Testing theoretical minimal genomes using whole-cell models

Joshua Rees-Garbutt [1,2], Lucia Marucci [1,3,4 +*] and Claire Grierson [1,2 +*]

1. BrisSynBio, University of Bristol, Bristol BS8 1TQ, UK;
2. School of Biological Sciences, University of Bristol, Bristol Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK;
3. Department of Engineering Mathematics, University of Bristol, Bristol BS8 1UB, UK;
4. School of Cellular and Molecular Medicine, University of Bristol, Bristol BS8 1UB, UK;

+ Co-last authors * Corresponding authors

Corresponding authors: Joshua Rees-Garbutt (joshua.rees@bristol.ac.uk), Prof. Claire Grierson (claire.grierson@bristol.ac.uk), Dr. Lucia Marucci (lucia.marucci@bristol.ac.uk)

Abstract

Numerous authors have pondered what the minimal gene set for life might be and many hypothetical minimal gene sets have been proposed, including at least 10 for *Mycoplasma genitalium (M.genitalium)*. None of these have been reported to be tested *in-vivo* or *in-silico*. *In-vivo* testing would be extremely difficult as *M.genitalium* is very difficult to grow in the laboratory and laborious to engineer. However, the *M.genitalium* whole-cell model provides the first opportunity to test them *in-silico*. We simulated the eight smallest hypothetical minimal gene sets from the literature, and found none could produce *in-silico* cells that could grow and divide. Using knowledge gleaned from previous research on *in-silico* gene essentiality in *M.genitalium*, we were able to repair the sets by reintroducing select genes, so that they produced dividing *in-silico* cells.

Introduction

Genome engineering builds on historical gene essentiality research. The sequencing of small bacterial genomes [1,2] led to comparative genomics, initially between pairs of bacteria [3], then greater numbers of bacteria as genome sequencing increased, which led to the development of minimal gene sets [3–12].

Minimal gene sets are lists of genes selected to produce a minimal genome, but have yet to be tested. A minimal genome is a reduced genome containing only the genetic material essential for survival, with an appropriately rich medium and no external

stresses. No single gene can be removed without loss of viability [13]. Minimal gene sets and minimal genomes focus on protein-coding genes ignoring: essential promoter regions, tRNAs, small noncoding RNAs [14], regulatory noncoding sequences [10], and the physical layout of the genome [10,15]. Predictions for the size of a viable, generic, bacterial minimal genome range from 151 genes [7], to between 300 and 500 genes, though up to 1000 genes if the cell is required to survive on minimal media [15]. As the ratio of genes to base pairs is approximately one gene per kilobase [16], the size expectation of a minimal genome is between 0.15mb and 0.5mb.

M.genitalium is the focal point of minimal gene set creation due it is naturally small genome size and available sequenced genome [2]. It has the smallest genome of an independent organism in nature at 0.58mb and 525 genes [2]. As a human parasite that has shed functional redundancies over evolutionary time it has proved a useful starting point for comparative genomics. The estimated number of essential genes ranges from: 256 by comparative genomics [3]; 388 by global transposon mutagenesis and comparative genomics [4]; and 381 by single gene knockout [6]. Extrapolating a comparison of single gene deletions of *Mycoplasma* genomes and the genome of *JCVI-syn3.0*, resulted in a prediction of 413 genes for a minimal *Mycoplasma* genome [13]. The number of genes that do not have an annotated function has been reported as 111 [4] and 134 [6]. The *M.genitalium* genome has been reproduced elsewhere, constructed from 25 synthetic parts within yeast [17]. It is also the subject of the first computational whole-cell model [8].Minimal gene sets are constructed using three different approaches: protocell development [18]; universal minimal genome theory or comparative genomics; and single gene essentiality research.

A simple definition of a cell as “living” is if it can reproduce, an “essential” gene being indispensable for cell division. A “non-essential” gene can be removed without preventing division [19,20]. But a cell’s need for specific genes and their products is dependent on the external cellular environment (i.e. how cells are grown [21]) and on the genomic context [19], which is the presence or absence of other genes in the genome, and resulting gene products, that can change each time a gene is removed. A recent review [19] updates gene essentiality from a binary categorisation to a gradient with four

categories: no essentiality (if dispensable in all contexts), low essentiality (if dispensable in some contexts, i.e. redundant essential and complexes), high essentiality (if indispensable in most contexts, i.e. protective essential), and complete essentiality (if indispensable in all contexts).

Minimal gene sets designed as protocells are not expected to function as full cells, instead functioning as a self-replicating, membrane-encapsulated collection of biomolecules [7], with the sets containing very small numbers of genes.

The universal minimal genome concept is a theory that comparing bacterial genome sequences for common genes will give a gene list that represents essential functions of the cell, and may resemble LUCA (the Last Universal Common Ancestor for life on Earth) [22]. This has been used to construct minimal gene sets, however, as the number of genomes sequenced have increased, the hope around discovering a universal minimal genome strictly from genetic sequences has decreased. Lagesen [23] found that only four genes are recognisably conserved among 1000 bacterial genomes, and even among the evolutionarily reduced *Mycoplasmas* only 196 orthologs (genes that have evolved differently from an ancestral gene but are still recognisable related and retain the same function) were found across the 20 species sequenced [13]. This apparent low conservation of cellular functions is due to non-orthologous gene displacements, independently evolved or diverged proteins that perform the same function but are not recognisably related [3,13]. This means that minimal gene sets designed by the universal minimal genome concept or comparative genomics (subsequently referred to as comparative genomics) could remove a large numbers of genes essential to *M.genitalium* depending on the number of bacterial genomes compared, as the genes are not required by the other bacterial species. Fewer genes may be removed if a smaller number of genomes are compared. This comparative work continues to be built on computationally, analysing the growing number of genomic data sets for key features that could be used to match non-orthologous gene displacements [24].

Minimal gene sets designed using single gene essentiality experiments should, in theory, not remove any essential genes, but do fall prey to issues with transposon

mutagenesis,with differing transposon variants, antibiotic resistance genes, and growth periods producing different essentiality classification for genes [25,26].

There are eight minimal gene sets in the literature with detailed gene lists (Table 1). Two are designed as protocells: Tomita *et al.* [5] and Church *et al.* [7]. Three are designed from comparative genomics: Mushegian and Koonin [3], Huang *et al.* [9], and Gil [10,11]. Three are designed from single gene essentiality experiments: Hutchison *et al.* [4], Glass *et al.* [6], and Karr *et al.* [8]. Due to the difficulty of using *M.genitalium* in the lab [27], combined with its long replication time of 12 - 15 hours [13,28,29], none of these minimal gene sets have been tested as minimal genomes, even with modern techniques [28].

Results

Ten minimal gene sets were found in the literature that were designed with *M.genitalium* genes [3–12], however two sets [11,12] were excluded as they were considered derivative of the Gil 2014 set [10] being identical apart from four genes in the Shuler *et al.* set (MG_056, MG_146, MG_388, MG_391) and four genes not in the Gil *et al.* 2004 set (MG_009, MG_091, MG_132, MG_460).

Minimal Gene Set	Code Name	Design Methodology
Mushegian and Koonin 1996 [3]	Bethesda	Comparative Genomics
Hutchison *et al.* 1999 [4]	Rockville	Single Gene Deletions
Tomita *et al.* 1999 [5]	Fujisawa	Protocell
Glass *et al.* 2006 [6]	Rockville 2	Single Gene Deletions
Forster and Church 2006 [7]	Nashville	Protocell
Karr *et al.* 2012 [8]	Stanford	Single Gene Deletions
Huang *et al.* 2013 [9]	Guelph	Comparative Genomics
Gil 2014 [10]	Valencia	Comparative Genomics

Table 1. Code names for the minimal gene sets from the literature.

To prevent confusion, we named the sets after the main location where the set was constructed (Table 1). The Bethesda set is a comparison of *M.genitalium* protein sequences to *Haemophilus influenzae* (representatives of gram-positive and gram-negative bacteria respectively) [3]. The Rockville set is the result of applying global transposon mutagenesis to *M.genitalium in-vivo* to identify non-essential genes [4]. The

Fujisawa set is an *in-silico* model of a hypothetical cell constructed from 127 *M.genitalium* genes using the E-Cell software [5]. The Rockville 2 set is an expansion of the original global transposon mutagenesis research on *M.genitalium*, with properly conducted isolation and characterisation of pure clonal populations [6]. The Nashville set is a list of 151 *E.coli* genes (compared to *M.genitalium* genes within the paper) to produce a chemical system capable of replication and evolution [7] . The Stanford set is the result of *in-silico* single gene knockouts conducted using the *M.genitalium* whole-cell model [8]. The Guelph set is the result of a comparative genomics analysis of 186 bacterial genomes [9]. The Valencia set compared the genome of *M.genitalium* with genetic data of five insect endosymbionts [10] .

To begin testing the eight minimal gene sets to see if they produced *in-silico* dividing cells, we had first to adapt each set for use in simulations, removing any genes unmodelled in the *M.genitalium* whole-cell model. The sets Nashville, Fujisawa and Stanford are as they appear in the literature, the others needed adapting by removing genes as follows (Table 2). Guelph had seven genes removed: six because they are not in the whole-cell model and one copy of MG_231 was removed as it was originally listed twice. Valencia had eight genes removed: seven that are not present in the *M.genitalium* whole-cell model and one copy of MG_231 was removed as it was originally listed twice. Bethesda had 15 genes removed: 13 as they are not in the whole-cell model and MG_297 and MG_336 were reduced to a single copy each as they were both originally listed twice. Rockville had 41 genes removed as they are not in the whole-cell model. Rockville 2 had 44 genes removed as they were not in the whole-cell model.

