

# Interactions Between the Frequency of the Duffy Antigen and the Dynamics of *P. vivax* Malaria Infections

*Elizabeth Gharthey*<sup>1,a,\*</sup>, *Gautam Rai*<sup>2,a,\*</sup>, *Dasha Selivonenko*<sup>3,a,\*</sup>, *Rachel K. Wissenbach (Mentor)*<sup>2,a</sup>, *Lucero Rodriguez Rodriguez (Mentor)*<sup>2,a</sup>, and *Joan Ponce (Mentor)*<sup>2,a</sup>

*\*These authors contributed equally to the work.*

<sup>1</sup>*University of Arizona, Tucson AZ, United States*

<sup>2</sup>*Arizona State University, Tempe AZ, United States*

<sup>3</sup>*Rice University, Houston TX, United States*

<sup>a</sup>*Simon A. Levin Mathematical, Computational, and Modeling Sciences Center: Quantitative Research for the Life & Social Sciences Program, Arizona State University, Tempe AZ, United States*

July 19, 2024

## Abstract

The malarial parasite *Plasmodium vivax* has infected humans for millennia. As such, alleles like the FY\*O allele in the Duffy Antigen Receptor for Chemokines (DARC) and the sickle cell allele (HbS) have been naturally selected for in malaria-endemic regions because they confer resistance to malaria, thereby increasing the survival and reproductive success of individuals carrying these protective alleles. As resistance becomes more common in a population, malaria incidence is expected to decline, reducing the evolutionary pressure for additional resistance. In this work, we explore the interaction between these processes. We construct a model with seasonality that tracks the frequency of Duffy genotypes and leverages fast/slow dynamics to analyze the coupled dynamics of malaria transmission and changes in the gene frequency of the DARC genotype. Specifically, we investigate how the burden of malaria changes with the fractions of people with the various DARC genotypes. Additionally, we calibrate our model using data from the Amazonas region in Brazil, which has a polymorphic population with respect to DARC, and still reports a substantial number of *P. vivax* cases. Analysis of our model determines the proportion of the population that must be Duffy-negative in order for the entire population to be protected against *P. vivax* without any further interventions. Furthermore, we assess how different proportions of Duffy-negative individuals influence the monthly incidence of *P. vivax* cases.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Influence of Genetics on the Epidemiology of Malaria . . . . .	3
1.2	Previous Mathematical Models of Natural Protection Against Malaria . . . . .	4
<b>2</b>	<b>Model</b>	<b>5</b>
2.1	Model Formulation . . . . .	5
2.2	Human population genetic model . . . . .	9
2.3	Re-scaled model . . . . .	11
2.4	Slow Dynamical System . . . . .	11
2.4.1	Manifold of Slow System . . . . .	12
2.4.2	Determining Fitness of the X-gene . . . . .	12
2.5	Fast Dynamical System . . . . .	14
2.6	The Next Generation Matrix and the Basic Reproduction Number of the Fast System . . . . .	15
2.7	Conditions for Stability of Disease-Free Equilibrium . . . . .	16
2.8	Location of Nontrivial Equilibrium . . . . .	17
2.9	Conditions for Stability of Endemic Equilibrium . . . . .	19
<b>3</b>	<b>Sensitivity Analysis</b>	<b>21</b>
3.1	Calculating The Elasticity of $\mathcal{R}_0$ . . . . .	22
3.2	Global Sensitivity on Fast System with Biting Rate Seasonality . . . . .	22
<b>4</b>	<b>Parameter Estimation</b>	<b>23</b>
4.1	Relative Risk of Infection Based on Genotype . . . . .	23
4.2	Fitting the Fast System with Seasonality . . . . .	25
<b>5</b>	<b>Results</b>	<b>27</b>
<b>6</b>	<b>Discussion</b>	<b>30</b>
<b>7</b>	<b>Acknowledgements</b>	<b>31</b>
<b>8</b>	<b>Appendix</b>	<b>32</b>
8.1	Mosquito Model Re-scale (3) . . . . .	32

# 1 Introduction

## 1.1 Influence of Genetics on the Epidemiology of Malaria

Malaria is an infectious disease spread by mosquitoes that are infected with *Plasmodium* parasites. Whereas *P. falciparum* is the most dangerous strain and the most prevalent in Africa, *P. vivax* is the most geographically widespread strain of malaria and the most common in the rest of the world [13]. Malaria is a perennial affliction to many tropical countries: in 2022, it caused an estimated 249 million cases globally, resulting in more than six hundred thousand deaths and billions of dollars in global costs [23].

The selective pressure induced by malaria has demonstrably influenced the genetic composition of tropical populations. Examples include hemoglobinopathies like the sickle cell trait in Sub-Saharan Africa and  $\beta$ -thalassemia in the Eastern Mediterranean. The sickle cell trait is caused by a missense mutation of the  $\beta$ -globin gene, resulting in the abnormal synthesis of the  $\beta$ -globin protein. The vast majority of people with sickle cell trait and sickle cell disease live in Sub-Saharan Africa. Similarly,  $\beta$ -thalassemia is characterized by the obstructed synthesis of the  $\beta$ -globin chain, which causes numerous phenotypes that may be asymptomatic or debilitating. People with sickle cell trait and some phenotypes of  $\beta$ -thalassemia have significant protection against *P. falciparum* malaria compared to people without either condition [6, 8, 11].

Another example of an adaptation against malaria is the absence of the Duffy antigen receptor for chemokines (DARC) on human red blood cells. *P. vivax* protozoa rely on the interaction between DARC and the *Plasmodium vivax* Duffy-binding protein (PvDBP) to invade human reticulocytes. Having the Duffy-negative trait thereby removes a mode of invasion into reticulocytes by *P. vivax* protozoa. Nonetheless, infections by *P. vivax* have been observed in Duffy-negative individuals, but the mechanism of infection in these cases are not fully understood [20].

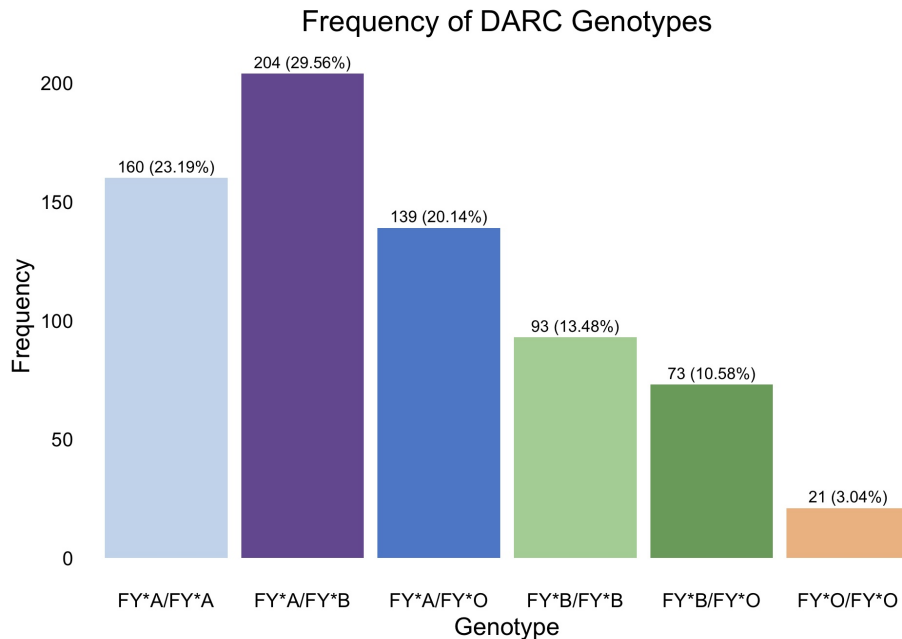
The transmission patterns of malaria are highly influenced by the Duffy-negative trait. Duffy-negativity is very common in sub-Saharan Africa, so the fraction of the population at risk of *vivax* malaria is very small. Consequently, the majority of malaria infections are caused by *P. falciparum*. Although *P. vivax* infections is exceedingly rare in sub-Saharan Africa, it is the most common malarial parasite in the rest of the world [10].

Our system models the dynamics of *P. vivax* malaria in the Amazonas state of Brazil, because this population is much more polymorphic with respect to the Duffy antigen trait than most parts of the world. Kano et al. [16] conducted a longitudinal study on the influence of DARC on people's susceptibility to *vivax* malaria in Rio Pardo, which is located in the Brazilian state of Amazonas. In this study, they sampled antibody response to *vivax* malaria, and disaggregated it by Duffy genotype to ascertain the influence of natural immunity. We used data from the longitudinal work of Kano et al to determine the populations of interest in our study.

## 1.2 Previous Mathematical Models of Natural Protection Against Malaria

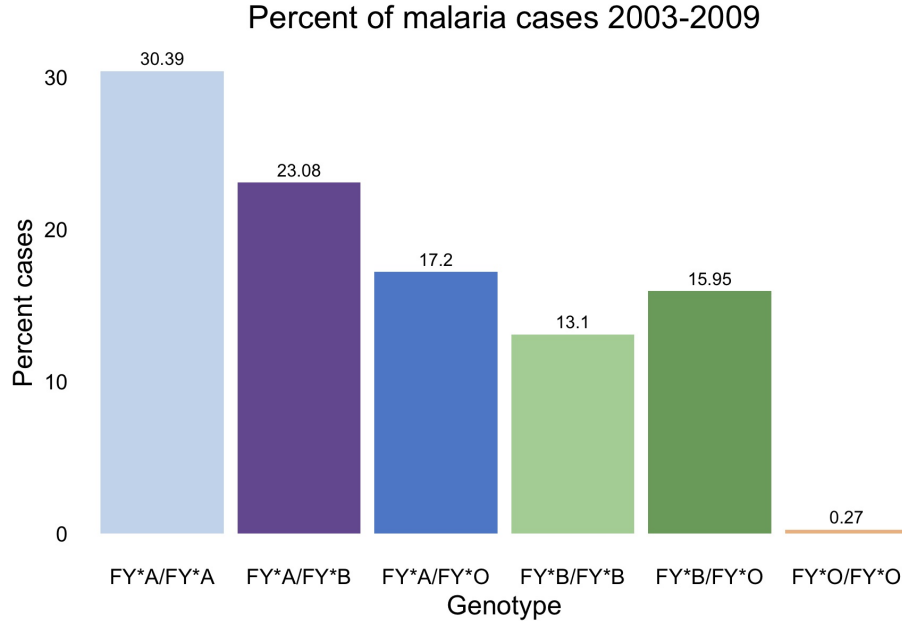
Previous mathematical models of malaria transmission dynamics frequently use epidemiological compartmental models like the SIR model. In this paper, we use ordinary differential equations to characterize the dynamics of *vivax* malaria in Rio Pardo, Brazil, with relation to the Duffy antigen. We reference previous population genetics models that relate to the sickle cell trait. Feng et al. [7] used a system of ordinary differential equations to describe the relationship between malaria transmission dynamics and the genetic frequency of the S-gene that causes sickle cell anemia. Furthermore, Feng and Castillo-Chavez [6] used an expanded system of ordinary differential equations with three elements to characterize the complex dynamics of the S-gene beyond the standard two-element system.

We use the Rio Pardo area studied by Kano et al. [16], because it is more polymorphic than most parts of the world. This has implications for transmission dynamics of *vivax* malaria in the region. Figure 1 denotes the genotypic frequencies of the Duffy trait in Rio Pardo.



**Figure 1: Distribution of Duffy antigen polymorphisms in Rio Pardo, Amazonas, Brazil.**

The above graph shows that the large majority of people in Rio Pardo are Duffy-positive, but a considerable minority are also homozygous Duffy-negative. The characteristics of the Duffy antigen suggest that people who are Duffy negative (FY\*O/FY\*O) are highly unlikely to contract *vivax* malaria. This is supported by Figure 2, which shows the *vivax* malaria cases by genotype in the Kano study [16].



**Figure 2: Malaria cases by genotype in a longitudinal study conducted in Rio Pardo, Amazonas, Brazil. Besides corroborating the protective effects of the Duffy-negative trait, this data also suggests that the FY\*B trait has limited protective effects against *vivax* malaria.**

The paper is structured as follows: we introduce the terms, parameters, and variables of our mathematical model. We then describe our model and its dynamics. We formulate the slow system, from which we deduce its manifold and the fitness of the X-gene. We use this information to simulate the fast system and to determine its next generation matrix, basic reproduction number, and equilibria. We conduct a global sensitivity analysis on the model, eventually including biting rate seasonality, and then estimate our parameters. We then conclude the paper by assessing the results and implications of our model.

## 2 Model

### 2.1 Model Formulation

There are four possible phenotypes of the Duffy antigen: Duffy-positive, where two of the alleles (FY\*A and/or FY\*B) alleles are present; the two heterozygous Duffy-positive traits, where one of the two alleles is present and the other is silent; and the homozygous Duffy-negative trait, where neither the FY\*A or FY\*B allele is expressed. For our paper, we use the following nomenclature for the Duffy blood group [12, 16, 18]:

Meaning	Phenotype	Alleles & Genotype	Classification
Duffy-Positive	Fy( $a + b+$ )	FY*A/FY*B	XX
	Fy( $a + b-$ )	FY*A/FY*A	XX
		FY*A/FY*O	XO
	Fy( $a - b+$ )	FY*B/FY*B	XX
		FY*B/FY*O	XO
Duffy-Negative	Fy( $a - b-$ )	FY*O/FY*O	OO

In this paper, we use the term XX to refer to the presence of two expressed Duffy alleles, XO for the presence of one expressed Duffy allele and one silent Duffy allele, and OO for the presence of two silent Duffy alleles. The subscript  $i$  denotes the genotype:  $i = 1$  for XX,  $i = 2$  for XO, and  $i = 3$  for OO.

