

BP-Tracer workflow

Species tracing of biopollutome

a. Pangenome construction

206876 RefSeq bacteria genomes



Chromosome 1.63%
Complete Genome 10.14%
Contig 53.92%
Scaffold 34.31%

Aggregate by species level

14051 Species genome clusters

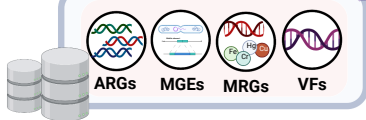
Gene prediction
(Prodigal)

14051 Species ORFs clusters



14051 Species pangenome ORFs

Mapping & extracting



b. Database and pipeline construction

kraken-build database

PGtax



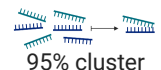
Adjusted WAAFLLE database

PGtrans



Encapsulated database

Gene analysis component



PGfunc

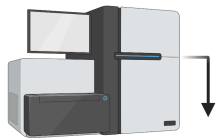
Species-functional gene database

```
>Specie1_ARG_1
DDRELTSLKELLEMEGFNVIVAHDGE
>Specie2_MGE_1
AAACHGLQLPLKQAMSLSFRWLYGPF
>Specie3_MRG_1
ADRFPRRTMLVCLDLVRAAVALLPFV
>Specie4_VF_1
ADGPHTTITVTATDAAGNVGNDTAVVTI
...
```



c. Pipeline input

MGS Sequencing



Raw reads



Quality control
and host removal

Clean reads



Assembly

Contig



d. BP-Tracer Results

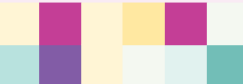
Pathogens detection

Pathogens

Pathogens profiles

Species1

Species2



Harmful genes

Type

Type-gene

Type1

Type2



Species

Host-gene

Species1

Species2



ARG/MRG/MGE/VF

HGT detection

ContigID

HGT events

Contig1

Contig2

