A randomized multiplex CRISPRi-Seq approach for the identification of critical combinations of genes

Nicole A. Ellis[1], Kevin S. Myers[2,3], Jessica Tung[1], Anne Davidson Ward[1], Kathryn Johnston[1], Katherine E. Bonnington[1], Timothy J. Donohue[2,3,4], Matthias P. Machner[1*]

[1]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA
[2]Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3]Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, Wisconsin, USA
[4]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA
*corresponding author: machnerm@nih.gov

Abstract

Identifying virulence-critical genes from pathogens is often limited by functional redundancy. To rapidly interrogate the contributions of combinations of genes to a biological outcome, we have developed a multiplex, randomized CRISPR interference sequencing (MuRCiS) approach. At its center is a new method for the randomized self-assembly of CRISPR arrays from synthetic oligonucleotide pairs. When paired with PacBio long-read sequencing, MuRCiS allowed for near-comprehensive interrogation of all pairwise combinations of a group of 44 *Legionella pneumophila* virulence genes encoding highly conserved transmembrane proteins for their role in pathogenesis. Both amoeba and human macrophages were challenged with *L. pneumophila* bearing the pooled CRISPR array libraries, leading to the identification of several new virulence-critical combinations of genes. *lpg2888* and *lpg3000* were particularly fascinating for their apparent redundant functions during *L. pneumophila* human macrophage infection, while *lpg3000* alone was essential for *L. pneumophila* virulence in the amoeban host *Acanthamoeba castellanii*. Thus, MuRCiS provides a method for rapid genetic examination of even large groups of redundant genes, setting the stage for application of this technology to a variety of biological contexts and organisms.

Introduction

There are many examples of synergistic processes in biology, often carried out by groups of redundant proteins that perform similar functions or whose activities result in the same biological outcome[1-3]. For example, the *Pseudomonas aeruginosa* genome encodes a total of 40 proteins for cyclic-di GMP synthesis and hydrolysis[4], *Legionella pneumophila* encodes 22[5], and *Vibrio cholerae* encodes at least 11[6]. Genes encoding proteins of similar function are commonly acquired either through gene duplication events or horizontal gene transfer and may have been selected for to provide fail safes to vital processes. While these redundancies generally provide benefits to the organism, they are obstacles towards gaining a fundamental understanding of biological processes. In these scenarios, the disruption of a single gene does not produce a detectable growth defect, hindering further biochemical and molecular analyses.

To dissect biological processes with intrinsic redundancy, genetic approaches must be multiplexed, allowing more than one gene to be disrupted at a time. In less tractable organisms, performing traditional deletion of individual genes, let alone combinations of genes, is laborious and time consuming. Newer genetic technologies, such as clustered regularly interspaced short palindromic repeats (CRISPR)-Cas technologies allow for fast, targeted, and multiplexed gene silencing or disruption[7, 8]. Developed from naturally occurring bacterial adaptive immune systems, CRISPR-Cas technologies are generally composed of a protein or group of proteins with enzymatic activity, usually nuclease activity, and CRISPR RNAs (crRNAs) yielding target gene specificity *via* homologous base pairing[9, 10]. *Streptococcus pyogenes* Cas9 is the most commonly used nuclease enzyme that, when guided by a crRNA to a complementary target DNA location, introduces a double-strand break[11, 12]. A catalytically inactive version of Cas9, dCas9, still localizes to complementary target genes upon crRNA direction, but instead of

inducing a double-strand break, precludes RNA polymerase activity, effectively silencing gene expression. This gene silencing approach, referred to as CRISPR interference (CRISPRi), is most effective upon binding the 5' region of the gene and is often used when a biological system lacks the machinery to repair double-strand breaks produced by the original Cas9 – as is true for most bacterial organisms[13, 14].

While silencing individual genes by CRISPRi has been achieved in various bacterial species[15, 16], suppressing two or more genes by CRISPRi has remained a major challenge as it requires the simultaneous expression of multiple gene-specific crRNAs. In nature, crRNAs are encoded by spacers separated by identical repeats in long stretches of DNA known as CRISPR arrays. Customizing CRISPR arrays in the laboratory has been a daunting task as repeat-containing DNA elements are often refractory to *in vitro* synthesis and cloning. Recently, we succeeded in building a multiplex CRISPRi platform in *L. pneumophila* that uses synthetic arrays capable of silencing up to ten bacterial genes at a time[17]. Despite much progress, a major short-coming of this and other existing multiplex CRISPRi approaches is their dependency on synthesis pipelines creating arrays of a defined set of crRNA-encoding spacers that target a predetermined set of genes[18-24]. In cases where predictions about the number and function of genes involved in a biological process are absent or incomplete, simultaneous gene silencing by CRISPRi will fail to detect synthetic lethal combinations of genes as the redundant genes not targeted by the array remain functional. This emphasizes the need for a protocol for randomized assembly of crRNA-encoding spacers into expansive libraries of diverse CRISPR arrays for unbiased probing of varied combinations of genes. Here, we develop a multiplex, randomized CRISPR interference sequencing (MuRCiS) approach that, when used in a proof-of-concept

experiment, discovered synthetic lethal combinations of genes from the virulence factor arsenal of *L. pneumophila*, the causative agent of Legionnaires' pneumonia[25, 26].

Results

Pre-determined CRISPR arrays fail to detect virulence-critical gene combinations

L. pneumophila is a Gram-negative bacterium that encodes over 300 predicted virulence factors[27, 28], known as effectors. The effectors are translocated *via* the Dot/Icm Type IV secretion system into either the natural host, free-living amoeba, or the disease host, human alveolar macrophages, to manipulate cellular processes and establish an intracellular replication compartment known as the *Legionella*-containing vacuolar (LCV)[29-31]. The study of *L. pneumophila* pathogenesis has been hindered by both the sheer number of effectors amassed by this pathogen[27, 28] and the apparent redundancy amongst them, as illustrated in the lack of detectable growth defects upon disruption of individual or even entire chromosome islands of effector-encoding genes[32, 33].

Since our earlier studies had shown that silencing of up to ten genes by multiplex CRISPRi is achievable in *L. pneumophila*[17], we built a library of 10-plex CRISPRi constructs (Table S1, MC, multiplex CRISPRi) capable of silencing more than 120 effector-encoding genes in groups of ten. The grouping of crRNA-encoding spacers into synthetic CRISPR arrays occurred based on two criteria: predicted protein function (groups 1-13, Table S1) and evolutionary conservation (groups A-D, Table S1). Protein functions were determined based on *in silico* predictions using Hidden Markov Model (HMM)-HMM-comparisons (HHPred; MPI Bioinformatics Toolkit)[34] and transmembrane predictors (TMHMM)[35]. Evolutionary conservation information was extracted from published bioinformatics analyses of effector-

encoding gene repertoires of 38 *Legionella* species[27]. To confirm the efficacy of multiplex gene silencing, we used quantitative polymerase chain reaction (qPCR) to assess knockdown efficiency of gene expression by six of the seventeen MC constructs in *L. pneumophila* bearing a chromosomal insertion of *S. pyogenes dcas9* at the *thyA* locus (Lp02(*dcas9*)[17]). We found that 10-plex gene silencing was reproducible for each construct and nearly all genes were knocked down at least 2-fold, with an average fold repression of one order of magnitude or more (Figure 1A).

Upon performing intracellular growth assays with this Lp02(*dcas9*) MC strain library in both U937 human macrophages and the amoeba *Acanthamoeba castellanii*, we found only a few intracellular growth phenotypes (Figure 1B). In U937 macrophages, the strains containing the MC array MC-A caused a growth defect as expected since it encoded crRNAs against *lpg2815* (MavN) and *lpg2300*, core effectors known to be vital for intracellular growth[27]. MC-10 improved growth of Lp02(*dcas9*) in U937 cells, likely due to the silencing of *lpg1368* (Lgt1), *lpg2862* (Lgt2), and *lpg1488* (Lgt3), glucosyltransferase effectors previously known to act in a redundant fashion[36]. MC-B caused a detectable growth defect of Lp02(*dcas9*) in U937 macrophages, suggesting it had silenced one or more virulence-critical effectors that had not been described before. Similar results were obtained during growth of the Lp02(*dcas9*) MC strain library in *A. castellanii,* where none of the MC arrays caused notable phenotypes. Together, these results show that silencing predetermined sets of genes, or even entire protein families, did not guarantee detection of synthetic lethal combinations of effectors.

Self-assembly of a randomized multiplex CRISPR array library

To screen for synthetic lethality more comprehensively, we developed a protocol where CRISPR arrays were assembled *de novo* from oligonucleotide pairs containing crRNA-encoding spacers. Since each spacer is flanked by 36 base pair (bp) repeats of identical sequence in a canonical CRISPR array, we rationalized that repeat sequences could be split and then re-linked during CRISPR array assembly. As such, we designed complementary DNA oligonucleotides, which we call R-S-Rs (repeat-spacer-repeat), composed of a 24 bp spacer flanked by the terminal 12 bps of the upstream repeat and the starting 20 bps of the downstream repeat, each with four bp "sticky" overhangs (TGAA = a common Golden Gate cloning overhang, Figure 2A). The coming together of different R-S-R building blocks would recreate complete repeats without spurious nucleotides (12 + 20 + 4 = classic 36 bp repeat). Since the spacer sequence was buried within the repeat sequences, competitive advantage of one R-S-R over another for integration into an array should be negligeable. Furthermore, since the four bp overhangs are identical for each R-S-R, spacers can recombine into arrays of any length and any order. To cap arrays on either end with Invitrogen Gateway sites for downstream cloning, two additional "dead end" oligonucleotide pairs were designed known as attB4r-R and R-attB3r (Figure 2B). Precursor-crRNA cleavage sites were not necessary to be incorporated into the R-S-R building blocks as the precursor-crRNA would be processed into individual crRNAs by RNases endogenous to *L. pneumophila.*

As target genes, we chose a total of 44 predicted *L. pneumophila* effectors with single or multiple transmembrane domains that our above mentioned *in silico* predictions had identified. These 44 effectors are likely incorporated either into the membrane of the LCV or that of surrounding organelles to assist in nutrient acquisition, detoxification, and membrane fusion (Figure S1, Table S2). Despite high conservation of these transmembrane effector-encoding

(TME) genes across all *Legionella* species[27, 37], only *lpg2815,* which encodes the metal ion transporter MavN[38, 39], has been shown to be virulence critical on its own. Hence, as proof-of-concept, we decided to probe this large group of genes with the randomized CRISPRi approach for synthetic lethality.