We structure our model similarly to the that Feng and Castillo-Chavez [6] used to describe the dynamics of the sickle cell trait. Distinct differences between our models exist: whereas the sickle cell model developed by Feng incorporates a death rate associated with the sickle cell trait, our model does not implement a death rate due to the Duffy antigen. Furthermore, our model considers the impact of seasonality on our numerical simulations. The variables of our initial model are described in Table 1 and a summary of the parameters is presented in Table 2.1.

**Table 1: Variables and description for original model.** Our values for  $P_i$  and  $w_i$  are based in Rio Pardo, Brazil.

State Variable	Description
$u_i$	Uninfected individuals of genotype $i$
$v_i$	Infected individuals of genotype $i$
$z$	Fraction of infected mosquitoes
$x_i$	Fraction of uninfected individuals of genotype $i$
$y_i$	Fraction of infected individuals of genotype $i$
$w_i$	Fraction of individuals in each genotype
$N$	Total population
$S_m$	Number of uninfected mosquitoes
$I_m$	Number of infected mosquitoes
$M$	Total number of mosquitoes

**Table 2: Parameters, descriptions, values, units, and source.** The values for  $P_i$  and  $w_i$  are for Rio Pardo, Brazil.

Human Parameters	Description	Value	Unit	Source
$P_1$	Fraction of total births with XX genotype	0.6657	people	[16]
$P_2$	Fraction of total births with XO genotype	0.3003	people	[16]
$P_3$	Fraction of total births with OO genotype	0.0339	people	[16]
$w_1$	Fraction of XX people	0.6623	N/A	[16]
$w_2$	Fraction of XO people	0.3072	N/A	[16]
$w_3$	Fraction of OO people	0.0304	N/A	[16]
$N$	Total Population	690	People	[16]
$b$	Base human birth rate	1.26e-5	1/day	[4]
$K$	Density dependent reduction in birth rate	10000	people	[6]
$\theta_1$	Probability of infection from infected bite for XX	0.78	N/A	[5]
$\theta_2$	Probability of infection from infected bite for XO	0.78	N/A	[5]
$\theta_3$	Probability of infection from infected bite for OO	0.00702	N/A	[5, 16]
$\gamma_1$	Recovery Rate for XX	1/24	1/day	[27]
$\gamma_2$	Recovery Rate for XO	1/24	1/day	[27]
$\gamma_3$	Recovery Rate for OO	1/24	1/day	[27]
$m$	Natural human death rate	3.78e-5	1/day	[4]
$\alpha_1$	Disease-induced death rate for XX People	0.0003	Deaths/infection	[19]
$\alpha_2$	Disease-induced death rate for XO People	0.0003	Deaths/infection	[19]
$\alpha_3$	Disease-induced death rate for OO People	0.0003	Deaths/infection	[19]
<b>Mosquito Parameters</b>				
$a$	Biting Rate	0.5	1/day	[24]
$c$	Mosquitoes per Human	2.3 - 4.36	Mosquitoes/Human	[25]
$\phi_1$	Probability a mosquito is infected from XX	0.326	N/A	[1]
$\phi_2$	Probability a mosquito is infected from XO	0.326	N/A	[1]
$\phi_3$	Probability a mosquito is infected from OO	0.30	N/A	[22]
$\delta$	Mosquito Death Rate	0.315	1/day	[3]

Let  $u(t)$  be the total number of uninfected people and  $v(t)$  be the total number of infected people at time  $t$ . Thus, the total population is given by  $N(t) = u(t) + v(t)$ . Furthermore, let the subscript  $i = 1$  describe individuals with the Duffy genotype XX;  $i = 2$  describe individuals with the XO genotype; and  $i = 3$  describe individuals with the OO genotype. For example,  $u_1$  denotes the number of uninfected people whose Duffy genotype is XX, and  $v_2$  denotes the number of infected people with the XO genotype. Lastly, let  $S_m(t)$  be the total number of susceptible mosquitoes and  $I_m(t)$  be the total number of infected mosquitoes, where the total number of mosquitoes in the population is given by  $M(t) \equiv S_m(t) + I_m(t)$  at time  $t$ .

The mathematical model is derived based on the following assumptions:

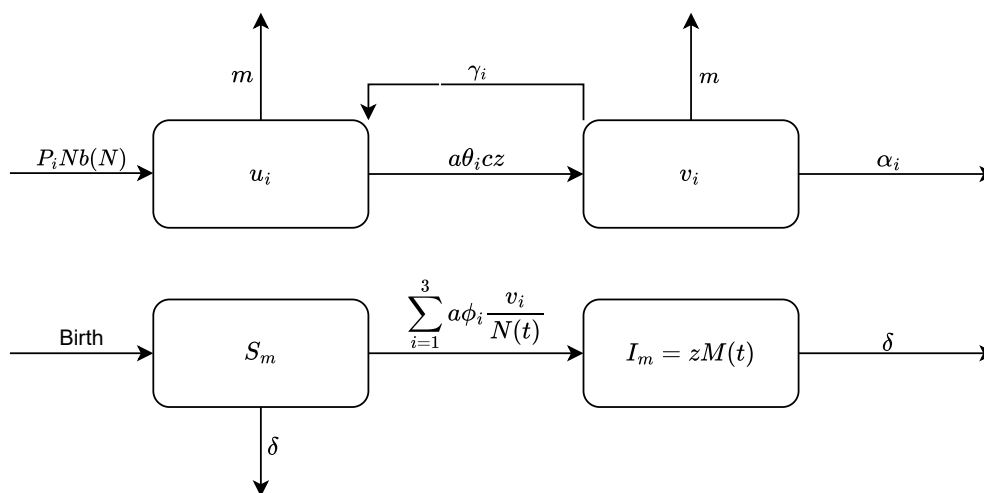
1. New uninfected people of genotype  $i$  are born at the rate  $P_i N b(N)$ , where  $b(N)$  represents the per capita birth rate as a function of the total population, and  $P_i$  is the fraction of births with genotype  $i$ . Arbitrarily, we define  $b(N)$  as  $b(1 - \frac{N}{K})$ , where  $b$  is the base birth rate and  $K$  is the carrying capacity. We assume that people die of non-disease-induced causes at the rate of  $m$ . Infection occurs in humans at the rate of  $a\theta_i c$ , where  $c$  is the mosquito-to-human ratio,  $a$  is the mosquito biting rate and  $\theta_i$  is the probability a human becomes infected after being bitten by an infected mosquito. Lastly, infected individuals recover and enter the uninfected class at the rate  $\gamma_i$ .
2. Individuals enter the infected compartment at a rate  $a\theta_i c$ . We assume individuals either die of natural causes at a rate  $m$ , or die due to malaria at a rate  $\alpha_i$ . Additionally, infected individuals may recover and leave the infected class at a rate  $\gamma_i$ .
3. The mosquitoes are born into the susceptible class at rate  $b_m$ . The term  $S_m \sum_{i=1}^3 \left[ a\phi_i \frac{v_i}{N} \right]$  describes the transition from susceptible to infected mosquitoes, where  $a$  is the mosquito biting rate, and  $\phi_i$  represents the probability that a mosquito becomes infected after biting an infected human of genotype  $i$ . Note that  $\frac{v_i}{N}$  can be thought of as the probability that a mosquito bites a person of genotype  $i$  given that they bite an individual. Lastly, we assume both susceptible and infected mosquitoes die at the rate of  $\delta$ . We then add the susceptible and infected mosquito classes to create the compartment  $\frac{dM}{dt} = b_m - \delta M$ , which we use to monitor the total population of mosquitoes.



Thus we have the system:

$$\begin{aligned}
\frac{du_i}{dt} &= \underbrace{P_i N b(N)}_{\text{Birth}} - \underbrace{m u_i}_{\text{Natural death}} - \underbrace{a \theta_i c z u_i}_{\text{Infection}} + \underbrace{\gamma_i v_i}_{\text{Recovery}} \\
\frac{dv_i}{dt} &= \underbrace{a \theta_i c z u_i}_{\text{New infections}} - \underbrace{m v_i}_{\text{Natural death}} - \underbrace{\gamma_i v_i}_{\text{Recovery}} - \underbrace{\alpha_i v_i}_{\text{Disease-induced death}} \\
\frac{dS_m}{dt} &= b_m - S_m \underbrace{\left( a \phi_1 \frac{v_1}{N(t)} + a \phi_2 \frac{v_2}{N(t)} + a \phi_3 \frac{v_3}{N(t)} \right)}_{\text{Rate of infection by blood meal}} - \underbrace{\delta z}_{\text{Natural death}} \\
\frac{dI_m}{dt} &= S_m \underbrace{\left( a \phi_1 \frac{v_1}{N(t)} + a \phi_2 \frac{v_2}{N(t)} + a \phi_3 \frac{v_3}{N(t)} \right)}_{\text{Rate of infection by blood meal}} - \underbrace{\delta I_m}_{\text{Natural death}}
\end{aligned} \tag{1}$$

The flowchart below describes how people and mosquitoes move between their respective classes.



**Figure 3:** Flowchart of the population dynamics of humans (above) and mosquitoes (below), as described in (3).

## 2.2 Human population genetic model

Our model requires distinguishing between the dynamics of each population's genotype, because malaria dynamics occur on two time scales. Thus, we used Punnett squares to determine the fraction of total births of each genotype, in the Rio Pardo locality of Amazonas, Brazil, denoted by  $P_i$ . We combine the expressed alleles FY\*A and FY\*B into a singular theoretical allele X. Likewise, we use O to represent the silent allele FY<sup>ES</sup> (where ES means "erythrocyte silent"). Let  $w_i$  be the fraction of the total population by genotype. An example calculation of  $P_1$  is below. Note that since  $P_1$  is the fraction of births with genotype XX, we only consider the possible crosses that can produce a child with XX when calculating  $P_1$ .

		Parent 2 ( $w_1$ )			
		X	X	$\Rightarrow$	
		XX	XX		$w_1 \times w_1$
Parent 1 ( $w_1$ )	X	XX	XX		Genotype Probability
	X	XX	XX		XX <span style="border: 1px solid black; padding: 2px;">1</span>
		Parent 2 ( $w_1$ )			
		X	X	$\Rightarrow$	
		XX	XX		$w_1 \times w_2$
Parent 1 ( $w_2$ )	X	XX	XX		Genotype Probability
	O	XO	XO		XX <span style="border: 1px solid black; padding: 2px;">1/2</span>
		XX	XO		XO <span style="border: 1px solid black; padding: 2px;">1/2</span>
		Parent 2 ( $w_2$ )			
		X	O	$\Rightarrow$	
		XX	XO		$w_2 \times w_1$
Parent 1 ( $w_1$ )	X	XX	XO		Genotype Probability
	X	XX	XO		XX <span style="border: 1px solid black; padding: 2px;">1/2</span>
		XX	XO		XO <span style="border: 1px solid black; padding: 2px;">1/2</span>
		Parent 2 ( $w_2$ )			
		X	O	$\Rightarrow$	
		XX	XO		$w_2 \times w_2$
Parent 1 ( $w_2$ )	X	XX	XO		Genotype Probability
	O	XO	OO		XX <span style="border: 1px solid black; padding: 2px;">1/4</span>
		XO	OO		XO <span style="border: 1px solid black; padding: 2px;">1/2</span>
		OO			OO <span style="border: 1px solid black; padding: 2px;">1/4</span>

$$\begin{aligned}
P_1 &= \boxed{1}(w_1w_1) + \boxed{\frac{1}{2}}(w_1w_2) + \boxed{\frac{1}{2}}(w_2w_1) + \boxed{\frac{1}{4}}(w_2w_2) \\
&= (w_1)^2 + (w_1w_2) + \frac{1}{4}(w_2)^2.
\end{aligned}$$

The other formulations of  $P_i$  are determined using a similar method. Thus, the formulas are as follows:

$$\begin{aligned}
P_1 &= (w_1)^2 + \frac{1}{2}(2w_1w_2) + \frac{1}{4}(w_2)^2, \\
P_2 &= \frac{1}{2}(2w_1w_2) + \frac{1}{2}(2w_2w_3) + \frac{1}{2}(w_2)^2 + (2w_1w_3), \\
P_3 &= (w_3)^2 + \frac{1}{4}(w_2)^2 + \frac{1}{2}(2w_2w_3).
\end{aligned}$$

To simplify our analysis, we let  $w_3 = 1 - w_1 - w_2$ . Then, our  $P_i$  expressions are reduced:

$$\begin{aligned}
P_1 &= (w_1)^2 + (w_1w_2) + \frac{1}{4}(w_2)^2, \\
P_2 &= (w_1w_2) + (w_2w_3) + 2(w_1w_3) + \frac{1}{2}(w_2)^2 \\
P_3 &= (w_3)^2 + (w_2w_3) + \frac{1}{4}(w_2)^2
\end{aligned} \tag{2}$$

The assumptions above lead to the following nonlinear ordinary differential equations, which model the disease dynamics of each genotype:

$$\begin{aligned}
\frac{du_i}{dt} &= P_i N b(N) - m u_i - a \theta_i c z u_i + \gamma_i v_i, \\
\frac{dv_i}{dt} &= a \theta_i c z u_i - m v_i - \gamma_i v_i - \alpha_i v_i, \\
\frac{dS_m}{dt} &= \text{Birth} - S_m \left( a \phi_1 \frac{v_1}{N(t)} + a \phi_2 \frac{v_2}{N(t)} + a \phi_3 \frac{v_3}{N(t)} \right) - \delta z, \\
\frac{dI_m}{dt} &= S_m \left( a \phi_1 \frac{v_1}{N(t)} + a \phi_2 \frac{v_2}{N(t)} + a \phi_3 \frac{v_3}{N(t)} \right) - \delta I_m, \quad i \in \{1, 2, 3\}.
\end{aligned} \tag{3}$$

### 2.3 Re-scaled model

The re-scaling of the model is done for the sake of simplification. By re-scaling our original model, we can use the proportion of genotypic populations in Rio Pardo, Brazil, and apply the model to different populations in other regions. After this, perturbation analysis allows us to separate the fast and slow dynamics in our system. As infection happens on a much faster scale than human birth, death and evolution, it is crucial to create this separation so we can determine the fitness of our genotypes, and to study the rate of infection by genotype.

