Streptococcus pneumoniae quorum sensing drives an asymmetric owner-intruder competitive strategy during host colonization via the competence regulon

Supplementary Information

Pamela Shen[1], John A Lees[1], Gavyn Chern Wei Bee[1], Sam P Brown[2], and Jeffrey N Weiser[1]*

[1]Department of Microbiology, New York University School of Medicine, New York, New York, USA.

[2]School of Biological Sciences, Georgia Institute of Technology. Atlanta, Georgia 30332, USA.

Supplementary Methods

Computational efficiency

ODE formulation

To improve computation speed when solving the ODE formulation, we used the analytic form of the Jacobian:

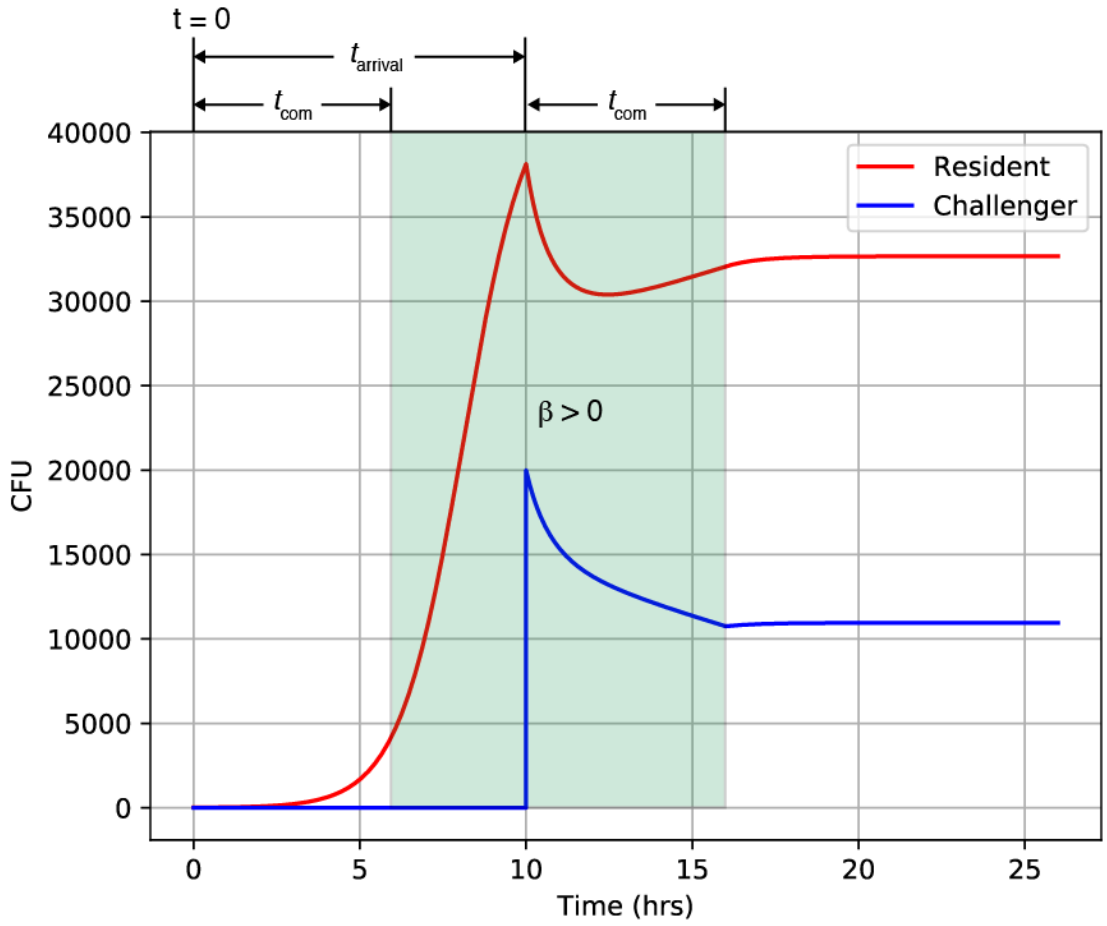
$$J = \begin{bmatrix} r_R\left(\frac{K-2R-\alpha_{CR}C}{K}\right) & -\frac{r_R R\alpha_{CR}}{K} \\ -\frac{r_C C\alpha_{CR}}{K} & r_C\left(\frac{K-2C-\alpha_{RC}R}{K}\right) \end{bmatrix}$$

CTMC formulation

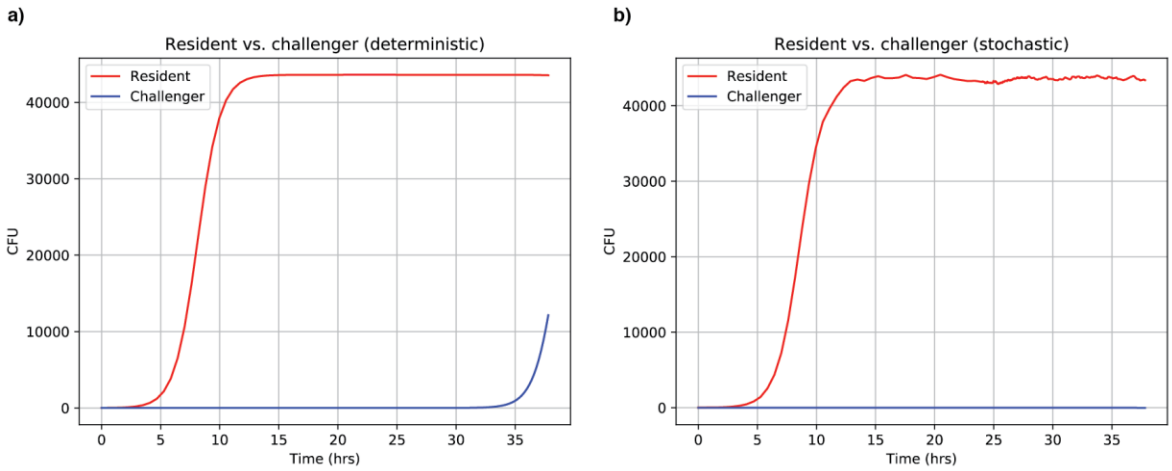
Event	Change	Notation	Probability
Resident cell duplication	$\Delta R = +1$	B_R	$r_R R$
Challenger cell duplication	$\Delta C = +1$	B_C	$r_C C$
Resident cell death	$\Delta R = -1$	D_R	$r_R R\left(\frac{R+\alpha_{CR}C}{K}\right)$
Challenger cell death	$\Delta C = -1$	D_C	$r_C C\left(\frac{C+\alpha_{RC}R}{K}\right)$

The Gillespie algorithm uses exponentially distributed time steps with a mean equal to the total rates in the above table. At each time step one event from the above table is selected with a probability proportional to its rate. This algorithm, even when we optimised it with just-in-time compilation[1], was very slow to run across the experimental conditions we used, and was therefore difficult to perform inferences or predictions with.

