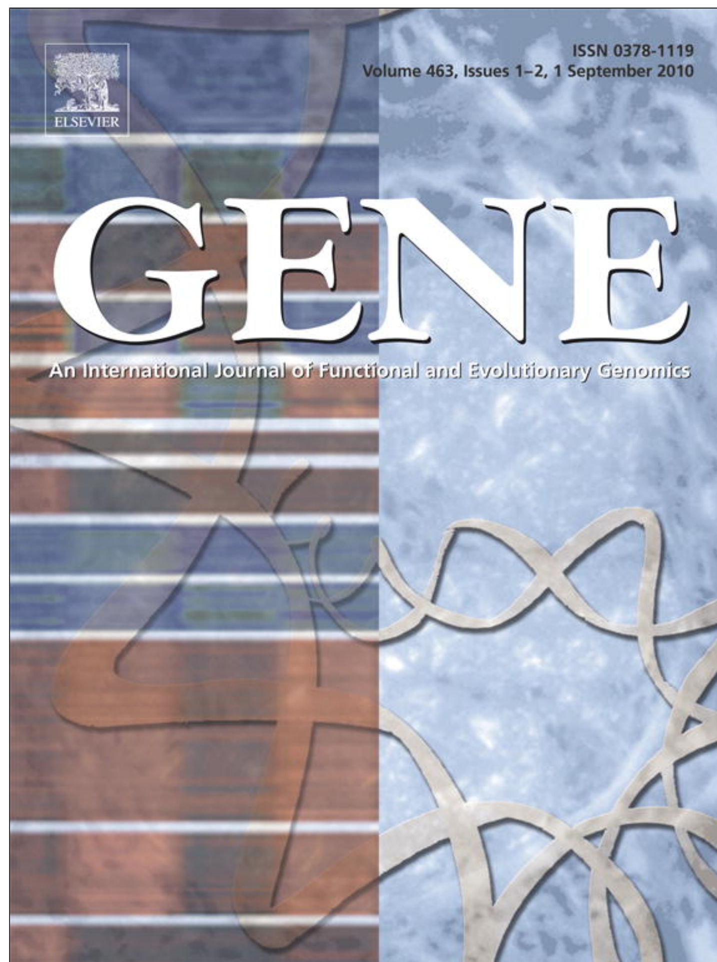
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Gene 463 (2010) 18–20

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Drosophila melanogaster recombination rate calculator

Anna-Sophie Fiston-Lavier [a,*,1], Nadia D. Singh [b,1], Mikhail Lipatov [a], Dmitri A. Petrov [a]

[a] Department of Biology 371 Serra St.,Stanford University Stanford CA 94305-3020, USA
[b] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-2703, USA

ARTICLE INFO

Article history:
Received 13 March 2010
Received in revised form 26 April 2010
Accepted 28 April 2010
Available online 7 May 2010

Received by C. Feschotte

Keywords:
Meiotic recombination
Drosophila

ABSTRACT

Recombination rate is a key evolutionary parameter that determines the degree to which sites are linked. Estimating recombination rates is thus of crucial importance for population genetic and molecular evolutionary studies. We present here a user-friendly web-based tool that can be used to retrieve recombination rate estimates for single and/or multiple loci in the *Drosophila melanogaster* genome given a user-defined choice of the genome release. We used the Marey map approach that is based on comparing the genetic and physical maps to infer recombination rates along the major chromosomes of the *D. melanogaster* genome. Our implementation of this approach is based on building third-order polynomials which are used to interpolate recombination rates at all points on the chromosome except for telomeric and centromeric regions in which such polynomials are known to provide particularly poor estimation.

Published by Elsevier B.V.

1. Introduction

Recombination is a key process in genetics and evolution. The rate of recombination is known to be non-uniform along chromosomes and among organisms (Nachman and Churchill, 1996; Broman et al., 1998; Yu et al., 2001; Kong et al., 2002). Recombination rate is usually measured in terms of the expected number of recombination events between two loci, and is often reported as the frequency of exchange per unit physical distance (e.g., cM/Mb).