Let  $x_i \equiv \frac{u_i}{N}$  and  $y_i \equiv \frac{v_i}{N}$  to represent the fraction of people with genotype  $i$  who are uninfected and infected, respectively. Thus,  $w_i \equiv x_i + y_i$  is the fraction of the population with genotype  $i$ . We define  $z \equiv \frac{I_m}{M}$  as the fraction of mosquitoes that are infected. Lastly, we let  $\frac{dw_i}{dt} = \frac{dx_i}{dt} + \frac{dy_i}{dt}$  and substitute  $\frac{dx_i}{dt}$  and  $\frac{dy_i}{dt}$ ; similarly,  $\frac{dN}{dt}$  is found using the relation  $N = \sum_{i=1}^3 [u_i + v_i]$ . We then have the following re-scaled model:

$$\begin{aligned}
\frac{dx_i}{dt} &= P_i b(N) - m x_i - a \theta_i c z x_i + \gamma y_i - x_i \frac{\dot{N}}{N}, \\
\frac{dy_i}{dt} &= \beta_{hi} z (w_i - y_i) - (m + \gamma_i + \alpha_i) y_i - y_i \frac{\dot{N}}{N}, \\
\frac{dz}{dt} &= (1 - z) (\beta_{v1} y_1 + \beta_{v2} y_2 + \beta_{v3} y_3) - \delta z, \\
\frac{dw_i}{dt} &= P_i b(N) - m w_i - \alpha_i y_i - w_i \frac{\dot{N}}{N}, \\
\frac{dN}{dt} &= N [b(N) - m - \alpha_1 y_1 - \alpha_2 y_2 - \alpha_3 y_3].
\end{aligned} \tag{4}$$

### 2.4 Slow Dynamical System

We use a fast-slow decomposition to analyze the dynamics of our model. The *slow system* employs the time-scale  $\tau = \epsilon t$ , for some small positive value  $\epsilon$ . As such, large changes in  $t$  result in small changes to  $\tau$ , making analysis of the slow system well suited to investigate dynamics which unfold over long periods of time, such as evolution. Model (4) provides a template for modeling the slow and fast dynamics of malaria transmission in the Rio Pardo. In the slow system, time is parameterized to be measured in  $\tau = \epsilon t$ . From this equation, we

determine that

$$\frac{dt}{d\tau} = 1/\epsilon.$$

For further simplification, we define  $\beta_{hi} = a\theta_i c$  to replace the infection rate of humans after being bitten by mosquitoes, and  $\beta_{vi} = a\phi_i$  to replace the infection rate of mosquitoes after biting humans. By letting  $m = \epsilon\tilde{m}$ ,  $\alpha_i = \epsilon\tilde{\alpha}_i$  and  $b = \epsilon\tilde{b}$  we obtain the following,

$$\frac{dy_i}{d\tau} = \frac{dy_i}{dt} \cdot \frac{dt}{d\tau} \implies \epsilon \frac{dy_i}{d\tau} = \frac{dy_i}{dt}.$$

A similar pattern applies to all derivatives with respect to  $t$ . In this fashion, the original model of the system is simplified to the slow system as follows:

$$\epsilon \frac{dy_i}{d\tau} = \beta_{hi} z(w_i - y_i) - \gamma y_i - \epsilon y_i \left[ \tilde{b}(N) + \epsilon\tilde{\alpha}_i - \epsilon\tilde{\alpha}_1 y_1 - \epsilon\tilde{\alpha}_2 y_2 - \epsilon\tilde{\alpha}_3 y_3 \right], \quad (5a)$$

$$\epsilon \frac{dz}{d\tau} = (1 - z)(a\phi_1 y_1 + a\phi_2 y_2 + a\phi_3 y_3) - \delta z, \quad (5b)$$

$$\frac{dN}{d\tau} = N \left[ \tilde{b}(N) - \tilde{m} - \tilde{\alpha}_1 y_1 - \tilde{\alpha}_2 y_2 - \tilde{\alpha}_3 y_3 \right], \quad (5c)$$

$$\frac{dw_i}{d\tau} = (P_i - w_i)\tilde{b}(N) - \tilde{\alpha}_i y_i + w_i [\tilde{\alpha}_1 y_1 + \tilde{\alpha}_2 y_2 + \tilde{\alpha}_3 y_3]. \quad (5d)$$

### 2.4.1 Manifold of Slow System

This system has a three-dimensional slow manifold:

$$M = \{(y_1, y_2, y_3, z, N, w_1, w_2, w_3) : y_i = y_i^*(w_1, w_2, w_3, N), i = 1, 2, 3, z = z^*(w_1, w_2, w_3, N)\}.$$

The functions  $y_i^*, i = 1, 2, 3$  and  $z^*$  are given in (21) and (22). The slow dynamics on  $M$  are described by the following equations:

$$\begin{aligned} \frac{dN}{d\tau} &= N \left[ \tilde{b}(N) - \tilde{m} - \tilde{\alpha}_1 y_1^* - \tilde{\alpha}_2 y_2^* - \tilde{\alpha}_3 y_3^* \right], \\ \frac{dw_i}{d\tau} &= (P_i - w_i)\tilde{b}(N) - \tilde{\alpha}_i y_i^* + w_i [\tilde{\alpha}_1 y_1^* + \tilde{\alpha}_2 y_2^* + \tilde{\alpha}_3 y_3^*]. \end{aligned} \quad (6)$$

### 2.4.2 Determining Fitness of the X-gene

The fitness of the X gene, that being the presence of the  $Fy^a$  and/or  $Fy^b$  antigens in reticulocytes, determines the capacity of the Duffy-positive gene to be passed down to future generations. By calculating the fitness of the X gene, we determine the influence of selection on the prevalence of the Duffy-negative trait in our region of study. We denote the frequency of the X-gene by  $p = w_1 + \frac{1}{2}w_2$ , where  $w_1$  is the frequency of XX individuals and  $w_2$  is the frequency of XO individuals. The invasion ability of the X-gene is described by

$$\left. \frac{1}{p} \frac{dp}{d\tau} \right|_{p=0}. \quad (7)$$

We will use this quantity to define the fitness of the X-gene,  $\mathcal{F}$ . We assume all genotypes have the same death rate,  $m$ . We define the following out of convenience:

$$T_{hi} = \frac{\beta_{hi}}{\gamma_i}. \quad (8)$$

Using equations from the slow system dynamics for  $w_i$  we obtain

$$\begin{aligned} \frac{dp}{d\tau} &= \left( P_1 + \frac{P_2}{2} - p \right) \tilde{b}(N) - \frac{\tilde{\alpha}_2}{2} y_2^* - \tilde{\alpha}_1 y_1^* + p \left( \sum_{i=1}^3 \tilde{\alpha}_i y_i^* \right), \\ &= a + b + c + d. \end{aligned} \quad (9)$$

We have that  $0 \leq w_1, w_2 \leq q$ . So, when  $p = 0$ , we must have that  $w_1 = w_2 = 0$  and subsequently  $w_3 = 1$ . Substituting the expressions for  $P_1$  and  $P_2$  in (2) in (9) we obtain:

$$\frac{1}{p} a|_{p=0} = 0.$$

Assume  $\tilde{\alpha}_1 = \tilde{\alpha}_2 = \tilde{\alpha}$  and  $T_{h_1} = T_{h_2} = T_h$ . Note also that  $z^*|_{p=0} = z^*|_{w_1=w_2=0} = \frac{\mathcal{R}_{03} - 1}{\mathcal{R}_{03} + T_{h_3}}$ . Then,

$$\begin{aligned} \frac{1}{p} (b + c)|_{p=0} &= \frac{1}{p} \left( -\frac{\tilde{\alpha}_2}{2} y_2^* - \tilde{\alpha}_1 y_1^* \right) = \frac{1}{p} \left( -\frac{\tilde{\alpha}}{2} \frac{w_2 T_{h_2} z^*}{1 + T_{h_2} z^*} - \tilde{\alpha} \frac{w_1 T_{h_1} z^*}{1 + T_{h_1} z^*} \right), \\ &= \frac{-\tilde{\alpha} T_h z^*}{1 + T_h z^*} = \frac{-\tilde{\alpha} T_h (\mathcal{R}_{03} - 1)}{\mathcal{R}_{03} (1 + T_h) - T_h + T_{h_3}}. \end{aligned} \quad (10)$$

Further, given that  $y_1|_{p=0} = 0$ ,  $y_2|_{p=0} = 0$  and

$$y_3|_{p=0} = \frac{z^* T_{h_3}}{z^* T_{h_3} + 1},$$

we have that

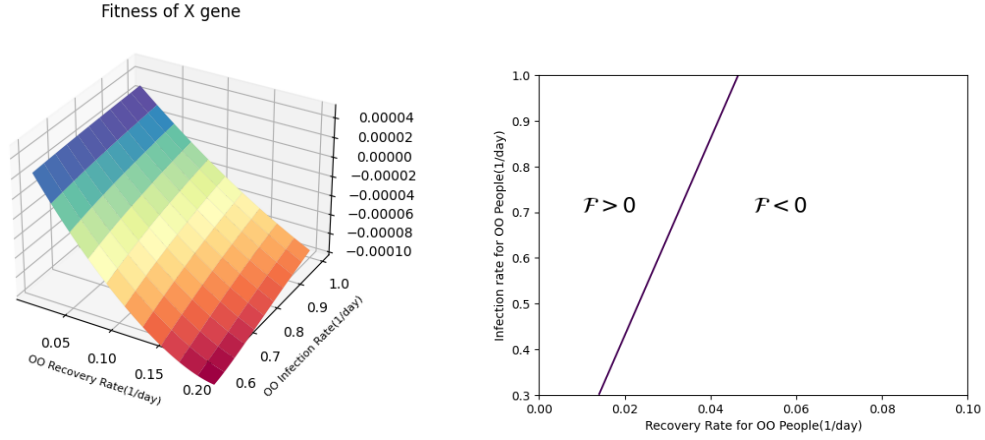
$$\frac{1}{p}|_{p=0} = \frac{\tilde{\alpha}_3 z^* T_{h_3}}{z^* T_{h_3} + 1} = \frac{\tilde{\alpha}_3 T_{h_3} (\mathcal{R}_{03} - 1)}{\mathcal{R}_{03} (T_{h_3} + 1)}.$$

Combining the terms yields

$$\mathcal{F} = -\tilde{\alpha} W_1 + \tilde{\alpha}_2 W_2,$$

$$\text{where } W_1 = \frac{T_h (\mathcal{R}_{03} - 1)}{\mathcal{R}_{03} (1 + T_h) - T_h + T_{h_3}} \text{ and } W_2 = \frac{T_{h_3} (\mathcal{R}_{03} - 1)}{\mathcal{R}_{03} (T_{h_3} + 1)}.$$

We plot the fitness of the X-gene utilizing the parameters obtained from the literature in Fig. 4. Notice that as the infection rate for *OO* people decreases, the fitness or the invasion ability of the X gene decreases. Also, notice that as the recovery rate for *OO* people increases, the fitness of the X gene decreases meaning the X gene invades less.



**Figure 4:** The left plot shows the fitness of the X gene plotted with respect to  $\gamma_3$  (recovery rate for Duffy-negative people) and  $\beta_{h3}$  (infection rate for Duffy-negative people). The right plot is the contour plot that shows the boundary where the fitness is greater or less than zero or rather when the X-gene invades.

