# CRyPTIC Release One numbers

PWF, 27 August 2020

CRyPTIC

## Pheno     Geno-Pheno-Intersection (GPI)     Geno

**21,055 UKMYC plates inoculated** ⟶ **15,211** ⟵ **70,372 genomes**

All belong to MBTC
15,188 (99.8%) *M. tuberculosis*
(44% Lineage 4, 36% Lineage 2)
12 different *M. bovis,* 10 *M. orygis*

Each has been processed by Clockwork
v0.8.3 and passed various basic
quality checks and a per-sample VCF
produced. The GPI samples ALSO have
been regenotyped and therefore have
TWO VCF files. Both are parsed, but
the regenotyped VCF is used to
populate the rows in the various
genetics tables for the GPI samples.
Includes the ENA FASTAQ files and 17
high-quality samples from Comas/
Gagneux. Does not include Seq&Treat.

**20,637 UKMYC readable plates** ⟶ **15,039**

UKMYC5: 7,965 / UKMYC6: 12,672

19,834 (96.1%) at day 14,
remainder mainly day 21

Predicted to be 43% susceptible (NEJM2018
definition), 32% MDR, 5% XDR, 3% RIF-only
and 18% unknown at present.

UKMYC5: 7,233 / UKMYC6: 7,806

**19,695 UKMYC plates ok to read** ⟶ **<span style="color:red">14,159</span>**   UKMYC5: 6,355 / UKMYC6: 7,804

Due to high levels of discrepancy exclude
all of SITE13 and 01575≤SUBJID≤02350 for
SITE04. No rows for these in
UKMYC_PHENOTYPES.

Note that SUBJID>02350 for SITE04 has an
elevated discrepancy rate and individual
analysis projects may wish to also remove
these 1,797 samples. In general, they have
more growth on the plate than you expect
genetically.

All of these samples have been regenotyped
using Clockwork and have a readable UKMYC
plate whose readings are not suspect.
12,984 (92%) have images and all have been
processed by AMyGDA and, where necessary,
by BashTheBug. They all have kmer counts,
species, lineage and SNP distances to all
other samples in this set are recorded.

The mean number of HIGH quality readings
is 9,857 per drug (76%, excl. PAS) and
ranges from MXF (8,093) to RFB (11,685).

To be added:
BashTheBugPRO is arbitrating of
the LOW quality results and
about 90% are completed.