Assembly of CRISPR arrays followed simple, canonical cloning procedures (Figure 2C). Forty-four complementary R-S-R oligonucleotide pairs were incubated together with an aliquot of the attB4r-R and R-attB3r "dead end" oligonucleotide pairs and allowed to self-link after brief heat-denaturation and slow cooling. Arrays were then locked into place by treatment with T4 ligase. At this point, we found it critical to size-select arrays, as shorter arrays of only one or two spacers efficiently out-competed larger arrays during subsequent cloning steps. Size-selection proved challenging (Supplemental Note S1) but was ultimately accomplished through DNA gel purification of arrays that had first been introduced to an interim plasmid by Invitrogen Gateway cloning by way of the attB4r/attB3r sequences of the "dead ends". To select for longer arrays, two gel sections were individually isolated (550-650 bps and 650-800 bps), corresponding to arrays of 2-4 and 4+ spacers in practice. Lastly, we used an Invitrogen Multisite Gateway Pro cloning strategy to create a final vector containing the tracrRNA-encoding sequence, and the pooled, size-selected randomized CRISPR array library flanked by a promoter (P_{tet}) and terminator (*rrnB* T1) sequence. Keeping the tracrRNA separate from crRNAs, unlike in single guide (sg)RNAs, kept the R-S-R building blocks at a manageable length. Upon navigating this assembly protocol twice, the final vector libraries were introduced into Lp02(*dcas9*) by electroporation and advanced to the next stage.

Self-assembled CRISPR arrays are diverse in length, order, and composition

To assess the diversity of array composition, CRISPR array-containing vector libraries (Library 1 and 2) were harvested from two Lp02(*dcas9*) subpopulations, linearized, and analyzed by next-generation sequencing. Importantly, given the repetitive nature of the CRISPR arrays, canonical short-read sequencing technologies (such as Illumina or 454 sequencing) would not have been adequate as they would not have provided enough unique sequence overlap outside of the repeat sequence to map the reads back to a specific array within the mixed vector population. Instead, we made use of a high-throughput, low error rate long-read sequencing technology (PacBio Sequel). The average length of reads containing R-S-R elements was ~10,400 bp, correlating to the size of the vector backbone plus a multiplex array (sequencing metrics reported in Table S3).

We harvested vectors four times from each Library 1 and Library 2 stock to assess array diversity and library reproducibility. We designed a custom bioinformatics pipeline (discussed below) to condense data from all arrays with identical spacer content into one data set, regardless of spacer order; as such we found that Library 1 contained 784 unique spacer combinations and Library 2 contained 1251 unique spacer combinations (overlap = 250 combinations found in both). When requiring a stringent arbitrary cut-off of five or more raw (pre-sequence-depth normalized) read counts per spacer combination, a total of 231 and 519 spacer combinations were observed in Library 1 and 2, respectively, with 121 and 300 of those spacer combinations being present in each of the four replicates (A-D and E-H; Figure 3A). We found that arrays of spacers were most often 2- to 3-plex in nature, though some reached 5-, 6-, or even up to 11-plex to give an average of 3.3 spacers per array (Figure 3B, Table S4).

Each library contained a diversity of arrays that would assess silencing of nearly all 946 possible pairwise combinations of the 44 TME genes. Chord diagrams plotting each time spacers

targeting two different TME genes were found in the same array provided a visualization of the remarkable pairwise comprehensiveness of each library, with the weight of the linking lines indicating frequency of observation (Figure 3C). Both libraries showed vast diversity in array composition, with Library 2 showing the greatest randomization of arrays leading to the most comprehensive pairwise coverage. Library 1, surprisingly, had an over-representation of arrays targeting the same five genes (*lpg0963, lpg2223, lpg2552, lpg2888,* and *lpg3000)* that accounted for 20% of all arrays in that library (indicated by the thick network of linking lines). Notably, from an array building vantage, the 5-plex arrays targeting these five genes were found to have seven different spacer orders, indicating that they had been generated multiple times during the assembly step. Together, these two libraries provide a vast collection of multiplex arrays to test for combinatorial TME gene contributions to *L. pneumophila* pathogenesis strategies.

Decoding randomized CRISPR array libraries by high-throughput sequencing

To identify combinations of genes that, upon silencing, were detrimental to intracellular *L. pneumophila* replication, vectors from Lp02(*dcas9*) cultures bearing the pooled input array libraries were purified, linearized, and PacBio Sequel-sequenced to determine the identity and composition of arrays before infection (Figure 4A). From the same cultures, pooled bacteria were simultaneously introduced to a selective pressure, in this case host cell infection, and surviving bacteria were harvested from host cells after 72 h (for U937 macrophage infections) or 48 h (*A. castellanii* infections) and grown as single colonies on media plates. Vectors from arising bacterial colonies were purified, linearized, and analyzed again by long-read sequencing to determine the identity of arrays in the output vector pool.

Since each PacBio sequencing read reported the complete spacer composition of an array, barcoding was dispensable. A custom read count-based bioinformatics pipeline was designed to determine arrays that were under-represented in the output pool, and therefore, must have silenced genes vital for successful intracellular replication (Figure 4B, https://github.com/GLBRC/MuRCiS_pipeline). Taking a most stringent approach, only reads that were an exact match to the repeat sequence were analyzed. Spacers within an array were identified by searching each read for exact matches to all possible forward and reverse combinations of spacer sequences and the repeat sequence. The number of arrays with the same spacer composition were counted and an arbitrary cut-off of five raw counts in the input pool was required for further analysis of that unique spacer combination. The remaining unique array counts were normalized for run-to-run sequence depth variation, and the proportion of each unique array compared to the total number of sequenced reads containing a repeat sequence was determined for each sample. For each experiment, the proportion of normalized counts for each unique spacer composition in the input sample was compared to the normalized counts for the same spacer composition in the output sample to determine a fold reduction. The larger the reduction, the more disadvantageous silencing of the targeted genes was for intracellular *L. pneumophila* replication. We excluded any subset of spacers with an input *vs.* output reduction of less than five-fold. Together, this experimental and bioinformatics pipeline forms the basis for MuRCiS.

MuRCiS elucidates novel combinations of virulence-critical genes

We first interrogated the contributions of TME genes to successful intracellular replication of *L. pneumophila* in human U937 macrophages. We performed four rounds of

infections: two with Lp02(*dcas9*) bearing Library 1 (U-1A and U-1B) and two with Lp02(*dcas9*) bearing Library 2 (U-2A and U-2B). Each pair of infections, A and B, occurred independently of each other serving as biological replicates. To test technical reproducibility, we also collected output samples twice (T1 and T2) when performing experiments "A". The four inputs for these experiments were A, C, E, and G (Figure 3A). In total, the four input cultures used to challenge U937 macrophages allowed for the probing of 710 of a possible 946 pairwise combinations (75%) of 44 TME when requiring a stringent arbitrary cut-off of five or more raw read counts (Figure 5A, Table S4). In fact, pairs were often in the input pool between eleven and 500 times allowing for in-experiment replication. Silencing *mavN,* which encodes the essential metal ion transporter, served as a control for the infection experiments. While hundreds to thousands of reads bearing the crRNA^{mavN}-encoding spacer were present in each input pool, each of the output samples (T1/T2 summed) had only ten or fewer reads (Figure 5B). These results indicated that MuRCiS could identify virulence-critical genes by monitoring changes in spacer abundance during host cell infection.

In total, 641 unique spacer combinations with five or more raw input reads were tested using this protocol (Figure 5C). Subsequent bioinformatic analyses found that 231 of those combinations showed a five-fold or greater reduction in read counts. All 59 combinations which contained the crRNA^{mavN}-encoding spacer were removed from the list, and 39 spacer combinations were identified as having a five-fold or greater read count reduction in both replicates of each library experiment (including both technical replicates). These 39 spacer combinations were examined further by gathering shorter versions of the arrays to track down the minimal number of spacers needed for the strain to show a fitness defect (Table S5). Twenty critical combinations of spacers were identified, of which most were 2- to 3-plex combinations

(Figure 5D and 5E). Because of the uniqueness of Libraries 1 and 2, different gene combinations were discovered by each library to be virulence-critical, suggesting that the assembly and assay of more than one library was advantageous for comprehensiveness. Targeting of *lpg2888* and *lpg3000* was identified to be deleterious to *L. pneumophila* intracellular growth in experiments with both libraries. Table S6 shows data for each of the subsets of the 20 critical gene combinations, providing evidence that it was truly the silencing of the reported combination of genes, and not fewer or individual genes, that was needed for strain to show a fitness defect.

Multiple TME pairs are critical for *L. pneumophila* virulence in macrophages

The nature of CRISPRi gene silencing and the novelty of the screen performed here mandated confirmation of targets by construction of strains with true chromosomal gene deletions, as the degree of target silencing can be influenced by a range of factors related to (1) crRNA design, (2) genetic environment, and on occasion (3) off-target effects. Individual and pairwise deletions of the candidate genes were made for five of the combinations identified in Figure 5E. All deletion strains were whole genome sequenced for confirmation of gene deletion and absence of background mutations and then used to challenge macrophages. Their ability to survive and grow over a period of three days was determined by plating assay.

This analysis confirmed the expected growth-inhibitory phenotype for three of the five gene pair deletions (Figure 5F), namely *lpg2888-lpg3000*, *lpg1959-lpg2804*, and *lpg0096-lpg2804*. The severity of the phenotype varied amongst the different pairs, ranging from a ~75-fold reduction in growth to complete inability to replicate in the host. For two pairs, *lpg2552-lpg3000* and *lpg0096-lpg2888*, we found that deleting only one of the genes, *lpg3000* and *lpg2888*, respectively, was sufficient for limiting *L. pneumophila* intracellular replication (Figure

5G). While we aimed to find critical combinations of genes, these "false positive hits" still identified genes important to *L. pneumophila* intracellular replication demonstrating much can be learned from this assay.

