Center for Inference and Dynamics of Infectious Diseases Seminar

## Uncovering the properties and limitations of models of sequence evolution

Dr. Stephanie J. Spielman, PhD
Institute for Genomics and Evolutionary Medicine at Temple University

Thursday July 21, 2016, 10:30–11:30AM Pelton Auditorium, Weintraub Building, B1-065 Fred Hutch

Refreshments will be served; Registration is not necessary to attend this event



ABSTRACT: Natural selection leaves signatures of its activity in DNA. As sequencing increasingly assumes a central role in modern biological research, we have an unprecedented opportunity to elucidate the myriad ways in which evolution shapes the diversity of life's genomes. Most commonly, we study these "fingerprints" of natural selection using statistical models of sequence evolution. In the context of of protein-coding sequence evolution, two popular complementary models have emerged: dN/dS-based models, which measure the relative rate of nonsynonymous to synonymous substitutions, and mutation—selection models, which measure site-specific amino acid propensities (or "fitness") using population genetics principles. Importantly, these models have been constructed and studied independently from one another. As a consequence, it has been entirely unknown whether different models reveal similar or incompatible information about the strength and direction of natural selection, in turn hindering our ability to draw robust conclusions about evolutionary pressures. In this talk, I will discuss how we can bridge this gap by deriving a precise relationship between dN/dS-based and mutation—selection models. I use this relationship to identity previously unappreciated limitations and behaviors of these models. For example, this work reveals that dN/dS inferences in the popular software PAML (codeml) are strongly biased, and further that standard metrics of model selection (AIC and BIC) may be positively misleading. I offer specific recommendations for how to most reliably apply models of protein-coding sequence evolution and interpret their results.

## Seminar Sponsors:



