Supplementary Material:

Two-Parameter Characterization of Chromosome-Scale Recombination Rate

Wentian Li and Jan Freudenberg

1 Marey Map for human chromosomes (Fig.S1)

2 Checking the normality assumption of the regression (Fig.S2)

3 Testing the robustness of regression result by adding random noise to the genetic map (Fig.S3)

4 AIC and BIC of two-parameter regression models vs. one-parameter models (Table S1)

5 Chicken genetic length vs. physical length in log-log scale (Fig.S4)

6 Two-parameter regression model of Opossum chromosome genetic length (Fig.S5)

7 Recombination rates of six genomes as a function of the smallest chromosome size (Fig.S6)

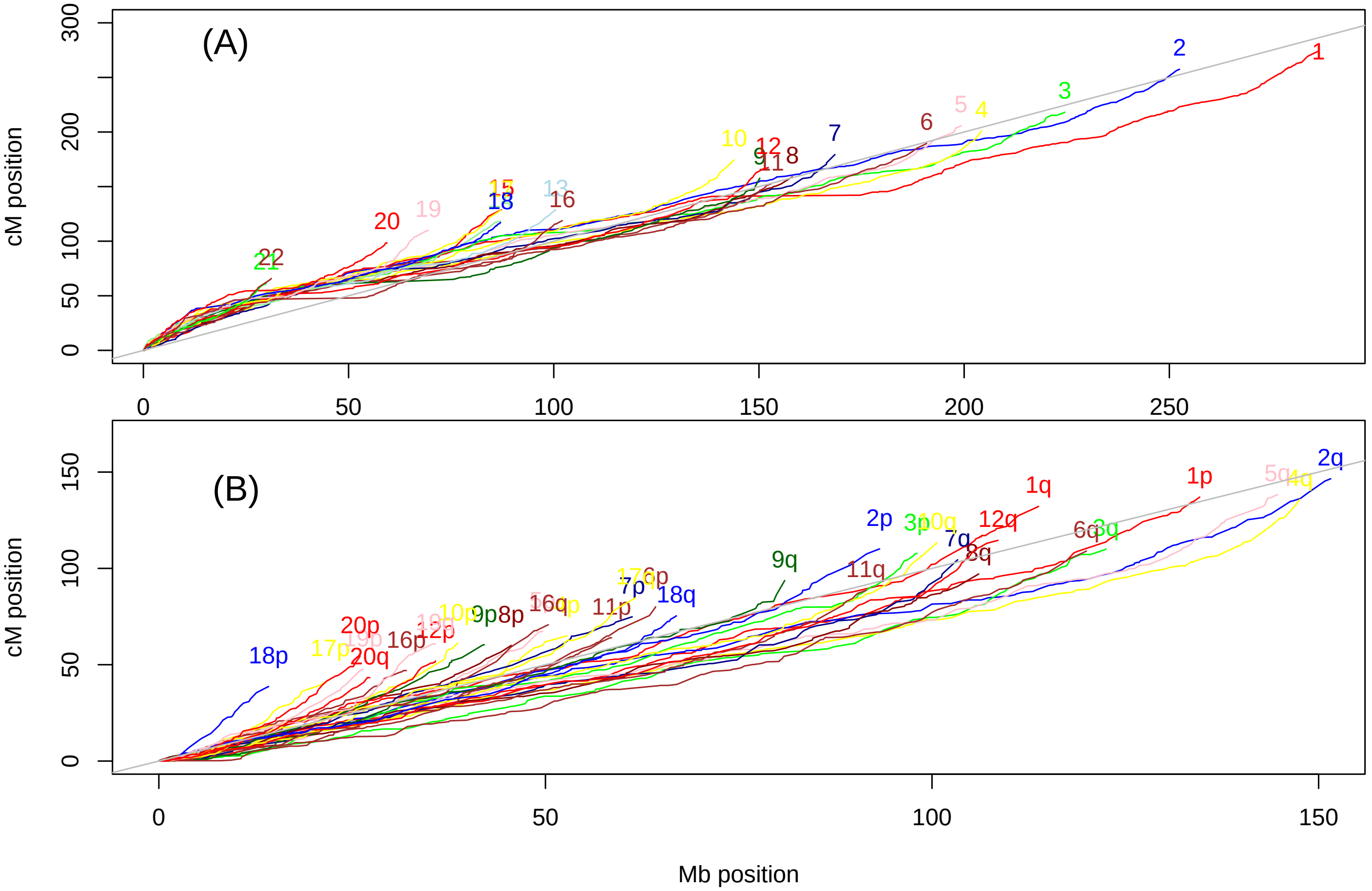

Figure S1: Marey map for human genetic data (Supplementary Table E of (Kong et al., 2002), with y-axis showing the genetic distance (in cM) from the first marker to the current marker, and x-axis showing the physical distance (in Mb). (A) Each line traces a chromosome (chromosome name is shown as label). (B) Each line traces an arm of a meta-centric chromosome (p- and q-arms are shown as label). The straight line indicates cM=Mb.

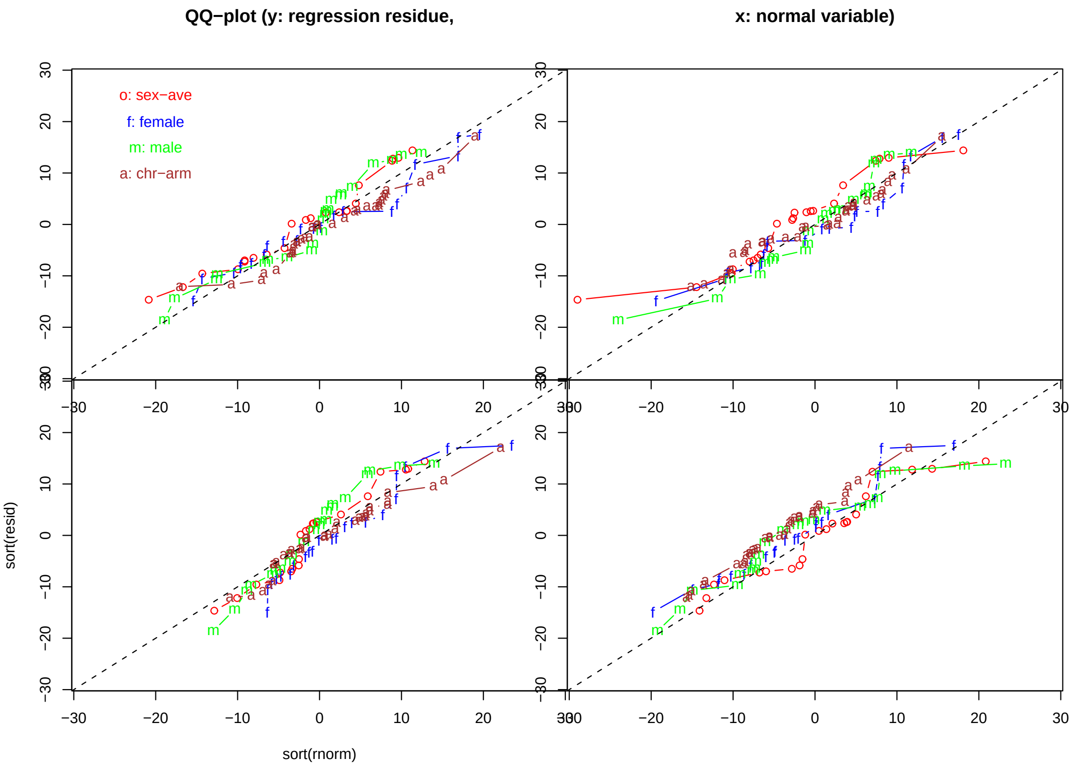

Figure S2: QQ-plots of the residuals of regression ($\epsilon = G - G_0 - kP$) against simulated normal variables. The variance of the normal variables is chosen to be equal to that of the regression residuals. Due to the small number of sample points (22 for human chromosomes, 34 for human chromosome arms), four sets of normal random variables were generated to test the robustness. In each QQ-plot, "o" denotes the regression for the sex-averaged map of human chromosomes, "f" for the human female map data, "m" for human male map data, and "a" for sex-averaged human map data for chromosome arms (see Eqs.(1,2)).