As expected, the minimal gene sets designed as protocells (Nashville, Fujisawa) have the smallest predicted *in-silico* genome size. Of the comparative genomics minimal gene sets, Guelph is substantially smaller than Valencia and Bethesda due to comparing 186 bacterial species for common genes [9], with Valencia only six species [10], and Bethesda directly comparing two species [3]. The single gene deletion minimal gene sets (Stanford, Rockville, Rockville 2) have similar numbers of *in-vivo* deletions, but Rockville and Rockville 2 have the highest numbers of genes that are missing from the

M.genitalium whole-cell model. This is due to the nature of exploratory genetic work, genes can be disrupted *in-vivo* whether or not the gene function is known. To be implemented *in-silico*, however, the function of the genes also must be known. All the genes in Stanford are contained in the whole-cell model, as the single gene deletions were conducted *in-silico* using the model and so did not target unmodelled genes. The gene content of the minimal gene sets (Table 3) and the required gene deletions to produce these minimal gene sets in the *M.genitalium* whole-cell model (Table 4) are listed.

	Design approach	*in-vivo* genome design size*	Unmodelled genes^ in genome design	Single *in-vivo* gene deletions*	Unmodelled genes^ in single gene deletions	Predicted *in-silico* genome size*	Predicted gene deletions *in-silico**
Nashville	Protocell	89	0	-	-	89	270
Fujisawa	Protocell	98	0	-	-	98	261
Guelph	Comparative Genomics	123	5	-	-	118	241
Valencia	Comparative Genomics	180	6	-	-	174	185
Bethesda	Comparative Genomics	253	12	-	-	241	118
Stanford	Single Gene Deletions	-	-	117	0	242	117
Rockville 2	Single Gene Deletions	-	-	101	44	302	57
Rockville	Single Gene Deletions	-	-	94	41	306	53
M.genitalium whole-cell Model*	-	-	124	-	-	359	-
M.genitalium in-vivo *	-	483		-	-	-	-

Table 2. Minimal Gene Sets from the literature, compared with *M.genitalium in-vivo* and the whole-cell model. *M.genitalium* has 42 RNA-coding genes that are not included in this table. * = protein-coding genes. ^ = due to unknown function.

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_001	I	I		I		I	I	I	I
MG_002						I			
MG_003	I	I		I		I	I	I	I
MG_004	I	I		I		I	I	I	I
MG_005	I	I	I	I	I	I	I	I	I
MG_006	I	I		I		I		I	
MG_007		I		I		I		I	
MG_008	I	I		I		I	I	I	I
MG_009	I								
MG_010						I			
MG_011						I			
MG_012	I	I				I			
MG_013	I	I		I		I			
MG_014				I					
MG_015	I	I		I					
MG_018						I			
MG_019	I	I		I		I	I	I	I
MG_020		I		I		I			
MG_021	I	I	I	I	I	I	I	I	I
MG_022		I		I		I			
MG_023	I	I	I	I		I		I	
MG_024	I	I				I	I	I	
MG_025				I		I			
MG_026	I	I		I	I	I		I	
MG_027		I		I					
MG_028		I		I		I			
MG_029				I					
MG_030	I	I		I				I	
MG_031		I		I		I		I	
MG_032						I			
MG_033	I		I						
MG_034		I		I		I			
MG_035	I	I	I	I	I	I	I	I	I
MG_036	I	I	I	I	I	I	I	I	I
MG_037		I		I		I			
MG_038	I	I	I	I		I			
MG_039		I							
MG_040		I							
MG_041		I	I	I		I		I	
MG_042	I	I		I		I			
MG_043	I	I		I		I			
MG_044	I	I		I		I			
MG_045	I	I		I		I			
MG_046		I		I		I	I	I	

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_047	I	I		I		I	I	I	I
MG_048	I	I		I		I	I	I	I
MG_049	I			I		I			
MG_050	I	I		I					
MG_051						I			
MG_052	I			I					
MG_053	I	I		I		I			
MG_054	I	I		I		I	I	I	I
MG_055		I		I				I	
MG_473		I		I		I			
MG_474		I		I		I			
MG_056	I	I				I	I	I	
MG_057		I		I		I			
MG_058	I	I		I		I	I	I	I
MG_059	I	I		I			I	I	
MG_060	I	I		I		I			
MG_061		I							
MG_062									
MG_063	I	I							
MG_064		I		I					
MG_065	I	I		I					
MG_066	I	I				I	I	I	
MG_067						I			
MG_068		I		I		I			
MG_069		I	I	I		I		I	
MG_070	I	I	I	I	I	I	I	I	I
MG_071	I	I		I		I			
MG_072	I	I		I		I	I	I	I
MG_073	I	I		I					
MG_074		I		I		I			
MG_075		I		I					
MG_076		I		I		I			
MG_077	I	I		I		I			
MG_078	I	I		I		I			
MG_079	I	I		I		I			
MG_080	I	I		I		I			
MG_081	I	I	I	I	I	I	I	I	I
MG_082	I	I	I	I	I	I	I	I	I
MG_083	I	I		I			I	I	
MG_084		I		I		I	I	I	
MG_085				I					
MG_086	I	I		I					
MG_087	I	I	I	I	I	I	I	I	I
MG_088	I	I	I	I	I	I	I	I	I
MG_089	I	I	I	I	I	I	I	I	I

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_090	I	I	I	I	I	I		I	
MG_091	I	I		I		I			
MG_092	I	I	I	I	I	I	I	I	I
MG_093	I	I	I	I	I	I		I	
MG_094	I	I		I		I	I	I	I
MG_095		I		I		I			
MG_096						I			
MG_097	I	I		I				I	
MG_098		I		I		I			
MG_099		I		I	I	I			
MG_100		I		I	I	I			
MG_101		I		I					
MG_102	I	I		I		I	I	I	I
MG_103						I			
MG_476		I		I		I			
MG_104	I	I		I		I			
MG_105		I		I					
MG_106	I	I		I		I			
MG_107	I	I		I		I	I	I	I
MG_108		I		I		I			
MG_109		I		I		I			
MG_110						I			
MG_111	I	I	I	I		I	I	I	I
MG_112	I	I				I	I	I	
MG_113	I	I	I	I	I	I	I	I	I
MG_114	I	I	I			I			
MG_115		I				I			
MG_116		I				I			
MG_117		I		I		I			
MG_118	I	I		I		I			
MG_119	I	I		I					
MG_120	I	I		I					
MG_121		I							
MG_122	I	I		I		I			
MG_123		I		I					
MG_124	I	I		I		I	I	I	I
MG_125	I	I		I		I			
MG_126	I	I	I	I	I	I	I	I	I
MG_127	I	I		I					
MG_128		I		I		I			
MG_129		I		I		I			
MG_130				I					
MG_131						I			
MG_132				I					
MG_133		I		I		I			

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_134		I				I			
MG_135		I		I		I			
MG_136	I	I	I	I	I	I	I	I	I
MG_137		I		I		I			
MG_138	I	I				I	I	I	
MG_139		I		I		I			
MG_140						I			
MG_141	I	I		I		I	I	I	I
MG_477				I		I			
MG_142	I	I	I	I	I	I	I	I	I
MG_143	I	I		I		I		I	
MG_144		I		I		I			
MG_145	I	I		I		I	I	I	I
MG_146		I		I		I			
MG_147		I		I		I			
MG_148		I		I		I			
MG_149									
MG_478		I				I			
MG_150	I	I	I	I	I	I	I	I	I
MG_151	I	I	I	I	I	I	I	I	I
MG_152	I	I	I	I	I	I	I	I	I
MG_153	I	I	I	I	I	I		I	
MG_154	I	I	I	I	I	I	I	I	I
MG_155	I	I	I	I	I	I	I	I	I
MG_156	I	I	I	I	I	I	I	I	I
MG_157	I	I	I	I	I	I	I	I	I
MG_158	I	I	I	I	I	I	I	I	I
MG_159	I	I	I	I	I	I		I	
MG_160	I	I	I	I	I	I	I	I	I
MG_161	I	I	I	I	I	I	I	I	I
MG_162	I	I	I	I	I	I	I	I	I
MG_163	I	I	I	I	I	I	I	I	I
MG_164	I	I	I	I	I	I		I	
MG_165	I	I	I	I	I	I	I	I	I
MG_166	I	I	I	I	I	I	I	I	I
MG_167	I	I	I	I	I	I	I	I	I
MG_168	I	I	I	I	I	I	I	I	I
MG_169	I	I		I	I	I		I	
MG_170	I	I		I		I	I	I	I
MG_171	I	I		I	I	I	I	I	I
MG_172	I	I		I		I	I	I	I
MG_173	I	I	I	I	I	I	I	I	I
MG_174	I	I	I	I	I	I		I	
MG_175	I	I	I	I	I	I	I	I	I
MG_176	I	I	I	I	I	I	I	I	I

