A genetic toolkit and gene switches to limit *Mycoplasma* growth for biosafety applications

Alicia Broto\(^1\), Erika Gaspari\(^2,3\), Samuel Miravet-Verde\(^4\), Vitor A. P. Martins dos Santos\(^2,5\) & Mark Isalan\(^1\) \(^\ast\)

*Mycoplasmas* have exceptionally streamlined genomes and are strongly adapted to their many hosts, which provide them with essential nutrients. Owing to their relative genomic simplicity, *Mycoplasmas* have been used to develop chassis for biotechnological applications. However, the dearth of robust and precise toolkits for genomic manipulation and tight regulation has hindered any substantial advance. Herein we describe the construction of a robust genetic toolkit for *M. pneumoniae*, and its successful deployment to engineer synthetic gene switches that control and limit *Mycoplasma* growth, for biosafety containment applications. We found these synthetic gene circuits to be stable and robust in the long-term, in the context of a minimal cell. With this work, we lay a foundation to develop viable and robust biosafety systems to exploit a synthetic *Mycoplasma* chassis for live attenuated vectors for therapeutic applications.

\(^1\)Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. \(^2\)Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, the Netherlands. \(^3\)European & Developing Countries Clinical Trials Partnership (EDCTP), The Hague, The Netherlands. \(^4\)Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, 08003 Barcelona, Spain. \(^5\)LifeGlimmer GmbH, Berlin, Germany. \(^\ast\)email: m.isalan@imperial.ac.uk
Mycoplasmas are perhaps best known for their reduced genome sizes, making them appealing minimal-cell models. When whole-genome sequencing began in the 1990s, Mycoplasma genitalium and Mycoplasma pneumoniae were among the first bacteria to be sequenced\textsuperscript{1,2}. Since then, Mycoplasmas have been intensively exploited in systems biology studies and whole-cell modelling projects\textsuperscript{3}. Moreover, they have been at the centre of significant milestones in synthetic biology, including the genome transplantation technique\textsuperscript{4}, the first chemically-synthesised bacterial genome\textsuperscript{5}, and the first synthetic cell\textsuperscript{6}; achievements that culminated in the design of a minimal bacterial genome\textsuperscript{6}.

Due to their genomic austerity, Mycoplasmas exhibit limited biosynthetic capabilities. Their strict nutritional needs result in a parasitic lifestyle for most Mycoplasma species\textsuperscript{7}, which has made them a major problem in animal husbandry. For instance, every year infections caused by Mycoplasmas in poultry, cows and pigs produce multimillion-euro losses worldwide. These bacteria lack cell walls and so are resistant to many of the common antibiotics\textsuperscript{8,9}. Furthermore, there are no vaccinations available against many Mycoplasma species infecting humans and animals (e.g. M. bovis)\textsuperscript{10}. In fact, there are only three effective vaccines on the market, all of them based on large-scale production of attenuated or inactivated Mycoplasma\textsuperscript{11,12}. Increasing the repertoire of such vaccines is, therefore, desirable.

Recently, efforts have been focused on engineering a Mycoplasma-based cell chassis for vaccines and other applications. One of the most advanced projects\textsuperscript{13,14} is capitalising upon the omics data generated for M. pneumoniae\textsuperscript{2,15,16}. This is being used to design a universal Mycoplasma chassis that can be deployed as a single- or multi-species vaccine in a range of animal hosts, by displaying selected surface antigens. The main idea is to use genome engineering of M. pneumoniae to remove pathogenic determinants and so develop a non-pathogenic chassis, which will also retain some interesting Mycoplasma features, such as the lack of cell walls, limited metabolic capabilities and a special genetic code\textsuperscript{33,34}. Such a cell chassis will have the potential for synthetic biology applications\textsuperscript{36} (e.g. drug delivery systems\textsuperscript{37}) as well as providing a scaffold for vaccines. However, developing an M. pneumoniae chassis as a live vector will require the implementation of biosafety measures to maximise containment of the resulting strain. In this regard, synthetic gene switches or kill-switches\textsuperscript{38} can provide genetic barriers limiting Mycoplasma growth to bioreactors during bioproduction. One of the main aims of the present work is therefore to develop such switches using transcription factors.

M. pneumoniae presents a very limited set of endogenous transcription factors: only nine in total (compared with around 400 in Escherichia coli), including only two sigma factors\textsuperscript{33}. The study of Mycoplasma biology is also limited, given the reduced amount of molecular genetic tools available. For example, while inducible gene expression systems have been widely exploited in model bacteria, like E. coli or Bacillus subtilis, that is not the case for Mycoplasma, despite their huge potential as cell chassis. With new advances in genome engineering techniques\textsuperscript{39–43}, it is more important than ever to increase the gene regulation toolkit for these microorganisms.

In terms of developing gene regulation tools for Mycoplasma, a decade ago the tetracycline-inducible promoter Ppyl/tetO\textsubscript{2} from B. subtilis was shown to be functional in the animal pathogen Mycoplasma agalactiae\textsuperscript{44}. Since that first report, this tetracycline-inducible promoter has only been exploited in two other Mycoplasma species: M. genitalium\textsuperscript{45} and M. pneumoniae\textsuperscript{39}. Nonetheless, the use of this system is still rather limited, as it is the only inducible promoter available. Increasing the amount and versatility of such tools, and porting other well-used tools from other model bacteria, will contribute to the exploitation of Mycoplasma as a synthetic biology asset.

In this study, we present the adaptation to M. pneumoniae of three well-characterised inducible systems: AraR from Bacillus subtilis\textsuperscript{46}, LacI systems from E. coli\textsuperscript{47,48} and the CI system from phage lambda\textsuperscript{49,50}. We then expanded the versatility of the TetR system from E. coli, and tested synthetic theophylline-riboswitches in M. pneumoniae. With this genetic toolkit at hand, we subsequently engineered synthetic gene switches to control and limit Mycoplasma growth for biosafety containment applications. Finally, we ascertained the long-term stability and robustness of these synthetic gene circuits in the context of a minimal cell. With this work, we provide the initial steps to developing viable biosafety systems to help exploit a synthetic Mycoplasma chassis for live attenuated vaccines\textsuperscript{51–53} or even for a live biotherapeutics (LBT) vector\textsuperscript{54–58}.

Results

A gene regulation toolkit for Mycoplasma. A choice of orthogonal gene expression systems is currently missing from the genetic toolkit available for M. pneumoniae. Until now, only one tetracycline-inducible promoter from B. subtilis has been reported to function in Mycoplasma. To overcome this limitation, we set out to engineer three well-known repressor systems to be orthogonal genetic switches in M. pneumoniae: the AraR system from B. subtilis\textsuperscript{46}, the LacI system from E. coli\textsuperscript{47,48} and the CI system from phage lambda\textsuperscript{49,50}.

First, we set out to establish the expression of the transcription factors (TFs) AraR, LacI and CI in M. pneumoniae. Genes coding for the repressor AraR and the L-arabinose permease AraE were isolated from genomic DNA from B. subtilis and expressed suitably (Supplementary Fig. 1), although later we showed that AraE is not needed for AraR to function in M. pneumoniae (Supplementary Fig. 3). By contrast, we used codon-optimised genes for the expression of LacI and CI proteins (Supplementary Fig. 1). The constitutive promoter p\textsubscript{438} from M. genitalium\textsuperscript{59} was successfully used for the expression of all these proteins, except for LacI. To reach a suitable expression level of the LacI protein, it was necessary to use a stronger synthetic promoter (p\textsubscript{S}) and to engineer the 5′-end of the gene to insert the coding sequence for a short leader-peptide to improve translation and stability of the protein in M. pneumoniae. This leader aims to reduce the secondary structure of the 5′-end of the gene, as reduced secondary structure correlates with increased expression in bacteria\textsuperscript{60} and this correlation may be even more important in a bacterium that does not depend on RBSs for translation initiation\textsuperscript{61}. As shown in Supplementary Fig. 1, all proteins were thus successfully expressed in M. pneumoniae.

