THE INFLUENCE OF INFECTIOUS DISEASES ON POPULATION GENETICS

ZHILAN FENG

Department of Mathematics
Purdue University, West Lafayette, IN 47907

CARLOS CASTILLO-CHAVEZ

Department of Mathematics and Statistics
Arizona State University, Tempe, AZ 85287

ABSTRACT. Malaria is the vector-transmitted disease that causes the highest morbidity and mortality in humans. Motivated by the known influence of sickle-cell anemia on the morbidity and mortality of malaria-infected humans, we study the effect of malaria on the genetic composition of a host (human) population where sickle-cell anemia is prevalent and malaria is endemic. The host subpopulations are therefore classified according to three genotypes, AA, AS, and SS. It is known that AA malaria-infected individuals experience higher malaria-induced mortality than AS or SS individuals. However, individuals carrying the S gene are known to experience a higher mortality rate in a malaria-free environment than those who lack such a gene. The tradeoffs between increased fitness for some types in the presence of disease (a population level process) and reduced fitness in a disease-free environment are explored in this manuscript. We start from the published results of an earlier model and proceed to remove some model restrictions in order to better understand the impact on the natural hosts' genetics in an environment where malaria is endemic.

1. **Introduction.** Efforts to understand a host's evolutionary dynamics (often slow) in the context of disease dynamics in "chemically" treated environments (e.g., chemotherapy) or the influence of diseases on the host's genetic variability or both have been carried out recently (see, for example, Andreasen [1], Beck [2], Castillo-Chavez and Feng [3], Feng et al. [5, 6], Galvani et al. [8, 9], Hsu-Schmitz [13], Kribs et al., May and Anderson [10], McKenzie [11, 12]) in the context of human diseases such as HIV, tuberculosis, malaria and others. Feng et al. [5] studied the influence of malaria dynamics (assumed to be fast) on hosts' fitness in their efforts to assess the effect of vector-borne diseases on the genetic composition of a host population. A typical setting was adopted by assuming that the human population is characterized by a two-allele single loci system. Here, we remove some of their model simplifications and show that the qualitative dynamics do not change. However, we briefly illustrate the potentially significant effect that the incorporation of additional host characteristics (age) may have on disease dynamics and on

---

the genetic structure of the host population. In other words, incorporating the life-history dynamics of the host can have a dramatic effect on the qualitative dynamics of the disease and the gene distribution in the host population. Here, we have only scratched the surface.

It was assumed (see [5]) that the human population could be properly characterized by a two-allele single loci system in the study of a malaria-human system. It is assumed that only three genotypes are possible: AA, AS, and SS, where the letters A and S denote the two alleles. Under a simplifying assumption that the SS individuals do not survive, the following model was introduced and analyzed in [5]:

$$
\begin{aligned}
\dot{u}_i &= \bar{P}_i b(N) N - m_i u_i - \beta_{hi} z u_i + \gamma_i v_i, \\
\dot{v}_i &= \beta_{hi} z u_i - (m_i + \gamma_i + \alpha_i) v_i, \qquad i = 1, 2, \\
\dot{z} &= (1 - z) \sum_{i=1}^{2} \beta_{vi} \frac{v_i}{N} - \delta z.
\end{aligned}
\tag{1}
$$

Here, $u_i(t)$ denotes the host population of uninfected i-individuals ($i = 1, 2$ with $AA = 1$, $AS = 2$), $v_i(t)$ denotes the host population of infected i-individuals ($i = 1, 2$), and $N(t) = \sum_{i=1}^{2}(u_i(t) + v_i(t))$ denotes the total host population at time t. $\bar{P}_i(t)$ ($i = 1, 2$) denotes the fraction of each genotype born into the host population and it is assumed that these proportions satisfy the following relationships (Hardy-Weinberger proportions):

$$
\bar{P}_1(t) = \bar{p}^2(t), \qquad \bar{P}_2(t) = 2\bar{p}(t)\bar{q}(t),
\tag{2}
$$

where $\bar{q}(t) = \bar{r}(t)/2$ is the frequency of the S-gene, $\bar{p}(t) = 1 - \bar{q}(t)$ is the frequency of the A-gene, and

$$
\bar{r}(t) = \frac{u_2(t) + v_2(t)}{N(t)}
$$

denotes the fraction of AS individuals.

Following the approach of the classical Ross McDonald model for the spread of malaria, we use the fraction of infected mosquitoes, $z(t)$, instead of the total number of infected mosquitoes in our model. The rationale behind this assumption comes from the fact that what typically determines "observed" vector densities is the number of vector breeding sites where there is fierce competition and an oversupply of eggs. Susceptible human hosts of type i become infected at the rate $\beta_{hi} z u_i$ ($i = 1, 2$), while the fraction of susceptible mosquitoes becomes infected (from biting an infected human host of type i) at the rate $(1-z)\beta_{vi} v_i / N$. Here, β_{hi} and β_{vi} ($i = 1, 2$) denote the transmission coefficients from humans to mosquitoes and mosquitoes to humans, respectively. The malaria disease-induced death rate is α_i while the average time before a victim of malaria recovers is denoted by $1/\gamma_i$. The documented resistance of AS individuals to malaria infection is modeled by a reduction in susceptibility to disease invasion. Hence, it is assumed that

$$
\beta_{h1} > \beta_{h2}, \qquad \gamma_1 \leq \gamma_2.
\tag{3}
$$

The hosts' demography is important, so $b(N)$ denotes the human per capita birth rate, possibly density dependent, and $m_i = m + \nu_i$, where m is the assumed constant per capita natural mortality of humans and ν_i ($i = 1, 2$) the excess per capita death

TABLE 1. Definition of variables and parameters

Name	Description	Notes:
$i = 1$	AA individuals	
$i = 2$	AS individuals	
$i = 3$	SS individuals	
u_i	number of uninfected humans of genotype i	
v_i	number of infected humans of genotype i	
N	total population size	
x_i	fraction of uninfected humans of genotype i	u_i/N
y_i	fraction of infected humans of genotype i	v_i/N
z	fraction of inected mosquitoes	
r	frequency of AS individuals	$x_2 + y_2$
s	frequency of SS individuals	$x_3 + y_3$
w	sum of frequencies of AS and SS individuals	$r + s$
q	frequency of S-gene	$r/2 + s$
p	frequency of A-gene	$1 - q$
β_{hi}	malaria infection rate of humans of type i	$\beta_{h1} > \beta_{h2} = \beta_{h3}$
γ_i	rate of recovery from malaria of humans of type i	$\gamma_1 \leq \gamma_2 = \gamma_3$
β_{vi}	infection rate of mosquitoes from biting type i humans	
δ	mosquito death rate	
m	natural human death rate	
α_i	malaria-induced death rate of humans of type i	$\alpha_1 \geq \alpha_2 = \alpha_3$
ν_i	S-gene related death rate of humans of type i	$\nu_1 = 0 < \nu_2 \leq \nu_3$
$b(N)$	per capita birth rate of humans	
P_1	fraction of total births of genotype AA	$(1 - r/2 + s)^2$
P_2	fraction of total births of genotype AS	$(1 - r/2 + s)(r + 2s)$
P_3	fraction of total births of genotype SS	$(r/2 + s)^2$

rate ascribed to S-gene carriers. Here, we take $\nu_1 = 0$, $\nu_2 > 0$. Definitions of all variables and parameters can be found in Table 1.

