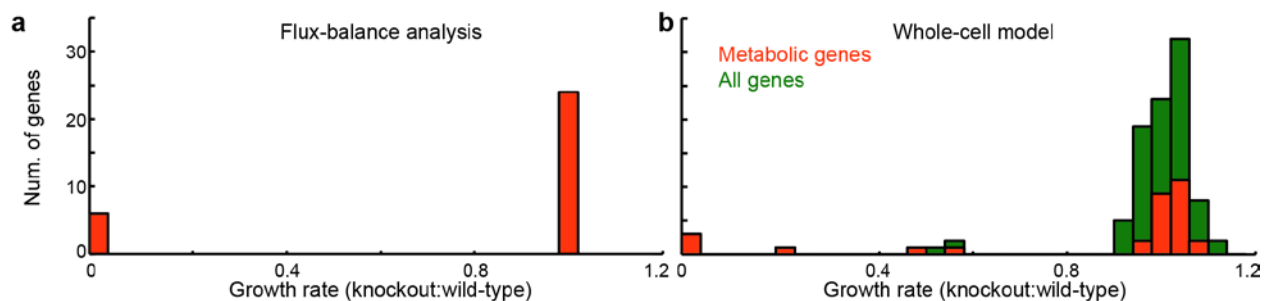# Supplementary Material

Supplementary Figure 1 | Flux balance analysis applied to predicting specific growth rates in *M. genitalium*.


Flux balance analysis applied to predicting specific growth rates in *M. genitalium*. (**a**) Distribution of non-essential metabolic genes (essentiality determined experimentally, see Ref. 6) in *M. genitalium* predicted by flux balance analysis. (**b**) A similar distribution using the whole-cell model (*n* = 5). Metabolic genes are colored orange; all other non-essential genes are colored green.

**Supplementary Figure 2 | Data to motivate hypotheses about discrepant strains.**
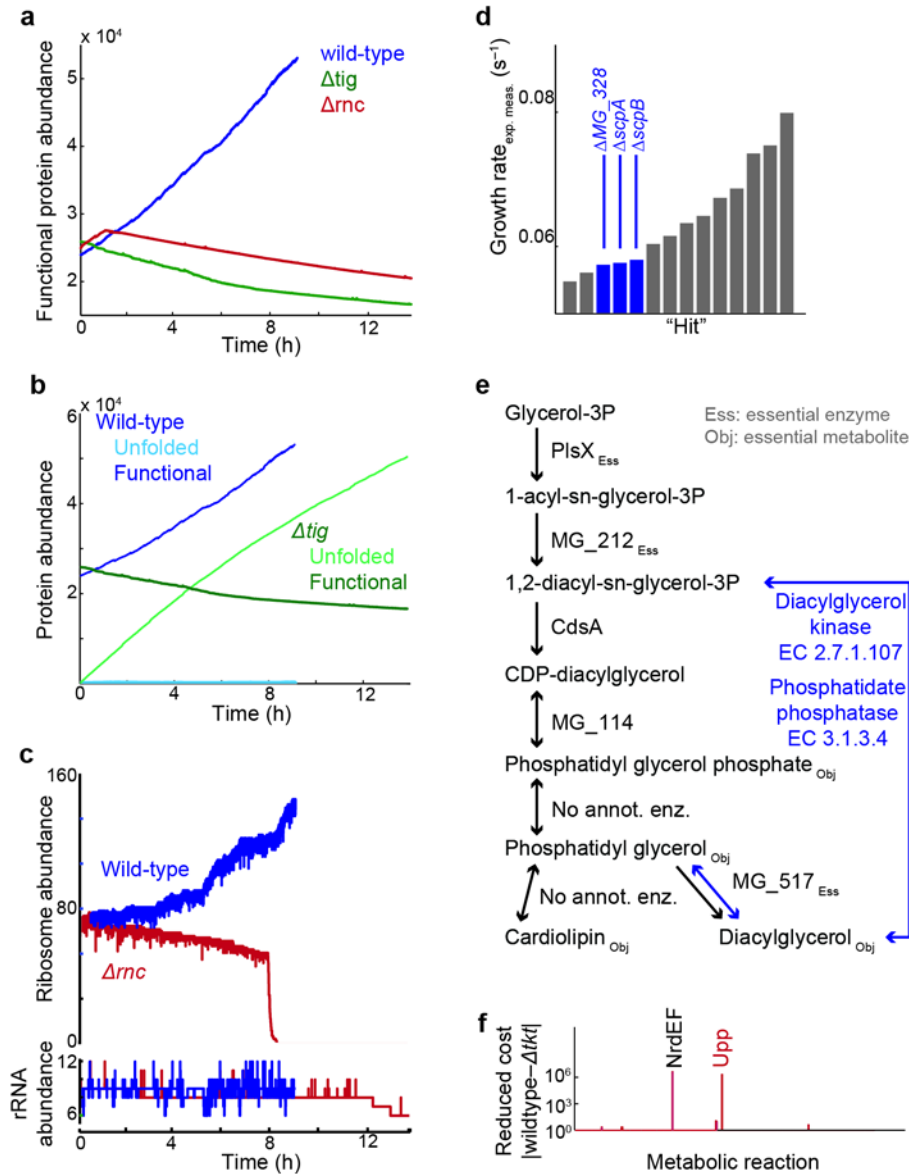
Data to motivate hypotheses about discrepant strains. (**a**) A model simulation shows that the absence of trigger factor (Tig) or ribonuclease III (Rnc) affects the cell's ability to make functional protein. (**b**) In particular, the model shows that *Δtig* is unable to fold protein to their functional form. (**c**) The model results indicate that *Δrnc* is unable to assemble ribosomes (top) even though it is able to synthesize rRNAs to the same degree as the wild-type (bottom). (**d**) The measured growth rates of all gene hits. The growth rates of *ΔMG_328*, *ΔscpA*, and *ΔscpB* all cluster together, indicating that they may have similar functions. (**e**) The lipid biosynthesis pathway as it currently exists in the *M. genitalium* reconstruction (black). Experimentally determined essential genes are indicated as "Ess" and metabolites in the objective function (required for growth) are indicated as "Obj." A possible alternative pathway to produce diacylglycerol is shown in blue. (**f**) Reduced cost analysis for the *Δtkt* strain (similar to that presented in **Fig. 2a,c**).

