www.sciencemag.org/cgi/content/full/333/6045/984/DC1

Supporting Online Material for

Partitioning Regulatory Mechanisms of Within-Host Malaria Dynamics Using the Effective Propagation Number

C. J. E. Metcalf,* A. L. Graham, S. Huijben, V. C. Barclay, G. H. Long, B. T. Grenfell, A. F. Read, O. N. Bjørnstad

*To whom correspondence should be addressed. E-mail: charlotte.metcalf@zoo.ox.ac.uk

Published 19 August 2011, *Science* **333**, 984 (2011)
DOI: 10.1126/science.1204588

This PDF file includes:

Materials and Methods

Effective propagation P_E

The model defined by [1] can be fit using any standard statistical software (e.g., using the 'lm' function in R (*45*) with $\log(\mathbf{I}_t) + \log(\mathbf{S}_t)$ as an 'offset'). More exact estimation would use a generalized linear model, with normal or lognormal errors, as the log of the expectation will differ from the expectation of the log. However, results are robust to the error distribution and link used, so we retained the linear regression for simplicity here.

RBC dynamics

Change in susceptible (uninfected) RBCs, $\mathbf{S}_t$ follows

$$\mathbf{S}_{t+1} = \mathbf{S}_t - \mathbf{M}_{t+1} + b_t, \quad \text{(a)}$$

where $\mathbf{M}_t$ is total infected RBCs (encompassing both asexual infected RBCs, $\mathbf{I}_t$, capable of continued within-host infection; and sexual infected RBCs, which cannot), and b_t is change in RBC densities beyond loss to infection. Erythropoiesis is thought to track RBC densities, i.e., $b_t = c_1 + c_2\mathbf{S}_{t-\tau} + \varepsilon$ where τ is an unknown response lag. This can be fit via regression. The RBC age profile over the course of the infection can be reconstructed by noting that the equilibrium age structure consists predominantly of day 4 and older RBCs (since RBC survival probability is >0.9 each day), and by assuming that b_t calculated for day t corresponds to 'new' one day old RBCs at t if b_t>0. One day olds at t can then be projected forward to obtain two day olds present at t+1, etc.

Partitioning effective propagation

We fit a nested series of increasingly complex models with the same structure as [1], e.g. models assuming constant effective propagation,

$$\log(\mathrm{E}[(\mathbf{M}_{t+1})]) = \log(P_{E,c}) + \log(\mathbf{I}_t) + \log(\mathbf{S}_t), \quad \text{(b)}$$

where $P_{E,c}$ is the constant effective propagation rate; models allowing different invasion rates for different aged susceptible RBCs,

$$\log(\mathrm{E}[(\mathbf{M}_{t+1})]) = \log(P_{E1}\sum_1^3 \mathbf{S}_{a,t} + P_{E2}\mathbf{S}_{4,t}) + \log(\mathbf{I}_t) \quad \text{(c)}$$

where P_{E1} and P_{E2} indicate effective propagation into the 1-3rd, and 4th RBC age class (both uninfected RBC invasion and infected RBC burst size may be age-specific, but

uninfected RBC age is accessible, so we focused on invasion rates here); models incorporating a dose-dependent response to innate immunity with the fraction of infected cells surviving innate immunity defined as $s_t = \log(\mathbf{M}_t)/(a + \log(\mathbf{M}_t))$ for $t \le t_c$; $s_t = 1$ otherwise, yielding

$$\log(E[(\mathbf{M}_{t+1})]) = \log(P_{E1}\sum_1^3 S_{a,t} + P_{E2}\mathbf{S}_{4,t}) + \log(\mathbf{I}_t) + \log(s_t) \quad \text{(d)}$$

Throughout, discrepancies between $P_{E,t}$ predicted from (b-d) and the empirically derived pattern of $P_{E,t}$ from [1] indicate the relative importance of the components.

Inferring adaptive immunity

From the first day with observations of uninfected and asexual infected RBCs, we calculate infected RBCs expected one day later in the absence of adaptive immunity with parameter estimates for age specific effective propagation (P_{E1}, P_{E2}) and innate immunity (t_c, a), setting uninfected RBCs on day 1 to be predominantly 4 days old or more. Predictions of total infected RBCs will be larger than observed since adaptive immunity is not included (i.e., $\hat{\mathbf{M}}_2 > \mathbf{M}_2$ where the ˆ indicates the predicted vector), and we can calculate the fraction 1-p_t of infected cells required to survive adaptive immunity to obtain the observed, e.g., for t=2, $\mathbf{M}_2 = p_2\hat{\mathbf{M}}_2$, etc. Similarly, uninfected RBCs are projected forward using (a) with b_t predicted from estimates of c_1 and c_2 parameters (Figure S5), and $\mathbf{M}_t$ taken from the data. The smoothed surface is obtained using the 'gam' function in the 'mgcv' package in R (*45*); see Fig. S11 for the raw data.

Supporting Online Material Text: Description of the data, details of model fitting, uncertainties in modeling RBC age structure and dynamics

Description of the data

Plasmodium chabaudi parasites were originally isolated from thicket rats and subsequently cloned (*46*). The first dataset (Fig. S2) used clone AS_{12062} (letters denote the clone lineage, subscript denotes the precise clonal history) to infect $CD4^+$ T-cell depleted mice as described previously (*47*). Briefly, the rat monoclonal antibody, GK1.5, was used to deplete $CD4^+$ T cells. Five experimental mice were injected intra-peritoneally with 500µg of the appropriate purified antibody in phosphate-buffered saline (PBS) 5 days before parasite challenge, and then by 250µg 4 days and 1 day before parasite challenge and weekly after challenge. Hosts were inbred female C57Bl/6 mice (Harlan, England). A fluorescent activated cell sorter (FACS) was used to confirm $CD4^+$ T-cell depletion. The five resultant $CD4^+$ T-cell depleted mice were challenged intra-peritoneally with 10^6 AS parasites. During the course of infection (days 3-21) blood samples were taken from the tail to estimate RBC density (by flow cyometry; Beckman Coulter) and for genotype-specific real-time quantitative PCR (qPCR) assays as described previously (*48*).

The second dataset (Fig. S3) used the clone $AS_{6p(pyr\text{-}1A)}$, a clone that was made resistant to pyrimethamine by laboratory selection (Walliker et al. 1975). Hosts were 15 week old female C57Bl/6 laboratory mice (Charles River Laboratories). The inoculum of clone AS consisted of 10^6, 10^5, 10^3 or 10^1 parasites injected intra-peritoneally, as described previously (*23*). Each treatment group consisted of five mice, except for the groups with an inoculum of 10^1 parasites which consisted of 10 mice to allow for uninfected hosts. RBC density, asexual parasite density and gametocyte density were measured daily (day 3-21 post infection), for which respectively two µl, 5 µl and 10 µl of blood was taken by tail snip. Red blood cell density was measured using flow cytometry, asexual parasite density and gametocyte density were measured using respectively DNA and RNA extraction of the blood and subsequent quantitative PCR as described in (*49*). Estimates of numbers of gametocyte-dedicated infected RBCs rely on a gene expressed two days

after the allocation decision is made, so gametocyte allocation was estimated by lagging numbers by two days.

The third dataset (Fig. S4) used the AS clone AS_{12280} in 6-8 week old female inbred C57BL/6 mice (bred in house). IL-10 signalling was neutralized *in vivo* by the intra peritoneal injection of 20μg of neutralizing anti-IL-10R monoclonal IgG antibody (clone 1B1.3a; BD PharMingen) as previously described (*24*). Groups of mice were injected i.p. with control IgG (Sigma-Aldrich) or αIL-10R and, the following day, single-clone infections were initiated with an inoculum consisting of 10^6 parasites injected intra-peritoneally. Antibody treatments were applied again on days 1 to 4 post-infection inclusively. Red blood cell density, asexual parasite density and gametocyte density were measured daily (day 4-21 post infection). Red blood cell density was measured using flow cytometry, thin blood smears were used to measure asexual parasite density and gametocyte density as described in (*24*).

