Mathematical function for predicting recombination from map distance

Mikko Kivikoski[a,1,2], **Pasi Rastas**[b,1], **Ari Löytynoja**[b], **and Juha Merilä**[a,c]

[a]Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, FI-00014 University of Helsinki, Finland; [b]Institute of Biotechnology, HiLIFE, FI-00014 University of Helsinki, Finland; [c]Research Division for Ecology and Biodiversity, School of Biological Sciences, Kadoorie Biological Science Building, Pokfulam Road, The University of Hong Kong, Hong Kong, SAR

Map distance, measured in centimorgans, is used in genetics to represent the average number of crossovers between two loci in a gamete and can be translated to recombination frequency with mapping functions. Commonly used mapping functions, such as Haldane or Kosambi functions, are single-variable functions and yield one recombination prediction per map distance. Here we show that these approaches contain systematic error and formulate a new function to predict recombination frequency from the map distance. Our new function is paradigm shifting as it takes into account the variation in the number of crossovers and their localisation, yielding context-specific estimates of recombination frequency that can vary for a given map distance. The function is chromosome-specific, making it generalizable for different sexes and organisms. We validate the new approach using empirical data from stickleback fishes and humans and show that it outperforms the conventional mapping functions and produces more accurate predictions about recombination frequency between loci.

Crossover interference | *Gasterosteus aculeatus* | *Homo sapiens* | Mapping function | *Pungitius pungitius* | Recombination | Stickleback

Crossovers in meiosis break physical linkage among loci and allow formation of recombinant chromosomes. The recombination frequency of two loci – i.e. the likelihood of an odd number of crossovers in the gamete (1) – can be estimated empirically by studying offspring phenotypes (2), chiasmata under a microscope (3) or with genetic markers (4). Use of genetic markers has become the standard and applicable also for non-model organisms such as cattle (5), soay sheep (6), red deer (7), nine-spined stickleback (8) and the three-spined stickleback (9). Although recombinations can be investigated empirically, data for that are laborious to produce and researchers utilizing recombination frequency may rely on earlier reported map distances. Map distances represent the average number of crossovers in the gamete and are measured in units of centimorgans (cM) (10). A known map distance is converted to a recombination frequency between the markers using mapping functions (1). Essentially, these functions aim to account for the non-additivity of the recombination frequencies and, by making certain assumptions about crossover localisation, provide a generalized association between the map distance and recombination frequency.

Probably the two most widely used mapping functions are those of Haldane (11) and Kosambi (12) (e.g. (1, 13, 14)). The limitations of these mapping functions have been discussed in the literature and alternatives have been put forward (10, 15), but to the best of our knowledge, they have not been adopted to wider use (see (16, 17), for example). Moreover, we are not aware of any attempts to re-assess old mapping functions or formulation of new functions in the context of contemporary high-density marker data sets such as available for humans (18) and sticklebacks (19) for instance.

Here we introduce a novel mathematical function for predicting the recombination frequency between two loci of known map distance. Our function builds on three findings: (1) Recombination frequency, r, can be expressed as a function of probability of no crossovers between the markers in the bivalent, p_0, so that $r = \frac{1}{2}(1-p_0)$ (13, 17, 20). (2) Due to crossover interference, locations of crossovers in the chromosome depend on their number, and follow a certain periodicity so that the crossovers tend to be evenly distributed (21, 22). (3) The number of crossovers in meiosis varies across organisms, sexes and individuals and the probabilities for different numbers of crossovers can be estimated from the data (23).

For each marker pair, the estimate of p_0 considers the joint-effect of the marker positions and the probabilities for different numbers of crossovers in the chromosome (Fig. 1). As a result, the prediction depends on the location of the markers in the chromosome, and the recombination frequency can be different for two marker pairs with the same map distance or the same for two marker pairs with different map distances. We demonstrate the performance of the new function with high density sex-specific linkage maps of the nine-spined stickleback (*Pungitius pungitius*), the three-spined stickleback (*Gasterosteus aculeatus*) and human (*Homo sapiens*) crossover data. We compare the performance of the new function with that of the three conventional mapping functions: linear (hereafter called the Morgan function), Haldane and Kosambi.

