Title: Global assessment of *Mycobacterium avium* subspecies *hominissuis* genetic requirement for growth and virulence.

Short title: MAH genetic requirement for growth and virulence.

Marte S. Dragset[1,2,3¶*], Thomas R. Ioerger[4¶], Maja Loevenich[1], Markus Haug[1,5], Niruja Sivakumar[1], Anne Marstad[1], Pere Joan Cardona[2], Geir Klinkenberg[6], Eric J. Rubin[3], Magnus Steigedal[1,3], Trude H. Flo[1]

[1] Centre of Molecular Inflammation Research and Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway.

[2] Tuberculosis Research Unit, Germans Trias i Pujol Research Institute, Badalona, Barcelona, Spain.

[3] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

[4] Department of Computer Science and Engineering, Texas A&M University, College Station, Texas, USA.

[5] Department of Infection, St. Olavs University Hospital, Trondheim, Norway.

[6] Department of Biotechnology and Nanomedicine, SINTEF Materials and Chemistry, Trondheim, Norway.

[*]Corresponding author

E-mail: marte.dragset@ntnu.no (MSD)

[¶]These authors contributed equally to this work.

Abstract

Nontuberculous mycobacterial infections caused by the opportunistic pathogen *Mycobacterium avium* subsp. *hominissuis* (MAH) are currently receiving renewed attention due to increased incidence combined with difficult treatment. Insights into the disease-causing mechanisms of this species have been hampered by difficulties in genetic manipulation of the bacteria. Here, we identified and sequenced a highly transformable, virulent MAH clinical isolate susceptible to high-density transposon mutagenesis, facilitating global gene disruption and subsequent investigation of MAH gene function. By transposon insertion sequencing (TnSeq) of this strain, we defined the MAH genome-wide genetic requirement for virulence and *in vitro* growth, and organized ~3500 identified transposon mutants for hypothesis-driven research. The majority (71 %) of the genes we identified as essential for MAH *in vitro* had a growth-essential mutual ortholog in the related and highly virulent *M. tuberculosis* (*Mtb*). However, passaging our library through a mouse model of infection revealed a substantial number (54% of total hits) of novel virulence genes. Strikingly, > 97 % of the MAH virulence genes had a mutual ortholog in *Mtb*. Two of the three virulence genes specific to MAH (i.e. no *Mtb* mutual orthologs) were PPE proteins, a family of proteins unique to mycobacteria and highly associated with virulence. Finally, we validated novel genes as required for successful MAH infection; one encoding a probable MFS transporter and another a hypothetical protein located in immediate vicinity of six other identified virulence genes. In summary, we provide new, fundamental insights into the underlying genetic requirement of MAH for growth and host infection.

Author summary

Pulmonary disease caused by nontuberculous mycobacteria is increasing worldwide. The majority of these infections are caused by the *M. avium* complex (MAC), whereof >90% arise from *Mycobacterium avium* subsp. *hominissuis* (MAH). Treatment of MAH infections is currently difficult, with a combination of antibiotics given for at least 12 months. To control MAH by improved therapy, prevention and diagnostics, we need to understand the underlying mechanisms of infection. While genetic manipulation of pathogens is crucial to study pathogenesis, *M. avium* (*Mav*) has been found notoriously hard to engineer. Here, we identify an MAH strain highly susceptible to high-density transposon mutagenesis and transformation, facilitating genetic engineering and analysis of gene function. We provide crucial insights into this strain's global genetic requirements for growth and infection. Surprisingly, we find that the vast majority of genes required for MAH growth and virulence (96% and 97%, respectively) have mutual orthologs in the tuberculosis-causing pathogen *M. tuberculosis* (*Mtb*). However, we also find growth and virulence genes specific to MAC species. Finally, we validate novel mycobacterial virulence factors that might serve as future drug targets for MAH-specific treatment, or translate to broader treatment of related mycobacterial diseases.

Introduction

Mycobacterium avium complex (MAC) is a group of genetically related and ubiquitously distributed opportunistic mycobacteria that can cause nontuberculous infections collectively called MAC disease [1]. *M. avium* (*Mav*), one of the MAC species, has been classified into subspecies *avium*, *paratuberculosis, silvaticum* and *hominissuis* based on molecular characterizations, prevalent hosts and diseases caused [2, 3]. The latter, *M. avium* subsp. *hominissuis* (MAH), can infect humans and lead to pulmonary and disseminated disease, particularly in immunocompromised individuals [4]. MAH infections are currently hard to treat, with a combination of antibiotics typically given for at least 12 months [5]. Similar to its relative *M. tuberculosis* (*Mtb*), the causative agent of tuberculosis, MAH proliferates within macrophages by hijacking normal phagosomal trafficking, overcoming the host's elimination strategies [6-12]. Mechanisms of infection may therefore partly be conserved between the two species. MAH lacks the Type VII ESX-1 secretion system crucial for full *Mtb* virulence [13], suggesting they also differ in virulence strategies. While *Mtb* is an obligate human pathogen in nature, with limited survival outside the host, *Mav* is environmental and opportunistic, found in a variety of niches (e.g. soil, fresh water, showerheads) and a range of prevalent hosts [3]. MAH isolates exhibit high genetic variation [14], perhaps as an adaption to diverse niches and hosts. It is currently not known to what degree MAH and *Mtb* depend on similar mechanisms for growth and virulence, given the same selective conditions. Even so, MAH genes encoding factors required for basic proliferation and virulence may be attractive targets for improved MAH therapies, and may translate to the treatment of related mycobacterial diseases.

Transposon insertion sequencing (TnSeq), which combines transposon mutagenesis and massive parallel sequencing, has been widely used to determine the

conditional requirement of bacterial genes on a genome-wide scale [15]. By massive parallel sequencing of libraries consisting of more than 100 000 transposon mutants, the genetic requirement of *Mtb*, *M. marinum* and *M. avium* subsp. *paratuberculosis* has been defined, for *in vitro* growth or during infection [16-20]. For MAH, transposon mutagenesis has also been of great importance to identify virulence genes [21-28]. However, studies in MAH have to date been limited in library sizes, identifying a smaller number of potential virulence genes per screen. The majority of past MAH mutant libraries were constructed using the clinical isolate MAH 104 [21-25, 28]. 104 was the first MAH strain with a publicly available, fully assembled genome [29], and has thus naturally been widely studied, also by us (exemplified in [8, 10, 11, 29, 30]). MAH 104 is, however, resistant to transposon mutagenesis by ϕMycoMarT7, the phagemid preferred to generate high-density mycobacterial libraries [16-20, 31, 32]. Moreover, MAH 104 (along with many other *Mav* strains) transforms with low efficiency [28, 33-35], complicating the use of this strain in hypothesis-driven genetic manipulation. To facilitate genome-wide investigation of MAH gene function, we identified an MAH strain (MAH 11) highly susceptible to both ϕMycoMarT7-mediated transposon mutagenesis and transformation. We used this strain to generate a transposon insertion library with ~66% saturation density, which we profiled for genes required *in vitro* and in a mouse model of infection. In fact, MAH 11 is currently used as a screening strain in mycobacterial drug discovery programs [36, 37], adding further value to determining the growth requirements of this particular strain. Finally, we constructed an ordered subset of our transposon library in 384-well format (with a unique clone in each well) and used a multiplex sequencing strategy to identify the disrupted gene in each well, providing access to ~3500 MAH mutants. With these, we validated novel MAH virulence factors. Fig 1

120 summarizes the overall experimental setup of our study.

121

122 ## Results

123 **Identification of a ϕMycoMarT7-transducible MAH strain.** We aimed to find an
124 MAH strain in which we could create high-density transposon mutant libraries. The
125 transposon donor phagemid ϕMycoMarT7 is widely used (and recommended over
126 Tn*5367* transposition [31]), for efficient mycobacterial transposon mutagenesis [16-
127 20, 32]. ϕMycoMarT7 is derived from ϕAE87, which originates from
128 mycobacteriophage TM4 [38, 39]. TM4 has been inconsistent in its ability to
129 transduce various *Mav* strains and unable to infect the commonly studied genome
130 sequenced strain MAH 104 [40]. In agreement with these observations, we failed to
131 obtain kanamycin resistant mutants (marker for successful transposition of the
132 ϕMycoMarT7-encoded *Himar1* transposon) when attempting to transduce MAH 104.
133 Hence, to identify a ϕMycoMarT7-transducible strain of MAH, we screened seven in-
134 house clinical *Mav* isolates originating from patients at the National Taiwan
135 University Hospital, Taiwan. Around 70% of the strains resulted in kanamycin
136 resistant colonies after transduction, indicating *Himar1* transposition. Based on
137 particularly efficient transducibility (up to 300 000 kanamycin-resistant colonies per
138 ml starting culture in our initial small-scale screen) and general ease to handle, we
139 focused on a strain isolated from an HIV positive patient's bone marrow, MAH 11.
140 By PCR, we confirmed the presence of transposon inserted in 10 out of 10 colonies
141 tested (S1 Fig). Hence, this MAH strain is highly susceptible to ϕMycoMarT7-
142 mediated transposon mutagenesis.

143

MAH 11 genome sequence. We sequenced the genome of the ϕMycoMarT7-transducible MAH 11 strain on an Illumina HiSeq 2500 instrument in paired-end mode with a read length of 125 bp, yielding a mean depth of coverage of 55.4. The sequence was assembled by a comparative assembly strategy, using MAH 104 as a reference sequence, augmented with contig-building to build large-scale indels. The Illumina data was supplemented with long reads (up to 40kb) from a PacBio sequencer, which were used to confirm the connectivity of the chromosome. The length of the genome is 5,105,085 bp, and is GC-rich (69.2%) like that of other mycobacteria (Fig 2A). The genome sequence is highly concordant with the recently reported draft genome of MAH 11 (70 contigs) [41]. The primary discrepancies involve repetitive sequence elements, such as differences in locations of native transposons and copy-numbers of MIRU tandem repeats, for which assembly from short reads can be ambiguous.

MAH 11 is positively identified as *M. avium* subsp. *hominissuis*, based on 16S rRNA and *hsp65* sequences that are identical to MAH 104 (but distinct from other subspecies, like *M. avium* subsp. *avium*) [2]. However, relative to MAH 104, there are substantial numbers of SNPs (approximately 10 SNPs per 1kb) and a cumulative loss of ~377 kb, showing it is a distinct lineage (Fig 3A). The reductions are clustered in several large-scale deletions, the largest of which (>40kb) are listed in Fig 3B. This variability has been seen in other *Mav* isolates [42, 43], and several of the large-scale deletions correspond to known large-scale polymorphisms [42]. In addition, there are several large-scale insertions (Fig. 3B), including a prophage (56 genes, *b6k05_17725-18015*, inserted at coordinate 3.77 Mbp). This phage is almost identical to a prophage previously reported in the genome of *M. chimaera* str. MC045 (RefSeq NZ_LT703505.1). Also, a cluster of 48 non-phage metabolic genes (*b6k05_03885-*

169 *04170*; 57kb) is inserted in the tail of existing MAH 104 prophage phiMAV-1
170 (*mav_0779-0841*; [44]), containing a variety of hydrolases, reductases,
171 dehydrogenases, and monooxygenases. This gene cluster has been previously
172 observed in other MAH strains, e.g. strain H87 [45]. The MAH 11 genome contains at
173 least 14 copies of IS1245 (similar to MAH 104; [46]), but none of IS901 (associated
174 with members of the MAC complex primarily infecting birds; [42, 47]). S2 Fig shows
175 the position of MAH 11 in a phylogenetic tree (created using PHYLIP,
176 http://evolution.genetics.washington.edu/phylip.html) together with 21 other *M.*
177 *avium* genomes obtained from NCBI GenBank, including *M. avium* subsp.
178 *paratuberculosis* (MAP) K10 and three *M. avium* subsp. *avium* (MAA) as outgroup
179 strains.