At any point in an infection, gametocytes typically constitute less than 1% of parasites (e.g. (*50*)). In the first data set, where gametocyte data were not available, modeled data include all parasites. In the second and third data sets, modeled data are total parasite numbers minus gametocytes which gives asexual parasite density. Sensitivity analyses show that the inclusion or exclusion of gametocyte data made no difference to our conclusions (data not shown) because gametocyte densities are usually at least two orders of magnitude lower than the densities of asexual parasites. Estimates of parasite densities and uninfected red cell densities are highly repeatable across replicate samples taken from individual mice (*11, 12, 50*), showing that parasites and red blood cells are well mixed in peripheral blood.

Details of fitting models to data

To estimate time-varying effective propagation, we grouped mice by strain genotype (Dataset 1) or treatment group (Dataset 2, 3) and fitted an ANCOVA to resulting groups, with log infected and uninfected RBCs as offsets, and "day" fitted as a factor. For Dataset 1, we adjusted the time-course to align timing of peak parasitaemia across mice, and since data was not available to distinguish $\mathbf{M}_t$ and $\mathbf{I}_t$ we set $\mathbf{M}_t=\mathbf{I}_t$ (the difference should be slight as gametocyte production is very rare relative to

merozoite production (*51*)). For Dataset 2, mice were measured every other day for $t>21$, so we fitted smooth splines to time-series, then estimated $P_{E,t}$ with missing values replaced by predictions from the smoothing. In all three datasets, we found no evidence for individual host differences (no significant individual variation identified via a mixed effects model) and disregarded this subsequently.

Unlike the models defined by (a-b) in the Methods, the models defined by (c-d) cannot be directly fitted using regression approaches. Instead, we can maximize the log likelihood of the predicted values of $\mathbf{M}_{t+1}$ given the observed values using an estimate of observation error. Using additional data consisting of 5 repeated measures on parasite densities at 7 different time points over the course of an infection, we subtracted the mean log measure for each mouse or time-point, then estimated the standard deviation and took the overall mean standard deviation as the global observation error. The value obtained was 0.14.

Note that the models used are based on the dynamics of an ideal gas, where infected and uninfected RBCs move and meet randomly. If local density was important to transmission between RBCs (effectively one RBC was interacting with a limited number of other RBCs due to spatial packing), the dynamics might be expected to behave as if 'frequency dependent' rather than 'density dependent' (1). Frequency dependency can be captured by modeling $I_{t+1}=\beta I_t S_t/N_t$, where I_t and S_t are number of susceptible and infected individuals, N_t is the total population size and β captures the scale of transmission. Identifying where a process lies on the continuum between frequency and density dependence can be tackled by modeling $I_{t+1}=\beta I_t S_t/[N_t]^k$ - if k is close to zero, this suggests that density dependence is predominant (2), and the model collapses to the form fitted throughout the manuscript (see eqn 1); frequency dependence is suggested when k is close to 1. In our framework, this can be addressed by expanding eqn [1] to fit the regression:

$$\log(\mathrm{E}[(I_{t+1})]) = \log(P_{E,t}) + \log(I_t) + \log(S_t) + k\log(N_t)$$

We fit this model to the data available in CD4+ T cell-depleted mice (Dataset 1; chosen to minimize interference from known dose effects on immunity). We compared the fit with that obtained via eqn [1], shown in Fig. 1. Including the extra

parameter did not significantly improve the model (likelihood ratio test, LR=0.14, df=1,p>0.1); and furthermore, the value of *k* was not significantly different from zero, with an estimate of 0.11(±0.32); brackets indicate standard errors. We therefore focus on the density dependent formulation in the main text. Biologically, we might also expect this since there is no large-scale evidence of compartmentalization or other types of spatial or temporal clumping effects of malaria in rodent blood. Nevertheless, this is an interesting area for future investigations, especially in the context of 'rosetting', a behaviour whereby infected RBCs bind to and surround themselves with uninfected RBCs. This process should in principle leave a detectable signature of 'frequency-dependence-like' transmission. Comparison of statistical models like the above across strains known to differ in the frequency of resetting could provide insight into its relative importance.

Modeling RBC dynamics

Since the range of b_t incorporates zero, fitted process models of RBC dynamics (eqn a) reflect a combination of RBC replenishment and destruction. The analysis supports a model in which RBC homeostasis operates with a one-day lag in mice with impaired immune function (Fig. S5), and up to a 3-day lag for intact mice (although differences between the r^2 of models of different lags are slight). Overall, the relationship between RBC density and change in RBC (corrected for infection) suggests that a simple linear model of RBC production and regulation may be adequate to reflect up-regulation of RBC production during times of anemia (*11, 12*). However there is considerable variation that is not explained by a simple homeostatic model (Fig. S5). The source of this variation merits future attention.

The pattern over time of RBC destruction and replenishment relative to RBC density hints that the slope of transmission may increase at later time-steps, e.g., for a lag of $\tau=1$, in intact mice, at equivalent low levels of RBCs, the level of replenishment is greater for colours further into the green spectrum, i.e. later in the time-series (see legend, Fig. S5). We can test this statistically by fitting 'day' as a covariate into the regression. For a lag of one day, the minimum adequate model

included an interaction with day *d* and RBC density (N=39, r^2=0.19, p<0.01) of the form

$$b_t = -2939 - 0.013\mathbf{S}_{t-\tau} + 1401\,day - 0.016\,day * \mathbf{S}_{t-\tau} + \varepsilon_t \ .$$

This equation indicates that more new RBCs are produced later during the time-series, but even more are produced at lower RBC densities later. The mechanisms contributing to this may include increased up-regulation of RBC production in response to anemia dipping below a threshold (*11, 12, 41*), and decreases or increases in the rate of bystander killing (*4, 11*). However, resolving which contribute when and to what degree is complex as the timing of both processes may vary (*11*) and are essentially statistically confounded. As disentangling the mechanistic determinants of this effect is intractable, and since the model improvement is slight, we do not incorporate this to our analysis, but it is an interesting question for future work and experiments.

Uncertainties in inferred RBC age structure

Our proxy for age structure will be most accurate if clearing of uninfected RBCs and bursting of RBCs mostly affects older RBCs, so that one-day-olds can be projected a day forwards to obtain two-day-olds without too great a loss via RBC mortality. Effective propagation is estimated to be greater in older RBCs (Fig. S5), and clearing by e.g. the spleen is thought to affect mostly older RBCs (*30, 41*), so the inferred pattern is likely to reasonably capture the true age structure, although discrepancies may be present.

For the lowest starting density group in Dataset 2, including the age effect in modeling effective propagation (eqn c) still leads to underestimation of the empirically estimated effective propagation early in the time-series (Fig. S9, grey squares are present above and below for the last panel). Parasitaemia increases most slowly in this group, so that the shift towards younger age classes of RBCs may be less abrupt, and therefore harder to detect, biasing the timing of the balance of RBCs entering the model described by eqn c to earlier or later than it should be; resulting also in slightly under-estimation in the forward simulations in this low

density group (Fig. S11). Direct data on age structure could identify whether some other process is occurring in low density mice, or mis-estimation of the age-structure is responsible for model deviances.

Uncertainties in inferred killing surface by later induced immunity

The established six-parameter model used to capture fraction of infected cells killed (p_t) slightly under-estimates the numbers of parasites starting at low densities across the expansion phase (Fig. S11, last row, beyond day 5). Potential explanations include underestimation of RBCs (for which this inoculum size shows the greatest variability, see above). The result may be an under-estimate of immune efficacy early during the time-series in individuals inoculated with small doses.

Irregularities in the smoothed surface, particularly between peaks in parasitaemia, may partly reflect statistical artifacts due to small numbers of parasites, but most dips are associated with transient parasite release events (Fig. S3, second waves, see Fig. S11 for their timing) and therefore are likely to capture the biological reality. The magnitude of immune activation required later in the time-series may be slightly under-estimated, since the abundance of uninfected RBCs is under-estimated (Fig. S11, by up to 20% at the higher densities).