**Supplementary Figure 3 | Additional controls and quantification of kinetic assay data.**
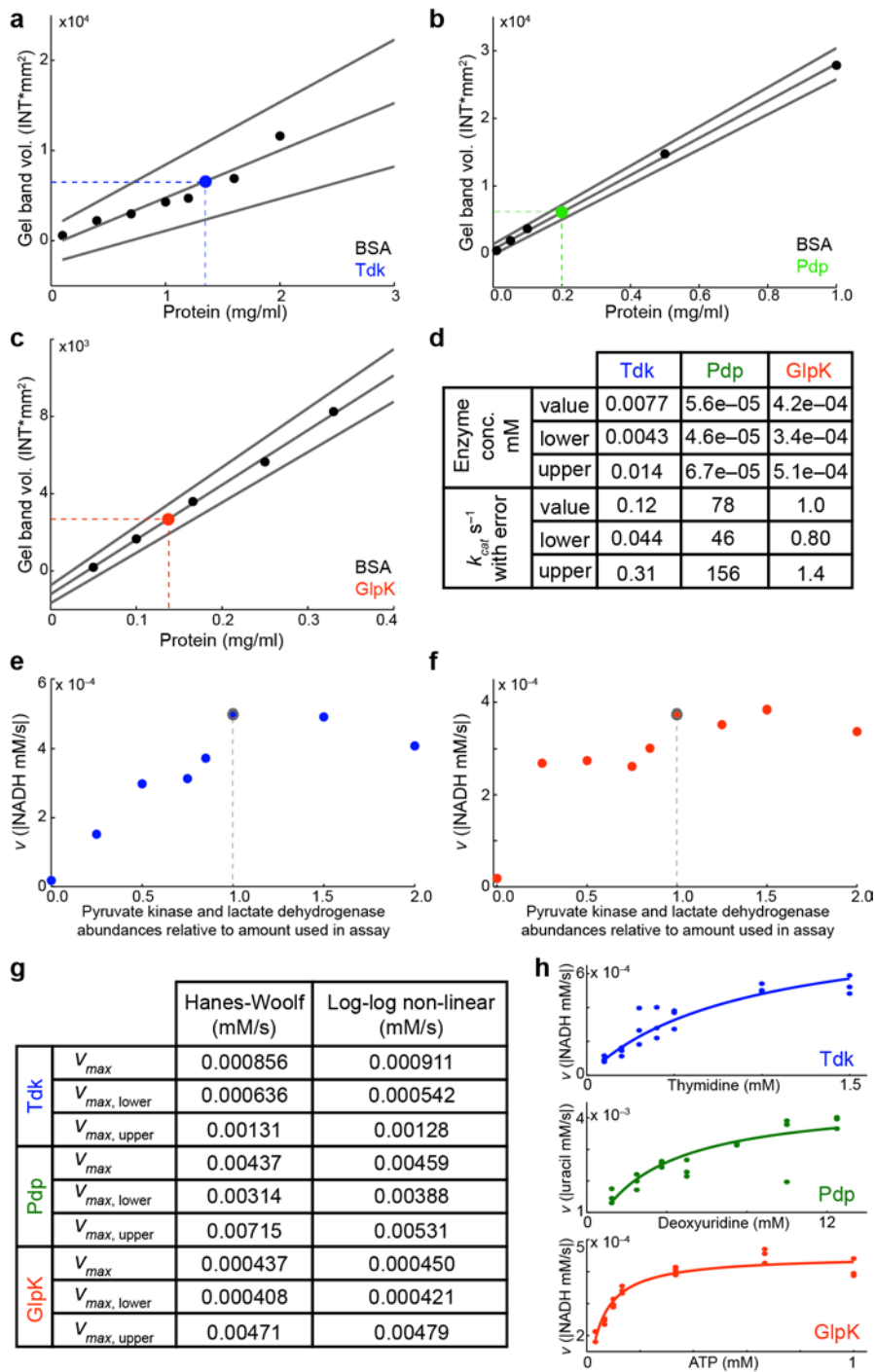
Additional controls and quantification of kinetic assay data. (**a-c**) Linear regressions of BSA standard curves used to quantify Tdk (**a**), Pdp (**b**), and GlpK(**c**) isolated proteins, with 95% confidence intervals of the regressions. (**d**) Enzyme concentrations used in the kinetic assays as quantified in (**a-c**) and the resulting k_{cat}s. (**e-f**) Controls for Tdk (**e**) and GlpK (**f**) assays with varying amounts of pyruvate kinase and lactate dehydrogenase indicate that both of the product-consuming enzymes were in excess during the v_{max} determining assays. (**g**) Comparison of Hanes-Woolf and log-log Michaelis Menten non-linear fits to the kinetic assay data indicate that the method used for the fit does not greatly impact the calculated v_{max}s. (**h**) Plots showing that the log-log Michaelis Menten non-linear regressions appropriately fit the data. The Hanes-Woolf regressions are shown in Fig. 3.