Recombination rates in Drosophila have been estimated using a variety of statistical estimators. Early estimators, such as the adjusted coefficient of exchange, were based on cytological markers in combination with DNA content estimates from optical densities of polytene chromosomes (Kindahl, 1994; Sorsa, 1988). Other techniques used repeated sequences such as transposable elements as markers on the physical map (Hey and Kliman, 2002; Ising and Block, 1984). However, the availability of the whole genome sequence of *Drosophila melanogaster* (Adams et al., 2000) and thus a complete and accurate physical map has facilitated a much more accurate estimation of recombination rates (Hey and Kliman, 2002; Marais et al., 2001). Estimates of the recombination rate can be obtained using the "Marey map" approach, an approach originally proposed by Chakravarti (1991).

We previously estimated recombination rates in *D. melanogaster* by generating "Marey maps" of the genetic positions of molecular markers (in cM) as a function of their physical positions (in Mb). Recombination rate was estimated as the derivative of this estimated function at any position on the chromosome (Singh et al., 2005). These maps provided accurate estimates of recombination rate throughout the genome with the exception of those regions at the telomeres and centromeres. Here we discuss an improved statistical procedure specifically focused on correcting the telomeric and centromeric estimates. We have updated our free, user-friendly web-based tool to incorporate these statistical advances to our methodology. This web-based tool, implemented in Perl, provides the user with easy access to our estimates of the rates of recombination for any position in the *D. melanogaster* genome. This recombination rate calculator is fast and can be tailored to any specific request. The user is free to choose the release of the genomic sequence and to input single or multiple locus coordinates.

2. Approach

The recombination rate at any given nucleotide coordinate is estimated by taking the slope of the curve relating the genetic to the physical maps in one of two ways. One could employ a sliding window approach, in which some function is used to fit the genetic position as a function of physical position within a window of a given size along a chromosomal arm. This function may be as simple as a linear function, or could be as complex as cubic splines (Berloff et al., 2002; Yu et al., 2001). Alternatively, the genetic position of the markers can be fitted as a function of physical position using a polynomial curve across an entire chromosomal arm. Under this approach, recombination is estimated at an individual nucleotide coordinate as the derivative of the polynomial curve. While the polynomial curve approach is less sensitive to regional variation in recombination rates than the sliding window approach, it is more robust to errors in the physical and genetic maps.

Abbreviations: bp, base pair(s); cM, centiMorgan(s); Mb, Mega base pair(s).

* Corresponding author. Tel.: +1 650 7361169; fax: +1 650 7236132.
E-mail address: afiston@stanford.edu (A.-S. Fiston-Lavier).

[1] These authors contributed equally and are listed alphabetically.