The next step was to engineer promoters with operator sites for the TFs (AraR, LacI and CI). Additionally, we designed promoters with TetR-operator boxes to gain the versatility of the tetracycline-inducible system for M. pneumoniae\textsuperscript{62}. We used the promoter consensus sequence of M. pneumoniae\textsuperscript{62} in the design of promoter/operator pairs. Moreover, we took into account each TF-operator consensus sequence, the amount and position of the operator boxes inside the promoter region, the structure of the repressor binding to the DNA and the influence of DNA-bending\textsuperscript{66,48,50,62–73}. The promoter candidates that we designed were first evaluated in silico using the same algorithm developed to predict promoters in the genome of M. pneumoniae\textsuperscript{62}. We also included in the analysis a few heterologous inducible promoters extracted from the original hosts of the systems, but most of them did not meet the threshold of the test (Supplementary Data 1A). Subsequently, candidates predicted to be a promoter sequence in
M. pneumoniae were synthesised and cloned into a minitransposon (MTn) platform. The latter includes a transcriptional terminator to insulate the promoter activity and a gene encoding an mCherry reporter (Supplementary Fig. 2A). These MTns were used to generate polyclonal M. pneumoniae mutant strains. Thus, we analysed the activity of the synthetic promoters in a mCherry time-course assay (Supplementary Fig. 2). To determine the background noise in these experiments, we used a strain carrying an empty MTn. Also, we used a strain with an MTn with the constitutive promoter p438 as a positive control for the mCherry signal. All synthetic promoters tested produced mCherry expression in M. pneumoniae. Additionally, we confirmed mCherry expression in the late-exponential phase of growth by western blot (Supplementary Fig. 2B). Importantly, we achieved a wide range in the strength of the synthetic promoters (Supplementary Fig. 2). Also, the results confirmed that all predictions made in silico turned out to be functional promoters in M. pneumoniae.

To study the complete inducible systems in M. pneumoniae, we generated polyclonal strains by re-transforming the strains carrying the TFs with the different MTns expressing mCherry via the synthetic promoters. In this case, we performed the mCherry time-course assays at different concentrations of the corresponding inducer. Similar to the previous time-course assays, we used a strain without the mCherry gene as a negative control and a strain with the mCherry gene with the constitutive promoter p438 as a positive control. In the positive control, as expected, the expression of mCherry is independent of any of the TFs or inducers (Supplementary Fig. 4). By contrast, strains expressing the relevant TF alone produce little mCherry signal with the inducible promoters, suggesting a limited basal level, which is a desirable feature. The addition of the inducer to the medium reveals a clear mCherry signal, and that signal increases with increasing amounts of inducer. We confirmed the variations in mCherry expression depending on the inducer concentration in the late-exponential phase by western blot or fluorimetry (Fig. 1). In this way, we showed that all inducible systems are functional in M. pneumoniae. Induction of the reporter is dose-dependent in M. pneumoniae and all systems appear to work with promoters with very different strengths (Fig. 1).

To complete the toolkit, we evaluated several synthetic theophylline-dependent riboswitches2,4,76 in M. pneumoniae, using the MTn-reporter platform. The results show that theophylline-dependent riboswitches are functional in M. pneumoniae but with relatively high basal levels and narrow dynamic ranges (Supplementary Fig. 5). Recently, a transcriptional study in M. pneumoniae used a high-throughput screening system based on ELM-seq methodology24. It would be interesting to apply this methodology to find better riboswitches for Mycoplasma in the future.

Engineering synthetic gene circuits in Mycoplasma. We applied the tools developed above in the design of gene switches to limit Mycoplasma growth, for example for live vaccine chassis safety switches. The first case studied was a kill-switch designed to generate fatal toxicity by producing multiple double-strand breaks (DSB) on the Mycoplasma chromosome. Such damage is produced by a nuclease and, as M. pneumoniae cells lack an efficient system to repair DSB, the cells consequently die. The design of a kill-switch is usually a challenging case for circuit engineering, because of the need for absolute ON and OFF states, but it is even more difficult when working with an organism that has a promiscuous transcription and translation systems, which generate a large amount of noise19,23. We, therefore, tested several kill-switches and nucleases. Notably, the toxicity burden imposed on the system by the slightest basal expression of a toxin, a nuclease, in this case, was the main limitation for a successful kill-switch in Mycoplasma (see Supplementary Fig. 6 for examples). We only overcame this problem when using an RNA-dependent site-targeted nuclease, the Cas9 protein from Streptococcus pyogenes76. In this case, the need of two elements (the nuclease and the RNA) to generate a DSB increases the toxicity threshold, and the system tolerates a more-achievable basal expression of the toxin. Nonetheless, for a successful kill-switch, we still had to optimise the induction of Cas9. For that, we engineered a weak inducible promoter (pLG64) and, more importantly, we changed the initiation codon from ATG to a weak CTG start codon. Thus, we managed to reduce basal Cas9 expression to below detectable levels by western blot while still inducing enough Cas9 to produce toxicity in the presence of a guide-RNA (gRNA)77 (Fig. 2). This kill-switch (KS1) is an IPTG switch for Cas9 expression, combined with the constitutive expression of a gRNA that targets a repeated-sequence in the chromosome (Fig. 2). Polyclonal strains with the KS1 inserted in the genome by transposition have similar growth with or without gRNA in the construct, indicating a good silencing of cas9 before induction. Conversely, severe growth impairment (reflected in an elongated lag-phase) is only observed in the strain carrying a gRNA upon the addition of the inducer IPTG in the medium, indicating some successful induction of cell death. After this initial phase, escape mutants appear to restore growth and after the first round of IPTG treatment, the population becomes insensitive to further IPTG-induced killing (cf. later experiments below, with sequence analyses of escape mutants and the eventual implementation of designs to obviate these effects). The circuit with multiple single-target gRNAs that promote DSB in different genes (instead of a single gRNA with multiple targets) has similar performance (Supplementary Fig. 7). We also re-engineered the kill-switch in a way that the LacI represses the cas9 and two gRNAs simultaneously (Supplementary Fig. 7). Again, this design is fully inducible but performs similarly to the original KS1. Therefore, the silencing and induction of a single cas9 gene appears sufficient to reduce growth, but only for a certain period of time, and so we decided to search for additional kill-switch designs to complement KS1.

We engineered a second kill-switch (TKS2) using the same nuclease, Cas9. This time, we focused on engineering a synthetic addiction that would maximise the containment of M. pneumoniae when combined with a kill-switch. For that, theophylline (Theo, a methylxanthine chemically-related to caffeine) is a good candidate since it is effectively absent in animal tissue in vivo78. Thus, a theophylline-dependent switch could prevent unwanted growth in a vaccinated host. The TKS2 was designed as a positive feedback loop that incorporates a theophylline-riboswitch to regulate the expression of the main TF in the circuit, the TetR (Fig. 3; for intermediate optimisation steps see Supplementary Figs. 9, 10). The inhibition of TetR with tetracycline, or its analogue anhydrotetracycline (ACT), activates the circuit. Similarly, the lack of theophylline is designed to achieve similar results by decreasing TetR expression (Fig. 3). Either way, the loss of TetR triggers the expression of the nuclease Cas9, the Cre recombinase and the CI repressor, with all three elements arranged in a single operon (Fig. 3). The Cre recombinase in the circuit acts here as an indirect repressor of the TetR. To achieve that, we included a pair of lox sites flanking the gene encoding the TetR repressor (tetR). Hence, the induction of Cre eliminates TetR from the system by recombinating the pair of lox sites. Moreover, we made the activation of the circuit irreversible with a pair of mutant lox sites (lox66 and lox71). Finally, the TKS2 also includes a CI for its potential to be exploited as a repressor of a Mycoplasma conditional-survival gene. Although we successfully engineered some Mycoplasma promoters to include CI-operators (Supplementary Fig. 8), we found that these did not improve
circuit function (Supplementary Fig. 10) and so these were not included in the final TKS2 design. For now, the cl gene has a passive role in the stabilisation of the circuit before induction (Fig. 3b).