Supplementary Figure 4 | Model-predicted and experimentally measured protein abundance.
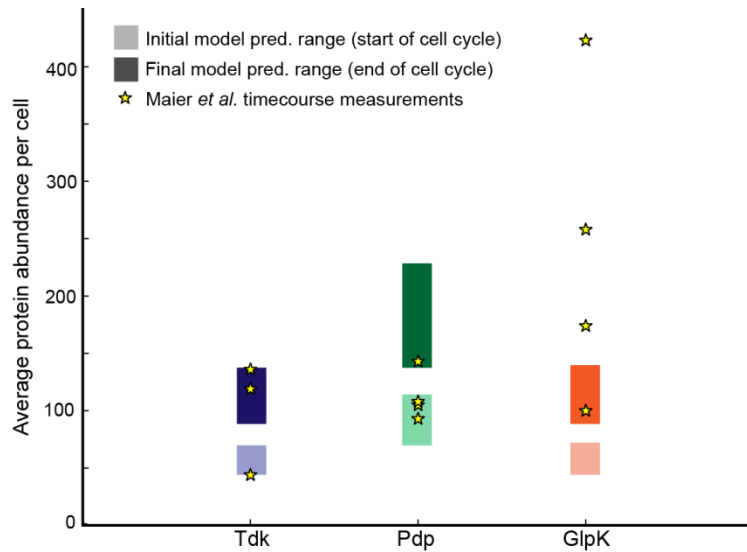
Model-predicted and experimentally measured protein abundance. The whole-cell model calculates the maximum flux through a particular reaction, or v_{max}, as the product of (enzyme abundance × k_{cat}). This means that model predictions of both enzyme abundance and k_{cat} should be accurate in order to make valid predictions. Here, we compare our model predictions of protein abundance for Tdk, Pdp, and GlpK, to experimental measurements made in the close homolog, *Mycoplasma pnuemoniae* (Maier *et al.*, 2011). For the model, the predictions are expressed as two ranges—one at the beginning (light box) and the other at the end (dark box) of the cell cycle—which represent the mean value ± one standard deviation. The experimental measurements (yellow stars) were obtained from a time course of M. pneumoniae growth over four days (Maier *et al.*, 2011). Two caveats should be mentioned with regard to this comparison: 1) *M. pneumonaie* cells have a larger volume than *M. genitalium* cells and therefore may have higher protein counts, and 2) that *M. pnuemoniae* cells have a larger genome and express more gene products, which could make the protein concentration lower relative to *M. genitalium*.