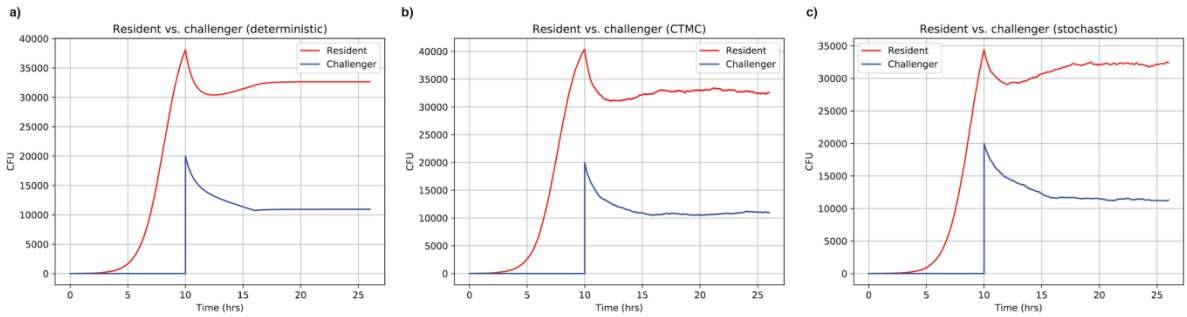
We then considered two alternatives to improve the runtime (Supplementary Table 6). The first was tau-leaping[2], an extension to the Gillespie algorithm which simulates forward in discrete chunks of time, thereby allowing multiple transitions from the above table in each step by generating random Poisson deviates with a mean equal to the expected number of transitions in the time interval. Tau can be set programmatically to avoid negative population sizes[3], however using the maximum rates for the parameters we estimated was slower than the basic Gillespie algorithm. We instead used tau = 10^{-3} hrs, which empirically gave equivalent results in a much shorter computational time.



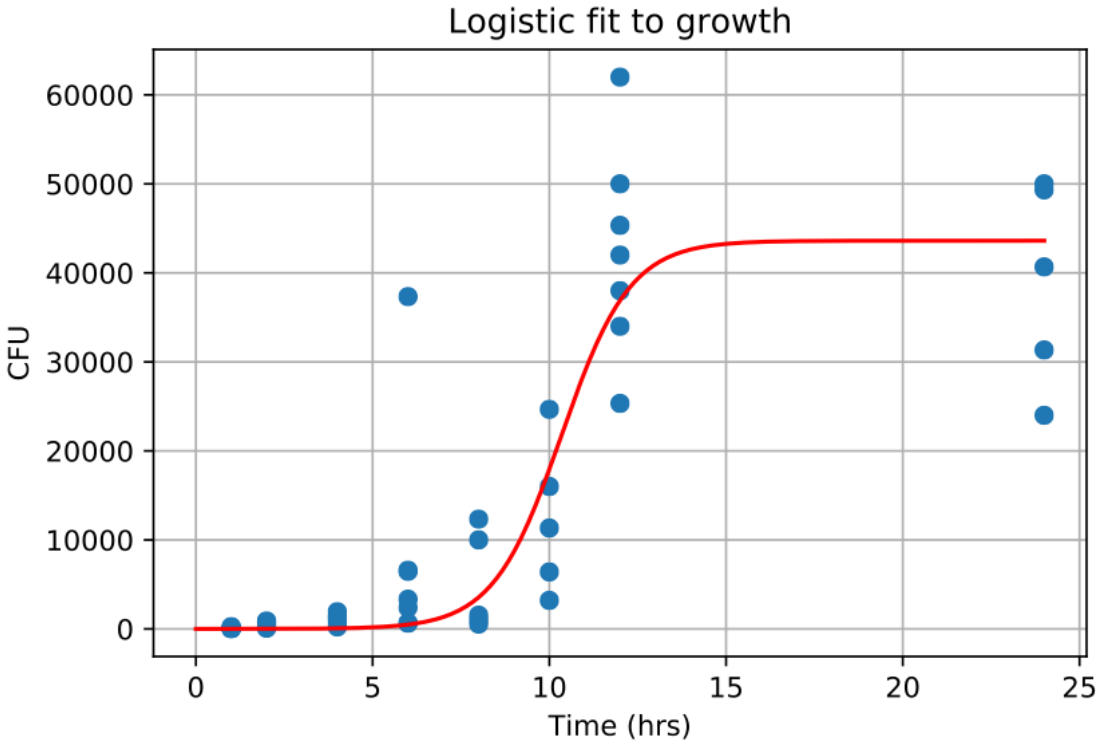
Supplementary Figure 1: Times and Com regulon activation in the mathematical model. This shows an example trajectory of resident and challenger density over time with t_{com} = 6hrs and $t_{arrival}$ = 10hrs with the times relevant to the model marked at the top. The region shaded in green is when the Com regulon is active in the resident but not the challenger, giving it a competitive advantage.



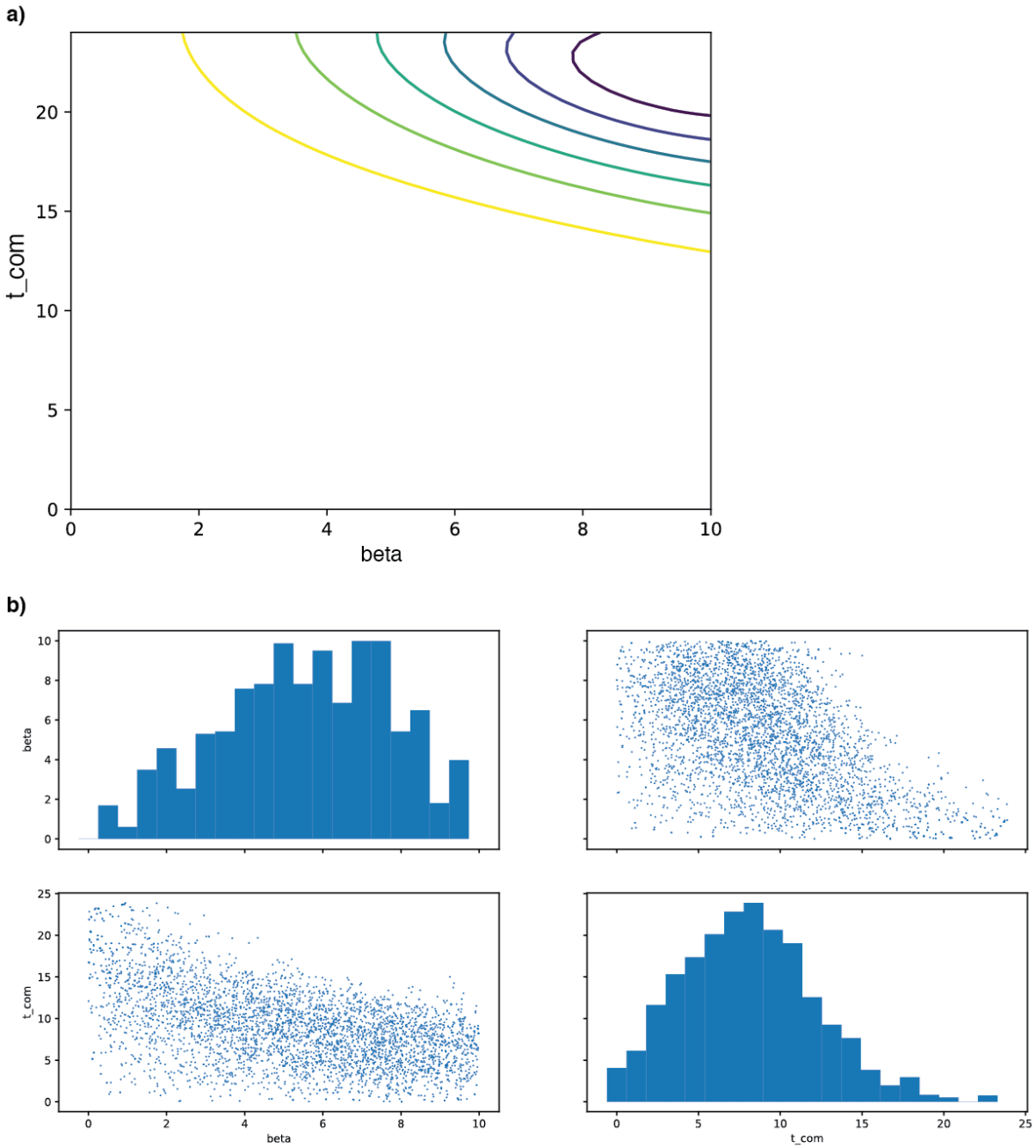
Supplementary Figure 2: The ODE formulation does not accurately model zero populations at long times. The resident and challenger population sizes as a function of time for a) the ODE formulation and b) the SDE formulation. The model was run with the same parameters in each case: α_{CR} and α_{RC} both set to 0.01; t_{arrival} = 24h; initial population sizes 10 CFU. In b) zero is an absorbing state for each population, so once reached (the challenger is excluded) it cannot increase again.



