

Tool	What is it used for	Documentation	Example of where to install it	Notes regarding AWS instance type	Exercise with info on how to install it and examples of how to use the tool
bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. We use it specifically to identify variants (SNPs and INDELS)	http://www.htslib.org/doc/bcftools.html	AWS with AMI: Ubuntu Server 16.04 LTS (HVM), SSD Volume Type		Ex. 5
BWA	Mapping reads to reference genomes	http://bio-bwa.sourceforge.net/bwa.shtml	AWS with AMI: Ubuntu Server 16.04 LTS (HVM), SSD Volume Type	BWA does not require a lot of RAM, but some of the downstream tools used for calling SNPs do, so you might as well select, e.g., a t2.large	Ex. 5
Cutadapt	Trimming of reads	http://cutadapt.readthedocs.io/en/stable/index.html	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type	t2.micro (free)	Ex. 3
FastQC	Assessing the quality of raw reads	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	Local computer (this workshop), but can also be installed on Linux machine	NA	Ex. 3
Miniasm	Assembling long (PacBio or ONT) reads	https://github.com/lh3/miniasm	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type		Ex. 4-extra
Minimap2	Sequence alignment, typically long reads (PacBio or ONT)	https://github.com/lh3/minimap2	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type		Ex. 4-extra
ndTree	Creating SNP-based phylogenetic trees	https://bitbucket.org/mettevoidbylarsen/ndtree	AWS with AMI: Ubuntu Server 16.04 LTS (HVM), SSD Volume Type	t2.micro (free)	Ex. 7
PICARD	Picard is a set of Java command line tools for manipulating high-throughput sequence data and formats such as SAM/BAM. We specifically use it to remove duplicates (reads that map with identical 5' coordinates and orientations)	https://broadinstitute.github.io/picard/command-line-overview.html#Overview	AWS with AMI: Ubuntu Server 16.04 LTS (HVM), SSD Volume Type	t2.large	Ex.5
Prinseq	Trimming of reads	http://prinseq.sourceforge.net/ (for online version. Can also be downloaded to run locally from this site)	NA	NA	Supplementary material to Ex. 3 (https://www.goseqit.com/wp-content/uploads/2018/09/PRINSEQ_How_to.pdf)
prinseq_lite	We use this tool for generating summary statistics of sequence and quality data	http://prinseq.sourceforge.net/manual.html	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type	t2.micro (free)	Ex. 3
Quast	Assessing the quality of draft genomes	http://quast.bioinf.spbau.ru/manual.html#sec2.1	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type		Ex. 4
Racon	Polishing minimap/miniasm assemblies made on the basis of long (PacBio or ONT) reads	https://github.com/isovic/racon	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type		Ex. 4-extra
Samtools	A library and software package for parsing and manipulating alignments in the SAM/BAM format	http://www.htslib.org/doc/samtools.html	AWS with AMI: Ubuntu Server 16.04 LTS (HVM), SSD Volume Type		Ex. 5

Tool	What is it used for	Documentation	Example of where to install it	Notes regarding AWS instance type	Exercise with info on how to install it and examples of how to use the tool
SPAdes	Assembling short (Illumina) reads to draft genomes. Can also perform hybrid assembly using both short (Illumina) and long (PacBio or ONT) reads	http://cab.spbu.ru/files/release3.12.0/manual.html	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type	SPAdes needs <i>a lot</i> of RAM. As a rule of thumb, 32 GiB RAM should be enough if you have around 50X coverage, otherwise select an instance type with more RAM	Ex. 4 and Ex. 4-extra
Trimmomatic	Trimming of reads	https://www.goseqit.com/wp-content/uploads/2018/09/Trimmomatic_tutorial.pdf	AWS with AMI: Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type	t2.micro (free)	Supplementary material to Ex. 3 (https://www.goseqit.com/wp-content/uploads/2018/09/Trimmomatic_tutorial.pdf)