobacter pylori* is a common commensal of humans. Its presence is associated with ulcers and gastric cancer. In its absence in childhood its associated with the development of asthma. It is kind of a double-edged sword - the idea that the same organism is protective in one case but then is causative in another. Virulence for ulcers and cancer is associated with two genes called *CagA* and *VacA*.

**Slide 5:** The other partner in the horizontal gene transfer is the phage, phiHP33, related to *E. coli* phage . The genome structure, with 27 genes is shown in this slide.

**Slide 6:** Alex looked at the 335 *H.pylori* complete genomes that were then available in GenBank and IMG. She found that a third of them do contain sequences related to phiHP33, the other two-thirds don't.

**Slide 7:** The other thing that she found which was really interesting was that if she looked at the correlation between which strains have these phage orthologus genes or **POGs,** and the ones that had *CagA* and *VacA*, the two virulence factors, there is a really high statistical correlation. So what we don't know at this point is whether the POGs are virulence factors themselves or whether they mediate the virulence somehow.

**Slide 8:** We believe that this is Horizontal Gene Transfer because Alex took a known phylogeny of *Helicobacter* strains and then mapped which phiHP33 gene orthologs were present showed the phage genes that are present do not follow the phylogeny, indicating that this must be a form of horizontal gene transfer.

**Slide 9:** More broadly, why is this interesting? The SEA-phages program has found hundreds of bacteriophages that infect one single bacterium, *Mycobacterium smegmatis*. What that suggests is that the virome is even more extensive and diverse than the bacteria on which these bacteriophages live. There is lot of structural diversity among the phages from EM studies, a lot of genetic diversity from the annotation of the phage genomes. So what we would like to understand is how the virome influences evolution of bacteria and maybe vice versa.

**Slide 10:** This figure comes from a paper published in Nature a few years ago. It seems that many bacteria have genes that have come from somewhere else, not necessarily just from bacteriophages but from other bacteria too. The pink here shows the extent to which there are genes from other place; almost 13% in our favorite bacterium*, E. coli*. This phenomenon is therefore important for bacterial diversity and evolution.  We would like to know more specifically the extent to which this bacteriophage - bacteria interaction is driving evolution.

**Slide 11:** There are many examples of phage orthologus genes (POGs) found in bacteria. In Part II, we discussed the databases where bacterial genomic data are stored. In Parts III, IV, and V, we talked about some of the tools you can use for comparative genomic analysis and by doing the associated exercises, you got some practice in their use.

**Slide 12:** You too can contribute instances of phage gene orthologs that you have found in bacteria.

We have a workflow that really just involves BLAST. The workflow document is included on the Genome Solver QUBES Hub Project site. There are also two Excel templates for the data. You’ll upload your data using a Google Forms link that is found in the workflow.

**Slide 15:** The developers of Genome Solver are listed on this slide. We would also like to thank the National Science Foundation for support (2011-2018).