Statistical analysis and simulations that account for simultaneous effects of positive, negative, and no crossover interference in multilocus recombination data

Shaul Sapielkin
shaul@evo.haifa.ac.il
University of Haifa
Eyal Privman
University of Haifa
Abraham B. Korol
University of Haifa
Zeev Frenkel
University of Haifa

Research Article

Keywords: Genetic recombination, crossover interference, coefficient of coincidence, extension of the gamma-sprinkled model, negative crossover interference

Posted Date: July 23rd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7129389/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on December 13th, 2025. See the published version at https://doi.org/10.1186/s12864-025-12313-2.

Original article

Statistical analysis and simulations that account for simultaneous effects of positive, negative, and no crossover interference in multilocus recombination data

Shaul Sapielkin[1,2*#] [0000-0001-9375-3850], Zeev Frenkel[1,2*] [0000-0003-0211-6966]

Eyal Privman[1,2] [0000-0002-3186-7024], Abraham B. Korol[1,2#] [0000-0001-9411-7602]

[1]The Department of Evolutionary and Environmental Biology,

University of Haifa, Haifa, 3498838, Israel

[2]Institute of Evolution, University of Haifa, Haifa, 3498838, Israel

[*] = These authors contributed equally to this work

= corresponding authors

Email addresses: Shaul Sapielkin: shaul@evo.haifa.ac.il Abraham B. Korol: akorol@evo.haifa.ac.il

Abstract. Crossover interference (COI) is a widespread feature of homologous meiotic recombination. It can be quantified by the classical coefficient of coincidence (CoC), but this characteristic is highly variable and specific to the pair of chromosomal intervals considered. Several models have been proposed to characterize COI at the chromosomal level. In the *gamma* model, the strength of interference is characterized by a shape parameter *v*. In contrast, the *gamma-sprinkled* two-pathway model (GS) accounts for both interference-dependent and independent crossover (CO) events by fitting a mixture of gamma distributions with $v>1$ and $v=1$ with proportions 1-*p* and *p*. In reality, COI may vary along the chromosome, resulting in a poor fit of the employed model to the analysed data. Additional inconsistency can be caused by the common neglect of possible negative crossover interference in the model, although the phenomenon was previously reported for several organisms. We present an extension of the GS model to account for possible negative COI and provide suitable software for estimating the parameters of the extended GS model. On real chromosome data from a conifer (larch, *Larix principis-rupprechtii*), sugar beet (*Beta vulgaris*), and wheatgrass (*Thinopyrum intermedium*), we demonstrate a distinctive separation of three crossover (CO) processes on chromosomes: CO repulsion (positive COI), CO clustering (negative COI), and independent CO events (no COI).

Keywords: Genetic recombination, crossover interference, coefficient of coincidence, extension of the gamma-sprinkled model, negative crossover interference.

Introduction

Crossover interference (COI) refers to the non-random distribution of crossover (CO) events (chiasmata) along chromosomes during meiosis. COI is reflected in deviations of the observed distribution from the expected pattern under the hypothesis of COs independence. In other words, COI leads to a lower or higher rate of double COs in adjacent regions than expected assuming no interference between COs. The phenomenon was discovered in the early experiments on recombination in *Drosophila melanogaster* [1-3]. The degree of COI can be quantified by the coefficient of coincidence (*CoC*), the ratio of the observed to the expected number of double COs for a pair of chromosomal intervals: $CoC < 1$, $CoC > 1$, and $CoC = 1$, corresponding to positive, negative, and no COI, respectively. The nature of COI was explained in two ways: (i) as a result of interaction, either repulsion or clustering, of COs (referred to as chiasmata interference) [4], and (ii) as non-random involvement of chromatids in neighboring CO exchanges

(referred to as chromatid interference) [2]. Studies of many species

[5-11] have found no significant effect of chromatid interference on COI. Therefore, most proposed COI models assume no chromatid interference [8,12-14]. However, analysis in recent years of data from various organisms and mathematical modelling suggests that chromatid interference may exist, at least in some species [15,16]. However, at the level of genetic mapping analysis, it is impossible to distinguish between the contributions of crossover and chromatid interference, since direct testing of chromatid interference requires analysing all four products of meiotic division. In this article, we do not consider chromatid interference, assuming that it may be present only occasionally in some organisms/chromosomes.

To describe CO interference as a single mechanism of CO formation operating at the chromosome or even the entire genome level, one of the most convenient formulations is the gamma model. [12,17,18-20]. It considers the dependence of the number of COs per gamete or the distances between COs along a chromosome on the interference "strength" parameter (v). The values $v>1$ correspond to positive COI ("repulsion" between COs), while $v=1$ implies no interference. The gamma model was further modified in accordance with the discovery of the interference-independent metabolic pathway of meiotic recombination along with the common interference-dependent pathway [21]. Thus, the resulting "gamma-sprinkled" (GS) model is a mixture of two gamma distributions, with $v>1$ and $v=1$ (corresponding to *positive COI* and *no COI*) with proportions 1-p and p, respectively, appeared to be more consistent with empirical data [22]. The case of CO clustering along the chromosome (i.e. *negative COI*) can be modelled by gamma-distribution with $v<1$. Cases of possible negative COI in some genomic regions were reported in different organisms (*Neurospora*, yeast, *Drosophila*, *Arabidopsis*, maize, barley, wheat) [23-36]. Applying to such cases a standard GS model may result in biased estimates of both parameters, p and v, due to the presence of three rather than two components, with $v>1$, $v=1$, and $v<1$. Therefore, to account for possible negative COI, we extend the standard GS model by fitting a mixture of the corresponding three distributions.

Materials and Methods:

Statistical analysis of COI landscape using *CoC* calculations on high-density linkage maps

In general, for a pair of adjacent intervals on a genetic map flanked by three points (genetic markers, $m1$-$m2$-$m3$), the log-likelihood function for a testcross progeny can be presented as:

$$Log\text{-}likelihood\ (CoC) = \sum n_{ij} \cdot ln\ p_{ij}(CoC), \quad (1)$$

where p_{ij} are the probabilities of the classes of gametes: $p_{00} = (1\text{-}r_1\text{-}r_2\text{+}CoC\cdot r_1\cdot r_2)$ for no COs on both intervals, $p_{10} = (r_1\text{-}CoC\cdot r_1\cdot r_2)$ and $p_{01} = (r_2\text{-}CoC\cdot r_1\cdot r_2)$ for CO on the 1st and the 2nd intervals, respectively, and $p_{11} = (CoC\cdot r_1\cdot r_2)$ for double COs; while $r_{1=}\hat{r}_1$ and $r_{2=}\hat{r}_2$ are the recombination rates for the intervals *m*1-*m*2 and *m*2-*m*3 (calculated for each interval independently), and n_{ij} are the observed numbers of cases for the corresponding classes.

With typical modern datasets consisting of a large number of markers and a relatively small sample size, using too short intervals implies insufficient statistical power to distinguish between the alternatives $H_1(CoC\neq1)$ versus $H_0(CoC=1)$. Increasing the interval size by using only a subset of markers is wasteful and may lead to the problem of classifying a genotype as "CO" or "non-CO" for longer intervals, because genotypes with double COs in such intervals will be classified as "non-COs". A slight modification of the "CO genotype" definition for each new interval solves this problem. Specifically, we consider a genotype to be recombinant in an extended interval if it was formed as a result of one or more CO events within the corresponding subintervals (considering all markers). We refer to this type of classification as the "crossover superposition principle" (or CSP).

Accounting for heterogeneity in the COI landscape sliding windows on high-density linkage maps

- To estimate *CoC,* we maximize the likelihood function [37] using the simplex method [38,39] for three hypothetical situations: (1) according to Haldane mapping function, i.e. with *CoC*=1 [40]; (2) with COI according to the Kosambi [41] mapping function, i.e. with *CoC*=2*r*; and (3) with

arbitrary *CoC*. As a real-world example for illustration, we use the high-density linkage data for Norway spruce (Picea abies) published by Bernhardsson [42] (their cluster #3, segregating population with n=842 genotypes). For the analysis of interference, we constructed high-quality maps using the MultiPoint-ultradense software [43]. To classify the genotypes into groups 00, 01, 10, and 11, we used the approach described in the previous section with different sizes of the sliding window (l_1+l_2) along the chromosome, from 10 to 70 cM (which we considered a reasonable limit on the maximum l_1+l_2 map length). The obtained vectors of frequencies $P = (p_{00}, p_{01}, p_{10}, p_{11}) = P(l_1, l_2)$ were employed for local maximum likelihood estimation of *CoC* with sliding windows along the linkage map. We used a likelihood-ratio test [44] to compare the alternative COI models in each run (with an increasing sliding window size) and the Benjamini-Hochberg procedure [45] for false discovery rate correction. On the linkage group LG7 (according to the authors' designations) we found a region with relaxation of positive COI followed by no-COI and even negative COI at some map points (Fig.1). A candidate place with significant negative COI was detected, was represented by sliding windows of 14-15 cM length. These windows were formed by triplets of markers with different coordinates: left – at 79 cM, central – 82.1-83.5 cM, and right – 94.2-94.9 cM; with mean *CoC* = 2.39±0.68 (cumulative p-value = 0.013, calculated by Fisher's (1932) combined probability method with permutation correction to account for correlations between tests [47,48]. The permutation test was run with 1000 iterations, where haplotypes were randomly shuffled to generate an empirical distribution of the pooled statistics. The analysis of this region of the map with an increased sliding window size resulted in p-values that did not

allow us to distinguish between the obtained *CoC*>1 and the Haldane model (i.e., no COI); yet, positive interference was not detected. This case may serve as an example of heterogeneity of local chromosomal regions for crossover interference.

