In many organisms, such as humans, recombination tends to occur at a few specific locations in the genome – "hotspots". These hotspots are caused by "hot" alleles, which increase the chances of a double-strand break of the DNA at their location. The repair of that break does two things: 1) it initiates recombination and 2) it converts the allele that caused the break into the other allele at that locus. The perplexing thing is that this means that over time the hot alleles will be converted into cold alleles (this is called biased gene conversion). Why then do we see hotspots? This is the hotspot conversion paradox.

Let's consider one proposed solution to the hotspot conversion paradox (Peters 2008 *Genetics*). The model has two loci. At the potential hotspot **H** we have two alleles, the cold allele h and the hot allele H. The other locus, **M**, with alleles m and M, determines whether H has any effect. Only if there is an M will H actually increase the rate of recombination. This increased rate of recombination is assumed to be selected for but also converts H to h. The main question is whether M can invade and therefore create a hotspot.

We have four genotype frequencies to track:

genotype	frequency
hm	x_1
Hm	x_2
hM	x_3
HM	x_4

And there are four main forces to consider: gene conversion, selection, mutation, and recombination. Let the probability of a double-strand break at either h or H in the absence of an M allele be β_h. In the presence of at least one M allele the probability of a double-strand break increases to β_H. When a double-strand break occurs in Hh heterozygotes, whichever allele caused the break gets converted to the other. This conversion event will also convert the neighbouring allele at the **M** locus with probability τ. The strength of selection for increased recombination is $\sigma > 0$. The h allele mutates to the H allele with probability ν and the H allele mutates to the h allele with probability μ. Finally, recombination between the **H** and **M** loci occurs with probability r.

In continuous-time the dynamics are described by some rather big equations,

$$\begin{aligned}
\frac{dx_1}{dt} =& x_1x_4(\beta_H\tau/2 - \beta_h/2) + x_2x_3(1-\tau)\beta_H/2 \\
&+ x_1\left(2\beta_h(x_1 + x_2 + x_3 + x_4/2) + \beta_H x_4 - \bar{\beta}\right)\sigma \\
&+ \mu x_2 - \nu x_1 + r(x_2x_3 - x_1x_4) \\
\frac{dx_2}{dt} =& x_2x_3(\beta_h\tau/2 - \beta_H) + x_1x_4(1-\tau)\beta_h/2 \\
&+ x_2\left(2\beta_h(x_1 + x_2 + x_3/2) + 2\beta_H(x_3/2 + x_4) - \bar{\beta}\right)\sigma \\
&+ \nu x_1 - \mu x_2 - r(x_2x_3 - x_1x_4) \\
\frac{dx_3}{dt} =& x_2x_3(\beta_H\tau/2 - \beta_h/2) + x_3x_4(\beta_H/2 - \beta_h/2) + x_1x_4(1-\tau)\beta_H/2 \\
&+ x_3\left(2\beta_h(x_1 + x_2/2 + x_3 + x_4/2) + \beta_H(x_2 + x_4) - \bar{\beta}\right)\sigma \\
&+ \mu x_4 - \nu x_3 - r(x_2x_3 - x_1x_4) \\
\frac{dx_4}{dt} =& x_1x_4(\beta_h\tau/2 - \beta_H/2) + x_3x_4(\beta_h/2 - \beta_H/2) + x_2x_3(1-\tau)\beta_h/2 \\
&+ x_4\left(\beta_h(x_1 + x_3) + 2\beta_H(x_1/2 + x_2 + x_3/2 + x_4) - \bar{\beta}\right)\sigma \\
&+ \nu x_3 - \mu x_4 + r(x_2x_3 - x_1x_4),
\end{aligned}$$

where $\bar{\beta} = 2\beta_h(x_1^2 + 2x_1x_2 + 2x_1x_3 + x_1x_4 + x_2^2 + x_2x_3 + x_3^2) + 2\beta_H(x_1x_4 + x_2x_3 + 2x_2x_4 + 2x_3x_4 + x_4^2)$ is the mean probability of recombination.

