Supplementary Information

Our model for the emergence of infectious diseases can be formulated as a multi-type branching process[1]. The multi-type branching process has been used previously to describe biological phenomena, such as the evolution of polynucleotides[2], cell proliferation[3], the dynamics of viral quasispecies[4], and the evolution of drug resistance and immune escape[5,6]. The advantage of the multi-type branching formulation is that the an expression for the probability of emergence can be easily derived.

Consider the case where there are m pathogen strains with $m-1$ mutations required to attain $R_0 > 1$ (note that n in Figure 3 is the number of intermediate strains, and thus $n = m-2$). The m-type branching process is given by m probability generating functions:

$$f_i(s_1, s_2, \ldots, s_m) = \sum_{j_1, j_2, \ldots, j_m = 0}^{\infty} p_i(j_1, j_2, \ldots, j_m) {s_1}^{j_1} {s_2}^{j_2} \cdots {s_m}^{j_m}, \quad i \in \{1, 2, \ldots, m\} \tag{1}$$

Hereby, $p(j_1, j_2, \ldots, j_m)$ denotes the probability that a single infection of type i gives rise to j_i secondary infections with a pathogen with $i-1$ mutations.

The extinction probability of a transmission chain initiated by one infection with wildtype q_1 can be calculated from the fixed point equation:

$$f_i(q_1, q_2, \ldots, q_m) = q_i, \quad i \in \{1, 2, \ldots, m\} \tag{2}$$

To formulate our model as a multi-type branching process, we need to define the probability generating functions. Let us assume we have m different variants of the pathogen, one wild-type and $m-1$ mutants. We denote the basic reproductive number of the wild-type pathogen as $R_0^{(1)}$, that of the $m-2$ intermediate mutants as $R_0^{(2)}, \ldots R_0^{(m-1)}$ and that of the fully-evolved pathogen as $R_0^{(m)}$. We make the following assumptions about the mutation of the pathogen and spread of the infection:

- the mutation rate, μ, of all the variants is the same
- only single mutations occur, i. e. the ith variant can only mutate into an $i+1$ mutant
- the total number of secondary infections arising from an individual infected with variant i is Poisson-distributed with mean $R_0^{(i)}$
- the number of secondary infections of type i arising from an individual infected with variant i is Poisson-distributed with mean $(1-\mu)R_0^{(i)}$
- the number of secondary infections of type $i+1$ arising from an individual infected with variant i is Poisson-distributed with mean $\mu R_0^{(i)}$

With these assumptions, the probability generating functions, $f_1, f_2, \ldots f_m$ are given by:

$$\begin{aligned}
f_1(s_1, s_2, \ldots, s_m) &= \exp[-(1-\mu)R_0^{(1)}(1-s_1)] \exp[-\mu R_0^{(1)}(1-s_2)] && (3)\\
f_2(s_1, s_2, \ldots, s_m) &= \exp[-(1-\mu)R_0^{(2)}(1-s_2)] \exp[-\mu R_0^{(2)}(1-s_3)] && (4)\\
&\vdots \qquad \vdots && (5)\\
f_{m-1}(s_1, s_2, \ldots, s_m) &= \exp[-(1-\mu)R_0^{(m-1)}(1-s_{m-1})] \exp[-\mu R_0^{(m-1)}(1-s_m)] && (6)\\
f_m(s_1, s_2, \ldots, s_m) &= \exp[-R_0^{(m)}(1-s_m)] && (7)
\end{aligned}$$

Using Equation 2, the probability of emergence can be easily calculated numerically.

An intuitive look at the processes we have described can be obtained if we consider the deterministic approximation of the stochastic processes. We first describe the case where a single mutation is required for the evolution of the infection to have R_0 greater than one. The probability of evolution per introduction can be computed from the probability that no evolution occurs. The latter equals $(1-\mu)$ raised to the power of the number of secondary cases. The number of secondary cases equals the sum of the geometric series $R_0^{(0)} + (R_0^{(0)})^2 + (R_0^{(0)})^3 ..... = \frac{R_0^{(0)}}{1-R_0^{(0)}}$. The probability of evolution thus equals

$$1 - (1-\mu)^{\left(\frac{R_0^{(0)}}{1-R_0^{(0)}}\right)} \approx \frac{\mu R_0^{(0)}}{1-R_0^{(0)}} \tag{8}$$

the second approximation requiring that there be very few cases when the probability of evolution is small (i.e. $\mu << 1$ and $R_0^{(0)}$ is not too close to one).

Under these conditions a rough estimate of the probability of an epidemic when m mutations are required to progress from an infection with $R_0^{(0)}$ to $R_0^{(1)}, ..., R_0^{(m-1)}$ all < 1 to one with $R_0^{(m)} > 1$ is

$$p = \left(\frac{\mu R_0^{(0)}}{1-R_0^{(0)}}\right)\left(\frac{\mu R_0^{(1)}}{1-R_0^{(1)}}\right) ... \left(\frac{\mu R_0^{(m-1)}}{1-R_0^{(m-1)}}\right)\left(1 - \frac{1}{R_0^{(m)}}\right) \tag{9}$$

The first term is approximately equal to the probability that the infection with $R_0^{(0)}$ gives rise to an infection with one mutation and the second term is the probability that this infection with $R_0^{(1)}$ gives rise to an infection with two mutations and so on. The last term is the probability that the infection with $R_0^{(m)} > 1$ goes extinct due to stochastic events. Clearly the probability of an epidemic after N introductions will be approximately $P = 1 - \exp(-Np) \approx Np$. This makes clear the observations that:

1. The probability of an epidemic is directly proportional to the number of introductions N.
2. The probabillity of an epidemic increases with mutation rate and the number of mutations required in a manner described by $\mu^{m-1} = \mu^{n+1}$.

In the model described above mutations only result in secondary infections. The above framework can be extended to consider what happens if mutations result in the "conversion" of an infection (with probability μ) from one with $R_0 < 1$ to one with $R_0 > 1$. In this case the initial transmission event can convert to one having a higher R_0. In the case that only one mutation is required for R_0 to exceed 1, the probability of evolution will be approximately

$$\mu\left(1 + \frac{R_0^{(0)}}{1-R_0^{(0)}}\right) \tag{10}$$

In this case the probability of emergence is similar when R_0 is close to one but falls only to μ as R_0 approaches zero.

References

1. Athreya, K. B. & Ney, P. *Branching processes*. Springer-Verlag, Berlin, New York,, (1972).
2. Demetrius, L., Schuster, P. & Sigmund, K. Polynucleotide evolution and branching processes. *Bull Math Biol* **47**, 239–62 (1985).

3. Stivers, D. N., Kimmel, M. & Axelrod, D. E. A discrete-time, multi-type generational inheritance branching process model of cell proliferation. *Math Biosci* **137**, 25–50 (1996).

4. Wilke, C. O. Probability of fixation of an advantageous mutant in a viral quasispecies. *Genetics* **163**, 467–74 (2003).

5. Ywasa, Y., Michor, F. & Nowak, M. A. Evolutionary dynamics of escape from biomedical intervention. *Proc. R. Soc* **in press** (2003).

6. Ywasa, Y., Michor, F. & Nowak, M. A. Evolutionary dynamics of invasion and escape. *J. Theor. Biol.* **in press** (2003).