- Similar landscape patterns have been found by many authors in pericentromeric regions [24,27,29,49,50].

<<Figure 1 here>>

In order to characterize the chromosomal landscape of COI, we also performed a chromosome-wide analysis using simulations with CODA software [51], allowing us to estimate the parameters of the standard gamma-sprinkled model (Fig. 2). The solution obtained based on simulations points to the presence of a non-interfering component in the distribution of recombination events in the considered data, with a proportion $p = 0.29$. Likewise, the data showed an excess of gametes with CO number > 3 (Fig. 2, right) and an excess of short inter-crossover (CO-CO) distances (Fig. 2, left). The observed deviation of the CODA simulations with the obtained estimates of v and p parameters, i.e., the excess of both multiple exchanges and short inter-CO intervals, remarkably coincides with our finding of possible negative COI for the same considered linkage group (see Fig. 1). However, a more objective conclusion about how to interpret these deviations from the standard GS model requires an extension of the GS model allowing for the possibility of negative interference.

<<Figure 2 here>>

Extension of the GS model to account for negative COI:

COI model based on gamma distribution

Let crossover events occur on the chromosome with a distance from the previous CO event, depending only on the position of this previous event. Let the distances D on the genetic map between crossover events follow a gamma distribution with parameters v (shape) and θ (scale) [18]:

$$f_{v,\theta}(x) = \frac{1}{\Gamma(v)\theta^v} x^{v-1} e^{-x/\theta},\ v>0 \quad (2)$$

This model can be viewed as a generalization of the classical Haldane model, where CO events are independent so that the distances between successive events follow an exponential distribution:

$$f_{1,1/\lambda}(x) = \frac{1}{\Gamma(1)\cdot 1/\lambda} x^0 e^{-x\lambda} = \lambda e^{-\lambda x} \quad (3)$$

From $\Gamma(1, 1/\lambda) \sim Exp(\lambda)$ follows the Haldane model [40]. If $2v$ is an integer, then the gamma distribution coincides with χ^2 distribution with $2v$ degrees of freedom ($v = df/2$, $\theta = 2$):

$$f_{df/2,2}(x) = \frac{1}{\Gamma(df/2)2^{df/2}} x^{\frac{df}{2}-1} e^{-x/2} \quad (4)$$

With df=4, v=2, and θ=100/2, the gamma distribution coincides with the χ^2 distribution with df=2. Such distribution of distances between COs is close to the Kosambi model [41]. In general, the average and variance of the distances between consecutive COs fitting the gamma model can be presented as:

$$E(D) = \int_{-\infty}^{\infty} x f(x) dx = v\theta = \frac{\alpha}{\beta} \quad (5)$$

and

$$Var(D) = v\theta^2,\ \sim\Gamma(v,\theta) => cD \sim \Gamma(v, c\theta)\ , \quad (6)$$

where c is a scaling coefficient.

Genetic distance (in Morgans) is the expected number of CO events between two points on the genetic map; in terms of 'cM distances': $E(D)=100=v\theta$. Hence, $\theta=100/v$ and to specify the model we need only parameter v. The case with $v=1$ corresponds to the situation of no COI, $v>1$ implies positive COI, and $0<v<1$ - negative COI.

Based on the symmetry, the doubled distance from the chromosome "begin" to the first CO event should also follow the model (2). Hence, the distribution of this distance can be presented as:

$$g_{v,\theta}(x) = 2f_{v,\theta}(2x),\ v>0 \qquad (7)$$

Simulation of distances between two CO events based on gamma distribution with parameter *v*

Define $G_v(x)$ and $F_v(x)$ as distribution functions of the genetic distances from the "beginning" of the chromosome to the first CO event and from the position of the observed CO to the next CO under the gamma distribution model with the corresponding shape parameter, respectively:

$$G_v(x) := \int_0^x 2f_{v,100/v}(2t)dt = F_v(2x)$$
$$F_v(x) := \int_0^x f_{v,100/v}(t)dt = \frac{1}{\Gamma(v)(100/v)^v}\int_0^x t^{v-1}e^{-vx/100}dt \qquad (8)$$

Let u be a random variable with uniform distribution on the interval (0, 1): $U(0,1)$. We simulate D_v as $D_v := F_v^{-1}(u)$. For calculations of $F_v(x)$, we use the following approximation:

$$\int_a^b f(x)dx \approx \sum f(x_i)\Delta x \qquad (9)$$

Note that $\lim_{x\to 0} f_{v,\theta}(x) = \begin{cases} 0, & v \geq 1 \\ \infty, & 0 < v < 1 \end{cases}$

Therefore, this approximation is problematic for $v<1$ and small x_i. In such cases, we can use the following asymptotic:

$$e^{-\frac{vt}{100}} \approx 1 - \frac{vt}{100},$$

$$\int_0^x t^{v-1} e^{-vt/100} dt \approx \int_0^x t^{v-1} dx - \frac{v}{100}\int_0^x t^v dt = \frac{x^v}{v} - \frac{vx^{v+1}}{100(v+1)} \qquad (10)$$

We simulated CO events on a chromosome with genetic length L (scored in cM) based on equations 8-10 so that the CO map positions are values from $[0, L]$.

Let $u(i)$~U[0,1], $D(0) \coloneqq F1{-}1(u(0))$, $x(0) \coloneqq 0.5\, D(0)$. Then

$$D^{(j)} \coloneqq F_k{}^{-1}(u^{(j)}),\ x^{(j)} \coloneqq x^{(j-1)} + D^{(j)},\ j > 0 \qquad (11)$$

We stop the simulation when $x^{(j)} + D^{(j+1)} > L$. The resulting ordered set of values $x^{(i)}$ is denoted by vector X.

Simulation of CO events based on a mixture of three processes (with parameters v_0, v_1, and v_2)

Let three types of CO events obey the models with shape parameters v_0, v_1, and v_2, while events of one type do not depend on events of other types. Let p_1 and p_2 be the proportions of events of types 1 and 2 on the chromosome, $0 \le p_1, p_2 \le 1$, $p_1 + p_2 \le 1$. If $0<p_1, p_2, p_1+p_2<1$, then to maintain on average 0.01 CO rate per 1 cM, the following condition should hold: $p_1 v_1 \theta_1 + p_2 v_2 \theta_2 + (1 - p_1 - p_2) v_0 \theta_0 = 100$, so that

$$\theta_1 = \frac{100}{p_1 v_1},\ \theta_2 = \frac{100}{p_2 v_2},\ \theta_0 = \frac{100}{(1-p_1-p_2)v_0}. \qquad (12)$$

$$\text{Let } u^{(i)}\sim \text{U}[0,1],\ D_t^{(j)} \coloneqq F_{v_t,\theta_t}^{-1}\left(u^{(j)}\right),\ t = 1,2,3 \qquad (13)$$

Our simulations are based on iterations $x_t^{(n+1)} := x_t^{(n)} + D_t^{(n+1)}$ with $x_t^{(0)} = 0.5D_t^{(0)}$. Simulations continue while $x_t^{(n)} < L$. We unite the three sets of the obtained values and sort them in increasing order. The resulting vector is denoted by *X*.

Simulation of CO events with changing parameter *v* along the chromosome

Let parameter *v* vary along the genetic map: $v=v(x)$.