Was the omission of SS individuals justified in model (1)? Here, we remove such a restriction, making our work relevant to vector-transmitted infectious diseases for which all three genotypes of individuals need to be considered. Although there are obvious quantitative differences, we show that the qualitative results are equivalent to those generated by the simpler model. In other words, on the fast time scale of malarial dynamics, the disease level reaches an equilibrium and, consequently, the host's gene frequency distribution is influenced only by the equilibrium level of malaria. However, the generality of this conclusion comes into question when the host's population structure is incorporated. The rest of the manuscript is organized as follows: Section 2 describes the extended model; Sections 3 and 4 justify the simplifications that result from the identification of two processes whose dynamics take place at highly distinct temporal scales; Section 5 looks at the fitness of the S-gene, and Section 6 provides a critical discussion of the results while suggesting that a key missing component, the host's population structure, may indeed make the robustness of the above results questionable.

2. **The generalized model.** The model (1) does not include the SS individuals. To include these individuals we need to introduce new variables, u_3 and v_3, to denote the host populations of susceptible and infected SS individuals, respectively. Then

(2) will be changed to

$$P_1(t) = p^2(t), \qquad P_2(t) = 2p(t)q(t), \qquad P_3(t) = q^2(t),$$

where

$$q(t) = \frac{r(t)}{2} + s(t), \qquad p(t) = 1 - q(t)$$

and

$$r(t) = \bar{r}(t) = \frac{u_2(t) + v_2(t)}{N(t)}, \qquad s(t) = \frac{u_3(t) + v_3(t)}{N(t)}.$$

In this case the system (1) can be extended to become the following seven-dimensional system:

$$\begin{aligned}
\dot{u}_i &= P_i b(N) N - m_i u_i - \beta_{hi} z u_i + \gamma_i v_i, \\
\dot{v}_i &= \beta_{hi} z u_i - (m_i + \gamma_i + \alpha_i) v_i, \qquad i = 1, 2, 3, \\
\dot{z} &= (1 - z) \sum_{i=1}^{3} \beta_{vi} \frac{v_i}{N} - \delta z.
\end{aligned} \qquad (4)$$

All parameters have the same meaning as in (1), with some additional assumptions concerning the SS individuals; i.e., $\nu_2 \le \nu_3$, and

$$\beta_{h1} > \beta_{h2} = \beta_{h3}, \qquad \gamma_1 \le \gamma_2 = \gamma_3. \qquad (5)$$

As in [5], we take advantage of the fact that the time scales of evolutionary processes compared to those of the disease are very different, which allows us to use a singular perturbation approach to separate the fast and slow dynamics of the system. For this purpose we introduce new variables, $x_i = u_i/N$ and $y_i = v_i/N$. These are the fractions of corresponding subpopulations of human hosts. Using the new notation, we can rewrite the frequencies of the AS and SS individuals as $r = x_2 + y_2$ and $s = x_3 + y_3$, respectively. We note for clarification that $x_1 + y_1 + x_2 + y_2 + x_3 + y_3 = 1$ and $x_1 + y_1 = 1 - r - s$. We can replace the set of variables in the system (4) with the new set of variables: y_1, y_2, y_3, z, r, s, and N, which obviously describe important epidemiological, demographic, and population genetic quantities. System (4) is equivalent to the following system:

$$\begin{aligned}
\dot{y}_1 &= \beta_{h1} z (1 - r - s - y_1) - (m_1 + \gamma_1 + \alpha_1) y_1 - y_1 \dot{N}/N, \\
\dot{y}_2 &= \beta_{h2} z (r - y_2) - (m_2 + \gamma_2 + \alpha_2) y_2 - y_2 \dot{N}/N, \\
\dot{y}_3 &= \beta_{h3} z (s - y_3) - (m_3 + \gamma_3 + \alpha_3) y_3 - y_3 \dot{N}/N, \\
\dot{z} &= (1 - z) \sum_{i=1}^{3} \beta_{vi} y_i - \delta z, \\
\dot{r} &= P_2 b(N) - \alpha_2 y_2 - m_2 r - r \dot{N}/N, \\
\dot{s} &= P_3 b(N) - \alpha_3 y_3 - m_3 s - s \dot{N}/N, \\
\dot{N} &= N \left(b(N) - m_1 (1 - r - s) - m_2 r - m_3 s - \sum_{i=1}^{3} \alpha_i y_i \right).
\end{aligned} \qquad (6)$$

Following the approach in [5] our mathematical analysis is for the specific density dependent per capita birth function, $b(N) = b(1 - N/K)$, where b is a constant (the maximum birth rate when the host population size is small) and K is approximately

the density dependent reduction in birth rate. We remark that the full model in [5] is a five-dimensional system, and the slow system consists of two equations. Here, the full system (4) is seven-dimensional, and we will show next that the slow system consists of three equations.

3. **Fast dynamics of epidemics.** The relevant parameters vary across many orders of magnitude. For example, the demographic parameters (b and m_i) and the genetic parameters (α_i) are on the order of 1/decades, and the malaria disease parameters (β_{hi}, γ_i, β_{vi}, and δ) are on the order of 1/days. Hence, although the malaria disease dynamics and the changes in genetic composition are two coupled processes, the former occurs on a much faster time scale than the latter. Let $m_i = \epsilon\tilde{m}_i, \alpha_i = \epsilon\tilde{\alpha}_i$, and $b = \epsilon\tilde{b}$ with $\epsilon > 0$ being small. We can use this fact to simplify the mathematical analysis of the full model with the use of singular perturbation techniques, which allows us to separate the time scales of the different processes. By letting $\epsilon = 0$ we obtain the following system for the fast dynamics:

$$
\begin{aligned}
\dot{y}_1 &= \beta_{h1}z(1 - r - s - y_1) - \gamma_1 y_1, \\
\dot{y}_2 &= \beta_{h2}z(r - y_2) - \gamma_2 y_2, \\
\dot{y}_3 &= \beta_{h3}z(s - y_3) - \gamma_3 y_3, \\
\dot{z} &= (1 - z)\sum_{i=1}^{3}\beta_{vi}y_i - \delta z,
\end{aligned}
\tag{7}
$$

where r and s are considered parameters which represent genotype frequencies of AS and SS individuals. Thus, the fast system (7) describes the epidemics of malaria for a given distribution of genotypes determined by r and s. Let $w = r + s$. Then, $q \le w \le 2q$. Note that $q(t) \to 0$ if and only if $w(t) \to 0$ as $t \to \infty$. Therefore, the variable w provides a good measure of the S-gene frequency when studying persistence properties.