We first optimised the circuit TKS2 in the absence of a gRNA. The TetR expression, crucial to prevent leaky expression of the toxic elements, was initially insufficient even in the presence of theophylline (Fig. 3). We found we had to duplicate the riboswitch-tetR cassette in the circuit to improve its expression significantly to provide the desired silencing of the kill-switch (Fig. 3). As explained above, for a successful kill-switch, limiting the basal expression of the toxins is the most important thing. In this circuit, preventing the leaky expression of the Cre recombinase was even more important. Similarly to Cas9 induction, we managed to decrease Cre expression, below the western blot detection limit, by mutating the initiation codon to CTG, adding nine suboptimal codons among the twenty codons at the beginning of the cre gene and by avoiding putting cre in the first position of the polycistron (Supplementary Figs. 9, 10). We observed the induction of Cre by the reduction of the size of the PCR product generated when amplifying the fragment of the circuit flanked by the lox sites (Fig. 3b, c). We also confirmed the ATc-induction of the three elements (CI, Cas9 and Cre) by western blot or PCR (Fig. 3c, d). Moreover, the dilution of the theophylline that occurs when passaging the strain in a theophylline-free medium also produced increasing amounts of the three elements and resulted in the programmed loss of the TetR (Fig. 3d). Therefore, the network components appeared functional.

We next inserted the circuit TKS2 by transposition into a strain that already had the KS1 circuit with a functional gRNA. The resulting double kill-switch strain induces strong growth impairments with either IPTG or ATc. The best result is achieved when both inducers are added together to trigger both circuits simultaneously (KS1 and TKS2). In that case, no growth was observed even during 10 days of incubation, indicating a strong kill-switch effect (Supplementary Fig. 11). Transferring the strain into a theophylline-free medium by itself has a small impact on growth. Nevertheless, this is to be expected as in this case only one of the two kill-switches in the strain would be activated. Overall, the double circuit functions as a strong and efficient Mycoplasma kill-switch, in the presence of the two small-molecule inducers.

**Long-term kill-switch circuit stability and robustness in the context of a minimal cell.** The optimisation and analysis of the circuits were initially studied using polyclonal strains with the circuits inserted in the genome by transposition. Later, we isolated clonal strains and determined the location of the circuits in the genome by sequencing. We grew the clonal strains by sequential passaging every 48 h for a month (15th passages) to study the long-term stability and robustness of the biosafety systems.
As explained above, a polyclonal strain with the single biosafety circuit KS1 has severe growth impairments in the IPTG-medium, but killing is not complete as some individuals within the strain restore growth and so the population becomes immune to the kill-switch. We analysed this further with the clonal strain C5, which has a single minitransposon with the single KS1 kill-switch inserted into the mpn462 gene (encoding a hypothetical protein), classified as non-essential in *M. pneumoniae* [3]. As expected, the IPTG-induced killing produces a much greater growth impairment in the C5 strain compared to the polyclonal strain (Fig. 4, Supplementary Fig. 12, and Supplementary File 2). Nonetheless, the escape frequency (E.F.) of this strain, calculated as the ratio between the IPTG-insensitive CFUs and the total CFUs obtained in rich medium, is initially $7 \times 10^{-6}$ but increases rapidly with repeated growth, becoming $2.5 \times 10^{-4}$ at the 15th passage.

When sequencing the circuits of some IPTG-insensitive clones, we found *cas9* frameshift mutations that explain the inactivation of this single integrated circuit. Mycoplasmas have high mutation rates [8,9] and, the accumulation of mutations that inactivate the circuit is expected. However, the question is how resilient is the kill-switch to mutation and whether it is more sensitive to the IPTG concentration. Therefore, we did the whole-genome sequencing of the strain C5 at different passages (p2, p3 and p15), affecting one of the three critical residues for IPTG binding (R208 in *lacI4*). LacI4 is essential in *E. coli* and *M. pneumoniae* [17]. As expected, the kill-switch for non-induced samples presents fractions comparable to those found in genes and intergenic regions (median fraction of 1.49%, 1.55%, 1.47%, respectively). In all passages analysed, we observed that the kill-switch fractions tend to be equal or lower than in the other genomic distributions suggesting that mutations in the kill-switch are not more frequent than expected in the genome (Supplementary Fig. 13).

Focusing on the kill-switch, the analysis of the two p3IPTG replicates revealed a group of more-frequent variants present with a fraction equal to or higher than the population median (i.e., 1.5%) (Supplementary Data 3). These variants, specified in Table 1, affect either the *lacI4* or *cas9* genes, including missense (n = 16), frameshift (n = 3) and stop-gaining mutations (n = 4) (Supplementary Fig. 14). The most frequent variant is reproduducible between the two replicates. It is found in the *lacI4* gene and affects one of the three critical residues for IPTG binding (R208 in LacI4, corresponding to R197 in *E. coli* LacI4): a stop-gaining mutation in one replicate (R208*, fraction = 25.3%), which truncates the protein but preserves the DNA-binding domain, and the other a missense mutation preventing IPTG binding [47].
Fig. 3 Design and characterisation of circuit TKS2. a Schematic representation of the circuit TKS2. Blunt arrows indicate repression and pointy arrow activation. b Schematic representation of relevant DNA segments of the MTnct with TKS2 used to generate a polyclonal strain (chloramphenicol used for selection). White triangles for the inverted repeats (IR); Bend-arrows for constitutive promoters (black) and ATc-inducible (white) promoters. Arrows for genes (grey); Black semicircles show the RBSS; A white square for the riboswitch (RS); Black triangles indicate the position and orientation of the lox sites ([lox66 and lox71]; cat, selection cassette. Blue brackets indicate the segment that is duplicated in the analysis of the circuit with 2xTetR cassette. c DNA gel showing the results from the PCR analysis of Cre/lox recombination in the TKS2 from genomic DNA samples from passages 1 and 2 in theophylline medium with Cre basal expression (p1+ and p2+) and passage 2 in theophylline-free medium and Cre induction (p2–). PCR product size is expected for the full fragment, before Cre/lox recombination, and the distance between lox sites are indicated in (b) between the small head arrows S5 and S17, showing the position of the primers pair used in the assay (in blue for 2xTetR). d Western blot results showing TetR, CI and Cas9 expression in strains with the circuit TKS2. Expression of the proteins was analysed using 10 µg of total protein extracts from incubations of the strains in the growth permissive condition (0.5 mM theophylline (+); after a single passage in theophylline-free medium producing slow activation of the TKS2 (−); after two passages in theophylline-free medium (−) or in 50 ng/mL of ATc to trigger the kill-switch (ATc). Results presented from the circuit engineered with one or two copies of the p5ori-riboswitch-tetR cassette between the lox pair sites in the circuit (1xTetR or 2xTerR, respectively). DNA gel and western blot results are representative of two independent experiments in each case.

(R208C; fraction = 20.9%). The following most frequent variants correspond to frameshift mutations in the cas980 gene truncating the protein and abolishing cleavage activity80 and missense mutations in the lacI4 gene affecting the hydrophobic surface of the IPTG-binding pocket47 (Table 1).

It should also be noted that we find a wide range of scarce variants in all samples, most of them represented by a minimal number of reads and sometimes just one (Supplementary Data 2 and Supplementary Fig. 14). However, considering the reproducibility of the most frequent variants and their expected impact in abolishing the killing activity of the circuit, they alone could explain the population that survived and became insensitive to the IPTG stress.

Taken together, these results indicate that we managed to silence the basal levels of the kill-switch very effectively and, consequently, it is not producing a significant selection pressure in the strain without chemical induction, which is a remarkable achievement. After induction, we still find a limited group of variants that affect crucial elements of the kill-switch mechanism and together explain the gain of resistance to the kill-switch. Given that the accumulation of mutations is an intrinsic property of the strain, the only way to increase the resilience of the kill-switch is to introduce redundancy in the system, as follows.