Define $F_{v(x)}(y) := \int_0^y f_{v(x),100/v(x)}(t)dt$. The distribution density of the distance to the next CO event (after the CO event at x_0) is

$$f(y) = f_{v(x0+y)}(y)\frac{1-F(y)}{1-F_{v(x0+y)}(y)}, \text{ where } F(y) = \int_0^y f(t)dt.$$

To simplify the calculations, we can use the following equation:

$$\frac{F'(y)}{1-F(y)} = \frac{f_{v(x0+y)}(y)}{1-F_{v(x0+y)}(y)}$$

By integration of this equation, we obtain the following formula for $F(y)$:

$$F(y) = F_{x0}(y) = 1 - \exp(-\int_0^y \frac{f_{v(x0+t)}(t)}{1-F_{v(x0+t)}(t)}dt) \qquad (14)$$

Analogously to the case with constant *v*, we simulate positions of CO events on the map [0, *L*] as: $D^{(0)} := F_0^{-1}(u^{(0)})$, $x^{(0)} := 0.5\,D^{(0)}$; and $D^{(j+1)} := F_{x^{(j)}}^{-1}(u^{(j+1)})$, $x^{(j+1)} := x^{(j)} + D^{(j+1)}$, where $u^{(j)}$~U[0,1] , $j \geq 0$.

Simulation study demonstrating the application of the extended GS model

As an illustration of the application of the described approach, we generated a testcross dataset with characteristics similar to the Norway spruce real data mentioned above (section 3). We simulated several populations (*n*=200, 500, 1000, and 5000) encompassing chromosome zones with varying *v,* including strong positive COI (*v*=5), Kosambi COI model (*v*=2), Haldane COI model (*v*=1; no COI), negative COI (*v*=0.5), and strong negative COI (*v*=0.2). We analysed

271 the distribution of *CoC* using two approaches: the "crossover superposition
272 principle" CSP (described in section 3) and standard classification (based on flank
273 markers). For population sizes N=200 and 500, we employed window sizes $l_1 + l_2$
274 = 20 cM and 30 cM, while 10 cM and 20 cM windows were employed for N=1000
275 and 5000, bearing in mind the increase in resolution with a higher sample size
276 (Fig. 3). We found that for cases of strong negative COI (NCOI), *CoC* estimates
277 based on the CSP classification are more accurate than those obtained from the
278 standard classification. This underscores the efficacy of the CSP approach in
279 improving the accuracy of *CoC* estimates, especially in the context of strong
280 NCOI. This approach will be most effective due to the ever-increasing trends in
281 improving the quality of the genetic mapping data (including higher marker
282 density and population size). Our simulations confirmed that for the LG7 Norway
283 spruce data considered in section 3 (population size n=842, map length ~150 cM),
284 using a sliding window size of ~15 cM, we could detect a robust zone of negative
285 COI with a possible transition to no COI. A notable drawback of employing *CoC*
286 calculations to assess local interference heterogeneity lies in their sensitivity to
287 the interval length, population size, and the local marker density in the target
288 region of the genetic map. To characterize the effect of NCOI along the genome,
289 between different genomes, and sexes, more effectively would be to employ a
290 model allowing simultaneous NCOI component alongside the "classical" positive
291 COI and no COI.

292 <<Figure 3 here>>

Statistical analysis of COI using the extended GC model:

Basic statistics for a single vector and a set of vectors of CO positions

Let X be a vector of consequent CO positions on a chromosome for an individual of a segregating population. The dimension of this vector ($\dim(X)\geq 0$) corresponds to the number of CO events for the individual. Let further denote the set of lengths between consecutive CO events by

$$vlIntPart(X) = \{x_{i+1} - x_i | i = 1,2, \ldots, \dim(X) - 1\};$$

if $\dim(X)\leq 1$ then the set $vlIntPart(X)$ is empty. Let $\boldsymbol{X}$ be the set of vectors X for the entire population. Then we denote by $N_m(\boldsymbol{X})$ the number of vectors X from set $\boldsymbol{X}$ such that $\dim(X) = m$, and $p_m := \hat{p}_m(\boldsymbol{X})$ is the proportion of vectors X with dimension m and $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{X}) = \boldsymbol{p}^{(\mathrm{obs})}(\boldsymbol{X}) = (p_0^{(obs)}, p_1^{(obs)}, \ldots)$ is a vector of the observed proportions of individuals with different numbers of COs. We unite all cases with $m \geq 5$ because such cases are extremely rare in most genetic mapping studies [52]. We denote the observed cumulative distribution function of distances between consequent COs within such subset of $\boldsymbol{X}$ as $F(x) = F(\mathrm{obs})(x) = F(\mathrm{obs})(x, \boldsymbol{X})$.

Distribution of the considered statistics

For a given chromosome length L and vector of parameters $V=(v_0, v_-, v_+, p_0, p_-, p_+)$, we generate a large population $\boldsymbol{X}_V$ of N independently simulated gametes (e.g., N=5000). Each gamete X_i is represented by a list of CO map positions $X_i = (x_1, \ldots, x_{n(i)})$ generated according to the method described in paragraph 4.3. Based on set $\boldsymbol{X}_V$ we estimate distributions $\boldsymbol{p}^{(V)} = \boldsymbol{p}(\boldsymbol{X}_V)$ and $F^{(V)}(x) = F(x, \boldsymbol{X}_V)$.

315 **Testing the model adequacy**

316 To compare alternative models for the test data, we employ the statistics $\boldsymbol{p}$ and F
317 described above and calculate statistical support (p-values) using standard
318 criteria: the χ^2-test and Kolmogorov-Smirnov test. To calculate these statistics,
319 first, we need to estimate map positions $X_j^{(i)}$ of all CO events observed on N
320 individuals. Then we estimate $\boldsymbol{p}$, based on the observed proportions of individuals
321 with different numbers of COs, and F, based on the observed distribution of
322 CO-CO distances scored on the subset of individuals with exactly two COs. These
323 statistics are used to compare two alternative hypotheses for the considered
324 dataset - H_0: "the data are consistent with the model with the defined parameters",
325 and H_1: "the data can be explained by some alternative model". For simplicity,
326 we only discriminate between models described in Section 4 with different sets
327 of parameters $V = (v_0, v_-, v_+, p_0, p_-, p_+)$, i.e., H_0: $V = V^{(0)}$ versus H_1: $V \neq V^{(0)}$. We
328 specifically focus on comparing the null hypothesis of the *absence of COI* with
329 different alternative hypotheses. These alternatives include scenarios with only
330 positive COI, only negative COI, a combination of positive COI and no-COI, a
331 combination of negative COI and no-COI, and a mixture involving all three
332 distribution components. For example, this way we can distinguish between the
333 GS model (p_-=0, i.e., no negative COI) and our extended GS model, which allows
334 for some negative COI.

335 To discriminate between these hypotheses, we use the likelihood ratio
336 test. Instead of the quite complicated likelihood function for the set $\boldsymbol{X}$, we used
337 likelihood functions for two simple statistics, $\boldsymbol{p}$ and $F(x)$, of this set. The
338 likelihood for the statistic $\boldsymbol{p}$ can be calculated based on the distribution function
339 for the statistics in the χ^2-test to compare two discrete distributions. The
340 likelihood for the statistic $F(x)$ can be calculated based on the distribution

function for the test statistic in the Kolmogorov-Smirnov test for the comparison of two continuous distributions [53].

χ^2 test based on the observed distribution of CO numbers: This test assesses the differences between the observed and expected proportions of CO number classes per chromosome. The test statistic,

$$X^2 = X^2(\boldsymbol{p}^{(\mathrm{obs})}, V) = N \sum_{m=0}^{5} \frac{\left(p_m^{(\mathrm{obs})} - p_m^{(V)}\right)^2}{p_m^{(0)}}$$

is computed using the distributions $\boldsymbol{p}^{(\mathrm{obs})}$ and $\boldsymbol{p}^{(V)}$ expected under model V. Under H_0, the test statistic X^2 should follow a *chi*-square distribution with 5 degrees of freedom (since we employed 6 categories for the observed CO number). The *p*-value is calculated as the probability of observing a value of X^2 that is as extreme or more extreme than the observed value: $p = \Pr(\chi_5^2 \geq X^2)$. For the defined set of the model parameters V, the distribution $\boldsymbol{p}^{(V)}$ can be estimated by intensive machine simulations. The likelihood function based on statistic $\boldsymbol{p}$ can be defined by

$$L_{\boldsymbol{p}}(\boldsymbol{X}, V) = \Pr(\chi_5^2 \geq X^2(\boldsymbol{p}, V)) \qquad (15)$$

The χ^2 test has limitations when applied to data with a small number of categories, such as in cases of short chromosomes where the average number of crossovers is low. In such cases, the assumptions underlying the χ^2 test may not hold, leading to unreliable results, so it is important to use an additional test.