Materials and Methods

Mathematical function for the absence of crossovers. Following the principle of Weinstein (24) and the idealized example (Fig. 1), we formulated a new function for predicting the likelihood of no crossovers between the markers in the bivalent, p_0, given the map distance. This function accounts for alternative numbers of the crossovers in the chromosome. The recombination frequency, r, is then derived with equation $r = \frac{1}{2}(1-p_0)$. Let d be the map length of the whole chromosome. If there are k (k>0) crossovers in the bivalent, they are assumed to occur in k non-overlapping regions of equal size, $\frac{d}{k}$, so that every crossover occurs in a different region. Within each region, the localisation is uniformly distributed and the localisations are independent across the regions. For a given marker pair, $p_0 = \sum_{k=0}^{n} P(\text{no crossovers}|k) = \sum_{k=0}^{n} p_k p_0(k)$, where n is the highest possible number of crossovers in the bivalent, p_k is

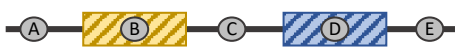
M.K. formulated the mathematical function introduced here, executed the data analyses and was responsible for designing the manuscript. P.R. made the stickleback linkage maps and was involved in the formulation of the mathematical function. A.L. and J.M. supervised the study and participated in writing the manuscript.

The authors declare no competing interests.

[1]M.K. and P.R. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: mikko.kivikoskihelsinki.fi

A



B

	A-B, B-C, C-D, D-E	A-C, C-E	A-D, B-E	A-E	B-D
p_0	0.5	0	0	0	0.25
p_1	0.5	1	0.5	0	0.5
p_2	0	0	0.5	1	0.25
Recombination frequency	0.25	0.5	0.5	0.5	0.375
Additive map distance (cM)	25	50	75	100	50
Haldane map distance (cM)	34.7	Inf	Inf	Inf	69.3
Kosambi map distance (cM)	27.5	Inf	Inf	Inf	48.7

C

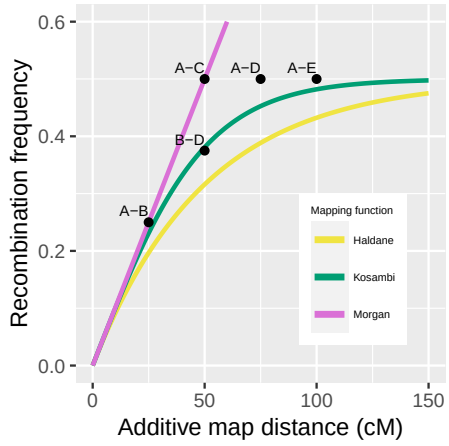


Fig. 1. Number and location of crossovers affects the recombination frequency. (A) An example of a chromosome with always two crossovers *in the bivalent* so that one occurs in the yellow and the other one in the blue area. The locations of the crossovers within their distinctive regions are independent. (B) The table shows the probability for 0, 1, or 2 crossovers between the markers (p_0, p_1 and p_2, respectively) *in the bivalent*, the resulting recombination frequency of the marker pair (*in the gamete*) and the estimated map distance for every marker pair using different mapping functions. (C) Graph showing the relationship between the recombination frequency and map distance with different mapping functions. Notably, intervals A-B, A-C and A-D have the same recombination frequencies but different map distances, whereas, marker pairs A-C and B-D have the same map distances but different recombination frequencies.

the probability for k crossovers in the bivalent and $p_0(k)$ is the probability of no crossovers between the markers when there are k crossovers in the bivalent. For markers at map positions m_i and m_j ($m_j \geq m_i$), $p_0(k)$ is defined as:

$$p_0(k) = \begin{cases} 1 & \text{if } k = 0 \\ 1 - \frac{|m_j - m_i|}{\frac{d}{k}} & \text{if } \left\lceil \frac{m_j}{\frac{d}{k}} \right\rceil = \left\lceil \frac{m_i}{\frac{d}{k}} \right\rceil \text{ and } k > 0 \\ \left(1 - \frac{|b_i - m_i|}{\frac{d}{k}}\right) \frac{|b_j - m_j|}{\frac{d}{k}} & \text{if } \left\lceil \frac{m_j}{\frac{d}{k}} \right\rceil - \left\lceil \frac{m_i}{\frac{d}{k}} \right\rceil = 1 \text{ and } k > 0 \\ 0 & \text{if } \left\lceil \frac{m_j}{\frac{d}{k}} \right\rceil - \left\lceil \frac{m_i}{\frac{d}{k}} \right\rceil > 1 \text{ and } k > 0, \end{cases} \quad [1]$$

where $\left\lceil \frac{m_i}{\frac{d}{k}} \right\rceil$ and $\left\lceil \frac{m_j}{\frac{d}{k}} \right\rceil$ are the crossover regions of the markers m_i and m_j, and b_i and b_j are the upper boundaries for these regions, so that $b_i = \left\lceil \frac{m_i}{\frac{d}{k}} \right\rceil \frac{d}{k}$ and $b_j = \left\lceil \frac{m_j}{\frac{d}{k}} \right\rceil \frac{d}{k}$.