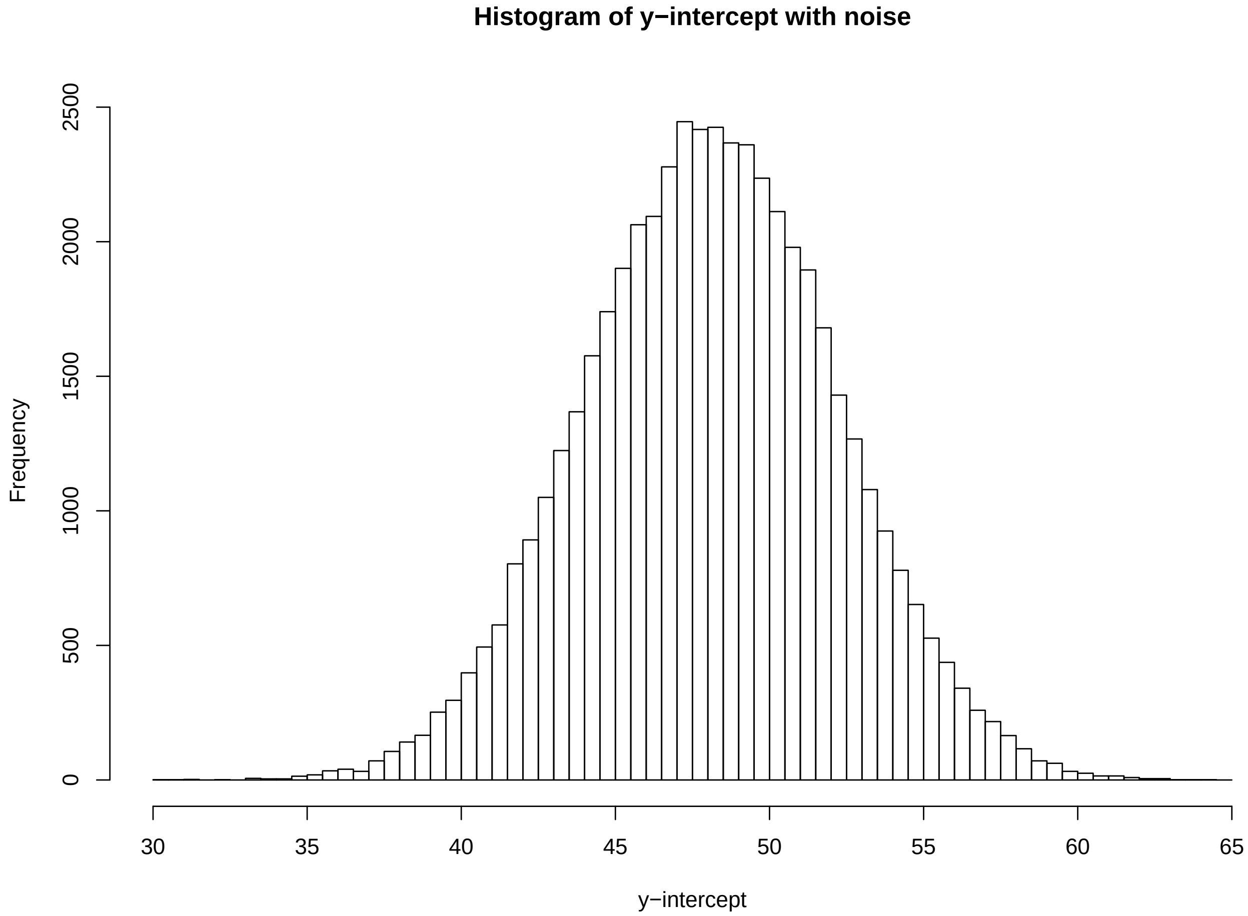

Figure S3: Histogram of G_0 for human sex-averaged chromosome regression $G_{ch,sex-ave,human} = G_0 + kP$ when noise is added to the genetic length G. The noise was modeled as a normally distributed variable with zero-mean and standard deviation (sd) of 8.51517, which is the observed sd of regression residuals $\epsilon = G - G_0 - kP$, for $G_{ch,sex-ave,human} \sim P$ in Eq.(1).

source of genetic length data	ΔAIC= AIC$_2$−AIC$_1$	ΔBIC= BIC$_2$−BIC$_1$
human chromosome, sex-averaged	−42.5	−41.4
female	−45.3	−44.2
male	−34.5	−33.4
human ch. arm, sex-averaged	−52.9	−51.4
female	−43.8	−42.3
male	−43.6	−42.1
rat, chromosome, sex-averaged	−9.9	−8.9
mouse, chromosome, sex-averaged	−1.1	−0.2
chicken, chromosome, sex-averaged	−28.7	−27.5
honeybee, chromosome, sex-averaged	1.9	2.7
remove chromosome 1	1.0	1.7
yeast, chromosome, sex-averaged	−6.1	−5.3
worm,central region	1.4	0.96
distal region	−7.9	−7.6

Table S1: Difference of Akaike and Bayesian information criterion (AIC and BIC) between the two- and one-parameter regression models for modeling chromosome-scale recombination in 7 genomes. A negative ΔAIC or ΔBIC value favors the two-parameter regression model.

