Systems Biology of Epidemiology:
From Genes to Environment

MBI CTW: From Within-Host Dynamics to the Epidemiology of Infectious Disease

Juan B. Gutierrez

UGA’s Department of Mathematics & Institute of Bioinformatics

April 10, 2014
Collaborators

MaHPIC Project. PI: Mary Galinski (Emory), Jessica Kissinger (UGA), Alberto Moreno (Emory). NIAID Grant HHSN272201200031C (2012-2017).

ICEMR Project. PI: Socrates Herrera (Centro de Investigación Caucaseco, Colombia), Myriam Arevalo (Centro de Investigación Caucaseco, Colombia), Martha Quinones (Universidad Nacional de Colombia). NIAID Grant U19AI089702-01 (2010-2017).
The Pioneers

Left to right: Ronald Ross, Robert Koch, Giovanni Battista Grassi. With the use of a microscope, the pathogens that inspired them look identical.
Epidemics are Driven by Cellular and Molecular Interactions

- The circulation of pathogens is the result of hosts shedding infectious agents in the environment.
- The subsequent pathogen uptake and colonization of suitable hosts is the fundamental cause of epidemics.
- Due to a host’s immune response, exposure to low levels of a pathogen may not result in infection; furthermore, pathogens may be unable to find a within-host substrate before degradation.
- Therefore, the onset of infection requires the direct or indirect (vector-driven) uptake of a minimum dose of infectious agents (virus, bacteria, fungi, protozoa, or helminths).
Epidemics are Driven by Cellular and Molecular Interactions

- The minimum infectious dose that colonizes a host with 50% probability of success is known as $ID_{50}$. It has a very large range of variation, from a minimum of one infectious agent for e.g. *Coxiella burnetii* to $>2 \times 10^{10}$ for *Gardnerella vaginalis* (Gama JA, Abby SS, Vieira-Silva S, Dionisio F, Rocha EPC (2012) Immune Subversion and Quorum-Sensing Shape the Variation in Infectious Dose among Bacterial Pathogens. PLoS Pathog 8(2): e1002503. doi:10.1371/journal.ppat.1002503).

- Hence, $ID_{50}$ can be interpreted as an Allee effect for the onset of infection.
Epidemics are Driven by Cellular and Molecular Interactions

- Microdiversity among strains of the vast majority of pathogens is extensive; each genotype infecting a host can present significant differences in virulence, immunogenicity, and antigenic variation.

- Pathogens in circulation are not uniform; instead, they form a distribution defined by the expression of different genetic, pathogenic and population dynamic traits.

- The circulation of pathogens with different genotypes is multifactorial and can depend heavily on human movement dynamics, and, in some situations, vector availability and competence.
Immunoepidemiology


- Since then, there have been multiple attempts to link within-host dynamics and between-host transmission.
Challenges of ’omic data

- Thousands of variables, but significantly less samples.
- Traditionally, a t-test produces a p-value. The null hypothesis (there is no relationship) is rejected (there is relationship) when $p < 0.05$.
- Multiple tests have to use a smaller p-value. Naive example: 15,000 genes per sample. Then, the null hypothesis would be rejected when $p < 0.05/15,000 \approx 0.000003$. We could use only a dozen variable out of 15,000 with this approach.
- False Discovery Rate (FDR) is used instead. FDR procedures are designed to control the expected proportion of incorrectly rejected null hypotheses. In the previous example, with FDR $= 5\%$, we could use 150 variables.

Molecular Changes vs. Physiological State

Image source: MaHPIC Project, Gutierrez Lab
Molecular Changes vs. Physiological State

Image source: MaHPIC Project, Gutierrez Lab
Molecular Changes vs. Physiological State

![KEGG](http://www.genome.jp/dbget-bin/www_bget?hsa:7167)

<table>
<thead>
<tr>
<th>Entry</th>
<th>7167</th>
<th>CDS</th>
<th>T01001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene name</td>
<td>TP1, HEL-S-49, TIM, TPI, TPID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definition</td>
<td>triosephosphate isomerase 1 (EC:5.3.1.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthology</td>
<td>K01803 triosephosphate isomerase (TIM) [EC:5.3.1.1]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>hsa Homo sapiens (human)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa00010 Glycolysis / Gluconeogenesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa00051 Fructose and mannose metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa00562 Inositol phosphate metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa01100 Metabolic pathways</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa01200 Carbon metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa01230 <strong>Biosynthesis of amino acids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module</td>
<td>hsa_M00001 Glycolysis (Embden-Meyerhof pathway), glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module</td>
<td>hsa_M00002 Glycolysis, core module involving three-carbon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module</td>
<td>hsa_M00003 Gluconeogenesis, oxaloacetate =&gt; fructose-6P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>H00664 Anemia due to disorders of glycylcotic enzvmes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Molecular Changes vs. Physiological State