This function is also applicable for multiple loci. In case of three markers (m_i, m_j, m_k), the recombination frequencies derived with this function meet the criteria $r_{ij} + r_{jk} \geq r_{ik}$, where r_{ij} is the the recombination frequency between markers m_i and m_j (16, 25) (see Supplementary material for proofs).

We applied the function for sex-specific recombination data in the 21 chromosomes of the nine- and three-spined sticklebacks, and in the 22 human autosomes. The total map lengths were derived from the linkage maps and the likelihoods for different numbers of crossovers in the bivalent were inferred from the observed number of crossovers (see 'Inference of crossover frequency' below)

Stickleback linkage maps. The crossover frequencies and locations were estimated from the linkage maps described in Kivikoski et al. (19). For the nine-spined stickleback (*Pungitius pungitius*), high density (22,468 informative markers) linkage maps were reconstructed with Lep-MAP3 software (26) from a data set of 938 F_1-offspring. The parental fish, 46 females and 87 males, were wild caught individuals from Finland (Helsinki, 60°13'N, 25°11'E) which were artificially crossed in laboratory to produce the aforementioned F_1-offspring (19), see also (8). Five females were each crossed with a different male, forming five full-sib families and other 41 females were each crossed with two different males, which formed 41 half-sib families. For the three-spined stickleback (*Gasterosteus aculeatus*), linkage maps were generated with Lep-MAP3 and based on 517 F_1-offspring from 60 families (30 males and 60 females, so that each male was crossed with two different females) as described in Kivikoski et al. (19). The parents were caught from the Baltic Sea and the crossing scheme and their sequence data are described in detail in Pritchard et al. (27).

For all linkage maps (maternal and paternal maps of the nine- and three-spined stickleback), the genetic distance between the adjacent markers was calculated from recombination frequency with the Haldane mapping function. For non-adjacent markers the distances are additive. There were, on average, 1,070 and 1,342 markers per chromosome in the nine- and three-spined stickleback maps, respectively. Hence, the inter-marker distances were short: on average 19,571 bp corresponding to 0.054 cM and 0.106 cM in the paternal and maternal maps of the nine-spined stickleback, and 15,461 bp corresponding to 0.043 cM and 0.075 cM in the paternal and maternal maps of the three-spined stickleback, respectively. As all conventional functions yield very similar results for small recombination frequencies, the choice of the mapping function has a minor impact on the map distances.

Analysis of human recombination data. To evaluate the general applicability of the new function, we analysed human data from Halldorsson et al. (18). This consisted of sex-specific linkage maps (their supplementary Data S1 and S2) and the crossovers data (their supplementary Data S4). Crossover locations and counts were obtained from the column 'medsnp' of the sex-specific linkage maps for the 41,092 probands with both paternal and maternal crossover information. All crossovers were used irrespective of their status regarding the gene conversions (complex, non-complex or not assessed; see Halldorsson et al. (18)). Moreover, no probands were discarded based on the total number of crossovers in them: the highest number of crossovers per proband per chromosome was 17 maternal crossovers in chromosome 13. The number of markers in the linkage maps was 17,894-90,036 depending on the chromosome. We sampled 1.5% of the markers of every chromosome (seed value 2021 in R (28) was used to ensure reproducibility), which yielded 268–1,351 markers, i.e. 35,778–911,925 marker pairs, per chromosome. For every marker pair, we estimated the sex-specific recombination frequency by calculating the proportion of the studied probands (n=41,092) with an odd number of crossovers between the markers.