The basic reproductive number for the malaria epidemics can be calculated using the next generation matrix (see [4]). We keep in mind that, on the fast time scale, r and s are treated as parameters. Let $E_0 = (0, 0, 0, 0)$ be the disease-free equilibrium (DFE) of the fast system (7). Then from the Jacobian matrix at E_0 we obtain the next generation matrix:

$$
G = \begin{pmatrix}
0 & 0 & 0 & \frac{\beta_{h1}(1-r-s)}{\delta} \\
0 & 0 & 0 & \frac{\beta_{h2}r}{\delta} \\
0 & 0 & 0 & \frac{\beta_{h3}s}{\delta} \\
\frac{\beta_{v1}}{\gamma_1} & \frac{\beta_{v2}}{\gamma_2} & \frac{\beta_{v3}}{\gamma_3} & 0
\end{pmatrix},
$$

whose leading eigenvalue is

$$
\sqrt{\frac{\beta_{v1}\beta_{h1}(1 - r - s)}{\delta\gamma_1} + \frac{\beta_{v2}\beta_{h2}r}{\delta\gamma_2} + \frac{\beta_{v3}\beta_{h3}s}{\delta\gamma_3}}.
$$

This quantity gives the basic reproductive number. However, to simplify the notation we define our $\mathcal{R}_0$ to be this number squared:

$$
\begin{aligned}
\mathcal{R}_0 &= \frac{\beta_{v1}\beta_{h1}(1 - r - s)}{\delta\gamma_1} + \frac{\beta_{v2}\beta_{h2}r}{\delta\gamma_2} + \frac{\beta_{v3}\beta_{h3}s}{\delta\gamma_3} \\
&=: \mathcal{R}_1(1 - r - s) + \mathcal{R}_2 r + \mathcal{R}_3 s,
\end{aligned}
$$

where

$$\mathcal{R}_i = \frac{\beta_{vi}\beta_{hi}}{\delta\gamma_i}, \qquad i = 1, 2, 3,$$

which involves parameters associated with malaria transmission between mosquitoes and humans of genotype i. In fact, $\mathcal{R}_i$ (or $\sqrt{\mathcal{R}_i}$) is the basic reproductive number when the host population consists of entirely humans of genotype i. When $\mathcal{R}_0 < 1$, all eigenvalues of G are bounded by 1, and at least one eigenvalue of G will exceed 1 when $\mathcal{R}_0 > 1$. This implies that the Jacobian at E_0 has all eigenvalues with negative real parts if $\mathcal{R}_0 < 1$ and has at least one eigenvalue with positive real part if if $\mathcal{R}_0 > 1$. We have proved the following result.

Theorem 1. *The DFE E_0 of the fast system (7) is l.a.s. if $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$.*

The following result also holds as in most endemic models:

Theorem 2. *The fast system (7) has no endemic equilibrium when $\mathcal{R}_0 < 1$. When $\mathcal{R}_0 > 1$, a unique endemic equilibrium exists and is l.a.s.*

Proof. To simplify notations we introduce the variable $w = r + s$, which represents the frequency of individuals carrying the S-gene. Using (5) we have $\mathcal{R}_3 = \mathcal{R}_2$. Then we can rewrite $\mathcal{R}_0$ as

$$\mathcal{R}_0 = \mathcal{R}_1(1 - w) + \mathcal{R}_2 w.$$

Let $E^* = (y_1^*, y_2^*, y_3^*, z^*)$ be a nontrivial equilibrium of the fast system (7); i.e., all components of E^* are positive. Then

$$y_1^* = \frac{T_{h1}z^*(1 - r - s)}{1 + T_{h1}z^*}, \quad y_2^* = \frac{T_{h2}z^*r}{1 + T_{h2}z^*}, \quad y_3^* = \frac{T_{h3}z^*s}{1 + T_{h3}z^*}, \tag{8}$$

where

$$T_{hi} = \frac{\beta_{hi}}{\gamma_i}, \qquad i = 1, 2, 3,$$

and z^* is a positive solution of a quadratic equation

$$k_0 z^2 + k_1 z + k_2 = 0, \tag{9}$$

with

$$
\begin{aligned}
k_0 &= T_{h1}T_{h2} + \mathcal{R}_1 T_{h2}(1 - w) + \mathcal{R}_2 T_{h1}w, \\
k_1 &= T_{h1} + T_{h2} + \mathcal{R}_1(1 - T_{h2})(1 - w) + \mathcal{R}_2(1 - T_{h1})w, \\
k_2 &= 1 - \mathcal{R}_1(1 - w) - \mathcal{R}_2 w.
\end{aligned}
\tag{10}
$$

Here we have used the fact that $T_{h3} = T_{h2}$ which is due to (5).

Let $\mathcal{R}_0 = \mathcal{R}_1(1 - w) + \mathcal{R}_2 w < 1$. Then $k_2 = 1 - \mathcal{R}_0 > 0$. To show that the equation (9) has no positive solution, it suffices to show that $k_1 > 0$. Let $T_0 = \max\{T_{h1}, T_{h2}\}$. Then $T_{h1} + T_{h2} - T_0\mathcal{R}_0 \geq 0$, and

$$
\begin{aligned}
k_1 &= T_{h1} + T_{h2} + \mathcal{R}_0 - \mathcal{R}_1 T_{h2}(1 - w) - \mathcal{R}_2 T_{h1}w \\
&\geq T_{h1} + T_{h2} + \mathcal{R}_0 - T_0\mathcal{R}_0 \\
&> 0.
\end{aligned}
$$

Since k_0 is always positive, it follows that both solutions of (9) are negative. Therefore, E^* is not biologically feasible.

Let $\mathcal{R}_0 > 1$. Then $k_2 = 1 - \mathcal{R}_0 < 0$. Since $k_0 > 0$, the equation (9) has a unique positive solution z^*. Let $h(z)$ denote the function of z given by the left-hand side of (9). Notice that $h(0) = k_2 < 0$, $h(1) = 1 + T_{h1}T_{h2} + T_{h1} + T_{h2} > 0$, and $h(z^*) = 0$. Hence, $0 < z^* < 1$. From (8) we also have that $0 < y_i^* < 1$, $i = 1, 2, 3$. It follows that $E^* = (y_1^*, y_2^*, y_3^*, z^*)$ exists and is unique.