Virulence-critical gene combinations are host-specific

Having found critical combinations of genes required for growth of *L. pneumophila* in its disease host, we next interrogated the cohort of TME genes for their necessity during *L. pneumophila* growth in the natural host, *A. castellanii*. As before, we performed four rounds of infections: two with Lp02(*dcas9*) bearing Library 1 (AC-1A and AC-1B) and two with Lp02(*dcas9*) bearing Library 2 (AC-2A and AC-2B). The four inputs for these experiments were B, D, F, and H (Figure 3A). In total, these four inputs allowed for probing 672 of a possible 946 pairwise combinations (71%) of 44 TME genes when requiring a stringent arbitrary cut-off of five or more raw read counts (Figure 6A, Table S4).

This time, 571 unique spacer combinations with five or more raw input reads were tested during the experiments, leading to the identification of 71 spacer combinations with five-fold or greater reduction in read counts (Figure 6B). Of those, 21 emerged from both replicates of a given library experiment. We examined these spacer candidates further by tracking the most minimal number of spacers needed to attenuate *L. pneumophila* intracellular growth (Table S7). Six critical combinations of TME genes were ultimately identified, of which all were 2-plex combinations (Figure 6C and 6D, Table S8). Notably, silencing of *lpg2888* and *lpg3000* was again found to be deleterious to *L. pneumophila* growth in *A. castellanii,* as it was in U937 macrophages (Figure 5E). Having already constructed *L. pneumophila* stains bearing deletions in *lpg2888* and/or *lpg3000*, we tested these for growth in *A. castellanii.* Surprisingly, in *A.*

castellanii deletion of *lpg3000* alone was sufficient to cause a growth defect as dramatic as that observed for the double mutant (Figure 6E). This contrasted with U937 macrophage experiments where only deletion of both *lpg2888* and *lpg3000* resulted in attenuated growth (Figure 5F).

Two of the other critical virulence-gene combinations identified in the *A. castellanii* experiments also contained *lpg3000: lpg0086-lpg3000* and *lpg0621-lpg3000,* suggesting we would again observe deletion of *lpg3000* sufficient of a maximum intracellular growth defect if tested. Though each library included 1-plex arrays expressing only crRNA^{lpg3000} for silencing *lpg3000*, we did not identify those 1-plex arrays as being sufficient to attenuate virulence, likely because their decrease was not consistent between Library 1 and Library 2 experiments, although there was consistency between replicates A and B of each library (Table S8). A possible explanation was that incomplete knockdown (*vs.* complete knockout) of a seemingly important process was at a tipping point of phenotype extremes, and that silencing of an additional effector may have tipped the scale. If requiring a five-fold or greater reduction in only one library and not the other, an additional three (two of which are 1-plex; *lpg0086* and *lpg2223*) and six (one of which is 1-plex; *lpg0086*) spacer combinations may qualify as critical to U937 macrophage and *A. castellanii* infection, respectively.

Single-round CRISPRi experiments are of sufficient sensitivity to detect virulence-critical genes

For any virulence-critical gene combination to be listed as hit during U937 macrophages infection (Figure 5) or *A. castellanii* infection (Figure 6), the CRISPR arrays silencing them had to pass the most stringent cutoff criteria, including that the spacer combination had to be represented in both biological replicates, A and B. There were instances though where CRISPR

arrays were reduced in only one replicate because arrays bearing the same spacer combinations were absent in the other replicate (input overlap shown in Figure 3A). When requiring a five-fold or greater reduction in only one replicate, an additional 18 and 14 virulence-critical spacer combinations were observed in the U937 macrophage and *A. castellanii* experiments, respectively (Table S9). For example, in the infection experiment AC-2A, arrays silencing both *lpg0096* and *lpg2888* were nine-fold under-represented in the output pool after *L. pneumophila* growth in *A. castellanii*. *L. pneumophila* strains bearing deletions in one or both of these genes were examined for their growth in *A. castellanii.* Interestingly, unlike in U937 macrophages, we did observe a synergistic phenotype upon deletion of *lpg0096* and *lpg2888* (Figure 6F). These data represent yet another example of a host-specific phenotype and suggest that the proof-of-concept study performed here may have identified even more virulence-critical combinations of *L. pneumophila* genes than initially thought warranting examination of gene combinations that are considered hits in only one of the experimental replicates.

Discussion

In this study, we expanded the versatility of CRISPRi platforms in bacteria by devising an approach capable of "shuffling" the spacer composition of CRISPR arrays, thus creating a randomized, multiplexed genetic tool adept to cross-examine even large cohorts of *L. pneumophila* effectors for synthetic lethality. A critical innovation of our approach was the quantitative (abundance output vs input pool) and qualitative (spacer composition) analysis of CRISPR arrays by high-throughput long-read sequencing. Indubitably, MuRCiS will be applicable to the study of any group of redundant genes in a wide range of CRISPRi-capable organisms.

Our *de novo* assembly protocol created distinct libraries of arrays each time it was executed, exemplifying its truly random nature (Figure 3). As mentioned, Library 1 unexpectedly amassed arrays targeting five specific genes (*lpg0963, lpg2223, lpg2552, lpg2888,* and *lpg3000)*, such that these arrays accounted for 20% of the input read counts for this library. These arrays did decrease in abundance during *L. pneumophila* infection of both U937 macrophages and *A. castellanii,* consistent with our finding that *lpg2888* and *lpg3000* are important for replication in both these hosts. We do not fully understand why these particular arrays were so prevalent only in Library 1, even more so given that assembling arrays of fewer than five spacers proved more efficient (Figure 3B). It is possible that silencing these TME genes provided *L. pneumophila* with a fitness advantage during axenic growth causing them to be overrepresented in the input pool.

A unique feature of MuRCiS is its ability to quickly assign phenotypes to exact spacer combinations simply by tracking phenotypes to shorter arrays with subsets of spacers and, hence, target genes for subsequent follow-up studies (Table S5 and Table S7). For example, the 5-plex arrays containing *lpg0963, lpg2223, lpg2552, lpg2888,* and *lpg3000* can be subdivided into 30 subsets of spacers - all of which were present in the experiment and allowed assigning the phenotype to just two genes, *lpg2888* and *lpg3000* (Table S5). Notably, not all phenotypes caused by long arrays could be condensed to just one or two spacers. Some phenotypes required 4- or 5-plex gene silencing (Table S6), showing that although only a minor fraction of all possible combination of three or more spacers were represented in our array library, the approach was capable of identifying these synthetic lethal gene silencing combinations as well.

Examining the PacBio DNA sequence data, we found that the number of unique arrays identified in a given sample increased as the read number increased (Table S3). For this proof-of-concept experiment we chose to limit the initial pool of arrays for analysis to those with a

minimum of five raw input counts. Maintaining this same cut-off with increased sequence depth undoubtedly would increase the input-to-input array overlay (Figure 3A) and the comprehensiveness of surveying pairwise silencing of genes (Figures 5A and 6A). We anticipated the gains of increased sequencing depth by recalculating the coverage of the correlation plots in Figure 5A and 6A such that the observed coverage of 710/946 (75%) or 672/946 (71%) pairs with the requirement of five raw input counts increased to 789/946 (83%) or 782/946 (83%) pairs when the requirement was lowered to just one raw input count, respectively. Hence, the ability to make comprehensive libraries was even better than first reported.

Array length poses the biggest bottleneck to performing MuRCiS on larger groups of genes. Here, each of the libraries individually probed an average of ~1000 different spacer combinations (independent of order). If all spacer combinations were represented by 2-plex arrays, theoretically one could investigate pairwise combinations of ~45 genes using just 990 arrays. But as reported, the assembly protocol produced longer arrays of up to 11 spacers, averaging 3.3 spacers. Each time a spacer is added to an array, the number of pairwise combinations able to be probed dramatically increases. If all arrays were 3-plex, 3000 pairwise combinations would be possible within just 1000 arrays allowing comprehensive interrogation of a group of ~75 genes for synthetic lethality, and if all arrays were 4-plex, 6000 pairwise combinations would be possible within just 1000 arrays allowing comprehensive interrogation of a group of ~110 genes for synthetic lethality. As it stands, the theoretical upper limit of MuRCiS seems between 75-100 genes, but with the construction of more than one library, deeper sequencing, and lowered input count requirements this number could easily be expanded.

Excitingly, the combinatorial effector gene deletions that were assayed in our experiments produce some of the largest phenotypes yet identified pertaining to *L. pneumophila* intracellular replication (Figure 5 and 6). This is especially true of the simultaneous deletion of *lpg2888* and *lpg3000* producing a truly synthetic lethal phenotype during U937 macrophage infection. The only other pair of *L. pneumophila* genes known to have such a large combinatorial impact on *L. pneumophila* intracellular replication are *icmW* and *icmR*[40], genes that encode components of the Dot/Icm Type IV secretion system itself; hence their deletion has a global effect limiting translocation of many effector proteins into the host. HHpred analysis of *lpg2888* revealed numerous high confidence hits to tripartite pore-forming toxin components, while *lpg3000* showed high confidence hits to ABC transporter proteins. Additional studies are ongoing to uncover why deletion of their encoding genes caused synthetic lethality. From an evolutionary standpoint, it was interesting that synergy of these genes was only seen during *L. pneumophila* infection of U937 macrophages. During *A. castellanii* infections, deletion of just *lpg3000* was already lethal, whereas deletion of *lpg2888* alone caused just a minor growth defect (Figure 6E). These data are in agreement with data from a transposon sequencing (Tn-Seq)-based screen that identified a growth defect upon disruption of *lpg3000* in *L. pneumophila* during infection of *A. castellanii* and another protist, *Hartmannella vermiformis*[41]. Altogether, these results suggest that the biological process promoted by *lpg3000* is evolutionarily essential for growth in the natural amoeban host. Clearly, MuRCiS identified exciting biology pertaining to *L. pneumophila* pathogenesis; and if applied to other groups of genes, other selective pressures or other microorganisms, it undoubtedly has the potential to do the same.

Methods

Strains and cell lines

All *L. pneumophila* strains are listed in Table S10 and were derived from the parent strain *L. pneumophila* Philadelphia-1 Lp02 (*thyA hsdR rpsL*). Host cell lines included U937 monocytes (ATCC CRL-1593.2) which were differentiated into macrophages with 12-O-tetradecanoylphorbol-13-acetate (TPA, Sigma-Aldrich P1585) and *A. castellanii* (ATCC 30234).