Assessment of mapping function performance. The three mapping functions (Morgan, Haldane, Kosambi) and the new function formulated here are based on certain assumptions of crossover localisation and are not fitted to the observed data of recombination frequencies and genetic distances. As parameter fitting is not involved, the models cannot be compared with likelihood ratio test or Akaike information criterion (AIC). Instead, the performance of the four functions were assessed by calculating the likelihood of the empirical data under each function. The likelihood of the data (i.e. empirical recombination frequency of a marker pair) was calculated as the likelihood of binomial distribution so that the number of trials corresponded to the number of offspring studied (n=938, n=517,

n=41,092 in the three species studied), the number of successes to the observed number of recombinants and the likelihood of success under the null-hypothesis to the recombination rate predicted by the mapping function. For human data, the number of trials and successes were divided by 100 to avoid numerical errors in R programme. The Morgan function was applied here so that the predicted recombination frequency was the map distance in centimorgans divided by 100 or 0.5 for map distances greater or equal to 50 cM. Instead of letting the prediction increase linearly with mapping distance, recombination frequencies above 0.5 typically do not exist (apart from sampling error) and this formulation was chosen to get a more impartial assessment of the method. We assessed separately every chromosome in both maternal and paternal datasets. The mapping function with the highest likelihood (the sum of log-likelihoods) was considered the best. These results are reported for the whole genome pooled and separately for each chromosome. For chromosome level comparisons we also report the p-values of binomial tests where the number of successes is the number of chromosomes where the new function outperforms the three others. The null-hypothesis is that the new function is equally good as the best of the three others (Haldane, Kosambi, Morgan), and the probability of success under the null-hypothesis is 0.5.

Inference of crossover frequency. According to Weinstein (24), probability for observing k crossovers in a randomly sampled meiotic product, depends on the number of crossovers in the bivalent so that $P(k) = \frac{\binom{n}{k}}{\sum_{i=0}^{n}\binom{n}{i}}$, where n is the number of crossovers in the bivalent and $0 \leq k \leq n$. Number of crossovers in the bivalent varies, not only between the sexes and chromosomes, but also between individuals and individual meioses (see ref. (29) for an example). Therefore, the observed crossovers in the gamete pool are a sample of crossovers from meioses with a different number of crossovers in the bivalent. As the sampling function above is known, the multinomial distribution of the number of crossovers in the bivalent can be estimated. We used the expectation-maximization (EM) algorithm of Yu and Feingold (23) to estimate the multinomial distribution for different numbers of crossovers in bivalent. The algorithm approximates the multinomial distribution that maximizes the likelihood of the data consisting of the numbers of meiotic products with 0...N observed crossovers in a given chromosome. According to Yu and Feingold (23), for data where the highest number of observed crossovers in a single meiotic product is N, it is sufficient to estimate the multinomial distribution between 0 and $2N-1$. The algorithm was applied separately for maternal and paternal crossovers and for each chromosome, by pooling all meiotic products (n=938, n=517, n=41,092 for nine- and three-spined stickleback and human data, respectively). The estimated multinomial distribution includes the maximum-likelihood estimate for no crossovers in the bivalent. We applied the bootstrapping test of Yu and Feingold (23) to estimate if, in case of an estimate above zero, the deviation from that is statistically significant ($p<0.05$). For the chromosomes with non-significant p-values, we used a restricted multinomial distribution, that restricts the likelihood of no crossovers to zero, for further analyses. This choice was made in order to assume the obligate crossover as a null-hypothesis. Chromosomes with $p<0.05$ were chromosome 15 of the three-spined stickleback (paternal meioses) and in humans chr21 and chr22 (maternal and paternal meioses) and chr3 (paternal meioses).

Data Availability. The stickleback linkage maps and computer code for replication of the analyses of this study are available in Github https://github.com/mikkokivikoski/recombinationStudies. In order to facilitate reproducibility of the analyses, analytical steps following the linkage map reconstruction step are integrated into an Anduril2 workflow (30).

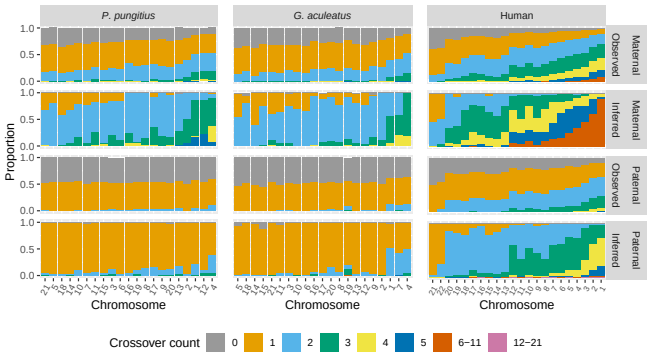


Fig. 2. Observed and inferred crossover frequency distributions in the nine-spined stickleback (*P. pungitius*), three-spined stickleback (*G. aculeatus*) and human (*H. sapiens*). Each bar shows the proportion of offspring (n=938, n=517, n=41,092) with a certain number of crossovers (Observed) and inferred proportions of meiosis with a certain number of crossovers in the bivalent in maternal and paternal meioses (Inferred). Chromosomes are ordered from shortest to longest (in base pairs). Crossover counts 6–11 and 12–21 are grouped for readability.