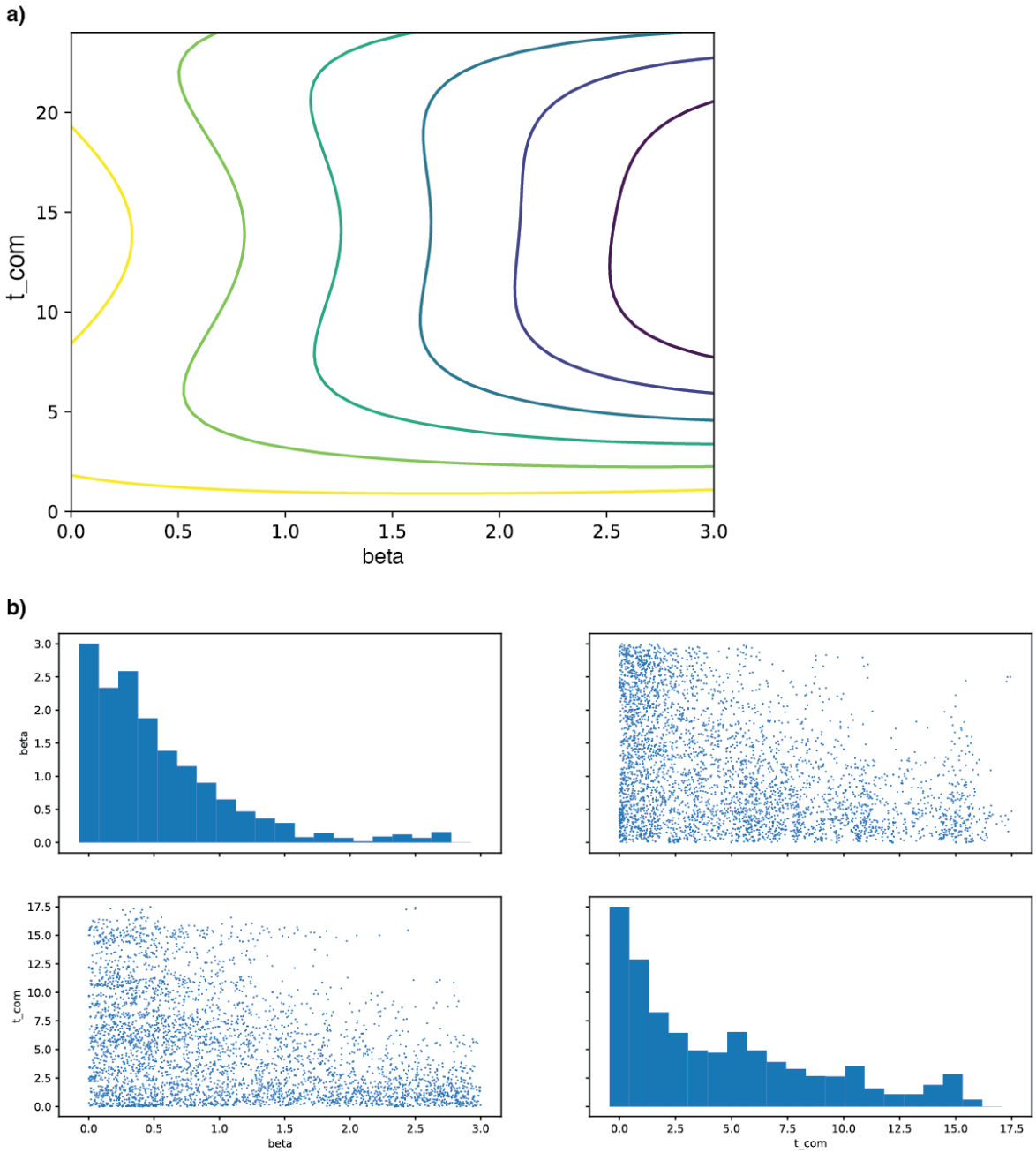
Supplementary Figure 3: Example population trajectories for the three model formulations. The model was run for isogenic resident and challenger, with $t_{arrival}$ = 10 a challenger inoculum of $2x10^4$ CFU and β = 0.1. a) ODE formulation; b) CTMC formulation (simulated with the Gillespie algorithm); c) SDE formulation.