180 We identified 4653 ORFs (along with 1 copy of the rRNAs (16S, 23S, 5S) and
181 42 tRNAs, similar to MAH 104) using the NCBI Prokaryotic Genome Annotation
182 Pipeline [48]. 4211 genes have mutual orthologs with MAH 104 (where each gene in
183 one organism is the best match for the ortholog in the other organism, with BLAST E-
184 value < 10^{-10}, S1A Dataset). Almost all of these orthologs (4123) have ≥ 96% amino
185 acid identity, and nearly half (1730) have 100% amino acid identity. For simplicity,
186 we will, where applicable, refer to MAH ORFs using the MAH 104 locus tags
187 (*mav_xxxx*) from here on. Compared to *Mtb* (H37Rv), MAH 11 has ~500 more genes,
188 and over half the genes in each genome (2681) has a mutual ortholog in the other
189 genome (S1B Dataset). Most of the remaining genes also have orthologs, but their E-
190 values are above the stringent threshold of 10^{-10} (i.e. 1052 MAH genes have no clear
191 *Mtb* ortholog, S1C Dataset), or their specific partner in the other genome is
192 ambiguous, as is sometimes the case in duplicated gene families.
193

MAH 11 plasmids. Two large extra-chromosomal contigs that appear to represent circular plasmids were detected. One, pMD1 (193kb, 162 ORFs) (Fig 2B and S1D Dataset), bears weak similarity (based on BLAST search) to parts of plasmids in a wide range of other mycobacteria. The other, pMD2 (78kb, 66 ORFs) (Fig 2C and S1E Dataset), is nearly identical to conjugative plasmid pMA100 from MAH strain 88Br (though reduced, since pMA100 is 116kb, S1F Dataset) [49], and bears partial similarity to the conjugative pRAW-like plasmids found in several slow-growing mycobacterial species [50].

MAH 11 is highly transformable. *Mav* is notoriously hard to transform [28, 33-35], complicating introduction of new DNA and thus genetic engineering of this species. In our and others' experience, MAH 104 is transformable with low efficiency [34, 51]. We investigated the transformation frequency of MAH 11 and found that this strain is around 100 times more susceptible to obtain plasmid DNA compared to the 104 strain, using optimized protocols for *Mav* electroporation (Fig 4A) [33]. We did not observe a notable difference in cell wall integrity (potentially explaining different susceptibility to electroporation) between MAH 11 and MAH 104 when subjecting them to sodium dodecyl sulfate (SDS) stress (S3 Fig). In summary, MAH 11 might be particularly apt for hypothesis-driven genetic approaches.

MAH *in vitro* essential gene set. To define genes required for MAH *in vitro* growth we generated, by ϕMycoMarT7-mediated transduction, a library of ~170,000 transposon mutants selected on 7H10 medium. The *Himar1*-based mariner transposon of ϕMycoMarT7 inserts randomly at TA dinucleotides [52], of which there are 55,516 sites in the MAH 11 genome (excluding plasmids). We sequenced the transposon

219 junctions of two independent libraries, mapped the genomic position of the
220 transposon insertion sites (insertion counts), and counted insertions (reduced to
221 unique templates using barcodes [53]) (S1 Table). The library had a saturation of
222 66.3%, with insertions at 36,813 out of 55,516 TA sites. By gene requirement
223 analysis, using a Hidden Markov Model incorporated into the Transit platform [54],
224 we defined 362 genes as essential for *in vitro* growth, 312 as genes causing growth-
225 defect when disrupted, 278 as genes causing growth-advantage when disrupted and
226 3651 genes as non-essential for growth (Fig 4B, S1G Dataset). 71% (258/362) of
227 MAH 11's essential genes had an essential ortholog in *Mtb* (as defined by DeJesus *et*
228 *al*. [55]). Remarkably, very few MAH (15) and *Mtb* (10) essential genes (~4% and
229 ~2% of total essentials, respectively) did not have a mutual ortholog in the other
230 species (Fig 4C, S1G Dataset), suggesting that the vast majority of genes required for
231 *in vitro* proliferation are conserved between MAH and *Mtb*.

232 Almost all genes found on the two MAH 11 plasmids were non-essential for *in*
233 *vitro* growth (S1H and S1I Datasets). However, on pMD1, four genes caused growth
234 defects of MAH 11 when disrupted; two encoding hypothetical proteins, one
235 homologous to a gene encoding chromosome partitioning protein, *parB*, and one
236 homologous to DNA processing protein-encoding *dprA*. On pMD2, a gene
237 homologous to *rep*, involved in plasmid replication, caused growth defect when
238 disrupted. Taken together, we identified 674 chromosomal and five plasmid MAH
239 genes that were essential or caused a growth defect *in vitro* when disrupted.

240

241 **MAH 11 establishes infection in mice.** We and others have shown that MAH 104 is
242 virulent in mice [8, 11, 12, 22]. We thus examined whether MAH 11 would be
243 suitable to study the role of MAH genes *in vivo*. C57BL/6 mice were infected

244 *intraperitoneally* with MAH 11 or MAH 104, and organ bacterial load was analyzed
245 in the chronic phase of infection. As we have previously shown, bacterial loads
246 remained relatively constant in liver and spleen from 22 to 50 days after initial MAH
247 104 infection [12]. The same trend was seen for MAH 11, albeit with an overall lower
248 bacterial load compared to MAH 104, especially in the spleen (Fig 5A and B). MAH
249 11 and 104 grew comparably in 7H9 medium (Fig 5C), suggesting MAH 11 is less
250 virulent than MAH 104 in mice.

251 T cells produce effector cytokines upon activation to elicit an adaptive immune
252 response towards infections. To control *Mav* infection, production of IFNγ effector
253 cytokine by CD4+ T helper 1 cells is of particular importance [56]. We have
254 previously monitored anti-mycobacterial T cell responses to MAH 104 [12]. To
255 investigate whether MAH 11 is suitable to study MAH-specific host immune
256 responses, we measured mycobacteria-specific CD4+ T cell responses after mouse
257 infection. Fig 5D shows frequencies of MAH-specific CD4+ T helper 1 cells
258 producing IFNγ effector cytokine after infection. Interestingly, the frequencies of
259 IFNγ-producing CD4+ T cells were found lower in MAH 11-infected mice compared
260 to MAH 104-infected mice after 50 days of infection, possibly reflecting the lower
261 organ bacterial loads (Fig 5A and B). Nevertheless, MAH 11 appears suitable to study
262 mycobacterial disease mechanisms, as well as host responses, in a mouse infection
263 model.

264

265 **MAH virulence gene set.** To identify genes required for MAH virulence we infected
266 six mice with our MAH library and analyzed bacterial load from the spleen and liver
267 (organs from two animals each were pooled, resulting in three spleen libraries and
268 three liver libraries) after 26 days of infection. We sequenced the harvested libraries,

which yielded saturation of 61.9% and 71.0% (combined over replicates) for spleen and liver, respectively (S1 Table). We then defined the genetic requirement for infection using a Transit-incorporated resampling algorithm for comparative analysis [54], comparing output data from sequenced libraries before and after infection. We identified 144 and 128 genes as required for spleen and liver infection, respectively (S1B Dataset). 112 genes were required for survival in both organs (~80% overlap). Direct comparison of the spleen and liver datasets by resampling did not reveal any statistically significant differences, and hence no genes uniquely required for colonization of either organ were identified. Among the core genes identified (found in both spleen and liver) were 51 previously identified in *Mtb* mouse model TnSeq experiments [16], including well-established mycobacterial virulence genes like *uvrABC* (the UvrABC endonuclease complex [57]), *secA2* (alternative ATPase of Sec secretion pathway [58]), *icl* (isocitrate lyase [59]), *bioA* (within the biotin biosynthesis pathway [60]) and *glcb* (malate synthase [61]). However, importantly, we identified 61 core genes (92 genes with spleen and liver genes combined) not previously detected by *Mtb* TnSeq virulence gene screening (Fig. 4D)[16]. Some of these genes were found in genetic clusters, like six genes within the region encoding the Type VII ESX-5 secretion system. Other genes were found in operons (for instance *prcA*/*prcB*, *mav_3300*/*ripA*, *mav_3691*/*rbfA*) or in close genomic vicinity (*mav_4154*/*4158*/*4159*/*4160*/*4163*). Strikingly, only three (< 3%) core MAH virulence genes did not have a mutual ortholog in *Mtb* (Fig 4D). Two of them, *mav_4273* and *mav_4274*, are potentially co-expressed, and both encode proline-proline-glutamic acid (PPE) family proteins. PPE proteins might show ambiguity in ortholog matching due to large duplications within the family. However, by BLAST search, the two PPE proteins have clear orthologs in MAC species [62], but not in other well-known

mycobacterial species like *Mtb*, *M. bovis*, *M. abscessus* nor *M. leprae* (albeit both PPE proteins show a weak similarity to *M. marinum* PPE14 (*mmar_1235*) with 53 and 57% amino acid identity, respectively). The last of the three MAH virulence genes without *Mtb* mutual orthologs, *mav_4409*, encodes a putative acyltransferase.

As could be expected, most genes found on the two MAH 11 plasmids were non-essential for infection (S1J-M Datasets); the only exceptions were an AAA-family ATPase on pMD1 (out of 162 ORFs), and two ORFs of unknown function on pMD2 (66 ORFs).

Identification of mutants in an organized MAH library. Organized libraries of insertion mutants are of great value for hypothesis-driven research. Vandewalle *et al.* developed a method where sequence tagging transposon library pools was used to bulk-identify, by TnSeq, both the gene disrupted and the location of the mutant within an organized (plated) *M. bovis* BCG transposon library [63]. We employed such a strategy to identify mutants within an ordered MAH 11 library of 9216 transposon-insertion mutants, plated in 384-well plates using a colony-picking robot. We tagged the library by plates, rows and columns, then pooled and sequenced the samples, and analyzed the barcode-to-genome coordinate maps to determine the location of the mutants and the transposon insertion sites in bulk. We were able to map the specific location of 2696 unique transposon insertion mutants within 1697 (34.8%) of the 4881 MAH 11 ORFs (plasmid ORFs included) (S1N Dataset). Transposon insertion sites that mapped ambiguously to more than one plate, column and/or row were disregarded; these might be due to a relatively high number of mutant duplicates in the picked library. 3161 wells had a unique clone, and only 155 wells had more than one mutant assigned to them (S1O Dataset). The latter might be due to mutants

319 clumping in colonies picked, incomplete sterilization of the robotic picking device
320 between rounds of colony picking, or transposon insertions in repetitive or duplicated
321 regions. To experimentally verify the correct location of mutants, we sequenced 11
322 mutants picked from 11 wells (S2 Table). All mutants had the transposon inserted at
323 the location predicted by our TnSeq approach. Taken together, we identified the
324 transposon insertion site and mapped the unambiguous location of 3489 clones
325 (including those in intergenic regions), providing access to a plethora of mutants to
326 study the role of the respective MAH genes.
327
328 ***uvrB* is required for MAH virulence.** UvrABC is an enzyme complex involved in
329 *Escherichia coli* nucleotide excision repair [64]. The genes encoding the
330 mycobacterial homologues of the three members of the complex, *uvrA*, *uvrB*, and
331 *uvrC*, were all defined as virulence genes in our screen. *uvrB* has previously been
332 implicated in *Mtb* virulence, via protection against host-mediated reactive nitrogen
333 and oxygen intermediates [57]. To verify the involvement of *uvrB* in MAH virulence,
334 we infected mice with an available *uvrB* transposon mutant (*uvrB::tn*) and
335 complemented mutant for 26 days. *uvrB::tn* showed reduced bacterial burden in
336 infected mice compared to MAH 11 wild-type (wt) and complement-infected animals
337 (Fig 6A and B). Neither the *uvrB* mutant nor the complemented mutant showed
338 reduced fitness *in vitro* (Fig 6E), hence, our results suggest that UvrB is required for
339 full virulence in MAH.
340
341 **A probable MFS transporter is required for MAH virulence.** Next, we aimed to
342 validate MAH determinants not previously implicated in mycobacterial virulence. A
343 mutant of a probable MFS transporter, MAV_1005 (ortholog of Rv0876c), showed

attenuated growth in our virulence screen. When we subjected an MAH 11 transposon mutant of this transporter (*1005::tn*) to mouse model infection, we saw a strong attenuation after 26 days of infection compared to wt and the complemented mutant (Fig 6A and B). Furthermore, the mutant but not the complemented mutant elicited a reduced MAH-specific CD4+ effector T cell response from the host (Fig 6C and D). The mutant did not show reduced fitness *in vitro* (Fig 6E). Taken together, our results suggest that *mav_1005* is crucial for MAH virulence.