Molecular Changes vs. Physiological State

![Graph showing the relationship between days and the number of significantly changed metabolites.](image-source: MaHPIC Project, Gutierrez Lab)
Molecular Changes vs. Physiological State

Image source: MaHPIC Project, Gutierrez Lab
Within-Host Dynamics
Within-Host Dynamics

Image source: Yan, Gutierrez, in preparation (2013)
A Change in Perspective

Image source: http://www.biology.arizona.edu/cell_bio/tutorials/cells/cells2.html
A Shift in Paradigm

A simple model that takes into account the immune system. Let $p$ be pathogen population density, and $m$ represent immune cell population.

\[
\begin{align*}
\dot{p} &= \nu p(1 - p)(p - \sigma) - \mu mp, \\
\dot{m} &= \frac{mp}{\xi + p} - \eta m,
\end{align*}
\]

- $\nu$ is the rate of growth of the pathogen population.
- $\sigma$ is the Median Infective Dose (MID50).
- $\mu$ is the rate of removal of pathogens by the immune response.
- $\xi$ is the pathogenic load that elicits half the immune response.
- $\eta$ is the rate of removal of immune cells.
A simple model that takes into account the immune system. Let $p$ be pathogen population density, and $m$ represent immune cell population.

Image source: Gutierrez (2013, in preparation)
Now consider the system

\[
\frac{\partial p}{\partial t} = D \Delta p + \nu p(1 - p)(p - \sigma) - \mu mp,
\]

(2)

\[
\frac{\partial m}{\partial t} = \frac{mp}{\xi + p} - \eta m,
\]

(3)

Multiply Equation 2 through by \( p \) and integrate both sides over \( \Omega \) to yield

\[
\int_{\Omega} p \frac{\partial p}{\partial t} \, dx = D \int_{\Omega} p \Delta p \, dx + \nu \int_{\Omega} (1 - p)(p - \sigma)p^2 \, dx - \mu \int_{\Omega} mp^2 \, dx.
\]

(4)

Green’s first identity,

\[
\int_{\Omega} p \Delta p \, dx = \int_{\partial \Omega} p \nabla p \cdot \mathbf{n} \, dS - \int_{\Omega} |\nabla p|^2 \, dx = \int_{\partial \Omega} p \frac{\partial p}{\partial n} \, dS - \int_{\Omega} |\nabla p|^2 \, dx,
\]

is applied, where \( \mathbf{n} \) is the outward pointing unit vector normal to the surface element \( dS \), and \( \partial p/\partial n \) is the directional derivative. Note that under Dirichlet and Neumann boundary conditions, the boundary integral becomes zero.
Basic Propagation Number

Therefore,
\[ \int_{\Omega} p \Delta p d\mathbf{x} = - \int_{\Omega} |\nabla p|^2 d\mathbf{x} = - |\nabla p|^2. \]  
(5)

Also note that
\[ \int_{\Omega} p \frac{\partial p}{\partial t} d\mathbf{x} = \frac{1}{2} \frac{\partial}{\partial t} \int_{\Omega} p^2 d\mathbf{x} = \frac{1}{2} \frac{\partial |p|^2}{\partial t}. \]  
(6)

Substitute Equations 5 and 6 into Equation 4 to obtain
\[ \frac{1}{2} \frac{\partial |p|^2}{\partial t} = -D |\nabla p|^2 + \nu \int_{\Omega} (1 - p)(p - \sigma)p^2 d\mathbf{x} - \mu \int_{\Omega} mp^2 d\mathbf{x}, \]  
(7)

from where
\[ \frac{1}{2} \frac{\partial |p|^2}{\partial t} \leq -D |\nabla p|^2 + \nu \int_{\Omega} (1 - p)(p - \sigma) d\mathbf{x}. \]  
(8)
Basic Propagation Number

