

Table 1: Comparaison des workflows suggérés comme alternatives à Bactopia.

Outil	Données	Nettoyage	Assemblage	Alignement	Variant calling	Annotations/infos données	Autres
Bactopia	<ul style="list-style-type: none"> - Générables (art) - Téléchargement (ena-dl : ENA, NCBI) - Qualité : fastq-scan (stats check si contient assez de séquences) - FastQC - Estime taille génome : Mash 	<ul style="list-style-type: none"> - BBTools : adapteurs contaminants - Lighter : erreurs seq - McCortex : supp 31-mer singletons 	<ul style="list-style-type: none"> - Shovill : MEGAHIT, SKESA, SPAdes or Velvet - Unicycler (reads longs) - Quality : assembly-scan, CheckM, 	<ul style="list-style-type: none"> Bedtools (coverage), BWA, Samtools (convert) Sourmash/ : comparaison séquences contre GenBank, RefSeq et PLSDB (utilisés pour sélection auto de génomes de réf pour le calling.) 	<ul style="list-style-type: none"> Snippy (caller) 	<ul style="list-style-type: none"> Prokka, Bakta AMRFinderPlus/ ARIBA/ BLAST (résistance, type) : run fastq sur BD mlst puis aligne locus sur génome assemblé, vcf-annotator 	<ul style="list-style-type: none"> BLAST : aligne sur génome assemblé (gène, primer, protéine à fournir) - Illumina, Hybrid (Nanopore, Pacbio)
AQUAMIS	<ul style="list-style-type: none"> - fournir fastq 	<ul style="list-style-type: none"> - fastp (QC + trim) 	<ul style="list-style-type: none"> - Shovill - QUAST 	<ul style="list-style-type: none"> - Mash : recherche ref base = contigs + téléchargement NCBI 	<ul style="list-style-type: none"> / 	<ul style="list-style-type: none"> - Kraken2/ bracken (classification sur base MiniKraken ou fournir), - ConFindr : contaminations (mlst) 	<ul style="list-style-type: none"> Rapport html : info assemblage/ trim, coverage, map sur contigs, BD perso pour Campylobacter Illumina or IonTorrent
ASA3P	<ul style="list-style-type: none"> - Qualité : FastQC, fastQ screen (contamination) 	<ul style="list-style-type: none"> Trimmomatic, Fitlong (filtre reads longs par qualité) 	<ul style="list-style-type: none"> SPAdes (Illumina), HGAP (PacBio), Unicycler (ONT) 	<ul style="list-style-type: none"> Bowtie2 (Illumina), BWA, pbalgn (PacBio), minimap2 (ONT), MeDuSa : réarrangement et ordonner les contigs 	<ul style="list-style-type: none"> Samtools 	<ul style="list-style-type: none"> - Prokka (annot génomes assemblés), - SNPeff - BD RefSeq : pseudogénomes - kraken (recherche de profils kmer) / mumer (calcul d'une identité nucléotidique moyenne) / blast (basée sur 16S + MLST + facteurs virulence) : classification, - RGI : antibio res, 	<ul style="list-style-type: none"> - fait pour les isolats bactériens étroitement apparentés, Comparaison génomes: Roary (détecte gènes isolés entre génomes)/ FastTreeMP (arbre max vrais) - Illumina, Hybrid (Nanopore, Pacbio)
MicroPIPE	/	<ul style="list-style-type: none"> Porechop : rm adapters pour ONT ! unsupported 2018, Fitlong/Japsa : filtre en fonction du type, de la longueur et de la qualité 	<ul style="list-style-type: none"> Flye QUAST 	/	/	/	<ul style="list-style-type: none"> - commence à l'appel des bases : guppy - qcat/guppy : demultiplexing - pycoQC : qualité séquençage - Polishing : Racon/Medaka (long read) NextPolish(short read) corrige erreurs de seq / améliore précision, Circulator : circularise/ visualise génomes - Illumina
Nullarbor	<ul style="list-style-type: none"> Fournir fastq paired-end 	<ul style="list-style-type: none"> Trimmomatic 	<ul style="list-style-type: none"> SKESA, SPAdes, Megahit, shovill, Velvet 	<ul style="list-style-type: none"> Snippy-core (contre ref) 	<ul style="list-style-type: none"> Snippy@ (caller) Snippy-core (dans core genome) 	<ul style="list-style-type: none"> Kraken, Kraken2, Centrifuge (classification métagénomique) Prokka MLST ABRicate (résistance antimicrobienne /virulence) snp-dists (matrice distance de snp) 	<ul style="list-style-type: none"> - Roary - IQTree, FastTree (arbres) - Rapports stats - Illumina
ProkEvo	<ul style="list-style-type: none"> - parallel fastq-dump (download) - Qualité : FastQC 	<ul style="list-style-type: none"> Trimmomatic 	<ul style="list-style-type: none"> SPAdes, QUAST, custom script (supprimer assemblages avec 0 ou plus de 300 contigs, ou avec N50 inf 25 000) 	/	/	<ul style="list-style-type: none"> - MLST (type) - ABRicate (résistance antimicrobienne /virulence) - Sistr (classification Salmonella) - plasmidfinder (id plasmique dans assemblages) - Prokka 	<ul style="list-style-type: none"> - Roary (identifier le pan-génome et générer des alignements core-génome à partir des annot Prokka gff3) - fastbaps (regrouper hiérarchiquement les séquences génomiques de l'alignement de séquences multiples) Illumina