Synthetic MC array construction

Synthetic 10-plex MC array constructs were synthesized by GenScript and moved into pMME977 (*thyA*+) by Multisite Gateway Pro cloning (Invitrogen 12537-100) as previously described[17]. The MC arrays tested here bear a *boxA* sequence -58 bp upstream of the first repeat which was added after vector completion by quick-change PCR as described previously[17]. Final plasmids were introduced to Lp02(*dcas9*) by electroporation and the strains containing the plasmids were selected for on CYE plates[42] without thymidine. Spacer composition of each array and their corresponding nucleotide sequences are listed in Table S1.

Axenic growth of *L. pneumophila* 10-plex CRISPRi strains

As described previously[17], *L. pneumophila* cultures were grown overnight in AYE +Fe +Cys (10 g ACES, 10 g yeast extract per liter, pH 6.9 with 0.4 mg/ml cysteine and 0.135 mg/ml ferric nitrate) under non-inducing conditions. On the second day, cultures were sub-cultured twice (AM and PM, ~6-7 h apart) to OD600 0.2–0.3 with 2–3 mL fresh AYE +Fe +Cys containing 40 ng/mL anhydrous tetracycline (aTC, Clontech 631310). On the third day, cultures that had reached OD600 3-5 (post-exponential growth) were collected for mRNA analyses and/or used in host cell infections.

RNA extraction and qPCR

RNA extraction was performed on bacteria pellets after aTC induction using the Trizol Max Bacterial RNA Isolation Kit (Invitrogen 16096040). Contaminating DNA was removed using the Turbo DNA-free Kit (Invitrogen AM1907) and RNA was converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems 4368814). qPCR was performed using the SYBR Green Master Mix (Applied Biosystems 4367659) on a StepOne Plus Real-Time PCR System (Applied Biosystems) using comparative CT. qPCR primers are listed in Table S1. mRNA levels from different samples were normalized to a house-keeping gene (*rpsL*) and mRNA levels in CRISPRi strains were compared to that of a vector-bearing Lp02(*dcas9*) control strain using the ΔΔCT method[43] to determine fold repression.

CRISPRi intracellular growth assays

U937 monocytes (ATCC CRL-1593.2) were maintained in DMEM (Mediatech 15-013-CV) + 10% FBS + glutamine at 37 °C. Two days prior to challenge with *L. pneumophila*, cells were plated on 24-well plates at $3x10^5$ cells/well with 0.1 μg/mL 12-O-tetradecanoylphorbol-13-acetate (TPA, Sigma-Aldrich P1585) to promote differentiation. *L. pneumophila* strains bearing each MC array were incubated under inducing conditions, as described above, and added to differentiated U937 macrophages in DMEM + 10% FBS + glutamine containing 40 ng/mL aTC at a multiplicity of infection (MOI) of 0.05. Plates were centrifuged for five minutes at 200 x g to increase bacteria-macrophage contact. After a 2 h incubation at 37 °C, extracellular bacteria were removed by washing the macrophages twice with DMEM + FBS + glutamine media containing 40 ng/mL aTC. Both 2 and 72 hours post infection (hpi) intracellular bacteria were collected upon macrophage cell lysis by addition of digitonin (0.02% final concentration), serially diluted, and spotted on CYE plates. Results are given as fold growth (CFUs harvested 72 hpi *vs.* 2 hpi) compared to that of the vector-bearing Lp02(*dcas9*) control.

A. castellanii (ATCC 30234) were maintained in PYG media[44] at 25 °C. The day before challenge with *L. pneumophila*, cells were plated on 24-well plates at $3x10^5$ cells/well. The morning of challenge, the media in the *A. castellanii* plates was changed from PYG to AC buffer[44] to promote starvation for 2 h at 25 °C. *L. pneumophila* strains bearing each MC array were incubated under inducing conditions, as described above, and added to the starved *A. castellanii* in AC buffer containing 40 ng/mL aTC at a MOI of 0.03. Plates were centrifuged for five minutes at 200 x g to increase bacteria-amoeba contact. After a 2 h incubation at 37 °C, extracellular bacteria were removed by washing amoeba twice with AC buffer containing 40 ng/mL aTC. Both 2 and 48 hpi intracellular bacteria were collected upon amoeba lysis by addition of saponin (0.05% final concentration), serially diluted, and spotted on CYE plates. Results are given as fold growth (CFUs harvested 48 hpi *vs.* 2 hpi) compared to that of the vector-bearing Lp02(*dcas9*) control.

Multiplex random CRISPR array self-assembly protocol

R-S-R oligonucleotides were designed as the building blocks for array assembly. For each gene, the reverse complement of the 24 nucleotides downstream of the first PAM sequence (NGG) after the transcription start site was identified and would become the spacer sequence. This sequence was checked to be null of cut sites for SacI (GAGCTC), NotI-HF (GCGGCCGC), and AflII (CTTAAG) as these are used for downstream steps; if not, the sequence adjacent the subsequent PAM sequences was checked. Spacer sequences were then flanked on either side with one half of the repeat according to the following scheme: top oligonucleotide (5’ to 3’): tgaatggtcccaaaac-24 nt spacer-gttttagagctatgctgttt. Bottom oligonucleotide (5’ to 3’): ttcaaaacagcatagctctaaaac-24 nt complement-gttttgggacca. Standard oligos were ordered from Eurofins Genomics already resuspended in water at 100 µM concentration. Dead end

oligonucleotide pairs for R-attB4r and R-attB3r did not contain a spacer but did contain the upstream or downstream repat sequence fused to the attB4r or attB3r Invitrogen Gateway cloning sequence, respectively.

For array assembly, top and bottom oligonucleotides for a single R-S-R or dead end oligonucleotide pairs were combined 1 µL:1 µL and were phosphorylated in a 50 µL reaction using T4 Polynucleotide Kinase (NEB M0201L) for 2 h at 37 °C. The reaction was stopped by addition of NaCl to a final concentration of 50 mM. Next, 2 µL of each phosphorylated R-S-R oligonucleotide pair for all 44 gene targets and 10 µL of the dead end oligonucleotides were combined in a single microcentrifuge tube. The microcentrifuge tube was placed in a heat block at 95 °C for 5 minutes and then the heat block was turned off and allowed to cool to room temperature for 2 h. Assembled arrays were preserved through ligation with T4 DNA ligase and its corresponding buffer (NEB M0202L) for 1.5 h at room temperature. The reaction was purified using a NucleoSpin PCR Clean-up Kit (Macherey-Nagel 740609.250) and eluted in 30 µL 1X TE buffer.

To size select arrays of interest, four 6.5 µL aliquots of the purified arrays were moved into the interim pDonor4r-3r in parallel by the standard Invitrogen Gateway BP reaction recipe (ThermoFisher 11789-020) which was allowed to remain at room temperature for 72 h prior to stopping the reactions with Proteinase K and transformation of all material into *E. coli* GC5 (Genesee Scientific 42-650) with selection on LB+Kan (30 µg/mL) plates. All colonies were scraped off plates using an inoculation loop and added directly to the A1 buffer of a Nucleospin Plasmid Purification Kit (Macherey-Nagel 740499.250) for plasmid isolation and elution in water. A NotI-HF (NEB R3189S) and AflII (NEB R0520S) double digest was performed on the four now recombined eluates, as well as one aliquot of the circular empty pDonor4r-3r plasmid,

at 37 °C overnight. Next, a 1% agarose gel was used to separate the vector backbone from the excised size-ordered arrays, and we collected bands containing arrays ≥2-plex at 550-650 bps and 650-800 bps to be purified separately using a NucleoSpin Gel Purification Kit (Macherey-Nagel 740609.250). We also collected the cut vector backbone, but only from the empty pDonor4r-3r sample. For all steps forward, the 550-650 bps and 650-800 bps samples were kept separate, but procedures were carried out in parallel. As such, purified arrays were ligated back into the pDonor4r-3r backbone using T4 DNA ligase for 2.5 h at room temperature, all material was transformed with the plasmid into *E. coli* GC5, and re-circulated plasmids were selected for on LB+Kan plates. All colonies were scraped off plates using an inoculation loop and added directly to the A1 buffer of a Nucleospin Plasmid Purification Kit (Macherey-Nagel 740499.250) for plasmid isolation and elution in 20 μL 1X TE buffer. At this point, individual plasmids can be sequenced using a primer with the sequence: GTTTTCCCAGTCACGAC, if individually purified from a single colony, to confirm successful array assembly.

A Multisite Gateway Pro kit (Invitrogen 12537-100) was used to move arrays into the final vector. To generate the promoter-bearing donor plasmid pMME2162, the *tet* promoter was amplified from pMME1996 with primers containing attB5/attB4 ends for recombination into the pDonorP5-P4 via a standard Invitrogen Gateway BP reaction. To generate the terminator-bearing donor plasmid pMME2163, the *rrnB* T1 terminator was amplified from pMME1996 with primers containing attB3/attB2 ends for recombination into the pDonorP3-P2 via a standard Invitrogen Gateway BP reaction. Next, the size-selected arrays in the pDonor4r-3r plasmid were combined with pMME985 (tracrDNA in pDonorP1-P5r, described previously[17]), pMME2162, and pMME2163, and were added to the pMME977 destination vector (attR1/attR2, +*thyA*) by using three-times the standard Invitrogen LR reaction recipe (ThermoFisher 11791-020), with

addition of β-mercaptoethanol to a final concentration of 1 mM, and allowed to remain at room temperature for 72 h prior to stopping the reactions with Proteinase K and electroporation into *E. coli* GC5 with selection on LB+Amp (100 µg/mL) plates. Next, all colonies (generally thousands for 550-650 bps samples and hundreds for 650-800 bps samples) were scraped off plates using an inoculation loop and added directly to the A1 buffer of a Nucleospin Plasmid Purification Kit (Macherey-Nagel 740499.250) for plasmid isolation and elution in 20 µL water. At this point, individual plasmids can be sequenced using a primer with the sequence: CAACCACTTTGTACAAGAAAGCTGGG, if individually purified from a single colony, to confirm successful construct assembly.

