Figure S1  Estimated values of $\nu$ for the Single-Pathway Gamma model. The $\nu$ estimates (black circles) for the 5 chromosome pairs (1 to 5, in red) for male (x-axis) and female (y-axis) meiosis with their 95% confidence intervals (black solid lines). The diagonal (black dashed line) is for $y=x$. 
Figure S2  Scatter plot of successive CO positions for all gametes having 3 COs on their chromosome 1 (male backcross population). Each point gives the positions of the pair (CO 1, CO 2), i.e., (first CO, second CO) or of the pair (CO 2, CO 3), i.e., (second CO, third CO) where the positions are genetic and rescaled to lie in [0,1]. CO positions in adjacent or nearby inter-marker intervals are close to the diagonal (black dashed line, y=x).
Figure S3  Scatter plot of successive CO positions for all gametes having 2 COs on their chromosome 2 (male backcross population). Each point gives the positions of the pair \((CO_1, CO_2)\), i.e., (first CO, second CO) where the positions are genetic and rescaled to lie in \([0,1]\). CO positions in adjacent or nearby inter-marker intervals are close to the diagonal (black dashed line, \(y=x\)).
Figure S4  Genome-wide map of interval hotness for the non-interfering P2 pathway. x-axis: marker intervals in Mb (Mb) along the chromosome considered (corresponding genetic positions also provided in centiMorgan or cM), y-axis: minus the natural logarithm of the p-value of Pearson's chi-square comparison test for that interval. This p-value corresponds to the null hypothesis that the two-pathway Gamma model fits the data and in particular that the P2 COs are uniformly distributed in genetic positions along the chromosomes. The dashed horizontal line shows the FWER (family-wise error rate) of 5% when using the Bonferroni correction for the multiple tests on the chromosome considered. Compared to Figure 5 of the main text, the data set has been filtered: all cases where a gamete is doubly recombinant in two adjacent intervals have been removed.
Figure S5  Comparison of data for chromosome 4 during male meiosis between Drouaud et al. (2007) and Giraut et al. (2011). A. Scatter plot of successive CO positions for gametes having 2 COs on their chromosome 4 (male backcross population) in Drouaud et al. (2007) data set. B. Scatter plot of successive CO positions for gametes having 2 COs on their chromosome 4 (male backcross population) in Giraut et al. (2011) data set. Each point gives the positions of the pair (CO 1, CO 2), i.e., (first CO, second CO) where the positions are genetic and rescaled to lie in [0,1]. CO positions in adjacent or nearby inter-marker intervals are close to the diagonal (black dashed line, $y=x$).
Figure S6  Comparison of data for chromosome 4 during male meiosis between Drouaud et al. (2007) and Giraut et al. (2011). x-axis: mid-points of inter-marker intervals in relative genetic position along the length of chromosome 4M. y-axis: coefficient of variation all inter-crossover distances, weighted (see detailed method in Supporting Information) using the distance between the mid-point of the inter-marker interval under consideration and the mid-point of each of the inter-CO distances (Drouaud in red; Giraut in black). The position of the centromere is indicated in a blue dot along the x-axis.
File S1

SUPPORTING INFORMATION

In the following text we use the symbol ‘\( \nu \)’ instead of ‘\( \nu \)’ to denote the interference strength (Pathway 1, P1) parameter of the Gamma distribution for ease of handling mathematical expressions, the figures maintain the ‘\( \nu \)’ as it is. \( \nu \) must be strictly positive. \( \nu > 1 \) corresponds to positive interference effects, \( \nu < 1 \) corresponds to negative interference effects, while no interference is produced by setting \( \nu = 1 \).

METHODS

Models

COs are formed via two pathways: the first (P1) is interfering and depends on the ZMM family of genes, while the second (P2) has little or no interference and depends on the Mus81 pathway. Following previous two-pathway modeling studies (Copenhaver et al. 2002), we worked with the hypothesis that the two pathways produce COs independently and that P2 has no interference at all. Within such a framework, to simulate meiosis, COs can be produced from each pathway separately and the two corresponding lists merged to get the complete set of CO positions. Simulating P2 is particularly simple: one puts down COs at random with a uniform density in genetic space that depends only on the proportion of COs coming from P2. The main challenge is to produce COs in pathway P1; the procedures are nearly identical in the single and two-pathway approaches.