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_177	I	I	I	I		I	I	I	I
MG_178	I	I	I	I	I	I	I	I	I
MG_179		I		I		I			
MG_180	I	I		I		I			
MG_181		I		I		I			
MG_182	I			I		I			
MG_183						I			
MG_184		I		I		I			
MG_185						I			
MG_186				I					
MG_187	I	I		I					
MG_188		I		I					
MG_189		I		I					
MG_190		I		I					
MG_191				I		I			
MG_192				I		I			
MG_194	I	I	I	I	I	I	I	I	I
MG_195	I	I		I	I	I	I	I	I
MG_196	I	I	I	I	I	I	I	I	I
MG_197	I	I	I	I	I	I		I	
MG_198	I	I	I	I	I	I	I	I	I
MG_199		I		I		I			
MG_200		I		I					
MG_201	I	I		I		I		I	
MG_202				I		I			
MG_203	I	I		I					
MG_204	I	I		I		I			
MG_205		I		I					
MG_206	I	I		I					
MG_207						I			
MG_208		I		I		I			
MG_209	I			I					
MG_210	I	I				I			
MG_480		I		I		I			
MG_481		I		I		I			
MG_211		I		I		I			
MG_482		I		I					
MG_212	I	I		I		I		I	
MG_213									
MG_214		I							
MG_215	I	I	I	I		I		I	
MG_216	I	I	I	I		I		I	
MG_217		I		I		I			
MG_218		I		I		I			
MG_491		I		I		I			

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_219		I		I		I			
MG_220		I				I			
MG_221	I	I		I		I			
MG_222	I	I		I		I	I	I	
MG_223		I		I		I			
MG_224	I	I		I		I	I	I	
MG_225		I		I					
MG_226									
MG_227	I	I						I	
MG_228	I	I		I		I		I	
MG_229	I	I		I		I		I	
MG_230		I		I		I			
MG_231	I	I		I		I	I	I	I
MG_232	I	I	I	I	I	I		I	
MG_233		I		I		I			
MG_234	I	I	I	I	I	I	I	I	I
MG_235		I		I				I	
MG_236		I		I					
MG_237						I			
MG_238	I	I				I			
MG_239	I	I		I		I			
MG_240		I		I					
MG_241		I		I		I			
MG_242		I		I		I			
MG_243		I		I		I			
MG_244	I	I							
MG_245	I	I		I		I			
MG_246		I		I		I			
MG_247	I	I		I		I			
MG_248		I				I			
MG_249	I	I	I	I		I	I	I	I
MG_250	I	I		I			I	I	
MG_251	I	I	I	I	I	I		I	
MG_252	I	I							
MG_253	I	I	I	I	I	I	I	I	I
MG_254	I	I		I		I	I	I	I
MG_255						I			
MG_494						I			
MG_256						I			
MG_257	I	I	I	I	I	I	I	I	I
MG_258	I	I		I	I	I	I	I	I
MG_259	I	I		I	I			I	
MG_260		I				I			
MG_261	I	I		I		I	I	I	I
MG_262	I	I		I			I	I	

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_498	I	I							
MG_263		I		I		I			
MG_264								I	
MG_265	I	I		I					
MG_266	I	I	I	I	I	I	I	I	I
MG_267		I		I		I			
MG_268	I					I			
MG_269						I			
MG_270	I	I		I		I			
MG_271	I	I				I			
MG_272	I	I		I		I			
MG_273	I	I		I		I			
MG_274	I	I		I		I			
MG_275	I	I		I		I			
MG_276	I	I		I		I			
MG_277		I		I		I			
MG_278	I	I		I		I			
MG_279						I			
MG_280						I			
MG_281						I			
MG_282	I	I		I		I	I	I	I
MG_283	I	I	I	I	I	I		I	
MG_284		I				I			
MG_285						I			
MG_286						I			
MG_287	I	I		I		I			
MG_288									
MG_289		I							
MG_290		I							
MG_291									
MG_505		I		I		I			
MG_292	I	I	I	I	I	I	I	I	I
MG_293	I								
MG_294						I			
MG_295	I			I		I	I	I	
MG_296				I		I			
MG_297	I	I		I		I	I	I	I
MG_298		I							
MG_299	I	I		I		I			
MG_300	I	I	I	I		I	I	I	I
MG_301	I	I	I	I		I	I	I	I
MG_302		I		I		I			
MG_303		I		I		I			
MG_304		I		I		I			
MG_305	I	I		I		I	I	I	I

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_306		I		I		I			
MG_307		I		I		I			
MG_308				I		I			
MG_309		I		I					
MG_310				I					
MG_311		I	I	I	I	I	I	I	
MG_312		I		I		I			
MG_313		I		I		I			
MG_314		I		I		I			
MG_315		I				I		I	
MG_316									
MG_317				I		I			
MG_318		I		I		I			
MG_319		I		I		I			
MG_320		I		I		I			
MG_321		I		I		I			
MG_322	I	I		I		I			
MG_323		I		I		I			
MG_515				I		I			
MG_324		I		I					
MG_325	I	I	I	I	I	I		I	
MG_326		I		I		I			
MG_327				I					
MG_328						I			
MG_329		I		I		I	I	I	
MG_330	I	I		I		I			
MG_331		I		I		I			
MG_332	I	I		I		I			
MG_333	I	I		I					
MG_334	I	I	I	I	I	I	I	I	I
MG_335	I			I		I			
MG_516		I		I		I			
MG_517		I		I		I			
MG_336	I	I		I				I	
MG_337		I		I		I			
MG_338				I		I			
MG_339	I								
MG_340	I	I	I	I		I	I	I	I
MG_341	I	I	I	I		I	I	I	I
MG_342		I		I		I			
MG_343						I			
MG_344		I	I	I					
MG_345	I		I	I	I	I	I	I	
MG_346	I							I	
MG_347	I	I		I		I			

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_348		I		I		I			
MG_349		I		I					
MG_350		I		I		I			
MG_521		I		I		I			
MG_351	I	I	I	I	I	I		I	
MG_352									
MG_353	I	I		I				I	
MG_354		I		I		I			
MG_355	I	I							
MG_356		I		I					
MG_357	I	I		I		I			
MG_358	I	I							
MG_359	I	I							
MG_360		I				I			
MG_361	I	I	I	I	I	I	I	I	I
MG_362	I	I	I	I	I	I	I	I	I
MG_363	I	I	I	I	I	I		I	
MG_522	I	I	I	I	I	I		I	
MG_364		I		I		I			
MG_365	I	I	I	I	I	I			
MG_366				I		I			
MG_367	I	I				I	I	I	
MG_368		I		I		I			
MG_369		I		I					
MG_370									
MG_371		I		I		I			
MG_372				I		I			
MG_373		I		I		I			
MG_374		I		I		I			
MG_375	I	I	I	I	I	I		I	
MG_376		I		I					
MG_377		I		I		I			
MG_378	I	I	I	I	I	I	I	I	I
MG_379	I	I		I		I		I	
MG_380	I								
MG_381		I		I		I			
MG_382	I	I		I		I			
MG_383	I	I		I		I			
MG_384	I	I		I		I	I	I	I
MG_524		I		I		I			
MG_385									
MG_386		I		I		I			
MG_387	I	I		I		I		I	
MG_388		I		I		I		I	
MG_389		I		I		I			

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_390									
MG_391	I	I		I		I		I	
MG_392	I	I		I	I	I	I	I	I
MG_393	I	I		I	I	I		I	
MG_394	I			I		I	I	I	
MG_395				I		I			
MG_396		I		I		I		I	
MG_397		I				I			
MG_398	I	I						I	
MG_399	I	I		I			I	I	
MG_400	I	I		I			I	I	
MG_401	I	I		I			I	I	
MG_402	I	I		I				I	
MG_403	I	I		I				I	
MG_404	I	I		I				I	
MG_405	I	I		I				I	
MG_406	I	I		I		I			
MG_407	I	I	I	I		I	I	I	I
MG_408	I								
MG_409		I		I					
MG_410	I								
MG_411	I								
MG_412									
MG_414						I			
MG_525						I			
MG_417	I	I	I	I	I	I	I	I	I
MG_418	I	I	I	I	I	I	I	I	I
MG_419		I		I		I			
MG_421	I			I					
MG_422		I		I		I			
MG_423				I		I			
MG_424	I	I	I	I	I	I	I	I	I
MG_425	I	I		I		I	I	I	I
MG_426	I		I	I	I	I		I	
MG_427		I		I					
MG_428									
MG_429		I	I	I		I	I	I	
MG_430	I	I	I	I		I		I	
MG_431	I	I	I	I		I	I	I	I
MG_432		I		I		I			
MG_433	I	I	I	I	I	I	I	I	I
MG_434	I	I		I		I			
MG_435	I	I		I	I	I	I	I	I
MG_437	I	I	I			I		I	
MG_438									

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Common
MG_439		I		I		I			
MG_440				I		I			
MG_441		I		I		I			
MG_442				I		I			
MG_443		I		I		I			
MG_444	I	I	I	I	I	I	I	I	I
MG_445	I	I		I	I	I	I		
MG_446	I	I	I	I	I	I	I	I	I
MG_447		I		I					
MG_448	I	I		I					
MG_449	I					I			
MG_450		I		I		I			
MG_451	I	I	I	I	I	I	I	I	I
MG_452						I			
MG_453	I	I		I		I			
MG_454		I		I					
MG_455	I	I	I	I	I	I	I	I	I
MG_456		I				I			
MG_457	I	I		I			I	I	
MG_458	I	I		I		I		I	
MG_459		I		I		I			
MG_460		I	I	I					
MG_461		I		I		I			
MG_462	I	I	I	I	I	I	I	I	I
MG_463	I	I		I			I	I	
MG_464		I		I		I		I	
MG_465	I	I		I	I	I		I	
MG_466	I	I	I	I	I	I		I	
MG_467				I					
MG_468				I					
MG_526		I		I					
MG_469	I	I		I		I	I		
MG_470				I					

Table 3. Comparing the gene content of the minimal gene sets. Light grey genes are unmodelled in the *M.genitalium* whole-cell model. Dark grey genes cause the simulation to crash when deleted on our implementation (using Matlab R2013B on University of Bristol BlueGem's supercomputer). I = gene included in the minimal gene set. Common = genes that the minimal gene sets agree upon, but excluding the protocell designed sets.