1. **[4/20 points]** Let's first consider a population without any M alleles, i.e., $x_3 = x_4 = 0$ and $x_2 = 1 - x_1$. Show how to solve for the equilibrium frequency of Hm genotypes, $\hat{x}_2 = \nu/(\mu+\nu)$. This is called mutation-mutation equilibrium.

When $x_3 = x_4 = 0$ we know $x_2 = 1 - x_1$ and there is only one variable to track, x_1. Making these substitutions gives $\bar{\beta} = 2\beta_h$ and

$$\frac{dx_1}{dt} = \mu(1 - x_1) - \nu x_1.$$

Setting this equation to zero and solving for $\hat{x}_1$ gives $\hat{x}_1 = \mu/(\mu + \nu)$, and thus $\hat{x}_2 = 1 - \hat{x}_1 = \nu/(\mu + \nu)$.

2. **[4/20 points]** To determine stability we look at the Jacobian evaluated at this equilibrium, which can be written in block form,

$$\mathbf{J}|_{\text{eq}} = \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}.$$

Explain why we can determine stability by considering only $\mathbf{A}$ and $\mathbf{B}$.

Stability is determined by the eigenvalues of the Jacobian. In this case the Jacobian is a block triangular matrix. The eigenvalues of a block triangular matrix are the eigenvalues of the submatrices on the diagonal, which in this case are $\mathbf{A}$ and $\mathbf{B}$.

3. **[4/20 points]** The stability of the equilibrium in the absence of any M alleles is determined by $\mathbf{A}$, which has trace $-\mu-\nu-2\beta_h\sigma$ and determinant $(\mu + \nu)2\beta_h\sigma$. Use these facts to determine when the equilibrium is stable in the absence of any M alleles.

Given that μ, ν, and β_h are described as probabilities we know that they are non-negative. We've also been told $\sigma > 0$, i.e., there is selection for more recombination. This implies that, unless μ, ν, and β_h are all zero (a case we'll ignore), the trace is negative and the determinant is positive. By the Routh-Hurwitz criteria we therefore know this equilibrium is always stable in the absence of any M alleles.

4. **[4/20 points]** Whether or not the M allele will invade is determined by $\mathbf{B}$. Unfortunately the eigenvalues and even the trace and determinant are difficult to interpret. To get more intution we do a perturbation analysis, assuming both β_h and β_H are small (i.e., recombination at any particular

hotspot in a given meiosis is unlikely), specifically $\beta_h = \beta'_h\epsilon$ and $\beta_H = \beta'_H\epsilon$, where $\epsilon \ll 1$. We expand the leading eigenvalue as $\lambda = \lambda_0 + \lambda_1\epsilon$. We find that $\lambda_0 = 0$ and

$$\lambda_1 = \frac{\nu}{\mu + \nu} 2(\beta_H - \beta_h)\sigma.$$

Will the M allele invade?

Again, we know that ν and μ are non-negative, that $\sigma > 0$, and we've also been told $\beta_H > \beta_h$, i.e., that the hot allele increases the probability of recombination. So, as long as $\nu > 0$, i.e., there are some H alleles around, the M allele will invade.

5. [4/20 points] Explain, in biological terms, what λ_1 tells us about why (or why not) hotspots may continually arise despite their inevitable destruction via gene conversion.

We see that hotspots will continually arise because $\lambda_1 > 0$ implies alleles like M will invade. The expression for λ_1 is the frequency of the hot allele at equilibrium, $\hat{x}_2 = \nu/(\mu + \nu)$, times twice the increase in recombination rate, times the strength of selection for increased recombination. We can interpret this as saying that the M allele will invade whenever there are some hot alleles segregating ($\hat{x}_2 > 0$) because the increased recombination it causes is favoured by selection. The destruction of the hot alleles via biased gene conversion will eventually cause h to fix, at which point M loses its selective advantage. This process can repeat at many sites in the genome, causing hotspots to appear and disappear over time.