0378-1119/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.gene.2010.04.015

Author's personal copy

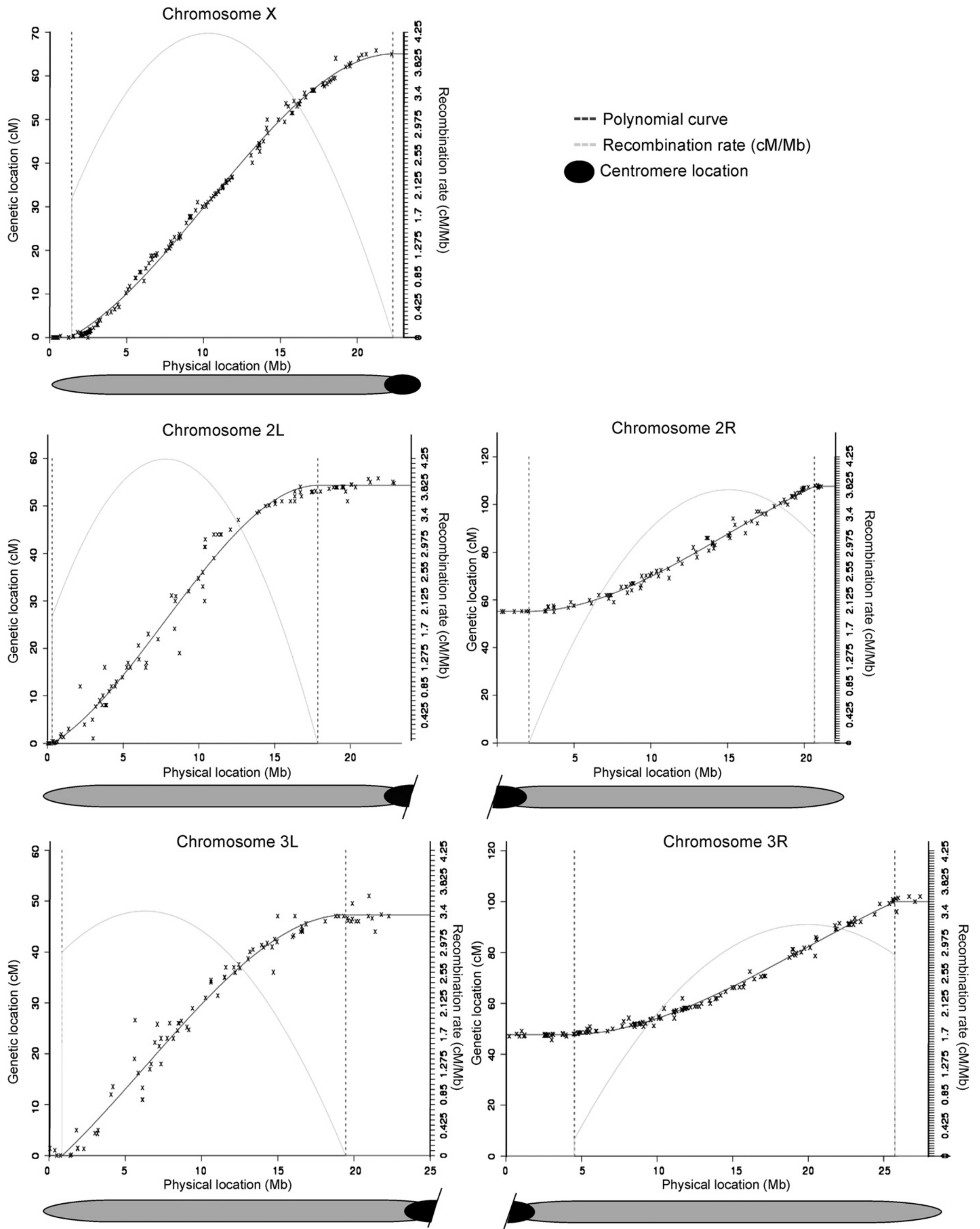


Fig. 1. Marey maps for the five chromosomal arms of *Drosophila melanogaster*. For each chromosomal arm, a third polynomial curve representing the genetic positions of molecular markers (cM) against their physical position (Mb) is plotted in dark gray. The gray curve corresponds to the derivative of the estimated relationship as the estimate of the recombination rate at every position.

Author's personal copy

We retrieved the 644 genes that had been localized on both the physical and genetic maps in the Release 5.19 of the *D. melanogaster* genome (http://flybase.org/). Using these genes as markers, we plotted the Marey maps (see Fig. 1) for each of the five major chromosomal arms with detectable recombination (2L, 2R, 3L, 3R and X). Because chromosome 4 is known to be largely heterochromatic, recombination rate is thought to be equal to zero or negligible. Thus, the four markers located on this chromosome have been eliminated. We visually identified outliers and removed them ($n = 2, 4, 0, 0$, and 1 outliers on chromosomal arms: 2L, 2R, 3L, 3R and X, respectively). We then fitted a 3rd order polynomial curve to the genetic map position as a function of physical position for the remaining markers on each chromosomal arm ($n = 113, 104, 90, 151$, and 175 genes for chromosomal arms 2L, 2R, 3L, 3R and X, respectively; see Fig. 1).

Some telomeric and centromeric regions of the *D. melanogaster* genome are experimentally known to have a highly reduced recombination rate. This absence of recombination cannot be extrapolated from the Marey maps built using the internal regions of chromosomes, which necessitated setting the transition areas separately. Specifically, we started from each telomere and fit the data with a 3rd order polynomial with an increasing number of data points.

The R^2 starts off close to 1 near the telomere where recombination is similarly low for all points. Then, at some point, when we start adding markers more distal to the telomere with different recombination behavior, the fit becomes substantially worse as the 3rd degree polynomial cannot fit the two distinct behaviors well. By consequence, the R^2 value decreases creating a “valley” in the R^2 distribution (see Supplemental Fig. 1) and then starts climbing again as the additional points in the middle of the chromosomes start exhibiting self-consistent behavior that is different from that of the points at telomeres. We define the regions of exactly zero recombination for the telomeric regions at the points proximal to the lowest point in the R^2 valley (Supplemental Fig. 1). The telomeric points defined in this way indeed all have roughly the same genetic map positions indicating that our approach is reasonable.