The stability of E^* is determined by the eigenvalues of the following matrix H:

$$H = \begin{pmatrix} -(\beta_{h1}z^* + \gamma_1) & 0 & 0 & \beta_{h1}(1 - r - s - y_1^*) \\ 0 & -(\beta_{h2}z^* + \gamma_2) & 0 & \beta_{h2}(r - y_2^*) \\ 0 & 0 & -(\beta_{h3}z^* + \gamma_3) & \beta_{h3}(s - y_3^*) \\ \beta_{v1}(1 - z^*) & \beta_{v2}(1 - z^*) & \beta_{v3}(1 - z^*) & -(\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta) \end{pmatrix}.$$

The matrix H can be written in the form $H = M - D$, where

$$M = \begin{pmatrix} 0 & 0 & 0 & \beta_{h1}(1 - r - s - y_1^*) \\ 0 & 0 & 0 & \beta_{h2}(r - y_2^*) \\ 0 & 0 & 0 & \beta_{h3}(s - y_3^*) \\ \beta_{v1}(1 - z^*) & \beta_{v2}(1 - z^*) & \beta_{v3}(1 - z^*) & 0 \end{pmatrix},$$

and

$$D = \begin{pmatrix} \beta_{h1}z^* + \gamma_1 & 0 & 0 & 0 \\ 0 & \beta_{h2}z^* + \gamma_2 & 0 & 0 \\ 0 & 0 & \beta_{h3}z^* + \gamma_3 & 0 \\ 0 & 0 & 0 & \sum_{i=1}^{3}\beta_{vi}y_i^* + \delta \end{pmatrix}.$$

Notice that $M \geq 0$; i.e., all elements of M are nonnegative (recall that $1 - r - s - y_1 = x_1 > 0$, $r - y_2 = x_2 > 0$, $s - y_3 = x_3 > 0$) and D is a diagonal matrix with positive diagonal elements. It is known that all eigenvalues of H have negative real parts if and only if the dominant eigenvalue of the matrix MD^{-1} is less than one. MD^{-1} has a double zero eigenvalues and two other eigenvalues given by

$$\lambda_\pm = \pm\sqrt{A_1\left(\frac{\beta_{v1}(1 - z^*)}{\beta_{h1}z^* + \gamma_1}\right) + A_2\left(\frac{\beta_{v2}(1 - z^*)}{\beta_{h2}z^* + \gamma_2}\right) + A_3\left(\frac{\beta_{v3}(1 - z^*)}{\beta_{h3}z^* + \gamma_3}\right)},$$

where

$$A_1 = \frac{\beta_{h1}(1 - r - s - y_1^*)}{\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta}, \quad A_2 = \frac{\beta_{h2}(r - y_2^*)}{\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta}, \quad A_3 = \frac{\beta_{h3}(s - y_3^*)}{\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta}.$$

Obviously, $\lambda_- < 1$. Next we show that $\lambda_+ < 1$. Using the following equalities (which are obtained by setting the right-hand side of the y_1 and z equations in (7)

equal to zero),

$$z^* = \frac{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta} = \frac{\gamma_1 y_1^*}{\beta_{h1}(1 - r - s - y_1^*)},$$

or

$$\frac{\beta_{h1}(1 - r - s - y_1^*)}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^* + \delta} = \frac{\gamma_1 y_1^*}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*},$$

and noticing that $\beta_{hi}z^* + \gamma_i > \gamma_i$, $i = 1, 2, 3$, we obtain

$$A_1\left(\frac{\beta_{v1}(1 - z^*)}{\beta_{h1}z^* + \gamma_1}\right) = \frac{\gamma_1 y_1^*}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*}\left(\frac{\beta_{v1}(1 - z^*)}{\beta_{h1}z^* + \gamma_1}\right) < \frac{\beta_{v1}y_1^*(1 - z^*)}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*}.$$

Similarly we can obtain the following inequalities:

$$A_2\left(\frac{\beta_{v2}(1 - z^*)}{\beta_{h2}z^* + \gamma_2}\right) < \frac{\beta_{v2}y_2^*(1 - z^*)}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*},$$

$$A_3\left(\frac{\beta_{v3}(1 - z^*)}{\beta_{h3}z^* + \gamma_3}\right) < \frac{\beta_{v3}y_3^*(1 - z^*)}{\displaystyle\sum_{i=1}^{3}\beta_{vi}y_i^*}.$$

Substitution of these inequalities into the expression of λ_+ yields

$$\lambda_+^2 < 1 - z^*.$$

Since $0 < z^* < 1$, it follows that $\lambda_+ < 1$ and that E^* is l.a.s.

This finishes the proof.

4. **Slow dynamics of population genetics.** By using the rescaled time $\tau = \epsilon t$, we can rewrite the full system (6) as

$$\epsilon \frac{dy_1}{d\tau} = \beta_{h1} z (1 - r - s - y_1) - \gamma_1 y_1 - \epsilon \, y_1 F_1(y_1, y_2, y_3, z, r, s, N),$$

$$\epsilon \frac{dy_2}{d\tau} = \beta_{h2} z (r - y_2) - \gamma_2 y_2 - \epsilon \, y_2 F_2(y_1, y_2, y_3, z, r, s, N),$$

$$\epsilon \frac{dy_3}{d\tau} = \beta_{h3} z (s - y_2) - \gamma_3 y_3 - \epsilon \, y_3 F_3(y_1, y_2, y_3, z, r, s, N),$$

$$\epsilon \frac{dz}{d\tau} = (1 - z) \sum_{i=1}^{3} \beta_{vi} y_i - \delta z,$$

$$\frac{dr}{d\tau} = (P_2 - r)\tilde{b}(N) - \tilde{\alpha}_2 y_2 + r\left(\tilde{m}_1(1 - r - s) \right. \tag{11}$$

$$\left. -\tilde{m}_2 + \tilde{m}_2 r + \tilde{m}_3 s + \sum_{i=1}^{3} \tilde{\alpha}_i y_i \right),$$

$$\frac{ds}{d\tau} = (P_3 - s)\tilde{b}(N) - \tilde{\alpha}_3 y_3 + s\left(\tilde{m}_1(1 - r - s) \right.$$

$$\left. -\tilde{m}_3 + \tilde{m}_2 r + \tilde{m}_3 s + \sum_{i=1}^{3} \tilde{\alpha}_i y_i \right),$$

$$\frac{dN}{d\tau} = N\left(\tilde{b}(N) - \tilde{m}_1(1 - r - s) - \tilde{m}_2 r - \tilde{m}_3 s - \sum_{i=1}^{3} \tilde{\alpha}_i y_i \right),$$

where

$$F_i = \tilde{b}(N) - \tilde{m}_1(1 - r - s) + \tilde{m}_i - \tilde{m}_2 r - \tilde{m}_3 s + \tilde{\alpha}_i - \sum_{j=1}^{3} \tilde{\alpha}_j y_j, \quad i = 1, 2, 3.$$

This system has a three-dimensional slow manifold:

$$M = \left\{ (y_1, y_2, y_3, z, r, s, N) : y_i = y_i^*(r, s, N), i = 1, 2, 3, \; z = z^*(r, s, N) \right\},$$

which is normally hyperbolically stable, as it consists of a set of such equilibria E^* of the fast system (7). The functions y_i^*, $i = 1, 2, 3$, and z^* are given in (8) and (9). The slow dynamics on M are described by the equations

$$\frac{dr}{d\tau} = (P_2 - r)\tilde{b}(N) - \tilde{\alpha}_2 y_2^* + r\left(\tilde{m}_1(1 - r - s) \right.$$

$$\left. -\tilde{m}_2 + \tilde{m}_2 r + \tilde{m}_3 s + \sum_{i=1}^{3} \tilde{\alpha}_i y_i^* \right),$$

$$\frac{ds}{d\tau} = (P_3 - s)\tilde{b}(N) - \tilde{\alpha}_3 y_3^* + s\left(\tilde{m}_1(1 - r - s) \right. \tag{12}$$