Lastly, 10 µL of these final plasmids were introduced into Lp02(*dcas9*) by electroporation and recipient cells were selected for on CYE plates. Colonies were scraped off plates using an inoculation loop, added to 15 mL AYE +Fe +Cys, vortexed, and frozen in 1 mL aliquots with glycerol (final concentration 15%) to be stored at -80 °C. Freezer stocks of Lp02(*dcas9*) bearing each of the size-selected portions of a given library (Library 1 and 2) are listed in Table S10.

MuRCiS intracellular growth assays

L. pneumophila bearing each size-selected portion of either Library 1 or 2 were patched from -80 °C freezer stocks onto CYE plates. Two days prior to infection *L. pneumophila* cultures of each were grown overnight in 3 mL AYE +Fe +Cys under non-inducing conditions. On the second day, cultures were sub-cultured twice first in the AM to OD600 0.2 with 3 mL fresh AYE +Fe +Cys containing 40 ng/mL aTC and then ~6-7 h later, to OD600 0.2 with 4 mL fresh AYE +Fe +Cys containing 40 ng/mL aTC (two cultures each). On the third day, cultures pertaining to either Library 1 or 2 were combined in one tube to serve as a single culture for infection (input

vector pool). For Library 1, two different 550-650 bps portions and one 650-800 bps portion was combined. For Library 2, one 550-650 bps portion and one 650-800 bps portion was combined.

U937 monocytes were maintained in DMEM + 10% FBS + glutamine and allowed to differentiate into macrophages with 0.1 μg/mL TPA for three days prior to *L. pneumophila* challenge. *A. castellanii* were maintained in PYG and starved in AC buffer 2 h prior to challenge with *L. pneumophila*. Ultimately, differentiated U937 macrophages or starved *A. castellanii* were plated in two to three 10 cm dishes at a density of $1x10^7$ cells per dish.

Infections were carried out with 40 ng/mL aTC in the host cell media at an MOI of 0.05 for differentiated U937 macrophages and 0.03 for starved *A. castellanii*. After addition of bacteria, the 10 cm dishes were spun down at 200 x g to increase bacteria-host cell contact and incubated for 2.5 h at 37 °C. Next, host cells were washed in their corresponding media with 40 ng/mL aTC to remove extracellular bacteria and then placed back at 37 °C for the duration of the infection, 72 h or 48 h for U937 macrophage and *A. castellanii* infections, respectively. To collect input vector pools, bacteria from the cultures used for infection were immediately pelleted by centrifugation and resuspended in 8 mL of RES-EF of the NucleoBond Xtra Midi EF kit prior to proceeding with the standard kit protocol (Macherey-Nagel 740420.50). Precipitated vectors were resuspended in 120 μL of water.

Following the infection duration, *L. pneumophila* were harvested from either the U937 macrophages or *A. castellanii* 10 cm dishes (15 mL vol) by addition of 300 μL of 1 % digitonin or 600 μL of 1% saponin, respectively. Next, 10 cm dishes were incubated for 10 minutes at 37 °C, followed by pipet agitation of the cells and collection of the media in 50 mL conical tubes. The conical tubes were shaken vigorously to promote host cell lysis and the *L. pneumophila* were pelleted by centrifugation for 10 min at 3500 x g. The resulting bacteria pellets were resuspended

in a minimal amount of supernatant and plated on CYE plates. All colonies were scraped off plates using an inoculation loop and resuspended in 8 mL of RES-EF of the NucleoBond Xtra Midi EF kit prior to proceeding with the standard kit protocol for collection of the output vector pools. Precipitated vectors were resuspended in 120 µL of water.

PacBio long-read sequencing

Purified vector populations from input and output samples were linearized by addition of 24 µL of SacI restriction enzyme (NEB R0156L) with 16 µL of NEB r1.1 buffer and overnight incubation at 37 °C. To purify the linearized vectors, the AMPureXP Bead protocol (Beckman Coulter A63880) was followed using 100 µL of bead slurry. Final eluent volumes varied between samples based on the ability to get a bead-free eluent. Generally, ~10 µL of water was needed for input sample elution and ~40 µL of water was needed for output sample elution. Input DNA concentrations were generally ~200-300 ng/uL and output DNA concentrations were usually 300-700 ng/uL. Vector linearization was confirmed on a 1% agarose gel.

PacBio long-read sequence of the linearized vectors was performed by the NIH NICHD Molecular Genomics Core. Samples were prepared for sequencing on a Sequel instrument (PacBio) using the Sequencing Primer v4 and the Sequel Binding Kit 3.0. Samples were multiplexed and run on multiple SMRT cells simultaneously (# SMRT cells = # multiplex samples divided by 2 or 3) with 10 h of sequencing camera time.

Bioinformatics pipeline

PacBio sequencing files (ccs_bam files) were processed using a custom bioinformatics pipeline (https://github.com/GLBRC/MuRCiS_pipeline). Total read number and lengths were determined using Samtools[45] (version 1.9) and standard Linux commands (see GitHub for specific commands). All spacer and repeat forward and reverse orientation sequence

combinations were used to identify the unique array sequence in the sequencing files using the "count_spacers_NIH.py" custom Python script (https://github.com/GLBRC/MuRCiS_pipeline). Reads lacking an exact match to the repeat sequence were removed from further analysis. Counts for unique spacers were normalized by dividing the read count plus a pseudocount of 1 by the total number of reads with repeat sequences multiplied by 10,000 in a modified Counts Per Million (CPM) equation. The 10,000 value was used instead of 1 million due to the lower number of overall reads from this PacBio sequencing.

$$\frac{(Read\ Count + Pseudocount) * 10{,}000}{(Total\ reads\ with\ repeat\ sequences)}$$

To determine the pairwise correlation of two spacers and plot Chord diagrams, the "chord_correlation_plot_script.R" script was used. All plots were generated in R (version 4.1.0) and all scripts and command descriptions are available on GitHub (https://github.com/GLBRC/MuRCiS_pipeline).

Deletion strain construction and whole genome sequencing

L. pneumophila gene deletion strains were constructed using gene-specific versions of the pSR47s suicide plasmid listed in Table S10 and described recently[46]. In our hands, each pSR47s plasmid was introduced into the *L. pneumophila* strain background of choice with the help of pRK600 by tripartite mating as described previously[47]. Successful strains were confirmed by failure to grow on CYET+Kan, colony PCR, and whole genome sequencing.

For whole genome sequencing, genomic DNA of *L. pneumophila* was extracted using the Wizard Genomic DNA Purification Kit (Promega A1120), following the standard Gram-negative bacteria protocol, and resuspended in 10 mM Tris-HCl, 1 mM EDTA buffer. The Illumina Nextera XT DNA Library Preparation Kit (#FC-131-1096) was used to prepare libraries for sequencing on an Illumina MiSeq System by the NIH NICHD Molecular Genomics Core.

Nucleotide variants were called using HaplotypeCaller (version GATK3, Broad Institute). Integrative Genomics Viewer (version 2.4.14, Broad Institute) was used to visually confirm chromosome deletions. Final deletion strains are listed in Table S10.

Deletion strain intracellular growth assays

Intracellular growth assays with deletion strains were performed exactly as for the 10-plex CRISPRi intracellular growth assays except that U937 monocytes were allowed to differentiate for three days prior to *L. pneumophila* challenge and *L. pneumophila* cultures were simply grown overnight without the inducer aTC. To overcome the thymidine auxotrophy of Lp02 strains, thymidine at a final concentration of 0.1 mg/mL was added directly to *L. pneumophila* cultures and to the DMEM + 10% FBS + glutamine media to promote growth in U937 macrophages with CFU plating on CYET plates, whereas pMME2400 (+*thyA*) was added to *L. pneumophila* strains by electroporation before growth in *A. castellanii* with selection on CYE plates. For U937 macrophage experiments the control strain was Lp02. For *A. castellanii* experiments that control strain was MML854 (Lp02 + pMME2400, G->A mutation at chromosome position 2134550).

Data availability

PacBio CCS BAM sequencing files are available at NCBI SRA under BioProject PRJNA902385 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA902385).

Acknowledgements

We thank the members of the Machner lab for thoughtful scientific discussion during the development of the MuRCiS workflow. We thank members of the NIH NICHD Molecular Genomics Core, especially Tianwei Li and James Iben, for executing the Illumina whole genome

sequencing, PacBio long-read sequencing, and for performing initial processing and transfer of data files.

Funding acknowledgement

This work was funded by the Intramural Research Program of the National Institutes of Health, USA under Project Number 1ZIAHD008893-12 (to M.P.M.) and the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409 (to T.J.D.).

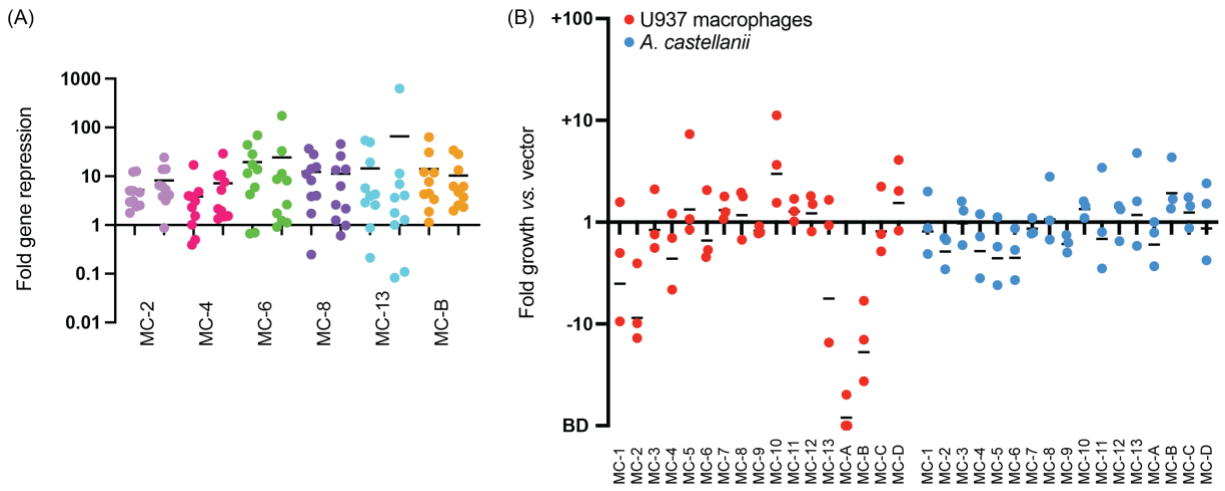


Figure 1: Intracellular growth of Lp02(*dcas9*) upon CRISPRi gene silencing with pre-determined 10-plex arrays

(A) Gene silencing efficiency assay. Each of the ten genes targeted by one of the six indicated MC arrays was assayed for gene repression by qPCR. RNA levels in the Lp02(*dcas9*) strain bearing the CRISPR array were compared to that bearing an empty vector. Data are a summary of two replicates shown side-by-side. Horizontal bars indicate mean fold repression of all targets of the array. (B) *L. pneumophila* intracellular growth assay. Host cells were challenged at a multiplicity of infection (MOI) of 0.05 (U937 macrophages) or 0.03 (*A. castellanii*) with Lp02(*dcas9*) bearing the indicated MC arrays. Colony forming units (CFUs) of samples taken 2 hours post infection (hpi) and either 72 (U937 macrophages) or 48 (*A. castellanii*) hpi were determined, and fold growth of Lp02(*dcas9*) bearing the indicated MC arrays relative to that of Lp02(*dcas9*) bearing the empty vector was plotted. Numbers below 1 indicate a growth defect upon gene silencing. BD, below detection (arbitrarily set to -100). Horizontal bars indicate mean fold growth *vs.* vector using data from three replicates.