References

45. R-Core-Development-Team, *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2009).
46. G. H. Beale, D. Walliker, R. Carter, in *Rodent malaria.,* R. Killick-Kendrick, W. Peters, Eds. (Academic Press, London, 1978), pp. 213-245.
47. V. C. Barclay *et al.*, *Proc R Soc Lond B Biol Sci* **275**, 1171 (2008).
48. A. S. Bell, J. C. Roode, D. Sim, A. F. Read, *Evolution* **60**, 1358 (2006).
49. D. R. Drew, S. E. Reece, *Molecular and Biochemical Parasitology* **156**, 199 (2007).
50. S. Huijben *et al.*, *Evolution* **64**, 2952 (2010).
51. A. F. Read, L. H. Taylor, *Parasitology Today* **13**, 135 (1997).

Figure S1: *Plasmodium chabaudi* in the blood. Dynamics follow a Susceptible-Infected-like process (i.e., no persistence of infected cells between time-steps), but with infected mortality via immunity, and complicated patterns of susceptible replenishment and removal, as well as age effects on RBC invasion and burst size with the consequence that the relative importance of factors shaping effective propagation vary through time.

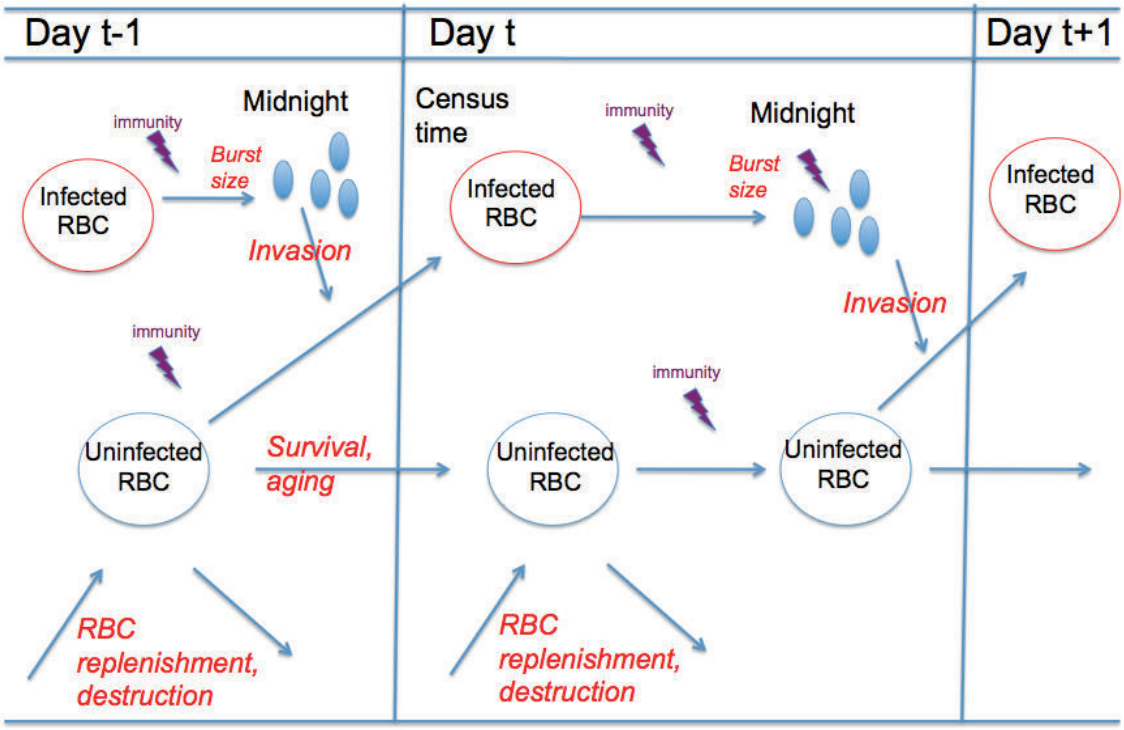

Figure S2: Time-series in Dataset 1. Each point corresponds to a measurement on a CD4+ T-cell depleted mouse on a particular day for the AS strain (n=5 mice). Axes of time-series are logged, with values shown $x10^{-2}$ per µl on an unlogged scale.

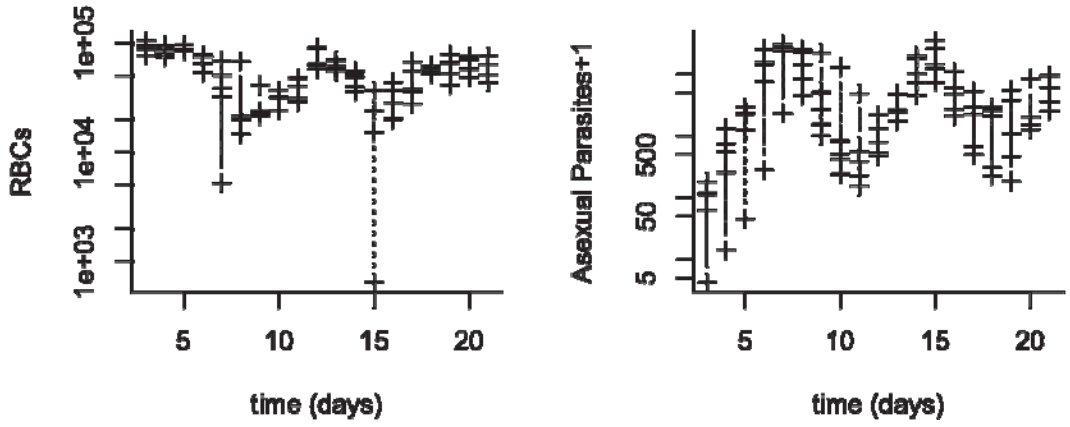

Figure S3: Time-series in Dataset 2. Each point corresponds to a measurement on an intact mouse on a particular day, and each row corresponds to a different starting density (10^1 in black, 10^3 in red, 10^5 in green, 10^6 in blue); axes of time-series are logged, with unlogged values shown $x10^{-2}$ per µl.

