PathogenDx

Simplified pipeline diagram



Input data: Metadata Raw reads Optional references

nextflow

PDF

Fast initial ID bbmap sendsketch Trim adapters Filter by kmer fastp abundance Search for assemblies khmer trim low abund **Entrez esearch** Assemble genomes spades Make signatures of **Download assemblies** user references datasets download Filter assemblies sourmash sketch custom script annotation Create signatures Make signatures of assemblies of reads Annotate genomes sourmash sketch sourmash sketch bakta Compare signatures **Extract BUSCO genes** sourmash compare Read2Tree phylogeny (Eukaryotes busco Identify shared orthologs genome Assign references pirate Make Read2Tree custom script reference database custom script Extract core genes phylogeny (Prokaryotes custom script Align reads Infer phylogeny bwa mem read2tree Variant calling Make gene multiple **Identify duplicates** sequence alignments picard MarkDuplicates mafft **Call variants** graphtyper genotype Combines VCF files Infer Phylogeny graphtyper igtree vcf concatenate **Filter variants** vcflib vcffilter Infer phylogeny **IO-TREE** Find and download references Select reference for variant calling **Assess read quality** Make read Genome assembly and annotation fastqc quality report Variant calling multiac Assess assembly quality Read2Tree phylogeny quast Create reports Core genome phylogeny custom script Quality control and reporting