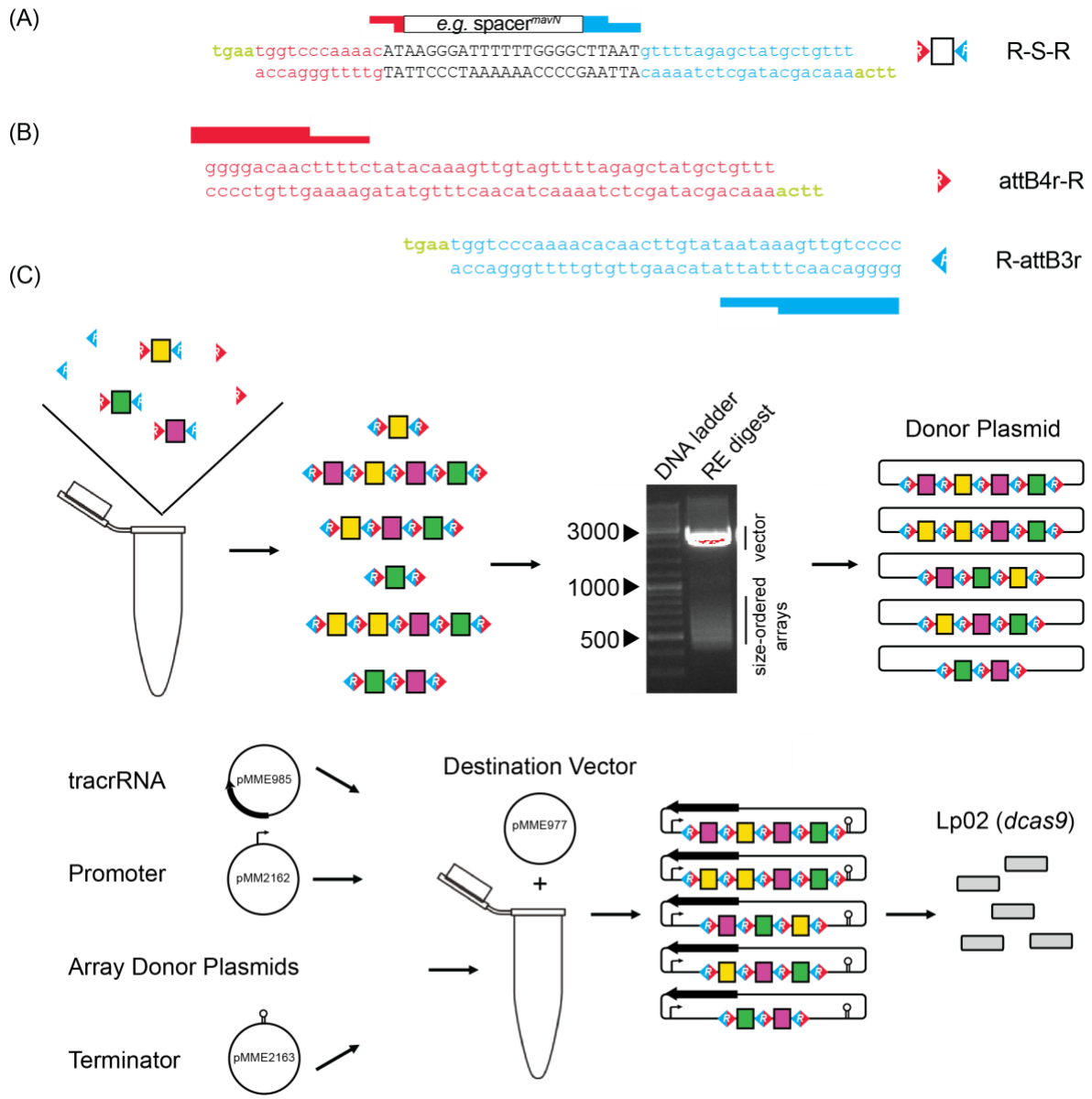


Figure 2: Cloning strategy for *de novo* self-assembly of CRISPR array libraries

(A) Nucleotide sequence of R-S-R building blocks. Each R-S-R element is composed of a top and bottom oligonucleotide. The 24 bp spacer sequence (black font; shown for crRNA^{mavN} as an example) is flanked by sequences of the upstream (red) and downstream (blue) repeat element, with each end containing sticky overhangs (green). (B) Nucleotide sequence of attB4r-R and R-

attB3r “dead ends” with only one sticky overhang each (green). (C) Array self-assembly, size selection, and cloning. R-S-Rs and “dead end” elements were allowed to self-assemble in a single tube and then ligated together. Arrays excised from an interim cloning vector were subjected to gel electrophoresis to separate them according to size. RE, restriction enzyme. After gel extraction, arrays of 550-650 bps and 650-800 bps in size were cloned into a donor plasmid such that three additional elements, namely the tracrRNA, a *tet* promoter, and a *rrnB* T1 terminator, could assemble with the arrays into the destination vector by Invitrogen Multisite Gateway Pro cloning. Final constructs were introduced into Lp02(*dcas9*) by electroporation.

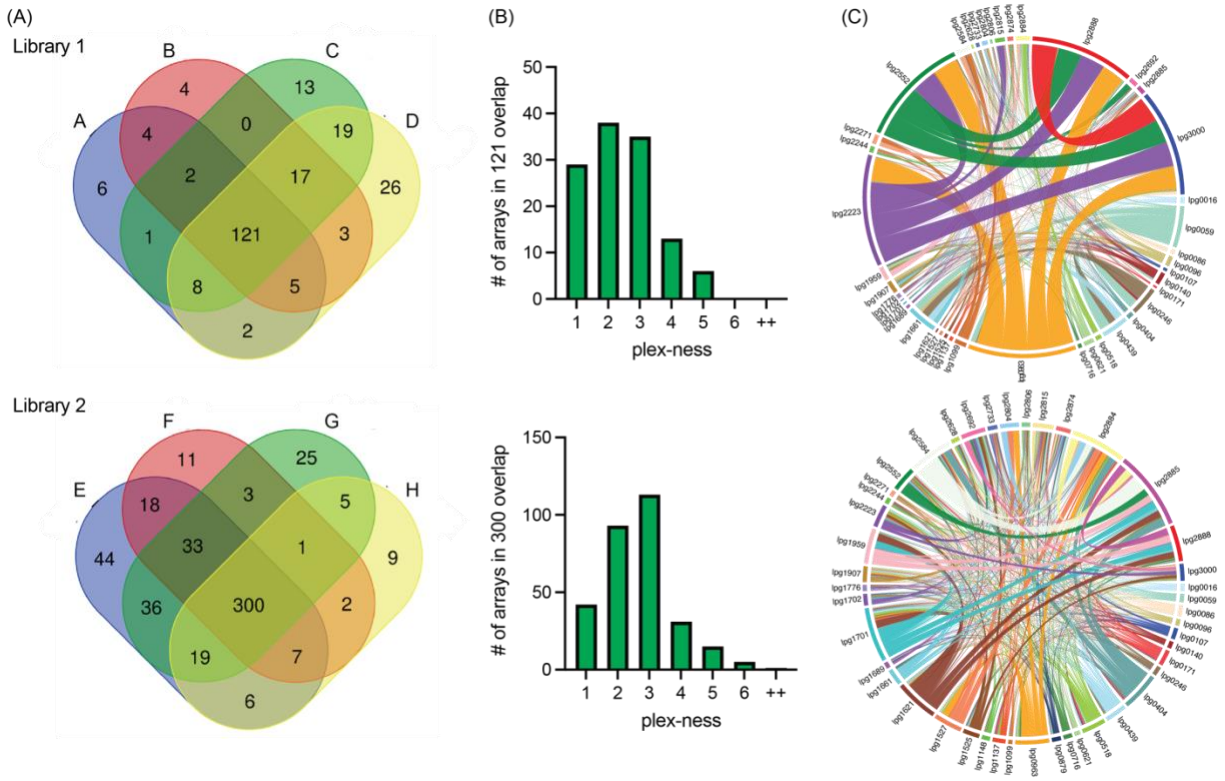


Figure 3: Self-assembled CRISPR arrays are diverse in length and spacer composition

(A) Two CRISPR array libraries were built from 44 R-S-R building blocks. Vectors from each library were harvested four times and sequenced. Venn diagrams show the overlap of unique spacer combinations, with the requirement of five or more raw (pre-sequence-depth normalized) read counts per array in each round of sequencing. (B) The distribution of plex-ness (spacer count) for arrays found in all four rounds of sequencing. (C) Chord diagrams link TME genes, listed around the outside of the circle, each time spacers targeting each is present in an array. Link line width is weighted according to the number of times the combination of spacers was observed. Link line color is unique for each spacer and is constant between the two diagrams.