Kolmogorov–Smirnov test: This test compares the cumulative distribution functions (CDFs) $F^{(\mathrm{obs})}(x) = F(x, \boldsymbol{X})$ and $F^{(V)}(x)$ of the observed and expected (under the model with the vector of parameters V) distances between chromosomal positions of CO events in individuals. To assess the adequacy of our model, we employed a method that handles separately groups of individuals

with any fixed number of crossovers per considered chromosome (two, three, four, five, and more). For chromosomes with *n* crossovers, we analysed the (*n*-1)-dimensional vector of map distances between consecutive crossovers. First, we used a multivariate generalization of the classical Kolmogorov–Smirnov test proposed by Fasano and Franceschini [54]**.** This method extends the univariate test by constructing hyper-quadrants around each data point and comparing the fraction of points falling within each region across the two samples. This approach is particularly well-suited for our analysis as it handles multivariate data without assuming any specific parametric distribution - an important consideration given that crossover interference can induce complex, non-Gaussian dependencies between inter-crossover intervals. In our implementation, we used the R package *fasano.franceschini.test* [55]**.** To further validate our findings and account for different statistical sensitivities, we also employed the test of Biau and Györfi [56], available in the *DataSimilarity* R package [57]. This test assesses the homogeneity of two multivariate distributions based on the L1-distance between histogram-based density estimates. A key strength of the Biau–Györfi test is its robustness to the curse of dimensionality through its reliance on adaptive partitioning, allowing it to detect differences in distributional shape even in moderate-dimensional data. We evaluate its performance via permutation testing, which generates an empirical null distribution to calculate accurate *p*-values. Since the p-values for the considered groups with different number of crossovers (2, 3, etc) are dependent, to obtain the cumulative p-value, we use the previously mentioned Fisher's [46] combined probability method with permutation correction to account for correlations between tests [47,48]. This comprehensive, nonparametric, and multidimensional testing framework strengthens our ability to rigorously evaluate the fit of the

extended GS3 model across diverse recombination scenarios. The test statistic $D(\boldsymbol{X}, V) := \sup_x \left| F(x, \boldsymbol{X}) - F^{(V)}(x) \right|$ represents the maximum discrepancy between $F^{(\text{obs})}(x)$ and $F^{(V)}(x)$. Under certain assumptions, this statistic follows the Kolmogorov distribution. Thus, the p-value is computed as: $p_{value} = 1 - F_K(\sqrt{n} D(\boldsymbol{X}, V))$, where $F_K(x)$ is the cumulative distribution function for the Kolmogorov distribution. The likelihood function based on the statistic $F(x)$ can be defined as:

$$L_F(\boldsymbol{X}, V) = 1 - F_K(\sqrt{n} D(\boldsymbol{X}, V)) \qquad (16)$$

Searching for a suitable set of parameters

The motivation of our approach is to estimate the contribution of the three putative mechanisms of CO localization (COI-free, positive COI, and negative COI) to the observed patterns of multilocus recombination in real data. We hypothesize that the three mechanisms may operate simultaneously in the same chromosome or even chromosomal region. Thus, we model such situations by a mixture of three processes (section 4), with parameters v_0=1 (no COI), $0<v_-<1$ (COI negative), and $v_+>1$ (COI positive). For different vectors (v_0, v_-, v_+, p_0, p_-, p_+) we calculate the significance level for CO-number and CO-CO distances based on statistical tests (section 5.3). We only consider models with p-values that are not too small for both statistical tests to be a relevant description of the data being analysed. Such an approach enables estimating model parameters assuming homogeneity of the entire process along the chromosome or large enough chromosomal segments.

We simulated a testcross progeny with 2000 genotypes and 200 markers on a chromosome of 200 cM length, with the following parameter values: v_0=1, v_-=0.5, v_+=2, p_-=0.2, p_+=0.4, p_0=1-p_--p_+=0.4. This combination of parameters was

chosen assuming a situation where moderate positive interference (v_+=2) is mixed with no-COI (v_0=1) in equal proportions (40%), and there are also 20% events affected by intense negative COI (v_-=0.5). Using the described procedure (sections 5.1-5.3), the vector of maximum likelihood parameter estimates (maximizing the min $\{L_{\boldsymbol{p}}(\boldsymbol{X}, V), L_F(\boldsymbol{X}, V)\}$) was: $\hat{v}_-$=0.5, $\hat{v}_+$=1.9, $\hat{p}_-$=0.25, $\hat{p}_+$=0.45. We tested hypothesis H_1 that parameter values are equal to the estimated ones against the alternative H_0={p_-=0, p_+=1} (i.e., assuming unimodal distribution with the presence of only positive COI). Under H_0, $\hat{v}_+$ =1.10. The distribution of CO numbers in our simulated data significantly (p-value<0.001) differed from the one expected under H_0. Therefore, the proposed analysis can unravel the mixed nature of interference patterns.

Following the previous section, we compared our "extended" GS model (v_0, v_-, v_+, p_0, p_-, p_+) with the standard GS model (v_+, v_0, p_+, p_0). Our objective was to gain insights into the influence of complex relationships between chromosome length, population size, and relative proportions of the three mechanisms on the outcome of COI models comparison. To ensure the rigor and robustness of the analysis, we conducted a comprehensive investigation using Monte-Carlo simulations. The simulations were performed for each scenario, wherein we systematically varied the population size. Furthermore, to explore the effect of different genetic map lengths, we assessed various scenarios, spanning from 20 to 400 centimorgans (cM) (Table 1). The main objective was to evaluate the relationship between population size and the power of the test (1-β). A similar approach was applied to sets of chromosome lengths, with population size serving as the independent variable. In each simulation, three critical p-value thresholds were considered (0.1, 0.05, and 0.01).

441 <<Table 1 here>>

442 Our numerical tests showed that for achieving an acceptable level of test power
443 in comparisons between different COI models (even in situations of high marker
444 density) and distinguish between the standard GS model (v_+, v_0, p_+, p_0) and the
445 extended GS model (v_0, v_-, v_+, p_0, p_-, p_+), samples of more than 150-200
446 individuals would be more relevant. Notably, the power of the test to distinguish
447 between the COI models was influenced more by the chromosome length than the
448 population size. This suggests that researchers should collect larger samples when
449 conducting COI analysis for species with short genetic maps.

450 **Extended gamma sprinkling model: application on real data**

451 To assess the performance of our modified GS model, we selected a few suitable
452 publicly available datasets on plants: wheatgrass [58], conifer larch [59], and
453 sugar beet [60]. Using a sliding window and *CoC* calculation method, we first
454 looked for areas of negative COI in the COI landscape along the chromosome
455 (Table 2).

456 <<Table 2 here>>

457 We selected chromosomes on which we identified "robust" zones of NCOI with
458 high *CoC* >1, represented by groups of flanking (L and R) markers and central
459 (C) markers of the sliding window. The groups of central-markers (C) displayed
460 a heterogeneity pattern of the interference landscape on the chromosome in the
461 form of an NCOI "valley" similar to the patterns discussed in section 4.4. These
462 data fit the *extended GS* model (v_0, v_-, v_+, p_0, p_-, p_+). For the subsequent analysis,
463 we found it appropriate to normalize the model parameters to account for the
464 simultaneous presence of two COI components with opposite deviations from the
465 no-COI pattern. We associate the extended GS model (v_0, v_-, v_+, p_0, p_-, p_+) with a

singe-shape parameter v (either <1, =1, or >1) of gamma model. This normalization allows for the comparison of various mixture models on a common scale. Such simplification facilitates the evaluation of the relative contribution of clustering (v^-) and repulsion (v^+) components. We propose the following normalization for the shape parameter:

$$v = 1 - (1 - v^-) \cdot p^- + \frac{(v^+ - 1) \cdot p^+}{{v_{max}}^+ - 1} \quad (17)$$

This formula accounts for different scenarios: (*i*) $p^- = p^+ = 0$, v =1; (*ii*) $p^+ = 0$, $p^- = 1$, $v = v^-$; and (iii) $p^- = 0$, $p^+ = 1$, $v = 1 + \frac{v^+ - 1}{{v_{max}}^+ - 1}$. Correspondingly, the normalized parameters are defined as:

$$\hat{v}^- = 1 - (1 - v^-) \cdot p^-, \ \hat{v}^+ = 1 + \frac{v^+ - 1}{{v_{max}}^+ - 1} \cdot p^+$$

The results of this normalization and application of the extended GS model to the analysed chromosomes (see Table 2) are presented in Table 3.

<<Table 3 here>>

In our formula for the model normalization, the denominator ${v_{max}}^+ - 1$ allows us to remove the imbalance of v^- and v^+ ranges. The best value for v_{max} can be estimated from simulated data for each specific experimental situation. The presented algorithm is implemented in software tools (https://github.com/stingerfun/Extended-Gamma-Sprinkled-Model).

