

## Next Generation Sequencing

— haphpipe\_assemble\_01  
— haphpipe\_assemble\_02

## READS

manipulate reads,  
perform quality  
control

*sample\_reads*

subsample reads with SEQTK

FASTQ > FASTQ

*trim\_reads*

trim reads with Trimmomatic

FASTQ > FASTQ

*ec\_reads*

error correction with SPAdes

FASTQ > FASTQ

*join\_reads*

join reads with FLASH

FASTQ > FASTQ

## ASSEMBLE

assemble consensus  
sequences

*assemble\_denovo*

de novo assembly with  
SPAdes  
FASTQ > FASTA

*assemble\_amplicons*

assemble contigs with  
MUMMER 3+  
FASTA > FASTA

*assemble\_scaffold*

scaffold contigs against  
reference with MUMMER 3+  
FASTA > FASTA

*align\_reads*

align reads to reference with  
Bowtie2 and Picard  
FASTQ > BAM

*call\_variants*

variant calling with GATK  
BAM > VCF

*vcf\_to\_consensus*

generate consensus sequence  
from VCF file  
VCF > FASTA

*refine\_assembly*

three-step assembly  
refinement  
FASTQ > FASTA

*finalize\_assembly*

finalize consensus sequence  
FASTQ > FASTA

## ANNOTATE

consensus sequence  
annotation

*pairwise\_align*

apply coordinate system to  
final sequences  
FASTA > JSON

*extract\_pairwise*

extract sequence regions  
JSON > FASTA

*annotate\_from\_ref*

annotate consensus  
sequence  
JSON > GTF

*summary\_stats*

report assembly and  
haplotype summary statistics  
output files > TXT, TSV

## PHYLO

phylogenomics

*multiple\_align*

multiple sequence alignment  
with MAFFT  
FASTA > FASTA

*build\_tree*

phylogeny reconstruction  
with RAxML  
FASTA > TRE

## HAPLOTYPE

haplotype  
reconstruction  
and assembly

*predict\_haplo*

identify haplotypes with  
PredictHaplo  
FASTQ > FASTA

*ph\_parser*

format PredictHaplo output  
file  
FASTA > FASTA