## 2.5 Fast Dynamical System

The *fast system*, which describes short-term disease dynamics, uses  $t$  as the time variable applied to disease transmission. Whereas the slow system operates in a matter of decades depending on the value of  $\epsilon$ , the fast system operates on a more limited scale in a matter of days. This neglects birth rates and death rates, and instead focuses on *P. vivax* transmission and recovery. Let  $\epsilon > 0$  be small, and let  $m = \epsilon\tilde{m}$ ,  $\alpha_i = \epsilon\tilde{\alpha}_i$  and  $b = \epsilon\tilde{b}$ . Substituting into system (4) yields:

$$\begin{aligned}
\frac{dy_i}{dt} &= \beta_{hi}z(w_i - y_i) - \epsilon\tilde{m}y_i - \gamma_i y_i - \epsilon\tilde{\alpha}_i y_i - \frac{y_i}{N} \frac{dN}{dt}, \\
\frac{dz}{dt} &= (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2 + \beta_{v3}y_3) - \delta z, \\
\frac{dw_i}{dt} &= P_i \left[ \epsilon\tilde{b} \left( 1 - \frac{N}{K} \right) \right] - \epsilon\tilde{m}w_i - \epsilon\tilde{\alpha}_i y_i - \frac{w_i}{N} \frac{dN}{dt}, \\
\frac{dN}{dt} &= N \left[ \epsilon\tilde{b} \left( 1 - \frac{N}{K} \right) - \epsilon\tilde{m} - \epsilon\tilde{\alpha}_1 y_1 - \epsilon\tilde{\alpha}_2 y_2 - \epsilon\tilde{\alpha}_3 y_3 \right].
\end{aligned} \tag{11}$$

In the fast system, the parameter  $\epsilon$  approaches zero so that the birth and death rate are negated. Because  $\lim_{\epsilon \rightarrow 0} \left( \frac{\dot{N}}{N} \right) = 0$ , then as  $\epsilon \rightarrow 0$ , the fast system then becomes

$$\frac{dy_i}{dt} = \beta_{hi}z(w_i - y_i) - \gamma_i y_i, \quad (12a)$$

$$\frac{dz}{dt} = (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2 + \beta_{v3}y_3) - \delta z, \quad (12b)$$

$$\frac{dw_i}{dt} = 0, \quad (12c)$$

$$\frac{dN}{dt} = 0. \quad (12d)$$

## 2.6 The Next Generation Matrix and the Basic Reproduction Number of the Fast System

The next generation matrix is used to calculate the *basic reproduction number*, which predicts the number of *P. vivax* cases caused by one preceding incidence. We use the remaining equations in the fast system to form the initial matrices

$$\mathcal{F} = \begin{bmatrix} \beta_{h1}z(w_1 - y_1) \\ \beta_{h2}z(w_2 - y_2) \\ \beta_{h3}z(w_3 - y_3) \\ (1 - z)(\beta_{v1}y_1 + \beta_{v2}y_2 + \beta_{v3}y_3) \end{bmatrix}^T \quad \text{and} \quad \mathcal{V} = \begin{bmatrix} \gamma_1 y_1 \\ \gamma_2 y_2 \\ \gamma_3 y_3 \\ \delta z \end{bmatrix}^T. \quad (13)$$

Taking the Jacobian of  $\mathcal{F}$  and  $\mathcal{V}$ , respectively, yields:

$$F := J[\mathcal{F}] = \begin{bmatrix} -\beta_{h1}z & 0 & 0 & \beta_{h1}(w_1 - y_1) \\ 0 & -\beta_{h2}z & 0 & \beta_{h2}(w_2 - y_2) \\ 0 & 0 & -\beta_{h3}z & \beta_{h3}(w_3 - y_3) \\ -\beta_{v1}(z - 1) & -\beta_{v2}(z - 1) & -\beta_{v3}(z - 1) & -\beta_{v1}y_1 - \beta_{v2}y_2 - \beta_{v3}y_3 \end{bmatrix},$$

$$V := J[\mathcal{V}] = \begin{bmatrix} \gamma_1 & 0 & 0 & 0 \\ 0 & \gamma_2 & 0 & 0 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & 0 & 0 & \delta \end{bmatrix}.$$

Thus, at the disease free equilibrium  $E_0$  where  $(y_1, y_2, y_3, z) = (0, 0, 0, 0)$ , the next generation matrix is:

$$FV^{-1} = \begin{bmatrix} 0 & 0 & 0 & \beta_{h1}w_1 \\ 0 & 0 & 0 & \beta_{h2}w_2 \\ 0 & 0 & 0 & \beta_{h3}w_3 \\ \beta_{v1} & \beta_{v2} & \beta_{v3} & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 & 0 & 0 & 0 \\ 0 & \gamma_2 & 0 & 0 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & 0 & 0 & \delta \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 0 & (\beta_{h1}w_1)/\delta \\ 0 & 0 & 0 & (\beta_{h2}w_2)/\delta \\ 0 & 0 & 0 & (\beta_{h3}w_3)/\delta \\ \beta_{v1}/\gamma_1 & \beta_{v2}/\gamma_2 & \beta_{v3}/\gamma_3 & 0 \end{bmatrix}.$$

The spectral radius of the next generation matrix is our basic reproduction number, denoted by  $\mathcal{R}_0$ .

$$\mathcal{R}_0 = a \sqrt{\frac{c(\gamma_2\gamma_3\theta_1\phi_1w_1 + \gamma_1\gamma_3\theta_2\phi_2w_2 + \gamma_1\gamma_2\theta_3\phi_3w_3)}{\delta\gamma_1\gamma_2\gamma_3}}, \quad (14)$$

$$(\mathcal{R}_0)^2 = \frac{a^2c\gamma_2\gamma_3\theta_1\phi_1w_1}{\delta\gamma_1\gamma_2\gamma_3} + \frac{a^2c\gamma_1\gamma_3\theta_2\phi_2w_2}{\delta\gamma_1\gamma_2\gamma_3} + \frac{a^2c\gamma_1\gamma_2\theta_3\phi_3w_3}{\delta\gamma_1\gamma_2\gamma_3} \quad (15)$$

$$= w_1 \frac{(ac\theta_1)(a\phi_1)}{\delta\gamma_1} + w_2 \frac{(ac\theta_2)(a\phi_2)}{\delta\gamma_2} + w_3 \frac{(ac\theta_3)(a\phi_3)}{\gamma_3} \quad (16)$$

$$= w_1\mathcal{R}_{01} + w_2\mathcal{R}_{02} + w_3\mathcal{R}_{03}, \quad (17)$$

where  $\mathcal{R}_{0i} = \frac{\beta_{hi}\beta_{vi}}{\delta\gamma_i}$ . For convenience, we define

$$\mathcal{R}_0 = w_1\mathcal{R}_{01} + w_2\mathcal{R}_{02} + w_3\mathcal{R}_{03},$$

which doesn't affect the validity of the inequalities  $\mathcal{R}_0 > 1$  and  $\mathcal{R}_0 < 1$ .

## 2.7 Conditions for Stability of Disease-Free Equilibrium

**Theorem 1.** *The disease free equilibrium  $E_0$  is locally asymptotically stable if  $\mathcal{R}_0 < 1$  and unstable if  $\mathcal{R}_0 > 1$ .*

*Proof.* To determine the stability of  $E_0$ , we first find the Jacobian of the *fast system* and evaluate it at  $E_0$ .

$$J = \begin{bmatrix} -(\beta_{h1}z + \gamma_1) & 0 & 0 & \beta_{h1}(w_1 - y_1) \\ 0 & -(\beta_{h2}z + \gamma_2) & 0 & \beta_{h2}(w_2 - y_2) \\ 0 & 0 & -(\beta_{h3}z + \gamma_3) & \beta_{h3}(w_3 - y_3) \\ \beta_{v1}(1 - z) & \beta_{v2}(1 - z) & \beta_{v3}(1 - z) & -(\beta_{v1}y_1 + \beta_{v2}y_2 + \beta_{v3}y_3 + \delta) \end{bmatrix} \quad (18)$$

$$J(E_0) = \begin{bmatrix} -\gamma_1 & 0 & 0 & \beta_{h1}w_1 \\ 0 & -\gamma_2 & 0 & \beta_{h2}w_2 \\ 0 & 0 & -\gamma_3 & \beta_{h3}w_3 \\ \beta_{v1} & \beta_{v2} & \beta_{v3} & -\delta \end{bmatrix} \quad (19)$$

Note that  $J(E_0)$  can be expressed as  $M - D$  where  $M$  and  $D$  are both non-negative matrices and  $D$  is additionally a diagonal matrix.

$$J(E_0) = M - D \quad \text{where} \quad M = \begin{bmatrix} 0 & 0 & 0 & \beta_{h1}w_1 \\ 0 & 0 & 0 & \beta_{h2}w_2 \\ 0 & 0 & 0 & \beta_{h3}w_3 \\ \beta_{v1} & \beta_{v2} & \beta_{v3} & 0 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} \gamma_1 & 0 & 0 & 0 \\ 0 & \gamma_2 & 0 & 0 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & 0 & 0 & \delta \end{bmatrix}.$$

From here, we use the following theorem.

**Theorem 2.** *Let  $M$  and  $D$  be matrices such that every entry is nonnegative and  $D$  is a diagonal matrix and let  $H = M - D$ . If the dominant eigenvalues of  $MD^{-1}$  are less than 1, then the real part of the eigenvalues of  $H$  are less than 0. [6]*



It can be shown that the matrix  $MD^{-1}$  has two eigenvalues of 0 and the remaining two are

$$\lambda_1, \lambda_2 = \pm \sqrt{w_1 \frac{\beta_{h1}\beta_{v1}}{\delta\gamma_3} + w_2 \frac{\beta_{h2}\beta_{v2}}{\delta\gamma_2} + w_3 \frac{\beta_{h3}\beta_{v3}}{\delta\gamma_3}}.$$

Making the substitution  $\mathcal{R}_0 = w_1 \frac{\beta_{h1}\beta_{v1}}{\delta\gamma_1} + w_2 \frac{\beta_{h2}\beta_{v2}}{\delta\gamma_2} + w_3 \frac{\beta_{h3}\beta_{v3}}{\delta\gamma_3}$  yields

$$\lambda_1, \lambda_2 = \pm \sqrt{\mathcal{R}_0}.$$

Note that  $\lambda_2 = -\sqrt{\mathcal{R}_0}$  is always less than 1 since  $\mathcal{R}_0$  is always greater than or equal to 0. Additionally,  $\lambda_1 < 1$  when  $\mathcal{R}_0 < 1$  and  $\lambda_1 > 1$  when  $\mathcal{R}_0 > 1$ . By [Theorem 2](#), the real part of all the eigenvalues of  $J(E_0)$  are negative when  $\mathcal{R} < 1$ , meaning the disease-free equilibrium is locally asymptotically stable when  $\mathcal{R}_0 < 1$ . Likewise,  $E_0$  is unstable when  $\mathcal{R}_0 > 1$ .  $\square$

## 2.8 Location of Nontrivial Equilibrium

The nontrivial equilibrium  $E^* = (y_1^*, y_2^*, y_3^*, z^*)$  is a solution to the following system.

$$\begin{aligned} 0 &= \beta_{h1}z^*(w_1 - y_1^*) - \gamma_1y_1^* \\ 0 &= \beta_{h2}z^*(w_2 - y_2^*) - \gamma_2y_2^* \\ 0 &= \beta_{h3}z^*(w_3 - y_3^*) - \gamma_3y_3^* \\ 0 &= (1 - z^*)(\beta_{v1}y_1^* + \beta_{v2}y_2^* + \beta_{v3}y_3^*) - \delta z^* \end{aligned} \tag{20}$$

For convenience, we define  $T_{hi} = \frac{\beta_{hi}}{\gamma_i}$ . Substituting and simplifying yields the following expression for  $y_i^*$ :

$$y_i^* = \frac{\beta_{hi}w_i z^*}{\gamma_i + \beta_{hi}z^*} = \frac{T_{hi}w_i z^*}{T_{hi}z^* + 1}. \tag{21}$$

To find the solutions for  $z^*$ , we substitute out  $y_1^*, y_2^*$ , and  $y_3^*$  in [\(20\)](#) and divide out the trivial solution of  $z^* = 0$ . Additionally, we let  $w = w_1 + w_2$  and assume  $T_{h1} = T_{h2}$  and  $\mathcal{R}_{01} = \mathcal{R}_{02}$ . This yields the quadratic:

$$0 = k_2(z^*)^2 + k_1(z^*) + k_0,$$

where

$$\begin{aligned} k_2 &= T_{h1}T_{h3} + \mathcal{R}_{03}T_{h1}(1 - w) + \mathcal{R}_{01}T_{h3}w, \\ k_1 &= T_{h3} + T_{h1} + \mathcal{R}_{03}(1 - T_{h1})(1 - w) + \mathcal{R}_{01}w(1 - T_{h3}), \\ k_0 &= 1 - \mathcal{R}_{03}(1 - w) - \mathcal{R}_{01}w. \end{aligned} \tag{22}$$

Note that  $k_0$  can be rewritten as  $k_3 = 1 - \mathcal{R}_0$ , and let  $h(z) = k_2(z)^2 + k_1(z) + k_0$ . Note that  $z^*$  is by definition a solution to the equation  $h(z) = 0$ .