Discussion

Variation of crossover interference along a genetic map can be investigated at local and chromosome-wide levels. Both these approaches have limitations. In practical situations, COI can vary along the chromosome, resulting in low compliance of the chromosome-wide model to real data. For a qualitative assessment of the local COI, the standard analysis using *CoC* strongly depends

on the population size and the length of the map interval under study. To address the common problem of insufficient sample size for COI detection, we propose, as an option, to employ larger size intervals. With this approach, we consider an individual as '*recombinant*' in a larger-length interval if one or more CO (either odd or even) events have been registered in the sub-intervals of the larger interval. We demonstrated that *CoC* estimates for such classification of genotypes have a certain advantage over the standard classification. This makes it possible to better assess the impact of negative COI on the local scale of the genetic map. To check for the robustness of the derived conclusions, we perform the analysis on a range of sizes of the consequent intervals along a chromosome.

For the chromosome-wide analysis of COI, the currently widely employed GS model (v_+, v_0, p_+, p_0) does not allow testing for the presence of negative COI, which may exist as demonstrated by some empirical evidence (23-25, 61-63]. Due to the absence of such a possibility in the GS model (v_+, v_0, p_+, p_0), negative COI cases (when present) are mixed with no-COI cases in the output results.

In this paper, we proposed an extension of the GS-model aimed at considering cases of possible (local) negative COI. For this purpose, we analysed the real data by fitting the whole distribution of crossover events as a mixture of three rather than two COI processes (positive, negative, and no-COI). In this way, we estimate distributions of the CO number per genotype and inter-CO distances expected under this model. Comparisons of the distributions observed in the analysis of real data with the massive simulations enable estimating the parameters of the model and testing the corresponding hypotheses. This approach should assist in more objectively addressing the question of the presence of negative COI in each suitable experimental data, rather than ignoring it outright. Usually, COI is considered a factor preventing the formation of excessive COs [62]. Recent studies in *Arabidopsis* and some crops have shown that loss-of-

function mutations in anti-recombination genes involved in DNA repair of excessive double-strand breaks during meiosis can result in a very high increase in CO rates [63-66]. What is the role of deregulation of COI in this increase remains an open question. It is important to mention that in some earlier reports, including drosophila and wheat, negative COI was remarkably associated with pericentromeric regions [24;27;29;49,50], where normally recombination is strongly suppressed.

Abbreviations

cM - centiMorgan
CO - crossover
COI - crossover interference
CoC - coefficient of coincidence
CODA - software for crossover distribution analysis
CSP - crossover superposition principle
df - degrees of freedom
GS - gamma-sprinkled
LG - linkage group
NCOI - negative crossover interference

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The software tools developed in this study are available at: https://github.com/stingerfun/Extended-Gamma-Sprinkled-Model. The datasets

analysed during the current study are available from the corresponding authors on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

The reported study was supported by the Israel Science Foundation (Grant/Award Number: #1844/17); the US-Israel Binational Agricultural Research and Development Fund (Grant/Award Number: IS-5116-18); the Graduate Studies Authority of the University of Haifa; and the Israeli Ministry of Aliyah and Integration.

Authors' contributions

SS and ZF contributed equally to this work. SS, ZF, EP and ABK designed the study. SS and ZF developed the methodology and performed the analyses. SS implemented the software. SS, ZF and ABK wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Morgan TH, Bridges CB, Sturtevant AH. The genetics of Drosophila. Bibliogr. Genet. 1925.
2. Haldane JBS. The Cytological Basis of Genetical Interference. Cytologia (Tokyo). 1931; 3: 54-65.
3. Stevens WL. The analysis of interference. J Genet. 1936; 32: 51-64.
4. Muller HJ. The Mechanism of Crossing-Over. Am Nat. 1916; 50: 193-221.
5. Lindegren CC, Lindegren G. Locally Specific Patterns of Chromatid and Chromosome Interference in Neurospora. Genetics. 1942; 27: 1-24.
6. Strickland WN. An analysis of interference in Aspergillus nidulans. Proc R Soc London Ser B - Biol Sci. 1958; 149: 82-101.

572 7. Hawthorne DC, Mortimer RK. Chromosome Mapping in Saccharomyces:
573 Centromere-Linked Genes. Genetics. 1960; 45: 1085.
574 8. Zhao H, Speed TP, McPeek MS. Statistical analysis of crossover interference
575 using the chi-square model. Genetics. 1995; 139: 1045-1056.
576 9. Copenhaver GP, Browne WE, Preuss D. Assaying genome-wide recombination
577 and centromere functions with Arabidopsis tetrads. Proc Natl Acad Sci U S A.
578 1998; 95: 247-252.
579 10. Chen SY, Tsubouchi T, Rockmill B, Sandler JS, Richards DR, Vader G, et al.
580 Global Analysis of the Meiotic Crossover Landscape. Dev Cell. 2008; 15: 401-
581 415.
582 11. Li X, Li L, Yan J. Dissecting meiotic recombination based on tetrad analysis by
583 single-microspore sequencing in maize. Nat Commun. 2015; 6: 1-9.
584 12. Foss E, Lande R, Stahl FW, Steinberg CM. Chiasma interference as a function of
585 genetic distance. Genetics. 1993; 133: 681-691.
586 13. Foss EJ, Stahl FW. A test of a counting model for chiasma interference. Genetics.
587 1995; 139: 1201-1209.
588 14. McPeek MS, Speed TP. Modeling interference in genetic recombination.
589 Genetics. 1995; 139: 1031-1044.
590 15. Ferreira MTM, Glombik M, Perničková K, Duchoslav M, Scholten O,
591 Karafiátová M, et al. Direct evidence for crossover and chromatid interference in
592 meiosis of two plant hybrids (Lolium multiflorum×Festuca pratensis and Allium
593 cepa×A. roylei). J Exp Bot. 2021; 72: 254-267.
594 16. Sarens M, Copenhaver GP, De Storme N. The Role of Chromatid Interference in
595 Determining Meiotic Crossover Patterns. Front Plant Sci. 2021; 12: 656691.
596 17. Bailey NTJ. Introduction to the mathematical theory of genetic linkage (Oxford
597 University Press, Ed.). The Clarendon Press: London. 1961.
598 18. Zhao H, Speed TP. On Genetic Map Functions. Genetics. 1996; 142: 1369-1377.
599 19. Browning S. The relationship between count-location and stationary renewal
600 models for the chiasma process. Genetics. 2000; 155: 1955-1960.
601 20. Falque M, Anderson LK, Stack SM, Gauthier F, Martin OC. Two types of meiotic
602 crossovers coexist in maize. Plant Cell. 2009; 21: 3915-3925.
603 21. Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C. Genome-Wide
604 Crossover Distribution in Arabidopsis thaliana Meiosis Reveals Sex-Specific
605 Patterns along Chromosomes (M Lichten, Ed.). PLoS Genet. 2011; 7: e1002354.

606 22. Otto SP, Payseur BA. Crossover Interference: Shedding Light on the Evolution
607 of Recombination. Annu Rev Genet. 2019; 53: 19-44.
608 23. Sinclair DA. Crossing over between closely linked markers spanning the
609 centromere of chromosome 3 in Drosophila melanogaster. Genet Res. 1975; 26:
610 173-185.
611 24. Denell RE, Keppy DO. The nature of genetic recombination near the third
612 chromosome centromere of Drosophila melanogaster. Genetics. 1979; 93: 117-
613 130.
614 25. Colombo PC, Jones GH. Chiasma interference is blind to centromeres. Heredity
615 (Edinb). 1997; 79(Pt 2): 214-227.
616 26. Bowring FJ, Catcheside DEA. Evidence for negative interference: Clustering of
617 crossovers close to the am locus in Neurospora crassa among am recombinants.
618 Genetics. 1999; 152: 965-969.
619 27. Peng J, Korol AB, Fahima T, Ro MS, Ronin YI, Li YC, et al. Molecular Genetic
620 Maps in Wild Emmer: genome-wide coverage, massive negative interference, and
621 putative quasi-linkage. Genome Res. 2000; 10: 1509-1531.
622 28. Auger DL, Sheridan WF. Negative crossover interference in maize translocation
623 heterozygotes. Genetics. 2001; 159: 1717-1726.
624 29. Boyko E, Kalendar R, Korzun V, Fellers J, Korol A, Schulman AH, et al. A high-
625 density cytogenetic map of the Aegilops tauschii genome incorporating
626 retrotransposons and defense-related genes: insights into cereal chromosome
627 structure and function. Plant Mol Biol. 2002; 48: 767-790.
628 30. Esch E, Weber WE. Investigation of crossover interference in barley (Hordeum
629 vulgare L.) using the coefficient of coincidence. Theor Appl Genet. 2002; 104:
630 786-796.
631 31. Tsubouchi T, Zhao H, Roeder GS. The Meiosis-Specific Zip4 Protein Regulates
632 Crossover Distribution by Promoting Synaptonemal Complex Formation
633 Together with Zip2. Dev Cell. 2006; 10: 809-819.
634 32. Drouaud J, Mercier R, Chelysheva L, Bérard A, Falque M, Martin O, et al. Sex-
635 specific crossover distributions and variations in interference level along
636 Arabidopsis thaliana chromosome 4 (M Lichten, Ed.). PLoS Genet. 2007; 3:
637 1096-1107.
638 33. Jahns MT, Vezon D, Chambon A, Pereira L, Falque M, Martin OC, et al.
639 Crossover Localisation Is Regulated by the Neddylation Posttranslational
640 Regulatory Pathway (M Lichten, Ed.). PLoS Biol. 2014; 12: e1001930.