Single-pathway modeling

Framework: To incorporate interference in the P1 pathway, we used the Gamma model (McPeek and Speed 1995). It is formulated at the level of the bivalent, involving the two homologues (each having 2 sister chromatids) for a total of 4 chromatids during the meiosis. This is a statistical model based on a stationary renewal process in which successive crossovers are separated by genetic distances that are independent, identically distributed random variables following the so-called Gamma distribution. The Gamma model has an interference (strength) parameter such that the larger this parameter, the greater the intensity of the “suppressive” effect, i.e., the less likely it is to find two COs close to one another. When working in genetic space, by definition, CO density is 2 per Morgan at the bivalent level, and as a consequence the average distance between adjacent COs is 0.5 Morgan. Because this average distance between adjacent COs is fixed, when interference suppresses short distance events, this effect is accompanied by a reduction in the frequency of long distance events. As a result, increasing interference strength will reduce the coefficient of variation of distances between adjacent COs.
In the Gamma model, the parameter quantifying interference strength is called $\nu$: it is the shape parameter of the (Gamma) distribution of distances between adjacent COs, while $2 * (\nu)$, that is, $2\nu$, is the “rate” of that Gamma distribution on the bivalent when the density of COs is 2 per Morgan. A convenient feature is that the COs generated along the bivalent are directly specified via their genetic positions (in Morgans or centiMorgans) for this model.

In the present work, the experimental data are not given at the bivalent level: we only have one of the four gametes produced during each meiosis. In this situation one says that the CO data are “thinned”. There is evidence that COs arise between non-sister chromatids without any particular bias in favor of any of these non-sisters (Zhao et al. 1995; Copenhaver 1998). Because of this evidence, which also simplifies modeling, most work is performed enforcing no bias at all. This assumption is referred to as “no chromatid interference”. It is then possible to derive the statistics of the CO patterns at the level of the gametes using the properties at the level of the bivalents even in the absence of knowledge of which chromatids were involved in the genetic exchange.

**Likelihood Computation:** Since the Gamma model produces COs using a stationary renewal process, it is possible to construct the exact likelihood function for any list of COs assuming a given value of the interference strength $\nu$. This likelihood takes into account the effects of thinning (Broman and Weber 2000). Since each backcross is associated with a different (independent) meiosis, the likelihood of the whole data set is the product of likelihoods of each backcross plant. Thus one may obtain the “best” value of the interference strength by maximizing this product; and this is what we do in the maximum likelihood method.

The likelihood computation becomes involved mathematically when the chromosome portions under consideration do not form a continuous stretch; this is the situation when estimating the interference parameters for the distal regions of the chromosomes because the central region has been removed and has to be treated as missing data. Now we explicitly address this situation, using the notations in Broman and Weber (2000).

**Likelihood of data in distal regions of a chromosome (no thinning):** We begin with the COs on the bivalents assumed to be formed by the P1 pathway. If there is at least one CO, let $X_1$, $X_2$, ..., $X_n$ be the $n$ corresponding genetic positions. Then the inter-CO distances, $d_i = X_{i+1} - X_i$ are independent random variables that follow a Gamma distribution with shape $\nu$ and rate parameter $2\nu$, so have probability density, $f(d; \nu) = e^{-2\nu d} (2\nu)^\nu d^{\nu-1} / \Gamma(\nu)$. We furthermore define $d_0 = X_0$ which has its own distribution corresponding to having an interval size at least as large as $d_0$. Following Broman and Weber (2000), the density of $d_0$ is, $g(d; \nu) = 2 (1 - F(d; \nu))$, where $F(d; \nu)$ is the cumulative density function (cdf) of $f(d; \nu)$. The likelihood of these bivalents with at least one CO contains a last factor for the last interval of length $d_n = L - X_n$ where $L$ is the genetic length of the chromosome; this factor is $(1 - F(d_n; \nu))$. If the bivalent has no COs, its likelihood is the probability that the whole chromosome of length $L$ can be placed in an interval of density, $f(d; \nu)$. Again following Broman and Weber (2000), this probability can be obtained in closed form. To summarize, the likelihoods for different values of $n$ is given in (Eqn. 1(a), (b), (c)) along with the corresponding diagrams.
L(ν; [d_i]) = 1 - G(L; ν), when n = 0 (G is the cdf of g)

L(ν; [d_i]) = g(d_0; ν)(1 - F(d_i; ν)), when n = 1

L(ν; [d_i]) = g(d_0; ν)[\prod_{i=1}^{n-1} f ((d_i; ν))] (1 - F(d_n; ν)), when n > 1

We provide a schematic representation following each case to provide intuition towards the more involved likelihood computations that follow. The straight line from 0 to L denotes the bivalent here and the shaded circles show the CO genetic positions while an arc between adjacent COs (or across the entire structure) indicates the length to be used as argument to the function given there.