**Lemma 1.** *If  $0 < \mathcal{R}_0 < 1$  then  $k_1 > 0$ .*

*Proof.* Let  $0 < \mathcal{R}_0 < 1$  and define  $T_0 = \max(T_{h1}, T_{h3})$ . Based on these two statements,

$$T_{h3} + T_{h1} - T_0 \mathcal{R}_0 \geq 0. \quad (23)$$

Additionally, since  $T_0 \geq T_{h1}$  and  $T_0 \geq T_{h3}$ , we know

$$\mathcal{R}_{03}T_0(1-w) + \mathcal{R}_{01}T_0w \geq \mathcal{R}_{03}T_{h1}(1-w) + \mathcal{R}_{01}T_{h3}w. \quad (24)$$

From our definition of  $k_1$  (22),

$$\begin{aligned} k_1 &= T_{h3} + T_{h1} + \mathcal{R}_{03}(1-T_{h1})(1-w) + \mathcal{R}_{01}w(1-T_{h3}), \\ &= T_{h3} + T_{h1} + \mathcal{R}_{03}(1-w) + \mathcal{R}_{01}w - \mathcal{R}_{03}T_{h1}(1-w) - \mathcal{R}_{01}T_{h3}w, \\ &= T_{h3} + T_{h1} + \mathcal{R}_0 - \mathcal{R}_{03}T_{h1}(1-w) - \mathcal{R}_{01}T_{h3}w. \end{aligned}$$

From here, using relation (24) yields

$$\begin{aligned} k_1 &\geq T_{h3} + T_{h1} + \mathcal{R}_0 - \mathcal{R}_{03}T_0(1-w) - \mathcal{R}_{01}T_0w, \\ k_1 &\geq T_{h3} + T_{h1} + \mathcal{R}_0 - T_0 \mathcal{R}_0, \end{aligned}$$

and relation (23) implies

$$T_{h3} + T_{h1} - T_0 \mathcal{R}_0 + \mathcal{R}_0 \geq \mathcal{R}_0.$$

Combining the two results above yields  $k_1 \geq \mathcal{R}_0$ . Therefore, if  $0 < \mathcal{R}_0 < 1$ , then  $k_1 > 0$ .  $\square$

**Theorem 3.** *Assuming all parameters are biologically possible, the fixed point  $E_0$  in the **fast system** (12a) is biologically impossible when  $\mathcal{R}_0 < 1$  and biologically possible when  $\mathcal{R}_0 > 1$ .*

*Proof of Theorem 3.* Let all parameters be biologically possible and  $0 < \mathcal{R}_0 < 1$ . From Equation 22 it follows that  $k_2 > 0$  and  $k_0 > 0$ , due to the respective conditions above. Additionally, Lemma 1 implies,  $k_1 > 0$ .

Let  $z_1^*, z_2^*$  be the two solutions to the quadratic  $h(z) = 0$ . Then, by Vieta's formulas, the solutions to  $h(z) = 0$  satisfy

$$z_1^* \cdot z_2^* = \frac{k_0}{k_2} \qquad z_1^* + z_2^* = -\frac{k_1}{k_2}$$

and using the conditions  $k_0, k_1, k_2 > 0$ , they simplify to

$$z_1^* \cdot z_2^* > 0 \qquad z_1^* + z_2^* < 0$$

The inequality on the left rules out the possibility that exactly one of the roots of  $h(z)$  is positive and the inequality on the left implies that the roots cannot both be positive. Therefore, neither  $z_1^*$  or  $z_2^*$  are positive. As such, when  $0 < \mathcal{R}_0 < 1$ , the nontrivial equilibrium is not biologically possible since  $E^*$  would require a negative or complex fraction of mosquitoes to be infected, neither of which are possible in the real world.

To prove the second part of Theorem 3, let all parameters be biologically possible and let  $\mathcal{R}_0 > 1$ . Similar to earlier, these two conditions imply  $k_2 > 0$ ,  $k_1 < 0$  and  $k_0 < 0$ . Similar to before, applying these conditions to Vieta's Formulas yields

$$z_1^* \cdot z_2^* = \frac{k_0}{k_2} \implies z_1^* \cdot z_2^* < 0.$$

Since  $k_1, k_2, k_3$  are all real numbers,  $z_1^*, z_2^*$  must also, so the inequality above must imply that either  $z_1^*$  or  $z_2^*$  is positive, but not both. Let  $z^*$  be this unique positive solution to  $h(z) = 0$ .

Next, we wish to show that  $0 < z^* < 1$ . Since  $\mathcal{R}_0 > 1$ , we know  $k_0 < 0$  by [Lemma 1](#). Additionally,

$$\begin{aligned} h(0) = k_0 & \implies h(0) < 0 \\ h(1) = 1 + T_{h3} + T_{h1} + T_{h3}T_{h1} & \implies h(1) > 0 \end{aligned}$$

since  $T_{h1}$  and  $T_{h3}$  are greater than 0 when all parameters are biologically possible. Because  $h(z)$  is a continuous function with  $h(0) < 0$  and  $h(1) > 0$ , the Intermediate Value Theorem implies that there exists a solution to  $h(z) = 0$  between  $z = 0$  and  $z = 1$ . Since  $z^*$  is the only positive solution to  $h(z) = 0$ , we conclude that

$$0 < z^* < 1. \tag{25}$$

Recall that  $z^*$  represents the fraction of mosquitoes at the equilibrium point  $E^*$ . As such,  $z^*$  is biologically possible when  $\mathcal{R}_0 > 1$  since  $0 < z^* < 1$ .

Lastly, we wish to show that  $y_1^*, y_2^*, y_3^*$  are also between 0 and 1 when all parameters are biologically possible. Recall that [Equation 21](#) tells shows

$$y_i^* = \frac{T_{hi}w_i z^*}{T_{hi}z^* + 1}$$

Since  $w_i$  represents the fraction of the population with genotype  $i$ ,  $w_i$  must satisfy  $0 \leq w_i \leq 1$  in the real world. As such, the denominator for  $y_i^*$  is necessarily greater than the numerator. Therefore, at the nontrivial equilibrium,  $0 \leq y_i^* < 1$ , meaning  $y_1^*, y_2^*$ , and  $y_3^*$  are all biologically possible. As such, we conclude that  $E^*$  is biologically possible when all parameters are biologically possible and  $\mathcal{R}_0$  is greater than 1.  $\square$

## 2.9 Conditions for Stability of Endemic Equilibrium

In the previous section, we proved that the endemic equilibrium  $E^*$  exists in biologically possible space only when  $\mathcal{R}_0 > 1$ . We now present the following theorem on the stability of  $E^*$ .

**Theorem 4.** *If  $\mathcal{R}_0 > 1$  and all parameters are biologically possible (i.e.  $E^*$  is biologically possible), then the fixed point  $E^*$  is stable.*

*Proof.* Taking the Jacobian of the *fast system* [\(18\)](#) and evaluating it at  $E^*$  yields

$$J(E^*) = \begin{bmatrix} -(\beta_{h1}z^* + \gamma_1) & 0 & 0 & \beta_{h1}(w_1 - y_1^*) \\ 0 & -(\beta_{h2}z^* + \gamma_2) & 0 & \beta_{h2}(w_2 - y_2^*) \\ 0 & 0 & -(\beta_{h3}z^* + \gamma_3) & \beta_{h3}(w_3 - y_3^*) \\ \beta_{v1}(1 - z^*) & \beta_{v2}(1 - z^*) & \beta_{v3}(1 - z^*) & -(\beta_{v1}y_1^* + \beta_{v2}y_2^* + \beta_{v3}y_3^* + \delta) \end{bmatrix}.$$

Note that  $J(E^*) = M - D$  where  $M$  and  $D$  are both non-negative matrices and  $D$  is a diagonal matrix.

$$M = \begin{bmatrix} 0 & 0 & 0 & \beta_{h1}(w_1 - y_1^*) \\ 0 & 0 & 0 & \beta_{h2}(w_2 - y_2^*) \\ 0 & 0 & 0 & \beta_{h3}(w_3 - y_3^*) \\ \beta_{v1}(1 - z^*) & \beta_{v2}(1 - z^*) & \beta_{v3}(1 - z^*) & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} \beta_{h1}z^* + \gamma_1 & 0 & 0 & 0 \\ 0 & \beta_{h2}z^* + \gamma_2 & 0 & 0 \\ 0 & 0 & \beta_{h3}z^* + \gamma_3 & 0 \\ 0 & 0 & 0 & \beta_{v1}y_1^* + \beta_{v2}y_2^* + \beta_{v3}y_3^* + \delta \end{bmatrix}.$$

Recall that [Theorem 2](#) states that the real part of the eigenvalues of  $J(E^*)$  are negative if the dominant eigenvalue of  $MD^{-1}$  is less than 1.

It can be shown that  $MD^{-1}$  has a double eigenvalue of 0 and the other two are

$$\lambda_+, \lambda_- = \pm \sqrt{A_1 \left( \frac{\beta_{v1}(1 - z^*)}{\beta_{h1}z^* + \gamma_1} \right) + A_2 \left( \frac{\beta_{v2}(1 - z^*)}{\beta_{h2}z^* + \gamma_2} \right) + A_3 \left( \frac{\beta_{v3}(1 - z^*)}{\beta_{h3}z^* + \gamma_3} \right)} \quad (26)$$

where

$$A_j = \frac{\beta_{hj}(w_j - y_j^*)}{\sum_{i=1}^3 [\beta_{vi}y_i^*] + \delta} \quad (27)$$

Note that  $\lambda_-$  is always less than 1 provided that all parameters are within biologically possible bounds. Next, we wish to show that  $\lambda_+$  is also less than 1.

From [Equation 20](#), recall that

$$0 = (1 - z^*)(\beta_{v1}y_1^* + \beta_{v2}y_2^* + \beta_{v3}y_3^*) - \delta z^*.$$

Manipulating this equation to isolate  $z^*$  yields

$$z^* = \frac{\sum_{i=1}^3 [\beta_{vi}y_i^*]}{\sum_{i=1}^3 [\beta_{vi}y_i^*] - \delta}.$$

Using the equations of form  $\gamma_j y_j^* = \beta_{hj} z^* (w_j - y_j^*)$  from [Equation 20](#), we obtain the following expression for  $z^*$ .

$$z^* = \frac{\gamma_j y_j^*}{\beta_{hj} z^* (w_j - y_j^*)}, \quad j \in \{1, 2, 3\}$$

Setting these two expressions for  $z^*$  equal and multiplying both sides by the fraction  $\frac{\beta_{hj}(w_j - y_j^*)}{\sum_{i=1}^3 [\beta_{vi}y_i^*]}$  yields

$$\frac{\beta_{hj}(w_j - y_j^*)}{\sum_{i=1}^3 [\beta_{vi}y_i^*] - \delta} = \frac{\gamma_j y_j^*}{\sum_{i=1}^3 [\beta_{vi}y_i^*]}, \quad j \in \{1, 2, 3\}$$

Note that the left side of this expression is equivalent to our definition of  $A_j$ , Equation(27). As such,

$$A_j = \frac{\gamma_j y_j^*}{\sum_{i=1}^3 [\beta_{vi} y_i^*]}, \quad j \in \{1, 2, 3\}.$$

Furthermore,

$$\begin{aligned} A_j \left[ \frac{\beta_{vj}(1-z^*)}{\beta_{hj}z^* + \gamma_j} \right] &= \frac{\gamma_j y_j^*}{\sum_{i=1}^3 [\beta_{vi} y_i^*]} \left[ \frac{\beta_{vj}(1-z^*)}{\beta_{hj}z^* + \gamma_j} \right], \quad j \in \{1, 2, 3\} \\ &= \frac{\beta_{vj} y_j^* (1-z^*)}{\sum_{i=1}^3 [\beta_{vi} y_i^*]} \left[ \frac{\gamma_j}{\beta_{hj}z^* + \gamma_j} \right]. \end{aligned}$$

From here, since all parameters are assumed to be biologically possible,  $\beta_{hj} + \gamma_j > \gamma_j$ . Using this inequality, we find

$$A_j \left[ \frac{\beta_{vj}(1-z^*)}{\beta_{hj}z^* + \gamma_j} \right] < \frac{\beta_{vj} y_j^* (1-z^*)}{\sum_{i=1}^3 [\beta_{vi} y_i^*]}, \quad j \in \{1, 2, 3\}. \quad (28)$$

Next, we sum Equation (28) over  $j = 1, 2, 3$  to obtain the following. Note that the left side is equivalent to  $(\lambda_+)^2$  by (26).

$$\begin{aligned} \sum_{j=1}^3 \left[ A_j \left[ \frac{\beta_{vj}(1-z^*)}{\beta_{hj}z^* + \gamma_j} \right] \right] &< \sum_{j=1}^3 \left[ \frac{\beta_{vj} y_j^* (1-z^*)}{\sum_{i=1}^3 [\beta_{vi} y_i^*]} \right] \quad j \in \{1, 2, 3\} \\ (\lambda_+)^2 &< 1 - z^* \end{aligned}$$

Recall from Equation (25) that  $0 < Z^* < 1$ . Therefore  $\lambda_+ < 1$ . Since both  $\lambda_-$  and  $\lambda_+$  are less than 1, Theorem 2 applies, meaning that the real part of the eigenvalues of  $J(E^*)$  are negative. As such, we conclude that  $E^*$  is stable if  $\mathcal{R}_0$  is greater than 1 and all parameters are biologically possible.  $\square$

### 3 Sensitivity Analysis

We conduct global sensitivity analysis on our model to determine which parameters are most influential to the model's behavior. We include biting rate seasonality on our global sensitivity analysis of the fast subsystem, as outlined by the inter-compartmental approach from Renardy et al [21].