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_001			x		x				
MG_002		x		x					
MG_003			x		x				
MG_004			x		x				
MG_005									
MG_006			x		x		x		
MG_007	x		x		x		x		
MG_008			x		x				
MG_009		x	x	x	x	x	x	x	
MG_010		x		x					
MG_011		x		x					
MG_012			x	x	x		x	x	
MG_013			x		x		x	x	
MG_014	x	x	x		x	x	x	x	
MG_015			x		x	x	x	x	
MG_018		x		x					
MG_019			x		x				
MG_020	x		x		x		x	x	
MG_021									
MG_022	x		x		x		x	x	
MG_023					x		x		
MG_024				x					
MG_025		x							
MG_026			x				x		
MG_027	x		x		x	x	x	x	
MG_028									
MG_029	x	x	x		x	x	x	x	
MG_030			x		x	x	x		
MG_031	x		x		x		x		
MG_032		x		x					
MG_033		x		x	x	x	x	x	
MG_034	x		x		x		x	x	
MG_035									
MG_036									
MG_037	x		x		x		x	x	
MG_038					x		x	x	
MG_039	x		x	x	x	x	x	x	
MG_040	x		x	x	x	x	x	x	
MG_041	x				x		x		
MG_042			x		x		x	x	
MG_043			x		x		x	x	
MG_044			x		x		x	x	
MG_045			x		x		x	x	
MG_046	x		x		x				
MG_047			x		x				

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_048			x		x				
MG_049		x	x		x		x	x	
MG_050			x		x	x	x	x	
MG_051	x	x	x	x	x		x	x	
MG_052		x	x		x	x	x	x	
MG_053			x		x		x	x	
MG_054									
MG_055	x		x		x	x	x		
MG_473	x		x		x		x	x	
MG_474									
MG_056				x					
MG_057									
MG_058			x		x				
MG_059			x		x	x			
MG_060									
MG_061	x		x	x	x	x	x	x	
MG_062	x	x	x	x	x	x	x	x	x
MG_063			x	x	x	x	x	x	
MG_064	x		x		x	x	x	x	
MG_065			x		x	x	x	x	
MG_066			x	x	x				
MG_067		x		x					
MG_068									
MG_069	x				x		x		
MG_070									
MG_071			x		x		x	x	
MG_072			x		x				
MG_073			x		x	x	x	x	
MG_074									
MG_075	x		x		x	x	x	x	
MG_076									
MG_077			x		x		x	x	
MG_078			x		x		x	x	
MG_079			x		x		x	x	
MG_080			x		x		x	x	
MG_081									
MG_082									
MG_083			x		x	x			
MG_084	x		x		x				
MG_085	x	x	x		x	x	x	x	
MG_086			x		x	x	x	x	
MG_087									
MG_088									
MG_089									
MG_090							x		

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_091			x		x		x	x	
MG_092									
MG_093							x		
MG_094			x		x				
MG_095									
MG_096		x		x					
MG_097			x		x	x	x		
MG_098	x		x		x		x	x	
MG_099	x		x				x	x	
MG_100	x		x				x	x	
MG_101	x		x		x	x	x	x	
MG_102			x		x				
MG_103		x		x					
MG_476	x		x		x		x	x	
MG_104			x		x		x	x	
MG_105	x		x		x	x	x	x	
MG_106			x		x		x	x	
MG_107			x		x				
MG_108									
MG_109	x		x		x		x	x	
MG_110	x	x	x	x	x		x	x	
MG_111					x				
MG_112			x	x	x				
MG_113									
MG_114				x	x		x	x	
MG_115				x					
MG_116				x					
MG_117									
MG_118			x		x		x	x	
MG_119			x		x	x	x	x	
MG_120			x		x	x	x	x	
MG_121	x		x	x	x	x	x	x	
MG_122			x		x		x	x	
MG_123	x		x		x	x	x	x	
MG_124			x		x				
MG_125									
MG_126									
MG_127			x		x	x	x	x	
MG_128	x		x		x		x	x	
MG_129									
MG_130	x	x	x		x	x	x	x	
MG_131		x		x					
MG_132	x	x	x		x	x	x	x	
MG_133									
MG_134				x					

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_135									
MG_136									
MG_137	x		x		x		x	x	
MG_138				x					
MG_139	x		x		x		x	x	
MG_140		x		x					
MG_141			x		x				
MG_477		x							
MG_142									
MG_143			x		x		x		
MG_144									
MG_145			x		x				
MG_146									
MG_147									
MG_148									
MG_149	x	x	x	x	x	x	x	x	x
MG_478				x					
MG_150									
MG_151									
MG_152									
MG_153							x		
MG_154									
MG_155									
MG_156									
MG_157									
MG_158									
MG_159							x		
MG_160									
MG_161									
MG_162									
MG_163									
MG_164							x		
MG_165									
MG_166									
MG_167									
MG_168									
MG_169			x				x		
MG_170			x		x				
MG_171			x						
MG_172			x		x				
MG_173									
MG_174							x		
MG_175									
MG_176									
MG_177					x				

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_178									
MG_179	x		x		x		x	x	
MG_180			x		x		x	x	
MG_181	x		x		x		x	x	
MG_182		x	x		x		x	x	
MG_183	x	x	x	x	x		x	x	
MG_184	x		x		x		x	x	
MG_185		x		x					
MG_186	x	x	x		x	x	x	x	
MG_187			x		x	x	x	x	
MG_188	x		x		x	x	x	x	
MG_189	x		x		x	x	x	x	
MG_190	x		x		x	x	x	x	
MG_191	x	x	x		x		x	x	
MG_192	x	x	x		x		x	x	
MG_194									
MG_195			x						
MG_196									
MG_197							x		
MG_198									
MG_199									
MG_200	x		x		x	x	x	x	
MG_201			x		x		x		
MG_202		x							
MG_203			x		x	x	x	x	
MG_204			x		x		x	x	
MG_205	x		x		x	x	x	x	
MG_206			x		x	x	x	x	
MG_207		x		x					
MG_208	x		x		x		x	x	
MG_209		x	x		x	x	x	x	
MG_210			x	x	x		x	x	
MG_480									
MG_481	x		x		x		x	x	
MG_211									
MG_482	x		x		x	x	x	x	
MG_212			x		x		x		
MG_213	x	x	x	x	x	x	x	x	x
MG_214	x		x	x	x	x	x	x	
MG_215					x		x		
MG_216					x		x		
MG_217	x		x		x		x	x	
MG_218	x		x		x		x	x	
MG_491									
MG_219									

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_220				x					
MG_221			x		x		x	x	
MG_222									
MG_223									
MG_224			x		x				
MG_225	x		x		x	x	x	x	
MG_226	x	x	x	x	x	x	x	x	x
MG_227			x	x	x	x	x		
MG_228			x		x		x		
MG_229			x		x		x		
MG_230	x		x		x		x	x	
MG_231			x		x				
MG_232							x		
MG_233									
MG_234									
MG_235	x		x		x	x	x		
MG_236	x		x		x	x	x	x	
MG_237		x		x					
MG_238			x	x	x		x	x	
MG_239			x		x		x	x	
MG_240	x		x		x	x	x	x	
MG_241									
MG_242									
MG_243									
MG_244			x	x	x	x	x	x	
MG_245			x		x		x	x	
MG_246									
MG_247									
MG_248				x					
MG_249					x				
MG_250			x		x	x			
MG_251							x		
MG_252			x	x	x	x	x	x	
MG_253									
MG_254			x		x				
MG_255		x		x					
MG_494		x		x					
MG_256		x		x					
MG_257									
MG_258			x						
MG_259			x			x	x		
MG_260				x					
MG_261			x		x				
MG_262			x		x	x			
MG_498			x	x	x	x	x	x	

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_263									
MG_264	x	x	x	x	x	x	x		
MG_265			x		x	x	x	x	
MG_266									
MG_267									
MG_268		x		x					
MG_269		x		x					
MG_270			x		x		x	x	
MG_271			x	x	x		x	x	
MG_272			x		x		x	x	
MG_273			x		x		x	x	
MG_274			x		x		x	x	
MG_275			x		x		x	x	
MG_276			x		x		x	x	
MG_277	x		x		x		x	x	
MG_278			x		x		x	x	
MG_279		x		x					
MG_280		x		x					
MG_281		x		x					
MG_282			x		x				
MG_283							x		
MG_284				x					
MG_285		x		x					
MG_286		x		x					
MG_287			x		x		x	x	
MG_288	x	x	x	x	x	x	x	x	x
MG_289	x		x	x	x	x	x	x	
MG_290	x		x	x	x	x	x	x	
MG_291	x	x	x	x	x	x	x	x	x
MG_505									
MG_292									
MG_293		x	x	x	x	x	x	x	
MG_294		x		x					
MG_295		x	x		x				
MG_296		x							
MG_297			x		x				
MG_298	x		x	x	x	x	x	x	
MG_299			x		x		x	x	
MG_300					x				
MG_301					x				
MG_302	x		x		x		x	x	
MG_303	x		x		x		x	x	
MG_304	x		x		x		x	x	
MG_305			x		x				
MG_306									