How does this framework for the likelihood generalize when data is available only for the distal parts of the chromosome (e.g., the first and last quarters of the genetic length of the bivalent)? In effect, one has “missing data” because the central region is hidden, and so we only know whether that region is recombinant or not; in particular, all information on CO positions is lost. CO positions are only known in the two visible regions that are disjoint: left (say the fraction 0 – 0.25) and right (say the fraction 0.75 – 1.0) each of which may or may not have COs. Since the recombination data (Giraut et al. 2011) has high marker density, we assign COs to the mid-point of the recombinant marker intervals in these visible regions. This enables us to retain a continuous picture. Considering the visible regions to begin with, there are four possible situations: (i) no COs visible to the left or to the right of the hidden region, (ii) at least one CO in the left visible region but no CO in the right visible region, (iii) no CO in the left visible region but at least one right visible CO and finally, (iv) at least one CO in the left as well as right visible regions.

In each of these four cases, the hidden region could be recombinant or non-recombinant. Consider first the case when it is non-recombinant. This implies that there is an even number of hidden COs, that is, 0 or 2 or 4 and so on. When the hidden region is recombinant, we know that the number of COs there is odd (1 or 3 and so on). The computation is cumbersome due to the uncertain number of these hidden COs and we must consider all possibilities compatible with the recombination state of the hidden region. When there are no hidden COs, the situation is the simplest. With 1 hidden CO, one has to calculate a one-dimensional integral. With at least 2 hidden COs (whether even or odd), one has to calculate a two-dimensional integral.
To proceed, we follow the previously described logic (Eqn. (1)) and treat successively the ‘\(n = 0\)’, ‘\(n = 1\)’ and ‘\(n > 1\)’ cases. Now, \(n\) denotes the total (combining the left and the right visible regions) number of visible COs. Some new notations need to be introduced. First, let \(W_l\) and \(W_r\) be the left and right ends of the hidden region in genetic units (the \(W\) stands for “window”). Second, if there is one hidden CO (respectively at least two, as the case maybe), let the integration variable associated with the first hidden CO genetic position be \(xCO_1\) (respectively the variable for the last hidden CO genetic position be \(xCO_2\)). We begin by considering the likelihood for the case \(n = 0\) and where the hidden region is non-recombinant. The likelihood is the sum of two likelihoods: \(P_1\), when there are no hidden COs and \(P_2\), when there are an even number of hidden COs. In the figures to follow, the dashed outlined box represents the hidden region and the shaded triangles are the first or last hidden COs while the shaded circles are the visible COs. The conventions are the same as before unless otherwise mentioned. When \(n = 0\) and the non-recombinant hidden region has no hidden COs as in Eqn. 1(a) above, we have:

\[
P_1(v; \{d_i\}) = (1 - G(L; v))
\]

... 2(a)

When \(n = 0\) and instead there are an even number (not 0) of COs in the hidden region, we have to integrate over all possible genetic positions of \(xCO_1\) and \(xCO_2\). \(P_2(v; \{d_i\})\) is thus given as:

\[
\int_{xCO_1}^{W_l} \int_{W_r}^{xCO_2} \left[ g(xCO_1; v) Weven(xCO_2 - xCO_1) \{1 - F(L - xCO_2; v)\} \right] dxCO1 \] dxCO2
\]

... 2(b)

In this expression,
\( \text{Even}(xcO2 - xcO1) = f(xcO2 - xcO1; \nu) + f(xcO2 - xcO1; 3\nu) + \ldots \)

corresponding to having 0, 2, 4, ... COs in addition to \( xcO1 \) and \( xcO2 \). In practice, this infinite sum is truncated as and when the required precision is obtained. Note that the rate \( 2 \nu \) remains the same for all the densities in Eqn. 2(b) but the shape parameter changes, a convenient feature of the Gamma distribution. Note also that for the hidden region, we have introduced a double arc between \( xcO1 \) and \( xcO2 \) to stress that there may be additional COs in that interval.

Thus, the likelihood for \( n = 0 \) and for the window to be non-recombinant is: \( L(\nu; \{d_i\}) = P_1(\nu; \{d_i\}) + P_2(\nu; \{d_i\}) \).

Continuing with the case, \( n = 0 \), but with the hidden region being recombinant, again there will be two classes of events to consider with an odd number of COs in the hidden region: having only one or at least three. The total likelihood will be the sum of the corresponding two probabilities.