A hypothetical gene required for MAH virulence. Five genes located in close genomic vicinity (region spanning from *mav_4154* to *mav_4163,* Fig 6F) appeared as hits in our virulence screen. An insertion mutant of one of the genes, hypothetical gene *mav_4160* (*4160::tn*), was subjected to mouse model infection. After 26 days of infection, *4160::tn* showed attenuated growth in both liver and spleen (Fig 6A and B), though the MAH-specific effector CD4+ T cell response was not significantly reduced (Fig 6C and D). The mutant did not show reduced fitness *in vitro* (Fig 6E). Hence, our findings suggest that *mav_4160* is required for full MAH virulence.

Inflammation and tissue pathology. We performed a broader characterization of the inflammation and tissue pathology in MAH 11-infected mice (including mutants and complemented mutants) at day 26 post infection. In brief, induction of organ homogenate cytokine production was low or not increased in response to infection (TNFα and IFNγ), except for IL-1β, which largely reflected organ bacterial loads (Fig 6 and S4 Fig). The overall low induction levels were not surprising; cytokines could be secreted by subsets of immune cells and act in an autocrine/paracrine manner e.g. in tissue granulomas. We have previously characterized the C57Bl/6 infection of

MAH 104 in great detail [12]. Similar to what we observed with MAH 104, MAH 11 infection induced organ pathology seen as disruption of splenic pulp structures, infiltration of immune cells, inflammatory foci (granulomas) and giant cell formation (S5 Fig). However, no obvious differences in cytokine production or tissue pathology were seen between mutants and wt, nor mutants and complemented mutants. Nevertheless, the overall organ pathology and IL-1β-production grossly reflected bacterial loads.

Discussion

To study the role of genes by loss of function is a powerful approach to understand how pathogens proliferate and avoid host elimination. We identified a virulent clinical isolate of MAH susceptible to genome-wide high-density gene disruption by ϕMycoMarT7-mediated transposon mutagenesis. The generated transposon library enabled us to define the MAH *in vitro* essential and virulence gene sets using a top-down discovery-based deep sequencing approach. A total of 674 genes were identified as required for normal growth *in vitro* (15% of total genes, similar to the proportion of required genes in *Mtb* [55]). There was a substantial overlap (71%) of *in vitro* essentials between MAH and *Mtb*, as well as many common virulence genes (e.g. *uvrABC*, *secA2*, *icl*) required for survival in a mouse model of infection. However, the majority of the virulence genes we identified were novel relative to *Mtb* TnSeq virulence screening [16]. Surprisingly, even with 1052 genes with no clear *Mtb* ortholog, only three of the MAH virulence genes were specific to MAH (i.e. did not have a mutual ortholog in the H37Rv genome). Two of these were PPE family proteins (*mav_4273/4*, found in MAC species but not *Mtb* [62]), which are unique to mycobacteria, associated with virulence, and some have been shown to be secreted by

394 the ESX-5 secretion system in *M. marinum* and *Mtb* [65, 66]. In fact, we observed six
395 genes within the MAH *esx-5* gene cluster among our virulence genes (*mav_2916-*
396 *2933*), strongly indicating a crucial role of ESX-5 during MAH infection. However,
397 when we subjected an ESX-5 mutant (*eccA5::tn*) to mouse infection, we did not see
398 attenuated growth compared to wt (S6 A, B and E Fig), perhaps due to insufficient
399 disruption of gene function in this mutant (transposon insertion in C-terminus), or that
400 this particular gene is not required for ESX-5-mediated virulence in mice, as
401 previously seen for *Mtb* [67]. Intriguingly though, we saw an increased CD4+ T cell
402 host response when *eccA5* was overexpressed (S6 C and D Fig). Even so, it is
403 possible that the two MAC-specific PPE proteins are secreted via the ESX-5 secretion
404 system of MAH. These PPE proteins could be excellent candidates for targeted drug,
405 vaccine and/or diagnostics discovery for improved control of MAC infections.

406 MAH and *Mtb* are both able to persist in human macrophages; however, *Mav* is
407 environmental and opportunistic, while *Mtb* is an obligate human pathogen. The
408 relatively modest overlap between MAH and *Mtb* in mutually orthologous virulence
409 genes (46%) compared to the overlap of *in vitro* essential genes (71%) might reflect
410 different mechanisms of virulence. However, interestingly, many of the genes that
411 were unique to our MAH virulence screen have been experimentally proven required
412 for *Mtb* virulence (exemplified by proteasome subunits *prcA*, *prcB* and genes
413 encoding the *esx-5* secretion system [66, 68]). This suggests that a portion of the
414 genes we found unique to MAH virulence, and that have *Mtb* mutual orthologs, might
415 be required for *Mtb* virulence as well. Even so, it is evident that when MAH and *Mtb*
416 are subjected to the same selective conditions (*in vitro* growth on 7H10 agar or *in vivo*
417 growth in C57BL/6 mice), MAH depends on very few genes that do not have mutual
418 orthologs in *Mtb* (~4% and ~3%, respectively). Whether this is true also for other

MAH isolates, remains to be investigated. None of the genes within the major MAH 11 insertions (relative to MAH 104, listed in Figure 3B) were required for infection, and only four were essential *in vitro*. It has been shown that *M. marinum* customizes its virulence mechanisms to infect different animal cells [19]. It is thus possible that, if subjected to infection of other animal models, a greater proportion of genes specific to MAH would be required.

Interestingly, we identified genes on the two plasmids, pMD1 and pMD2, that were required for growth *in vitro* or *in vivo*. The *in vitro* growth defect seen in disruption of plasmid replication (*rep*) and partition (*parB*) genes might indicate that the presence of the two plasmids is required for efficient MAH proliferation, or that disturbing plasmid replication/partition reduces the global fitness of the MAH cells. It is currently unclear which role the three plasmid genes we found required *in vivo* might play during infection.

Using defined transposon mutants from our organized (plated) library, we validated a subset of our MAH virulence hits. We verified that the excision repair protein UvrB, a probable MFS transporter, and a hypothetical gene located within a genomic region of several identified virulence genes, are required for full MAH virulence. UvrB has previously been implicated in mycobacterial virulence [57], while the MFS transporter (*mav_1005*) and the hypothetical gene (*mav_4160*) were first validated as mycobacterial virulence factors by us. Of the six other virulence genes identified (based on MAH and *Mtb* screening) in the immediate vicinity of *mav_4160*, two encode hypothetical proteins, one a hydroxylase, one RNA helicase E (RhlE), one isochorismate synthase (EntC) and one phosphoglycerate mutase (Gmp2) (Fig. 6E). EntC was shown to be required for siderophore production in *M. smegmatis* [69], and might thus play a role in iron acquisition during infection. However, the

mechanism of the MFS transporter, as well as *mav_4160* and the surrounding genes, in mycobacterial virulence remain to be elucidated.

Mav isolates exhibit high genetic variation [14, 42, 70, 71]. In accordance, we registered several large-scale insertions and deletions when we compared the genomic sequence of MAH 11 to MAH 104. The movement of genetic material between organisms is mediated by, among other mechanisms, phage transduction, natural transformation, and plasmid conjugation [72]. The cause (and effect) of *Mav* genomic plasticity is largely unknown. However, recently, mycobacterial conjugative plasmids have been identified, also in *Mav* [49, 50]. Interestingly, one of MAH 11's plasmids, pMD2, is almost identical to previously described *Mav* plasmid pMA100 [49]. pMA100 was shown to transfer via conjugation between the slow growing mycobacteria *Mav* and *M. kansasii* in a mixed infection patient [49]. pMD2 might thus partake in genetic exchange between MAH 11 and other mycobacteria.

In conclusion, we identified a highly transformable MAH strain susceptible to ϕMycoMarT7-mediated transposon mutagenesis. This strain enabled genome-wide identification of *in vitro* essential and virulence genes in this species. Based on our screens, we identified growth and virulence genes specific to MAH, as well as shared with *Mtb*. MAH-specific genes might be excellent targets for MAC disease control, while shared genes might target related mycobacterial diseases as well. We validated two novel MAH genes required for infection. Since MAH 11 is used as a screening strain in mycobacterial drug discovery programs [36, 37], a comprehensive understanding of the genetic requirement for growth and infection of this strain is a direct asset for current initiatives towards new anti-mycobacterial therapies.

Materials and methods

469 **Strains and growth conditions.** MAH strains used in this study were MAH 104 and
470 MAH 11 (NCBI GenBank accession number CP000479 and NBAW00000000,
471 respectively). MAH strains were cultured in Middlebrook 7H9 (BD Difco)
472 supplemented with 0.2% glycerol, 0.05% Tween 80 and 10% ADC (50 g BSA
473 fraction V, 20 g dextrose, 8.5 g NaCl, 0.03 g catalase, dH_2O up to 1 L) for liquid
474 growth and in Middlebrook 7H10 (BD Difco) supplemented with 0.5% glycerol and
475 10% ADC for solid growth. For selection of transposon mutants, 20 μg/ml kanamycin
476 and 0.1% Tween 80 was added to the 7H10 agar plate, the latter to simplify library
477 harvest.
478
479 **Genome sequencing and annotation.** DNA from late log phase cultures of MAH 11
480 was extracted using a Masterpure DNA Purification kit (Epicentre), prepared using
481 the TruSeq genome DNA sample preparation kit (llumina, Inc.), and sequenced on an
482 Illumina HiSeq 2500 instrument in paired-end mode with a read length of 125 bp.
483 The genome sequence was assembled using a comparative assembly method, using
484 MAH 104 as a reference sequence. SNPs were identified from mapped reads, which
485 were aligned to the reference sequence using BWA [73] allowing up to 5/125
486 mismatches, and insertions/deletions (indels) were detected and repaired using local
487 contig-building (as described in [74]). In addition, large-scale insertions and plasmid
488 sequences were assembled de novo using Newbler (Roche, Inc.), and integrated into
489 the genome where connectivity was supported by evidence from paired reads. The
490 genome was annotated using PGAP [48], the NCBI Prokaryotic Genome Annotation
491 Pipeline, which employs GeneMarkS+ to identify ORFs (along with RNAs and
492 pseudo-genes) and assigns putative functions and gene identification based on
493 homology. The sequences have been deposited in NCBI Genbank with accession

494 numbers NBAW00000000 (aviumMD30 assembly of MAH 11), CM009838.1
495 (pMD1) and CM009839.1 (pMD2). aviumMD30 is the assembly and annotation
496 analyzed in this paper. For genome completion, the MAH 11 strain was re-sequenced
497 on a PacBio Sequal instrument. A total of 650Mb of long reads (up to 40kb) was
498 collected. There were 130,570 subreads, with a mean read length of 4,937. Reads
499 were mapped to the aviumMD30 assembly using blasr (version 5.3.2) [75]. Coverage
500 of aligned segments was tabulated at each site. The PacBio data was used to confirm
501 the connectivity of the genome by showing that all sites with low coverage (0-10x) in
502 the Illumina data were spanned by PacBio reads, verifying that the chromosome
503 consists of a single 5.1Mb contig. Similarly, the PacBio reads were aligned to the
504 plasmid sequences pMD1 and pMD2 to confirm their continuity as circular, extra-
505 chromosomal DNA. The updated assembly of the MAH 11 genome sequence
506 (aviumMD36) is deposited in Genbank under accession number CP035744.
507
508 **Transformation.** Competent MAH cells were prepared as previously optimized for
509 *Mav* [33]. 100 µl of competent cells was electroporated with 2 µg plasmid DNA
510 (pMSP12::cfp, kind gift from Christine Cosma and Lalita Ramakrishnan [76]) in a 2
511 mm cuvette at settings 2.5 kV, 1000 ohm, and 25 µF. Cells were recovered overnight
512 and plated at serial dilutions for colony forming unit (CFU) counts. CFU counts of
513 pMSP12::cfp-transformed bacteria selected on 7H10 with 20 µg/ml kanamycin were
514 normalized to the CFU counts of transformed bacteria titered on 7H10 without
515 antibiotics.
516
517 **Growth curves.** MAH wt and mutants were grown in 7H9 medium until they reached
518 stationary phase, then diluted to OD_{600} 0.02 in triplicates of 200 µl 7H9 medium in