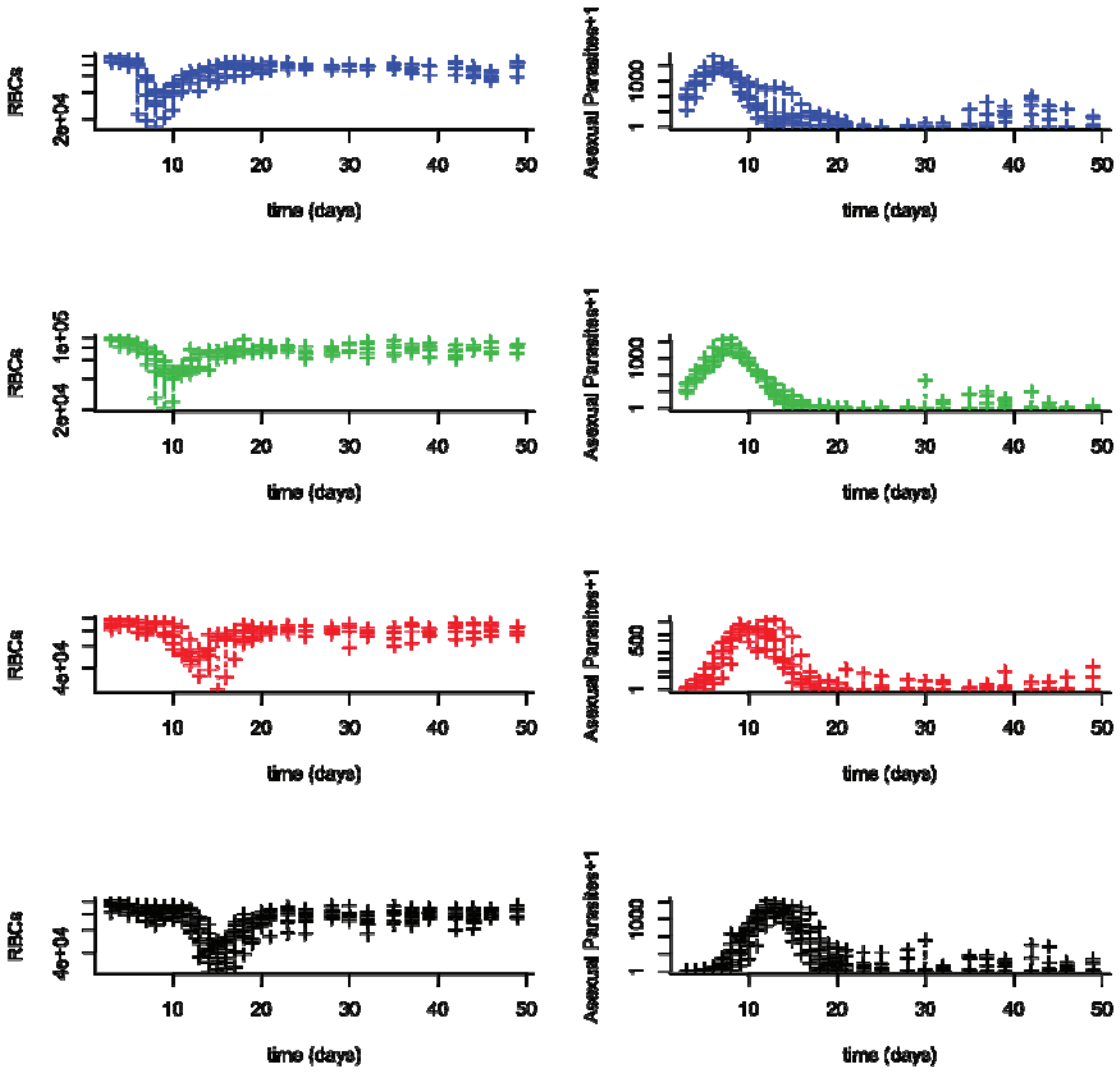

Figure S4: Time-series in Dataset 3. Each point corresponds to a measurement on a mouse on a particular day, and each column corresponds to a different treatment (control on the left, corresponding to IgG treatment on days 1-4; up-regulation of innate immunity on the right, corresponding to IL-10R treatment on days 1-4 on the right); axes of time-series are logged, with unlogged values shown $x10^{-2}$ per µl.

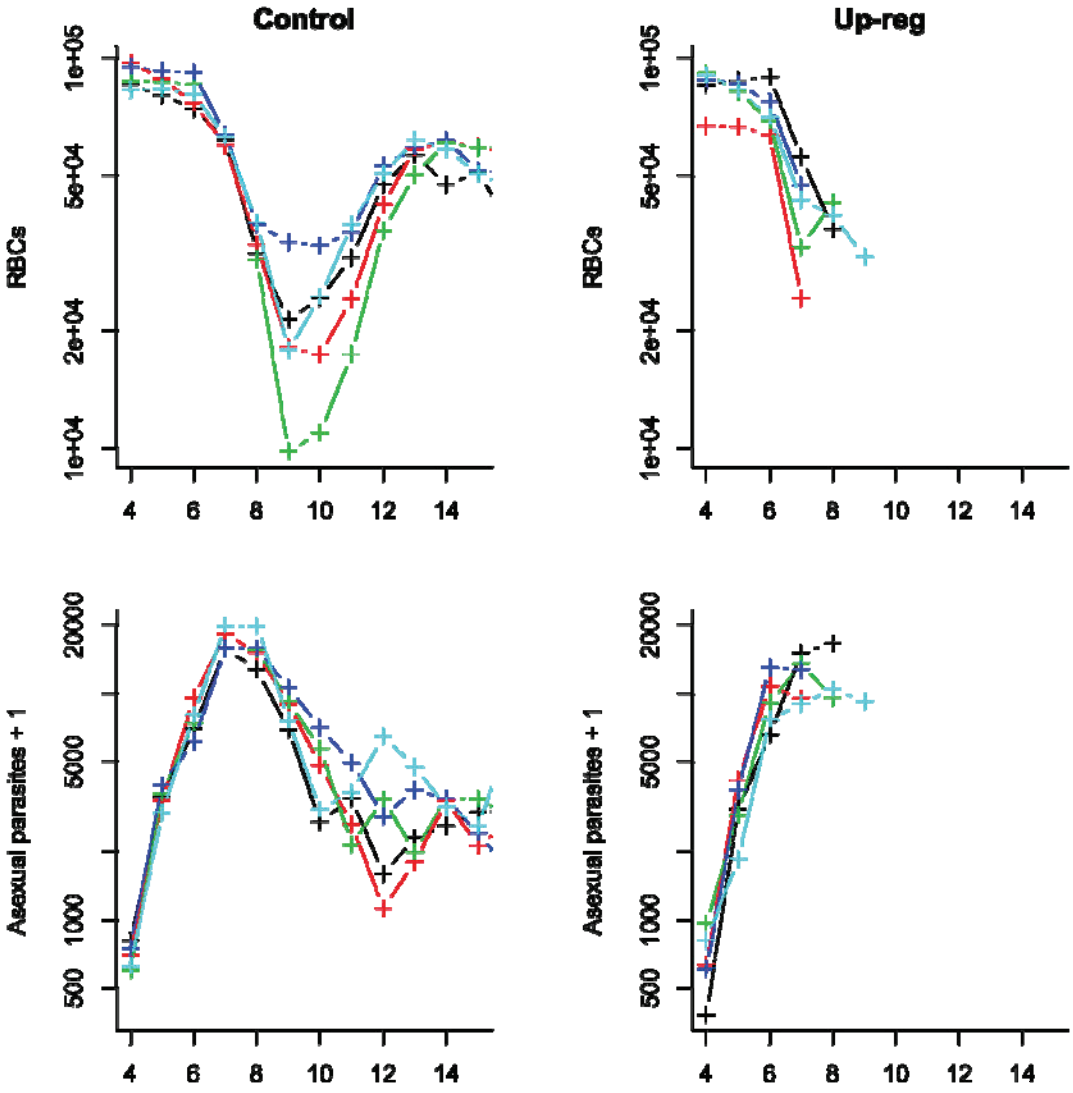

Figure S5: **RBC replenishment.** RBC change (b_t, y axis) against lagged RBCs (x axis, x10^{-2}) for CD-4+ T cell depleted mice infected with the AS strain (1st row); and intact mice infected with the AS strain (2nd row). Plot titles show the RBC lag (days), colours show day in the time-series (legend). The black line is the fitted regression $y \sim c_1 + c_2 x$; the r^2 value is provided bottom-left. For CD-4+ T-cell depleted mice, for the best-fitting model (r^2=0.26 corresponding to τ=1) parameters are c_1=21230 (±3261) and c_2=-0.30 (±0.05); for intact mice (r^2=0.22 for τ=3), parameters are c_1=13805 (±1685) and c_2=-0.19 (±0.02); brackets indicate standard errors.