$$\left. -\tilde{m}_3 + \tilde{m}_2 r + \tilde{m}_3 s + \sum_{i=1}^{3} \tilde{\alpha}_i y_i^* \right),$$

$$\frac{dN}{d\tau} = N\left(\tilde{b}(N) - \tilde{m}_1(1 - r - s) - \tilde{m}_2 r - \tilde{m}_3 s - \sum_{i=1}^{3} \tilde{\alpha}_i y_i^* \right),$$
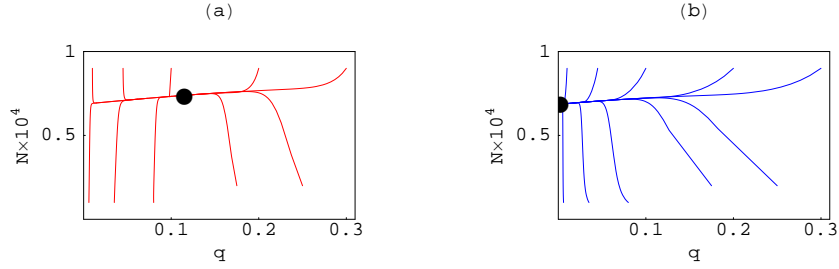
FIGURE 1. Phase portraits of system (12) in the (q, N) plane ($q = r/2 + s$). For (a) and (b) all parameters are equal except ν_2 and ν_3. In (a) $\nu_2 = \nu_3 = 5 \times 10^{-5}$ for which $\mathcal{F} > 0$. It shows that there is a stable interior equilibria (represented by a "$\bullet$"). In (b) $\nu_2 = \nu_3 = 6 \times 10^{-5}$ for which $\mathcal{F} < 0$. It shows that there is no interior equilibrium and that the only stable equilibrium is on the N-axis. Other parameter values are given in the text.

where

$$P_2 = \left(1 - \frac{r}{2} - s\right)\left(r + 2s\right), \qquad P_3 = \left(\frac{r}{2} + s\right)^2,$$

and again y_1^*, y_2^*, y_3^* and z^* are given in (8) and (9).

Since M is normally hyperbolically stable, geometric theory of singular perturbations due to Fenichel [7] allows us to study the system (11) by studying the reduced slow system (12) (see [6] for more details). In other words, if the dynamics of system (12) can be characterized with bifurcations, then the bifurcating dynamics on the slow manifold M are structurally stable and hence robust, subject to perturbations. Therefore, results from the slow system will provide bifurcation properties of the system (11) as well as the full system (6).

Some interesting scenarios are illustrated in Figure 1, which shows numerical solutions of the slow system for two sets of parameter values. The frequency of the S-gene is $q = r/2 + s$. The two sets of parameters used for the figures are chosen to give positive (Figure 1(a)) and negative (Figure 1(b)) values, respectively, of the S-gene fitness $\mathcal{F}$ (whose definition is given in the next section). Figure 1(b) shows that the boundary equilibrium on the N-axis of the slow system is a global attractor for negative fitness. When the fitness is positive, Figure 1(a) shows that it is possible for the S-gene to be maintained provided that the initial density of the S-gene is not too low. One of the important features of this coupled system is the possibility of multiple interior equilibria (see Figure 1(a)) and the bistability of two nontrivial equilibria (represented by the solid circle). The parameter values used are $\beta_{h1} = 0.04, \beta_{h2} = \beta_{h3} = 0.03, \beta_{v1} = \beta_{v2} = \beta_{v3} = 0.03,$ $\delta = 0.1, \gamma_1 = \gamma_2 = \gamma_3 = 0.01, \alpha_1 = 2.3 \times 10^{-4}, \alpha_2 = \alpha_3 = 2 \times 10^{-4}, m = 4 \times 10^{-5},$ $b = 3.4 \times 10^{-4}$, and $K = 10000$. These outcomes are usually absent in population genetic models without a dynamic disease process.

5. **Fitness of the S-gene.** Recall that $q = r/2 + s$ is the S-gene frequency. The invasion ability of the S-gene can be described by

$$\frac{1}{q}\frac{dq}{d\tau}\bigg|_{q=0}.$$

We will use this quantity to define the fitness of the S-gene, $\mathcal{F}$. Consider the case when $\nu_3 = \nu_2 > 0$; i.e., the extra death rate due to the S-gene in AS and SS individuals is equal. Then $m_3 = m_2$, and using the r and s equations in (12) we get

$$\frac{dq}{d\tau} = \left(\frac{1}{2}P_2 + P_3 - q\right)\tilde{b}(N) - \frac{1}{2}\tilde{\alpha}_2 y_2^* - \tilde{\alpha}_3 y_3^* + q\bigg(\tilde{m}_1(1 - r - s)$$

$$- \tilde{m}_2 + \tilde{m}_2 r + \tilde{m}_3 s + \sum_{i=1}^{3}\tilde{\alpha}_i y_i^*\bigg). \tag{13}$$

Since $0 \le r,\ s \le q$, we have $r = s = 0$ if $q = 0$. Using (8) and (9), and noticing that $\tilde{\alpha}_3 = \tilde{\alpha}_2$ and $T_{h3} = T_{h2}$, we have

$$\frac{1}{q}\left(\frac{1}{2}\tilde{\alpha}_2 y_2^* + \tilde{\alpha}_3 y_3^*\right)\bigg|_{q=0} = \frac{\tilde{\alpha}_2 T_{h2} z^*}{1 + T_{h2} z^*}\bigg|_{r=s=0}$$

$$= \frac{\tilde{\alpha}_2 T_{h2}(\mathcal{R}_1 - 1)}{(1 + T_{h2})\mathcal{R}_1 + T_{h1} - T_{h2}}. \tag{14}$$

Here, we have used the fact that $z^*|_{r=s=0} = (\mathcal{R}_1 - 1)/(\mathcal{R}_1 + T_{h1})$ and the following relation between k_i, i=0, 1, 2 (see (10)):

$$k_1^2 - 4k_0 k_2\big|_{r=s=0} = \left(T_{h1} - T_{h2} + \mathcal{R}_1(1 + T_{h2})\right)^2.$$

We can also show that

$$y_1^*\big|_{q=0} = \frac{T_{h1} z^*}{1 + T_{h1} z^*}\bigg|_{r=s=0} = \frac{T_{h1}(\mathcal{R}_1 - 1)}{(1 + T_{h1})\mathcal{R}_1}, \tag{15}$$

and

$$\frac{1}{q}\left(\frac{1}{2}P_2 + P_3 - q\right)\bigg|_{q=0} = 0. \tag{16}$$

It follows from (13)-(16) that the fitness of the S-gene is given by

$$\mathcal{F} = \frac{1}{q}\frac{dq}{d\tau}\bigg|_{q=0} = \tilde{m}_1 + W_1\tilde{\alpha}_1 - \left(\tilde{m}_2 + W_2\tilde{\alpha}_2\right), \tag{17}$$

where

$$W_1 = \frac{T_{h1}(\mathcal{R}_1 - 1)}{(1 + T_{h1})\mathcal{R}_1}, \quad W_2 = \frac{T_{h2}(\mathcal{R}_1 - 1)}{(1 + T_{h2})\mathcal{R}_1 + T_{h1} - T_{h2}}. \tag{18}$$

Clearly, $\mathcal{F}$ is determined by the difference of weighted death rates between the non-sickled (AA) and sickled (AS and SS) individuals, and the weights, W_1 and W_2, depend only on epidemiological parameters. This allows us to explore the effect of malaria epidemics on the distribution of genotypes.