Supplementary Figure 5 | Recalculation of the GlpK flux bound using a Michaelis Menten-based formulation.
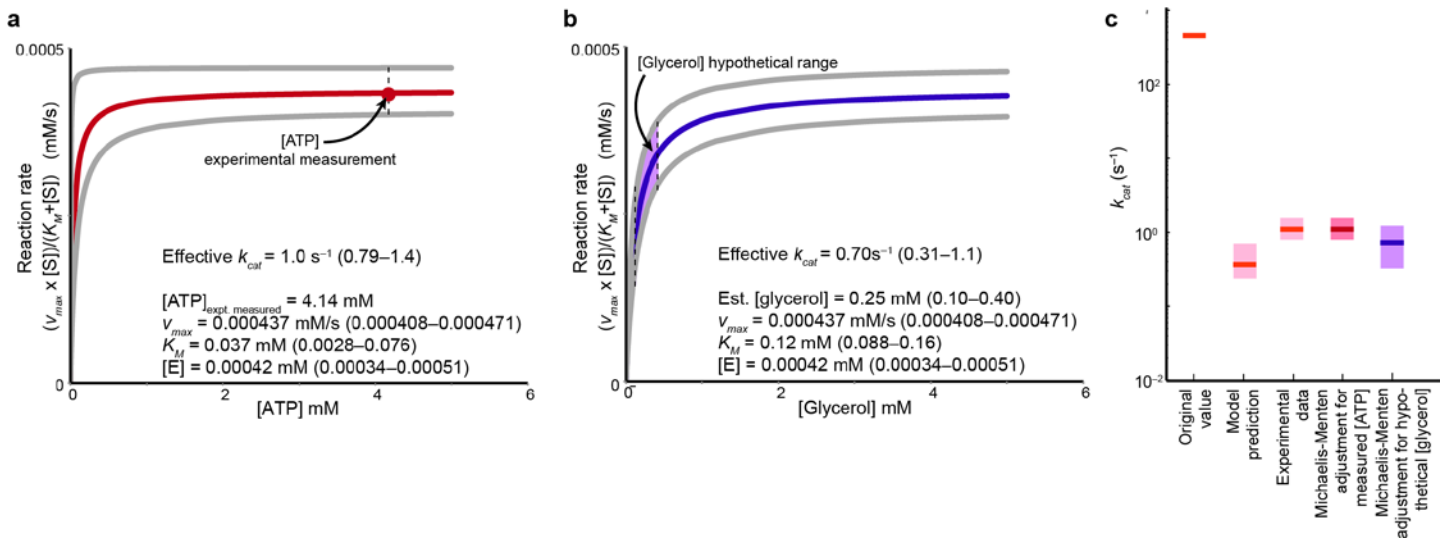
Recalculation of the GlpK flux bound using a Michaelis Menten-based formulation. The simplest explanation for the discrepancy between our predicted and experimentally measured range for k_{cat} is that the actual rate of an enzyme is often not only a function of the enzyme concentration and k_{cat} (as represented in the whole cell model), but also of many other parameters and variables, from the substrate concentration and K_m to the limits imposed by allosteric regulation, post-translational modifications and the like. Here, we consider a Michaelis-Menten formulation which also takes into account the concentration of substrate in determining the flux bound. (**a**) A plot showing the calculation of $v = v_{max} \times [\text{ATP}]/(K_m + [\text{ATP}])$, where the v_{max} and K_m are determined from the kinetic assay data (**Supplementary Figure 3c**). [ATP] was approximated by a recent measurement in *E. coli*[26]. (**b**) Same as (**a**), but now using $v = v_{max} \times [\text{Glycerol}]/(K_m + [\text{Glycerol}])$, where the v_{max} is the same as in (**a**), and the $K_m = 0.12$, an average of three values found in BRENDA[27]: 0.11 mM (*Pediococcus pentosaceus*), 0.088 mM (*Elizabethkingia meningoseptica*), and 0.162 mM (*Thermus thermophilus*). (**c**) The original model value of the k_{cat}, the new range predicted by the model, the k_{cat} range determined experimentally, and the effective rate determined using a Michaelis Menten formulation based on ATP and Glycerol concentrations. The adjustment for the measured [ATP] concentration has no impact on the effective k_{cat} range. No data or estimates of glycerol intracellular concentration in bacteria have been published, but a hypothetical range for [Glycerol] is identified in (**b**) which would be sufficient to lower the flux bound on this enzyme to that predicted by the model, as shown in (**c**).