There are multiple methods to perform global sensitivity analysis. For nonlinear and monotonic relationships, measures that work the best are Spearman rank correlation coefficient (RCC or Spearman's  $\rho$ ), partial rank correlation coefficient (PRCC), and standardized rank regression coefficients (SRRC) [17]. The Sobol method is effective for nonlinear non-monotonic trends, as do the Fourier amplitude sensitivity test (FAST) and its extended version (eFAST). The derivative based one at a time (OAT) method can be used on any continuous system where it isn't computationally difficult to calculate the partial derivatives. Since the relationship in the fast subsystem with seasonality is nonlinear and non-monotonic with computationally possible derivatives we can either utilize the eFAST or the

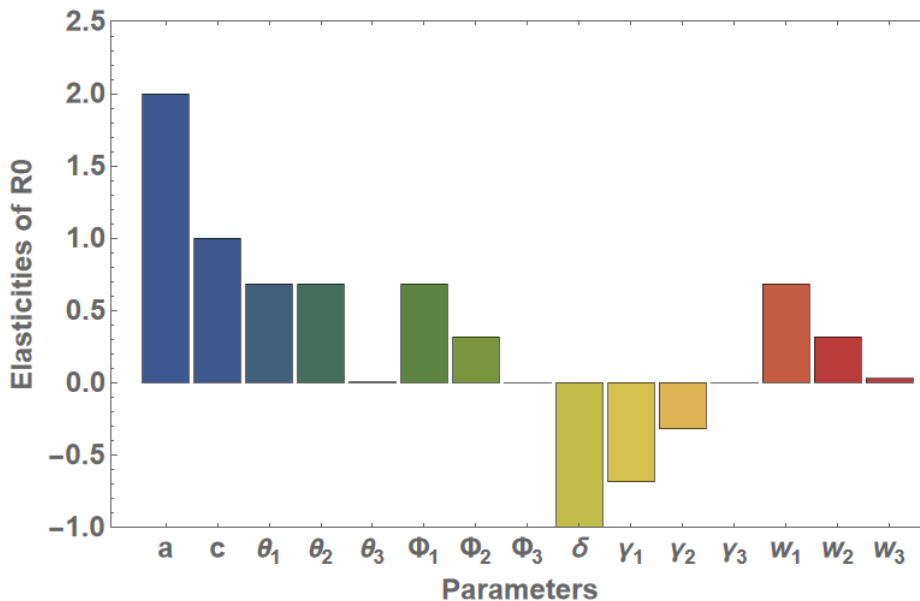
OAT method. We proceed with the OAT method at this time but will conduct eFAST in the future.

### 3.1 Calculating The Elasticity of $\mathcal{R}_0$

We compute the sensitivity indices of the parameters to determine how much each parameter influences  $\mathcal{R}_0$ . The sensitivity index for a parameter  $p$  is calculated as

$$\frac{\partial \mathcal{R}_0}{\partial p} \frac{p}{\mathcal{R}_0}.$$

The sensitivity indices are as shown below. A contour plot of the values of  $\mathcal{R}_0$  depending on the percentages of the population with XX and XO can be seen in [Figure 5](#). Since the biting rate  $a$  and the mosquito death rate  $\delta$  rely on the population of mosquitoes,  $\mathcal{R}_0$  is most drastically influenced by these parameters. The parameters  $\theta_3$ ,  $\phi_3$ , and  $\gamma_3$  all have low elasticities because of the low percentage of the population with absence of the Duffy trait.



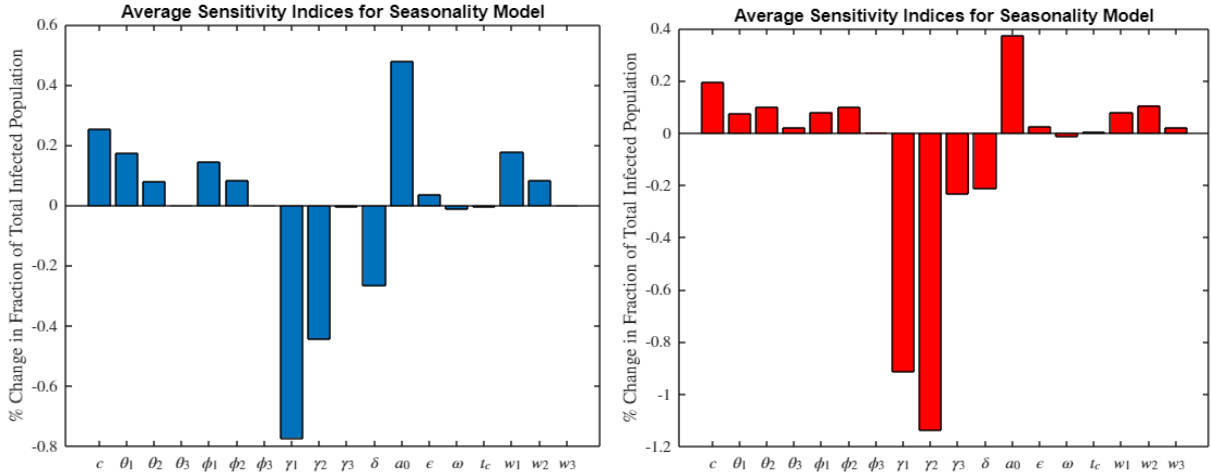
**Figure 5:** Sensitivity indices for  $\mathcal{R}_0$  for each parameter in Rio Pardo

### 3.2 Global Sensitivity on Fast System with Biting Rate Seasonality

The elasticity of  $\mathcal{R}_0$  substantiates that the mosquito biting rate is the most influential parameter to  $\mathcal{R}_0$ . We can then account for the seasonality of biting rate by equating  $a$  to the sinusoidal model determined by Iyaniwura et al. [15]:

$$a = a_0 [1 + \epsilon \cos(2\pi\omega(t - t_c))].$$

We use derivative-based sensitivity analysis and calculate the normalized sensitivity index from Arriola and Hyman [2] for the equation denoting the total change in the fraction of people getting infected. Thus, we are calculating  $\frac{\partial y}{\partial p} \frac{p}{y}$  for each parameter  $p$  in  $y$  where  $y = y_1 + y_2 + y_3$  where the other parameters are the baseline values determined by literature. The average sensitivity indices for the Rio Pardo and Macapa DARC distribution are shown in Fig. 6. Macapa has a larger percentage of OO people than Rio Pardo. Therefore,  $w_3$  has a greater sensitivity index in Macapa but is still smaller than  $w_1$  and  $w_2$  due to the reduced transmission rates for OO people.



**Figure 6:** Mean sensitivity index of each parameter. The left graph uses the initial conditions and DARC distribution in Rio Pardo, Brazil while the right graph uses the values from Macapa, Brazil.

## 4 Parameter Estimation

To estimate the transmission rates of malaria between mosquitoes and humans, we conduct parameter estimation on the fast subsystem using incidence data from Brazil and by assuming that Rio Pardo’s Duffy genotype frequency is representative of the wider Brazilian Amazon.

We begin by splitting up time series of cases in Brazil from August 2020 to May 2023 from Garcia et al. [9] to the number of people in each DARC genotype infected.

### 4.1 Relative Risk of Infection Based on Genotype

Due to the limitations of available literature, we use our known information to infer the infection rates of *P. vivax* malaria, based on the Duffy polymorphisms in Rio Pardo. For our model, we attempt to weight the relative risk of being infected with *P. vivax* malaria ( $n_i$ ) by an individual’s genotype. We presume that if everyone in the population contracts malaria at the same rate, then  $\frac{v_i}{v_1}$  is equal to  $\frac{w_i}{w_1}$ . We used the values of variables provided in the paper by Kano et al. [16] to adjust the risk according to the genotype of each population, resulting in this system of equations:

$$\begin{cases} \frac{v_2}{v_1} = \left( \frac{n_2}{n_1} \right) \left( \frac{w_2}{w_1} \right) \\ \frac{v_3}{v_1} = \left( \frac{n_3}{n_1} \right) \left( \frac{w_3}{w_1} \right) \\ v_1 + v_2 + v_3 = v_{total} \end{cases} \quad (29)$$

where  $n_i$  is the relative risk of infection of *P. vivax* malaria for someone with a genotype  $i$ . By calculating  $\frac{v_2}{v_1}$  and  $\frac{v_3}{v_1}$  and solving the system of equations, we can determine the proportion of the infected population by genotype.

The values of  $n_i$  must be calculated to determine the relative risk of infection for each genotype. As determined by Kano et al. [16], let  $n_{aa}$  be the risk of infection due to the FY\*A/FY\*A genotype;  $n_{bb}$  be the risk of infection due to the FY\*B/FY\*B genotype:

$$\begin{aligned} n_1 &= \left( \frac{FY * A / FY * A}{FY * A / FY * A + FY * B / FY * B + FY * A / FY * B} \right) (n_{aa}) + \\ &\quad \left( \frac{FY * B / FY * B}{FY * A / FY * A + FY * B / FY * B + FY * A / FY * B} \right) (n_{bb}) + \\ &\quad \left( \frac{FY * A / FY * B}{FY * A / FY * A + FY * B / FY * B + FY * A / FY * B} \right) (n_{ab}) \\ &= \left( \frac{160}{160 + 93 + 204} \right) (0.92) + \left( \frac{93}{160 + 93 + 204} \right) (1.00) + \left( \frac{204}{160 + 93 + 204} \right) (1.00) \\ &= 0.9719912 \end{aligned} \quad (30)$$

We define  $n_2$  as the weighted average of the relative risk of infection a heterozygous genotype with a silent allele. Let  $n_{ao}$  be the risk of infection for FY\*A/FY\*O genotype, and  $n_{bo}$  be the risk of the FY\*B/FY\*O genotype.

$$\begin{aligned} n_2 &= \left( \frac{FY * A / FY * O}{FY * A / FY * O + FY * B / FY * O} \right) (n_{ao}) \\ &\quad + \left( \frac{FY * B / FY * O}{FY * A / FY * O + FY * B / FY * O} \right) (n_{bo}), \\ &= \left( \frac{139}{139 + 73} \right) (0.81) + \left( \frac{73}{139 + 73} \right) (1.26), \\ &= 0.9649528. \end{aligned} \quad (31)$$

The relative risk  $\frac{n_2}{n_1}$  of the heterozygous genotype is therefore calculated as follows:

$$\frac{n_2}{n_1} = \frac{0.9649528}{0.9719912} = 0.992758. \quad (32)$$

Thus, from Equation 29 we solve for  $\frac{v_2}{v_1}$  and  $\frac{v_3}{v_1}$ :



$$\begin{aligned}
\frac{v_2}{v_1} &= \left(\frac{n_2}{n_1}\right) \left(\frac{w_2}{w_1}\right) & \frac{v_3}{v_1} &= \left(\frac{n_3}{n_1}\right) \left(\frac{w_3}{w_1}\right) \\
&= (0.9649528) \left(\frac{0.3072}{0.6623}\right) & &= (0.09) \left(\frac{0.0304}{0.6623}\right) \\
&= 0.4475819 & &= 0.00413
\end{aligned}$$

where  $\frac{v_3}{v_1}$  represents the homozygous Duffy-negative genotype.