The following result confirms that the fitness $\mathcal{F}$ defined above indeed determines the invasion ability of the S-gene.

Theorem 3. *The slow system (12) has a trivial equilibrium $U_0 = (0, 0, N_0)$, where $N_0 = K\left(1 - (\tilde{m}_1 + \tilde{\alpha}_1 y_1^*)/\tilde{b}\right)$. U_0 is l.a.s. if $\mathcal{F} < 0$ and unstable if $\mathcal{F} > 0$.*

Proof. We first notice that (see (14), (15), and (18))

$$\left.\frac{T_{h1}z^*}{1+T_{h1}z^*}\right|_{r=s=0} = W_1,$$

$$\left.\frac{T_{h3}z^*}{1+T_{h3}z^*}\right|_{r=s=0} = \left.\frac{T_{h2}z^*}{1+T_{h2}z^*}\right|_{r=s=0} = W_2.$$

Then the Jacobian matrix at U_0 can be written

$$J = \begin{pmatrix} \tilde{m}_1 + W_1\tilde{\alpha}_1 - \left(\tilde{m}_2 + W_2\tilde{\alpha}_2\right) & \tilde{b}(N_0) & 0 \\ 0 & \tilde{m}_1 + W_1\tilde{\alpha}_1 - \left(\tilde{m}_3 + W_2\tilde{\alpha}_3\right) & 0 \\ * & * & -\frac{\tilde{b}N_0}{K} \end{pmatrix},$$

where a "$*$" denotes a number that does not affect the eigenvalues of J. Since $\tilde{\alpha}_3 = \tilde{\alpha}_2$ and $\tilde{m}_3 = \tilde{m}_2$ we know that the three eigenvalues of J are

$$\lambda_1 = \lambda_2 = \mathcal{F}, \quad \lambda_3 = -\frac{\tilde{b}N_0}{K}.$$

Therefore, J has three negative eigenvalues if $\mathcal{F} < 0$ and has a positive eigenvalue if $\mathcal{F} > 0$. This completes the proof.

Theorem 3 implies that the S-gene cannot invade if the fitness is negative and that the invasion is possible if the fitness is positive. This result is confirmed by numerical computations of the slow system (see Figure 1). Figure 1(a) is for the case $\mathcal{F} > 0$, which shows that there is a locally asymptotically stable interior equilibrium, and Figure 1(b) is for the case $\mathcal{F} < 0$, which shows that there is no interior equilibrium and that the boundary equilibrium on the N-axis is a global attractor. These numerical results also suggest that the fitness measure given by $\mathcal{F}$ provides a criterion not only for invasion but also for the possible maintenance of the S-gene. Figure 1(a) shows that the S-gene can establish itself if its initial frequency is not too low. Notice from (17) and (18) that the fitness $\mathcal{F}$ is influenced by several of the parameters associated with malaria epidemiology. Figure 2(a) illustrates the joint effect of β_{h1} (malaria infection rate of type 1 humans) and γ_1 (recovery rate from malaria of type 1 humans) on the fitness. A contour plot of Figure 2(a) is shown in Figure 2(b), which can also be viewed as a bifurcation diagram in the parameter plane (γ_1, β_{h1}) indicating the region in which U_0 is stable ($\mathcal{F} < 0$) or unstable ($\mathcal{F} > 0$). This bifurcation is confirmed by our numerical simulations of the system (see Figures 1 and 3).

The slow system (12) can also have multiple interior equilibria and bistable non-trivial equilibria when the fitness is positive. Figure 3(a) is a phase portrait showing that there are two interior equilibria with one of them locally asymptotically stable (represented by a solid circle) while the other one unstable (a saddle, represented by a solid triangle). There is another stable equilibrium on the the N-axis. Figure 3(b) is a time plot showing the two locally stable equilibria. Parameter values for this figure are the same as those used for Figure 1 except that $\beta_{v2} = \beta_{v3} = 0.12$.

6. **Discussion.** We extended the model in [5] by including all three genotypes of individuals. More complexity is introduced by adding the additional subpopulation of SS individuals, which increased the dimension of the full system from 5 to 7 and the dimension of the slow system from 2 to 3. However, the qualitative behaviors
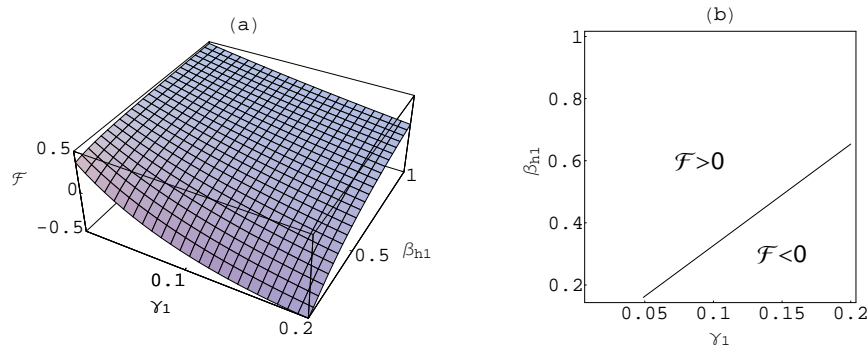
FIGURE 2. (a) Fitness $\mathcal{F}$ vs γ_1 and β_{h1}. (b) A contour plot showing the region in the (γ_1, β_{h1}) plane in which the fitness is positive ($\mathcal{F} > 0$) or negative ($\mathcal{F} < 0$).
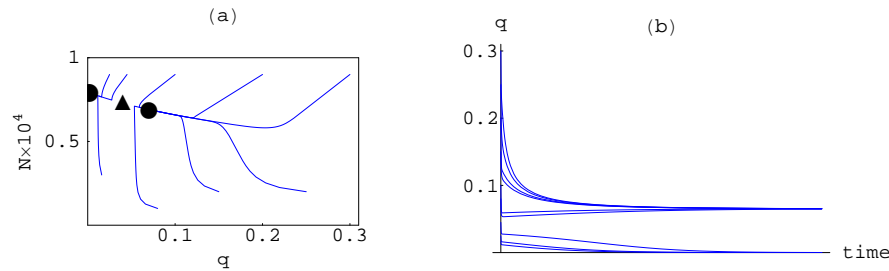

FIGURE 3. The slow system (12) can have multiple interior equilibria and bistable nontrivial equilibria when the fitness is positive. Figure 3(a) is a phase portrait showing that there are two interior equilibria, with one of them locally asymptotically stable It shows that (represented by a solid circle) and the other one unstable (represented by a solid triangle). There is another stable equilibrium on the the N-axis. Figure 3(b) is a time plot showing the two locally stable equilibria. Parameter values for this figure are the same as those used for Figure 1 except that $\beta_{v2} = \beta_{v3} = 0.12$.