Effective redundancy in the system was first attempted by including multiple copies of the whole circuit KS1. The clone C20, derived from the parental strain C5, has a second minitransposon with the circuit KS181 inserted in the intergenic region between mmp648 and mmp649 genes. This strain, with two copies of the kill-switch, shows no growth during a period of 10 days when treated with IPTG, indicating a strong kill-switch effect (Fig. 4 and Supplementary Fig. 12). The escape frequency for this strain is below the detection limit (3 × 10−9), as no colonies were found in plates with IPTG. We also performed a variant calling analysis from the whole-genome sequencing results of the strain C20, at early passages p2 and p3, and the late passage p15. For C20 samples, the frequency of variants per base decreased compared to C5 (0.9% on average), remaining significantly lower when compared to genes and intergenic regions (Mann–Whitney U one-tailed test, p value <0.05 in every comparison). We evaluated whether this lower frequency could be an effect of the shared sequences between the two copies of the kill-switch (KS1 and KS181) but the total coverage per base within the kill-switches for C20 samples was in fact higher than in C5 samples (Supplementary Fig. 13). As in the previous case C5, in the absence of any pressure, mutations appear to be accumulated evenly across the whole chromosome and so the chances of inactivating the kill-switch are fairly low. In this case, it was not possible to sequence samples after IPTG-stress because no sample could be produced with this very strong kill-switch. By applying regression models to the parameters of the growth curves, we estimated the arise of escape mutants over strain passages, so to predict the number of passages that M. pneumoniae would take to inactivate the kill-switches in the strains C5 and C20. The simulation shows that the clone C5 is predicted to inactivate the circuit after 20 strain passages, whereas clone C20 does so after

(2022)13:9130 | https://doi.org/10.1038/s41467-022-29574-0 | www.nature.com/naturecommunications
30 strain passages (Supplementary Fig. 15). The two predicted growth curves are compared to data for passages p2, p7 and p15 of clone C5 and for passages p2 and p15 of clone C20. Such high passages could be easily avoided during industrial production of vectors by keeping low-passage stocks and resequencing single clones if needed, as part of quality control. We were also unable to insert a third copy of the KS1 in the cells, even after several attempts at transformation. The third copy likely increases the basal expression of Cas9 to a point that is too toxic for the cells. This suggests that cells cannot overcome the system with three copies of the kill-switch.

Another potentially interesting biosafety system with some inbuilt redundancy is the combination of circuits KS1 and TKS2, which includes theophylline biocontainment. We studied the long-term stability of this system with the clone C26. This clonal strain, also derived from the parental C5, has a second minitransposon with TKS2 inserted into the mpn062 gene (purine nucleoside phosphorylase, deoD). This gene was classified as non-essential at least in one of the essentiality studies in *M. pneumoniae* 81. This was confirmed by simulating in silico the disruption of each gene comprised in the most updated metabolic model of *M. pneumoniae* (iEG158_mpn) 82, which showed that the disruption of mpn062 would decrease growth by only 1.2% (Supplementary Data 4). The C26 clone grows slightly slower compared to strains C5 and C20, which might be due to the disruption of this deoD gene. As expected for a strong double kill-switch, no growth was observed during 10 days of treatment with IPTG and ATc (Fig. 4). The escape frequency for clone C26, when both circuits are activated, is below the detection limit, as no colonies were found in plates with IPTG and ATc. However, the escape frequency of this clone C26 in ATc alone is higher, suggesting that circuit TKS2 is leakier compared to KS1 and produces a small selection pressure on the strain before proper circuit activation. Whole-genome sequencing results of samples from early and late passages of C26 (passages p2 and p15 in the presence of theophylline) revealed no significant differences in the sequences of the circuits in terms of frequency of variants per base and fraction of alternative read counts, compared to C5 (Mann–Whitney U one-tailed test, p value <0.05 in both p2 and p15 comparisons; Supplementary Fig. 13). In summary, the clonal double kill-switches developed here are capable of stably blocking *Mycoplasma* growth even after 15 rounds of prior passaging.

![Fig. 4 Study of the long-term performance of the three biosafety systems in clonal *M. pneumoniae* strains](https://example.com/fig4.png)

**Fig. 4 Study of the long-term performance of the three biosafety systems in clonal *M. pneumoniae* strains.** a Growth kinetics (growth index corresponding to the ratio Abs430 nm/Abs560 nm) of the clonal strains C5 (I), C20 (II) and C26 (III) analysed at an early (p2) and late passage (p15) at different induction conditions. We compare the growth of strains C5 and C20 (with one and two copies of the kill-switch KS1, respectively) when uninduced at the early passage p2 (0 mM IPTG, blue lines) or IPTG-induced that triggers the kill-switch at the passage 2 (1 mM IPTG; p15, orange line). We analysed the growth of strain C26 (with KS1 and TKS2) in theophylline-medium (0.5 mM Theophylline, permissive growth conditions with both circuits inactive, blue line), or theophylline-free medium (0 mM Theophylline, slow activation of TKS2, light brown line) in the early passage p2. The growth is also analysed when both circuits are activated simultaneously in ATc and IPTG medium at passage 2 (1 mM IPTG and 50 ng/mL ATc; p15, orange line). The black line is a negative control of the growth index reads from an empty medium and shows the baseline of these long experiments, as the Abs560 reads have a drift (SP4, black line). Mean values from three bio-replicates. Error bars indicate standard deviation. b Escape frequency (EF) for strain C5 at several strain passages, determined as the ratio between IPTG-resistant CFUs and the total CFUs. The dashed line shows the detection limit calculated in the analysis, considering the total CFUs plated. Source data are provided as a Source Data file.
Table 1 | List of the significant variants found after the IPTG-induction of K51 in strain C5 at passage three (sample p3IPTG).