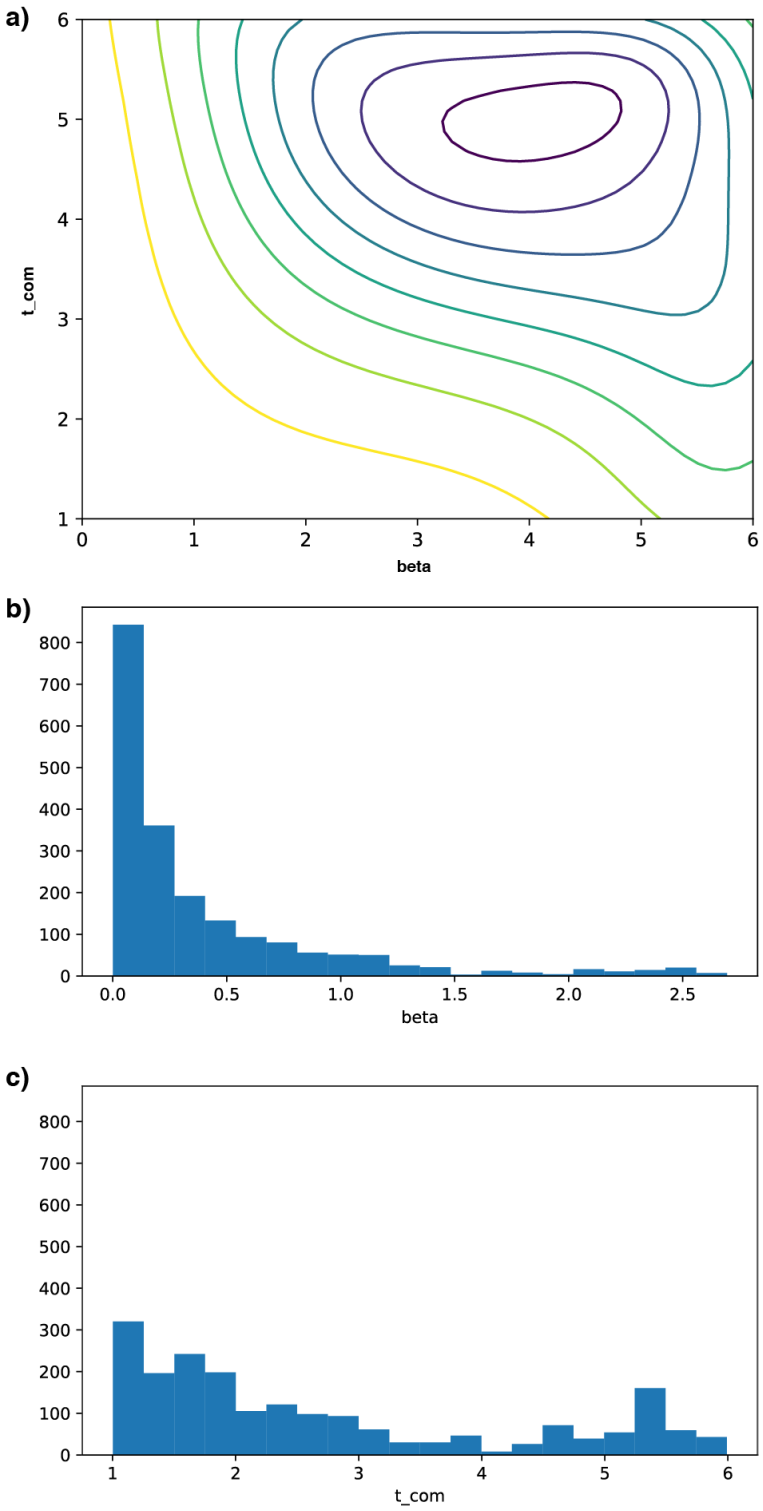
Supplementary Figure 4: Fit of logistic growth to *in vivo* time series data from Figure 3c. The blue dots are observations, and the red line is the least-squares fit with the estimated *r* and *K*.



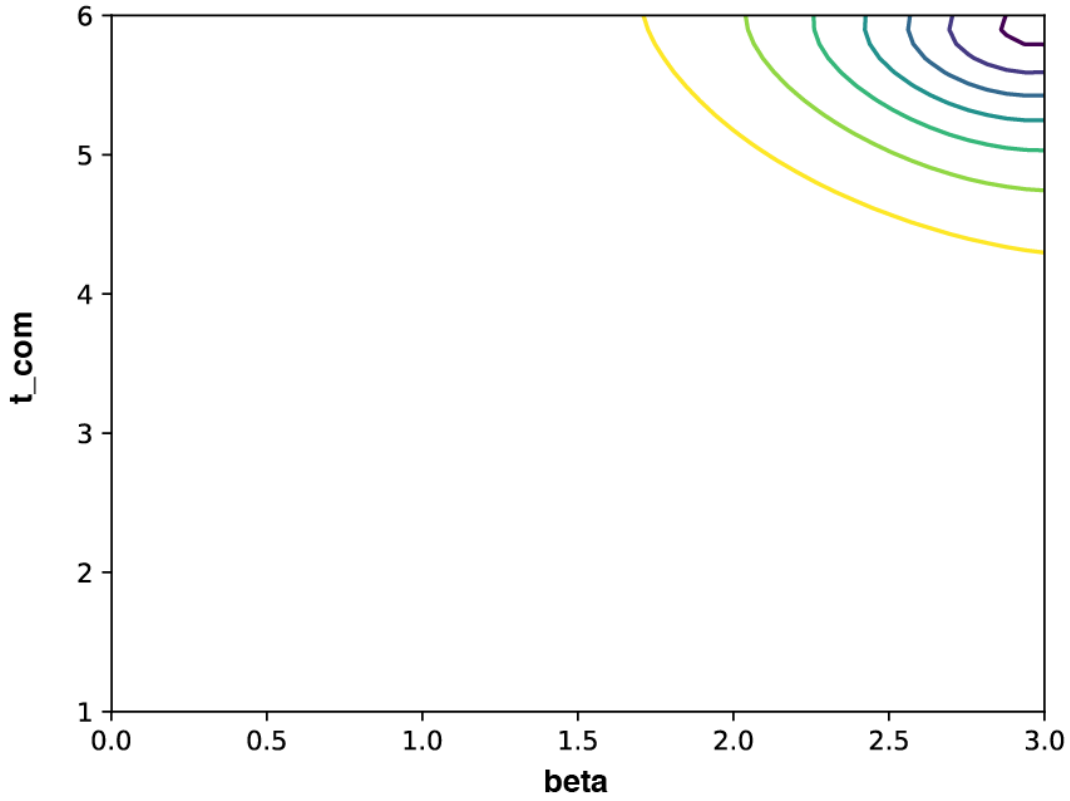
Supplementary Figure 5: BOLFI fit to simulated data with extended priors. Using priors $t_{com} \sim U(0,24)$ and $\beta \sim U(0,10)$ BOLFI fit to simulated data with $t_{com} = 3$ and $\beta = 1$. a) Approximate posterior estimated by BOLFI. b) Samples from the joint and marginal approximate posterior distributions for β and t_{com}. Top row: β. The left panel is a histogram of the approximate marginal posterior, the right panel shows the approximate joint posterior with β on the x-axis and t_{com} on the y-axis. Bottom row: t_{com}. The left panel shows the approximate joint posterior with t_{com} on the x-axis and β on the y-axis, the right panel is a histogram of the approximate marginal posterior.



