Comparative genomics has moved from the analysis of genome pairs to the analysis of several genomes at once. In the era of cheap genomes, whole clades of genomes will be sequenced not by large sequencing centers but by specialist researchers themselves. Researches will not be able to rely on the big genome servers like ENSEMBL or UCSC genome browser to annotate and serve their data. Our pipeline aims to fill this gap.

Our pipeline annotates a clade of genomes using a reference genome. The pipeline combines standard bioinformatic tools to predict genes, assign orthology and compute evolutionary rates. The pipeline detects multiple transcripts per gene and evaluates predictions based on exon/intron boundary conservation. We have applied our pipeline successfully to twelve Drosophila genomes. The results can be accessed on the web (http://wwwfgu.anat.ox.ac.uk). The site allows the user to browse between gene predictions and their orthologs in other species. Orthologous groups can be extracted, for example, by species composition and viewed as a multiple alignment. Each group is associated with a phylogenetic tree and synonymous and non-synonymous substitution rates.