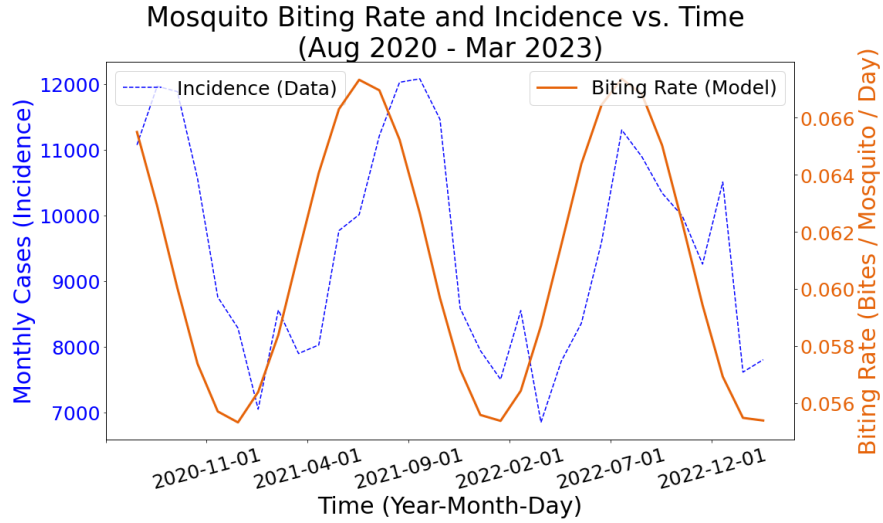
## 4.2 Fitting the Fast System with Seasonality

We begin the parameter estimation by fitting the sinusoidal seasonality of biting rate by fitting the below model from [15] to the incidence cases of Brazil:

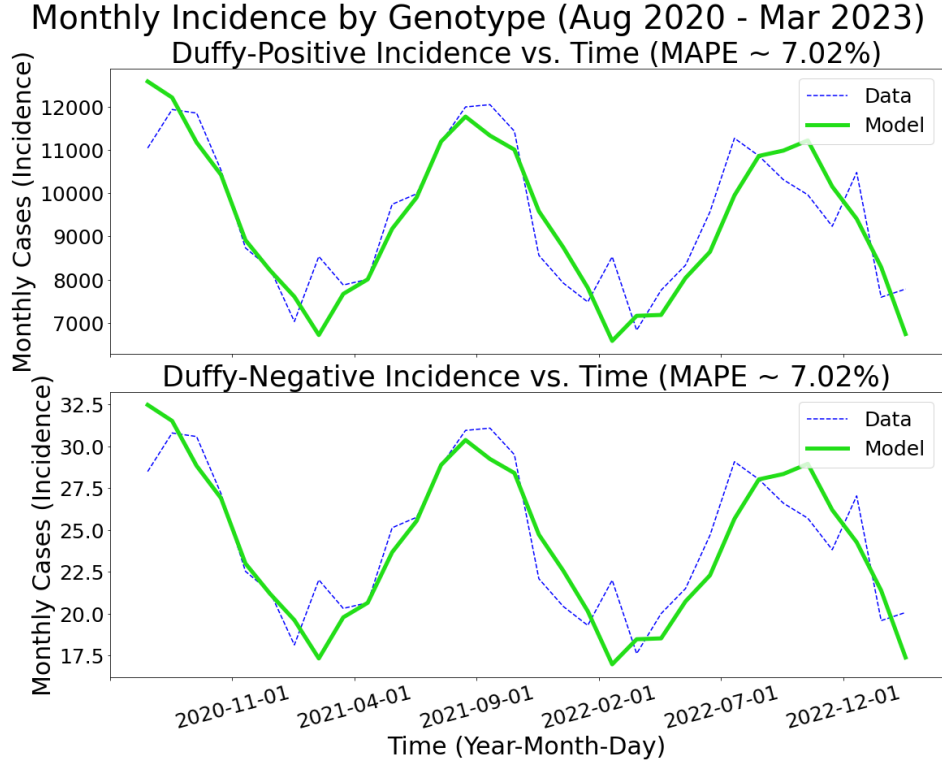
$$a = a_0 [1 + \epsilon \cos(2\pi\omega(t - t_c))].$$

Dual annealing was used to minimize the mean absolute percent error (MAPE) between  $a$  and the incidence of cases. This was used to fit the parameter  $\omega$ . We then continue to estimate every parameter except  $\omega$  with dual annealing where it minimizes the following value:

$$\begin{aligned}
\text{MAPE}(\text{Total incidence}) &= \text{MAPE}(\text{Incidence of XX cases}) + \\
&\text{MAPE}(\text{Incidence of XO cases}) + \text{MAPE}(\text{Incidence of OO cases}).
\end{aligned}$$



**Figure 7:** Modeled mosquito biting rate and *P. vivax* incidence in the Brazilian Amazon



**Figure 8:** Fit of model to the estimated breakdown of *P. vivax* by Duffy positivity in the Brazilian Amazon

In Fig. 7, the biting rate lags the incidence of *vivax* malaria cases. This is because it takes time for an increased number of bites to result in more infected mosquitoes, which would in turn increase in the number of infected people. In Fig. 8, the lines representing the data look very similar because they differ by a constant factor due to our assumptions when disaggregating the incidence data. Likewise, the model looks very similar because at any given time, the rate of Duffy-positive individuals become infected differs from the rate of infection of Duffy-negative individuals. We assumed that the distribution of people with each DARC genotype was equal to the distribution in Rio Pardo. The results of the parameter estimation can be seen in Table 3.

**Table 3: Estimated Parameters for Fast System with Biting Rate Seasonality**

Human Parameters	Description	Value
$\theta_1$	Probability of infection from infected bite for XX	0.729
$\theta_2$	Probability of infection from infected bite for XO	0.729
$\theta_3$	Probability of infection from infected bite for OO	0.060
$\gamma_1$	Recovery Rate for XX	0.0159
$\gamma_2$	Recovery Rate for XO	0.0159
$\gamma_3$	Recovery Rate for OO	0.0159
Mosquito Parameters		
$a_0$	Average Biting Rate	0.0613

$\epsilon$	Degree of Variation around Biting Rate	0.0986
$\omega$	Frequency for seasonal malaria epidemic	0.0025
$t_c$	Phase Shift in Periodicity	-50.07
$c$	Mosquitoes per Human	1.39
$\phi_1$	Probability a mosquito is infected from XX	0.520
$\phi_2$	Probability a mosquito is infected from XO	0.520
$\phi_3$	Probability a mosquito is infected from OO	0.025
$\delta$	Mosquito Death Rate	0.121

## 5 Results

We developed a multi-scale mathematical model that quantifies the impact of the DARC genotype on the epidemiology of *vivax* malaria, and the impact of *vivax* malaria on the long-term frequency of Duffy-positivity. From our sensitivity analysis of  $\mathcal{R}_0$  for the fast system without seasonal biting rate, we find that the mosquito biting rate,  $a$ , has the highest elasticity index, and therefore the largest impact on the value of  $\mathcal{R}_0$ . Thus, varying the mosquito biting rate heavily influences the dynamics of *P. vivax*.

Since malaria incidence data exhibits seasonal fluctuations, we implemented seasonality into our model by making the biting rate a sinusoidal function of time. Using this updated model, we then conducted sensitivity analysis on the model's prediction of the total number of people infected across genotypes. However, because different places have different percentages of the population with each genotype, we conducted this analysis using the genotype fractions from Macapa and Rio Pardo, where about 59% and 3% of people are Duffy-negative, respectively.

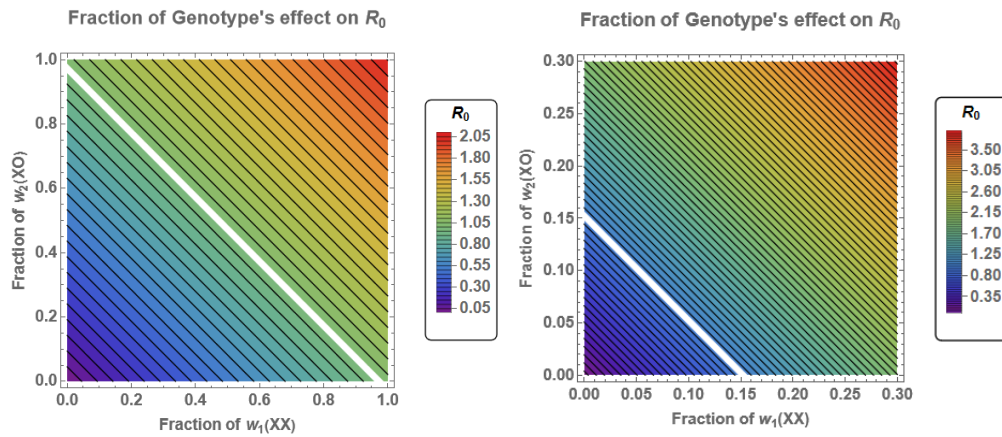
When comparing the results of the sensitivity analysis on the model incorporating seasonality, we find that in both Macapa and Rio Pardo, the recovery rates  $\gamma_1$  and  $\gamma_2$  have the largest impact on the number of total infected people, which reflects our model's assumption that people leave the infected class at these rates. Aside from these two parameters, the average biting rate  $a_0$  has the largest sensitivity index, with  $c$ , the ratio of mosquitoes to human, coming in second. The population of mature female mosquitoes per area is thereby a driving force of infection by *P. vivax*. Given that the life cycle of mosquitoes is dependent on temperature and humidity, we defined the biting rate as a sinusoidal function. By fitting this to our data, we assume that biting rate is the driving force of this cycle. Our sensitivity analysis confirms this positive correlation between the biological reality of the vector and our model.

Similarly, mosquitoes are infected by taking a blood meal from human populations; however, taking this blood meal from a Duffy-negative individual has no effect on the rate of mosquito infection. Indeed, only  $\phi_1$  and  $\phi_2$  have positive sensitivity indices in both areas. In Rio Pardo,  $w_3$  has a negligible sensitivity index due to the small percentage of Duffy-negative people in Rio Pardo. In Macapa,  $w_3$  has greater sensitivity due to the larger percentage of Duffy negative people in Macapa. However,  $w$  is still smaller than  $w_1$  and  $w_2$  due to the reduced transmission rates for Duffy-negative people.

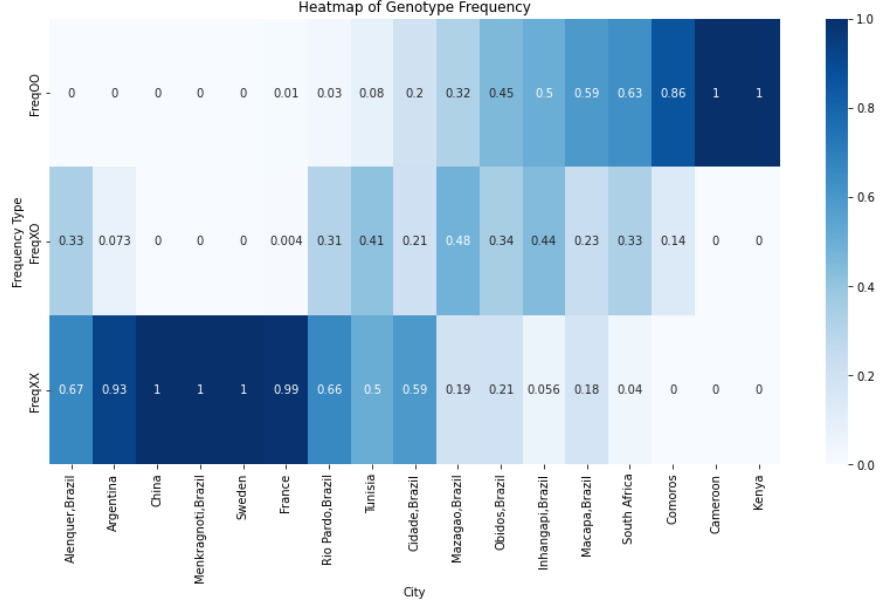
We conducted parameter estimation on the fast system by utilizing a time series of cases in Brazil [9] to fit the transmission rates of malaria between mosquitoes and humans. Sea-

sonality is incorporated through the mosquito biting rate in the fast subsystem so as to fit the seasonality of malaria cases in Brazil. We established a fit to the monthly incidence for Duffy-positive and Duffy-negative people, and determined a mean average percentage error of 7.02%. The transmission rates predicted by the parameter estimation are lower than the ones determined by literature. This leads to conservative  $\mathcal{R}_0$  values.

We investigate the impact of the fraction of people of XO and OO people on the reproduction number  $\mathcal{R}_0$ . Fig. 9 are formulated by utilizing transmission rates obtained through parameter estimation, and identical transmission rates multiplied by 2.5 to yield a high transmission scenario. The left plot suggests for the transmission rates found through parameter estimation that the fraction of homozygous or heterozygous Duffy-positive people doesn't affect  $\mathcal{R}_0$ . The right plot with higher transmission rates illustrates that at least 85% of the population needs to be OO for  $\mathcal{R}_0 < 1$  or in other words *P. vivax* is controlled.



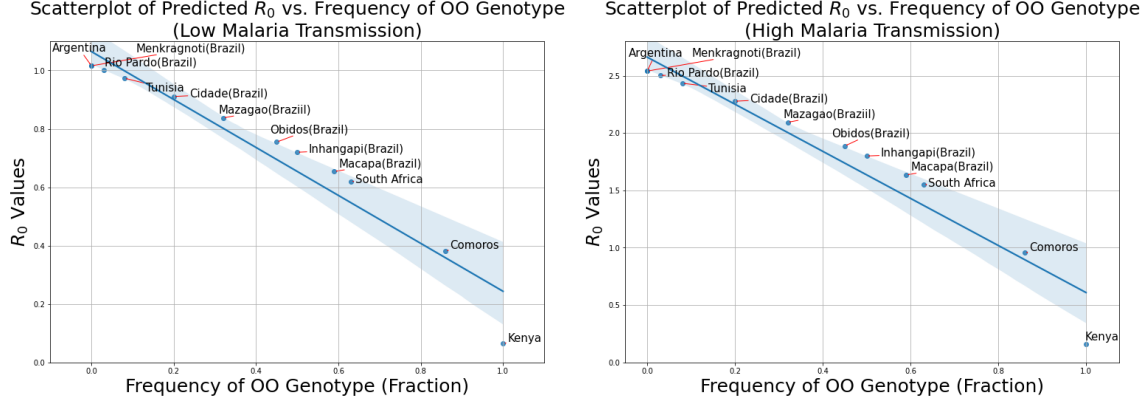
**Figure 9:** The white line through both figures is when  $\mathcal{R}_0 = 1$ . The left figure is a contour plot of how the fraction of homozygous or heterozygous Duffy-positive people affects  $\mathcal{R}_0$  for low transmission rates. The right figure is the same but for higher transmission rates.