Results.

Association between recombination frequency and map distance is context dependent. The idealized example (Fig. 1) leads to the conclusion that the association between recombination frequency and map distance is ambiguous, i.e. two marker pairs with the same map distances may have different recombination rates and two marker pairs with the same recombination rates may have different map distances. This is the case especially when there are multiple non-randomly located crossovers in the bivalent. Empirical data from the three species confirms this: females of all three species typically have two or more crossovers in the bivalent which is also the case in paternal meioses in the longest stickleback chromosomes and most human autosomes (Fig. 2). Furthermore, when multiple crossovers occur in the same bivalent, they are clearly separated (Fig. S1). Consequently, the association between the recombination frequency and map distance is ambiguous, especially in maternal data (Fig. 3, see also Fig. S3, S6, S9).

The new function outperforms the old ones. The predictions made with the new function resembled the empirical data (Fig. 3, see also Fig. S4, S7, S10). For paternal recombinations in sticklebacks, the predictions were more or less linear as there is typically one crossover per chromosome (Fig. 2). For human data and maternal recombinations in the sticklebacks, the predictions had a fan-shaped pattern as there are typically multiple crossovers per chromosome (Fig. 2). Notably, the three conventional mapping functions seemed to have poor fit with the maternal data. Interestingly, with paternal stickleback data, the linear Morgan function fitted the data quite closely (Fig. 3).

When taking all chromosomes into account, the new function yielded the highest likelihood in both sexes and all three species (Table 1). The number of marker pairs differs between species and therefore the sums of log-likelihoods are in different scales. In the chromosome-by-chromosome comparison, the new mapping function had the best performance in all chromosomes with human data and most chromosomes with maternal stickleback data (Table 2, binomial test $p < 0.05$). However, in paternal stickleback data, Morgan function had

the best performance in 13 nine-spined and 11 three-spined stickleback chromosomes (Table 2, see also Fig. S2).

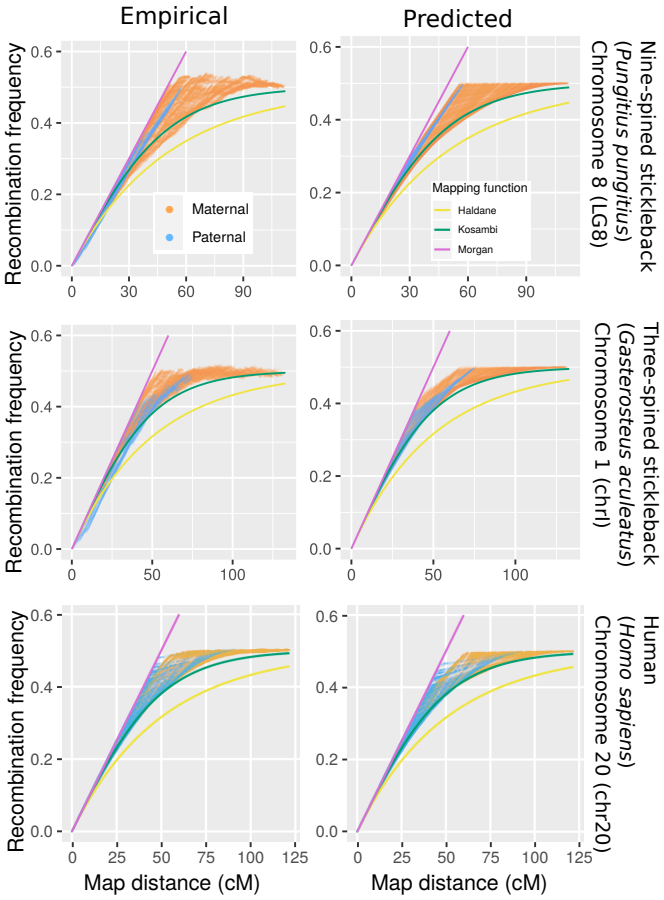


Fig. 3. Examples of empirical recombination frequencies (left) and predictions by the new function (right) in the nine-spined stickleback (top), three-spined stickleback (middle) and human (bottom). Solid lines show the three conventional mapping functions. Each orange and blue dot is a marker pair in maternal and paternal data, respectively.

