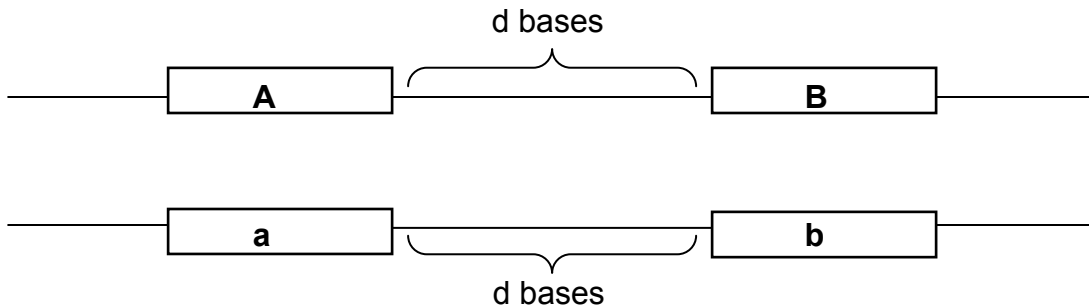


A simplified model that relates genetic distance to physical distance
David Gifford
February 28, 2007

How can we relate genetic distance (recombination frequency) to physical distance? In Meiosis I, homologous chromosomes can exchange genetic information by creating chiasmata and recombining, resulting in recombinant chromosomes. The recombination frequency between two loci is in part a function of the physical distance between the loci. This short note will develop a simple intuitive model of recombination, and provides a collection of caveats with references. However, a simple model can help us gain intuition even if it glosses over certain details.

The remainder of this document is not required material for 7.013 students, but for the curious, press fearlessly ahead!

Let's start by introducing the base flipping game. We have two homologous chromosomes with the genotype (AB / ab):



Let us suppose that we flip a biased coin at each base between the two loci, and the coin will come up heads with probability p and tails with probability $1 - p$. We will call q the probability $1 - p$. At a given base if the coin comes up heads a crossover will take place, and if the coin comes up tails a crossover will not occur. Thus the expected number of flips until we get a crossover event is $1 / p$.

In this highly simplified model we will get the recombinant genotype (Ab / aB) when there are an *odd number of crossovers* between the two loci.

The probability of an odd number of heads after d tosses is:

$$P(\text{odd number of heads} \mid d \text{ flips}) = \frac{1}{2}(1 - (q - p)^d)$$

(We derive this below)

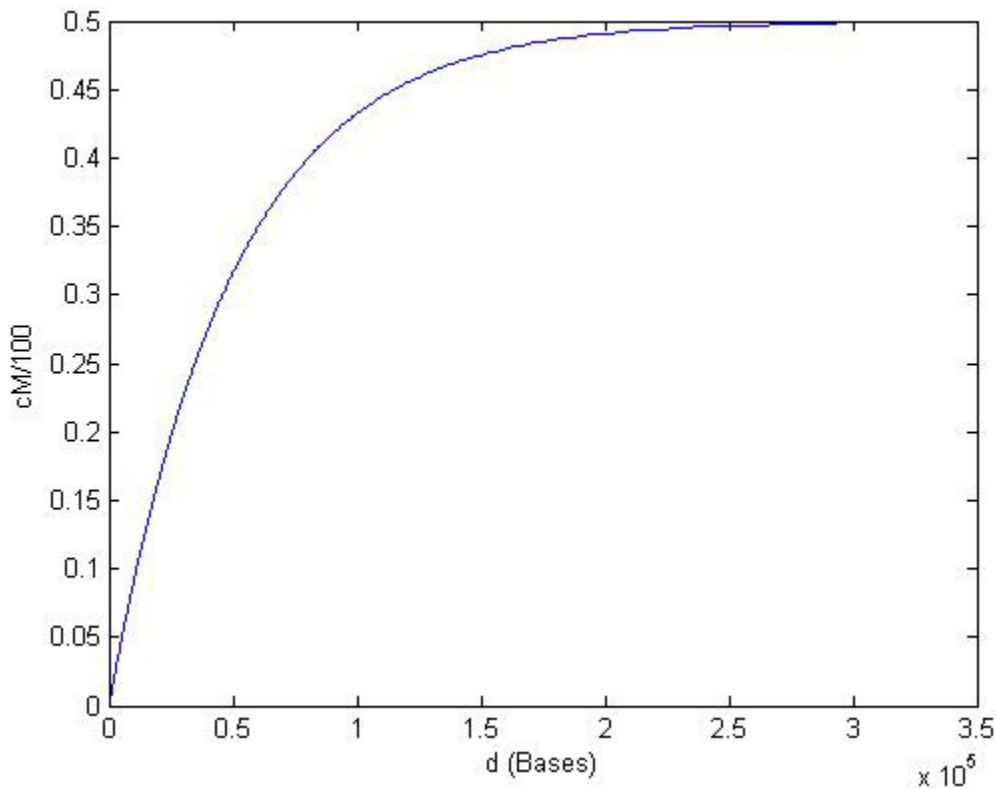
Thus the equivalent distance in cM (100 times the recombination frequency) between the loci is:

$$cM = 50(1 - (q - p)^d)$$

Yielding the equality

$$d = \log(1 - (cM / 50)) / \log(q - p)$$

which when plotted with $p = 10^{-5}$ (for demonstration purposes) looks like:



Note that the expected number of flips until the first head is 100,000 flips, but the probability that we would observe an odd number of heads after 100,000 flips is approximately 0.43, thus equating 100,000 bases to a genetic distance of 43 cM. Recall that we picked p at random, and thus these results are not intended to be representative of any specific organism.

We also note that the relationship between genetic distance (cM) and physical distance (bases) in this model is relatively linear up to about 20 cM. This is why it is possible to add, subtract, and compare short genetic distances without observing too much error, and why short distances can be used to create reasonably accurate genetic maps.

Caveats

- We assume that only two chromosomes participate in our coin flipping game. In fact *four chromosomes* might participate in crossing over in Meiosis I.
- We assume that **p** is fixed along a chromosome. However, *recombinational hot spots* are regions of the genome that have a higher number of crossovers than other portions of the genome. For example, see <http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371%2Fjournal.pbio.0020190>.
- We assume that coin flips are independent. However, *crossover interference* creates dependence between recombination events. For example, see <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1205352>.

By clearly articulating our assumptions, we can identify specific biological phenomenon that falsify specific assumptions. The clear articulation of assumptions is essential to the entire modeling enterprise, and permits us to revise assumptions as warranted – and at the same time revise our understanding of biology to better explain observed data.

Derivation: The probability of an odd number of heads in d flips, where

Given $P(\text{head} \mid 1 \text{ flip}) = p = (1 - q)$

Let $P(\text{odd number of heads} \mid n \text{ flips}) = O(n)$

$$\begin{aligned} O(1) &= p \\ O(n) &= p(1 - O(n-1)) + qO(n-1) \\ &= p + (q - p) O(n-1) \\ &= p + (q - p) (p + (q - p)O(n - 2)) \\ &= p + p(q - p)^1 + p(q - p)^2 + \dots + p(q - p)^{n-1} \\ &= p (1 + (q - p)^1 + \dots + (q - p)^{n-1}) \\ &= p ((1 - (q - p)^n) / (1 - (q - p))) \\ &= p ((1 - (q - p)^n) / 2p) \\ &= 1/2 (1 - (q - p)^n) \end{aligned}$$

Thus $P(\text{odd number of heads} \mid n \text{ flips}) = \frac{1}{2} (1 - (q - p)^n)$