| POS   | QUAL   | TOT | RFN   | ALTN | FRAC | REF     | ALT     | IMPACT | AFF   | MUT   | EFF    | INFO                                                                 |
|-------|--------|-----|-------|------|------|---------|---------|--------|-------|-------|--------|----------------------------------------------------------------------|
| 56603 | 1.96E-13 | 559 | 446   | 113  | 25.3 | CGCAAA  | TGCTAT  | HIGH   | LacI  | LR208* | stop gained | Expression of Nter fragment of LacI; Intact DNA-binding domain     |
| 569125| 0.00E+00 | 534 | 463   | 71   | 15.3 | ATCAAAC | GTCAAAC | HIGH   | cas98 | K896fs | frameshift variant | Deletion of RuvCII domain (affecting the Nuc lobe) and PAM-Interacting (PI) domain. The deletions of PI domain abolish cleavage activity |
| 56603 | 0.00E+00 | 271 | 224   | 47   | 21.0 | CGCAAA  | CACAT   | MODERATE | LacI  | A87V  | missense variant | Mutation of a critical residue in the LacI protein for IPTG binding. It does not affect LacI structure |
| 566149 | 1.76E-13 | 609 | 571   | 36   | 6.7  | A       | G       | MODERATE | LacI  | HAP86QER | missense variant | Mutation that affects the hydrophobic surface of the IPTG binding pocket in the LacI protein |
| 566164 | 3.47E-14 | 281 | 255   | 26   | 10.2 | GTTCGT  | CGTCTT  | MODERATE | LacI  | P897fs | frameshift variant | Expression of truncated Cas9; Deletion of RuvCII domain (affecting the Nuc lobe) and PAM-Interacting (PI) domain. The deletions of PI domain abolish cleavage activity |
| 566127 | 8.08E-15 | 235 | 210   | 25   | 11.9 | CACAT   | CAACAT  | HIGH   | cas98 | P897fs | frameshift variant | Loss of reading frame that prevents expression of Cas9. The mutation affects the protein from its first domain RuvCI |
| 569181 | 0.00E+00 | 481 | 469   | 12   | 2.6  | ATTTTTTGATA | ATTTTTTGATA | HIGH | cas98 | N466fs | frameshift variant |
| 566420 | 0.00E+00 | 607 | 596   | 11   | 1.8  | C       | A       | MODERATE | LacI  | A87S  | missense variant | Mutation that affects the hydrophobic surface of the IPTG binding pocket in the LacI protein |
| 566423 | 9.10E-15 | 291 | 284   | 7    | 2.5  | G       | A       | MODERATE | LacI  | H86Y  | missense variant | Mutation that affects the hydrophobic surface of the IPTG binding pocket in the LacI protein |
| 568288 | 9.10E-15 | 231 | 225   | 6    | 2.7  | ATTTTTACAAAGA | ATTTTTTCTATGGT | HIGH | cas98 | F1737TI | missense variant | Mutation in the PAM-recognition domain of the Cas9 protein; Deletion of the PI domain abolish Cas9 cleavage activity |
| 565983 | 2.26E-16 | 299 | 293   | 6    | 2.0  | C       | A       | MODERATE | LacI  | W232C | missense variant | Mutation that affects the hydrophobic surface of the IPTG binding pocket in the LacI protein |
| 569609 | 0.00E+00 | 206 | 201   | 5    | 2.5  | ATACCTTTATTAAG | TTACTTTTTATAG | MODERATE | cas98 | G1736SN | missense variant |
| 569662 | 1.25E-15 | 207 | 202   | 5    | 2.5  | ACTATCG | TCTTTG  | MODERATE | cas98 | D5718R* | missense variant | Mutation in the RuvCII domain of Cas9 |
| 569529 | 6.21E-15 | 232 | 227   | 5    | 2.2  | AAA     | AAGATCCCGCATCTTTAA | HIGH | cas98 | L1315* | stop gained | Expression of Nter fragment of LacI; Intact DNA-binding domain |
| 570703 | 0.00E+00 | 237 | 232   | 5    | 2.2  | AAA     | GAATCCGCTTCCATCTTTAC | HIGH | cas98 | F327I | missense variant | Mutation in the domain RCI of Cas9 |
| 566036 | 0.00E+00 | 243 | 238   | 5    | 2.1  | TAAAGGGTGGCCATCTTAGTCTGT | HIGH | cas98 | RL109* | stop gained | Expression of Nter fragment of LacI; Intact DNA-binding domain |
| 571004 | 0.00E+00 | 193 | 189   | 4    | 2.1  | CAAATAAGCCA | CTATTAAGCCA | MODERATE | cas98 | F238I | missense variant | Mutation in the domain REC2 of Cas9 |
| 567098 | 0.00E+00 | 246 | 242   | 4    | 2.2  | CCACCCAGAG | TGCCTCAGGT | MODERATE | cas98 | L255SH | missense variant | Mutation next to a Hotspot point mut |
| 568065 | 0.00E+00 | 200 | 197   | 3    | 1.5  | TATCCTT | CATCAT   | MODERATE | cas98 | EDN1250DD | missense variant | Mutation in the P1 domain of Cas9 |
| 568073 | 3.58E-15 | 200 | 197   | 3    | 1.5  | TAAAAG | CAAACCTT | MODERATE | cas98 | FFY137LYC | missense variant | Mutation in the RuvCII domain of Cas9 |

Columns correspond to genome loci (base pair position in the C5 genome; POS); Estimate of the probability of a polymorphism at the loci described by the record (QUAL). This value is presented in phred scale and takes into consideration both mapping and genotype quality for a specific variant; total number of reads obtained for the loci (TOT); Total number of reads matching the reference (REFN); number of reads presenting a variant (ALTN); frequency of the variant normalised by the total reads found mapping that specific loci (FRAC); Reference sequence in the genome (REF); variant sequence (ALT); Degree of impact of the variant (IMPACT), which is classified as "LOW" (synonymous mutations), "MODERATE" (missense), "HIGH" (nonsynonymous mutations, start or stop loss), or MODIFIER (variant found in an intergenic region); gene affected (AFF, in case the mutation is found intergenic, it is reported the closest downstream gene); Mutation or variant (MUT, an asterisk indicating stop codon); potential effect of the variant as given by snpeff102 (EFF); and, in last column (INFO), short explanation of the expected effect of the variant in the mechanism of the kill-switch based on the literature (see references 48 and 81). The full list of variants is provided in Supplementary Data 3.
These properties meet the challenging requirements of kill-switches and add to the utility of *M. pneumoniae* as a potential synthetic biology and vaccine chassis.

**Discussion**

For the first time, there is a set of inducible systems for *Mycoplasma*. This work presented the adaptation to *M. pneumoniae* of well-characterised TF-repressor systems from gram-positive and gram-negative bacteria and phage. The expression of the different TFs does not appear to be toxic. More importantly, we showed it is possible to engineer synthetic inducible promoters for *Mycoplasma*. The promoter design is the essential step to make the inducible systems functional, especially considering that most heterologous inducible promoters did not meet the threshold in the test for *Mycoplasma* promoters. Altogether, the systems presented here represent the first gene regulation toolkit for *M. pneumoniae* and will lead the way to expand such tools for *Mycoplasma*. In the future, also, it will be interesting to explore the activity of TF-activators in *Mycoplasma*.

In this study, we used the gene regulation toolkit to engineer gene switches to limit *M. pneumoniae* growth. We presented two kill-switches, KS1 and TK2S, with different complexities but both based on the toxicity that an RNA-dependent nucleic produces in a cell incapable of repairing genomic DSB. The most challenging step in engineering such switches was to overcome the basal toxicity that would compromise cell viability. To this end, we optimised the OFF state of the kill-switch. This allowed us to achieve a good balance between undetectable basal expression and high toxicity once the kill-switch is induced. Remarkably, we successfully included a synthetic Theo-riboswitch in the design of the more complex circuit TK2S, despite the narrow dynamic range that this RNA elements provide. The inclusion of this RNA regulation element created a synthetic Theo-addiction which can be exploited for *Mycoplasma* biomass containment.

This is the first time that circuit engineering has been applied in a minimal cell model. In this case, the genome-reduced bacterium *M. pneumoniae* exhibits a transcriptional regulation dominated by TF-independent mechanisms and with an important portion of unexplained variation associated with genetic stochasticity. Therefore, it was a challenge to engineer orthogonal circuits in *Mycoplasma* based on canonical TF-repressors and riboswitches. We succeeded in engineering tight kill-switches based on CRISPR/Cas9 with reasonable long-term stability, despite the high mutation rate of *Mycoplasma* and their limited DNA repair machineries. The durability of the biosafety circuits is a relevant concern for bioproduction, engineered systems will always be subject to the evolutionary pressure of the *Mycoplasma*, but, remarkably, we found that the kill-switches had been well silenced and do not increase that pressure, as the accumulation of mutations in the kill-switches is comparable to the rate found for the genome. Notably, we showed that a small built-in redundancy, like the duplication or combination of two kill-switches, is sufficient to decrease the escape frequency below the NIH recommended threshold for GMOs (>10<sup>-8</sup>). This significantly increases the long-term stability of the biosafety circuits, which is predicted to reach up to 30 passages in the case of the double KS1 included in strain C20. This would be more than adequate for industrial bioproduction. In the future, it will be necessary to study the compatibility of the biosafety circuits with the expression of heterologous bioproduction elements included in the products.

The gene regulation toolkit for *Mycoplasma* opens a new horizon. Not only will it contribute to better tools to improve the understanding of *Mycoplasma* essential genes with unknown function(s), but, more importantly, together with the new genome engineering tools, it will qualify *Mycoplasma* as a true synthetic biology asset.

Furthermore, the *Mycoplasma* biosafety circuits show successful efficacy and robustness in vitro. In a follow-up, it will be useful to corroborate these results in the context of large-scale production, and more importantly, in vivo, where an immune system will have a role to play, too.

Recent advances in *Mycoplasma* research have led to designing different chassis with potential as live attenuated vectors, whether for vaccination or drug delivery systems. For the aim of engineering live biotherapeutics for livestock or humans, it will be essential to ensure the biosafety of the products, during bioproduction as well. In this regard, we have presented two possible strategies to enhance the biocontainment of a MycoChassis: kill-switches that can be triggered upon demand and a synthetic addiction. Such genetic safeguards, which ensure confinement to controlled growth conditions, will contribute to gaining public confidence to bring the products to the market.