Table 1. Sum of the log-likelihoods of the full data.

Species	Sex	Haldane	Kosambi	Morgan	$p_0(k)$
Nine-spined stickleback	Maternal	-41,506,139	-15,154,609	-16,674,480	**-12,015,300**
	Paternal	-11,604,079	-5,364,491	-4,531,728	**-4,300,082**
Three-spined stickleback	Maternal	-11,545,404	-5,194,577	-5,071,404	**-4,328,440**
	Paternal	-2,609,774	-1,588,307	-1,632,652	**-1,370,573**
Human	Maternal	-49,610,858	-25,873,948	-27,776,004	**-23,790,878**
	Paternal	-55,057,174	-26,551,010	-27,444,547	**-22,935,403**

Table 2. Best performance on individual chromosomes

Species	Sex	Haldane	Kosambi	Morgan	$p_0(k)$	p-value†
Nine-spined stickleback	Maternal	0	0	1	20	**<0.0001**
	Paternal	0	0	13	8	0.38
Three-spined stickleback	Maternal	0	0	4	17	**0.0072**
	Paternal	0	0	11	10	1
Human	Maternal	0	0	0	22	**<0.0001**
	Paternal	0	0	0	22	**<0.0001**

† p-values are for two-tailed binomial test.

Discussion. Genetic recombination is a fundamental characteristic of sexual reproduction. Knowledge about the recombination tells how often two loci are inherited in the same molecule, which is essential for identifying loci under selection (31) and understanding evolution in general. Here, we have formulated a novel mathematical function for predicting recombination frequency from the map distance. Importantly, we have shown with empirical data from sticklebacks and humans that the association between recombination frequency and map distance varies considerably when there are multiple crossovers in the bivalent. This is because map distance characterizes the expected number of crossovers but does not account for the variation around the expectation and, in particular, for the probability of no crossovers. The new function implements these characteristics and outperforms Kosambi, Haldane and Morgan functions in predicting recombination frequency from the map distance.

When applied to the whole genome, the function proposed here yielded more accurate predictions for recombination frequency than any of the three classical functions in the nine-spined stickleback, three-spined stickleback and humans. Importantly, the predicted recombination frequencies varied within the same map distance, as they did in the empirical data (Fig. 3). At the level of a single chromosome, simple linear Morgan function yielded higher likelihood in paternal data of most stickleback chromosomes. However, in those chromosomes the differences to the new function were small and therefore the new function yielded the highest likelihood when all chromosomes were pooled (Table 1). Stickleback males typically have just one crossover in the bivalent (the longest chromosomes being notable exceptions, Fig. 2) and in those cases the additive map distance matches well with the recombination frequency, which makes the Morgan function to perform better. The function formulated here has a tendency to underestimate the recombination frequency when the likelihood for one crossover in meiosis is high (such as in male sticklebacks, Fig. 2). This comes from the model formulation where the size of the crossover region is defined as d/k, which means that when the recombination frequency is predicted for meiosis with one crossover (k=1), the denominator (d/k) becomes d, which is the map length of the chromosome. In chromosomes where multiple crossovers are rare and the crossover localisation depends on the number of crossovers, 50 cM would be closer to the 'correct' denominator value instead of the map length which is greater than 50 cM. Nonetheless, our empirical examples demonstrate the versatility of the new function and, by taking into account the map length of the chromosome and probabilities for different number of crossovers in the bivalent, it works for species, sexes and chromosomes with very different crossover rates.

Localisation of crossovers in a chromosome is not random and the process is influenced by crossover hotspots (32, 33), chromosomal inversions (34), sex-determining regions (35), repelling effects of centromere and telomere (9, 36, 37) as well as crossover interference. There are multiple approaches for modelling and characterizing crossover interference (reviewed in (38)). In the new function, the crossover interference is accounted for by adjacent crossovers appearing in close proximity less frequently than would be expected by chance. This

is implemented in the function by assuming that when there are multiple crossovers, they occur in distinctive regions of the chromosome (see Methods and Equation 1). This is motivated by the observation of periodicity in locations of multiple crossovers (Fig. S1), similar to those reported for example by Charles (21). We think this approach is useful as it does not require complex mathematical formulation for the crossover interference and with the resolution of modern data, the number of crossovers in the bivalent can be studied with high precision.