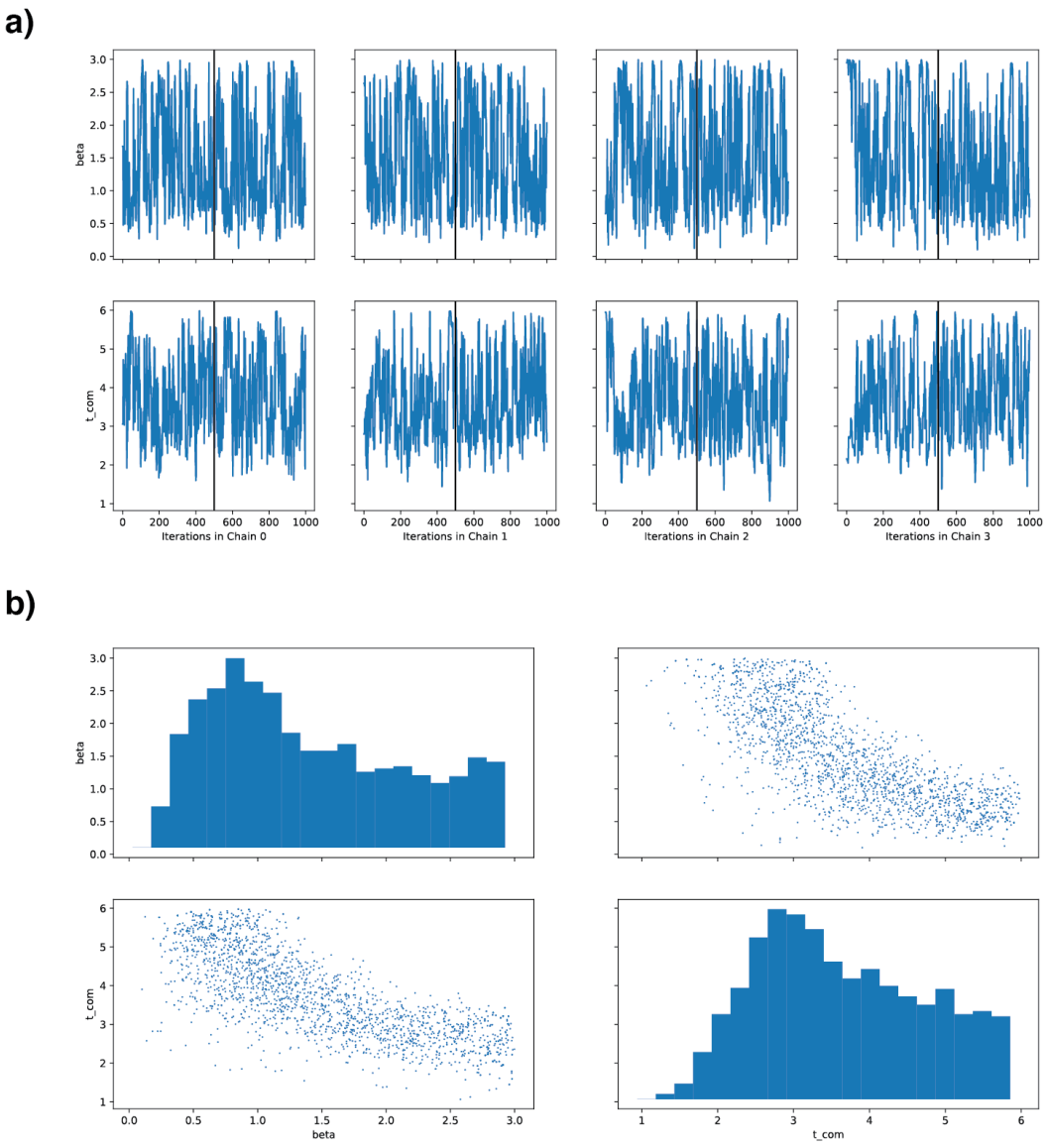
Supplementary Figure 6: BOLFI fit to real data with extended priors. Using priors $t_{com} \sim U(0,24)$ and $\beta \sim U(0,10)$ BOLFI fit to observed experimental data. Posterior means were $t_{com} = 5.11$ hrs and $\beta = 1.09$. a) Approximate posterior estimated by BOLFI. b) Samples from the joint and marginal approximate posterior distributions for β and t_{com}. Top row: β. The left panel is a histogram of the approximate marginal posterior, the right panel shows the approximate joint posterior with β on the x-axis and t_{com} on the y-axis. Bottom row: t_{com}. The left panel shows the approximate joint posterior with t_{com} on the x-axis and β on the y-axis, the right panel is a histogram of the approximate marginal posterior.



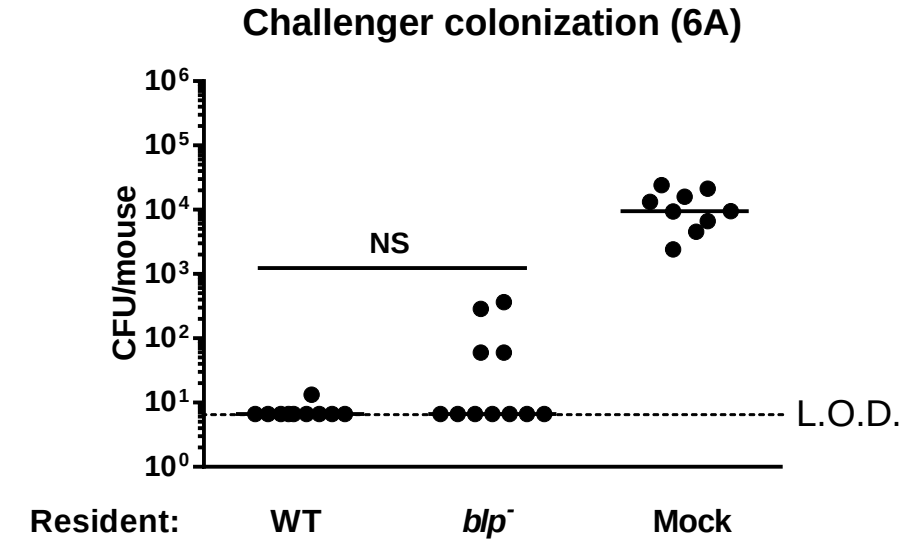
Supplementary Figure 7: Example of BOLFI fit to model-simulated data with actual values β = 0.1 and t_{com} = 5h. a) The approximate posterior likelihood function, shown as contours from yellow (high likelihood) to purple (low likelihood). b) 2000 samples from the posterior likelihood to give the marginal posterior for β. c) 2000 samples from the approximate posterior likelihood to give the marginal posterior for t_{com}.



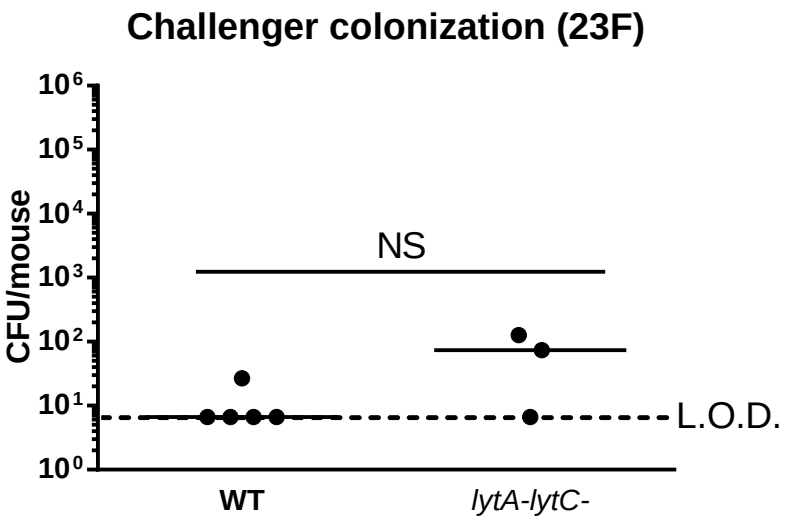
Supplementary Figure 8: The approximate posterior likelihood function given the experimental data in Figure 1. Shown as contours from yellow (high likelihood) to purple (low likelihood).