519 microplate honeycomb wells (Oy Growth Curves Ab Ltd.). The growth was
520 monitored over the course of 10 days in a Bioscreen growth curve reader (Oy Growth
521 Curves Ab Ltd.), shaking at 37 °C.
522
523 **Generation of MAH transposon mutant library.** The MAH high-density
524 transposon mutant library was prepared using ϕMycoMarT7 as previously described
525 for *Mtb* [77], with the exceptions of growing the bacterial culture to stationary as
526 opposed to exponential phase prior to transduction. The amount of ϕMycoMarT7
527 added for transduction was increased coordinately with the increased bacterial
528 density. Both the phagestock and bacterial culture were heated to 37°C before
529 transduction. The library was incubation at 37°C on 7H10 plates with Tween 80
530 (0.1%) and kanamycin (20 µg/ml) for 2-3 weeks.
531
532 **Transposon insertion sequencing (TnSeq).** The transposon library was harvested
533 and pooled by scraping 7H10 plates with 170 000 colonies. Total DNA was purified
534 using Masterpure DNA Purification kit (Epicentre), and prepared for TnSeq by PCR
535 amplification of transposon:genome junctions and adapter ligation following the
536 protocol in [53] . The samples were sequenced on an Illumina GAII instrument,
537 collecting around 10 million 54 bp paired-end reads per sample. The reads were
538 processed using TPP in Transit [54], which counts reads mapping to each TA
539 dinucleotide site (after eliminating reads sharing the same template barcode; [53]).
540
541 **MAH *in vitro* essential gene set.** Essential genes were identified using a Hidden
542 Markov Model (HMM, [78]), incorporated into Transit [54]. The HMM is a Bayesian
543 statistical model that parses the genome into contiguous regions labeled as one of 4

544 states - essential (ES), non-essential (NE), growth-defect (GD, suppressed insertion
545 counts), or growth-advantaged (GA, insertion counts higher than average) - based on
546 local insertion density and mean read count at TA sites. The output of the HMM was
547 processed by labeling each gene with the majority state among the TA sites spanned
548 by it.
549
550 **Mouse infection.** For MAH 104 and MAH 11 infection experiments, groups of four
551 C57BL/6 mice were infected *intraperitoneally* with $5x10^7$ CFU/mouse as previously
552 described [12]. On day 22 and day 50 post infection, MAH-specific effector T cell
553 responses and bacterial load were analyzed. Numbers of CFUs per gram organ were
554 measured by plating serial dilutions of spleen and liver homogenates on 7H10 plates.
555 For the virulence gene screen, six mice were infected *intraperitoneally* with $6x10^7$
556 CFU/mouse of the MAH 11 transposon mutant library. After 22 days of infection,
557 mice were sacrificed and the liver and spleen were harvested, homogenized and plated
558 on 7H10 (livers and spleens from two mice were pooled to make one library, resulting
559 in three liver and three spleen libraries in total from the six mice). After 2-3 weeks at
560 37°C, the colonies were scraped and DNA prepared for sequencing as described
561 above for TnSeq. For *in vivo* validation experiments of virulence genes, groups of 4-5
562 C57BL/6 mice were infected *intraperitoneally* with (approximately) $7.5x10^7$
563 CFU/mouse MAH 11 wt or MAH 11 transposon insertion mutants. On day 26 post
564 infection, bacterial load in liver and spleen, MAH-specific effector T cell responses,
565 cytokine levels in organs and serum and histopathology were analyzed.
566
567 **MAH-specific T cell response.** Isolated splenocytes from infected mice were
568 stimulated overnight with MAH (MOI 3:1); protein transport inhibitor cocktail

569 (eBioscience) was added for the last 4h of incubation. Unstimulated cells were used
570 as controls. Cells were harvested and stained with Fixable Viability Dye eFluor 780
571 (eBioscience) and fluorescence-labelled monoclonal antibodies against CD3 (FITC,
572 eBioscience) and CD4 (Alexa Fluor700 or Brilliant Violet 605, both from
573 BioLegend). After fixation and permeabilization, intracellular cytokine staining was
574 performed with fluorescent monoclonal antibody against IFNγ (Phycoerythrin,
575 eBioscience) and TNFα (Allophycocyanin, BioLegend). Cells were analyzed by flow
576 cytometry on a BD LSR II flow-cytometer (BD Biosciences) and data subsequently
577 analyzed with FlowJo (FlowJo, LLC) and GraphPad Prism (GraphPad Software, Inc.)
578 software. Frequencies of IFNγ- and TNFα-producing CD4+ effector T cells were
579 analyzed from FSC/SSC-gated, viable CD3+CD4+ T cells. The method is described
580 in further detail in [12].
581
582 **Cytokine measurements.** IL-1β, IFN-γ and TNF-α levels were analyzed in serum as
583 well as spleen and liver homogenates from infected mice using a custom-made
584 ProcartaPlex immunoassay panel (Affimetrix, eBioscience) according to the
585 manufacturer's protocol.
586
587 **Histopathology.** Standard hematoxylin & eosin staining of spleen and liver sections
588 was performed at the Cellular and Molecular Imaging Core Facility (CMIC) at
589 NTNU as described previously [12]. Images were acquired with a Nikon E400
590 microscope and NIS-Elements BR imaging software (Nikon Instruments, Melville,
591 NY, USA).
592

593 **MAH virulence gene set.** For comparative analysis between the *in vitro*-selected and
594 the *in vivo* (mouse infection)-selected transposon libraries, the TRANSIT-
595 incorporated 'resampling' algorithm was used [54]. Resampling is analogous to a
596 permutation test, examining if the sum of transposon insertion read counts differs
597 significantly between conditions.
598
599 **Bulk-identification of transposon insertion sites in the organized MAH library.**
600 *Organizing, pooling and sequence-tagging the library.* 9216 colonies were picked and
601 transferred to 24 384-well plates in duplicate using a Genetix QPixII colony picker
602 and software QSoft XP Picking. The cultures were incubated at 37°C for 3 weeks.
603 The library was pooled by plates (24), columns (24) and rows (16), giving a total of
604 64 culture pools using a Freedom EVO 200 (TECAN) liquid handling robot and
605 software EVOsim. Of the total 64 pools, each well should be represented three times;
606 in one plate pool, in one column pool and in one row pool. The various pools were
607 tagged by ligating barcoded adapters to the DNA fragments after DNA purification.
608 Barcodes 1-24 represent plates and columns 1-24, while barcodes 1-16 represent rows
609 1-16 (further details in S1 Materials and Methods and S3 Table).
610 *Analysis of arrayed library of Tn mutants.* The three tagged pools were sequenced
611 on an Illumina HiSeq 2500 with 125 bp paired-end reads, collecting 3.2-5.3 million
612 pairs of reads each pool. The genomic portions of the reads (in read 1) were mapped
613 to TA sites in the MAH 11 genome (including plasmids) using BWA [73]. The
614 barcodes (8 bp embedded in read 2) were extracted and tabulated for each insertion
615 coordinate. Subsequently, a script was written that compared the TA sites represented
616 by each combination of row-, column-, and plate-barcode to associate each TA site
617 with the most probable well. The count of each barcode for each site was normalized

by dividing by the barcode total over all sites. Barcodes with insufficient counts (<10,000 total) were excluded from the calculation. Then the relative frequencies of a plate-, row-, and column-barcodes for each site, $f_{i,p}^{P}$, $f_{i,r}^{R}$, and $f_{i,c}^{C}$ were computed by dividing by the total abundance of plate-, row-, and column barcodes represented by the site. Finally, a score $s(p,r,c) = f_{i,p}^{P} \cdot f_{i,r}^{R} \cdot f_{i,c}^{R}$ was computed for each possible combination of plate, row-, and column-barcodes that represented the likelihood of the well assignment for each TA site *i*. Wells with high probability ($s \geq 0.8$ for the maximal combination) were accepted as unique assignments; wells with $s < 0.8$ were rejected as ambiguous (i.e. potentially mapping to multiple wells).

Verification of mutants by PCR and Sanger sequencing. Total DNA was isolated using Masterpure Complete DNA Isolation kit (Epicentre). PCR to confirm presence of transposon was preformed using GoTaq® Green Master Mix (Promega) and primers KanF (TGGATTGCACGCAGGTTCTC) and KanR (CGTCAAGAAGGCGATAGAAG). Sanger sequencing was preformed using gDNA as template prepared with BigDye Terminator Cycle Sequencing kit v.1.1 (Thermo Fisher); with 60 rounds cycles of 95ºC 30 sec, 52ºC 30 sec, 60ºC 4 min, and primer KanSeq2 (CTTCCTCGTGCTTTACGG) reading directly into the gDNA. Samples were purified using BigDye XTerminator kit (Applied Biosystems) and sequenced on an ABI130xl.

Complementation of transposon insertion mutants. Plasmids for complementation of transposon insertion mutations were constructed by cloning the wt version of the disrupted gene into the mycobacterium-escherichia coli shuttle vector pMV261 ([79]) for constitutive expression. To select for the plasmids upon transformation into the

643 mutated strains, the kanamycin resistance gene from pMV261 was swapped with the
644 hygromycin resistance gene of pUV15TetORm [80], creating pMV261H.
645 *b6k05_04950* and *b6k05_04945*, *b6k05_18820*, *b6k05_12440*, and *b6k05_13510* were
646 amplified from the MAH 11 genome and cloned into pMV261H, resulting in plasmids
647 pMV261H *1005*, pMV261H *4160*, pMV261 *eccA5*, and pMV261 *uvrB*, respectively.
648 All clonings were performed using Gibson Assembly® Master Mix (New England
649 Biolabs). Primer sequences can be provided upon request. The complementing
650 plasmids were transformed into their respective MAH 11 mutant, resulting in strains
651 *1005::tn* compl., *4160::tn* compl., *eccA5::tn* compl., and *uvrB::tn* compl.
652
653 **Ethics statement.** The protocols on animal work were approved by the Norwegian
654 Animal Research Authorities (Forsøksdyrutvalget, FOTS ID 5955). All procedures
655 involving mice experiments were carried out in accordance with institutional
656 guidelines, national legislation and the Directive of the European Convention for the
657 protection of vertebrate animals used for scientific purposes (2010/63/EU).
658

659 ## Acknowledgements

660 We thank Jun-Rong Wei, Jung-Yien Chien and Po-Ren Hsueh for providing clinical
661 MAH isolates from the National Taiwan University Hospital, the Medical Genetics
662 Department at St. Olavs Hospital for assistance with Sanger sequencing, the Cellular
663 and Molecular Imaging Core Facility (CMIC) at NTNU for histopathology staining
664 and personnel at the Comparative Medicine Core Facility (CoMed) at NTNU for
665 assistance in animal experiments. CMIC and CoMed are funded by the Faculty of
666 Medicine and Health Science at NTNU and Central Norway Regional Health
667 Authority. We acknowledge the help of the Texas A&M Sequencing Center

(Genomics & Bioinformatics Service, Dr. Charlie Johnson, director).