In order to apply the new function, map length of the chromosome and likelihoods for different numbers of crossovers are needed. Ideally, the multinomial distribution of different numbers of crossovers would be estimated from empirical data as done here. However, if such data were not available, one could approximate the multinomial distribution by a linear combination of multiples of 50 cM (one crossover in the bivalent) that yields the chromosome's empirical map length d. This idea can be demonstrated by defining the linear combination so that the possible crossover counts are the largest multiple of 50 cM smaller than the chromosome map length $\lfloor d/50 \rfloor$, and that added by one $\lfloor d/50 \rfloor + 1$. The probabilities for those crossover counts are then $1 - (d/50 - \lfloor d/50 \rfloor)$ and $(d/50 - \lfloor d/50 \rfloor)$, respectively. For example, for map length d=110 cM one would assume an outcome of two or three crossovers in the bivalent with probabilities of 0.8 and 0.2, respectively. This solution is simple to derive, but an oversimplification for chromosomes where the crossover count varies between multiple values, which is the case in humans and especially for maternal crossovers in sticklebacks (Fig. 2). However, this approximation works reasonably well in paternal data and for some maternal chromosomes as well (Table S1, Fig. S5, S8, S11).

Relatedness between siblings (or other kinship excluding parent-offspring relationship) is a random variable, with expected value being the pedigree relatedness which is 0.5 for full-sibs of unrelated parents. The variance in the relatedness decreases with increasing recombination because the more and shorter the recombining units, the less there is variance in relatedness. Although this has been acknowledged in published estimators of relatedness variance (39), they implement the crossover localisation (and their number) in terms of map distance. This is not problematic only because it does not translate into base pairs (14), but because the recombination and map distance do not have an unambiguous relationship as shown here. Veller et al. (14) reported that the map distance–recombination frequency relationship relying on Kosambi function did not match the empirical data and, consequently, the predicted variance in relatedness did not match with their empirical results. We anticipate that the underlying reason for this is the one explained here: in high-density linkage maps (such as the ones used in this paper), the map distance between adjacent markers is practically equivalent to recombination frequency, as the possibility for multiple crossovers between closely positioned loci is negligible. Hence, the map distances are calculated only for short intervals and, over longer intervals they are additive. As a result, the map distance over longer intervals does not necessarily predict the recombination frequency, as it is not measured in such units.

The resolution of linkage maps keeps increasing (compare e.g. maps in Rastas et al. (8) and Kivikoski et al. (19); or Kong et al. (4) and Halldorsson et al. (18)), and there is an urgent need to re-think the meaning and implication of map distance. Here we provide a solution that is mathematically simple, scales and works for modern data from multiple species. We anticipate that our approach can stimulate further research on mapping functions. For example, it could be of interest to know how centromere interference and chromosome or species specific characteristics could be incorporated into them.

ACKNOWLEDGMENTS. Our research was supported by Academy of Finland (grants: 129662, 134728 and 218343 to JM; 322681 to AL; 343656 to PR), the Doctoral Programme in Wildlife Biology Research (LUOVA; funding to MK) as well support from the Helsinki Institute for Life Sciences (HiLife; grant to JM). We thank the Biodata Analytics Unit of the University of Helsinki, and Jukka Siren in particular, for advice in statistics. We wish to acknowledge CSC – IT Center for Science, Finland, for access to computational resources.