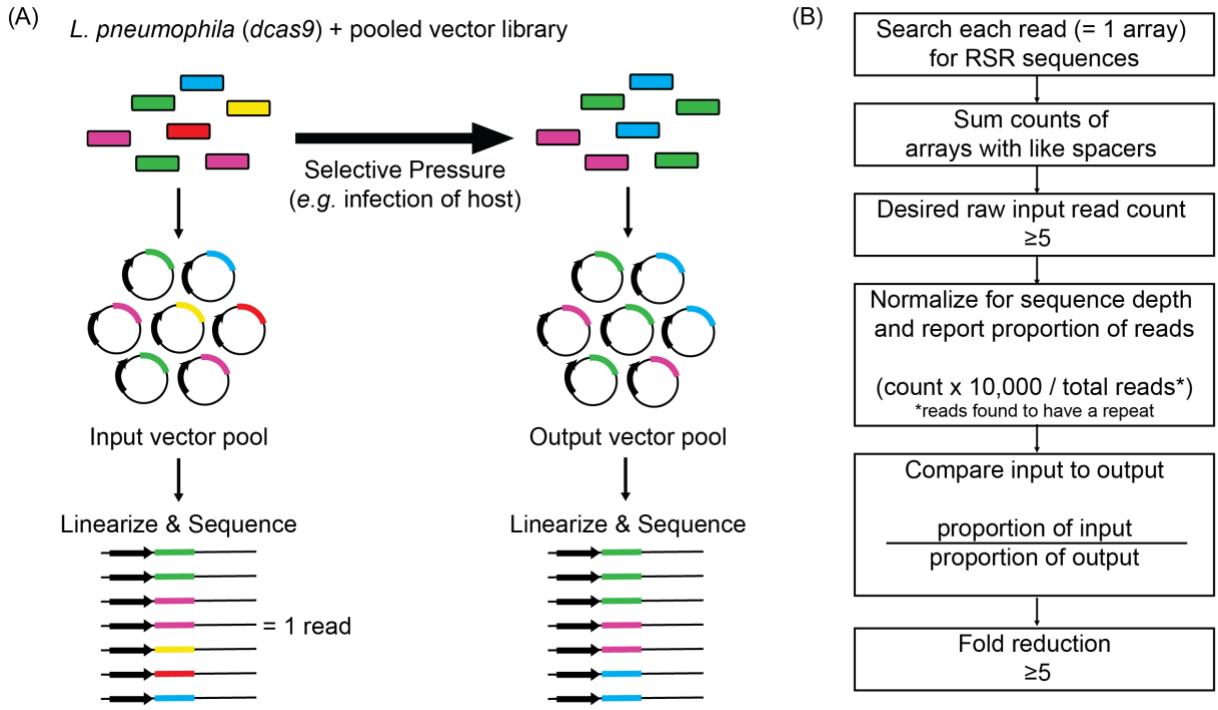
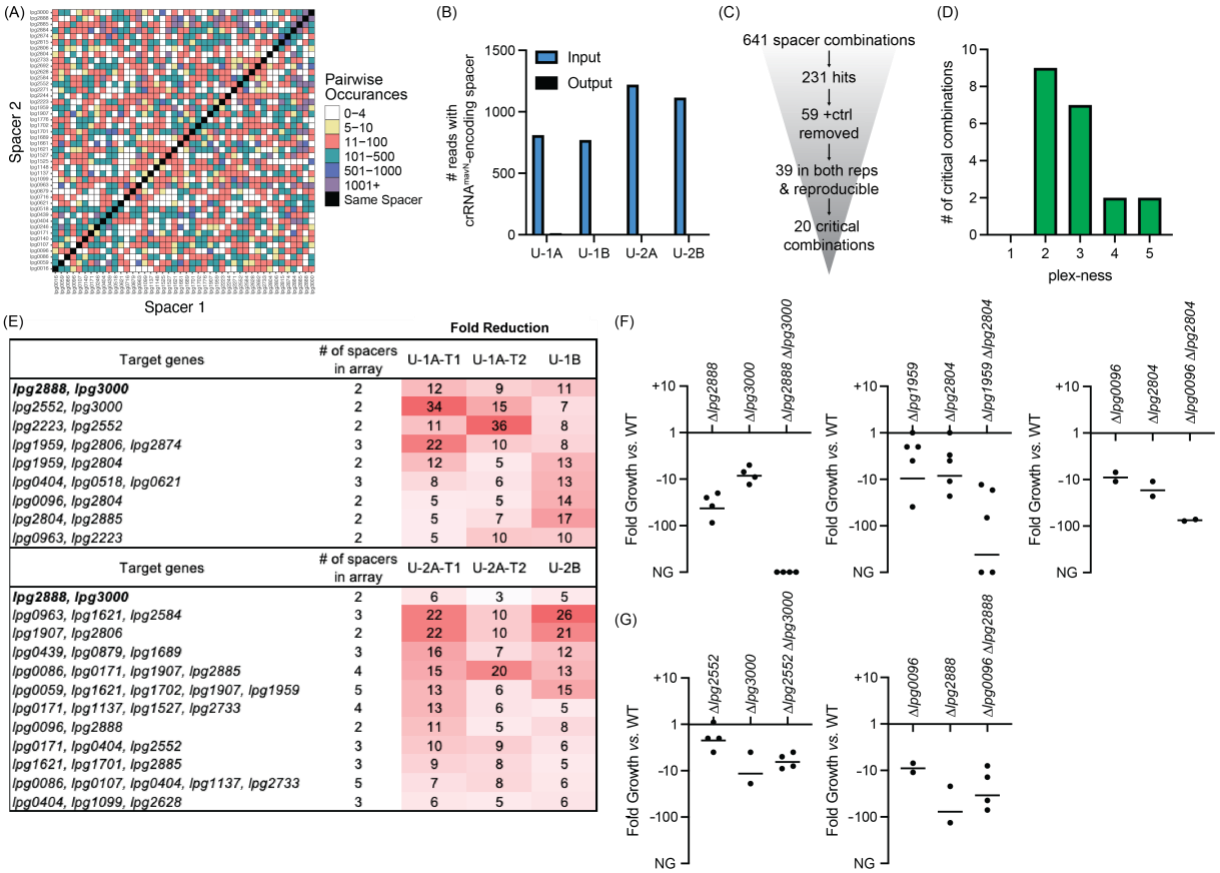


Figure 4: Schematic overview of the experimental and bioinformatics pipeline of MuRCiS (A) During MuRCiS, a pooled population of *L. pneumophila* (*dcas9*) bearing the multiplex random CRISPR arrays were subjected to a selective pressure (intracellularity). Vectors were purified from both input and output bacteria populations, linearized, and submitted for long-read PacBio Sequel sequencing. For simplification, an array of mixed spacer population is shown as a single stretch of color, each color representing a different combination of spacers. In this example, the yellow and red arrays are lost in the output suggesting they silence critical combinations of genes. (B) Overview of the custom bioinformatics pipeline used to identify unique spacer combinations causing growth attenuation as defined by a fold reduction of five or more (https://github.com/GLBRC/MuRCiS_pipeline).



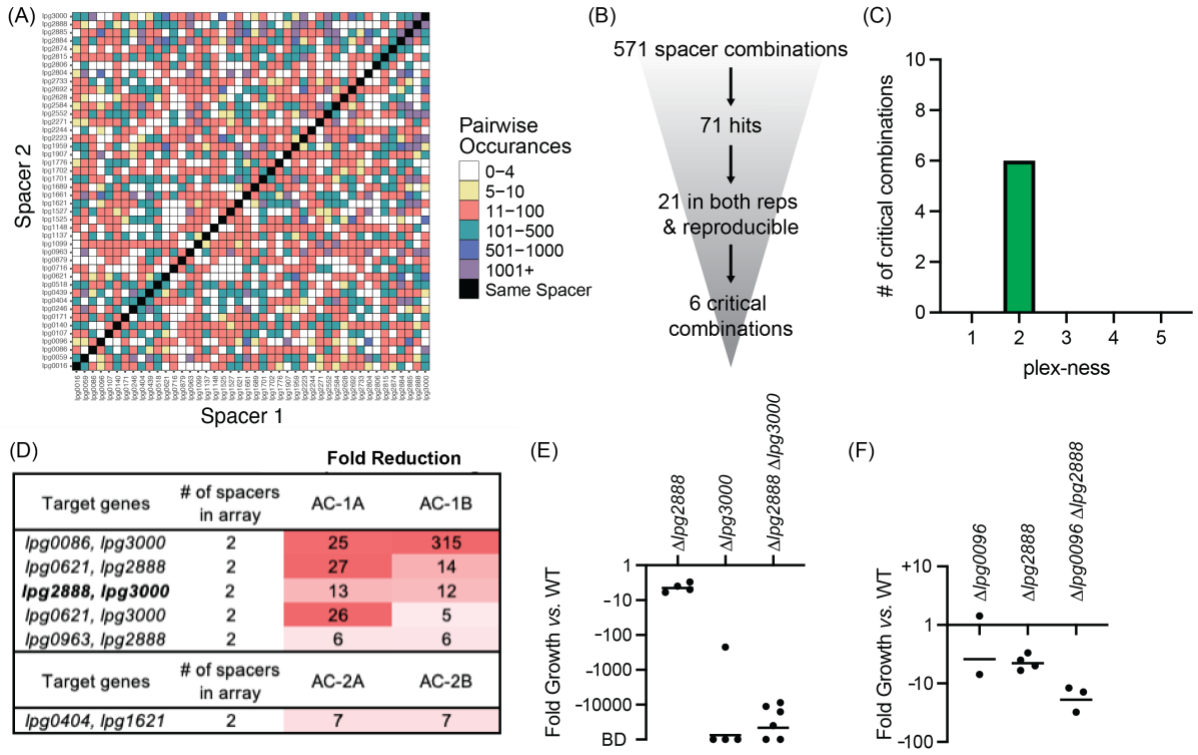
Target genes	# of spacers in array	U-1A-T1	U-1A-T2	U-1B
lpg2888, lpg3000	2	12	9	11
lpg2552, lpg3000	2	34	15	7
lpg2223, lpg2552	2	11	36	8
lpg1959, lpg2806, lpg2874	3	22	10	8
lpg1959, lpg2804	2	12	5	13
lpg0404, lpg0518, lpg0621	3	8	6	13
lpg0096, lpg2804	2	5	5	14
lpg2804, lpg2885	2	5	7	17
lpg0963, lpg2223	2	5	10	10