34. Aggarwal DD, Rashkovetsky E, Michalak P, Cohen I, Ronin Y, Zhou D, et al. Experimental evolution of recombination and crossover interference in Drosophila caused by directional selection for stress-related traits. BMC Biol. 2015; 13.
35. Aggarwal DD, Rybnikov S, Cohen I, Frenkel Z, Rashkovetsky E, Michalak P, et al. Desiccation-induced changes in recombination rate and crossover interference in Drosophila melanogaster: evidence for fitness-dependent plasticity. Genetica. 2019; 147: 291-302.
36. Aggarwal DD, Rybnikov S, Sapielkin S, Rashkovetsky E, Frenkel Z, Singh M, et al. Seasonal changes in recombination characteristics in a natural population of Drosophila melanogaster. Heredity (Edinb). 2021; 127: 278-287.
37. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via the EM Algorithm. J R Stat Soc Ser B. 1977; 39: 1-22.
38. Nelder JA, Mead R. A Simplex Method for Function Minimization. Comput J. 1965; 7: 308-313.
39. Wright MH. Direct Search Methods: Once Scorned, Now Respectable (G. Balint, B Antala, C Carty, J-MA Mabieme, IB Amar, and A Kaplanova, Eds.). Pittman Res Notes Math Ser Numer Anal 1995. 1996: 191-208.
40. Haldane JBS. The calculation of distances between the loci of linked factors. J Genet. 1919; 8: 299.
41. Kosambi DD. The Estimation of Map Distances From Recombination Values. Ann Eugen. 1943; 12: 172-175.
42. Bernhardsson C, Vidalis A, Wang X, Scofield DG, Schiffthaler B, Baison J, et al. An Ultra-Dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (Picea abies). G3 Genes, Genomes, Genet. 2019; 9: 1623-1632.
43. Ronin YI, Mester DI, Minkov DG, Akhunov E, Korol AB. Building ultra-high-density linkage maps based on efficient filtering of trustable markers. Genetics. 2017; 206: 1285-1295.
44. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann Math Stat. 1938; 9: 60-62.
45. Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. J Educ Behav Stat. 2000; 25: 60-83.

675 46. Fisher RA. Statistical methods for research workers, 4th ed. Oliver & Boyd:
676 Oxford, England. 1932.
677 47. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait
678 mapping. Genetics. 1994; 138(3): 963-971.
679 48. Westfall PH, Young SS. Resampling-Based Multiple Testing: Examples and
680 Methods for p-Value Adjustment. John Wiley & Sons. 1993.
681 49. Korol AB, Preygel IA, Preygel SI. Recombination Variability and Evolution.
682 London Chapman Hall. 1994: 361 pp.
683 50. Korol A, Mester D, Frenkel Z, Ronin Y. Methods for Genetic Analysis in the
684 Triticeae. In: C. Feuillet, G.J. Muehlbauer (eds.), Genetics and Genomics of the
685 Triticeae, Plant Genetics and Genomics: Crops and Models 7, Springer. DOI
686 10.1007/978-0-387-77489-3_6. 2009.
687 51. Gauthier, F., Martin, O. C., & Falque, M. (2011). CODA (crossover distribution
688 analyzer): quantitative characterization of crossover position patterns along
689 chromosomes. BMC bioinformatics, 12(1), 27.
690 52. Lynch M. The origins of eukaryotic gene structure. Mol Biol Evol. 2006; 23: 450-
691 468.
692 53. Smirnov N. Table for Estimating the Goodness of Fit of Empirical Distributions.
693 Ann Math Stat. 1948; 19: 279-281.
694 54. Fasano G, Franceschini A. A multidimensional version of the Kolmogorov-
695 Smirnov test. Monthly Notices of the Royal Astronomical Society. 1987; 225(1):
696 155-170.
697 55. Puritz C, Ness-Cohn E, Braun R. fasano.franceschini.test: An Implementation of
698 a Multivariate KS Test in R. The R Journal. 2023; 15(3): 159-171.
699 56. Biau G, Györfi L. On the asymptotic properties of a nonparametric L1-test
700 statistic of homogeneity. IEEE Transactions on Information Theory. 2005;
701 51(11): 3965-3973.
702 57. Stolte M, Kappenberg F, Rahnenführer J, Bommert A. Methods for quantifying
703 dataset similarity: a review, taxonomy and comparison. Statist. Surv. 2024; 18:
704 163-298.
705 58. Kantarski T, Larson S, Zhang X, Dehaan L, Borevitz J, Anderson J, et al.
706 Development of the first consensus genetic map of intermediate wheatgrass
707 (Thinopyrum intermedium) using genotyping-by-sequencing. Theor Appl Genet.
708 2017; 130: 137-150.

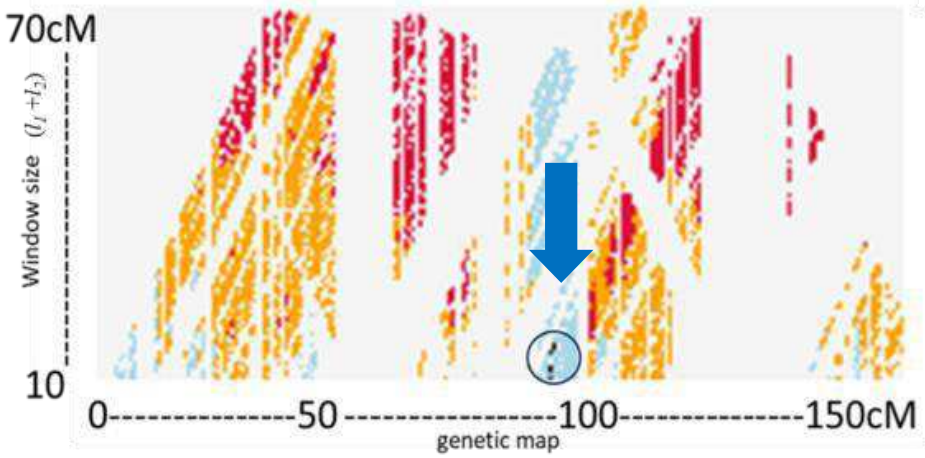
59. Dong M, He Q, Zhao J, Zhang Y, Yuan D, Zhang J. Genetic Mapping of Prince Rupprecht's Larch (Larix principis-rupprechtii Mayr) by Specific-Locus Amplified Fragment Sequencing. Genes (Basel). 2019; 10.
60. Arrieta M, Willems G, DePessemier J, Colas I, Burkholz A, Darracq A, et al. The effect of heat stress on sugar beet recombination. Theor Appl Genet. 2021; 134: 81-93.
61. Anderson LK, Reeves A, Webb LM, Ashley T. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. Genetics. 1999; 151: 1569-1579.
62. Bomblies K, Jones G, Franklin C, Zickler D, Kleckner N. The challenge of evolving stable polyploidy: could an increase in 'crossover interference distance' play a central role? Chromosoma. 2016; 125: 287-300.
63. Crismani W, Girard C, Froger N, Pradillo M, Santos JL, Chelysheva L, et al. FANCM limits meiotic crossovers. Science. 2012; 336: 1588-90.
64. Séguéla-Arnaud M, Crismani W, Larchevêque C, Mazel J, Froger N, Choinard S, et al. Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proc Natl Acad Sci U S A. 2015; 112: 4713-8.
65. Fernandes JB, Séguéla-Arnaud M, Larchevêque C, Lloyd AH, Mercier R. Unleashing meiotic crossovers in hybrid plants. Proc Natl Acad Sci U S A. 2018; 115: 2431-2436.
66. Mieulet D, Aubert G, Bres C, Klein A, Droc G, Vieille E, et al. Unleashing meiotic crossovers in crops. Nat Plants. 2018; 4: 1010-1016.

FIGURE AND TABLE LEGENDS

Fig. 1. COI landscape along the genetic map of Norway spruce LG7 from Bernhardsson [42]. We test the compliance of the obtained data to one of the four COI models: positive, stronger than in Kosambi's model – red; positive, Kosambi's model – orange; no-COI, i.e. Haldane model – blue; significant negative COI - dark blue [surrounded with a circle and marked by arrow]. For this purpose, scan the tested chromosomes using variable-sized sliding windows.

