

**Public *M. caprae*
genomes
downloaded from
NCBI
n = 81**

**Public *M. bovis*
genomes
downloaded from
NCBI
n = 3929**

N=457 excluded because part of a pre-publication release
N=130 excluded because called BCG on NCBI
N=1 excluded because misclassification
N=3 excluded because RNA-seq library

N=8 newly sequenced genomes
(PRJEB33773)

**Genomes analysed
with WGS pipeline
n = 81**

**Genomes analysed
with WGS pipeline
n = 3346**

N=4 genomics analysis failed
(low coverage, ratio het/homo > 1)

N=59 genomics analysis failed
(low coverage, ratio het/homo > 1)

**Final dataset
n = 3364**