Use Poincare’s inequality $|p|_2^2 \leq C|\nabla p|_2^2$ to yield

$$\frac{1}{2} \frac{\partial |p|_2^2}{\partial t} \leq - \frac{D}{C} |p|_2^2 + \nu \int_\Omega (1 - p)(p - \sigma) \, dx$$

$$\leq - \frac{D}{C} |p|_2^2 + \nu \int_\Omega (p - \sigma - p^2 + \sigma p) \, dx$$

$$\leq - |p|_2^2 \left( \nu + \frac{D}{C} \right) + \nu(1 + \sigma)P - \nu\sigma\Omega$$

where $P$ is the total amount of pathogen in the environment, and $\Omega$ is the size of the domain. This equation can be rearranged as an ODE on $|p|_2^2$:

$$|p|_2^2 + \left( \nu + \frac{D}{C} \right) |p|_2^2 + \nu[\sigma(\Omega - P) - P] \leq 0$$
Basic Propagation Number

The solution to the differential inequality \( \dot{p} + ap + b \leq 0 \) subject to \( p(0) = p_0 \) is

\[
p \leq -\frac{b}{a} + e^{-at} \left( p_0 + \frac{b}{a} \right).
\]

Then, there is a time given by

\[
t_0 = \frac{\ln \left( p_0 + \frac{b}{a} \right)}{a}
\]

such that the following uniform estimate holds

\[
p \leq \frac{a - b}{a}
\]
Basic Propagation Number

In terms of the original differential inequality,

$$|p|^2 \leq 1 - \frac{\sigma(\Omega - P) - P}{(1 + \frac{D}{\nu C})}$$

The interpretation of this result is that to minimize the density of pathogen, and hence stop transmission:

- Increase MID50.
- Decrease pathogen in the environment.
- Decrease size of the affected domain.
- Increase diffusivity.
- Decrease rate of infection.
Between-Host Dynamics

- Carey, H. C. (1867). *Principles of social science* (Vol. 3). JB Lippincott & Company. Ch 2 pg 41: “As planets gravitate towards each other, man tends towards his fellow man”; “Man, the molecule of society”

- Reilly, W. J. (1931). *The law of retail gravitation*. “Any city draws trade from its neighboring land. The trade between cities is inversely proportional to the square of the distance between the two cities and proportional to the population of each city”. Elder, RF (Ed.) The American economic review. Volume 21 (1931), pp. 528-530.


Implicit movement characterized by a gravity model can be made explicit through the use diffusion on a low-dimensional manifold reconstructed from effort of transportation between individuals.

Given a planar graph representing the distribution of human population in a geographic area, where the vertices represent the population centers, and the edges represent the transportation network, then it is possible to assign weights to the edges so that each weight represents the time required to travel between two vertices.
Figure: Manifold that results from considering the quality of roads between towns (left), and the spline smoothing (right) based on bivariate splines.
Consider $T_u M$, the tangent space of $M$ for each point $u$ in the following way

$$\frac{\partial p(x, t)}{\partial t} = D \nabla \cdot (C(|T_u M|) \nabla p(x, t)) + F(p(x, t), m(x, t)),$$

where $x = (x, y) \in \Omega \subset \mathbb{R}^2, t \geq 0$, and $|T_u M|$ represents a metric of the tangent space of $M$, such that

$$C(S) = \frac{1}{\sqrt{1 + |S|^2}}.$$
Key Aspects of Infectious Disease Can Hardly Be Studied with Existing Paradigms

- **Microdiversity**: Among strains of the vast majority of pathogens is extensive; each genotype infecting a host can present significant differences in virulence and immunogenicity.

- **Simultaneous infection**: Simultaneous presence of several distinct pathogen genomes, from the same or multiple species.

- **Antigenic diversity**: Antigenic differences between pathogens in a population.

- **Antigenic variation**: Ability of a pathogen to change antigens presented to the immune system during an infection.
Asymptomaticity is Very Important

Vector Behavior is Very Important

CORDOBA - NTV. Protection Loss = 23.6%

Population Density vs Time (hours)

Image source: CLAIM Project, 2014
Vector Behavior is Very Important

Image source: CLAIM Project, 2014
The Malaria Case
The Malaria Case

Consider the following populations:

- **Human**: A set \( H = \{ h_i \} \).
- **Mosquito**: A set \( M = \{ m_j \} \).
- **Plasmodium**: A set of classes \( P = \{ p_{kl} \} \) representing population density of different genotypes \( k \) and different life stages \( l \).
\[ \dot{p}_{ik1} = \sum_{n=1}^{J} \lambda_{ji} h(d_{ij}) p_{nk1} - \tau_2 p_{ik1} - \tau_3 p_{ik1} - \omega_1 \frac{p_{ik1} e_i}{1 + \gamma_1 p_{ik1}} \]  