Fig. 2. CODA simulations using the GS model for LG 7 Norway spruce data Bernhardsson [42].

741 Left - CO-to-CO distance distribution for gametes with CO number > 2; right -
742 CO number per gamete distribution. The obtained estimates of GS parameters:
743 v=7.90, p=0.29; green columns – real data; red line – simulation.
744 **Fig.** 3 **Simulation of COI landscape with variable *v*.** Simulated v is represented
745 by a step blackline, variation in *CoC* estimates based on flanked-markers and CSP
746 genotype classification are shown by solid and dotted lines, respectively, for each
747 size (w=10, 20, 30 cM) of the sliding window.
748 **Table 1**. Comparison of the standard GS model and the extended
749 GS model (v_0, v_-, v_+, p_0, p_-, p_+)
750 **Table 2.** Zones of negative COI were detected with a sliding window
751 **Table 3.** GS model (v_0, v_-, v_+, p_0, p_-, p_+) with normalization of the shape
752 parameter v
753
754



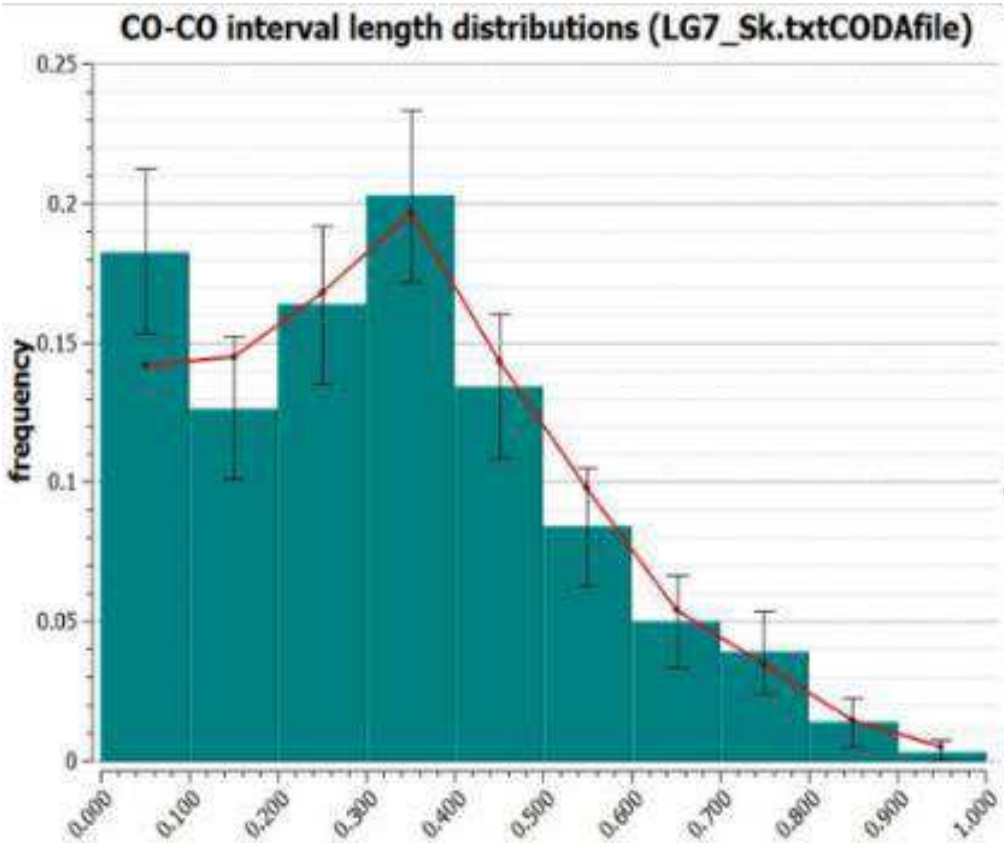
$COI_{Haldane}$: CoC ~ 1 | $LR_{(H1: Haldane)} < \alpha_{\chi 2}$, $LR_{(H1: Kosambi)} > \alpha_{\chi 2}$

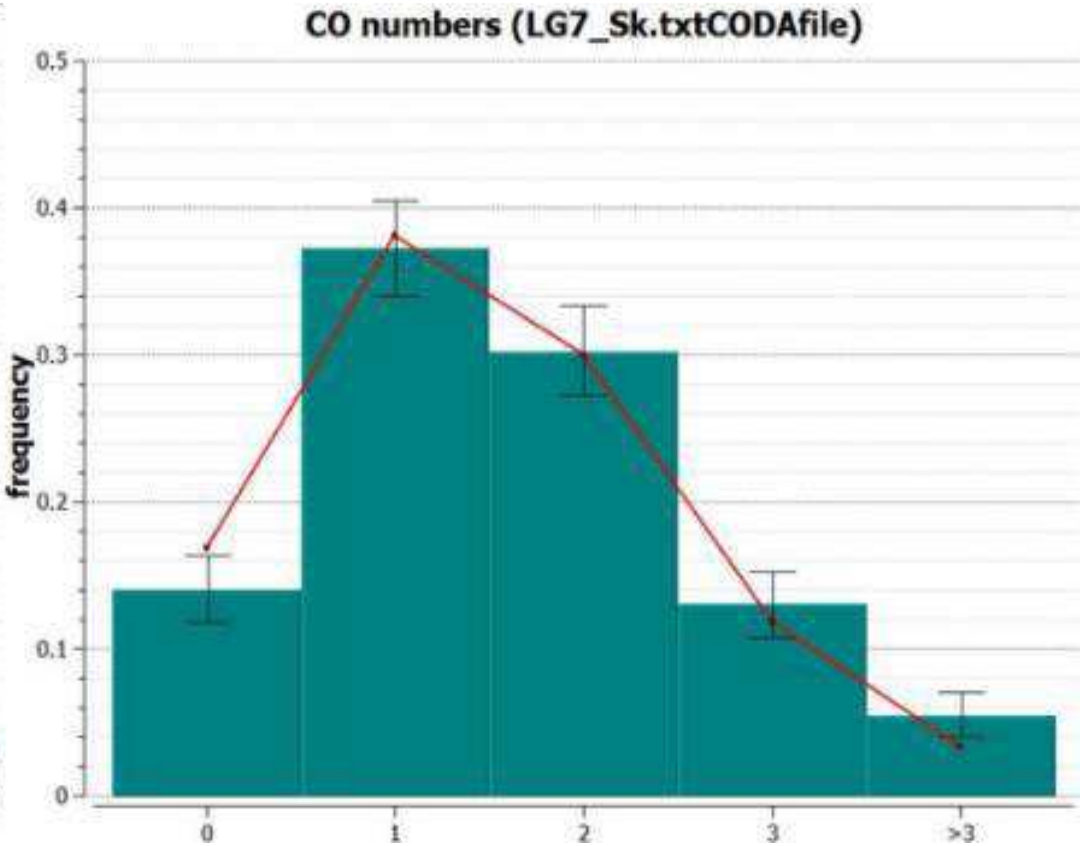
$COI_{Kosambi}$: CoC < 1 | $LR_{(H1: Haldane)} > \alpha_{\chi 2}$, $LR_{(H1: Kosambi)} < \alpha_{\chi 2}$

Significant different (SD):

positive COI (SDPI) : CoC < 1 | $LR_{(H1: Haldane)} > \alpha_{\chi 2}$, $LR_{(H1: Kosambi)} > \alpha_{\chi 2}$

negative COI (SDNI) : CoC > 1 | $LR_{(H1: Haldane)} > \alpha_{\chi 2}$, $LR_{(H1: Kosambi)} > \alpha_{\chi 2}$


CO-CO interval length distributions (LG7_Sk.txtCODAfile)
frequency
0.25
0.2
0.15
0.1
0.05
0
0.000
0.100
0.200
0.300
0.400
0.500
0.600
0.700
0.800
0.900
1.000


CO numbers (LG7_Sk.txtCODAfile)
frequency
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
>3