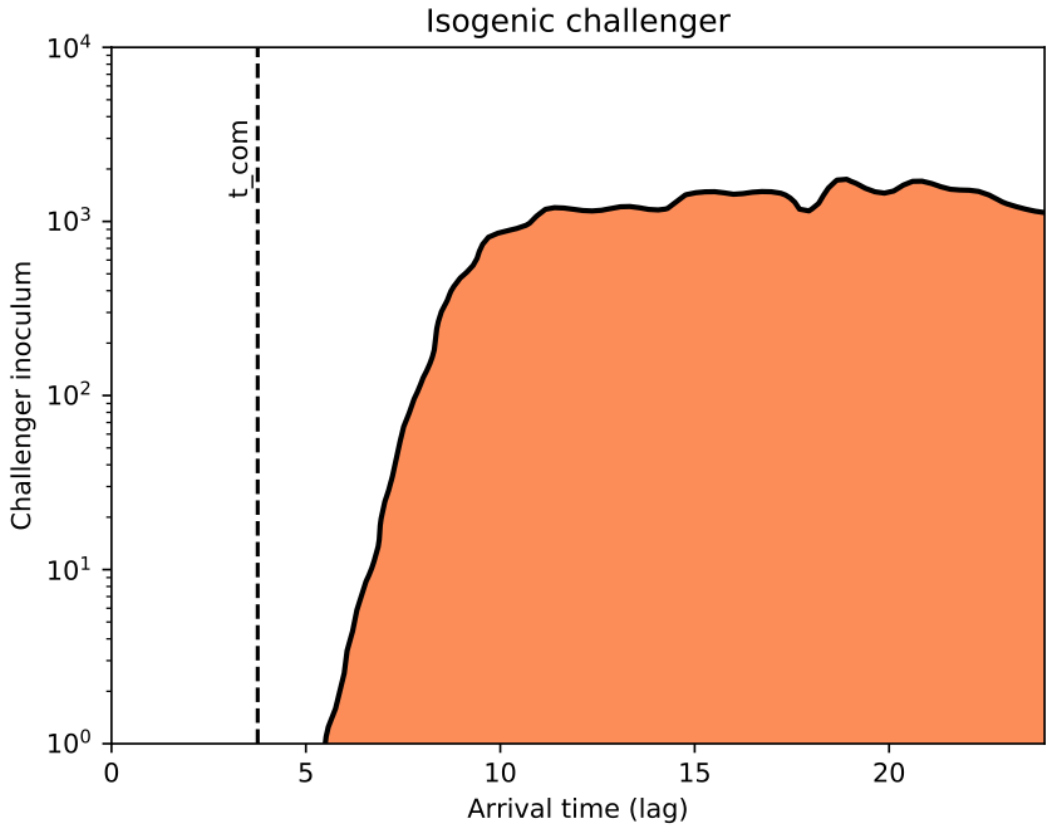
Supplementary Figure 9: Samples from the posterior in supplementary Figure 5. a) Values for β (top row) and t_{com} (bottom row) at each step in four chains (columns). Samples to the left of black vertical lines were discarded as burn-in. b) Samples from the joint and marginal approximate posterior distributions for β (top row) and t_{com} (bottom row).



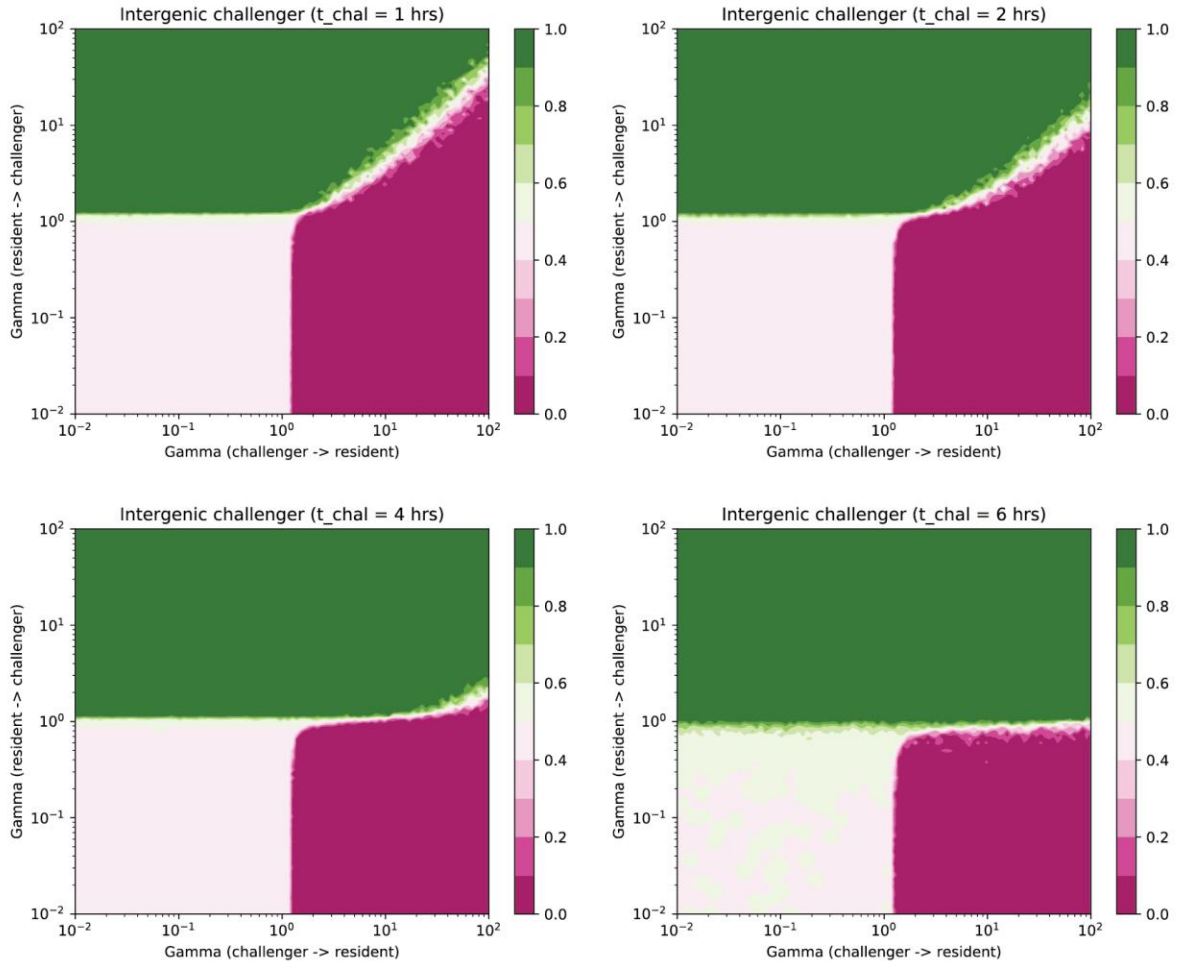
Supplementary Figure 10: Blp does not contribute to asymmetric competition. 4-5 day old pups were intra-nasally inoculated with 10^3 CFU of serotype 6A WT or a *blp*- mutant for 15h. An isogenic challenger was then introduced at 10^1 CFU. Challenger colonization density was determined 3 days later in nasal lavage samples. Groups were compared by Kruskal–Wallis one-way analysis of variance, NS: p>0.9999, n=9-11. Median values are shown. L.O.D., limit of detection, NS, non-significant.