References

1. Prevots DR, Marras TK. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: a review. Clin Chest Med. 2015;36(1):13-34.
2. Turenne CY, Wallace R, Jr., Behr MA. Mycobacterium avium in the postgenomic era. Clin Microbiol Rev. 2007;20(2):205-29.
3. Rindi L, Garzelli C. Genetic diversity and phylogeny of Mycobacterium avium. Infect Genet Evol. 2014;21:375-83.
4. Falkinham JO, 3rd. Epidemiology of infection by nontuberculous mycobacteria. Clin Microbiol Rev. 1996;9(2):177-215.
5. Thomson RM, Yew WW. When and how to treat pulmonary non-tuberculous mycobacterial diseases. Respirology. 2009;14(1):12-26.
6. Appelberg R. Pathogenesis of Mycobacterium avium infection: typical responses to an atypical mycobacterium? Immunol Res. 2006;35(3):179-90.
7. Awuh JA, Flo TH. Molecular basis of mycobacterial survival in macrophages. Cell Mol Life Sci. 2017;74(9):1625-48.
8. Awuh JA, Haug M, Mildenberger J, Marstad A, Do CP, Louet C, et al. Keap1 regulates inflammatory signaling in Mycobacterium avium-infected human macrophages. Proc Natl Acad Sci U S A. 2015;112(31):E4272-80.
9. Ernst JD. The immunological life cycle of tuberculosis. Nature Reviews Immunology. 2012;12(8):581-91.
10. Gidon A, Asberg SE, Louet C, Ryan L, Haug M, Flo TH. Persistent mycobacteria evade an antibacterial program mediated by phagolysosomal TLR7/8/MyD88 in human primary macrophages. PLoS Pathog. 2017;13(8):e1006551.
11. Halaas O, Steigedal M, Haug M, Awuh JA, Ryan L, Brech A, et al. Intracellular Mycobacterium avium intersect transferrin in the Rab11(+) recycling endocytic pathway and avoid lipocalin 2 trafficking to the lysosomal pathway. J Infect Dis. 2010;201(5):783-92.
12. Haug M, Awuh JA, Steigedal M, Frengen Kojen J, Marstad A, Nordrum IS, et al. Dynamics of immune effector mechanisms during infection with Mycobacterium avium in C57BL/6 mice. Immunology. 2013;140(2):232-43.
13. Champion PA, Champion MM, Manzanillo P, Cox JS. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol Microbiol. 2009;73(5):950-62.
14. Uchiya K, Tomida S, Nakagawa T, Asahi S, Nikai T, Ogawa K. Comparative genome analyses of Mycobacterium avium reveal genomic features of its subspecies and strains that cause progression of pulmonary disease. Sci Rep. 2017;7.
15. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009;6(10):767-U21.

16. Zhang YJJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, et al. Tryptophan Biosynthesis Protects Mycobacteria from CD4 T-Cell-Mediated Killing. Cell. 2013;155(6):1296-308.
17. Zhang YJJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, et al. Global Assessment of Genomic Regions Required for Growth in Mycobacterium tuberculosis. PLoS Pathog. 2012;8(9).
18. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism. PLoS Pathog. 2011;7(9).
19. Weerdenburg EM, Abdallah AM, Rangkuti F, El Ghany MA, Otto TD, Adroub SA, et al. Genome-Wide Transposon Mutagenesis Indicates that Mycobacterium marinum Customizes Its Virulence Mechanisms for Survival and Replication in Different Hosts. Infect Immun. 2015;83(5):1778-88.
20. Wang J, Pritchard JR, Kreitmann L, Montpetit A, Behr MA. Disruption of Mycobacterium avium subsp paratuberculosis-specific genes impairs in vivo fitness. BMC Genomics. 2014;15.
21. Motamedi N, Danelishvili L, Bermudez LE. Identification of Mycobacterium avium genes associated with resistance to host antimicrobial peptides. J Med Microbiol. 2014;63(Pt 7):923-30.
22. Li YJ, Danelishvili L, Wagner D, Petrofsky M, Bermudez LE. Identification of virulence determinants of Mycobacterium avium that impact on the ability to resist host killing mechanisms. J Med Microbiol. 2010;59(Pt 1):8-16.
23. Li Y, Miltner E, Wu M, Petrofsky M, Bermudez LE. A Mycobacterium avium PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol. 2005;7(4):539-48.
24. Danelishvili L, Wu M, Stang B, Harriff M, Cirillo SL, Cirillo JD, et al. Identification of Mycobacterium avium pathogenicity island important for macrophage and amoeba infection. Proc Natl Acad Sci U S A. 2007;104(26):11038-43.
25. Bermudez LE, Danelishvili L, Babrack L, Pham T. Evidence for genes associated with the ability of Mycobacterium avium subsp. hominissuis to escape apoptotic macrophages. Frontiers in cellular and infection microbiology. 2015;5:63.
26. Bermudez LE, Rose SJ, Everman JL, Ziaie NR. Establishment of a Host-to-Host Transmission Model for Mycobacterium avium subsp. hominissuis Using Caenorhabditis elegans and Identification of Colonization-Associated Genes. Frontiers in cellular and infection microbiology. 2018;8:123.
27. Rose SJ, Bermudez LE. Identification of Bicarbonate as a Trigger and Genes Involved with Extracellular DNA Export in Mycobacterial Biofilms. mBio. 2016;7(6).
28. Khattak FA, Kumar A, Kamal E, Kunisch R, Lewin A. Illegitimate recombination: an efficient method for random mutagenesis in Mycobacterium avium subsp. hominissuis. BMC Microbiol. 2012;12:204.
29. Horan KL, Freeman R, Weigel K, Semret M, Pfaller S, Covert TC, et al. Isolation of the genome sequence strain Mycobacterium avium 104 from multiple patients over a 17-year period. J Clin Microbiol. 2006;44(3):783-9.
30. Uchiya K, Takahashi H, Yagi T, Moriyama M, Inagaki T, Ichikawa K, et al. Comparative Genome Analysis of Mycobacterium avium Revealed Genetic

759 Diversity in Strains that Cause Pulmonary and Disseminated Disease. PLoS
760 One. 2013;8(8).
761 31. Rathnaiah G, Bannantine JP, Bayles DO, Zinniel DK, Stabel JR, Grohn YT, et al.
762 Analysis of Mycobacterium avium subsp. paratuberculosis Mutant Libraries
763 Reveals Loci-dependent Transposition Biases and Strategies to Novel Mutant
764 Discovery. Microbiology. 2016.
765 32. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of
766 conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A.
767 2001;98(22):12712-7.
768 33. Lee SH, Cheung M, Irani V, Carroll JD, Inamine JM, Howe WR, et al.
769 Optimization of electroporation conditions for Mycobacterium avium.
770 Tuberculosis. 2002;82(4-5):167-74.
771 34. Madiraju MVVS, Qin MH, Yamamoto K, Atkinson MAL, Rajagopalan M. The
772 dnaA gene region of Mycobacterium avium and the autonomous replication
773 activities of its 5 ' and 3 ' flanking regions. Microbiol-Sgm. 1999;145:2913-
774 21.
775 35. Foleythomas EM, Whipple DL, Bermudez LE, Barletta RG. Phage Infection,
776 Transfection and Transformation of Mycobacterium-Avium Complex and
777 Mycobacterium-Paratuberculosis. Microbiol-Sgm. 1995;141:1173-81.
778 36. Moreira W, Lim JJ, Yeo SY, Ramanujulu PM, Dymock BW, Dick T. Fragment-
779 Based Whole Cell Screen Delivers Hits against M-tuberculosis and Non-
780 tuberculous Mycobacteria. Front Microbiol. 2016;7.
781 37. Yang TM, Moreira W, Nyantakyi SA, Chen H, Aziz DB, Go ML, et al.
782 Amphiphilic Indole Derivatives as Antimycobacterial Agents: Structure-
783 Activity Relationships and Membrane Targeting Properties. J Med Chem.
784 2017;60(7):2745-63.
785 38. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of
786 conditionally essential genes in mycobacteria. Proceedings of the National
787 Academy of Sciences. 2001;98(22):12712-7.
788 39. Bardarov S, Kriakov J, Carriere C, Yu S, Vaamonde C, McAdam RA, et al.
789 Conditionally replicating mycobacteriophages: A system for transposon
790 delivery to Mycobacterium tuberculosis. Proceedings of the …. 1997.
791 40. Pope WH, Ferreira CM, Jacobs-Sera D, Benjamin RC, Davis AJ, DeJong RJ, et al.
792 Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of
793 Mycobacteriophage TM4. PLoS One. 2011;6(10):e26750.
794 41. Yee M, Klinzing D, Wei JR, Gengenbacher M, Rubin EJ, Chien JY, et al. Draft
795 Genome Sequence of Mycobacterium avium 11. Genome Announc.
796 2017;5(32).
797 42. Semret M, Zhai G, Mostowy S, Cleto C, Alexander D, Cangelosi G, et al.
798 Extensive genomic polymorphism within Mycobacterium avium. J Bacteriol.
799 2004;186(18):6332-4.
800 43. Wu CW, Glasner J, Collins M, Naser S, Talaat AM. Whole-genome plasticity
801 among Mycobacterium avium subspecies: Insights from comparative
802 genomic hybridizations. J Bacteriol. 2006;188(2):711-23.
803 44. Fan XY, Xie LX, Li W, Xie JP. Prophage-like elements present in
804 Mycobacterium genomes. BMC Genomics. 2014;15.
805 45. Ovrutsky AR, Chan ED, Kartalija M, Bai XY, Jackson M, Gibbs S, et al.
806 Cooccurrence of Free-Living Amoebae and Nontuberculous Mycobacteria in

807 Hospital Water Networks, and Preferential Growth of Mycobacterium avium
808 in Acanthamoeba lenticulata. Appl Environ Microbiol. 2013;79(10):3185-92.
809 46. Guerrero C, Bernasconi C, Burki D, Bodmer T, Telenti A. A Novel Insertion
810 Element from Mycobacterium-Avium, Is1245, Is a Specific Target for
811 Analysis of Strain Relatedness. J Clin Microbiol. 1995;33(2):304-7.
812 47. Kunze ZM, Wall S, Appelberg R, Silva MT, Portaels F, Mcfadden JJ. Is901, a
813 New Member of a Widespread Class of Atypical Insertion Sequences, Is
814 Associated with Pathogenicity in Mycobacterium-Avium. Mol Microbiol.
815 1991;5(9):2265-72.
816 48. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky
817 L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res.
818 2016;44(14):6614-24.
819 49. Rabello MCD, Matsumoto CK, de Almeida LGP, Menendez MC, de Oliveira RS,
820 Silva RM, et al. First Description of Natural and Experimental Conjugation
821 between Mycobacteria Mediated by a Linear Plasmid. PLoS One. 2012;7(1).
822 50. Ummels R, Abdallah AM, Kuiper V, Aajoud A, Sparrius M, Naeem R, et al.
823 Identification of a Novel Conjugative Plasmid in Mycobacteria That Requires
824 Both Type IV and Type VII Secretion. mBio. 2014;5(5).
825 51. Krzywinska E, Bhatnagar S, Sweet L, Chatterjee D, Schorey JS.
826 Mycobacterium avium 104 deleted of the methyltransferase D gene by allelic
827 replacement lacks serotype-specific glycopeptidolipids and shows
828 attenuated virulence in mice. Mol Microbiol. 2005;56(5):1262-73.
829 52. Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ. In vivo
830 transposition of mariner-based elements in enteric bacteria and
831 mycobacteria. Proc Natl Acad Sci U S A. 1999;96(4):1645-50.
832 53. Long JE, DeJesus M, Ward D, Baker RE, Ioerger T, Sassetti CM. Identifying
833 essential genes in Mycobacterium tuberculosis by global phenotypic
834 profiling. Methods Mol Biol. 2015;1279:79-95.
835 54. DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR. TRANSIT--A
836 Software Tool for Himar1 TnSeq Analysis. PLoS Comput Biol.
837 2015;11(10):e1004401.
838 55. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al.
839 Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis
840 Genome via Saturating Transposon Mutagenesis. mBio. 2017;8(1).
841 56. Saunders BM, Cheers C. Inflammatory response following intranasal
842 infection with Mycobacterium avium complex: role of T-cell subsets and
843 gamma interferon. Infect Immun. 1995;63(6):2282-7.
844 57. Darwin KH, Nathan CF. Role for nucleotide excision repair in virulence of
845 Mycobacterium tuberculosis. Infect Immun. 2005;73(8):4581-7.
846 58. Kurtz S, McKinnon KP, Runge MS, Ting JPY, Braunstein M. The SecA2
847 secretion factor of Mycobacterium tuberculosis promotes growth in
848 macrophages and inhibits the host immune response. Infect Immun.
849 2006;74(12):6855-64.
850 59. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan
851 WT, et al. Persistence of Mycobacterium tuberculosis in macrophages and
852 mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature.
853 2000;406(6797):735-8.