**Figure 10:** Frequencies of Duffy genotypes in different locations. Data from [14].

To quantify the relationship between disease burden and the frequencies of each Duffy genotype, we obtained Duffy genotype distributions for different locations from a handful of studies. In Fig. 10 is the map and the Duffy genotype breakdown of each sampled place. It can be seen that in the cities sampled from Cameroon and Kenya that the whole population is Duffy-negative while in the cities from China and Sweden the whole population is homozygous Duffy-positive. Some cities in Brazil have more diverse profiles such as Mazagão and Macapa.

We then use  $\mathcal{R}_0$  defined in (14) to plot the basic reproduction number vs. the fraction of people who are Duffy-negative. Additionally, we use the model parameters from Table 3 and  $w_1, w_2, w_3$  from the data mentioned above. Note that some locations present in Fig. 10 have the same Duffy genotype frequencies as other locations, and are thus redundant and are not included in the scatterplot. To represent scenarios with high malaria transmission, we multiply the infection rate parameters  $\beta_{hi}, \beta_{vi}, (i \in \{1, 2, 3\})$  by a factor of 2.5 to increase the rates of infection.



**Figure 11:** Scatterplots of predicted  $\mathcal{R}_0$  values versus the frequency of the OO genotype for low (Left) and high (Right) transmission rates with lines of best fit.

In Fig. 11, the lines of best fit for the low transmission and high transmission scatterplots are  $y = -0.8207x + 1.0644$  and  $y = -2.052x + 2.661$ , respectively. According to these slopes, a decrease of about 10 percentage points in the percentage of Duffy-negative people results in an increase of  $\mathcal{R}_0$  by about 0.08 when the malaria transmission is low. When we increase the transmission rates by a factor of 2.5, the same 10 percentage point difference causes a roughly 2.5 times larger increase to  $\mathcal{R}_0$  of about 0.21. Additionally, the relation is slightly nonlinear, as changes to smaller fractions of Duffy-negative people result in a smaller increase to  $\mathcal{R}_0$ .

## 6 Discussion

The purpose of our mathematical model is to characterize the epidemiology of *P. vivax* malaria in relation to the DARC genotype. We do so by separating the genotype into three classifications, each relating to the number of alleles that are silent or expressed in a Duffy polymorphism. Thus, we employ a three-genotype model similar to the one Feng and Castillo-Chavez [6] developed to describe the sickle cell trait.

Using the slow system of our model, we obtained the fitness of the X-gene, which represents the likeliness of the expressed Duffy alleles FY\*A and FY\*B to persist into future generations. As shown in Fig. 4, as the recovery rate of Duffy-negative people ( $\gamma_3$ ) increases, it becomes more advantageous to be Duffy-negative and the X-gene is less likely to be selected for than the silent O-gene. As the infection rate of Duffy-negative people decreases, the fitness of the X-gene also decreases since it's more beneficial to be Duffy-negative.

Through sensitivity analysis on both  $\mathcal{R}_0$  and the equation for the total number of infected people, we determined that the mosquito biting rate is highly influential parameter in both quantities, and thus highly influential in *P. vivax* malaria dynamics.

When we investigated the impact of the frequencies of the DARC genotypes and the transmission rates on  $\mathcal{R}_0$ , we found that as the fractions of homozygous and heterozygous Duffy-positive people increase,  $\mathcal{R}_0$  also increases. Based on the model, there is an nonlinear inverse relationship between the frequency of Duffy negative people and the basic reproductive number  $\mathcal{R}_0$ . To gather these results, we used a handful of Duffy genotype population

breakdowns from various locations. Additionally, increasing the transmission rates resulted in an increase to the smallest percentage of Duffy-negative people required to keep *P. vivax* outbreaks under control.

Our model quantifies how the diversity of DARC genotypes impacts the dynamics of *P. vivax* malaria infections. In this paper, we established a framework to determine what percentage of Duffy-negative people are required to keep *P. vivax* outbreaks under control, once given the transmission rates from a specific geographic area. With statistically representative genetic data with respect to Duffy genotypes, our model can more accurately determine which populations are at risk of *vivax* malaria. Furthermore, once given region-specific infection parameters, this model can be applied to geographic regions around the world with varying genotypic frequencies. This provides valuable insight for the delegation of resources to different populations, such as antimalarial drugs, insecticide-treated bed nets, and malarial vaccine trials. Future work includes conducting Markov chain Monte Carlo sampling to establish confidence intervals for our parameter estimation. In addition, we hope to apply Extended Fourier Analysis Test (eFAST) to the fast subsystem with biting rate seasonality to refine the sensitivity analysis. Additionally, we will incorporate evolutionary game theory into our study, as it presents an interesting approach to further investigate the evolutionary dynamics of *P. vivax* malaria. Different seasonality models such as the two peak model outlined by White et al. [26] need to be considered as well.

## 7 Acknowledgements

We would like to thank Dr. Fabio Milner, Director of the Simon A. Levin Mathematical, Computational, and Modeling Sciences Center (MCMSC), for the opportunity to participate in the Quantitative Research in the Life and Social Sciences Program (QRLSSP). We would also like to thank Dr. John Nagy, co-director of the QRLSSP, and coordinator Dr. Steffen Eikenberry for their efforts in planning and executing the program's instruction and activities. We recognize the work of the many faculty, graduate students, and tutors who helped facilitate the program. This research was conducted as part of the 2024 QRLSSP at the Levin Center (MCMSC) at Arizona State University (ASU). This project has been supported by a grant from the National Science Foundation (NSF Grant- DMS-2150492), the Office of the President of ASU, and the Office of the Provost of ASU.

## 8 Appendix

### 8.1 Mosquito Model Re-scale (3)

Taking the derivative of both sides with respect to time, substituting the expression for  $\frac{dI_m}{dt}$  from (3), and simplifying yields

$$\begin{aligned} z &= I_m \cdot M^{-1} \\ \frac{dz}{dt} &= M^{-1} \left[ \frac{dI_m}{dt} \right] + I_m \left[ -M^{-2} \frac{dM}{dt} \right] \\ &= M^{-1} \left[ S_m \left( a\phi_1 \frac{v_1}{N(t)} + a\phi_2 \frac{v_2}{N(t)} + a\phi_3 \frac{v_3}{N(t)} \right) - \delta I_m \right] - \frac{I_m}{M} M^{-1} \frac{dM}{dt} \end{aligned}$$

Note that  $z$  can be substituted for  $\frac{I_m}{M}$  and  $(1 - z)$  for  $\frac{S_m}{M}$  because of how we defined  $z$  and because  $M = S_m + I_m$ .

$$\frac{dz}{dt} = (1 - z) \left( a\phi_1 \frac{v_1}{N(t)} + a\phi_2 \frac{v_2}{N(t)} + a\phi_3 \frac{v_3}{N(t)} \right) - \delta z - z \frac{\dot{M}}{M}$$

where  $\dot{M}$  represents  $\frac{dM}{dt}$ . To maintain a roughly constant mosquito-to-human ratio (parameter  $c$ ), we assume that  $M$  is constant. Implementing this assumption yields

$$\frac{dz}{dt} = (1 - z) \left( a\phi_1 \frac{v_1}{N(t)} + a\phi_2 \frac{v_2}{N(t)} + a\phi_3 \frac{v_3}{N(t)} \right) - \delta z.$$

As such, our model can be expressed in the following equations.

$$\begin{aligned} \frac{du_i}{dt} &= P_i N b(N) - mu_i - a\theta_i c z u_i + \gamma_i v_i \\ \frac{dv_i}{dt} &= a\theta_i c z v_i - mv_i - \gamma_i v_i - \alpha_i v_i \\ \frac{dz}{dt} &= (1 - z) \left( a\phi_1 \frac{v_1}{N(t)} + a\phi_2 \frac{v_2}{N(t)} + a\phi_3 \frac{v_3}{N(t)} \right) - \delta z \end{aligned} \tag{33}$$

A similar process can be used to derive the equations for  $\frac{dx_i}{dt}$  and  $\frac{dy_i}{dt}$  starting with the relations  $x_i = \frac{u_i}{N}$  and  $y_i = \frac{v_i}{N}$ . This results in the following fractional expressions of (3).



## References

- [1] Andargie Abate et al. “Differential transmissibility to *Anopheles arabiensis* of *Plasmodium vivax* gametocytes in patients with diverse Duffy blood group genotypes”. In: *Malaria Journal* 22.1 (2023), p. 136.
- [2] Leon Arriola and James M Hyman. “Sensitivity analysis for uncertainty quantification in mathematical models”. In: *Mathematical and statistical estimation approaches in epidemiology*. Springer, 2009, pp. 195–247.
- [3] Fábio Saito Monteiro de Barros, Nildimar Alves Honorio, and Mercia Eliane Arruda. “Survivorship of *Anopheles darlingi* (Diptera: Culicidae) in relation with malaria incidence in the Brazilian Amazon”. In: *PloS one* 6.8 (2011), e22388.
- [4] *Brazil*. en. <https://data.who.int/countries/076>. Accessed: 2024-6-19.
- [5] Thomas S Churcher et al. “Probability of transmission of malaria from mosquito to human is regulated by mosquito parasite density in naive and vaccinated hosts”. In: *PLoS pathogens* 13.1 (2017), e1006108.
- [6] Zhilan Feng and Carlos Castillo-Chavez. “The influence of infectious diseases on population genetics”. In: *Mathematical Biosciences & Engineering* 3.3 (2006), pp. 467–483.
- [7] Zhilan Feng et al. “Coupling ecology and evolution: malaria and the S-gene across time scales”. In: *Mathematical biosciences* 189.1 (2004), pp. 1–19.
- [8] Renzo Galanello and Raffaella Origa. “Beta-thalassemia”. In: *Orphanet journal of rare diseases* 5 (2010), pp. 1–15.
- [9] Klauss Kleydmann Sabino Garcia et al. “Is Brazil reaching malaria elimination? A time series analysis of malaria cases from 2011 to 2023”. In: *PLOS Global Public Health* 4.1 (2024), e0002845.
- [10] Carlos A Guerra et al. “The international limits and population at risk of *Plasmodium vivax* transmission in 2009”. In: *PLoS neglected tropical diseases* 4.8 (2010), e774.
- [11] Terence J Hadley and Stephen C Peiper. “From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen”. In: *Blood, The Journal of the American Society of Hematology* 89.9 (1997), pp. 3077–3091.
- [12] Gabriela Höher, Marilu Fiegenbaum, and Silvana Almeida. “Molecular basis of the Duffy blood group system”. In: *Blood Transfusion* 16.1 (2018), p. 93.
- [13] Rosalind E Howes et al. “Global epidemiology of *Plasmodium vivax*”. In: *The American journal of tropical medicine and hygiene* 95.6 Suppl (2016), p. 15.
- [14] Rosalind E Howes et al. “The global distribution of the Duffy blood group”. In: *Nature communications* 2.1 (2011), p. 266.
- [15] Sarafa Adewale Iyaniwura et al. “Regional variation and epidemiological insights in malaria underestimation in Cameroon”. In: *medRxiv* (2023), pp. 2023–11.
- [16] Flora Satiko Kano et al. “Susceptibility to *Plasmodium vivax* malaria associated with DARC (Duffy antigen) polymorphisms is influenced by the time of exposure to malaria”. In: *Scientific reports* 8.1 (2018), p. 13851.

- [17] Simeone Marino et al. “A methodology for performing global uncertainty and sensitivity analysis in systems biology”. In: *Journal of theoretical biology* 254.1 (2008), pp. 178–196.
- [18] GM1 Meny. “The Duffy blood group system: a review”. In: *Immunohematology* 26.2 (2010), pp. 51–56.
- [19] Joseli Oliveira-Ferreira et al. “Malaria in Brazil: an overview”. In: *Malaria journal* 9 (2010), pp. 1–15.
- [20] Jean Popovici, Camille Roesch, and Virginie Rougeron. “The enigmatic mechanisms by which Plasmodium vivax infects Duffy-negative individuals”. In: *PLoS Pathogens* 16.2 (2020), e1008258.
- [21] Marissa Renardy et al. “Global sensitivity analysis of biological multiscale models”. In: *Current opinion in biomedical engineering* 11 (2019), pp. 109–116.
- [22] David J Sullivan, Peter Agre, et al. “Human Plasmodium vivax mosquito experimental transmission”. In: *The Journal of Clinical Investigation* 130.6 (2020), pp. 2800–2802.
- [23] Priya Venkatesan. “The 2023 WHO World malaria report”. In: *The Lancet Microbe* 5.3 (2024), e214.
- [24] Amy Yomiko Vittor et al. “The effect of deforestation on the human-biting rate of Anopheles darlingi, the primary vector of falciparum malaria in the Peruvian Amazon”. In: *American Journal of Tropical Medicine and Hygiene* 74.1 (2006), pp. 3–11.
- [25] Liping Wang et al. “Modeling the transmission and control of Zika in Brazil”. In: *Scientific reports* 7.1 (2017), p. 7721.
- [26] Michael T White et al. “Immunogenicity of the RTS, S/AS01 malaria vaccine and implications for duration of vaccine efficacy: secondary analysis of data from a phase 3 randomised controlled trial”. In: *The Lancet infectious diseases* 15.12 (2015), pp. 1450–1458.
- [27] Michael T White et al. “Plasmodium vivax and Plasmodium falciparum infection dynamics: re-infections, recrudescences and relapses”. In: *Malaria journal* 17 (2018), pp. 1–15.