For the first class, one has to integrate over all genetic positions of \( xcO1 \). This gives \( P_1(\nu; \{d_i\}) \) as:

\[
\int_{W_1}^{W_r} g(xcO1; \nu)[1 - F(L - xcO1; \nu)] dxcO1 \]

... 3(a)

Similarly for the second class one has to integrate over all genetic positions of \( xcO1 \) and \( xcO2 \). \( P_2(\nu; \{d_i\}) \) is thus computed as:

\[
\int_{W_1}^{W_r} \int_{xcO1}^{W_r} [g(xcO1; \nu)Wodd(xcO2 - xcO1)[1 - F(L - xcO2; \nu)] dxcO1] dxcO2 \]

... 3(b)

In this expression, in direct analogy with what we saw before, one has

\( Wodd(xcO2 - xcO1) = f(xcO2 - xcO1; 2\nu) + f(xcO2 - xcO1; 4\nu) + \ldots \)
corresponding to having 1, 3, 5, … COs in addition to \(xCO1\) and \(xCO2\). Just as for \(W\)even the series is truncated and the rate \(2 \nu\), remains the same for all the densities in 3(b) while the shape parameter varies.

This concludes the \(n = 0\) case.

Similar calculations apply to the cases \(n > 0\). Let us illustrate the calculations when there is only one visible CO and it is to the left of the hidden region. We will have the same sub-cases as before with regard to the hidden region which can be recombinant or non-recombinant.

Consider first the non-recombinant case. When the hidden region has no COs, we have a first likelihood as illustrated in the figure above:

\[
P_1(\nu; \{d_i\}) = g(d_0; \nu)(1 - F(d_1; \nu))
\]  
\[\ldots 4(a)\]

Note that \(d_1 = L - d_0\).

If instead the hidden region has COs, they must be an even number; the corresponding likelihood requires performing a double integral in analogy to the \(n = 0\) case. This leads to the formula for \(P_2(\nu; \{d_i\})\):

\[
g(d_0; \nu) \left[ \int_{L}^{Wl} \int_{xCO1}^{xCO2} f(xCO1 - d_0)W\text{even}(xCO2 - xCO1)(1 - F(L - xCO2; \nu))dxCO1\ dxCO2 \right]
\]  
\[\ldots 4(b)\]

where \(W\)even\((xCO2 - xCO1)\) is as defined in Eqn. 2(b).
The same logic applies to the situation when the hidden region is recombinant. The case of exactly one hidden CO leads to $P_1(\nu; \{d_i\})$, given by:

$$g(d_0; \nu) \int_{W_l}^{W_r} f(xCO1 - d_0; \nu)(1 - F(L - xCO1; \nu))dxCO1$$

... 5(a)

When there at least 3 COs, we see that the term, $P_2(\nu; \{d_i\})$ is the product of $g(d_0; \nu)$ and the double integral:

$$\left[\int_{W_l}^{W_r} \int_{W_l}^{W_r} [f(xCO1 - d_0; \nu)Wodd(xCO2 - xCO1)(1 - F(L - xCO2; \nu))]dxCO1 \right] dxCO2$$

... 5(b)

where $Wodd(xCO2 - xCO1)$ is as defined previously in Eqn. 3(b)

This case of there being exactly one visible left CO can be generalized very simply to any number of visible left COs (there is simply an additional factor $f(\cdot)$ for every arc between successive visible left COs).

Similarly if the bivalent has only right visible COs (none on the left), then the relevant diagrams are the mirror images of the ones presented above for visible left COs. The likelihood formulae then follow straightforwardly.

Finally when we have visible COs to the left as well as to the right, one has that (i) each of the end intervals contributes a factor $(1 - F(\cdot))$ and to one of these we associate an additional factor 2 (analogous to using $g(\cdot)$ on one side and $(1 - F(\cdot))$ on the other), (ii) all simple arcs connecting COs carry a factor $f(\cdot)$, (iii) all double arcs carry a factor $Weven$ (respectively $Wodd$) when the interval $[W_l, W_r]$ is non-recombinant (respectively recombinant), (iv) when there is one hidden CO, the integration range is $W_l \leq xCO1 \leq W_r$ and if
there are at least two COs, the integration range is $W_l \leq xCO1 \leq xCO2 \leq Wr$.

**Likelihood of data in the distal regions of a chromosome (with thinning):** Having treated what happens on the bivalent, we have to
c onsider now the likelihood of a gamete, one of the four products of the bivalent (cf. sub-head, ‘Fit of the Gamma Model’ in the
Materials & Methods section of Broman and Weber 2000). In the absence of chromatid interference, each CO on the bivalent has a 50%
probability of being passed to the considered gamete. This process of removing COs randomly is called “thinning” and leads to modified
formulae for the likelihoods. Broman and Weber derived these formulae when the whole chromosome is visible. The purpose of this
section is to generalize these formulae to the case where the chromosome contains a hidden region. For completeness, we begin by
recalling the results of Broman and Weber.