Target genes	# of spacers in array	U-2A-T1	U-2A-T2	U-2B
lpg2888, lpg3000	2	6	3	5
lpg0963, lpg1621, lpg2584	3	22	10	26
lpg1907, lpg2806	2	22	10	21
lpg0439, lpg0879, lpg1689	3	16	7	12
lpg0086, lpg0171, lpg1907, lpg2885	4	15	20	13
lpg0059, lpg1621, lpg1702, lpg1907, lpg1959	5	13	6	15
lpg0171, lpg1137, lpg1527, lpg2733	4	13	6	5
lpg0096, lpg2888	2	11	5	8
lpg0171, lpg0404, lpg2552	3	10	9	6
lpg1621, lpg1701, lpg2885	3	9	8	5
lpg0086, lpg0107, lpg0404, lpg1137, lpg2733	5	7	8	6
lpg0404, lpg1099, lpg2628	3	6	5	6

Figure 5: Discovery of *L. pneumophila* gene combinations critical for growth in U937 macrophages

(A) Correlation grid plotting all pairwise spacer combinations. The number of times two unique spacers were present in the sequenced array library is indicated by color-coding, ranging from light (rare) to dark (abundant). Black boxes, same spacer. (B) Total number of reads bearing the spacer which encodes crRNA^{mavN}. Counts in technical output replicates were summed. (C) Flow-chart going from the number of total spacer combinations to the number of critical combinations identified. (D) Histogram indicating the plex-ness of all critical spacer combinations identified. (E) A list of virulence-critical combinations of genes and the corresponding fold reduction observed for each in each experiment. Library 1 results are shown above Library 2 results. (F) Intracellular growth assays of strains with deletions in the indicated genes. Results are given as

fold growth (CFUs harvested 72 hpi *vs.* 2 hpi) compared to that of Lp02 (WT). Horizontal bars indicate mean fold growth of the deletion strain *vs.* WT strain for two or more experimental replicates. NG, no growth (arbitrarily set to -1000). (G) Intracellular growth assays of deletion strains that do not have a synergistic phenotype.



Target genes	# of spacers in array	AC-1A	AC-1B
lpg0086, lpg3000	2	25	315
lpg0621, lpg2888	2	27	14
lpg2888, lpg3000	2	13	12
lpg0621, lpg3000	2	26	5
lpg0963, lpg2888	2	6	6
Target genes	**# of spacers in array**	**AC-2A**	**AC-2B**
lpg0404, lpg1621	2	7	7

Figure 6: Discovery of *L. pneumophila* gene combinations critical for growth in *A. castellanii*

(A) Correlation grid plotting all pairwise spacer combinations. The number of times two unique spacers were present in the sequenced array library is indicated by color-coding, ranging from light (rare) to dark (abundant). Black boxes, same spacer. (B) Flow-chart going from the number of total spacer combinations to the number of critical combinations identified. (C) Histogram indicating the plex-ness of all critical spacer combinations identified. (D) A list of virulence-critical combinations of genes and the corresponding fold reduction observed for each in each experiment. Library 1 results are shown above Library 2 results. (E) Intracellular growth assays of strains +pMME2400 with deletions in the indicated genes. Results are given as fold growth (CFUs harvested 48 hpi *vs.* 2 hpi) compared to that of Lp02 +pMME2400 (WT). Horizontal bars indicate mean fold growth of the deletion strain *vs.* WT strain for two or more experimental

replicates. BD, below detection (arbitrarily set to -100,000). (G) Intracellular growth assays of strains with deletions in the indicated genes having been identified in single-round CRISPRi experiments.

References

1. Nowak, M.A., Boerlijst, M.C., Cooke, J. & Smith, J.M. Evolution of genetic redundancy. *Nature* **388**, 167-171 (1997).
2. Ghosh, S. & O'Connor, T.J. Beyond Paralogs: The Multiple Layers of Redundancy in Bacterial Pathogenesis. *Front Cell Infect Microbiol* **7**, 467 (2017).
3. Louca, S. et al. Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936-943 (2018).
4. Ryan, R.P., Fouhy, Y., Lucey, J.F. & Dow, J.M. Cyclic di-GMP signaling in bacteria: recent advances and new puzzles. *J Bacteriol* **188**, 8327-8334 (2006).
5. Levi, A., Folcher, M., Jenal, U. & Shuman, H.A. Cyclic diguanylate signaling proteins control intracellular growth of *Legionella pneumophila*. *MBio* **2**, e00316-00310 (2011).
6. Römling, U., Galperin, M.Y. & Gomelsky, M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**, 1-52 (2013).
7. McCarty, N.S., Graham, A.E., Studená, L. & Ledesma-Amaro, R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications* **11**, 1281 (2020).
8. Adiego-Perez, B. et al. Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS Microbiol Lett* **366** (2019).
9. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586 (2012).
10. Hille, F. et al. The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239-1259 (2018).
11. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
12. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911 (2018).
13. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).
14. Vigouroux, A. & Bikard, D. CRISPR Tools To Control Gene Expression in Bacteria. *Microbiol Mol Biol Rev* **84** (2020).
15. Peters, J.M. et al. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493-1506 (2016).
16. Peters, J.M. et al. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat Microbiol* **4**, 244-250 (2019).
17. Ellis, N.A., Kim, B., Tung, J. & Machner, M.P. A multiplex CRISPR interference tool for virulence gene interrogation in *Legionella pneumophila*. *Commun Biol* **4**, 157 (2021).
18. Liao, C. et al. Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat Commun* **10**, 2948 (2019).
19. Cress, B.F. et al. CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in E. coli. *ACS Synth Biol* **4**, 987-1000 (2015).
20. Hawkins, J.S., Wong, S., Peters, J.M., Almeida, R. & Qi, L.S. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). *Methods Mol Biol* **1311**, 349-362 (2015).

21. Vad-Nielsen, J., Lin, L., Bolund, L., Nielsen, A.L. & Luo, Y. Golden Gate Assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell Mol Life Sci* **73**, 4315-4325 (2016).
22. Deaner, M., Holzman, A. & Alper, H.S. Modular Ligation Extension of Guide RNA Operons (LEGO) for Multiplexed dCas9 Regulation of Metabolic Pathways in Saccharomyces cerevisiae. *Biotechnol J* **13**, e1700582 (2018).
23. McCarty, N.S., Shaw, W.M., Ellis, T. & Ledesma-Amaro, R. Rapid Assembly of gRNA Arrays via Modular Cloning in Yeast. *ACS Synth Biol* **8**, 906-910 (2019).
24. Shaw, W.M. et al. Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nat Commun* **13**, 4984 (2022).
25. McDade, J.E. et al. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med* **297**, 1197-1203 (1977).
26. Fraser, D.W. et al. Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med* **297**, 1189-1197 (1977).
27. Burstein, D. et al. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nat Genet* **48**, 167-175 (2016).
28. Gomez-Valero, L. et al. More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. *Proc Natl Acad Sci U S A* **116**, 2265-2273 (2019).
29. Ensminger, A.W. & Isberg, R.R. *Legionella pneumophila* Dot/Icm translocated substrates: a sum of parts. *Curr Opin Microbiol* **12**, 67-73 (2009).
30. Isberg, R.R., O'Connor, T.J. & Heidtman, M. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* **7**, 13-24 (2009).
31. Chauhan, D. & Shames, S.R. Pathogenicity and Virulence of *Legionella*: Intracellular replication and host response. *Virulence* **12**, 1122-1144 (2021).
32. O'Connor, T.J., Adepoju, Y., Boyd, D. & Isberg, R.R. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. *Proc Natl Acad Sci U S A* **108**, 14733-14740 (2011).
33. Shames, S.R. et al. Multiple *Legionella pneumophila* effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc Natl Acad Sci U S A* **114**, E10446-E10454 (2017).
34. Zimmermann, L. et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
35. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580 (2001).
36. Belyi, Y., Tabakova, I., Stahl, M. & Aktories, K. Lgt: a family of cytotoxic glucosyltransferases produced by *Legionella pneumophila*. *J Bacteriol* **190**, 3026-3035 (2008).
37. Wexler, M., Zusman, T., Linsky, M., Lifshitz, Z. & Segal, G. The *Legionella* genus core effectors display functional conservation among orthologs by themselves or combined with an accessory protein. *Curr Res Microb Sci* **3**, 100105 (2022).
38. Isaac, D.T., Laguna, R.K., Valtz, N. & Isberg, R.R. MavN is a *Legionella pneumophila* vacuole-associated protein required for efficient iron acquisition during intracellular growth. *Proc Natl Acad Sci U S A* **112**, E5208-5217 (2015).

39. Christenson, E.T. et al. The iron-regulated vacuolar *Legionella pneumophila* MavN protein is a transition-metal transporter. *Proc Natl Acad Sci U S A* **116**, 17775-17785 (2019).
40. Coers, J. et al. Identification of Icm protein complexes that play distinct roles in the biogenesis of an organelle permissive for *Legionella pneumophila* intracellular growth. *Mol Microbiol* **38**, 719-736 (2000).
41. Park, J.M., Ghosh, S. & O'Connor, T.J. Combinatorial selection in amoebal hosts drives the evolution of the human pathogen *Legionella pneumophila*. *Nature Microbiology* **5**, 599-609 (2020).
42. Feeley, J.C. et al. Charcoal-yeast extract agar: primary isolation medium for *Legionella pneumophila*. *J Clin Microbiol* **10**, 437-441 (1979).
43. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108 (2008).
44. Moffat, J.F. & Tompkins, L.S. A quantitative model of intracellular growth of *Legionella pneumophila* in *Acanthamoeba castellanii*. *Infect Immun* **60**, 296-301 (1992).
45. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
46. Wibawa, R.R., Li, P., McCaffrey, K. & Hartland, E.L. Using Genomic Deletion Mutants to Investigate Effector-Triggered Immunity During *Legionella pneumophila* Infection. *Methods Mol Biol* **2523**, 23-41 (2022).
47. Lee, P.C. & Machner, M.P. The *Legionella* Effector Kinase LegK7 Hijacks the Host Hippo Pathway to Promote Infection. *Cell Host Microbe* **24**, 429-438.e426 (2018).