Published in final edited form as:

Genetica. 2009 April ; 135(3): 305–307. doi:10.1007/s10709-008-9287-1.

Casares' map function: No need for a 'corrected' Haldane's map function

Daniel E. Weeks^{1,2}, Xinyu Tang², and Amy M. Kwon²

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

²Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract

A genetic map function $M(d) = RF$ provides a mapping from the additive genetic distance d to the non-additive recombination fraction RF between a given pair of loci, where the recombination fraction is the proportion of gametes that are recombinant between the two loci. Genetic map functions are needed because in most experiments all we can directly observe are the recombination events. However, since a recombination event is only observed if there are an odd number of crossovers between the two loci, recombination fractions are not additive. One of the most widely used map functions is Haldane's map function, which is derived under the assumptions of no chiasma and no chromatid interference, and has been in widespread use since 1919. However, Casares recently proposed a 'corrected' Haldane's map function – we show here that this 'corrected' map function is not correct due to faulty assumptions and mistakes in its derivation.

Introduction

The accurate construction of genetic maps and the mapping of disease genes is crucially dependent on genetic map functions, which define the relationship of additive genetic distances to the non-additive recombination fractions (McPeck 1996; Lange 2002; Speed 2005). To construct a useful genetic map function, one must use precise and clear definitions and accurately model the underlying biological phenomena. During meiosis, each chromosome replicates, forming a pair of sister chromosomes known as chromatids. Then the pair derived from the mother aligns with the pair derived from the father to form a four-strand bundle. Then homologous non-sister strands may break and recombine at exchange points called chiasmata. This may result in recombinant gametes that contain a mixture of maternally- and paternally-derived genetic material; the locations of the points of exchange in the single-strand gametes are known as crossovers. Note that the chiasmata are observed at the four-strand stage, while the crossovers are observed in the single-strand gametes. The genetic distance d between two loci is the average number of crossovers between the two loci per gamete. A genetic map function $M(d) = RF$ provides a mapping from the additive genetic distance d to the non-additive recombination fraction RF between a given pair of loci, where the recombination fraction is the proportion of gametes that are recombinant between the two loci. A recombination event is only observed if there are an odd number of crossovers between the two loci.

Corresponding author: Daniel E. Weeks, Ph.D., Professor of Human Genetics and Biostatistics, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Crabtree Hall, Room A303, 130 DeSoto Street, Pittsburgh, PA 15261, USA, 1 412 624-5388, FAX: 1 412 624-3020, E-mail: E-mail: weeks@pitt.edu.

We comment here on the recent brief report by Pelayo Casares entitled "A corrected Haldane's map function to calculate genetic distances from recombination data" (Casares 2007). Casares (2007) claims that Haldane's map function "is not correct" and instead develops a modified map function which generates shorter genetic distances than Haldane's map function does (Haldane 1919).

Theory and methods

Let us first review how Haldane's map function is derived. There are two basic approaches. The first approach employs Mather's formula (Mather 1938; Speed 1996; Lange 2002), which is derived under the assumption of no chromatid interference and which states that

$$RF = 1/2 [1 - P(\text{zero chiasmata})]. \quad [1]$$

Note that an equivalent form of Mather's formula was first derived by Emerson and Rhoades (1933). If there is no chiasma interference so that the chiasmata occur independently of each other, then the number of chiasmata in a given interval follows a Poisson distribution with mean $2d$, and we have

$$P(\text{zero chiasmata}) = e^{-2d}.$$

Inserting this into Mather's formula gives us Haldane's map function:

$$RF = 1/2 [1 - e^{-2d}].$$

The second approach for deriving Haldane's map function is based on recognizing that a recombination event is only observed if there are an odd number of crossovers between the two loci (Bailey 1961). If the number of chiasmata in a given interval follows a Poisson distribution with mean $2d$, and there is no chromatid interference, then the number of crossovers follows a Poisson distribution with mean d . So we have:

$$RF = \sum_{k=0}^{\infty} e^{-d} \frac{d^{(2k+1)}}{(2k+1)!} = \frac{1}{2} (1 - e^{-2d}).$$

The derivations of Casares (2007) appear to not make the important distinction between crossovers (observed in the gametes) and chiasmata (observed at the four-strand stage). For example, Casares first defines "m" as "the frequency of crossovers", and states that:

$$RF = 1/2 m. \quad [2]$$

However, if we compare Mather's formula in Equation 1 to Equation 2, we see that Equation 2 is true only if "m" is defined as the probability of one or more chiasmata in the interval. But later in the report, Casares redefines "m" as the mean number of "crossovers". A closer reading indicates that, in the majority of the report, Casares is really defining "m" as the expected number of chiasmata. If so, then under the assumption of no chromatid interference, as Speed (2005) states, the expected number of crossovers in the interval is half the expected number of chiasmata, or

$$d=1/2 m.$$

For the rest of this current discussion, let us assume that m is the expected number of chiasmata on the four-strand bundle between the two loci defining the interval.