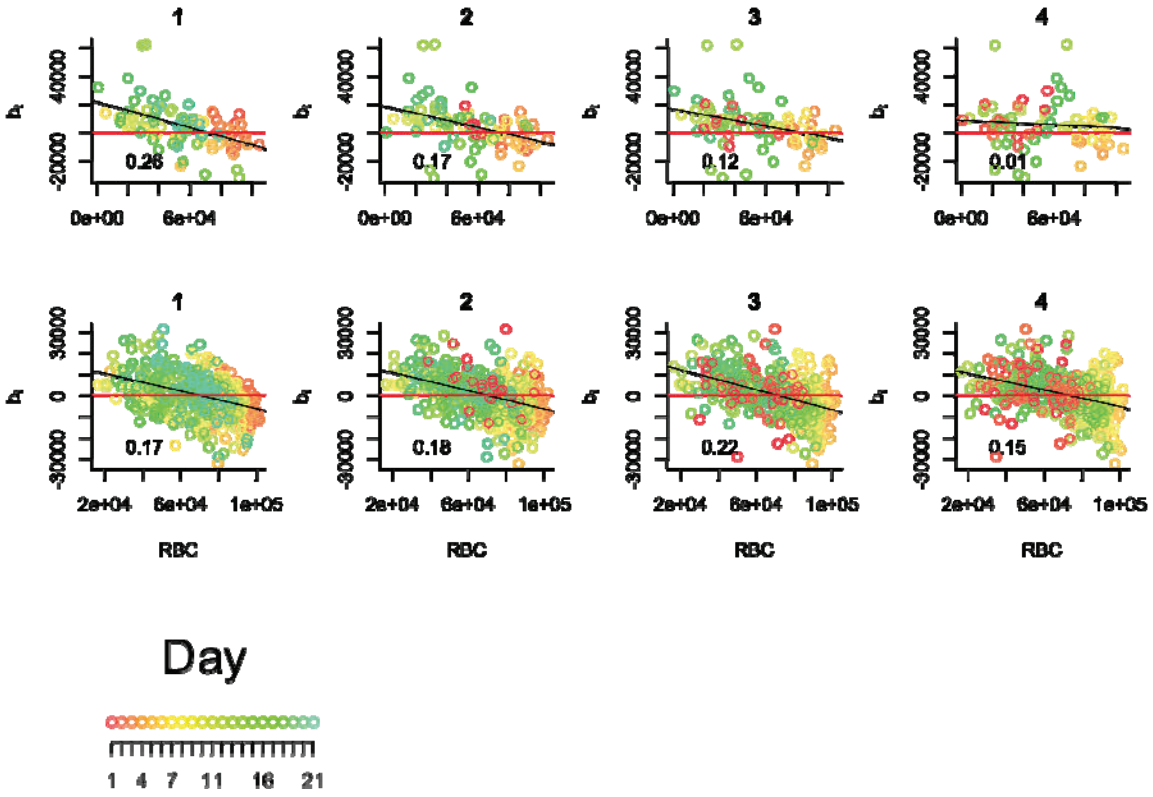

Figure S6: Within-host effective propagation in Dataset 3. The red line is the estimate of effective propagation for 4 intact AS infected mice treated with IL10-R (left) or controls (right); standard errors from the regression shown as dashed vertical lines. Grey polygons show mean approximate proportion of RBCs aged 1-3 days (highest), based on RBC-change reconstruction (right-hand axis). The bottom row shows the corresponding R_E, up to day 20 for untreated mice; treated mice die around day 7. Inoculum sizes are comparable to highest levels in Dataset 2.

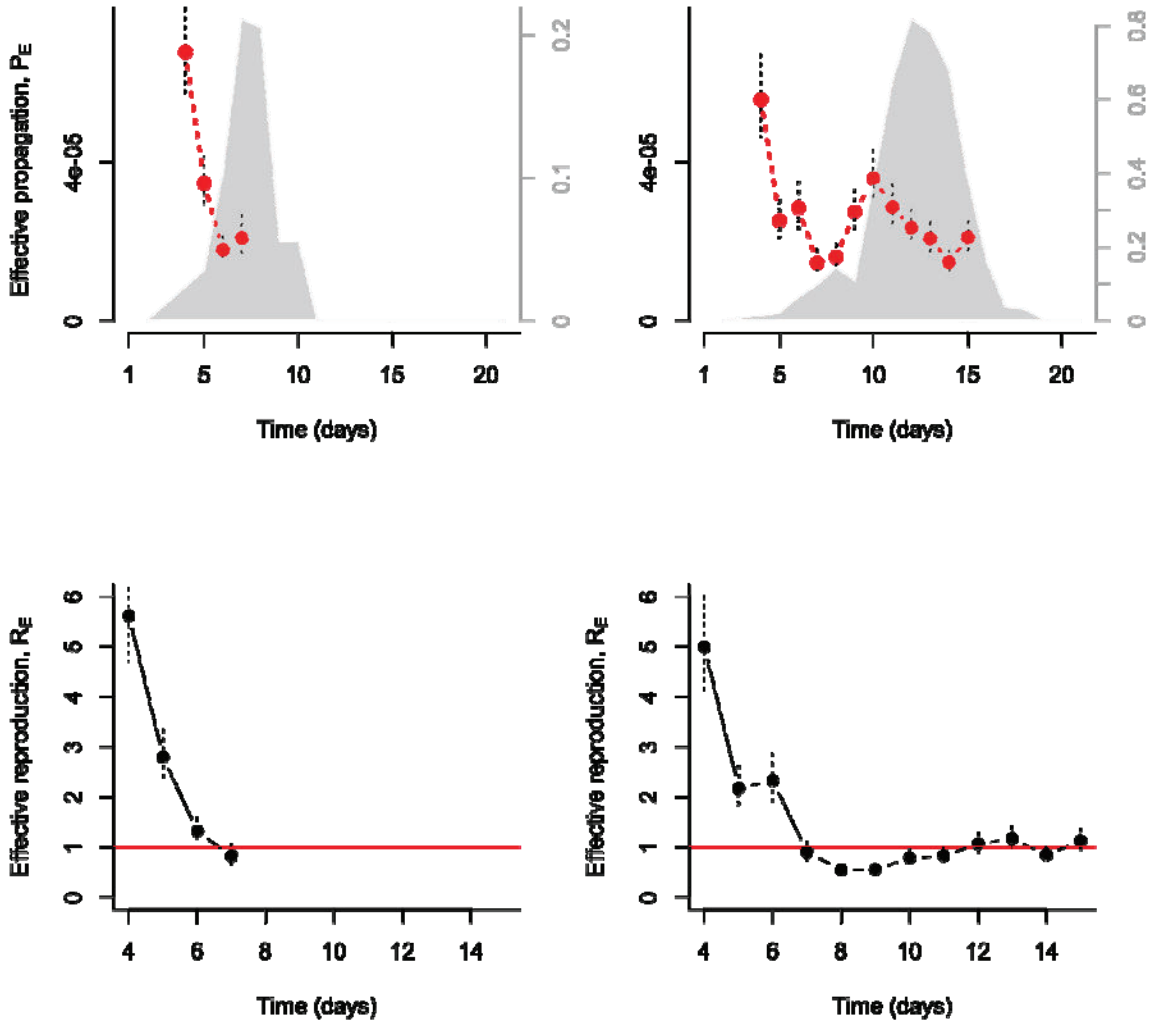

Figure S7: **Necessity of age structure for capturing fluctuations in effective propagation P_E.** For 5 CD4+ T-cell depleted mice infected with the AS strain, a one-step look-ahead model with a single effective propagation parameter (top row, $P_{E,c}$=2.86e-05), and a one-step look-ahead model with a different propagation parameter for each age class of susceptible individuals (bottom row, P_{E1}=1.34e-05 and P_{E2}=4.16e-05) showing the observed and expected time-series of infected RBCs plotted against time x 10^{-2} (a), c), and the associated effective propagation (b),d) predicted in red, observed in black).