60. Kar R, Nangpal P, Mathur S, Singh S, Tyagi AK. bioA mutant of Mycobacterium tuberculosis shows severe growth defect and imparts protection against tuberculosis in guinea pigs. PLoS One. 2017;12(6).
61. Puckett S, Trujillo C, Wang Z, Eoh H, Ioerger TR, Krieger I, et al. Glyoxylate detoxification is an essential function of malate synthase required for carbon assimilation in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2017;114(11):E2225-E32.
62. Mackenzie N, Alexander DC, Turenne CY, Behr MA, De Buck JM. Genomic comparison of PE and PPE genes in the Mycobacterium avium complex. J Clin Microbiol. 2009;47(4):1002-11.
63. Vandewalle K, Festjens N, Plets E, Vuylsteke M, Saeys Y, Callewaert N. Characterization of genome-wide ordered sequence-tagged Mycobacterium mutant libraries by Cartesian Pooling-Coordinate Sequencing. Nature communications. 2015;6.
64. Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: The UvrABC system. Chem Rev. 2006;106(2):233-52.
65. Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jimenez C, et al. PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Mol Microbiol. 2009;73(3):329-40.
66. Bottai D, Di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, et al. Disruption of the ESX-5 system of Mycobacterium tuberculosis causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol. 2012;83(6):1195-209.
67. Bottai D, Di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, et al. Disruption of the ESX-5 system of Mycobacterium tuberculosis causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol. 2012;83(6):1195-209.
68. Blumenthal A, Trujillo C, Ehrt S, Schnappinger D. Simultaneous analysis of multiple Mycobacterium tuberculosis knockdown mutants in vitro and in vivo. PLoS One. 2010;5(12):e15667.
69. Nagachar N, Ratledge C. Roles of trpE2, entC and entD in salicylic acid biosynthesis in Mycobacterium smegmatis. FEMS Microbiol Lett. 2010;308(2):159-65.
70. Oliveira RS, Sircili MP, Oliveira EMD, Balian SC, Ferreira-Neto JS, Leao SC. Identification of Mycobacterium avium genotypes with distinctive traits by combination of IS1245-based restriction fragment length polymorphism and restriction analysis of hsp65. J Clin Microbiol. 2003;41(1):44-9.
71. Kannan N, Lai YP, Haug M, Lilleness MK, Bakke SS, Marstad A, et al. Genetic variation/evolution and differential host responses resulting from in-patient adaptation of Mycobacterium avium. Infect Immun. 2019.
72. Soucy SM, Huang JL, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet. 2015;16(8):472-82.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
74. Ioerger TR, Feng YC, Ganesula K, Chen XH, Dobos KM, Fortune S, et al. Variation among Genome Sequences of H37Rv Strains of Mycobacterium tuberculosis from Multiple Laboratories. J Bacteriol. 2010;192(14):3645-53.

902 75. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic
903 local alignment with successive refinement (BLASR): application and theory.
904 BMC Bioinformatics. 2012;13:238.
905 76. Cosma CL, Humbert O, Ramakrishnan L. Superinfecting mycobacteria home
906 to established tuberculous granulomas. Nat Immunol. 2004;5(8):828-35.
907 77. Majumdar G, Mbau R, Singh V, Warner DF, Dragset MS, Mukherjee R.
908 Genome-Wide Transposon Mutagenesis in Mycobacterium tuberculosis and
909 Mycobacterium smegmatis. Methods Mol Biol. 2017;1498:321-35.
910 78. DeJesus MA, Ioerger TR. A Hidden Markov Model for identifying essential
911 and growth-defect regions in bacterial genomes from transposon insertion
912 sequencing data. BMC Bioinformatics. 2013;14:303.
913 79. Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, Bennett LT, et al.
914 New use of BCG for recombinant vaccines. Nature. 1991;351(6326):456-60.
915 80. Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, et al.
916 Controlling gene expression in mycobacteria with anhydrotetracycline and
917 Tet repressor. Nucleic Acids Res. 2005;33(2):e21.
918 81. Treangen TJ, Messeguer X. M-GCAT: interactively and efficiently constructing
919 large-scale multiple genome comparison frameworks in closely related
920 species. BMC Bioinformatics. 2006;7:433.
921 82. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the
922 comparison and visualization of biological lists using area-proportional Venn
923 diagrams. BMC Genomics. 2008;9.

924

925 Supporting information

926 **S1 Fig. Verification of transposon insertions.** 10 colonies were picked from a plated
927 MAH library and subjected to PCR using primers specific for the *Himar1* transposon.
928 Positive control (pos ctrl) was ϕMycoMarT7, negative control (neg ctrl) was un-
929 transduced MAH 11. Colony number 4 gave a band in the expected size when
930 resubjected to PCR.

931

932 **S2 Fig. MAH phylogenetic tree.** A phylogenetic tree containing MAH 11, MAH 104
933 and 19 other *M. avium* strains was constructed using PHYLIP. The subspecies MAH,
934 MAA, and MAP, were confirmed by their hsp60 sequence.

935

936 **S3 Fig. MAH cell wall integrity assay.** Sensitivity to SDS was assessed by spotting
937 dilutions of MAH 104 and MAH 11 (and transposon mutants) on 7H10 plates
938 prepared with or without 0.01% SDS.
939
940 **S4 Fig. Cytokine production during MAH 11 mouse infection.** Mice were infected
941 for 26 days with MAH 11 wt or MAH 11 transposon insertion mutants with and
942 without complementation. Levels of IL-1β (A), TNFα (B) and IFNγ (C) were
943 analyzed in spleen and liver homogenates and serum. Data show mean ± SEM of
944 three or four infected mice in each group. The dotted line represents the cytokine level
945 of uninfected mice.
946
947 **S5 Fig. Histopathology during MAH 11 mouse infection.** Mice were infected with
948 MAH 11 wt or MAH 11 transposon insertion mutants with and without
949 complementation. After 26 days, hematoxylin & eosin staining was performed on
950 spleen and liver sections. The panel shows representative 10x and 40x magnification
951 images of spleen and liver from one out of three or four infected mice in each group.
952
953 **S6 Fig. Mouse infection with MAH 11 eccA5::tn ESX-5 mutant.** Mice were
954 infected with MAH 11 wt or *eccA5::tn* mutant with and without complementation (4
955 mice in each group). After 26 days of infection, the bacterial burden in spleen (A) and
956 liver (B) was determined by CFUs per gram organ. (C and D) MAH-specific CD4+ T
957 cell response in mice infected for 26 days with MAH 11 wt or *eccA5::tn* mutant with
958 and without complementation. Splenocytes from all mice in each group were
959 stimulated with MAH overnight and frequencies of IFNγ- (C) and TNFα- (D)
960 producing CD4+ T helper cells were determined by flow-cytometry. Data show mean

± SEM. * P≤ 0.05; unpaired students *t* test, two-tailed, compared to wt. (E) *in vitro* growth (7H9 medium) of MAH 11 wt and eccA5::tn mutant with and without complementation. Data show mean ± SEM of three replicate samples per condition. The results are representative of three independent experiments.

S1 Table. Output statistics for the sequenced MAH libraries.

S2 Table. Verification of transposon insertion sites by Sanger sequencing.

S3 Table. Adapters used in TnSeq of the organized MAH library.

S1 Materials and Methods. Bulk-identification of transposon insertion sites in an organized MAH library.

S1 Datasets (A-O)

A – MAH 104 and MAH 11 mutual orthologs.

B – MAH 11 *in vivo* genetic requirement and MAH 11 and *Mtb* mutual orthologs.

C – MAH 11 and *Mtb* best orthologs.

D – Plasmid pMD1 annotation.

E – Plasmid pMD2 annotation.

F – pMA100 and pMD2 mutual orthologs.

G – MAH 11 *in vitro* genetic requirement.

H – pMD1 *in vitro* genetic requirement.

I – pMD2 *in vitro* genetic requirement.

J – pMD1 (spleen) *in vivo* genetic requirement.

986 K – pMD1 (liver) *in vivo* genetic requirement.

987 L – pMD2 (spleen) *in vivo* genetic requirement.

988 M – pMD2 (liver) *in vivo* genetic requirement.

989 N – Location of transposons in MAH 11 organized library.

990 O – Map of mutants/well in MAH 11 organized library.

991

992 ## Figure legends

993 **Fig 1: The experimental setup of the study.** An MAH transposon mutant library of
994 170 000 mutants (~66% coverage) was generated by selection on 7H10 agar medium.
995 This library was further subjected to mouse infection or organized mutant-by-mutant
996 in 384-well plates (24 plates). By TnSeq of the *in vitro*-selected (7H10), *in vivo*-
997 selected (C57BL/6 mice), and the organized library, the MAH essential- and virulence
998 gene sets were defined, and the location of ~3500 organized mutants were identified.
999 The output from the *in vitro* selected TnSeq'ed library was used to identify virulence
1000 genes (dashed line). Finally, a subset of virulence gene hits was validated by mouse
1001 infection experiments. Tn, transposon.

1002

1003 **Fig 2: Distribution of transposon insertion counts.** Transposon insertion counts
1004 across the MAH 11 genome (A) and the two plasmids, pMD1 (B) and pMD2 (C). The
1005 height of the black bars represents the number of insertion counts at the respective
1006 genome site.

1007

1008 **Fig 3: Synteny plot.** (A) Plot of synteny between MAH MD (top) and MAV 104
1009 (bottom), made with M-GCAT [81]. (B) The largest (>40kb) insertions and deletions
1010 in MAH 11 relative to MAH 104.

Fig 4: The MAH *in vitro* essential and *in vivo* virulence gene sets. (A) MAH 11 and MAH 104 transformation frequencies. Data show mean ± SEM of three individually electroporated samples. The result is representative of two independent experiments. **(**B) Number of MAH 11 genes defined as essential, growth defect, growth advantage and non-essential. (C) Venn diagram showing degree of overlap between MAH 11 and *Mtb* mutual orthologs (grey), MAH 11 *in vitro* essential genes as defined in this study (red), and *Mtb in vitro* essential genes (blue) as defined by DeJesus *et al*. [55]). (D) Venn diagram showing degree of overlap between MAH 11 and *Mtb* mutual orthologs (grey), MAH 11 virulence genes as defined in this study (red), and *Mtb* virulence genes (blue) as defined by Zhang *et al*. [16]. The Venn diagrams were created using Biovenn [82].

Fig 5: MAH 11 and MAH 104 mouse model infection. Bacterial burden in spleen (A) and liver (B) of mice after 22 and 50 days of infection with MAH 104 and MAH 11. Data show mean ± SEM of four infected mice in each group. (C) MAH 104 and MAH 11 *in vitro* growth (7H9 medium). Data show mean ± SEM of three replicate samples per condition. The result is representative of two independent experiments. (D) Frequencies of MAH-specific CD4+ T cells after 22 and 50 days of MAH 104 and MAH 11 infection. Splenocytes from all mice were stimulated overnight with MAH and frequencies of IFNγ-producing CD4+ effector T cells were analyzed by flow-cytometry. Data show mean ± SEM of four infected mice per group. Mice experiments are representative of two independent experiments (for 50 day time point).