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_307									
MG_308		x							
MG_309	x		x		x	x	x	x	
MG_310	x	x	x		x	x	x	x	
MG_311	x								
MG_312	x		x		x		x	x	
MG_313									
MG_314									
MG_315	x		x	x	x		x		
MG_316	x	x	x	x	x	x	x	x	x
MG_317	x	x	x		x		x	x	
MG_318	x		x		x		x	x	
MG_319									
MG_320									
MG_321	x		x		x		x	x	
MG_322			x		x		x	x	
MG_323	x		x		x		x	x	
MG_515		x							
MG_324	x		x		x	x	x	x	
MG_325							x		
MG_326									
MG_327	x	x	x		x	x	x	x	
MG_328		x		x					
MG_329	x		x		x				
MG_330			x		x		x	x	
MG_331									
MG_332									
MG_333			x		x	x	x	x	
MG_334									
MG_335		x	x		x		x	x	
MG_516									
MG_517	x		x		x		x	x	
MG_336			x		x	x	x		
MG_337									
MG_338		x							
MG_339		x	x	x	x	x	x	x	
MG_340					x				
MG_341					x				
MG_342	x		x		x		x	x	
MG_343		x		x					
MG_344	x				x	x	x	x	
MG_345		x							
MG_346		x	x	x	x	x	x		
MG_347			x		x		x	x	
MG_348									

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_349	x		x		x	x	x	x	
MG_350									
MG_521									
MG_351							x		
MG_352	x	x	x	x	x	x	x	x	x
MG_353			x		x	x	x		
MG_354									
MG_355			x	x	x	x	x	x	
MG_356	x		x		x	x	x	x	
MG_357			x		x		x	x	
MG_358			x	x	x	x	x	x	
MG_359			x	x	x	x	x	x	
MG_360				x					
MG_361									
MG_362									
MG_363							x		
MG_522							x		
MG_364									
MG_365							x	x	
MG_366		x							
MG_367			x	x	x				
MG_368	x		x		x		x	x	
MG_369	x		x		x	x	x	x	
MG_370	x	x	x	x	x	x	x	x	x
MG_371									
MG_372	x	x	x		x		x	x	
MG_373									
MG_374									
MG_375							x		
MG_376	x		x		x	x	x	x	
MG_377									
MG_378									
MG_379			x		x		x		
MG_380		x	x	x	x	x	x	x	
MG_381									
MG_382			x		x		x	x	
MG_383			x		x		x	x	
MG_384			x		x				
MG_524									
MG_385	x	x	x	x	x	x	x	x	x
MG_386	x		x		x		x	x	
MG_387			x		x		x		
MG_388									
MG_389									
MG_390	x	x	x	x	x	x	x	x	x

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_391			x		x		x		
MG_392			x						
MG_393			x				x		
MG_394		x	x		x				
MG_395		x							
MG_396	x		x		x		x		
MG_397				x					
MG_398			x	x	x	x	x		
MG_399			x		x	x			
MG_400			x		x	x			
MG_401			x		x	x			
MG_402			x		x	x	x		
MG_403			x		x	x	x		
MG_404			x		x	x	x		
MG_405			x		x	x	x		
MG_406									
MG_407					x				
MG_408		x	x	x	x	x	x	x	
MG_409	x		x		x	x	x	x	
MG_410		x	x	x	x	x	x	x	
MG_411		x	x	x	x	x	x	x	
MG_412	x	x	x	x	x	x	x	x	x
MG_414		x		x					
MG_525		x		x					
MG_417									
MG_418									
MG_419	x		x		x		x	x	
MG_421		x	x		x	x	x	x	
MG_422									
MG_423		x							
MG_424									
MG_425			x		x				
MG_426		x					x		
MG_427	x		x		x	x	x	x	
MG_428	x	x	x	x	x	x	x	x	x
MG_429	x				x				
MG_430					x		x		
MG_431					x				
MG_432									
MG_433									
MG_434			x		x		x	x	
MG_435			x						
MG_437				x	x		x		
MG_438	x	x	x	x	x	x	x	x	x
MG_439									

	Bethesda	Rockville	Fujisawa	Rockville 2	Nashville	Stanford	Guelph	Valencia	Agreed
MG_440		x							
MG_441									
MG_442	x	x	x		x		x	x	
MG_443									
MG_444									
MG_445			x					x	
MG_446									
MG_447	x		x		x	x	x	x	
MG_448			x		x	x	x	x	
MG_449		x		x					
MG_450									
MG_451									
MG_452		x		x					
MG_453			x		x		x	x	
MG_454	x		x	x	x	x	x	x	
MG_455									
MG_456				x					
MG_457			x		x	x			
MG_458			x		x		x		
MG_459									
MG_460	x			x	x	x	x	x	
MG_461									
MG_462									
MG_463			x	x	x	x			
MG_464	x		x		x		x		
MG_465			x				x		
MG_466							x		
MG_467	x	x	x		x	x	x	x	
MG_468	x	x	x		x	x	x	x	
MG_526	x		x		x	x	x	x	
MG_469			x		x			x	
MG_470	x	x	x		x	x	x	x	

Table 4. Comparing the gene deletions of the minimal gene sets. Light grey genes are unmodelled in the *M.genitalium* whole-cell model. Dark grey genes cause the simulation to crash when deleted on our implementation (using Matlab R2013B on University of Bristol BlueGem's supercomputer). x = a gene deletion required to produce the minimal gene set in the *M.genitalium* whole-cell model.

Analysing the Minimal Gene Sets

A comparison of six of the minimal gene sets (the protocell minimal gene sets (Nashville, Fujisawa) were excluded due to their much more reduced size) showed that they had 96 genes in common (Table 3, Common Column). Of these 96 genes, 87 were classified as essential by single gene deletion, eight were non-essential, and one gene was unmodelled (Supplementary Information 1). The 87 essential genes identify the cellular functional groups that the six minimal gene sets agree are essential: DNA (repair, supercoiling, chromosome replication, synthesis and modification of nucleotides, sigma factors, ligation, transcription termination, and DNA polymerase); RNA (ribosome proteins, initiation factors for translation, tRNA modification, ribonucleases, and RNA polymerase); and cellular processes (protein folding and modification, shuttling of proteins, protein membrane transport, production and recycling of metabolic substrates and intermediates, redox signalling, oxidation stress response, and the pyruvate dehydrogenase complex). Of the eight non-essential genes, four (MG_048, MG_072, MG_170, MG_297) are associated with the SecYEG complex [30] (which moves protein across or inserts them into the cell membrane), while MG_172 removes the start codon that initiates protein synthesis from synthesised nascent proteins, MG_305 and MG_392 assist in late protein folding, and MG_425 processes ribosomal RNA precursors. Although these eight genes are singly non-essential, by single gene deletion *in-silico* [8] and *in-vivo* [6], they all play a part in essential functions within the cell, hence their inclusion in published minimal gene sets.

A comparison of the genes deleted from the *M.genitalium* genome by all the eight minimal gene sets showed that they shared 14 gene deletions in common (Table 4, Agreed Column). The function of these genes includes: fructose import, activation of host immune response, chromosomal partition, amino acid transport, antibody binding, phosphonate transport, external DNA uptake, DNA repair, rRNA modification, membrane breakdown, toxin transport, quorum sensing, and a restriction enzyme. These have all been previously classified as non-essential for cell replication by single gene deletion *in-silico* [8] and *in-vivo* [6], hence their exclusion from the minimal gene sets. We placed these 14 common genes in an 'Agreed set' which was simulated in addition to the minimal gene sets from the literature.

Testing the Minimal Gene Sets

We simulated each minimal gene set from Table 4 in the *M.genitalium* whole-cell model and found that every set, including the Agreed set, produced a non-dividing *in-silico* cell. Each minimal gene set was simulated ten times per experiment, and each experiment was repeated three times (Table 5). All of these *in-silico* cells showed faults in metabolism, resulting in no DNA replication, RNA production, protein production, growth, or successful cell division.

An initial analysis found that every one of the sets from the literature deleted essential genes (genes that when removed stop the cell successfully dividing, previously classified by single gene deletion *in-silico* [8,31]). The number of essential genes deleted varied depending on the set: Nashville deleted 121, Fujisawa deleted 112, Guelph deleted 107, Valencia deleted 69, Bethesda deleted 34, Rockville and Rockville 2 both deleted 9, and Stanford deleted 3. Protocell minimal gene sets removed the largest number of essential genes and comparative genomics removed greater numbers of essential genes the higher the number of genomes compared. However, single gene deletion minimal gene sets also removed essential genes. This is likely due to transposon mutagenesis issues at the time of the minimal gene set design. Rockville labelled six genes as non-essential and so were excluded from the minimal gene set in 1999, but were later found to be essential by Rockville 2 in 2006. Even as recently as 2016, Hutchison *et al.* [29] had to make major improvements to their transposon mutagenesis protocol while producing *JCVI-Syn3.0*, due to incorrect identification of essentiality. Stanford removed MG_203, MG_250, and MG_470, likely due to averaging multiple simulation's data together before computationally assessing. These are three genes that our simulations found to be essential (Supplementary Information 1).

