

Mathematical & Computational Biology Seminar

Organizer: Valerie Hower

Wednesday, 2:00–3:00pm, 939 Evans

Sept. 16 **Alexander Schoenhuth**, UC Berkeley

A statistical framework for improved prediction of true indels in pair HMM based alignments

Since both evolutionary origins and consequences of indels have not yet been fully understood, indels as inferred by alignment algorithms usually need further evaluation. It would be desirable to be able to more reliably distinguish between, highly likely, true evolutionary indels and “gap noise” in large-scale alignment studies. We have developed a pair HMM based framework to assess the significance of indels in alignments with affine gap penalty scoring schemes. Tables containing the required probabilities can be efficiently computed. We validated the predictive power of our framework by benchmarking inferred indel positions against true indel positions in structural alignments as provided by BAliBASE. We found that indel significance, as assessed by our framework, is a substantially more powerful indicator of true indel positions than naive approaches. Although unintended, our framework also facilitates improved prediction of homology since alignments containing significant indel patterns also significantly more likely referred to homologous protein pairs.