Consider a random meiotic product, that is, a gamete. The COs of the bivalent have been thinned independently with probability 1/2. If
the inter-crossover genetic distances along the gamete are $l_1, l_2, \ldots$, then they have “similar” statistical properties as the $d_i$ introduced
above. Specifically, the $l_i$ are independent and $l_1, l_2, \ldots$ have density, $f^*(l; \nu) = \sum_{k=1}^{\infty} \frac{1}{2^k} f_k(l; \nu)$, where $f_k(l; \nu)$ is a Gamma
density with rate $2\nu$ and shape $k\nu$. Furthermore, the density of $l_0$(the distance between the left end of the chromosome and the first
CO) is $g^*(l; \nu) = 1 - F^*(l; \nu)$, where $F^*(l; \nu)$ is the cdf of $f^*(l; \nu)$ (details are given in Broman and Weber 2000). In effect, the
calculation of the likelihood of a gamete after thinning is as without thinning but where $f^*$ replaces $f$ in the derivation. Let us go over
the different cases. Suppose the gamete has $q$ COs after thinning and let the inter-crossover distances be $l_1, l_2, \ldots, l_{q-1}$. We furthermore
set $l_0$ to be the genetic position of the first CO and $l_q$ to be the length of the interval between the end of the chromosome and CO
number $q$. As before, $L$ is the genetic length of the gamete. Then depending on $q$, the likelihood of the gamete will be:

$L(v; \{l_i\}) = (1 - G^*(l_{q}; \nu))$, when $q = 0$ ($G^*$ is the cdf of $g^*$) \hfill \ldots (6a)

$L(v; \{l_i\}) = g^*(l_{0}; \nu) g^*(l_{1}; \nu)$, when $q = 1$ \hfill \ldots (6b)

$L(v; \{l_i\}) = g^*(l_{0}; \nu) \prod_{i=1}^{q-1} f^*(l_{i}; \nu)) g^*(l_{q}; \nu)$, when $q > 1$ \hfill \ldots (6c)

Given these formulae for the whole chromosome, we now generalize to the case of discontinuous regions as before. The notations are
the same as before. Thus, we have \( q \) visible COs ‘after’ thinning. Apart from the visible CO positions ‘after’ thinning, we must consider the possibility of there being a CO ‘before’ thinning in each visible inter-marker interval. So let there be \( m \) markers (namely, \( M(1), M(2), M(3), \ldots, M(m) \)) along the gamete in the visible region of which \( m_l \) are on the left of the hidden region and \( m_r \) are on the right. Finally, as far as the hidden region is concerned, no thinning is explicitly carried out there. In fact all COs we refer to in the hidden region are ‘before’ thinning.

Just as in the description without thinning, let us begin our explanations when there are no COs (‘after’ thinning) in the visible regions. There are 2 associated probabilities which are important to understand: (a) the probability \( P_a \) that the hidden region is recombinant ‘after’ thinning; note that this event can occur only if there is at least one hidden CO (not zero) ‘before’ thinning and (b) the probability \( P_b \) that the hidden region is non-recombinant ‘after’ thinning: this event is possible no matter what the number of hidden COs (including zero) ‘before’ thinning. These probabilities depend on 3 sub-cases in which the hidden region has zero, one or at least two COs (‘before’ thinning). If these 3 sub-cases have probabilities \( P_{0w} \), \( P_1 \) and \( P_2 \), then: \( P_a = (P_1 + P_2)/2 \) and \( P_a = P_0 + (P_1 + P_2)/2 \). For the calculation of \( P_{0w}, P_1 \) and \( P_2 \), there are 4 sub-sub-cases, each based on whether there are or not any (left or right) visible CO(s) ‘before’ thinning. In the representation below, the empty circles represent COs in the visible region (they could be either absent altogether or present only on one side) present ‘before’ thinning but which are thinned out, that is disappear ‘after’ thinning. The shaded triangles indicate hidden COs ‘before’ thinning (that may or may not be thinned out).

![Diagram](image)

To be specific, consider the sub-case of the probability \( P_1 \) and its sub-sub-case when there are visible COs to be thinned out on both sides of the hidden region. Let \( x \) be the genetic position of the last left visible CO to be thinned and \( y \) that of the first right visible CO to be thinned. The probability of this sub-sub-case can be written as follows:

\[
\int_0^{W_l} dx \int_{W_l}^{W_r} dy (1 - F_*(x; v)) f(xCO1 - x; v) f(y - xCO1; v)(1 - F_*(L - y; v))
\]

... (7)