of the two models seem to be the same. By coupling the malaria epidemiology and the sickle-cell genetics we investigate how the epidemiological and demographic parameters may affect the fitness of the sickle-cell gene. We derived threshold conditions which allow us to explore the impact of malaria disease dynamics on the genetic composition of the human population. From these threshold conditions, we can draw conclusions similar to those obtained in [5]. For example, whether the rare gene will go to extinction or persist in a population (on the slow time scale) is determined by the fitness coefficient of the gene, and this fitness coefficient depends not only on parameters related to the genetics of the S-gene but also on the parameters associated with malaria epidemiology which determine the disease prevalence. Our analytic results consider the ability of a rare gene to invade a host population composed of mainly wild-type individuals. This invasion ability

provides a measure for the fitness of the S-gene $\mathcal{F}$. Our numerical studies indicate that when $\mathcal{F} > 0$ the S-gene not only can invade but also will be maintained in the host population. Moreover, the slow system (host population genetics) can have multiple interior equilibria and multiple stable nontrivial equilibria. This occurs when $\mathcal{F}$ is negative but close to zero. In this case, although the S-gene cannot invade when its initial frequency is low, it may be able to establish itself if its value becomes large due to some stochastic perturbations (e.g., environmental changes). The main findings from this model are 1) the S-gene may be selected for if the selection force from malaria is strong; and 2) multiple stable equilibria are possible. These types of population dynamics are not observed in most models of population genetics models without coupling malaria epidemiology.

One would expect this malaria-sickle cell system to exhibit oscillatory behaviors under certain conditions for coexistence. Intuitively, the frequency of the sickle-cell gene will decrease in the absence of malaria due to a higher S-gene-related death rate in the SS and AS individuals than in the homozygote wild-type individuals (AA). On the other hand, the S-gene may be selected for if the endemic level of malaria is sufficiently high, and consequently polymorphism in the host population may be maintained in an oscillatory manner. However, as is seen in [5], the ODE system (12) or (4) does not produce periodic solutions. To observe such oscillatory dynamics we looked at a further extension of the model to include a time delay in the density dependence of the birth function. For example, if we choose the birth function in system (12) to be of the form $\tilde{b}[1 - N(t - T)/K]$, where T denotes a constant time delay, then the modified system is capable of producing oscillations (see Figure 4). From (13) and (16) we see that the introduction of delay will not affect the formula for the fitness $\mathcal{F}$ given in (17). Notice that $m_2 - m_1 = \nu$, where ν is the extra death rate of individuals with the S-gene. Rewrite this formula (17) as

$$\mathcal{F} = W_1 \tilde{\alpha}_1 - W_2 \tilde{\alpha}_2 - \nu.$$

The first two terms involve only malaria-associated parameters with the property $W_1 \tilde{\alpha}_1 - W_2 \tilde{\alpha}_2 > 0$. Therefore, the sign of $\mathcal{F}$ is determined by the genetic parameter ν if all the malaria related parameters are fixed.

Our numerical simulations of the slow system (12) show that there exists a critical value ν_c such that solutions converge to a positive equilibrium for $\nu > \nu_c$ (top panel of Figure 4) and that oscillations occur for $\nu < \nu_c$ (middle panel of Figure 4) for a given value of the delay $T = 0.25$. We can also fix the value of ν and vary T to observe the switching from a stable equilibrium to periodic oscillations. For example, the top and bottom figures in Figure 4 have the same ν value but different T values. Other parameter values used are $\beta_{h1} = 0.4, \beta_{h2} = \beta_{h3} = 0.1$, $\beta_{v1} = 0.3$, $\beta_{v2} = \beta_{v3} = 0.1$, $\delta = 0.05, \gamma_1 = 0.05, \gamma_2 = \gamma_3 = 0.06$, $\tilde{\alpha}_1 = 3, \tilde{\alpha}_2 = \tilde{\alpha}_3 = 1$, $\tilde{m}_1 = 0.1, \tilde{m}_2 = \tilde{m}_1 + \nu_1, \tilde{m}_3 = \tilde{m}_2 + 0.25$, $\tilde{b} = 7.8$, and $K = 10000$. Figure 5 shows that the full system (4) also exhibits similar dynamics as that of the slow system (12). From the oscillatory solutions (see Figure 6) we observe that for most of the time the change in the gene frequency $q\ (= r/2 + s)$ follows the change in the endemic level of malaria (i.e., the fraction of total infected people $y =: y_1 + y_2 + y_3$). For example, q increases with increasing y until y reaches its maximum. Soon after y starts decreasing, q also starts decreasing, until y reaches its minimum.

Other modifications of the model may also lead to periodic solutions. We have only presented a simple example. In any case, by explicitly coupling the malaria disease dynamics with changes in the frequency of the S-gene, our model allows
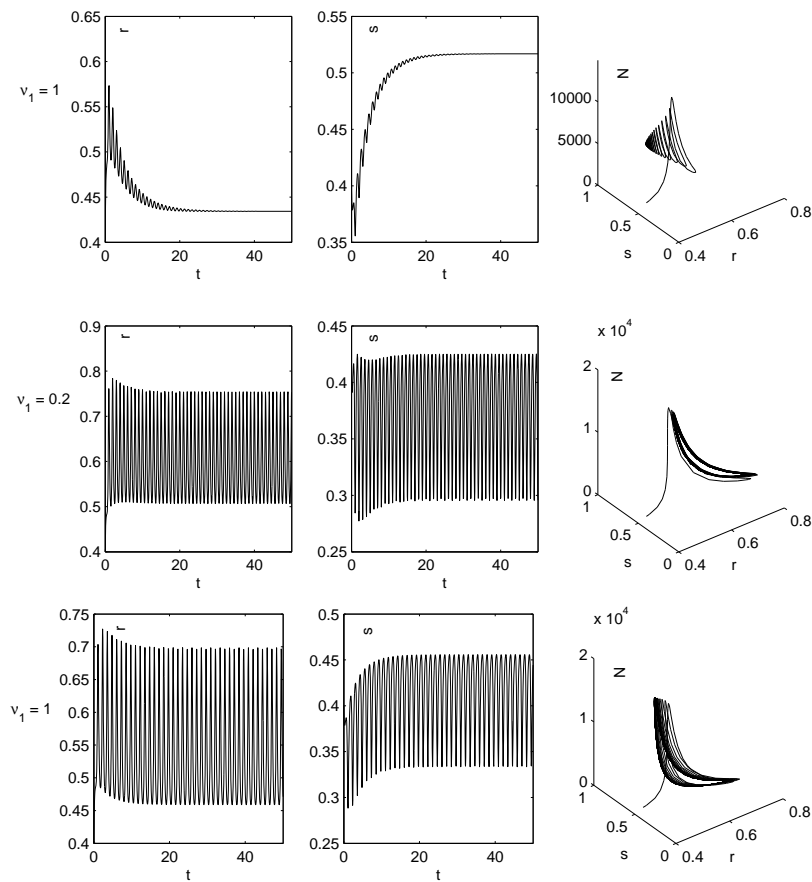
FIGURE 4. Numerical solutions of system (12). For the top and middle panels the time delay is $T = 0.25$. The top panel is for $\nu > 0.85$ ($\nu = 1$), in which case solutions converge to the interior equilibrium. The middle panel is for $\nu < 0.85$ ($\nu = 0.2$), in which case a stable periodic solution exists. The bottom panel is for the same ν value as that in the top panel ($\nu = 1$) but a different value for the delay ($T = 0.3$). A stable periodic solution also exists in this case. In fact, stable periodic solutions exist for all $T > 0.27$. The three plots on the right are phase portraits in the (s, r, N) space for corresponding sets of parameter values.