**Methods**

**DNA cloning and construct construction.** Subcloning was performed using restriction enzyme sites, Gibson assembly<sup>91</sup>, or reverse PCR and KLD reaction (NER M0554). All constructs derive from an MTn vector, which could be either an MTTpar<sup>92</sup> or MTTnacr<sup>93</sup>. Designed promoter/operator pairs were evaluated in silico using the algorithm developed to predict promoters in the genome of *M. pneumoniae*<sup>39</sup>. The MTn-regulator platform used to test promoter candidates or RSs, is an MTnpar derivative with a transcriptional terminator and a gene encoding a mCherry reporter<sup>94</sup> (Uniprot no. X5DSL3), with a 5’ end engineered with a short sequence coding for the leader peptide mp2007<sup>95</sup> to facilitate translation (Supplementary Fig. 2A). We constructed two versions of the platform, one with the terminator TrrnMPN (A) and the other with the T142 (B). We inserted the promoter p438 and an RBS site upstream of the gene encoding the reporter to obtain the MTn used in the positive control of mCherry expression in *Mycoplasma*. Then, we generated a collection of MTn-reporter vectors changing the p438 sequence by the different promoter candidates using Gibson cloning. For the test of the RSs, the RBS region was substituted by the different RS sequences tested.

**Genes coding for the AraE repressor** (Uniprot no. P96711) and the AraE l-arabinose permease (Uniprot no. A0A6M3ZGR6) were isolated from *B. subtilis* strain 168 genomic DNA (ATCC 23857D-5). *Mycoplasma* codon-optimised genes purchased from GenScript were used for the optimisation of the expression of LacI (Uniprot no. P30323) and CI (Uniprot no. P30304) repressor proteins and also, for the nuclease ApoRI (or Asel and Sppl) and Cas9 (Uniprot no. Q99ZW2) used in the kill-switches. The genes coding for TetR repressor (Uniprot no. B1VF0) and Cre recombinase (Uniprot no. P06936) proteins were extracted from the plasmid pBBR1MCS (kindly provided by Dr. J. Pinyol (IBB, UAB, Spain). Site-directed mutagenesis was used to generate the gene variants of cas9 (B) and cre (B and C) genes. Promoters, ribosomal binding sites (RBSs), terminators and the gRNAs were ordered as primers (Sigma). PCR primers were also used to insert a short sequence coding for a leader peptide at the 3’ end of the genes or a 3’ end tag for protein detection.

The synthetic gene switches KS1 and derivatives were cloned into an MTTpar vector. An MTTnacr vector was used to clone the circuits TK2S and KS1<sup>92</sup>, to generate strains C26 and C20, respectively.

All MTn constructs were confirmed by Sanger sequencing. All DNA sequences (primers, DNA parts, tools, and circuits) are listed in Supplementary Data 1B–F and GenBank files of the main plasmids are provided as supplementary data (all files are included in the supplementary file GenBank_MTns.zip).

**Bacterial strains and culture conditions.** Standard DNA cloning was performed with chemically competent cells TOP10 (Invitrogen), XL-1 Blue (Agilent), or DH5α F’Iq (NEB). All constructs and circuits designed for *M. pneumoniae* were assembled in an E. coli host. This step presents a bottleneck that impeded the generation of some constructs and their test in *M. pneumoniae*. To overcome this limitation, we exploited the use of codon TGA coding for W in *Mycoplasma* in relevant ORF sequences. We also used E. coli strains with constitutive expression of LacI or TetR to prevent unwanted expression from the synthetic IPTG- or Atc-inducible promoters included in some constructs.

All *M. pneumoniae* strains generated in this work are derived from M129 wild-type strain (ATCC 29342) and were cultivated in SP-4 medium, at 37 °C and 5% CO₂. Puromycin (3.3 μg/mL), chloramphenicol (20 μg/mL), l-arabinose (0.2%) or 0.5%, and hydrotetracycline (ATc; 10 ng/mL or 30 ng/mL), IPTG (0.4 or 1 mM) or Theophylline (Thee; 0.5 mM) were added when appropriate. Polymorphic *M. pneumoniae* strains were generated by electroporation using 1 μg of MTn vectors as previously described<sup>39</sup>. Briefly, *M. pneumoniae* cells were grown in SP4 medium in...
T25 (or T75) flasks at 37 °C and 5% CO2 to late-exponential phase. Then, cells were washed three times using electrophoresis buffer (272 mM Sucrose, 8 mM HEPES, pH 7.4) to disrupt and to digest the DNA (G25-Spin kit) in a final 1 mL volume of electrophoresis buffer and cell clumps digested passing the cell suspension through a 25 G needle ten times. For electrophoresis, 50 µL of the cell suspension was transferred to a 1 mm electrophoresis cuvette with 30 µL of electrophoresis buffer and 1 µg of the relevant Mn vector. The cells and DNA mix were kept on ice for 15 min before the electric shock (500 V, 2000 µF). After 1000 ohms, if no electric current was transferred back on the ice for another 15 min before the cells were transferred into an Eppendorf with 500 µL fresh SP4 for a 2 h recovery time at 37 °C and 5% CO2. Finally, a 1/10 dilution of the transformation was grown in SP4 with the relevant antibiosis (antibiotics and inducers) in T25 flasks at 37 °C and 5% CO2, for a week when the transformed cells were recovered in 1 mL of SP4. Given the Mn vectors are randomly inserted in the mycoplasma genome, a pool of transformant cells with the pertinent Mn is recovered in SP-4 with selection antibiotic (puromycin when using a Mnpar vector or chloramphenicol in the case of Mntacat) and theophylline when necessary. For clonal strains, after electrophoresis, individual clones were isolated in SP-4 plates with 0.8% agar and supplemented with the relevant antibiosis(s) and theophylline when pertinent. Specifically, M. pneumoniae clonal strains C5, C20 and C26 were isolated from SP-4 agar plates supplemented with either: puromycin; puromycin and chloramphenicol; or puromycin, chloramphenicol and theophylline. The location of the Mn in each clone was determined by A-PCR as previously described19. Briefly, genomic DNA from the clonal strain with an Mn unknown insertion is used in a low stringency PCR with the primers S78, with homology to a short fragment inside the MTn, and primer S79 which is a primer with a 5’ unique and the 3’ randomised sequence NNNNNNGATTA ending in a highly repeated-sequence in the M. pneumoniae genome. Then, PCR products generated were purified as a pool and used as the template in a second PCR to isolate a single fragment with primers S78 and S80. This final fragment is sequenced to determine the insertion site of the MnTn. The MnTn location was then confirmed by PCR and sanger sequencing. For long-term studies, M. pneumoniae clonal strains C5, C20 and C26 were serially passaged every 48 h for 15 passages in SP-4 medium supplemented with puromycin, chloramphenicol, and theophylline, respectively. All strains used in this work are listed in Supplementary Data 1G.

Western blot. For protein detection, total protein extracts were prepared from late-exponential phase cultures of M. pneumoniae strains for SDS-PAGE following standard procedures. Briefly, cells grown in 25 cm² flasks were washed three times with PBS and extracted in SP-4 plates with 0.8% agar and supplemented with the relevant antibiosis(s) and theophylline when pertinent. M. pneumoniae clonal strains C5, C20 and C26 were isolated from SP-4 agar plates supplemented with puromycin, puromycin and chloramphenicol, or puromycin, chloramphenicol and theophylline. The location of the Mn in each clone was determined by A-PCR as previously described19. Briefly, genomic DNA from the clonal strain with an Mn unknown insertion is used in a low stringency PCR with the primers S78, with homology to a short fragment inside the MTn, and primer S79 which is a primer with a 5’ unique and the 3’ randomised sequence NNNNNNGATTA ending in a highly repeated-sequence in the M. pneumoniae genome. Then, PCR products generated were purified as a pool and used as the template in a second PCR to isolate a single fragment with primers S78 and S80. This final fragment is sequenced to determine the insertion site of the MnTn. The MnTn location was then confirmed by PCR and sanger sequencing. For long-term studies, M. pneumoniae clonal strains C5, C20 and C26 were serially passaged every 48 h for 15 passages in SP-4 medium supplemented with puromycin, chloramphenicol, and theophylline, respectively. All strains used in this work are listed in Supplementary Data 1G.