In practice, the integrations involving \( x \) and \( y \) are replaced by summations (using the inter-marker interval sizes and their mid-points) which give good approximations in this case as the markers are densely placed along the length of the gamete. The other sub-sub-cases are treated analogously. If there is no visible CO to be thinned on one side (or both), then there is one (or two) less variable(s) to integrate over and the product \( (1 - F_*(\cdot)) f(\cdot) \) is replaced by \( (1 - F(\cdot)) \).
Moving to the sub-case where at least two COs arise in the hidden region, $P_2$ breaks down into the

\[ \int_0^{W_l} dx \int_{x_C01}^{x_C02} dy \left( 1 - F^*(x; v) \right) f(x_C01 - x; v) W(x_C02 - x_C01; v) f(y - x_C02) (1 - F^*(L - y; v)) \]

... (8)

Again, as before, here we actually do a double integral, the integrals over $x$ and $y$ are evaluated by summing over the inter-marker interval mid-points. Also, in case there is no visible CO to thinned on either side (or not at all), then, the product \( (1 - F^*(;)) f(\cdot) \) changes to only \( 1 - F(\cdot) \) and there is one (or two) less variable(s) to integrate for. This completes the cases when there are no visible COs 'after' thinning \( q = 0 \).

Lastly, there may be visible COs 'after' thinning \( q > 0 \). If so, consider the last left visible CO when there is one and the first right visible CO when it exists. These COs contribute to the $q$ visible COs ('after thinning). The computation of the corresponding likelihoods generalizes what we have seen before when there were left or right COs (without the hidden region) to include the sub-sub-cases based on the thinned that were just discussed when $q = 0$. Clearly, these thinned COs have a modified integration range. For instance, when integrating over $x$ (the genetic position of the last thinned CO on the left), the lower bound is the genetic position of the last left visible CO. The same comment applies to $y$ which must not go beyond the position of the first right visible CO. The expressions for the hidden region as well as the number of variables to be integrated over follow the same rules as for $q = 0$.

**Fitting Procedures, Confidence intervals:** Adjusting the model parameter to fit the experimental data requires searching for the maximum likelihood score $L$. We do this search via a “hill-climbing” procedure (Gauthier et al. 2011) as follows. Given a point in the search space (specified by the model's parameter value or values), we calculate $L_0$, the likelihood there, and also the score at neighboring points obtained by increasing and decreasing the parameter by a characteristic step. We also consider a trial point using polynomial interpolation and measure the score there. If at least one of these neighbors has likelihood larger than $L_0$, we move to the
point with the highest score. If none of the neighboring points has a score larger than $L_0$, we reduce the size of the characteristic step while also trying an intermediate point based on polynomial interpolation. The procedure is iterated and the position converges to a (local) maximum of $L$. We perform checks to verify that this maximum is in fact a global maximum, attainable from different initial positions. This fitting leads to the value ($\nu$) of the interference strength which gives the maximum likelihood in the parametric space. The confidence intervals are computed using Fisher’s Information matrix.

**Two-pathway modeling via sprinkling**

**Framework:** To include COs from the P2 pathway, the sprinkling procedure (Copenhaver et al. 2002) is used, which is simply the superposition of the non-interfering COs, generated using the Poisson distribution, onto the ones from the interfering P1 pathway. The fraction of P2 COs is thus an additional parameter to the one in the single-pathway modeling; on the bivalent, it is the proportion of non-interfering COs, that is $p$. When $p > 0$, the density of P1 COs is no longer 2 per Morgan, but $2 \times (1 - p)$. Comparing to the procedure for producing P1 COs in the single pathway model, we see that the shape parameter of the (Gamma) distribution of distances between adjacent COs is still $\nu$, but the rate parameter is changed from $2 \times (\nu)$ to $2 \times (\nu) \times (1 - p)$, that is from $2\nu$ to $2\nu(1 - p)$.

**Likelihood Computation:** As described above for the single pathway case, the process and logic remain the same except that in all likelihoods, the shape parameter changes (from $2\nu$ to $2\nu(1 - p)$).

**Fitting Procedures, Confidence Intervals:** The principles used here are the same as those mentioned for single pathway modeling; the main difference is that the likelihood (Falque et al. 2009) $L$ now is a function of two variables so the parameter space to search is two-dimensional. Again the hill-climbing algorithm was used (Gauthier et al. 2011). And the Fisher Information matrix was computed to obtain the confidence intervals.

**Statistical analyses and comparison tests**

**Comparing two datasets (separate chromosomes or different regions of one chromosome):** We performed three levels of comparisons to examine the variation in interference between and within chromosomes. Using mainly the two-pathway model, we compared the interference strength: (1) between male and female meiosis, (2) between the different chromosomes but for a given sex, and (3) between segments of the same chromosome, looking at variations in interference values between the two arms of a chromosome and also between the central and distal regions of a chromosome.