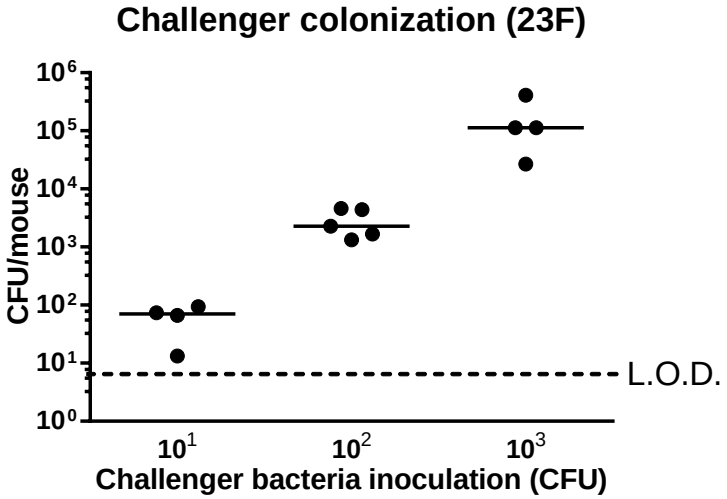
Supplementary Figure 11: LytA and LytC in the challenger do not contribute to asymmetric competition. 4-5 day old pups were intra-nasally inoculated with 10^3 CFU of serotype 23F WT resident for 6-8h. An isogenic WT or *lytA-lytC-* challenger was then introduced at 10^1 CFU. Challenger colonization density was determined 3 days later in nasal lavage samples. Groups were compared by two-tailed Mann-Whitney U test, NS: p=0.3, n=3-5. Median values are shown. L.O.D., limit of detection, NS, non-significant.



Supplementary Figure 12: Isogenic case for 'resident always wins' with the stochastic model. As in Figure 4a, the orange region shows the region of the parameter space for which the resident always wins. Plotted using the average of twenty runs of the stochastic model (with CTMC and tau-leaping), and Gaussian smoothing with variance 0.5 to produce a smooth interpolation in the plot.



Supplementary Figure 13: Domains of intergenic resident vs. challenger using the stochastic model. The same plot as Figure 4c, for earlier times, one per panel: top left, 1h; top right 2h; bottom left 4h; bottom right 6h. Axes are the strength of competition γ between the strains, in the absence of competence. Coloured by average domain over twenty runs of the model, at the extremes green is 'resident wins', pink is 'challenger wins' and in the center white is coexistence. The white region of coexistence in the upper right diagonal for strong competition is an artifact of plotting contours, and is actually a hard boundary (when $\gamma > 1$, one strain always wins).



Supplementary Figure 14: Challenger colonization density was proportional to the inoculum CFU when given at the same time as resident. 4-5 day old pups were intra-nasally inoculated with 10^3 CFU of serotype 23F WT resident or an isogenic challenger (10^1-10^3 CFU) at the same time. Challenger colonization density was determined 3 days later in nasal lavage samples, n=4-5. Median values are shown. L.O.D., limit of detection.

Supplementary Table 1: List of strains used in this study

Strain	Description	Reference or source
P1121	Serotype 23F clinical isolate	[4]
P1547	Serotype 6A clinical isolate	[5]
P1397	Streptomycin resistant P1121	[6]
P2499	P1121 made streptomycin resistant by transformation with a PCR product containing the *rpsL* gene constructed using the genome template of P1726[7]	This study
P2578	P2499 made kanamycin resistant by transformation with genomic DNA from P2405	This study
P2397	Spectinomycin resistant P1547	[5]
P2405	Kanamycin resistant P1547	[5]
P2500	P2499 with Janus cassette insertion into the *cbpD* locus	This study
P2575	P1121 with Janus cassette insertion into the *cibAB* locus	This study
P2516	P2500 with clean deletion of *cbpD*	This study
P2576	P2516 with Janus cassette insertion into the *cibAB* locus	This study
P2577	*cbpD* mutation corrected in P2500 with P2499 DNA, Janus cassette insertion into the *cibAB* locus, corrected *cibAB* mutation with P2499 DNA	This study
P2579	P2516 with Janus cassette insertion into the *comM* locus	This study
P2517	P2499 with Janus cassette insertion into the *lytC* locus and transformed with genomic DNA from a *lytA* mutant P2282[8]	This study
P2076	P1121 with Janus cassette insertion into the *comE* locus	This study
P2444	A streptomycin resistant 6A strain transformed with genomic DNA from a *blp* mutant and made kanamycin resistant with DNA from P2405	This study

Supplementary Table 2 List of primers used in this study

Gene target or purpose	Primer	Sequence (5'-3')
Construction of streptomycin resistant P2499	*rpsL* forward	GATAAGGACAGAACCAGTTCC
	rpsL reverse	GC ATCCTATCTTACCAACGG
Construction of P2500 *cbpD*::*Janus-cassette*	*cbpD* upstream forward	ACTAAGTTGGACAAAACGGTTGCTA
	cbpD upstream reverse:	ATTAAAAATCAAACTTTCATTCTTCCTCCTTGAAAAATAATATAA
	cbpD Janus forward:	TTATATTATTTTTCAAGGAGGAAGAATGAAAGTTTGATTTTTAAT
	cbpD Janus reverse:	AGGAAATTTCTCCTACTCCAATTTTCTATACCTTATGCTTTTGGAC
	cbpD downstream forward:	GTCCAAAAGCATAAGGTATAGAAAATTGGAGTAGGAGAAATTTCCT
	cbpD downstream reverse:	CTGGGATTTTAAAATGCCACAGGAT
Construction of P2575 *cibAB*:: *Janus-cassette*	*cibAB* upstream forward	TTTGTCAGACAAGAGTTCGATATATTCGATATTGTACTCTGGGCG
	cibAB upstream reverse:	ATTATCCATTAAAAATCAAACGGATTGCTAGATAAGAAACACATTTTTAG
	cibAB Janus forward:	CTAAAAATGTGTTTCTTATCTAGCAATCCGTTTGATTTTTAATGGATAAT
	cibAB Janus reverse:	AAAAAAGGAGGAAAGTTCAATGACATTATGCTTTTGGACGTTTAGTACCG
	cibAB downstream forward:	CGGTACTAAACGTCCAAAAGCATAATGTCATTGAACTTTCCTCCTTTTTT
	cibAB downstream reverse:	AATGCATACCAAGTCTGGTCTTGGGACAGATCTGCTTGGATTTGC
Construction of P2516 with clean deletion of *cbpD*	*cbpD* upstream forward	ACTAAGTTGGACAAAACGGTTGCTA
	cbpD upstream reverse:	TCTCCTACTCCAATTTTCTATACTTTCATTCTTCCTCCTTGAAAAA
	cbpD downstream forward:	TTTTTCAAGGAGGAAGAATGAAAGTATAGAAAATTGGAGTAGGAGA
	cbpD downstream reverse:	CTGGGATTTTAAAATGCCACAGGAT
Construction of P2579 *comM*:: *Janus-cassette*	*comM* upstream forward	AATTTCCCTTCTTCTATATATGCCCCACGCTCTTGGCTACCTTCA
	comM upstream reverse:	ATTATCCATTAAAAATCAAACGGATTTTTAGAGAAAGCCTGTTTTTTATG
	comM Janus forward:	CATAAAAAACAGGCTTTCTCTAAAAATCCGTTTGATTTTTAATGGATAAT
	comM Janus reverse:	GTAGGAAGGGAGAGAGAAGATGAAATTATGCTTTTGGACGTTTAGTACCG
	comM downstream forward:	CGGTACTAAACGTCCAAAAGCATAATTTCATCTTCTCTCTCCCTTCCTAC
	comM downstream reverse:	TTAGATGATAGAAATTATGAAAATTTGGAATATATTTATAGAATA
Construction of P2517 *lytC*:: *Janus-cassette*	*lytC* upstream forward	TCAAATTGAGGCCAAGAGAGCAGAA
	lytC upstream reverse:	ATTAAAAATCAAACTTTCAATCTTTCTCTCCTATAAAAAATGTAA
	lytC Janus forward:	TTACATTTTTTATAGGAGAGAAAGATTGAAAGTTTGATTTTTAAT
	lytC Janus reverse:	TCACATCCCTCTTTCAAATCATCGCTTAATATTATGCTTTTGGAC
	lytC downstream forward:	GTCCAAAAGCATAATATTAAGCGATGATTTGAAAGAGGGATGTGA
	lytC downstream reverse:	TATTCTATTTCTTACAAACCAGGTG
Primers for measuring *comX1* by SYBR green PCR	*comX1* forward	AGCAGGAAAGTCAGAAGCGT
	comX1 reverse	TCATCTAGCCAGAGACCCCC
Primers for measuring *comM* by SYBR green PCR	*comM* forward	TGGGACAAGATAGGCTGCAA
	comM reverse	CGTGCGCGATTTTCTTGCTA
Primers for measuring *cbpD* by SYBR green PCR	*cbpD* forward	CTCTGTAGCCATCCACCGTC
	cbpD reverse	GGGCAATGAAAACAGGCTGG
Primers for measuring *Tnfa* by SYBR green PCR	*Tnfa* forward	GACGTGGAACTGGCAGAAGAG
	Tnfa reverse	TTGGTGGTTTGTGAGTGTGAG
Primers for measuring *Ifnb* by SYBR green PCR	*Ifnb* forward	GCACTGGGTGGAATGAGACT
	Ifnb reverse	AGTGGAGAGCAGTTGAGGACA

