PYGR: A Python Graph Database Framework for Bioinformatics

PYGR makes it easy to do comparative genomics analyses of huge multigenome sequence alignments, using a disk-based query algorithm that is 5- 500x faster (100-300 microseconds/query) than other indexing methods such as MySQL, binning, or R-Trees. We will demonstrate how to use PYGR to write simple but powerful graph queries; store and query multiple sequence alignments in a form that can scale to alignments of whole genomes; provide a simple but powerful interface to large sequence databases and automation of tools like BLAST; store and query multi-genome alignments such as the UCSC BLASTZ alignments of many animal genomes.