To make these comparisons, we tested the null hypothesis ($H_0$) that the means ($\nu$ or $p$ here) of the populations, from which the two
samples under consideration have been drawn, are equal. Here the population variances are unequal which requires changing the formulae for the test statistic as well as the accompanying degrees of freedom for this modified two-sample \( t \)-test; for this we follow Welch’s \( t \)-test. Let the sample means be \( \bar{X}_1 \) and \( \bar{X}_2 \), the respective standard deviations be \( S_1, S_2 \) and lastly the sample sizes be \( n_1 \) and \( n_2 \). Then under the \( H_0: \mu_1 = \mu_2 \) (\( \mu \) are population means: interpreted here as \( \nu \) or \( p \)), the statistic \( t(\text{observed}) \) follows a \( t \)-distribution with degrees of freedom \( df \). Their corresponding formulae are given as follows:

\[
t(\text{observed}) = (\bar{X}_1 - \bar{X}_2) / \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}
\]

\[
df = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 / \left(\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}\right)
\]

Finally, the function \( pt(\cdot) \), of the R statistical package, was used to compute the two-sided \( P \)-value at the 5\% level of significance. When the \( P \)-value is significant, we reject the null hypothesis. This indicates that the difference between the samples under consideration may not be considered negligible.

**Comparing two discrete distributions (simulated and experimental):** We performed another kind of comparison between simulated datasets and experimental data. This was necessary when we tested for “hot” regions specific for the non-interfering (P2) pathway. The starting point is the distribution of COs given there is a CO in a “reference” interval under consideration. For each interval spanning adjacent markers (assuming at least 1 gamete has a CO in this interval), the frequency of COs in each of the other intervals is computed, using gametes which have multiple COs (treating separately those with 2 and 3 COs). The analogous frequencies are obtained in the context of the model’s predictions. Specifically, the model’s behavior is obtained from simulated data, generated using the `simdata` option of CODA (Gauthier et al. 2011) with \( \nu \) and \( p \) set to the values obtained from fitting the experimental data. Then, we tested for a significant difference between the expected (simulated or theoretical) and observed (experimental) frequency distributions of CO occurrences for each inter-marker interval at a time. We used the Pearson’s chi-square test function (Lindsey 2004) within the R statistical software, `chisq.test()` to test the null hypothesis that the observed distribution is not statistically different from the observed one, separately for all intervals. Furthermore, to obtain a global view of the results, we merged the values from the 2 COs and 3 COs cases by taking the sum of the corresponding chi-square values (for intervals having data for the 2COs and 3 COs cases). The new \( P \)-values were computed by the R function, `pchisq(.)`. And if an interval had data only for 2 COs or only for 3 COs, we retained the previous chi-square and \( P \)-values.

**Coefficient of variation of inter-CO distances along chromosomes:** Given a list of distances between adjacent COs, let \( \mu \) and \( \sigma \) be the associated mean and standard deviation. Then the coefficient of variation \( CV \) is defined as the ratio \( \sigma / \mu \). Low values of \( CV \) correspond to high levels of effective interference. To allow \( CV \) to be position dependent so as to get a picture of effective interference strength along a
We apply a weight to each element of the list (a distance between adjacent COs) as follows. Let $X$ be the current position where the local CV is to be measured and let $Y$ be the midpoint of the 2 COs under consideration. Then the weight assigned to the associated CO-CO distance is taken as $\exp(-10(X - Y)^2)$. Then the $X$-dependent mean $\mu$ and standard deviation $\sigma$ defining $\text{CV}(X)$ are simply computed using these weights. Explicitly, we have:

$$
\mu(X) = \frac{\sum w_i d_i}{\sum w_i} \quad \text{and} \quad \sigma^2(X) = \frac{\sum w_i (d_i - \mu(X))^2}{\sum w_i}
$$

where $i$ is the index of the element in the list of distances $d_i$ between adjacent COs. The corresponding $\text{CV}(X)$ curves are displayed in Fig S6. Note that the weights mimic a sliding window without having the drawback of a discontinuous behavior in $X$.

