SUPPLEMENTAL MATERIALS AND METHODS

Computational Requirements

The free software, Mauve (version 2.3.0 or more recent) can be downloaded by instructors and/or students (http://asap.ahabs.wisc.edu/mauve/download.php), and should be installed on either a PC (windows XP or more recent) or Macintosh computer (Operating System 10.4 or more recent). For students using personal computers, the minimum requirements are 386 Mb RAM, and computers may require a Java update, as Mauve runs on this platform.

Genome Alignments

Genome sequence files with annotations were obtained from the ASAP database for genomes *E. coli* O157:H7 EDL933, Sakai, and 4042 or *Y. pestis* CO92, KIM, 91001, Antiqua, and Nepal516 (Glasner *et al*. 2006). Genome alignments were conducted using the progressive alignment option of Mauve 2.2.0 (Darling *et al*. 2004) with the default settings, and can be downloaded (https://asap.ahabs.wisc.edu/~baumler/) for *E. coli* O157 (3 O157 alignment.zip) and *Y. pestis* (Yersinia pestis alignment 5 genome.zip). Figures were generated using the Mauve alignment viewer, which illustrates locally collinear blocks (LCBs) as regions without rearrangements in the homologous backbone sequence (Supplemental Figure 1 and Supplemental Figure 5). LCBs below a genome's center line represent the reverse complement orientation relative to the reference genome. Sequence similarity plots are displayed in the LCBs, and the height of the sequence identity plot reflects the degree of sequence similarity for the region of the respective alignment (Darling *et al.* 2004).

Analysis of genomic islands

The Mauve alignment visualization can be switched from the default “LCB” coloring scheme to the “Backbone” color scheme, which changes the colors of the genomic regions to reflect which subsets of genomes contain a particular aligned segment. In this exercise students use Mauve to identify genomic regions that are unique to a genome, and use the backbone view option to identify regions conserved in two of three genomes in the *E. coli* O157:H7 alignment (Supplemental Figure 2) and regions that are conserved in four of five genomes in the *Y. pestis* alignment (Supplemental Figure 6). By zooming in on these regions of the alignment, the student sees boxes representing individual ORFs that are clickable with links to additional information for detailed annotations in the ASAP database (Supplemental Figure 3). These annotations provide information about the function of the protein encoded by the ORF, such as its role in virulence. Students are encouraged to formulate new hypotheses regarding the presence and absence of genomic regions and the potential implications for explaining phenotypic differences between pathogenic bacterial strains.

Conservation of known or putative virulence factors

For *E. coli* O157:H7 EDL933 and *Y. pestis* CO92 genomes, a thorough survey of the scientific literature was performed and biological annotations were added to the ASAP database for genes believed to produce known or putative virulence factors based on conclusions drawn from experimental evidence, sequence homology, or the predicted physiological function. Overall, 394 genes were identified as known or potential virulence factors for *E. coli* EDL933 (supplemental data 1) and 148 genes for *Y. pestis* CO92 (supplemental data 2), thus representing a vast assortment of genes for student inquiry. For each microorganism, virulence factors were subcategorized based on the various functions that enable the organisms to cause disease in humans. A list of these categories with a brief description of each are provided, along with pie charts reflecting the proportion of the total number of potential virulence factors represented by that category, for *E. coli* O157:H7 (supplemental information 1 and supplemental instructional slides 1) and *Y. pestis* (supplemental information 2 and supplemental instructional slides 2).

Identification of single nucleotide polymorphisms in ORFs from Y. pestis

Once the ORFs for *glpD*, *napA*, and *araC* are identified for each of the five genomes represented in the Mauve genome alignment, clicking on a gene of interest and selecting the “view CDS” in ASAP opens a feature page for that coding sequence (CDS) containing a list of biological annotations in a web-browser. On this page, students can survey polymorphic sites contained within the ORF. Each polymorphic nucleotide position has a corresponding unique identifier. By copying and pasting the identifier into a SNP analysis tool “Snippy” (http://asap.ahabs.wisc.edu/~cabot/aep/snippy.php), students are able to analyze the consequence of each polymorphism on the protein encoded by the ORF from each *Y. pestis* genome (Supplemental Figure 4). The SNP data are displayed along with the corresponding alterations in the amino acid sequence.

Students are asked to identify strains that have permutations leading to production of a truncated, and hence non-functional, protein.