nature research

Corresponding author(s): Louis S Ates

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We used syngenic mice, at least 3 mice per group are necessary to do statistics, we used this number of mice per group and time point for bacilli quantification and morphometry studies in two independent experiments. A third experiment was added to comply with reviewers requests. Sample size estimation were made based on previous experience with hypervirulent Mycobacterium tuberculosis strains, but were not based on statistical power calculations.

2. Data exclusions

Describe any data exclusions.

Survival Data of the 3rd replicate of in vivo experiments have been excluded from the analyses. In this experiment all infection stocks seemed to be attenuated, which can be seen from lower bacterial lung burdens at later times in the infection. Correspondingly, only three animals died or where killed because of human end-point criteria and no survival data could be extracted from this experiment. This has been communicated to the reviewer/Editor in our latest revised version. Bacterial Lung burdens and lung inflammation were still assessed from this experiment.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental findings were consistent and reproducible (with the exception of the generally reduced virulence in the third replicate described above). All in vivo data are derived from 2 or 3 replicates and individual replicates have been added in the supplementary information.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Animals were distributed randomly, each cage had five mice. Then, 60 animals were infected with each strain. Animals were not pre-selected for any particular group in any way

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The principal investigator (RH-P) only possessed, strain names and no background on the hypothesized virulence (for replicates 1 and 2 replicates). Strains were further blinded for scientist performing the experiments. Strains were further blinded for scientists performing the experiments.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Statistics were calculated with GraphPad Prism software version 6.0. The software used for LC-MS/MS proteomics data analysis is: MaxQuant 1.5.2.8. Open-source software was used for the analysis of Mycobacterium tuberculosis genomes which is described in detail in the supplementary methods of the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Materials and strains are available upon contact with the corresponding authors

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	All antibodies used in these studies have been previously published and validated. anti-PGRS antibody 7C4.1F7 (Abdallah et al. 2009), anti-GroEL2 antibody CS-44 (kind gift from J. Belisle, Colorado state University and the NIH, Bethesda, MD, USA), polyclonal anti-PPE41 (Abdallah et al 2006), anti-EsxA antibody (Hyb76-8) (Harboe et al. 1998) and an anti-hemagglutinin antibody (HA.11, Covance).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used in this study
b. Describe the method of cell line authentication used.	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.*
c. Report whether the cell lines were tested for mycoplasma contamination.	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination OR state that no eukaryotic cell lines were used.*
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	*Provide a rationale for the use of commonly misidentified cell lines OR state that no commonly misidentified cell lines were used.*

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Mus musculus, Balb/C mice, only males with the same age (8 weeks) and weight (22 gr)

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	There were no human research participants in this study