**Comparison to Drouaud et al. results**

In a previous study (Drouaud et al. 2007) of chromosome 4M of *A. thaliana*, variability in interference strength was revealed by using three-point coefficients of coincidence (or “C3”s that use 2 adjacent windows). Those authors found that the effective interference was higher on the left side than on the right side. We reach the same conclusion from analysing our 4M data that was generated independently of that of Drouaud et al. The overall similarity of these data sets is demonstrated in Figure S5 which shows the distribution of COs in 4M gametes having exactly two COs, for the two data sets. (Figure S5 (A) uses the data of Drouaud et al. and thus provides quantitatively the same information as given in Figure 4 (A) of their paper). At a more mathematical level, we have extracted from each data set the distribution of distances between adjacent COs from which we determine the associated coefficient of variation (CV) as a function of position along the chromosome (see previous paragraph). In Figure S6 we display the associated curve for each data set: the two curves are qualitatively similar. Furthermore, CV is clearly lower on the left than on the right, pointing to a higher effective interference on the left side than on the right side, in agreement with the conclusion of Drouaud et al. Nevertheless, the “C3” values in Drouaud et al.’s work also suggested that interference strength was weakest in the middle, not all the way on the right end (see Figure 6 (A) in Drouaud et al. 2007 and Figure S6 here). Such differences can arise because “C3”s and CV are sensitive to different aspects of interference. Indeed, CV incorporates all adjacent inter-CO distances for each inter-marker interval to compute the CV value for that interval, whereas the “C3” values in Drouaud et al. depend on recombination within windows surrounding the point of interest. Thus the “C3” approach is less sensitive to nearby COs because only rarely do these lead to recombination in both of the associated windows. To illustrate this difference, consider again the COs displayed in Figure S5. On the far left side, the data of Drouaud et al. has a few close-by COs but nevertheless the “C3” values are tiny there, showing that indeed such close-by COs do not lead to double recombination events contributing to “C3”.

S. Basu-Roy et al.
REFERENCES


Table S1  Results (P-values) for Single-Pathway Gamma Comparisons. Comparisons between (a) male and female meioses for the same chromosome: diagonal values (enclosed in **bold**) and (b) different chromosomes’ male meiosis (upper triangle values) and female meiosis (lower triangle values) separately. The P-values were computed for the null hypothesis that the 2 meioses under consideration are associated with the same value of $\nu$ (or $\nu^\prime$).

<table>
<thead>
<tr>
<th></th>
<th>1M</th>
<th>2M</th>
<th>3M</th>
<th>4M</th>
<th>5M</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$\nu^* 4M$</td>
<td>-</td>
<td>1M</td>
</tr>
<tr>
<td>2F</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$\nu^* 4M$</td>
<td>$\nu^* 5M$</td>
<td>2M</td>
</tr>
<tr>
<td>3F</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$\nu^* 4M$</td>
<td>$\nu^* 5M$</td>
<td>3M</td>
</tr>
<tr>
<td>4F</td>
<td>$\nu^* 4F$</td>
<td>$\nu^* 4F$</td>
<td>$\nu^* 4F$</td>
<td>-</td>
<td>-</td>
<td>4M</td>
</tr>
<tr>
<td>5F</td>
<td>$\nu^* 5F$</td>
<td>-</td>
<td>$\nu^* 5F$</td>
<td>-</td>
<td>-</td>
<td>5M</td>
</tr>
<tr>
<td>Female</td>
<td>1F</td>
<td>2F</td>
<td>3F</td>
<td>4F</td>
<td>5F</td>
<td></td>
</tr>
</tbody>
</table>

Note: '*' indicates P-value is significant at the 5% threshold. For every significant comparison, the chromosome indicated has the higher $\nu$. 
Table S2  Results (P-values) for Intra-chromosomal Comparisons. Comparisons are (a) between the left (L) and right (R) arms (L/R) of each chromosome and (b) between the central (C) region (fraction 0.25 to 0.75 of the genetic length) and the distal regions or “extremities” (E, union of 0 to 0.25 and 0.75 to 1 of the genetic length) of the chromosome (C/E), for Gamma models: Single as well as Two-pathway. The P-values were computed for the null hypothesis that the 2 chromosome regions under consideration are associated with the same value of nu (or v) or p (separately).

<table>
<thead>
<tr>
<th></th>
<th>Single pathway</th>
<th>Two-pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L/R</td>
<td>C/E</td>
</tr>
<tr>
<td></td>
<td>Nu</td>
<td>nu</td>
</tr>
<tr>
<td>1M</td>
<td>-</td>
<td>* C</td>
</tr>
<tr>
<td>2M</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3M</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4M</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5M</td>
<td>-</td>
<td>* E</td>
</tr>
<tr>
<td>1F</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2F</td>
<td>-</td>
<td>* C</td>
</tr>
<tr>
<td>3F</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4F</td>
<td>* R</td>
<td>-</td>
</tr>
<tr>
<td>5F</td>
<td>-</td>
<td>* C</td>
</tr>
</tbody>
</table>

Note: ‘*’ indicates P-value is significant at the 5% threshold. For every significant comparison, the portion of the chromosome indicated (L/R for left/right or C/E for centre/extremities) is the one having the higher nu or p.