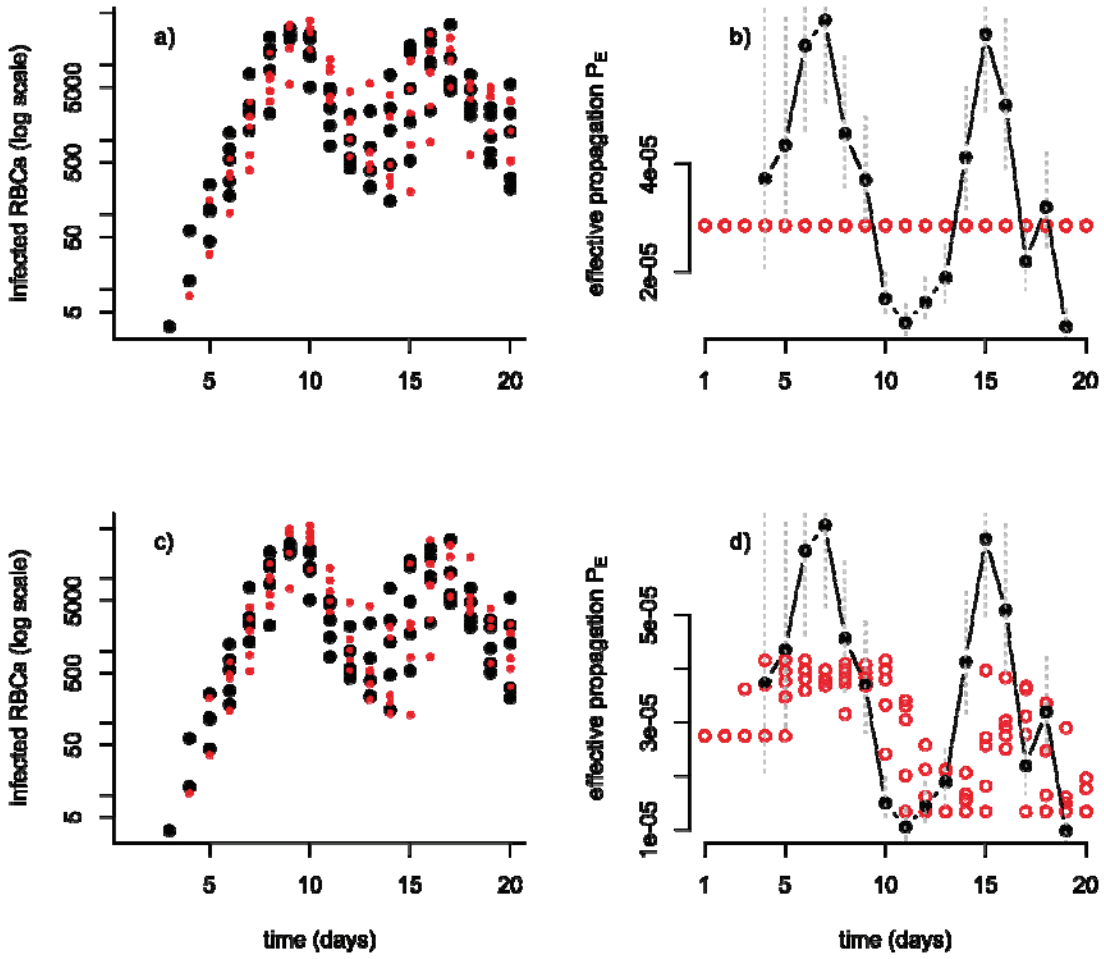

Figure S8: **Effect of induced immunity on effective propagation $P_{E,t}$.** Using parameters estimated from CD4+ T-cell depleted AS strain infected mice, observed and expected time-series of AS strain infected RBCs in intact mice plotted against time (left column, predicted in red, observed in black), and the associated effective propagation (right column, predicted in red, empirically observed in black, with standard errors shown as grey dashed vertical lines); rows show decreasing inoculum sizes (first row is 10^6; fourth row is 10^1). For low starting densities (bottom rows) the effect of innate immunity (first days) is not captured, and the effect of induced immunity (adaptive + innate) is visible later on for all starting densities, leading to over-estimates of incidence (first column).

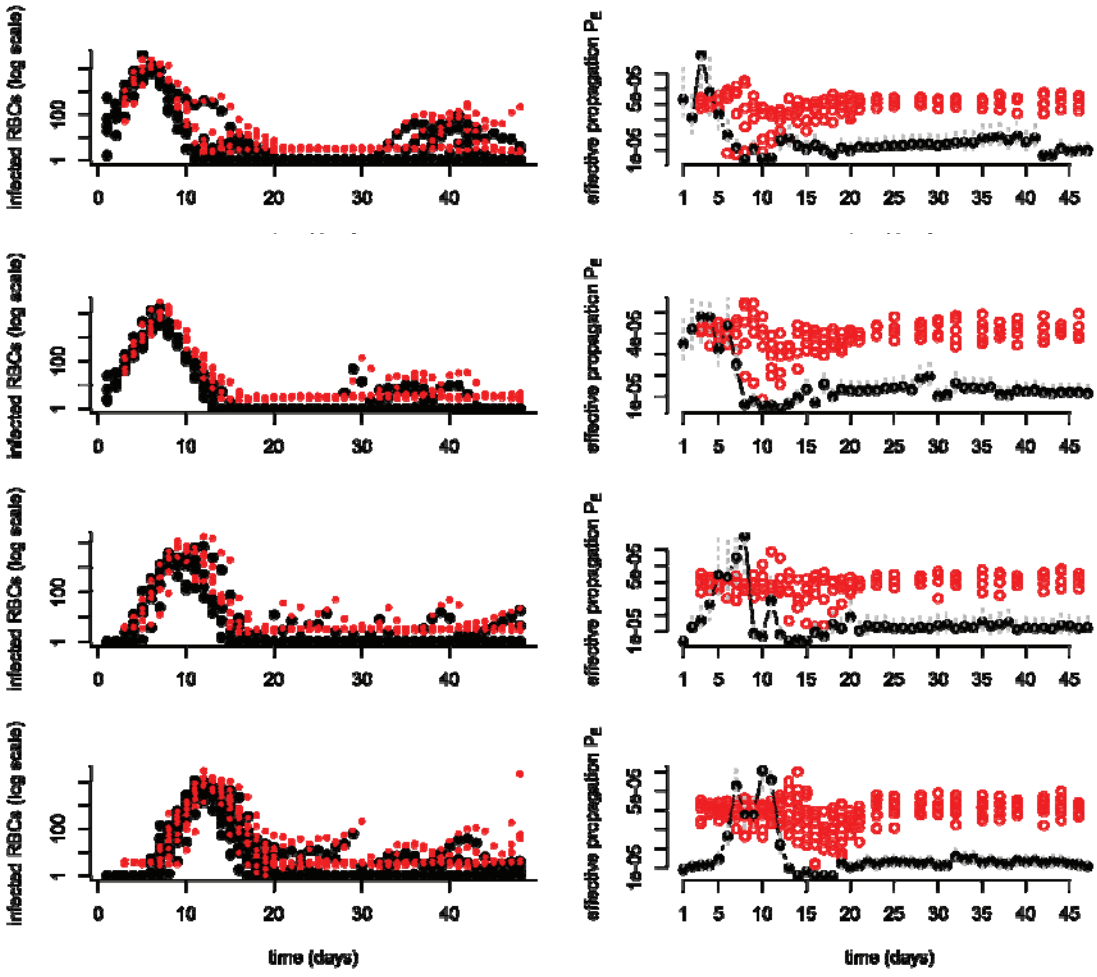

Figure S9: Partitioning effects of age structure, early and induced immunity. The log ratio of the empirically estimated $P_{E,t}$ (Fig. 1, intact mice) and mean $P_{E,t}$ predicted using age-specific parameters estimated from CD4+ T-cell depleted AS infected mice (grey) and with age structure and dose-dependent innate immunity (blue dashed lines) for 4 starting densities (titles). Polygons > 0 indicates the model's predicted $P_{E,t}$ is too high; polygons <0 indicate that it is too low. Squares show where the ratio remains >0 or <0 across the standard error envelope of $P_{E,t}$ (colours as above). Absence of squares indicates accurate prediction, squares above and below indicate robust over- and under-prediction respectively. The thick horizontal line shows where the log ratio is 0 (i.e., perfect prediction), green when error is predominantly due to missing bottom-up effects via age structure (grey polygons); red shows missing dose-dependent early top-down effects (blue dashed polygon closer to zero than grey polygon); black shows missing induced top down effects.