1. M Lynch, B Walsh, *Genetics and Analysis of Quantitative Traits*. (Sinauer Associates, Sunderland, MA), (1998).
2. AH Sturtevant, The behavior of chromosomes as studied through linkage. *Z. Abstam. Vererbung.* **13**, 234–287 (1915).
3. M Hultén, Chiasma distribution at diakinesis in the normal human male. *Hereditas* **76**, 55–78 (1974).
4. A Kong, et al., Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
5. L Ma, et al., Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLOS Genet.* **11**, e1005387 (2015).
6. SE Johnston, C Bérénos, J Slate, JM Pemberton, Conserved genetic architecture underlying individual recombination rate variation in a wild population of soay sheep (*Ovis aries*). *Genetics* **203**, 583–598 (2016).
7. SE Johnston, J Huisman, JM Pemberton, A genomic region containing REC8 and RNF212B is associated with individual recombination rate variation in a wild population of red deer (*Cervus elaphus*). *G3: Genes, Genomes, Genet.* **8**, 2265–2276 (2018).
8. P Rastas, FCF Calboli, B Guo, T Shikano, J Merilä, Construction of ultradense linkage maps with Lep-MAP2: Stickleback F_2 recombinant crosses as an example. *Genome Biol. Evol.* **8**, 78–93 (2016).
9. JM Sardell, et al., Sex differences in recombination in sticklebacks. *G3: Genes, Genomes, Genet.* **8**, 1971–1983 (2018).
10. H Zhao, TP Speed, On genetic map functions. *Genetics* **142**, 1369–1377 (1996).
11. JBS Haldane, The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**, 299–309 (1919).
12. DD Kosambi, The estimation of map distances from recombination values. *Annals Eugen.* **12**, 172–175 (1943).
13. TP Speed, *Genetic Map Functions*. (American Cancer Society), (2005).
14. C Veller, NB Edelman, P Muralidhar, MA Nowak, Variation in genetic relatedness is determined by the aggregate recombination process. *Genetics* **216**, 985–994 (2020).
15. P Casares, A corrected Haldane's map function to calculate genetic distances from recombination data. *Genetica* **129**, 333–338 (2007).
16. DE Weeks, Invalidity of the Rao map function for three loci. *Hum. Hered.* **44**, 178–180 (1994).
17. DE Weeks, X Tang, AM Kwon, Casares' map function: no need for a 'corrected' Haldane's map function. *Genetica* **135**, 305–307 (2009).
18. BV Halldorsson, et al., Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363** (2019).
19. M Kivikoski, P Rastas, A Löytynoja, J Merilä, Automated improvement of stickleback reference genome assemblies with Lep-Anchor software. *Mol. Ecol. Resour.* **21**, 2166–2176 (2021).
20. K Mather, Crossing-over. *Biol. Rev.* **13**, 252–292 (1938).
21. DR Charles, The spatial distribution of cross-overs in *X*-chromosome tetrads of *Drosophila melanogaster*. *J. Genet.* **36**, 103–126 (1938).
22. L Zhang, Z Liang, J Hutchinson, N Kleckner, Crossover patterning by the Beam-Film model: Analysis and implications. *PLoS Genet.* **10**, e1004042 (2014).
23. K Yu, E Feingold, Estimating the frequency distribution of crossovers during meiosis from recombination data. *Biometrics* **57**, 427–434 (2001).
24. A Weinstein, The theory of multiple-strand crossing over. *Genetics* **21**, 155–199 (1936).
25. S Karlin, U Liberman, Classifications and comparisons of multilocus recombination distributions. *Proc. Natl. Acad. Sci.* **75**, 6332–6336 (1978).
26. P Rastas, Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726–3732 (2017).
27. VL Pritchard, et al., Regulatory architecture of gene expression variation in the threespine stickleback *Gasterosteus aculeatus*. *G3: Genes, Genomes, Genet.* **7**, 165–178 (2017).
28. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), (2018).
29. KW Broman, JL Weber, Characterization of human crossover interference. *The Am. J. Hum. Genet.* **66**, 1911–1926 (2000).
30. A Cervera, et al., Anduril 2: upgraded large-scale data integration framework. *Bioinformatics* **35**, 3815–3817 (2019).
31. TR Booker, S Yeaman, MC Whitlock, Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Mol. Ecol.* **29**, 4274–4279 (2020).

32. L Kauppi, AJ Jeffreys, S Keeney, Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* **5**, 413–424 (2004).
33. F Baudat, et al., PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
34. MJ Thompson, CD Jiggins, Supergenes and their role in evolution. *Heredity* **113**, 1–8 (2014).
35. D Charlesworth, Evolution of recombination rates between sex chromosomes. *Philos. Transactions Royal Soc. B: Biol. Sci.* **372**, 20160456 (2017).
36. PB Talbert, S Henikoff, Centromeres convert but don't cross. *PLoS Biol.* **8**, e1000326 (2010).
37. Q Haenel, TG Laurentino, M Roesti, D Berner, Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* **27**, 2477–2497 (2018).
38. SP Otto, BA Payseur, Crossover interference: Shedding light on the evolution of recombination. *Annu. Rev. Genet.* **53**, 19–44 (2019).
39. N Risch, K Lange, Application of a recombination model in calculating the variance of sib pair genetic identity. *Annals Hum. Genet.* **43**, 177–186 (1979).