Repetitions	1	2	3
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Bethesda	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	-	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	-	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Rockville	No Division	No Division	No Division
Fujisawa	-	No Division	-
Fujisawa	-	-	-
Fujisawa	-	-	No Division
Fujisawa	No Division	-	-
Fujisawa	-	-	-
Fujisawa	No Division	No Division	-
Fujisawa	No Division	-	-
Fujisawa	-	-	-
Fujisawa	-	No Division	-
Fujisawa	-	-	-
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Rockville 2	No Division	No Division	No Division
Nashville	-	No Division	-
Nashville	-	No Division	-
Nashville	-	-	No Division
Nashville	-	-	-
Nashville	-	No Division	-

Repetitions	1	2	3
Nashville	-	-	-
Nashville	-	-	-
Nashville	-	-	-
Nashville	No Division	No Division	-
Nashville	-	-	-
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Stanford	No Division	No Division	No Division
Guelph	-	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Guelph	No Division	No Division	No Division
Valencia	No Division	No Division	No Division
Valencia	-	No Division	No Division
Valencia	-	-	-
Valencia	-	-	No Division
Valencia	-	-	No Division
Valencia	-	-	-
Valencia	No Division	-	-
Valencia	-	No Division	-
Valencia	-	-	-
Valencia	-	-	-
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division
Agreed	No Division	No Division	No Division

Table 5. Simulating the minimal gene sets as represented in the literature in the *M.genitalium* whole-cell model. - = a failed simulation due to supercomputer error.

Repairing the Minimal Gene Sets

We reintroduced the essential genes to the minimal gene sets, in an attempt to restore *in-silico* division. Based on previous research [31], we also reintroduced low essential genes (e.g. if the gene is dispensable in some contexts, i.e. redundant essential genes and gene complexes [19]), knowing that they would impact the *in-silico* cell's ability to divide. Each "repaired" minimal gene set was simulated ten times per experiment, and each experiment was repeated three times (Table 6). By reintroducing these genes, specific to each set (Table 7), each set produced a dividing cell *in-silico* (Table 8), including the Agreed set (Table 4).

Repetitions	1	2	3
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Bethesda	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Rockville	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	No Division	Dividing
Fujisawa	Dividing	Dividing	Dividing
Fujisawa	Dividing	No Division	Dividing
Fujisawa	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Rockville 2	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing

Repetitions	1	2	3
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	Dividing	No Division	Dividing
Nashville	Dividing	Dividing	Dividing
Nashville	No Division	Dividing	No Division
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	No Division	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Stanford	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	No Division
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	No Division
Guelph	Dividing	Dividing	Dividing
Guelph	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	No Division	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Valencia	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing
Agreed	Dividing	Dividing	Dividing

Table 6. Simulating the minimal gene sets with genes reintroduced to “repair” them in the *M.genitalium* whole-cell model. The reintroduced genes are listed in Table 7.

	Protocells		Comparative Genomics			Single Gene Knockouts		
	Nashville	Fujisawa	Guelph	Bethesda	Valencia	Stanford	Rockville 2	Rockville
MG_001	r	r						
MG_003	r	r						
MG_004	r	r						
MG_006	r	r	r					
MG_007	r	r	r	r				
MG_008	r	r						
MG_013	r	r	r		r			
MG_019	r	r						
MG_022	r	r	r	r	r			
MG_023	r		r					
MG_026		r	r					
MG_031	r	r	r	r				
MG_034	r	r	r	r	r			
MG_037	r	r	r	r	r			
MG_038	r		r		r			
MG_039	r	r	r	r	r	r	r	r
MG_041	r		r	r				
MG_042	r	r	r		r			
MG_043	r	r	r		r			
MG_044	r	r	r		r			
MG_045	r	r	r		r			
MG_047	r	r						
MG_049	r	r	r		r			r
MG_051	r	r	r	r	r		r	r
MG_053	r	r	r		r			
MG_058	r	r						
MG_066	r	r					r	
MG_069	r		r	r				
MG_071	r	r	r		r			
MG_077	r	r	r		r			
MG_078	r	r	r		r			
MG_079	r	r	r		r			
MG_080	r	r	r		r			
MG_084	r	r		r				
MG_090			r					
MG_091	r	r	r		r			
MG_093			r					
MG_094	r	r						
MG_098	r	r	r	r	r			

	Nashville	Fujisawa	Guelph	Bethesda	Valencia	Stanford	Rockville 2	Rockville
MG_099		r	r	r	r			
MG_100		r	r	r	r			
MG_102	r	r						
MG_104	r	r	r		r			
MG_106	r	r	r		r			
MG_107	r	r						
MG_111	r							
MG_112	r	r					r	
MG_114	r		r		r		r	
MG_118	r	r	r		r			
MG_124	r	r						
MG_128	r	r	r	r	r			
MG_137	r	r	r	r	r			
MG_141	r	r						
MG_145	r	r						
MG_153			r					
MG_159			r					
MG_164			r					
MG_169		r	r					
MG_171		r						
MG_174			r					
MG_177	r							
MG_179	r	r	r	r	r			
MG_180	r	r	r		r			
MG_181	r	r	r	r	r			
MG_182	r	r	r		r			r
MG_195		r	r					
MG_201	r	r	r					
MG_203	r	r	r		r	r		
MG_204	r	r	r		r			
MG_212	r	r	r					
MG_215	r		r					
MG_216	r		r					
MG_221	r	r	r		r			
MG_224	r	r						
MG_228	r	r	r					
MG_229	r	r	r					
MG_230	r	r	r	r				
MG_231	r	r						
MG_232			r					
MG_238	r	r	r		r		r	
MG_245	r	r	r		r			
MG_249	r							
MG_250	r	r				r		
MG_251			r					

	Nashville	Fujisawa	Guelph	Bethesda	Valencia	Stanford	Rockville 2	Rockville
MG_254	r	r						
MG_258		r						
MG_261	r	r						
MG_270	r	r	r		r			
MG_271	r	r	r		r		r	
MG_272	r	r	r		r			
MG_273	r	r	r		r			
MG_274	r	r	r		r			
MG_275	r	r	r		r			
MG_276	r	r	r		r			
MG_278	r	r	r		r			
MG_282	r	r						
MG_283			r					
MG_287	r	r	r		r			
MG_289	r	r	r		r	r	r	
MG_290	r	r	r		r	r	r	
MG_291	r	r	r		r	r	r	*r*
MG_295	r	r						r
MG_299	r	r	r		r			
MG_300	r							
MG_301	r							
MG_302	r	r	r	r	r			
MG_303	r	r	r	r	r			
MG_304	r	r	r	r	r			
MG_305	r	r						
MG_311				r				
MG_315	r	r	r	r		r	r	
MG_321	r	r	r	r	r			
MG_322	r	r	r		r			
MG_323	r	r	r	r	r			
MG_325			r					
MG_330	r	r	r		r			
MG_340	r							
MG_341	r							
MG_342	r	r	r	r	r			
MG_345								r
MG_347	r	r	r		r			
MG_351			r					
MG_357	r	r	r		r			
MG_363			r					
MG_365			r		r			
MG_367	r	r					r	
MG_368	r	r	r	r	r			
MG_372	r	r	r	r	r			r
MG_375			r					

	Nashville	Fujisawa	Guelph	Bethesda	Valencia	Stanford	Rockville 2	Rockville
MG_379	r	r	r					
MG_382	r	r	r		r			
MG_383	r	r	r		r			
MG_384	r	r						
MG_387	r	r	r					
MG_394	r	r						r
MG_396	r	r	r	r				
MG_407	r							
MG_419	r	r	r	r	r			
MG_424	r							
MG_426			r					r
MG_427	r	r	r		r	r		
MG_429	r			r				
MG_430	r		r					
MG_431	r							
MG_434	r	r	r		r			
MG_435		r						
MG_437	r		r				r	
MG_444					r			
MG_445		r						
MG_453	r	r	r		r			
MG_458	r	r	r					
MG_465		r	r					
MG_466			r					
MG_469	r	r			r			
MG_470	r	r	r	r	r	r		r
MG_473	r	r	r	r	r			
MG_481	r	r	r	r	r			
MG_517	r	r	r	r	r			
MG_522			r					

Table 7. Comparing the gene reintroductions made to make the minimal gene sets produce *in-silico* dividing cells. Light grey genes are low essential genes in the *M.genitalium* whole-cell model. r = a gene reintroduction.

	Design approach	*In-silico* gene deletions (cell did not divide)	*In-silico* gene deletions (cell divided)	Genes reintroduced	Size of *in-silico* genome
Nashville	Protocell	270	142	128	217
Fujisawa	Protocell	261	141	120	218
Guelph	Comparative Genomics	241	129	112	230
Valencia	Comparative Genomics	185	110	75	249
Stanford	Single Gene Deletions	117	109	8	250
Bethesda	Comparative Genomics	118	82	36	277
Rockville 2	Single Gene Deletions	57	45	12	314
Rockville	Single Gene Deletions	53	43	10	316
Agreed	-	14	13	1	346

Table 8. Minimal gene sets that produce dividing *in-silico* cells. After the reintroduction of essential and low essential genes.