Puromycin and chloramphenicol, or puromycin, chloramphenicol and theophylline. The location of the Mn in each clone was determined by A-PCR as previously described19. Briefly, genomic DNA from the clonal strain with an Mn unknown insertion is used in a low stringency PCR with the primers S78, with homology to a short fragment inside the MTn, and primer S79 which is a primer with a 5’ unique and the 3’ randomised sequence NNNNNNGATTA ending in a highly repeated-sequence in the M. pneumoniae genome. Then, PCR products generated were purified as a pool and used as the template in a second PCR to isolate a single fragment with primers S78 and S80. This final fragment is sequenced to determine the insertion site of the MnTn. The MnTn location was then confirmed by PCR and sanger sequencing. For long-term studies, M. pneumoniae clonal strains C5, C20 and C26 were serially passaged every 48 h for 15 passages in SP-4 medium supplemented with puromycin, chloramphenicol, and theophylline, respectively. All strains used in this work are listed in Supplementary Data 1G.

Quantiﬁcation of mCherry expression by ﬂuorimetry. Cell extracts were prepared from late-exponential phase cultures of M. pneumoniae strains grown in SP-4 medium with the relevant antibiosis and antibiotics, in the presence of the eGFP DNA kit (Sigma). DNA was quantiﬁed by qPCR using primers S124 and S125 (Supplementary Data 1E). One µL DNA sample, 5 pmols of each primer and 5 µL LightCycler 480 SYBR Green 1 Master mix (Roche) and a Roche Applied Science LightCycler 480 Instrument (384-wells) were used (10 µL reactions). A serial 1:10 dilution of M. pneumoniae genomic DNA sample, quantified by nanodrop, was used for the DNA standard curve.

The mCherry expression from synthetic promoters in M. pneumoniae was monitored by the fluorescence emission (Excitation 585 nm/Emission 625 nm) using a Tecan Infinite 200Pro plate reader. Samples were prepared following the protocol as explained for the mCherry fluorescence, except that for these experiments, the dilutions of the cell suspensions were prepared and 10 µL drops of all cell dilutions were spotted in 96-well-plates for a growth index reads every 2 h for 5 days. In these experiments, growth was monitored to ensure all samples behave similarly. Given the coloured SP-4 medium, the fluorescence reads have a high background signal at the beginning of the time-course, and it decreases with the acidification of the medium. In each experiment, a M. pneumoniae strain with an empty MTn was included to obtain the fluorescence background noise that is subtracted from the reads of the mCherry strains.

Background correction and regression model of clonal strains data to predict growth curve changes over strain passages. An effect of the medium volume reduction in the culture well is observed in the detection of the rate of absorbances Abs430/Abs560 for the clonal strains data types. Specifically, the volume reduction, due to evaporation, affects mainly Abs560 detection, while the effect on Abs430 is minimal. This is visible in SP4 data, consisting of the detection of Abs430/Abs560 in absence of the bacterium (Supplementary File 2). Linear regression was performed on SP4 data through the R (version 4.0.0) command lm for both Abs430 and Abs560, to capture the different entity of the effect of the volume reduction on the detection. Assuming both Abs430 and Abs560, in absence of the microorganism, should not change over time, the slope of the line obtained by linear regression was subtracted from the point-by-point mean of the original data, provided in triplicates. Slopes of the lines correspond to values of 1−0.08 for Abs430 and −2−0.07 for Abs560. The rates of absorbances CAbs430/CAbs560 for SP4 were recomputed with the background-corrected absorbances and standard deviations were calculated.

The difference between Abs430/Ab560 and CAbs430/CAbs560 was used to correct the data for all the strain types, for each condition and time point. All the
background-corrected curves, compared to the original data, are provided in Supplementary File 2. CAB430/CAB560 data on clonal strains were used for building regression models to predict curve’s parameters evolution over strain passages (Supplementary Fig. 15). The R package GrowthcurveR\(^{97}\) (version 0.3.1) was used to extract \(N_s\) (population size at the beginning of the exponential growth), \(K\) (carrying capacity) and \(r\) (growth rate) parameters for each strain passage. The population growth \(N_s\) over time \(t\) for each strain passage is described by Eq. (1):

\[
N_s(t) = \frac{K}{1 + \frac{K}{N_s} e^{-rt}}
\]

(1)

The death rate was estimated through linear regression with the R command `lm`.

**Simulation of gene disruption in *M. pneumoniae* through metabolic modelling.**

The most updated metabolic model of *M. pneumoniae* (iSIG158\(^{32}\)) comprehends all genes involved in the metabolism of this microorganism that have been annotated. Each gene included in the model has been disrupted in silico by allowing no flux through the reactions linked to the specific gene. For each disrupted gene, growth has been simulated through Flux Balance Analysis\(^{32}\), with biomass yield i.e. growth as an objective function, using cobrapy v. 0.5.11\(^{99}\) in Python v. 3.4.4. Growth-after-disruption of the single genes has been compared to the wild-type growth of *M. pneumoniae* strain M129 in iSIG158, mnp and percentages of decrease in growth have been computed (Supplementary Data 4).

**Whole-genome sequencing and variant calling analysis.**

Genomic DNA samples were prepared from late-exponential phase cultures of *M. pneumoniae* strains C5, C20 and C26 grown in SP-4 medium (and relevant antibiotics and inducers). Cells grown in 75 cm\(^2\) flasks were washed three times with PBS and followed by genomic extraction using the GenElute™ Bacterial Genomic DNA Kit (Sigma).

The London Biofoundry (Imperial College London) performed sequencing: Extracted genomes were prepared for sequencing with the Illumina Nextera DNA Flex kit (20187040) according to manufacturer recommendations via an automated implementation on an Opentrons OT-2 liquid handler. The resulting libraries were analysed and verified using a Fragment Analyzer (Agilent), Tapestation (Agilent) and Qubit (Thermo Fisher). Three libraries were pooled and diluted to 100 pM, and 20 µl loaded for sequencing and onboard de-multiplexing on an iSeq100 sequencer.

A list of single nucleotide polymorphisms (SNPs), and insertions and deletions (INDELs) present in each sequenced population was obtained by paired-end mapping of raw sequenced reads against BWA100. Then, alignments obtained were loaded for sequencing and onboard de-multiplexing on an iSeq100 sequencer.

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Code availability**

The pipeline for whole-genome sequencing and variant calling analysis, references and processing steps required to generate the intermediate and supplementary files to replicate the variant calling analysis can be found at https://github.com/SMV818VMS/kilowitch.