\[ \dot{p}_{ik2} = \tau_2 p_{ik1} + \phi_2 p_{ik2} \left( 1 - \frac{p_{ik2} + p_{ik3}}{C_L} \right) - \tau_4 p_{ik2} - \omega_2 \frac{p_{ik2} e_i}{1 + \gamma_2 p_{ik2}} \]  

\[ \dot{p}_{ik3} = \tau_3 p_{ik1} + \phi_3 p_{ik3} \left( 1 - \frac{p_{ik2} + p_{ik3}}{C_L} \right) - \zeta_4 p_{ik3} - \omega_3 \frac{p_{ik3} e_i}{1 + \gamma_3 p_{ik3}} \]  

\[ \dot{p}_{ik4} = \tau_4 p_{ik2} + \zeta_4 p_{ik3} + \phi_5 p_{ik5} \left( 1 - \frac{p_{ik4}}{C_M} \right) - \tau_5 p_{ik4} - \tau_6 p_{ik4} - \omega_4 \frac{p_{ik4} e_i}{1 + \gamma_4 p_{ik4}} \]  

\[ \dot{p}_{ik5} = \tau_5 p_{ik4} + \phi_5 p_{ik5} \left( 1 - \frac{p_{ik5}}{C_{RBC}} \right) + \rho \left( \sum_{s=1, s \neq j}^{K} p_{is5} - p_{ik5} \right) - \omega_5 \frac{p_{ik5} e_i}{1 + \gamma_5 p_{ik5}} \]  

\[ \dot{f}_{ik} = \frac{1}{2} \tau_6 p_{ik4} \left( 1 - \frac{f_{ik} + m_{ik}}{C_G} \right) - \sum_{n=1}^{J} \lambda_{ij} h(d_{nj}) f_{nk} - \omega_f \frac{p_{ikf} e_i}{1 + \gamma_f p_{ikf}} \]  

\[ \dot{m}_{ik} = \frac{1}{2} \tau_6 p_{ik4} \left( 1 - \frac{f_{ik} + m_{ik}}{C_G} \right) - \sum_{n=1}^{J} \lambda_{ij} h(d_{nj}) m_{nk} - \omega_m \frac{p_{ikm} e_i}{1 + \gamma_m p_{ikm}} \]  

\[ \dot{e}_i = \sum_{l=1}^{1..5,f,m} \kappa_l \frac{p_{ikl} e_i}{1 + \gamma_l p_{ikl}} - \epsilon e_i \]  

\[ \dot{p}_{jk1} = \sum_{q=1}^{J} \lambda_{ij} h(d_{ij}) (f_{qk} + m_{qk}) - \tau_1 p_{jk1} - \delta_6 (f_{qk} + m_{qk}) \]
The Malaria Case

- **Human**: A set $H = \{h_i\}$, with $I = \text{total number of humans}$,
- **Mosquito**: A set $M = \{m_j\}$, with $J = \text{total number of mosquitoes}$.
- **Plasmodium**: A set of classes $P = \{p_{kl}\}$ representing population density of different genotypes $k$ and different life stages $l$.

\[
\dot{p}_{ik_1} = \sum_{n=1}^{J} \lambda_{ji} h(d_{ij}) p_{nk_1} - \tau_2 p_{ik_1} - \tau_3 p_{ik_1} - \omega_1 \frac{p_{ik_1} e_i}{1 + \gamma_1 p_{ik_1}}
\]

\[\ldots\]

\[
\dot{p}_{jk_1} = \sum_{q=1}^{I} \lambda_{ij} h(d_{ij}) (f_{qk} + m_{qk}) - \tau_1 p_{jk_1} - \delta_6 (f_{qk} + m_{qk})
\]

**New data to be collected**: Human movement and robust detection of anopheline and Plasmodium species is required, e.g. using environmental mitochondrial DNA.
From Cells to Continent

Image source: Gutierrez, CLAIM Project (2013)
Within-host and between-models can be linked through:

- A variable in the between-host models that represents time since infection.
- Within-host dynamics determining pathogen shedding, and network dynamics to determine contact between hosts.
- Within-host dynamics takes into account immune response to generate host classes with different degrees of immunogenicity.
- etc.
The Bottom Line

- Traditional methods to study epidemiology can’t deal with emerging information about large families of pathogens.
- This research tries to condensate in a unified framework within-host dynamics and epidemiological processes.
- Within-host dynamics has to take into account the immune system and the cascade of molecular signaling, hence the “systems biology” in this approach to epidemiology.
THANKS FOR YOUR ATTENTION!

Juan B. Gutierrez

juan@math.uga.edu