us to explore how human population genetics respond to the prevalence of malaria which provides a dynamic environment. These results cannot be obtained with genetic models that assume a static environment.
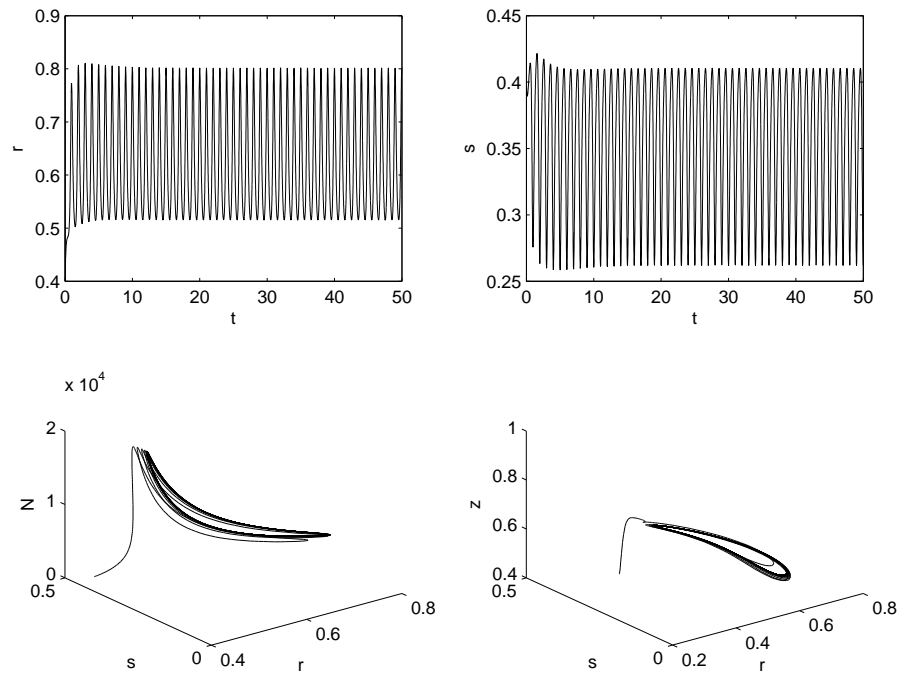
FIGURE 5. Numerical solutions of the full system (4). Behaviors of the solutions of the full system (e.g., the lower-right, is a phase portrait in the (s, r, z) space) are similar to those given by the slow system (12).
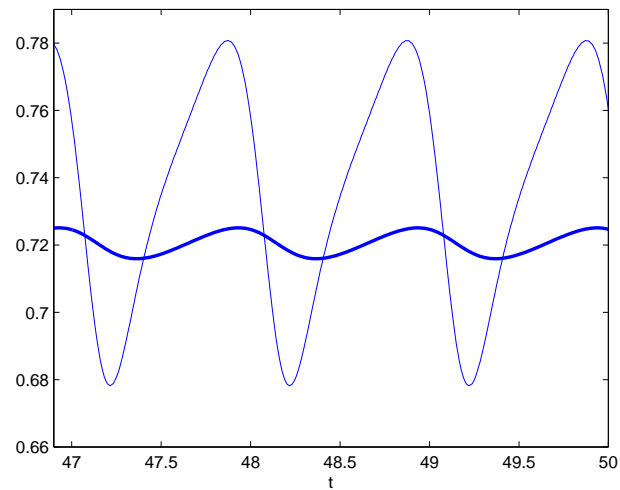

FIGURE 6. Plots of malaria endemicity ($y_1 + y_2 + y_3$ multiplied by 6, the thinner line) and S-gene frequency ($q = r/2 + s$, the thicker line) versus time t. The parameters have the same values as those used in the middle panel of Figure 3 except that $\nu = 0.45$.

REFERENCES

[1] Andreasen, V. 1993. Disease-induced natural selection in a diploid host. Theo. Population Biol., 44 (3), 261-298.

[2] Beck, K., J.P. Keener, and P. Ricciardi. 1984. The effect of epidemics on genetic evolution. J. Math. Bio., 19, 79-94.

[3] Castillo-Chavez, C. and Z. Feng. 1997. To treat or not to treat: the case of tuberculosis. Journal of Mathematical Biology, 35, 629-659.

[4] Diekmann, O., J.A.P. Heesterbeek and J.A.J Metz. 1990. On the definition and the computation of the basic reproduction ratio in models for infectious diseases in heterogeneous population. Journal of Math. Bio., 28, 365-382.

[5] Feng, Z., D.L. Smith, F.E. McKenzie and S.A. Levin. 2004a. Coupling ecology and evolution: malaria and the S-gene across time scales. Mathematical Biosciences, 189(1), 1-19.

[6] Feng, Z., Y. Yi and H. Zhu. 2004b. Malaria epidemics and the sickle-cell genetic dynamics. Journal of Dynamics and Differential Equations, 16(4), 869-896.

[7] Fenichel, N. 1979. Geometric singular perturbation theory for ordinary differential equations. J. Diff. Eqn., 31, 53-98.

[8] Galvani, A.P. and M.W. Slatkin. 2003. Evaluating plague and smallpox as historical selective pressures for the $CCR5 - \Delta32$ HIV-resistance allele. PNAS 100(25), 15276-15279.

[9] Galvani, A.P. and M.W. Slatkin. 2004. Intense selection in an age-structured population. Proceedings of the Royal Society London Series B 271(1535), 171-176.

[10] May, R.M. and R.M. Anderson. 1983. Epidemiology and genetics in the coevolution of parasites and hosts. Proc. Roy. Soc. London. B, 219 (1216), 281-313.

[11] McKenzie, F.E. et al. 1998. Discrete-event simulation models of *Plasmodium falciparum* malaria. Simulation, 71, 250-261.

[12] McKenzie, F.E. and W. H. Bossert. 1997. The dynamics of *Plasmodium falciparum* blood-stage infection. J. Theo. Bio., 188 (1), 127-140.

[13] Hsu-Schmitz, S. 2002. Effects of genetic heterogeneity on HIV transmission in homosexual populations. *in*: Mathematical approaches for emerging and reemerging infectious diseases: Models, methods, and theory, C. Castillo-Chavez et al. (eds.), IMA, 126, 245-260.