Supplementary Table 3: Ability to estimate t_{com} and β from model simulations. A simulation of the model with t_{com} and β specified as in the outer columns and rows, followed by a BOLFI fit to these simulations. The table entries show the mean posterior obtained for each parameter, with the numbers in brackets the 95% HPD. See supplementary figure 4 for a specific example with the posterior and samples.

	β = 0.3	**β = 1.1**	**β = 2.0**
t_{com} = 2 hrs	t_{com} = 3.48 (1.03-5.75) hrs β = 0.63 (0.001-1.85)	t_{com} = 3.03 (1.02 – 5.79) hrs β = 1.53 (0.12-2.90)	t_{com} = 3.41 (1.10-5.81) hrs β = 2.00 (0.673-3.00)
t_{com} = 4 hrs	t_{com} = 2.40 (1.00-5.25) hrs β = 1.10 (0.002-2.59)	t_{com} = 3.47 (1.36-5.98) hrs β = 1.94 (0.543 – 3.00)	t_{com} = 3.94 (1.49 – 5.99) hrs β = 1.87 (0.443-3.00)
t_{com} = 6 hrs	t_{com} = 2.70 (1.00-5.58) hrs β = 1.09 (0.009 – 2.49)	t_{com} = 4.10 (1.61-6.00) hrs β = 1.85 (0.533-2.99)	t_{com} = 3.96 (1.51- 6.00) hrs β = 1.83 (0.509 – 2.99)

Supplementary Table 4: See excel file SI Table 4.

Supplementary Table 5: Conservation of competence machinery in serotype 3 genomes. We have separately calculated the frequency and dN/dS of each gene in the entire Massachusetts population of 616 genomes and 93 serotype 3 only genomes.

Gene	TIGR4 ID	CLS ID	Isolates present (/616)	Isolates present (/93)	dN/dS (all)	dN/dS (serotype 3)
cbpD	SP_2201	CLS00029	616	93	0.35	0.32
comM	SP_1945	CLS01685	616	92	0.19	0.68
cibA	SP_0125	CLS00190	605	93	0.59	0.17
cibB	SP_0124	CLS00189	605	93	0.52	0.12
cibC	SP_0122	CLS00187	616	93	0.27	0.53
comA	SP_0042	CLS00122	616	93	0.16	0.22
comB	SP_0043	CLS00123	615	93	0.45	0.28
comC	SP_2237	CLS00064	614	93	0.71	0.74
comD	SP_2236	CLS00063	616	93	0.26	0.61
comE	SP_2235	CLS00062	616	93	0.34	0.40
comEB	SP_0744	CLS00678	615	93	0.12	0.32
comEA	SP_0954	CLS00859	614	93	0.24	0.16
comEC	SP_0955	CLS00860	608	93	0.34	0.25
comX	SP_0014/ SP_2006	CLS01734	605	93	0.31	1.43
comYF	SP_2048	CLS01771	616	93	0.30	0.23
comYE	SP_2049	CLS01772	616	93	0.38	0.06
comYD	SP_2050	CLS01773	614	93	0.37	0.24
comYC	SP_2051	CLS01774	616	93	0.84	0.32
comYB	SP_2052	CLS01775	616	93	0.18	0.23
comYA	SP_2053	CLS01776	616	93	0.14	0.22
comFC	SP_2207	CLS00035	616	93	0.12	0.24
comFA	SP_2208	CLS00036	616	93	0.10	0.80

Supplementary Table 6: Comparison of times taken to solve model equations using numerical integration. For each method the time taken to run 100 integrals with parameters '--t_com 3.8 --t_chal 1 --C_size 10 --beta 1.48 --R_size 10 --t_end 36 --g-RC 0.01 --g-CR 0.01 --resolution 2000' is shown. For the CTMC solutions, functions optimised by using just-in-time (JIT) compilation with numba, which increased their speed roughly five-fold.

Model	**Method**	**Time for 100 integrals**
ODEs	scipy.odeint	0.28 s
SDEs	sdeint.itoint	39.07 s
CTMC	Gillespie (JIT)	74.74 s
	Tau-leaping (JIT) [tau = 0.001h]	6.38 s

Supplementary References

1. Lam, S. K., Pitrou, A. & Seibert, S. Numba: A LLVM-based Python JIT Compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* 7:1-7:6 (2015). doi:10.1145/2833157.2833162
2. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115,** 1716–1733 (2001).
3. Cao, Y., Gillespie, D. T. & Petzold, L. R. Avoiding negative populations in explicit Poisson tau-leaping. *J. Chem. Phys.* **123,** 54104 (2005).
4. McCool, T. L., Cate, T. R., Moy, G. & Weiser, J. N. The immune response to pneumococcal proteins during experimental human carriage. *J. Exp. Med.* **195,** 359–365 (2002).
5. Kono, M. *et al.* Single Cell Bottlenecks in the Pathogenesis of Streptococcus pneumoniae. *PLoS Pathog.* **12,** (2016).
6. Zangari, T., Wang, Y. & Weiser, J. N. Streptococcus pneumoniae transmission is blocked by type-specific immunity in an infant mouse model. *MBio* **8,** (2017).
7. Zafar, M. A., Wang, Y., Hamaguchi, S. & Weiser, J. N. Host-to-Host Transmission of Streptococcus pneumoniae Is Driven by Its Inflammatory Toxin, Pneumolysin. *Cell Host Microbe* **21,** 73–83 (2017).
8. DeBardeleben, H. K., Lysenko, E. S., Dalia, A. B. & Weiser, J. N. Tolerance of a phage element by Streptococcus pneumoniae leads to a fitness defect during colonization. *J. Bacteriol.* **196,** 2670–2680 (2014).