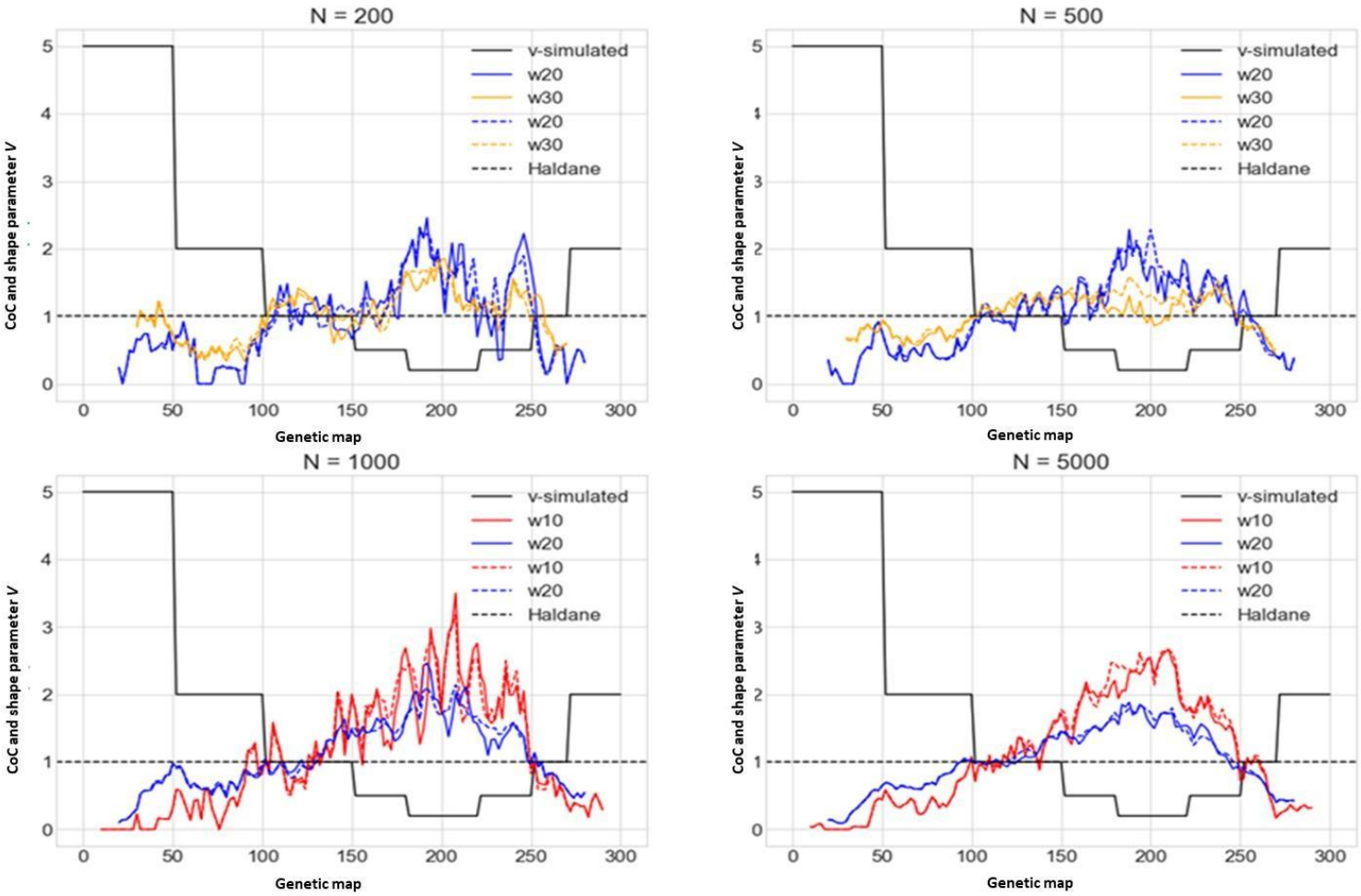

N = 200
v-simulated
w20
w30
w20
w30
Haldane
CoC and shape parameter V
Genetic map
N = 500
v-simulated
w20
w30
w20
w30
Haldane
CoC and shape parameter V
Genetic map
N = 1000
v-simulated
w10
w20
w10
w20
Haldane
CoC and shape parameter V
Genetic map
N = 5000
v-simulated
w10
w20
w10
w20
Haldane
CoC and shape parameter V
Genetic map

Table 1

Chromosome	Sample	*p*-value			Chromosome	Sample	*p*-value		
length	size	0.1	0.05	0.01	lenght	size	0.1	0.05	0.01
20	50	0.09	0.06	0.03	120	50	0.29	0.18	0.06
	100	0.14	0.1	0.05		100	0.47	0.32	0.14
	200	0.26	0.2	0.13		200	0.75	0.61	0.32
	300	0.35	0.28	0.19		**300**	**0.88**	0.77	0.5
	500	0.52	0.43	0.32		**500**	**0.97**	**0.94**	0.79
	750	0.69	0.62	0.47		**750**	**0.99**	**0.98**	**0.94**
	1000	**0.8**	0.71	0.58		**1000**	**1**	**0.99**	**0.98**
	2000	**0.98**	**0.96**	**0.88**		**2000**	**1**	**1**	**1**
50	50	0.17	0.1	0.03	150	50	0.26	0.16	0.04
	100	0.31	0.21	0.09		100	0.48	0.32	0.12
	200	0.55	0.43	0.21		200	0.75	0.59	0.3
	300	0.72	0.59	0.36		**300**	**0.87**	0.77	0.5
	500	**0.91**	**0.82**	0.62		**500**	**0.97**	**0.93**	0.77
	750	**0.96**	**0.94**	**0.84**		**750**	**0.99**	**0.98**	**0.93**
	1000	**0.99**	**0.98**	**0.93**		**1000**	**1**	**1**	**0.97**
	2000	**1**	**1**	**1**		**2000**	**1**	**1**	**1**
75	50	0.24	0.15	0.04	200	50	0.25	0.14	0.04
	100	0.41	0.3	0.12		100	0.45	0.29	0.09
	200	0.65	0.51	0.27		200	0.7	0.52	0.24
	300	**0.84**	0.72	0.48		**300**	**0.81**	0.7	0.39
	500	**0.96**	**0.92**	0.76		**500**	**0.93**	**0.86**	0.65
	750	**0.99**	**0.98**	**0.92**		**750**	**0.96**	**0.95**	0.84
	1000	**1**	**1**	**0.98**		**1000**	**0.98**	**0.97**	**0.91**
	2000	**1**	**1**	**1**		**2000**	**1**	**1**	**0.99**
100	50	0.25	0.16	0.05	400	50	0.23	0.14	0.03
	100	0.44	0.31	0.12		100	0.49	0.29	0.1
	200	0.71	0.55	0.28		200	0.75	0.61	0.29
	300	**0.86**	0.75	0.46		**300**	**0.88**	0.78	0.51
	500	**0.96**	**0.91**	0.75		**500**	**0.94**	**0.91**	0.77
	750	**0.99**	**0.98**	**0.92**		**750**	**0.97**	**0.96**	**0.91**
	1000	**1**	**1**	**0.98**		**1000**	**0.98**	**0.98**	**0.95**
	2000	**1**	**1**	**1**		**2000**	**1**	**1**	**0.99**

Table 2

Organism	Linkage map	Sample size	Marker groups	Positions on genetic map (cM)		
				Left flank	Center	Right flank
Wheatgrass	LG7 female	264	[15 - 21] [21, 25] [26 - 29]	56.1 - 67.5	67.5 - 74.4	81.6 - 93.1
				Average $CoC = 3.21 \pm 0.89$, zone p -value $=2.76\times10^{-25}$		
	LG9 female	264	[0, 10] [2, 12, 14] [15, 16]	0.0 - 57.8	7.2 - 72.8	19.7 - 82.9
				Average $CoC = 2.12 \pm 0.76$, zone p -value $=0.033$		
	LG18 male	342	[1- 8] [3 - 9] [8 - 11]	28.5 - 78.8	56.3 - 84.7	68.6 - 98.3
				Average $CoC = 3.32 \pm 1.47$, zone p -value $=0.0012$		
Larch	Female 10	144	[1 - 15] [8 - 27] [14 - 42]	0.0 - 49.9	17.0 - 50.6	25.4 - 73.3
				Average $CoC = 5.16 \pm 1.91$, zone p -value $=0.1\times10^{-25}$		
Sugar beat	C1 female	151	[1, 2] [2, 3, 4] [3 - 10]	52.4 - 109.3	74.5 - 123.5	92.8 - 142.4
				Average $CoC = 1.63 \pm 0.28$, zone p -value $=2.56\times10^{-18}$		

Table 3

Species	Linkage group	Extended GS model (v_0, v_-, v_+, p_0, p_-, p_+)					Normalized values	
		v_+	v_-	p_+	p_-	p_0	$\hat{v}^+$	$\hat{v}^-$
				%				
Wheatgrass	LG7 female	10	0.9	40	60	0	4.6	0.94
	LG9 female	1.2	0.7	20	70	10	1.04	0.79
	LG18 male	5	0.7	30	40	30	1.62	0.86
Larch	LG10 female	1.1	0.6	30	20	50	1.02	0.9
Sugarbeat	C1 female	1.2	0.8	30	20	50	1.03	0.96