PHBG	<ul style="list-style-type: none"> - fastq-scan (stats récap (longueurs de read, qualités par lecture, qualités par base)) - FastQC 	<ul style="list-style-type: none"> - BBDuk - fastp - Trimmomatic 	<ul style="list-style-type: none"> - Shovill - BUSCO (évalue assemblage, basé sur concept d'orthologues à copie unique : conservés parmi les espèces apparentées) - QUAST 	/	/	<ul style="list-style-type: none"> - ABRicate - AMRFinderPlus (identifie gènes de résistance) - Bakta - GAMMA.py (trouve correspondances de gènes par seq protéique/traduit-annote la correspondance en fournissant le type (mutant, troncature) et une description traduite) - plasmidfinder - Prokka - resfinder (résistance) - MUMmer, - Kraken2 	<ul style="list-style-type: none"> - iqtree, kSNP3 (identifier et annoter SNP sans alignement + arbre, mashtree (arbre sur distances mash)) - gambit (id taxo) - PIRATE (identifie/classe familles de gènes orthologues dans pangénomes bactériens) - snp-dists (Convert a FASTA alignment to SNP distance matrix) - annotations pour certains pathogènes MIDAS : pipeline, exploite 30 000 génomes de référence pour estimer l'abondance des espèces bactériennes et la variation génomique. Aligne reads sur BD avec HS-BLASTN, aligne avec bowtie2, profilage pan-génome
rMAP	<ul style="list-style-type: none"> - fastq-dump - FastQC - MultiQC 	<ul style="list-style-type: none"> - Trimmomatic 	<ul style="list-style-type: none"> - Shovill, megahit, Unicycler Stats : QUAST, assembly-stats 	<ul style="list-style-type: none"> - BWA - MAFFT - samclip (filtre SAM) - Samtools : trier, coverage 	<ul style="list-style-type: none"> - Freebayes, - Snippy - vt - ISMappeur (trouve insertions par rapport à une réf) 	<ul style="list-style-type: none"> - ABRicate - SnpEff - AMRFinderPlus - Prokka - MLST (détection de clones) 	<ul style="list-style-type: none"> - dev pour des pathogènes, - Prodigal (prédiction des gènes codants) - Kleborate (outil criblage des assemblages pour Klebsiella pneumoniae) - Fasttree/Iqtree (arbre) - BMGE (utilise entropie pour id régions avec le plus d'information phylogénétique) - Roary - Illumina paired-end data
TORMES	/	<ul style="list-style-type: none"> Trimmomatic, Prinseq/Sickle (filtre qualité) 	<ul style="list-style-type: none"> - SPAdes, megahit - QUAST - progressiveMauve (ordonner contigs par rapport à génome apparenté) 	/	/	<ul style="list-style-type: none"> - Kraken - Prokka - MLST - BLASTn (résistance/ gènes de virulence) - ABRicate (virulence) 	<ul style="list-style-type: none"> - Roary/FastTree - plus d'analyses possible (résistance, identification de réplicons de plasmides, sérotypage pour Coli et Salmonella) - Illumina)

**Public Health Bacterial Genomics*