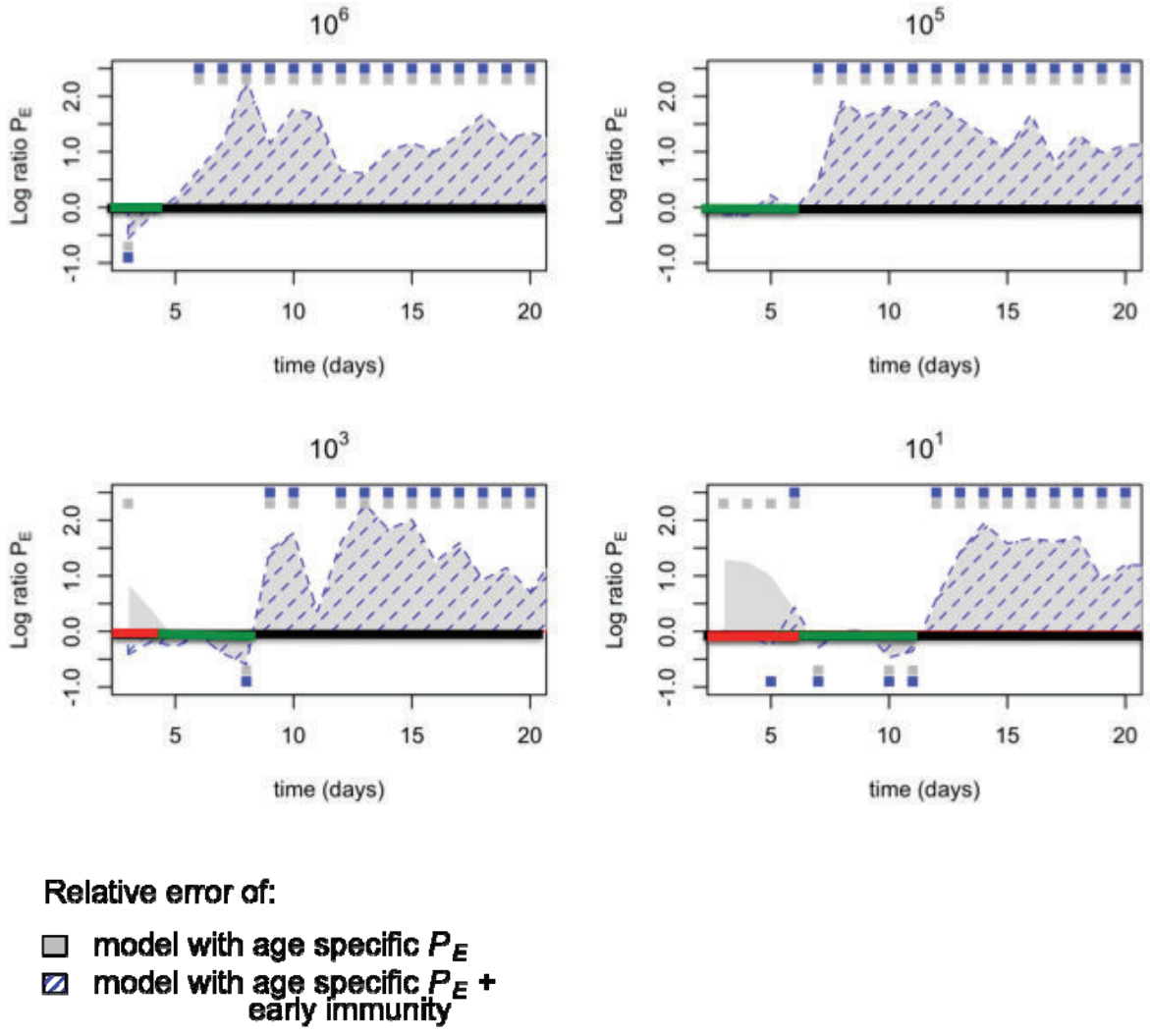

Figure S10: Timing and magnitude of dose dependent effect of innate immunity. Surfaces indicate the log likelihood (left) and sum of squares for empirical and predicted values of $P_{E,t}$ (right) for combinations of t_c (x axis) and the innate immunity parameter a (y axis, log scale) for intact mice with starting densities of 10^1 and 10^3; red indicates low values, and white high values. Note that the duration of up-regulated innate immunity may be underestimated, as its statistical signal is swamped on the onset of induced immunity, indicated by the long right tail in the profile likelihood surface. Nevertheless a global maximum, corresponding to t_c=5 and a=1.0027 is identified. These parameter values capture the pattern of effective propagation across dose observed in the data shown in Figure 2 (far right). The dashed line shows the predicted curve, with maximum effective propagation set $7e^{-5}$; and effective propagation across parasite abundance is calculate as the product of this maximum and the probability of survival defined by $\log(\mathbf{M}t)/(a+\log(\mathbf{M}t))$, see Methods.

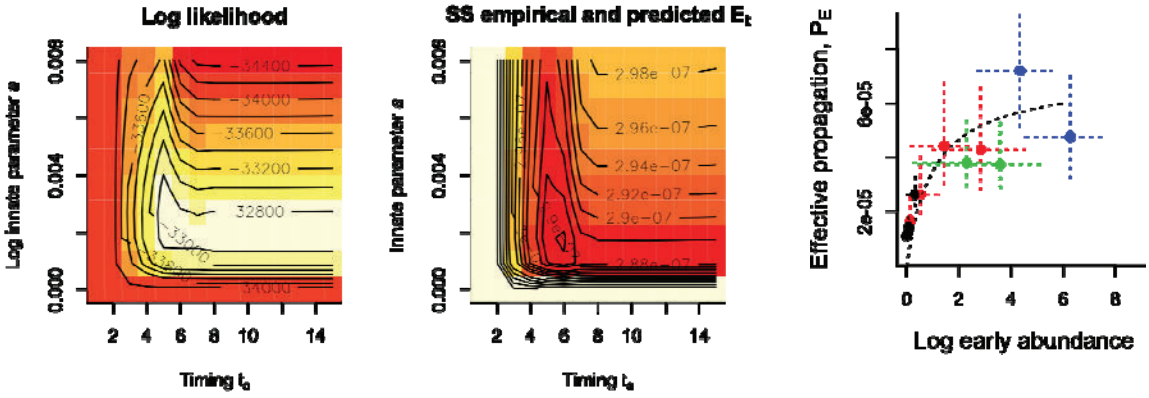

Figure S11: **Time course of immune activity.** Observed (points) and predicted (coloured lines) infected and uninfected RBCs for decreasing inoculum sizes (row 1 corresponds to 10^6; and row 4 to 10^1). Predictions are from 6 fitted parameters (see text) and an estimated fraction of infected cells killed by immunity (3rd col, p_t) for each individual. The grey polygon is the envelope across individuals, the black line is the median, rows range from the highest (top) to the lowest (bottom) starting density.