Supplementary Table 1 | Details of the chromosome map and growth assay data.

<Multi-page table. Please see excel document>

Details of the chromosome map and growth assay data. Table listing all of the genes in the M. genitalium genome, together with the model/experimental comparison category, as well as the model predicted and experimentally measured (where applicable) growth rates for each disruption strain in the study and wild-type. The sample size, standard deviation, t-test, and Wilcoxon test results are also listed. Six of the genes (*MG051*, *MG112*, *MG271*, *MG291*, *MG385*, *MG437*) were reported as isolated, but unculturable in our growth assay; we considered these genes essential for the purposes of our study. #N/A = not applicable, this was used in cases where the genes were essential and no quantitative growth rate data could be obtained.

Supplementary Table 2 | Model single-gene disruption specific growth rate mispredictions generated ten novel hypotheses.

	Gene ID	Gene name	Annotation	Cellular sub-model	Deviation from experiment	P-value	Model-driven hypothesis
Metabolism	*MG_227*	*thyA*	thymidylate synthase	metabolism	36.50%	0.01	New k_{cat} for Tdk
	MG_049	*deoD*	purine nucleoside phosphorylase	metabolism	-43.60%	0.006	New k_{cat} for Pdp
	MG_039	-	FAD-dependent glycerol-3-phosphate dehydrogenase, putative	metabolism	22.70%	0.0001	New k_{cat} for GlpK
	MG_066	*tkt*	transketolase	metabolism	-100.00%	0.0005	Related reaction, Upp, is reversible
	MG_114	-	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	metabolism	-81.60%	0.007	Phosphatidate phosphatase, diacylglycerol kinase, or similar activity exists in the cell
Chromosome condensation	*MG_213*	*scpA*	segregation and condensation protein A	chromosome condensation	38.70%	0.002	Enhances but is not essential for chromosome segregation
	MG_214	*scpB*	segregation and condensation protein B	chromosome condensation	30.70%	0.0005	Enhances but is not essential for chromosome segregation
	MG_328	-	coiled coil putative structural protein involved in cytoskeleton	--	32.90%	0.0001	May function similarly to ScpB
Gene expression	*MG_367*	*rnc*	ribonuclease III	RNA processing	-42.20%	1×10^{-6}	Another ribonuclease exists with redundant function
	MG_238	*tig*	trigger factor	protein folding	-41.10%	9×10^{-5}	DnaK has a redundant function to Tig
Under-characterized genes	*MG_032*	-	conserved hypothetical protein	--	32.80%	0.0004	--
	MG_115	*cinA*	competence/damage-inducible protein CinA domain protein	--	29.90%	0.0002	--
	MG_285	-	conserved hypothetical protein	--	24.30%	0.002	--
	MG_207	-	Ser/Thr protein phosphatase 2A	--	23.10%	0.002	--

Model single-gene disruption specific growth rate mispredictions generated ten novel hypotheses.
Gene disruption strains whose specific growth rate was not predicted accurately by the whole-cell model, categorized by function. Specific growth rates are presented as [(Model – Expt.) / Expt.] × 100%. The hypotheses generated by probing the discrepancies are presented.