**Received:** 30 September 2021; **Accepted:** 24 March 2022; **Published online:** 07 April 2022

**References**

1. Fraser, C. M. et al. The minimal gene complement of Mycoplasma genitalium. *Science* **270**, 397–399 (1995).
2. Himmelreich, R. et al. Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic Acids Res.* **24**, 4420–4449 (1996).
3. Karr, J. R. et al. A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).
4. Manog, E. et al. Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632–638 (2007).
5. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
6. Hutchison, C. A. III et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad253–aad253 (2016).
7. Marcone, C. Molecular biology and pathogenicity of phytoplasmas. *Ann. Appl. Biol.* **165**, 199–221 (2014).
8. Taylor-Robinson, D. Antibiotic susceptibilities of mycoplasmas and treatment of mycoplasmal infections. *Antimicrob. Chemother.* **40**, 622–630 (1997).
9. Bebèr, C., Pereyre, S. & Peuchant, O. Mycoplasma pneumoniae: susceptibility and resistance to antibiotics. *Future Microbiol.* **6**, 423–421 (2011).
10. Wang, X. & Luktenhaus, J. Characterization of the ftsZ gene from Mycoplasma pulmonis, an organism lacking a cell wall. *J. Bacteriol.* **178**, 2314–2319 (1996).
11. Perez-Casal, J., Prysišak, T., Maina, T., Suleman, M. & Jimbo, S. Status of the development of a vaccine against Mycoplasma bovis. *Vaccine* [https://doi.org/10.1016/j.vaccine.2017.03.095](https://doi.org/10.1016/j.vaccine.2017.03.095) (2017).
12. Nicholas, R. A. J., Ayling, R. D. & McAuliffe, L. Vaccines for Mycoplasma diseases in animals and man. *J. Comp. Pathol.* **140**, 85–96 (2009).
13. Ishfaq, M. et al. Current status of vaccine research, development, and challenges of vaccines for Mycoplasma gallisepticum. *Poult. Sci.* **99**, 4195–4202 (2020).
14. Maes, D. et al. Perspectives for improvement of Mycoplasma hyopneumoniae vaccines in pigs. *Vet. Res.* [https://doi.org/10.1186/s13567-021-00941-x](https://doi.org/10.1186/s13567-021-00941-x) (2021).
15. The MycoSynVac project. Engineering of Mycoplasma pneumoniae as a broad-spectrum animal vaccine. EU H2020 grant No 634942. [https://www.mycosynvac.eu/](https://www.mycosynvac.eu/)
16. Güell, M. et al. Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268 (2009).
17. Lluch-Senar, M. et al. Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol. Syst. Biol.* **11**, 780 (2015).
18. Junier, I., Unal, E. B., Yus, E., Llorens-Rico, V. & Serrano, L. Insights into the mechanisms of basal coordination of transcription using a genome-reduced bacterium. *Cell Syst.* **3**, 11–16 (2016).
19. Llorens-Rico, V. et al. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv.* **2**, e1501363–e1501363 (2016).
20. Trussart, M. et al. Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat. Commun.* **8**, 14665 (2017).
21. Yus, E., Yang, J., Sagues, A. & Serrano, L. A reporter system coupled with high-throughput sequencing unveils key bacterial transcription and translation determinants. *Nat. Commun.* **8**, 368 (2017).
22. Miravet-Verde, S. et al. Unraveling the hidden universe of small proteins in bacterial genomes. *Mol. Syst. Biol.* **15**, e8290 (2019).
23. Yus, E. et al. Determination of the gene regulatory network of a genome-reduced bacterium highlights alternative regulation independent of transcription factors. *Cell Syst.* **9**, 143–158.e13 (2019).
24. Montero-Blay, A., Piñero-Lambea, C., Miravet-Verde, S., Lluch-Senar, M. & Serrano, L. Inferring active metabolic pathways from proteomics and protein turnover in a bacterium. *Mol. Syst. Biol.* **9**, 31 (2013).
25. Kühner, S. et al. Proteome organization in a genome-reduced bacterium. *Science* **1235**, 1235–1240 (2009).
26. Yus, E. et al. Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263–1268 (2009).
27. Maier, T. et al. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* **7**, 511 (2011).
28. van Noort, V. et al. Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol. Syst. Biol.* **8**, 571 (2012).
29. Lluch-Senar, M. et al. Comprehensive methylene characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at single-base resolution. *PLoS Genet.* **9**, e1003191 (2013).
30. Wolde, J. A. H. et al. Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Mol. Syst. Biol.* 9, 653 (2013).
31. Luch-Senar, M. et al. Comparative ‘omics’ in Mycoplasma pneumoniae clinical isolates reveals key virulence factors. *PLoS ONE* 10, e0137354 (2015).
32. Llorens-Rico, V., Luch-Senar, M. & Serrano, L. Distinguishing between productive and abortive promoters using a random forest classifier in Mycoplasma. *Nucleic Acids Res.* 43, 3442–3453 (2015).
33. Razin, S. The mycoplasmas. *Microbiol. Mol. Biol. Rev.* 42, 414–470 (1978).
34. Razin, S. Molecular biology and genetics of mycoplasmas (Mollicutes). *Microbiol. Rev.* 49, 419–455 (1985).
35. Mariscal, A. M., González-Gonzalez, L., Querol, E. & Piñol, J. All-in-one bacterial vector. *F1000Res.* 5, 2118 (2016).
36. Pérez-Lambea, C. E. et al. Mycoplasma pneumoniae genome editing based on oligo recombineering and Cas9-mediated counterselection. *ACS Synth. Biol.* 9, 1693–1704 (2020).
37. García-Morales, L. et al. A RAGE based strategy for the genome engineering of the human respiratory pathogen Mycoplasma pneumoniae. *ACS Synth. Biol.* 9, 2737–2747 (2020).
38. Ruiz, E. et al. CreaPy-cloning: a method for simultaneous cloning and engineering of megabase-sized genomes in yeast using the CRISPR-Cas9 system. *ACS Synth. Biol.* 8, 2547–2557 (2019).
39. Tsarpoupolou, I. et al. In-yeast engineering of a bacterial genome using CRISPR/Cas9. *ACS Synth. Biol.* 5, 104–109 (2016).
40. Baban, C. K., Cronin, M., Oms, J. & Llorens-Rico, V. Live bacterial vaccine vector and delivery system. *ACS Synth. Biol.* 9, 301–306 (2020).
41. Mota, L. J., Tavares, P. & de Sá-Nogueira, I. Mode of action of AraR, the key regulator of L-arabinose catabolism of methylxanthines in plants. *Handb. Exp. Pharmacol.* 200, 11–31 (2011).
42. Topp, S. & Gallivan, J. P. Riboswitches in unexpected places – A synthetic riboswitch in a protein coding region. *RNA* 14, 2498–2503 (2008).
43. Delicheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III. *Nature* 471, 602–607 (2011).
44. Jinke, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821 (2012).
45. Schuster, S. et al. The EAGR box structure: a motif involved in mycoplasma antigenic variation and sequence synthesis during DNA assembly. *Cell. Mol. Life Sci.* 67, 1254 (1996).
46. Calisto, B. et al. The EAG BOX structure: a motif involved in mycoplasma motility. *Mol. Microbiol.* 86, 382–393 (2012).
94. Shaner, N. C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. 22, 1567–1572 (2004).
95. Zimmermann, C. U. & Herrmann, R. Synthesis of a small, cysteine-rich, 29 amino acids long peptide in Mycoplasma pneumoniae. FEMS Microbiol. Lett. 253, 315–321 (2005).
96. Green, M. R. & Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2012).
97. Sprouffske, K. & Wagner, A. Growcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinforma. 17, 1–4 (2016).
98. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? Nat. Biotechnol. 28, 245–248 (2010).
99. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: Constrainsts-based reconstruction and analysis for Python. BMC Syst. Biol. 7, 1–6 (2013).
100. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
101. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at arXiv:1207.3907v2 [q-bio.GN] (2012).
102. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92 (2012).
103. Broto, A., Gaspari, E., Miravet-Verde, S., dos Santos, V. A. P. M. & Isalan, M. A genetic toolkit and gene switches to limit Mycoplasma growth for biosafety applications (this paper). Zenodo https://doi.org/10.5281/zenodo.6323207 (2022).

Acknowledgements
We thank Marko Storch, Miles Priestman and Marta Ciechonska from the London Biofoundry for sequencing support. This work was funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 634942 (MycoSynVac).

Author contributions
A.B. and M.I. conceived the project and wrote the manuscript; A.B. performed all the experiments; A.B and S.M-V. analysed the iSeq data; E.G and V.A.M.d.S. performed the computational and modelling part. All authors reviewed and approved the manuscript. M.I. supervised.

Competing interests
A.B and M.I declare that have filed a patent on this technology, as PCT Application Number PCT/GB2021/050184 (International Publication Number WO 2021/152301 A1, 5 August 2021). E.G. is currently employed as Project Officer at the European & Developing Countries Clinical Trials Partnership (EDCTP). E.G. declares no Horizon 2020 funding distributed by EDCTP has been deployed in this project and that the object of this manuscript is not involved in EDCTP scopes or funded grants. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-29574-0.

Correspondence and requests for materials should be addressed to Mark Isalan.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022