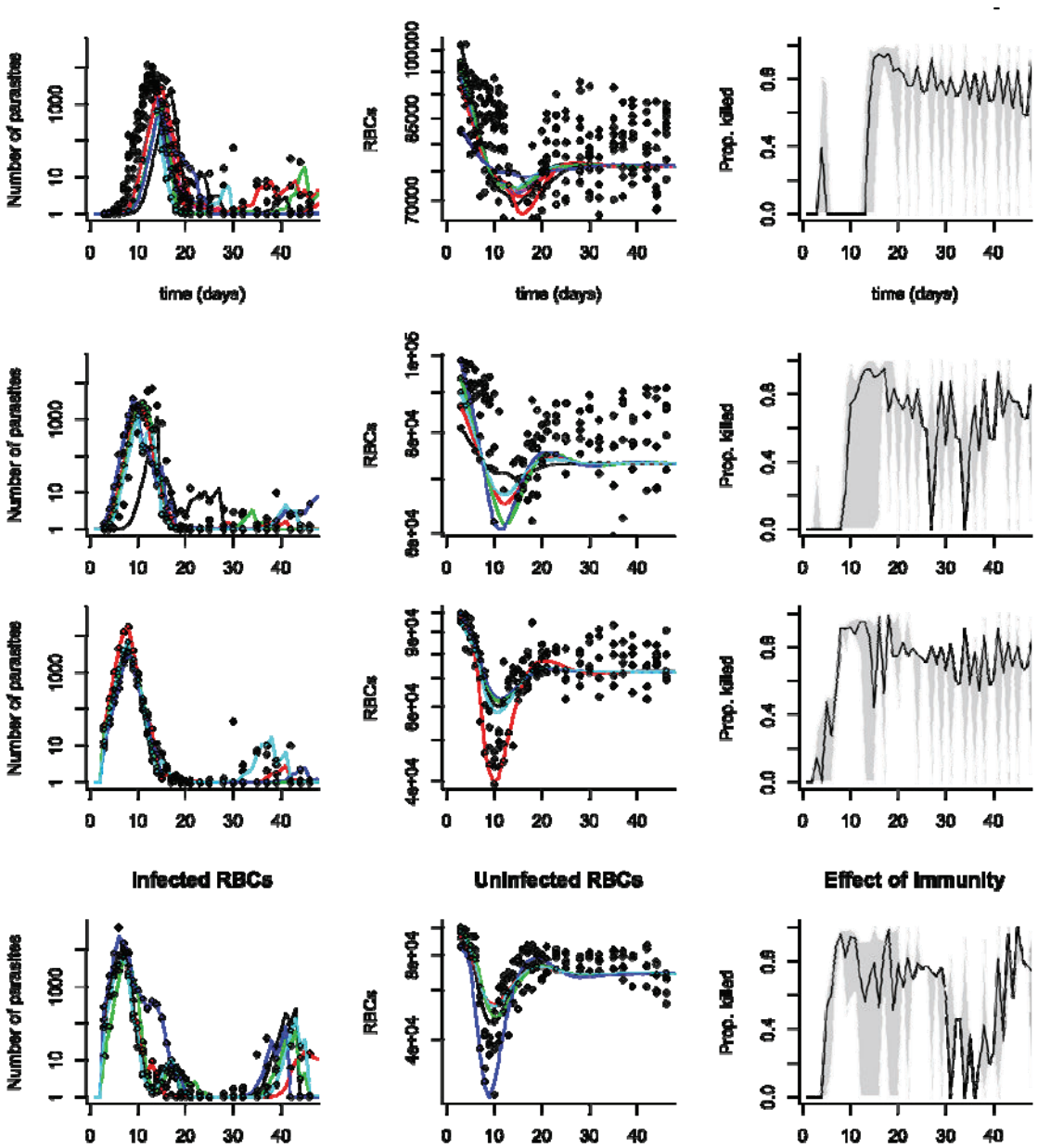

Figure S12: **Partitioning the effect of key components on effective reproduction** The pattern of R_E for the different inoculum sizes (10^1 in black, 10^3 in red, 10^5 in green, 10^6 in blue) where a) only a single effective propagation parameter is fitted, showing the effect of susceptible depletion alone; b) where only age parameters are fit, showing effects of susceptible depletion and age; and c) where only a mean and early non-induced immunity parameters are fit (a, t_c), showing the effect of dose dependent reduction of early R_E; and d) the observed pattern, showing the impact of induced immunity later in the time-series. Table 1 takes ratios of R_E at points indicated by dashed vertical lines, where the colours of the vertical lines show chosen inoculum size (taken for the largest effect).

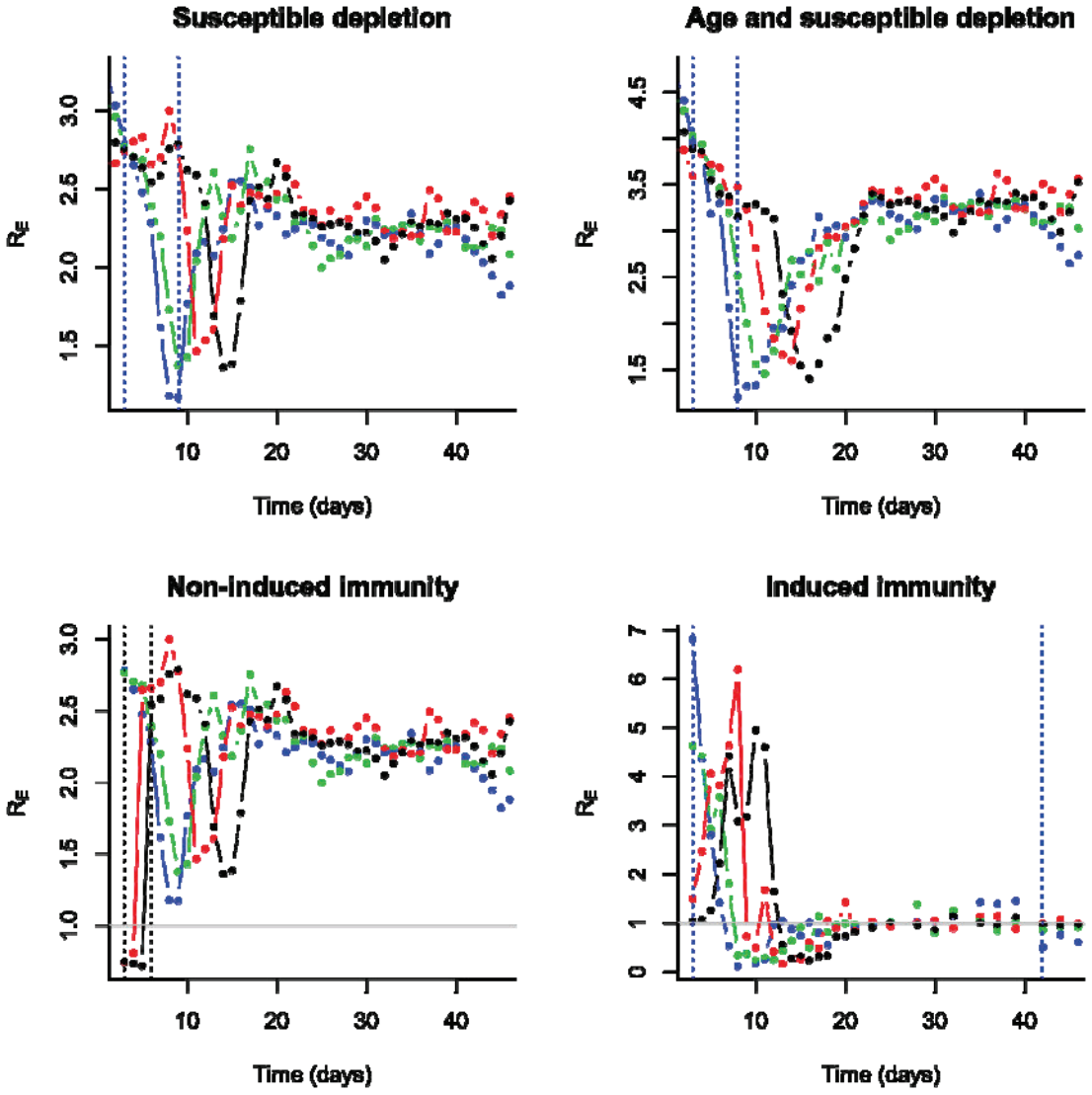

Table S1: Relative importance of malaria control mechanisms The degree to which each mechanism reduces the effective reproduction number R_E, obtained by fitting a range of different models and comparing resulting values of predicted R_E at relevant time-points (Fig. S12), dividing the larger observed value by the smaller value. Late-acting, adaptive immunity is the most powerful agent of control (13-fold reduction), followed by early innate immunity in mice with small inoculum sizes (note that the proximity of the earliest R_E for high inoculum sizes to the upper limit of burst size estimates for this clone suggest that no other control mechanisms are operating during this early phase, see text) and the effects of red blood cell age structure and depletion (both ~3 fold reduction), and then finally, the effects of susceptible RBC depletion alone.

Control	Strength	Description
RBC depletion	2.37	Change in R_E predicted between start of infection and when anemia is greatest. Only a single effective propagation parameter is fitted (P_E).
RBC depletion + RBC age structure	3.14	Change in R_E predicted between start of infection and when anemia is greatest and most dominated by reticulocytes. Only age structure parameters are fitted (P_{E1}, P_{E2}).
Innate immunity low dose	3.40	Change in R_E predicted between early time-steps with only early dose-dependent immunity parameters (a, t_c). Only an average effective propagation parameter fitted.
Innate immunity high dose	0.92	As above.
Adaptive immunity	13.51	Change in R_E predicted with directly estimated P_Es (see Fig. 1) between start of infection and during steady state controlled predominantly by adaptive immunity (since age structure and anemia are near-stable at later time-points).