Due to the variety of information sources we can now interrogate, we are able to start appreciating the true complexity of cells and the significance of the variables involved. In the study of gene expression, one now considers temporal expression in conjunction with what particular alleles are transcribed in what tissue and which splice variants are predominant. The analysis of translation state also became available and can provide essential pertinent information. Analyzing whole genome sequences in conjunction with sequence based expression profiles on large segregating populations should also allow for better QTL characterizations. The data permits the enhancement of usability of newly assembled genome by better defining gene structure annotation and thus allowing for better comparison across closely related species. Numerous protocols are available and examples of applications will be described.

Infectious disease researchers have to deal with diverse types of data to develop hypotheses about candidate macromolecules that can be used to design countermeasures (diagnostics, vaccines and therapeutics). Even considering only molecular data, it is a challenge to access all the public information available to them and implement workflows that support their analysis needs. I will use the example of Brucella spp. to illustrate how public, open, freely available resources can be designed, developed, and implemented in support of such infectious disease research and development goals. There are now 40 genomes of Brucella (Alphaproteobacteria; Rhizobiales) sequenced, sampling all known species and biovars of this facultative intracellular pathogen. A phylogenomic analysis of these genomes united all Brucella spp. to illustrate how public, open, freely available resources can be designed, developed, and implemented in support of such infectious disease research and development goals. There are now 40 genomes of Brucella (Alphaproteobacteria; Rhizobiales) sequenced, sampling all known species and biovars of this facultative intracellular pathogen. A phylogenomic analysis of these genomes united all Brucella spp. when they were compared with outgroups including Ochrobactrum, Bartonella, Mesorhizobium and Agrobacterium. Although the Brucella genomes are united, there is some interesting diversity. A well-studied group of Brucella species are united in a clade separated by a long branch. This large clade has little phylogenetic depth, but species within this group are known to have specific host preferences. Sub-branching patterns within this group reflect these host preferences and specific protein families unique to, and absent from, these groups were identified. Two novel strains, Brucella inapinata BO1T and BO2, were recently identified and isolated from human patients. Genomes from these strains, as well as 2 new isolates isolated from Australian rodents (Brucella spp. NF2653 and 83/13), are quite unique and separated from the rest of the Brucella. Although they share many similarities with the other Brucella species, they are missing many large areas of their genomes that can be seen in other Brucella. Analyses of these areas using new bioinformatic tools and data resources have shown that some of these missing areas include previously identified genomic islands and virulence factors, and there are novel findings as well that impact biochemical pathways and unique changes in the synthesis of lipopolysaccharide.