Supplementary Table 3 | Comparison of reduced cost analysis using the whole-cell model and FBA.

Reaction	Whole-cell Model Reduced cost |KO-WT|			FBA Reduced cost |KO-WT|		
	delta_thyA	delta_deoD	delta_039	delta_thyA	delta_deoD	delta_039
Tdk2	1541240	7	0	0	0	0
Pdp1	0	1541237	0	0	0	0
ThyA	0	1541230	0	0	0	0
CoA Transport	0	0	0	12499	0	0
TatD dAMP	0	0	0	0	11719	0
Fru Transport	0	0	0	7956	0	0
OppB Transport	0	0	0	6060	0	0
Udk1	0	0	0	0	6050	0
TatD dTMP	0	0	0	3027	0	0
Upp	0	0	0	3031	0	0
ComE Transport	0	0	0	0	756	0
NrdEF GDP	0	28	0	0	0	0
Hpt1	0	0	25	0	0	0
Tdk1	7	0	0	0	0	0
Dck	0	0	4	0	0	0

Comparison of reduced cost analysis using the whole-cell model and FBA. We performed the reduced cost analysis twice on our metabolic network, once using the constraints in the whole-cell model (dependent on all the other modules in the model), and once using simple unbounded FBA. This table shows that only the whole-cell model was able to highlight the redundant reactions for ThyA and DeoD.

Supplementary Results:

Detailed descriptions of model-driven hypotheses:
The comparison of model predictions to experimentally measured growth rates resulted in a series of hypotheses about misrepresented or missing behavior in the model (**Supplementary Table 2**). The hypotheses that were not immediately testable are explained in more detail here.

Gene and protein expression
We first considered both of the genes involved in gene expression. For both of the genes, the corresponding disruption strain is unable to produce functional protein (**Supplementary Fig. 2a**), but due to different mechanisms. For example, Trigger Factor (Tig) assists in protein folding. Since chaperone requirements in *M. genitalium* are not well characterized, and Tig has previously been characterized as a generally unspecific chaperone that may bind nascent polypeptide chains leaving ribosomes[17], our whole-cell model required Tig activity for all monomeric protein folding events. In doing so, the model overestimates Tig essentiality. The simulation results show that *Δtig* strains accumulate immature proteins as they are unable to fold them to their functional form (**Supplementary Fig. 2b**). Another chaperone, DnaK, is found in *M. genitalium*[18], and other bacteria such as *E. coli* have been identified in which single deletions of Tig or DnaK did not impact growth, while a double Tig/DnaK deletion was not viable[19]. We concluded that perhaps the chaperone functions of Tig and DnaK are also redundant in *M. genitalium.*

In contrast, the model predicts that Ribonuclease III (Rnc) is essential for protein production due to its role in ribosome assembly. Specifically, Rnc is responsible for cleaving 30S rRNA into the 5S, 16S, and 23S rRNA precursors without which the ribosomes cannot be formed. Simulated *Δrnc* strains are able to make rRNAs equivalently to wild-type strains, but unable to assemble the rRNAs into ribosomes (**Supplementary Fig. 2c**). The disruption strain grows at a wild-type rate, indicating that *M. genitalium* has a redundant mechanism for performing this essential function. It has been shown in *E. coli* that different ribonucleases, thought to act on distinct types of RNAs, actually may act on a variety of RNAs. For example, RNAse E can act on mRNAs, sRNAs, tRNAs, and shares rRNA cleavage sites with Rnc[20]. While *M. genitalium* does not have a gene annotated as RNAse E, possibly one of the several other ribonucleases (H, J, R, and P) found in the genome is redundant to Rnc.

Chromosome condensation
Interestingly, two of our "hits" reside in the same protein complex, structural maintenance of the chromosome complex (SMC). SMCs are tweezer-like protein complexes that physically pinch the DNA to compact it[21]. In the whole-cell model, SMCs have 6 subunits: two core proteins (tweezer arms) with a set of ScpA and ScpB proteins on each arm[22]. However, since all of the subunits are non-essential[6], we did not model chromosomal compaction by SMCs as essential for growth. The experiments show that both *scpA* and *scpB* disruptions have an impact on the growth rate of the cell. Previous work on *Streptomyces coelicolor* has shown that *scpA* and *scpB* mutants do not affect the ability of the chromosomes to segregate, but rather the morphology of the nucleoid[23]. Our experimental results indicate that *scpA* and *scpB* disruption strains have very similar growth rates, indicating that they may have similar functions (**Supplementary Fig. 2d**). Another "hit", *MG_328*, also has a similar growth rate. A conserved domain search of MG_328 indicated that it has domains similar to *scpB* (e-value: 1.18×10^{-8}, NCBI Conserved Domain Search[24]). This suggests that *MG_328* may have a similar function to *scpA* and *scpB*.