Fig 6: Validation of virulence genes. Mice were infected with MAH 11 wt or MAH 11 transposon insertion mutants with and without complementation. After 26 days of infection, the bacterial burden in spleen (A) and liver (B) was determined by CFUs per gram organ. Data show mean ± SEM of three or four infected mice in each group. *, P≤ 0.05; Mann Whitney U test, one-tailed, mutant compared to wt (blue asterisk) or mutant compared to complemented mutant (red asterisk). (C and D) MAH-specific CD4+ effector T cell response in mice infected with MAH 11 wt or MAH 11 transposon insertion mutants. 26 days post infection, splenocytes from all mice in each group were stimulated with MAH overnight and frequencies of IFNγ-producing (C) and TNFα-producing (D) CD4+ T helper cells were determined by flow-cytometry. Data show mean ± SEM. * P≤ 0.05, ** P≤ 0.01, and *** P≤ 0.001; unpaired students *t* test, two-tailed, compared to wt. (E) *in vitro* growth (7H9 medium) of MAH 11 wt and transposon insertion mutants with and without complementation. Data show mean ± SEM of three replicate samples per condition. The results are representative of three independent experiments. (F) Genetic region spanning from *mav_4154-4163*. Hits from our virulence screen (mav_41xx) and previously published *Mtb* virulence screen (Rv32xx) [16] are shown in bold red.

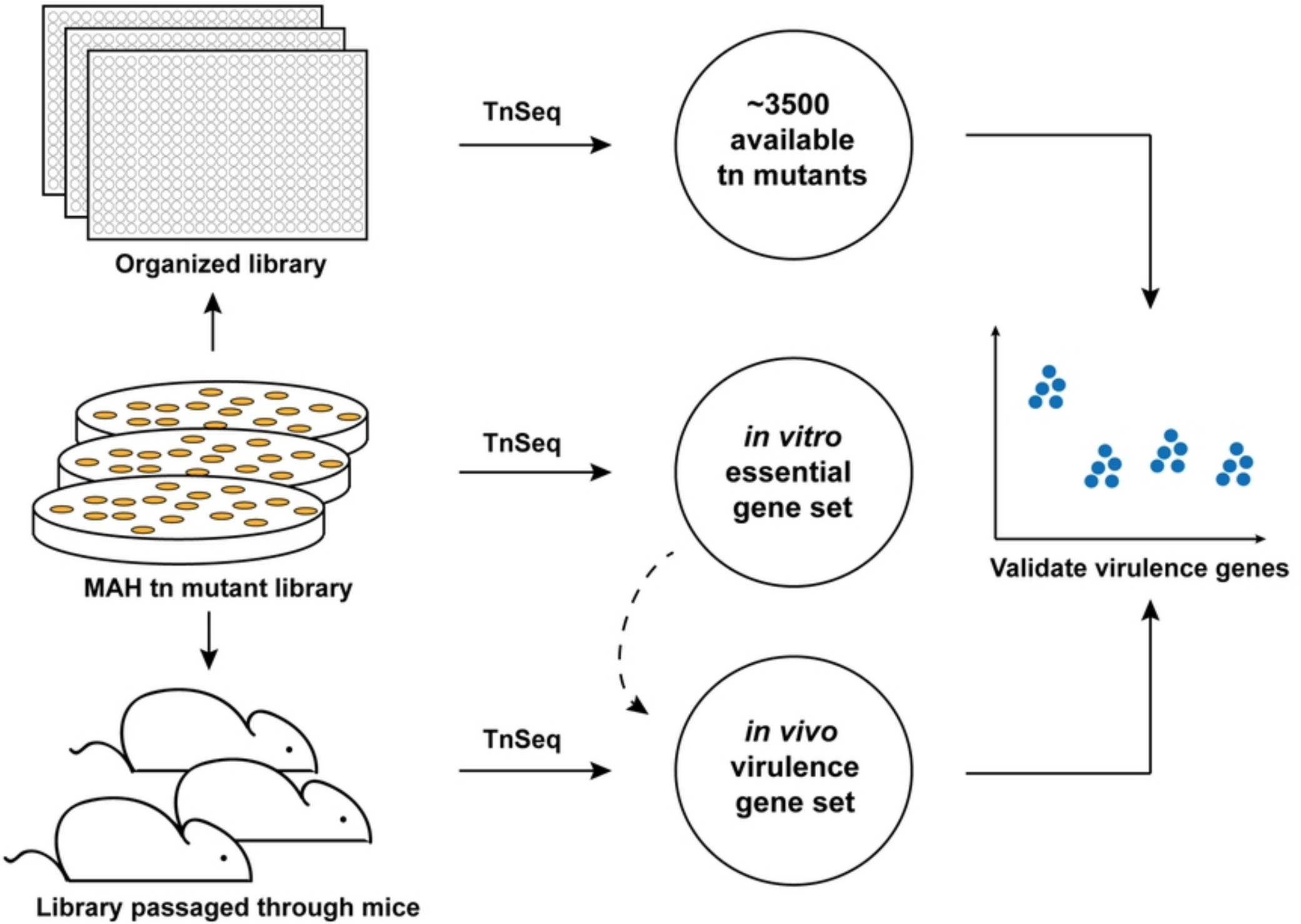
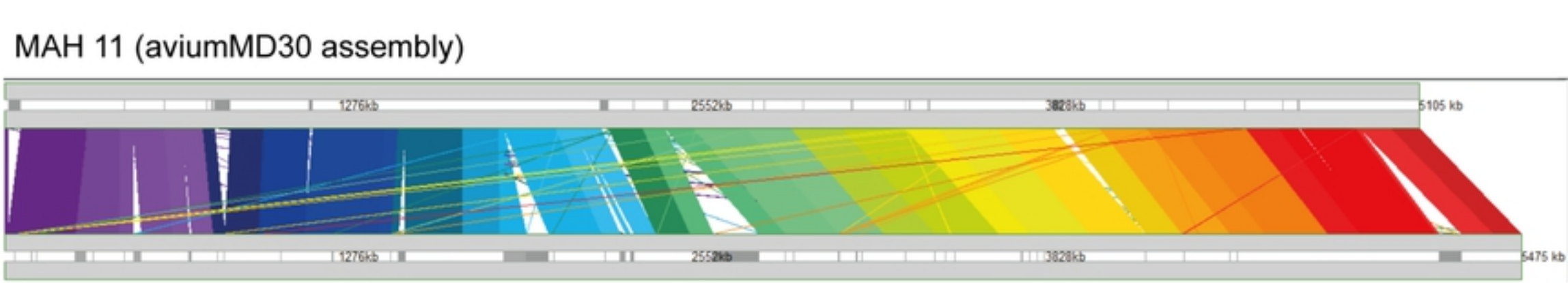

Organized library
MAH tn mutant library
Library passaged through mice
TnSeq
TnSeq
TnSeq
~3500 available tn mutants
in vitro essential gene set
in vivo virulence gene set
Validate virulence genes

Figure 1

A



B

	annotation	coordinates (in MAH 11)	size (bp)
deletions in MAH 11			
MAV_0253-MAV_0299 (46 genes)	LSP3	293034	40,100
MAV_1806-MAV_1974 (168 genes)	LSP4	1800387	162,713
MAV_2518-MAV-2691 (174 genes)	LSP1	2378550	173,202
MAV_5026-MAV_5107 (82 genes)	LSP6	4888896	81,635
insertions in MAH 11			
BK065_00050-BK065_00255 (37 genes)	hypotheticals, IS's	11154-57064	45,910
BK065_03885-BK065_04170 (48 genes)	metabolic genes	756785-814220	57,435
BK065_17725-BK065_18015 (56 genes)	new prophage	3779806-3821237	41,431