To repair the Agreed set, one low essential gene had to be reintroduced (MG_291), which is involved in phosphonate transport with two other genes (MG_289, MG_290). The agreed set also removed MG_412, a gene that along with MG_410 and MG_411, is involved with phosphate transport. By disrupting both these processes, the Agreed set produced an *in-silico* cell that did not have a source of phosphate. This relationship has been established previously [31]. MG_291 had to be reintroduced in every minimal gene set apart from Bethesda, which does not remove phosphate transport (MG_410, MG_411, MG_412). MG_289 and MG_290 also had to be reintroduced in six of the minimal gene sets.

We identified 31 genes that were reintroduced into at least five of the minimal gene sets (Table 9). 26 were classified as essential and 5 as low essential (Supplementary Information 2). The cellular functional groups that the minimal gene sets needed reintroducing were: DNA (polymerase delta / gamma / tau subunits, introduction of thymidine into DNA, rescue of pyrimidine bases for nucleotide synthesis, chromosome segregation); RNA (polymerase subunit delta, tRNA modification, 50S and 30S ribosomal subunits); transporters (cobalt, phosphonate, potassium); production (NAD, flavin, NADP, fatty acid/phospholipids); and dehydrogenation (glycerol and alpha-keto acids).

1	MG_022	11	MG_203	21	MG_321
2	MG_034	12	MG_238	22	MG_323
3	MG_037	13	MG_271	23	MG_342
4	MG_039	14	MG_289	24	MG_368
5	MG_051	15	MG_290	25	MG_372
6	MG_098	16	MG_291	26	MG_419
7	MG_128	17	MG_302	27	MG_427
8	MG_137	18	MG_303	28	MG_470
9	MG_179	19	MG_304	29	MG_473
10	MG_181	20	MG_315	30	MG_481
				31	MG_517

Table 9. 31 genes that were reintroduced into at least five minimal gene sets. They are all classified as essential (previously classified by single gene deletion *in-silico* [8,31]), apart from MG_039, MG_289, MG_290, MG_291, MG_427, which are classified as low essential (previously classified by single gene deletion *in-silico* [31]).

Of the 26 essential genes, 19 did not need to be reintroduced into the single gene deletion minimal gene sets (Stanford, Rockville, Rockville 2) as they had not been previously deleted. Two of these 19 genes, MG_137 (involved in cell wall production of mycobacteria) and MG_517 (produces fundamental components for plasma membrane stability) are specifically essential for *Mycoplasma* species, which were not identified as essential by the other, non species-specific, methodologies. Additionally, five of the 19 genes were involved in cobalt transport. Cobalt is used in the cell to produce cobalamin (otherwise known as vitamin B12), an enzyme cofactor which increases the reaction rates of DNA synthesis, fatty acid metabolism, and amino acid metabolism. This was also not identified as essential by the other design methodologies.

Interestingly, of the five low essential genes, Rockville and Bethesda did not delete four out of the five genes and Rockville 2 did not delete two of the five genes. It is likely that Bethesda did not remove the low essential genes due to the direct comparison with a closely related species, the genes with low essentiality being conserved to a greater degree than non-essential genes. We would speculate that Rockville removed less low essential genes than Rockville 2 due to Rockville using a transposon mutagenesis

protocol with less precise results (as stated by Rockville 2 [25]), cells were grown in mixed pools with DNA isolated from these mixtures rather than from isolated pure colonies of cells.

The protocell minimal gene sets (Nashville and Fujisawa) produced the smallest genomes *in-silico* (Table 8, Figure 1), once they were made functional, but they required the most number of genes to be reintroduced. The comparative genomics minimal gene sets (Guelph, Valencia, Bethesda) produce larger genomes *in-silico* but required less genes to be reintroduced. The smaller the number of genomes compared in their design, the less genes had to be reintroduced to make dividing *in-silico* cells. Interestingly, Stanford produced a smaller *in-silico* genome than Bethesda, as the Stanford minimal gene set did not target unmodelled gene deletions and only required eight genes to be reintroduced.

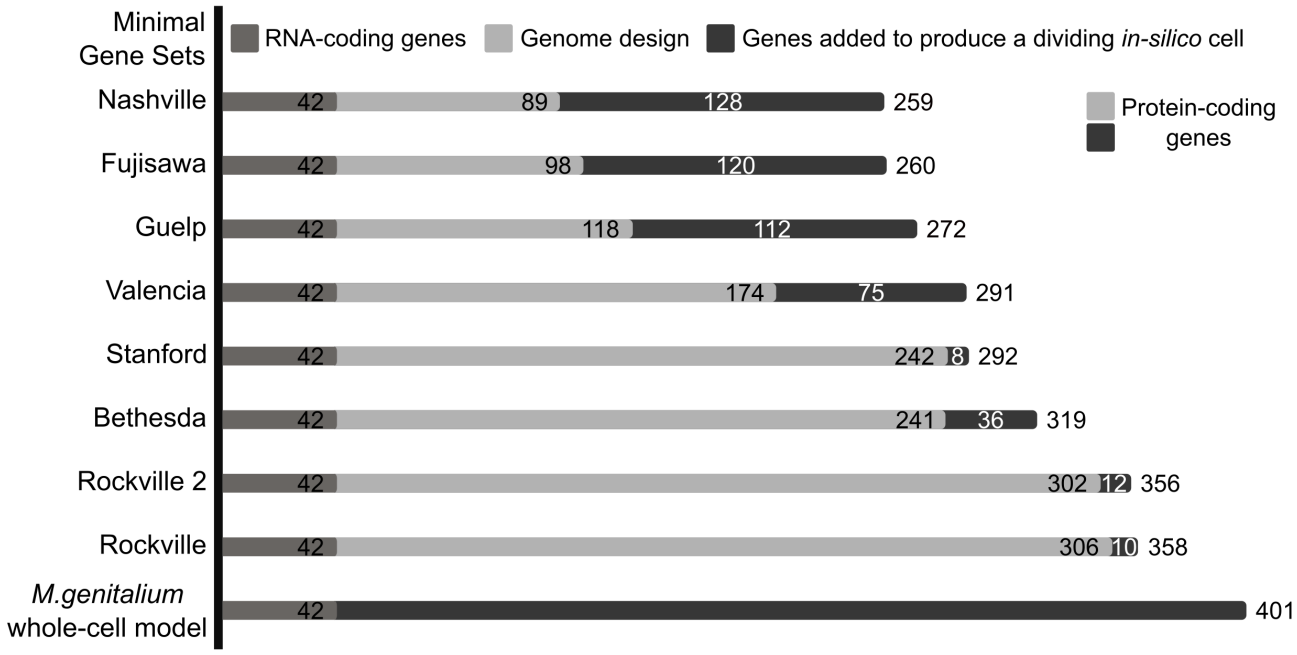

Figure 1. Repairing minimal gene sets from the literature to produce dividing *in-silico* cells. Minimal gene sets are shown from a range of design methodologies, produced across the past twenty years, and organised by the size of genome they produce in the whole-cell model. Original genome designs are light grey bars, genome designs with reintroduced essential and low essential genes are black bars.

Discussion

We tested minimal gene sets from the literature [3–10] for the first time, finding that they produced *in-silico* cells that did not divide, but by reintroducing specific essential and low essential genes the sets could produce dividing *in-silico* cells.

It currently appears that there is little to be learned for minimal genome research from modelling protocell minimal gene sets in the *M.genitalium* whole-cell model, given the amount of genetic modifications required to produce a viable *in-silico* cell. The comparative genomics minimal gene sets are closest to producing dividing cells *in-silico* when a lower number of genomes are compared. The single gene deletion minimal gene sets required the least genes to be reintroduced, so could be considered closest to viable *in-silico* minimal genomes from the outset, but without the ability to identify low essential genes, minimal gene sets designed with this methodology will still require correcting.

The identification of genes to reintroduce was straightforward given the previous labelling of *M.genitalium in-silico* essential and low essential genes, which would have been required if the information had not already been available. The results from this work reinforce a broader trend of moving away from universal minimal genomes [13, 23] (see Introduction), towards working on species-specific minimal gene sets and minimal genomes, specifically identifying low essential genes and their interactions.

This research has limitations associated with the use of the *M.genitalium* whole-cell model (models are not perfect representations of reality and, through necessity, the *M.genitalium* whole-cell model bases some of its parameters on data from other bacteria [8]), and uncertainty around the impact of the unmodelled genes on *in-vivo* experiments (e.g. if they perform a unique essential function with a gene / gene product that has already been removed, then the *in-vivo* cell will not survive), as we have stated previously [31].

This research advances minimal genome design. The computational predictions we have produced need to be tested in living cells, but with the advancement of gene synthesis and genome transplantation in other *Mycoplasma* species [17,28,32–35] this is becoming a more realistic proposition for *Mycoplasma* researchers.

References

1. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496–512 (1995).
2. Fraser, C. M. *et al.* THE MINIMAL GENE COMPLEMENT OF MYCOPLASMA-GENITALIUM. *Science* **270**, 397–403 (1995).
3. Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10268–10273 (1996).
4. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **286**, 2165–2169 (1999).
5. Tomita, M. *et al.* E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84 (1999).
6. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 425–430 (2006).
7. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, (2006).
8. Karr, J. R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).
9. Huang, C. H., Hsiang, T. & Trevors, J. T. Comparative bacterial genomics: defining the minimal core genome. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **103**, 385–398 (2013).
10. Gil, R. The Minimal Gene-Set Machinery. *Reviews in Cell Biology and Molecular Medicine* (2014).
11. Gil, R., Silva, F. J., Pereto, J. & Moya, A. Determination of the core of a minimal

bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–+ (2004).