Casares' main complaint regarding Haldane's map function is that it overestimates genetic distances. Casares arrives at this complaint by the following argument: Suppose there are exactly 2 chiasmata per chromosome, so that the chromosome has genetic length, L , of 100 cM. If we divide the chromosome in 5 equal intervals, then the expected number of chiasmata per interval, m , is 0.40, and so, applying the formula $d = 1/2 m$, the genetic distance d is 0.20 Morgans or 20 cM. Casares then claims that a d of 20 cM corresponds to a Haldane distance of 25.5 cM, which would imply that the corresponding chromosome length $L = 5 \times 25.5$ is too long. However, we do not obtain a Haldane distance of 25.5 cM if we use Casares' own formulae to convert m to RF and then to a Haldane distance by proceeding as follows:

$$RF=1/2 (1 - e^{-m}),$$

we have, for $m = 0.40$, $RF = 0.165$. If we then apply Haldane's map function to convert this RF to genetic distance d in Haldane cM, we obtain:

$$d= - 1/2 \ln(1 - 2RF)=0.20.$$

So each interval is 20 cM long, and so the length of the entire chromosome is $L = 5 \times 20 = 100$ cM as expected. It appears that Casares arrived at a Haldane distance of 25.5 cM for each interval by plugging 0.20 into the formula for the inverse of Haldane's map function:

$$d= - 1/2 \ln(1 - 2RF)= - 1/2 \ln(1 - 2(0.20))=0.255.$$

But this is incorrect, as the 0.20 is *not* a recombination fraction, but rather it is a genetic distance. Thus, we conclude that Casares' argument that the Haldane map function overestimates genetic distances is mistaken and spurious.

Casares (2007) then goes on to derive a 'corrected' map function. The derivation of Casares' map function appears to rest on the incorrect assumption that the "actual number of recombinant gametes" is somehow relevant to the recombination fraction. However, the actual number is not relevant at all, as the recombination fraction is measured only relative to the two loci defining the interval of interest. That is, a gamete is only considered to be recombinant on the basis of the grandparental origins of the gamete at the two loci, regardless of how many "actual" recombination events occurred along the chromosomal strand between the two loci. That is why a recombination event is only observed if there is an odd number of crossovers between the two loci. Thus, Casares claim that "for 4 crossovers, ... practically all gametes are recombinant" is false. Indeed, Lange (2002) proves that, under the assumption of no chromatid interference, the probability that a gamete is recombinant given n chiasmata in the interval is one half for all $n > 0$, and zero if $n = 0$ (Emerson and Rhoades 1933; Mather 1936; Mather 1938). So we have:

$$\begin{aligned}
 RF &= P(\text{crossover}) \\
 &= \sum_{n=0}^{\infty} P(\text{crossover} \mid n \text{ chiasmata}) P(n \text{ chiasmata}) \\
 &= \sum_{n=1}^{\infty} \frac{1}{2} P(n \text{ chiasmata})
 \end{aligned}$$

In contrast, Casares' map function appears to be given as (in the notation used here):

$$RF = \sum_{n=1}^{\infty} \left(1 - \left(\frac{1}{2}\right)^n\right) P(n \text{ chiasmata})$$

which implies that Casares is incorrectly assuming that

$$P(\text{crossover} \mid n \text{ chiasmata}) = \left(1 - \left(\frac{1}{2}\right)^n\right).$$

This assumption is false, as proven by Lange (2002).

Discussion

One of Casares' motivations for developing a new map function was the claim that the Haldane map function overestimates genetic distances. However, as shown above, this claim is incorrect, as it was based on mistakenly applying the inverse of Haldane's map function to a genetic distance. In addition, we have shown above that the derivation of the 'corrected' Casares map function is apparently based on an incorrect formula for the $P(\text{crossover} \mid n \text{ chiasmata})$.

Finally, Casares prefers the 'corrected' Casares map function to the Haldane map function simply because it generates shorter genetic distances than Haldane's does. However, there are a large number of map functions that, for a given recombination fraction, return a shorter genetic distance than that of Haldane's. Therefore the choice of map function should be based on more criteria than the simple one of returning a shorter genetic distance. These criteria should include what the map function assumes about stationarity, chiasma and chromatid interference, and whether or not the map function is multilocus feasible (Weeks et al. 1993; Weeks 1994; Weeks et al. 1994; Speed 1996).

In conclusion, since the derivation of Casares' map function is based on false assumptions and faulty reasoning, we suggest that it be promptly retired from use.

Acknowledgments

We would like to acknowledge the support of the University of Pittsburgh and NIH grant R01GM076667.

References

- Bailey, NTJ. Introduction to the mathematical theory of genetic linkage. Oxford: Clarendon Press; 1961.
- Casares P. A corrected Haldane's map function to calculate genetic distances from recombination data. *Genetica* 2007;129:333–338. [PubMed: 16900314]
- Emerson RA, Rhoades MM. Relation of chromatid crossing over to the upper limit of recombination percentages. *The American Naturalist* 1933;67:374–377.

- Haldane JBS. The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet* 1919;8:299–309.
- Lange, K. Mathematical and statistical methods for genetic analysis. Vol. 2nd edition. New York: Springer; 2002.
- Mather K. The determination of position in crossing-over. I. *Drosophila melanogaster*. *J Genet* 1936;33:207–235.
- Mather K. Crossing-over. *Biol Reviews Camb Phil Soc* 1938;13:252–292.
- McPeck, MS. An introduction to recombination and linkage analysis. In: Speed, TP.; Waterman, MS., editors. *Genetic mapping and DNA sequencing: IMA Volumes in Mathematics and its Applications*. New York: Springer-Verlag; 1996. p. 1-14.
- Speed, TP. What is a genetic map function?. In: Speed, TP.; Waterman, MS., editors. *Genetic mapping and DNA sequencing: IMA Volumes in Mathematics and its Applications*. New York: Springer-Verlag; 1996. p. 65-88.
- Speed, TP. *Encyclopedia of Biostatistics*. New York: John Wiley & Sons, Ltd; 2005. Genetic map functions.
- Weeks DE. Invalidity of the Rao map function for three loci. *Hum Hered* 1994;44:178–180. [PubMed: 8039803]
- Weeks DE, Lathrop GM, Ott J. Multipoint mapping under genetic interference. *Hum Hered* 1993;43:86–97. [PubMed: 8359820]
- Weeks DE, Ott J, Lathrop GM. Detection of genetic interference: simulation studies and mouse data. *Genetics* 1994;136:1217–1226. [PubMed: 8005426]