Figure 3

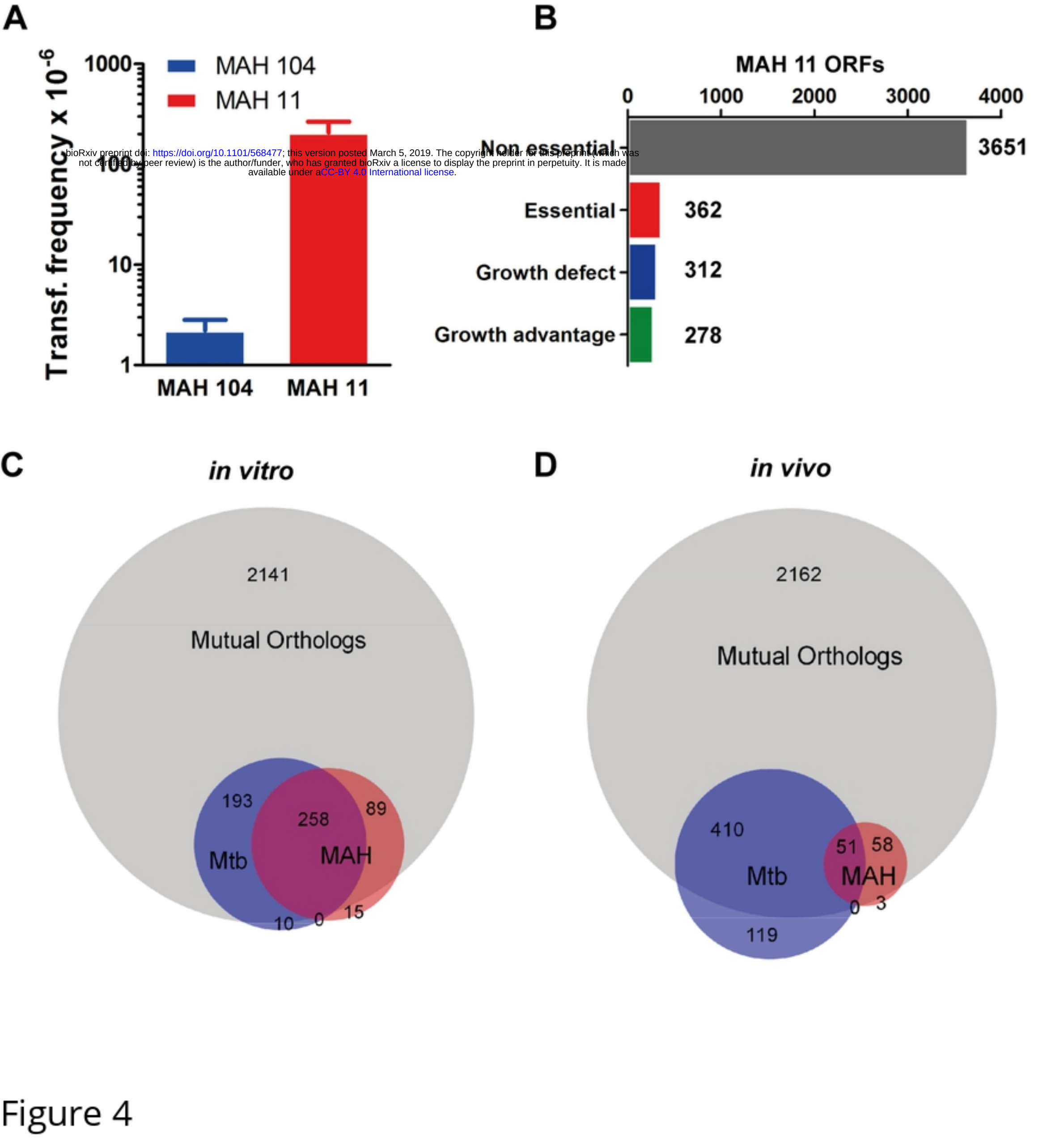


Figure 4

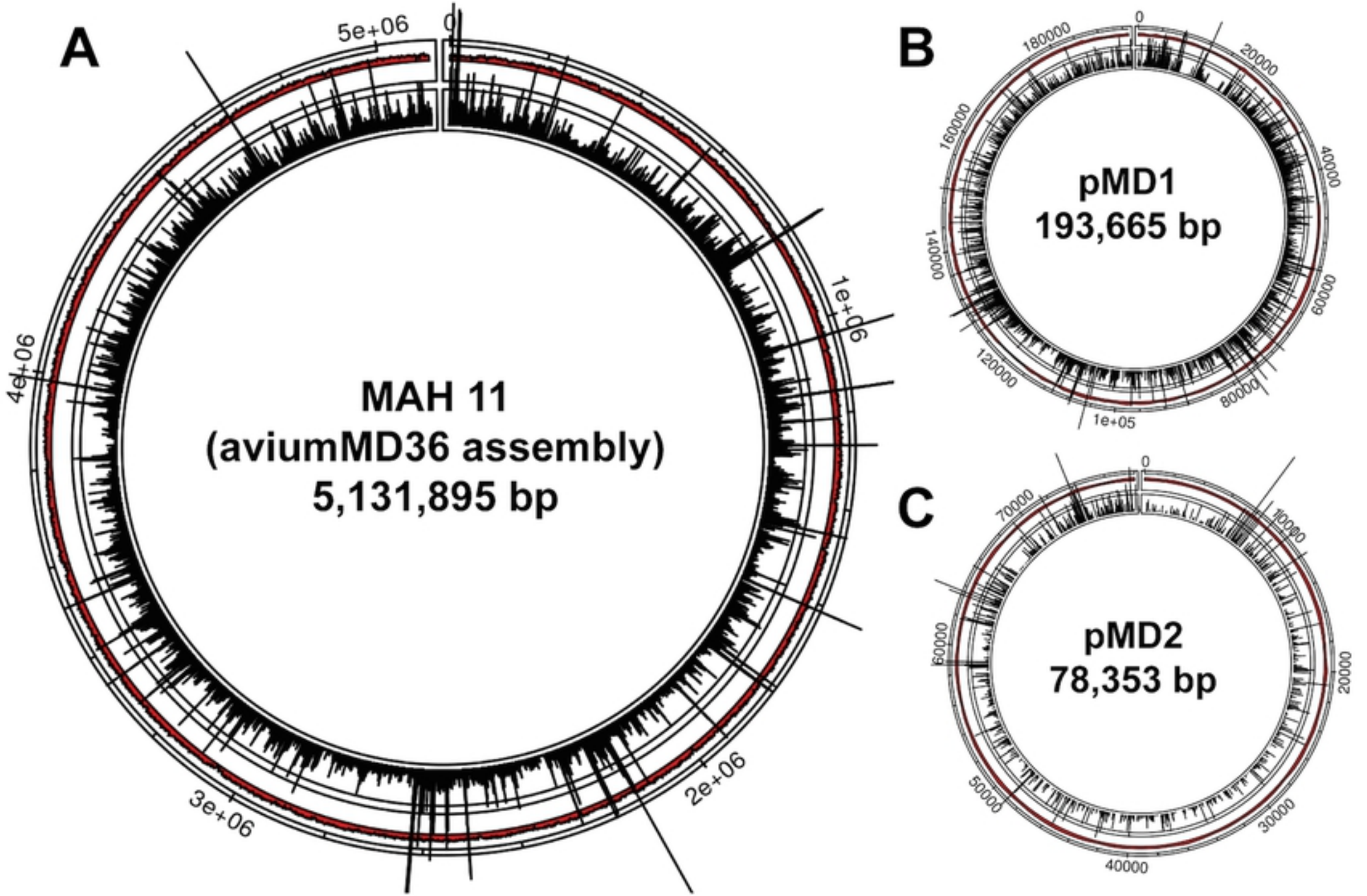

A
5e+06
0
1e+06
2e+06
3e+06
4e+06
MAH 11
(aviumMD36 assembly)
5,131,895 bp
B
0
20000
40000
60000
80000
1e+05
120000
140000
160000
180000
pMD1
193,665 bp
C
0
10000
20000
30000
40000
50000
60000
70000
pMD2
78,353 bp

Figure 2

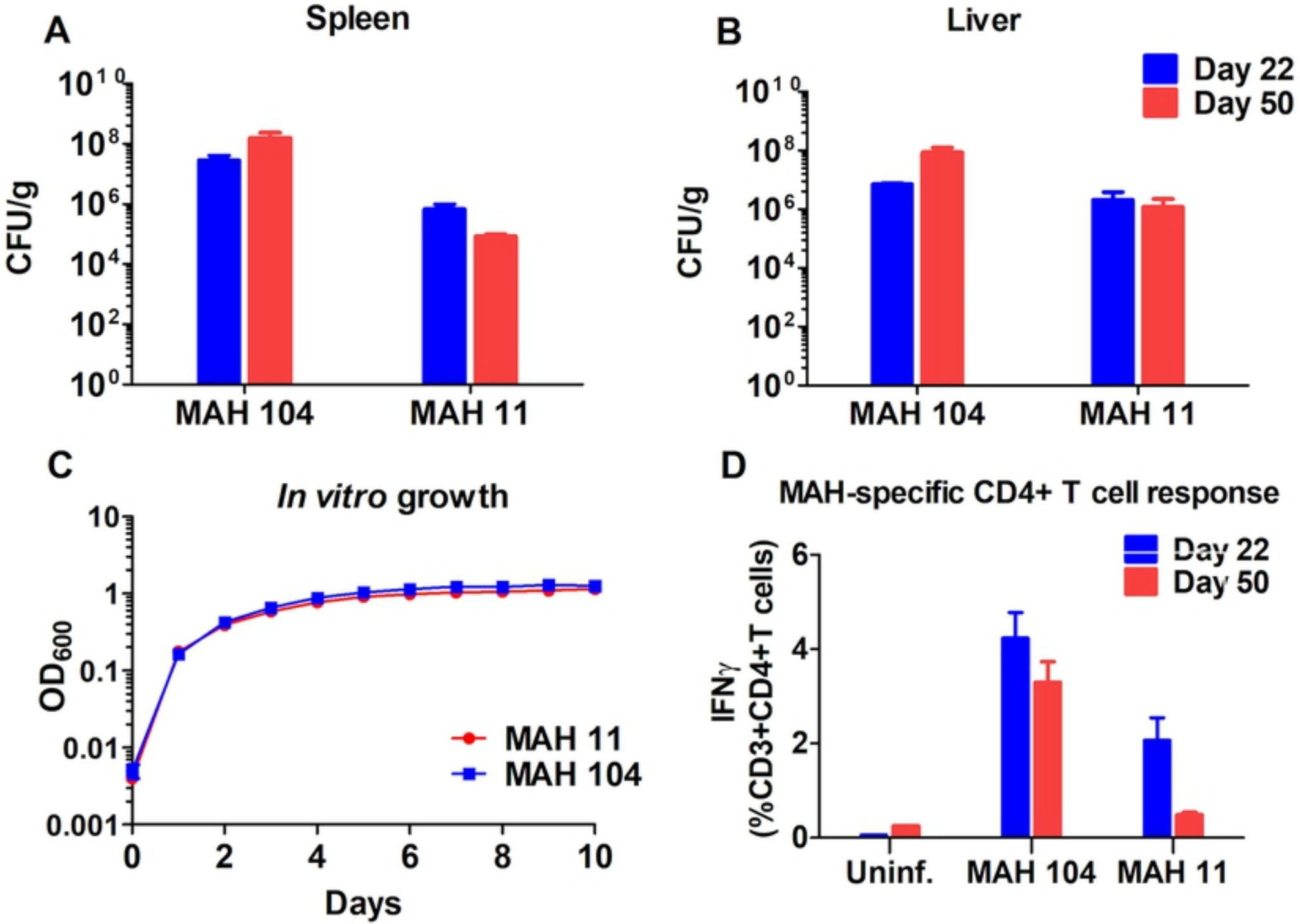

A
Spleen
CFU/g
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
MAH 104
MAH 11
B
Liver
Day 22
Day 50
CFU/g
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
MAH 104
MAH 11
C
In vitro growth
OD_{600}
10
1
0.1
0.01
0.001
0
2
4
6
8
10
Days
MAH 11
MAH 104
D
MAH-specific CD4+ T cell response
Day 22
Day 50
IFNγ
(%CD3+CD4+T cells)
6
4
2
0
Uninf.
MAH 104
MAH 11

Figure 5

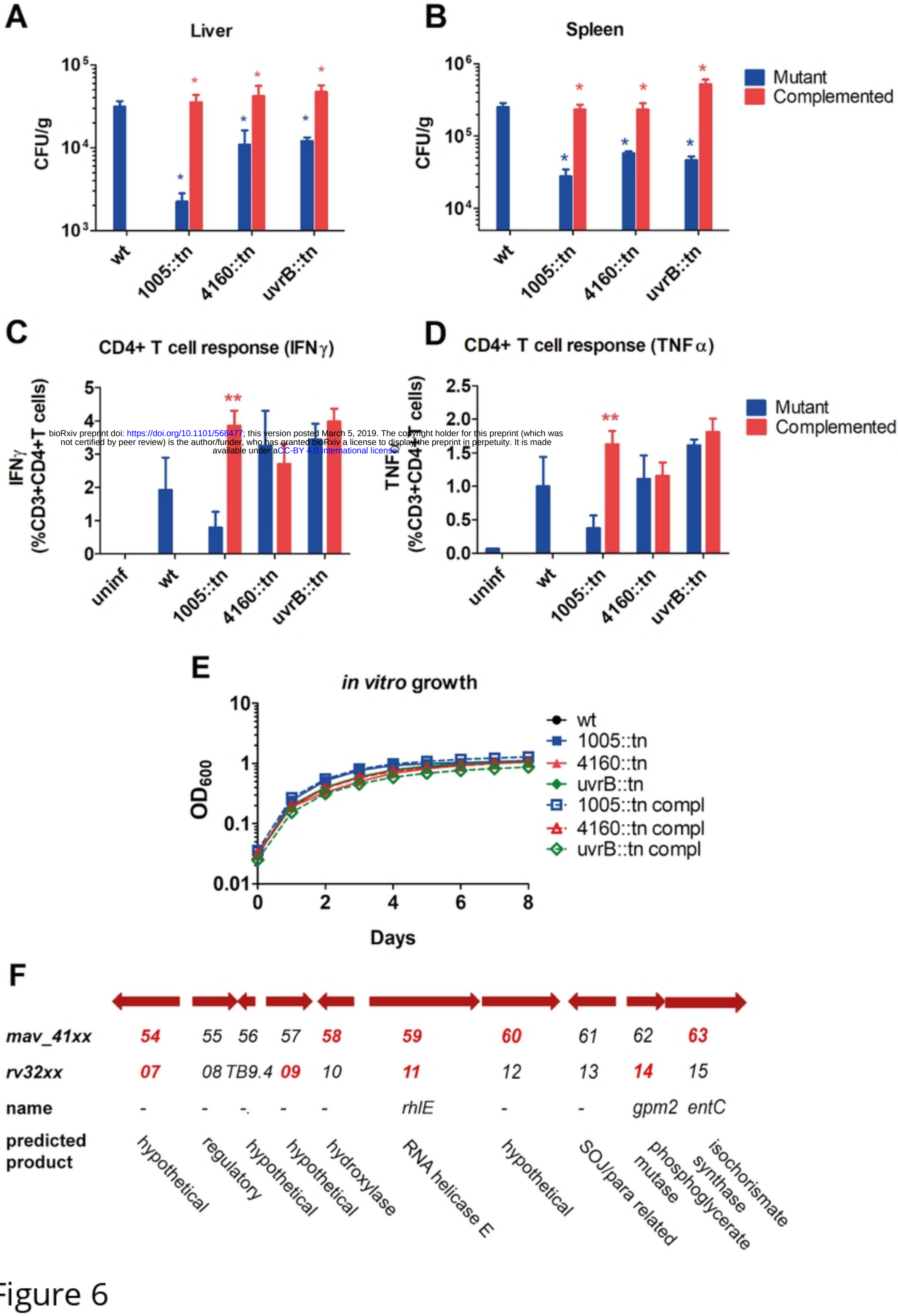


Figure 6