12. Shuler, M. L., Foley, P. & Atlas, J. Modeling a Minimal Cell. in *Microbial Systems Biology* (ed. Navid, A.) vol. 881 573–610 (Humana Press, 2012).
13. Glass, J. I., Merryman, C., Wise, K. S., Hutchison, C. A., 3rd & Smith, H. O. Minimal Cells-Real and Imagined. *Cold Spring Harb. Perspect. Biol.* (2017) doi:10.1101/cshperspect.a023861.
14. Xavier, J. C., Patil, K. R. & Rocha, I. Systems Biology Perspectives on Minimal and Simpler Cells. *Microbiol. Mol. Biol. Rev.* **78**, 487–509 (2014).
15. Mushegian, A. The minimal genome concept. *Curr. Opin. Genet. Dev.* **9**, 709–714 (1999).
16. Fraser, C. M., Eisen, J. A. & Salzberg, S. L. Microbial genome sequencing. *Nature* **406**, 799–803 (2000).
17. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215–1220 (2008).
18. Dzieciol, A. J. & Mann, S. Designs for life: protocell models in the laboratory. *Chem. Soc. Rev.* **41**, 79–85 (2012).
19. Rancati, G., Moffat, J., Typas, A. & Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19**, 34–49 (2018).
20. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).
21. D'Elia, M. A., Pereira, M. P. & Brown, E. D. Are essential genes really essential? *Trends Microbiol.* **17**, 433–438 (2009).
22. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **1**, 127–136 (2003).

23. Lagesen, K., Ussery, D. W. & Wassenaar, T. M. Genome update: the 1000th genome - a cautionary tale. *Microbiology-Sgm* **156**, 603–608 (2010).
24. Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W. & Danchin, A. From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet.* **29**, 273–279 (2013).
25. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 425–430 (2006).
26. Juhas, M., Eberl, L. & Glass, J. I. Essence of life: essential genes of minimal genomes. *Trends Cell Biol.* **21**, 562–568 (2011).
27. Reich, K. A. The search for essential genes. *Res. Microbiol.* **151**, 319–324 (2000).
28. Benders, G. A. *et al.* Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* **38**, 2558–2569 (2010).
29. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, 1414–U73 (2016).
30. du Plessis, D. J. F., Nouwen, N. & Driessen, A. J. M. The Sec translocase. *Biochim. Biophys. Acta* **1808**, 851–865 (2011).
31. Rees-Garbutt, J. *et al.* Designing minimal genomes using whole-cell models. *Nat. Commun.* **11**, 836 (2020).
32. Tsarmpopoulos, I. *et al.* In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9. *ACS Synth. Biol.* **5**, 104–109 (2016).
33. Karas, B. J. *et al.* Direct transfer of whole genomes from bacteria to yeast. *Nat. Methods* **10**, 410–+ (2013).
34. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).

35. Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404–20409 (2008).

36. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

37. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–9 (2004).

Methods

Code Availability

All code created as part of this paper will be made available on Github (github.com/squishybinary, github.com/GriersonMarucciLab) under a GNU General Public License v3.0 (gpl-3.0). For more information see choosealicense.com/licenses/lgpl-3.0/.

Data Availability

The databases used to design the *in-silico* experiments, and compare the results to, includes Karr *et al.* [8] and Glass *et al.* [6] Supplementary Tables, and Fraser *et al.* *M.genitalium* G37 genome [2] interpreted by KEGG [36] and UniProt [37] as strain ATCC 33530/NCTC 10195. The output .fig files for all simulations referenced will be made available at the group's Research Data Repository (data-bris) at the University of Bristol.

Model Availability

The *M.genitalium* whole-cell model is freely available: github.com/CovertLab/WholeCell. The model requires a single CPU and can be run with 8GB of RAM. I run the *M.genitalium* whole-cell model on Bristol's supercomputers using MATLAB R2013b, with the model's standard settings. However, we use our own version of the SimulationRunner.m. MGGRunner.m (github.com/GriersonMarucciLab/Analysis_Code_for_Mycoplasma_genitalium_whole-cell_model) is designed for use with supercomputers that start hundreds of simulations simultaneously. It artificially increments the starting time-date value for each simulation, as this value is subsequently used to create the initial conditions of the simulation. Our research copy of the whole-cell model was downloaded 10th January 2017.

M.genitalium in-silico Environmental Conditions

M.genitalium is grown *in-vivo* on SP4 media. The *in-silico* media composition is based on the experimentally characterized composition, with additional essential molecules added (nucleobases, gases, polyamines, vitamins, and ions) in reported amounts to support *in-silico* cellular growth. Additionally, the *M.genitalium* whole-cell model represents 10 external stimuli including temperature, several types of radiation, and

three stress conditions. For more information see Karr *et al.* Supplementary Tables S3F, S3H, S3R [8].

Equipment

For the *M.genitalium* whole-cell model we used the University of Bristol Advanced Computing Research Centres's BlueGem, a 900-core supercomputer, which uses the Slurm queuing system, to run whole-cell model simulations.

We used a standard office desktop computer, with 8GB of ram, to write new code, and interact with the supercomputer. We used the following GUI software on Windows 7: Notepad++ for code editing, Putty (ssh software) for terminal access to the supercomputer, FileZilla (ftp software) to move files in bulk to and from the supercomputer, and PyCharm (IDE software) as an inbuilt desktop terminal and for python debugging. The command line software used included: VIM for code editing, and SSH, Rsync, and Bash for communication and file transfer with the supercomputers.

Data Format

For the *M.genitalium* whole-cell model the majority of output files are state-NNN.mat files (Figure 2), which are logs of the simulation split into 100-second segments. The data within a state-NNN.mat file is organised into the 16 cellular variables. These are typically arranged as 3-dimensional matrices or time series, which are flattened to conduct analysis. The other file types contain summaries of data spanning the simulation. Each wild type simulation consists of 300 files requiring 0.3GB. Each gene manipulated simulation can consist of up to 500 files requiring between 0.4GB and 0.9GB. Each simulation takes 5 to 12 hours to complete in real time, 7 - 13.89 hours in simulated time.

Data Analysis Process

For the *M.genitalium* whole-cell model, the raw data is automatically processed as the simulation ends. runGraphs.m carries out the initial analysis, while compareGraphs.m overlays the output on collated graphs of 200 unmodified *M.genitalium* simulations.

Both outputs are saved as MATLAB .fig and .pdfs, though the .pdf files were the sole files analysed. The raw .mat files were stored in case of further investigation.

Further analysis, including: analysis of genetic content and similarity, gene ontology, and identification and investigation of high and low essentiality genes and groupings, were completed manually. The GO biological process terms were downloaded from Uniprot [37] (strain ATCC 33530/NCTC 10195), processed by a created script (github.com/squishybinary/Gene_Ontology_Comparison_for_Mycoplasma_genitalium_whole-cell_model), then organised manually into tables of GO terms that were unaffected, reduced, or removed entirely by gene deletions.

```
State-N.mat
├─ Chromosome
│  └─ 18 sub variables, graphed:  Ploidy
├─ FtsZRing
│  └─ 5 sub variables
├─ Geometry
│  └─ 9 sub variables, graphed:  Pinched Diameter
├─ Host
│  └─ 4 sub variables
├─ Mass
│  └─ 9 sub variables, graphed:  Total, RnaWt, ProteinWt
├─ MetabolicReaction
│  └─ 3 sub variables, graphed:  Growth
├─ Metabolite
│  └─ 4 sub variables
├─ Polypeptide
│  └─ 4 sub variables
├─ ProteinComplex
│  └─ 1 sub variable
├─ ProteinMonomer
│  └─ 1 sub variable
├─ Ribosome
│  └─ 8 sub variables
├─ Rna
│  └─ 1 sub variable
├─ RNAPolymerase
│  └─ 9 sub variables
├─ Stimulus
│  └─ 1 sub variable
├─ Time
│  └─ 1 sub variable
└─ Transcript
   └─ 5 sub variables
```

Figure 2. Anatomy of a state file (*M.genitalium* whole-cell model). The data within a *state-XXX.mat* file is organised into 16 cellular variables. The variables graphed in my analysis are: Ploidy as DNA replication, Pinched Diameter as Cell Division, Total as Mass, RnaWt as RNA production, ProteinWt as Protein Production, Growth as Growth.

Modelling Scripts

There are six scripts used to run the *M.genitalium* whole-cell model. Three are the experimental files created with each new experiment (the bash script, gene list, experiment list), and three are stored within the whole-cell model and are updated only upon improvement (MGGrunner.m, runGraphs.m, and compareGraphs.m). The bash script is a list of commands for the supercomputer(s) to carry out. Each bash script determines how many simulations to run, where to store the output, and where to store the results of the analysis. The gene list is a text file containing rows of gene codes (in the format 'MG_XXX',). Each row corresponds to a single simulation and determines which genes that simulation should knockout. The experiment list is a text file containing rows of simulation names. Each row corresponds to a single simulation and determines the final location of the simulation output and analysis results.