For the centromeric regions, we defined the transition to a recombination rate equal to 0 at the point for which the recombination rate estimate based on the polynomial became negative.

Based on the gene data and the two approaches to define the telomere and centromere regions, the precise points at which recombination rates are positive are: from 0.3 Mb to 17.84 Mb, from 2.05 Mb to 20.66 Mb, from 0.86 Mb to 19.45 Mb, from 4.53 Mb to 25.74 and from 1.53 Mb to 22.32 for the chromosomal arms 2L, 2R, 3L, 3R and X, respectively (see Fig. 1). We then recalculated the regression for each chromosomal arm without the defined zero-recombination telomeric and centromeric regions. The median of R^2 from fitted models for all chromosomal arms is equal to 0.99, which means that 99% of the selected data fit well with the model.

The recombination for any locus is estimated as the derivative of the resulting 3rd order polynomial curve at a given nucleotide coordinate (except for the telomeric and centromeric regions where it is set at zero). For any locus, our web-based tool estimates recombination at the start coordinate, end coordinate, and the midpoint. Because of the smoothing effect of this recombination estimation technique, there will be little difference in the recombination rate estimates at these three points for small loci. We provided all three estimations in the event that the locus of interest is sufficiently large for regional variation in recombination rate.

3. Conclusion/Discussion

Our web-based tool provides the local estimates of the meiotic recombination rates for any position in the *D. melanogaster* genome. Rezvoy et al. recently developed a tool called MareyMap offering a user-friendly graphical interface to estimate recombination rate in a variety of model systems (Rezvoy et al., 2007). This tool offers the choice of different interpolation methods. Moreover, user can upload personal data. These two tools are complementary. While our web tool is less flexible than the MareyMap tool and limited to Drosophila, it is based on a highly curated data set and corrects extrapolation of recombination rate at the chromosomal tips. We believe that there is much value in such curated web-based tools. The approach implemented in this web-based tool will be very useful for other organisms.

Acknowledgments

We would like to thank David Sutherland at Flybase who provided us the physical and genetic maps for the previous releases (3 and 4). We are grateful to Peter Andolfatto for his comments, Josefa González for her help in improving the manuscript and James J. Cai for website advice. We also thank the reviewers for their helpful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2010.04.015.

References

Adams, et al., 2000. The Genome Sequence of Drosophila melanogaster. Science 2185.
Berloff, N., et al., 2002. Spline methods for the comparison of physical and genetic maps. J. Comput. Biol. 9, 465–475.
Broman, K.W., et al., 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. 63 (3), 861–869 Sep.
Chakravarti, A., 1991. A graphical representation of genetic and physical maps: the Marey map. Genomics 11, 219–222.
Hey, J., Kliman, R.M., 2002. Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics 160, 595–608.
Ising, G., Block, K., 1984. A transposon as a cytogenetic marker in *Drosophila melanogaster*. Mol. Gen. Genet. 196, 6–16.
Kindahl, E.C., 1994. Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. Cornell University, Ithaca, NY.
Kong, A., et al., 2002. A high-resolution recombination map of the human genome. Nat. Genet. 31 (3), 241–247.
Marais, G., et al., 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. PNAS 98, 5688–5692.
Nachman, M.W., Churchill, G.A., 1996. Heterogeneity in rates of recombination across the mouse genome. Genetics 142 (2), 537–548 Feb.
Rezvoy, C., et al., 2007. MareyMap: an R-based tool with graphical interface for estimating recombination rates. Bioinformatics 23 (16), 2188–2189.
Singh, N.D., Arndt, P.F., Petrov, D.A., 2005. Genomic heterogeneity of background substitutional patterns in *Drosophila melanogaster*. Genetics 169 (2), 709–722 Feb.
Sorsa, V., 1988. Chromosome Maps of Drosophila. CRC Press, Boca Raton, FL.
Yu, A., et al., 2001. Comparison of human genetic and sequence-based physical maps. Nature 409, 951–953.