Metabolism

As the largest category of genes (almost 40%) in the whole-cell model are metabolic, it is not surprising that many of the "hits" are metabolic. A closer look at two of the metabolic hits, *MG_114* and *tkt*, suggested the presence of metabolic reactions not present in the Karr *et al.* or Suthers *et al.* reconstructions of the *M. genitalium* metabolic network[2,5]. *MG_114* is a gene in the CDP-diacylglycerol and cardiolipin biosynthesis pathway (**Supplementary Fig. 2e**, black)[2,6]. The implementation of this mostly linear pathway in the framework whole-cell model required a design choice. For cardiolipin and diacylglycerol to be essential lipids in the system, all of the enzymes upstream in the pathway are essential. However, enzyme *MG_114* has previously been characterized as non-essential[6], and our growth rate studies show that in its absence, cells are able to grow at near wild-type rates. Since cardiolipin and diacylglycerol are modeled as essential, the whole-cell model predicts MG_114 to be near essential. If there were an alternative way to make the lipid end products, MG_114 would no longer be a required enzyme in the model. Other organisms, such as *E. coli*, have enzymes, phosphatidate phosphatase (PgpB) and diacylglycerol kinase (DgkA), that can directly catalyze diacylglycerol from a 1,2-diacyl-sn-glycerol-3-phosphate precursor and its reverse reaction (**Supplementary Fig. 2e**, blue). One hypothesis is that *M. genitalium* has enzymes such as PgpB and DgkA that provide alternative routes in the lipid biosynthesis pathway, making MG_114 activity non-essential.

The largest growth discrepancy is of the disruption of *tkt* (transketolase), an enzyme in the pentose phosphate cycle. While experimentally non-essential, in the current metabolic framework of the whole-cell model, a *tkt* disruption is detrimental to growth. Tkt catalyzes a reaction downstream of 5-phosphoribosyl diphosphate production, and in the current framework, no alternative source of 5-phosphoribosyl diphosphate production exists. We performed a reduced cost analysis, similar to that done for *ΔthyA* and *ΔdeoD,* for *Δtkt* (**Supplementary Fig. 2f**), and found that uracil phosphoribosyltransferase (Upp), catalyzes a reaction that uses 5-phosphoribosyl diphosphate as a substrate to produce UMP. We found that if we allow Upp to catalyze its reverse reaction, producing 5-phosphoribosyl diphosphate from UMP, we can reconcile the *tkt* disruption phenotype. While the Upp reaction is not reversible in most organisms, it has been characterized as reversible in some organisms such as *Bacillus subtilis*[25].

References in Supplementary Material:

17. Hartl, F.U., Hayer-Hartl, M. *Science* **295**, 1852-1858 (2002).

18. Davidsen, T., *et al. Nucleic Acids Res.* **38**, D340-D345 (2010).

19. Deurerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., Bukau, B. *Nature* **400**, 693-696 (1999).

20. Stead, M.B., *et al. Nucleic Acids Res.* **39**, 3188-3203 (2010).

21. Mascarenhas, J., *et al. BMC Cell Biol.* **6**, 28 (2005).

22. Porter, I.M., Khoudoli, G.A., Swedlow, J.R. *Curr. Biol.* **14**, R554-R556 (2004).

23. Dedrick, R.M., Wildschutte, H., McCormic, J.R. *J. Bacteriol.* **191**, 320-332 (2009).

24. Marchler-Bauer, A., *et al. Nucleic Acids Res.* **39**, D225-D229 (2011).

25. Grabner, G.K., Switzer, R.L. *J. Biol. Chem.* **278**, 6921-6927 (2003).

26. Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D. *Nat. Chem. Biol*. **5**, 593-9 (2009).

27. Schomburg, I., *et al. Nucleic Acids Res.* **41**,764-772 (2013).