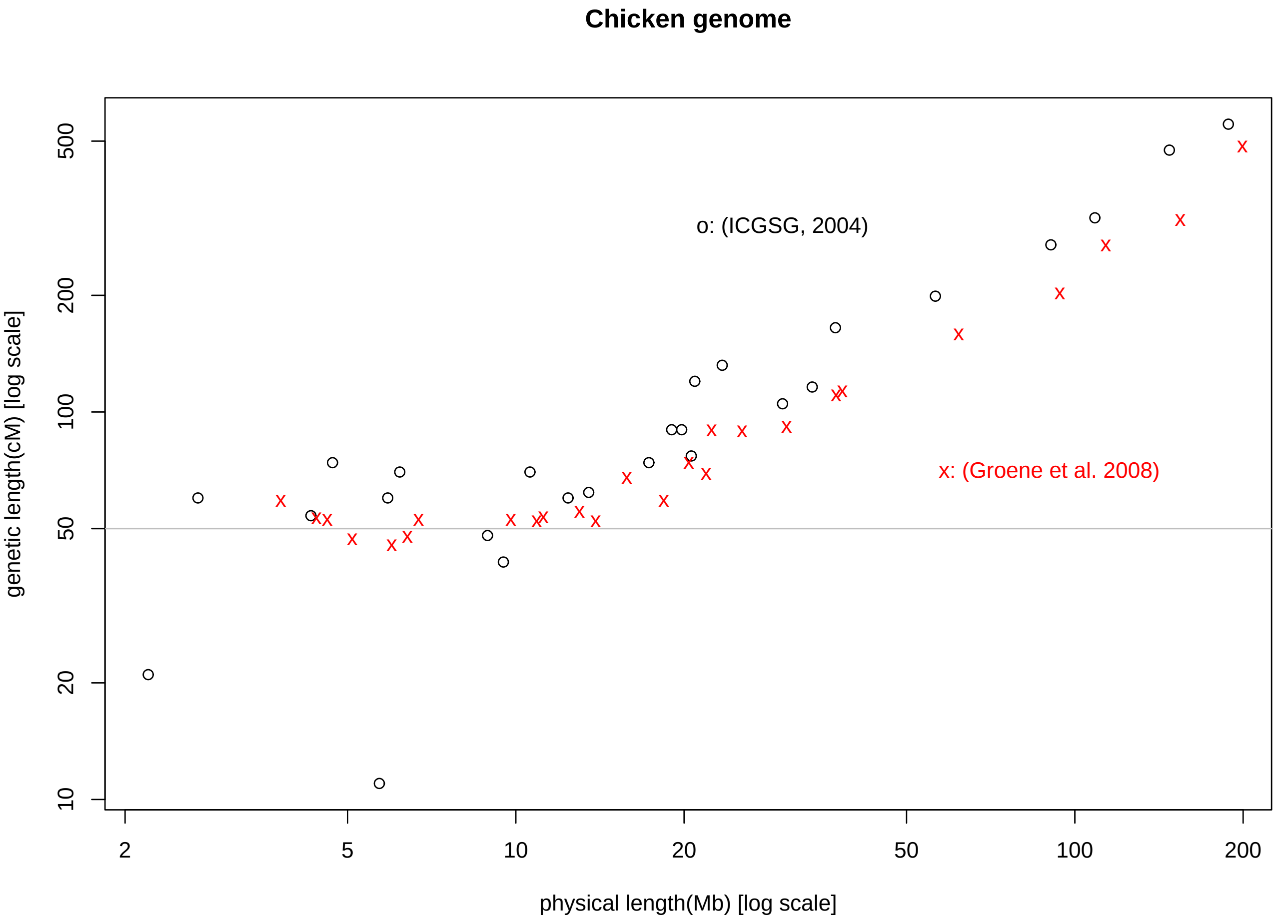

Figure S4: The genetic length (in cM) vs. physical length (in Mb) plots for chicken genome (*Gallus gallus*) in log-log scale (for linear-linear scale, see Fig.2(B)).

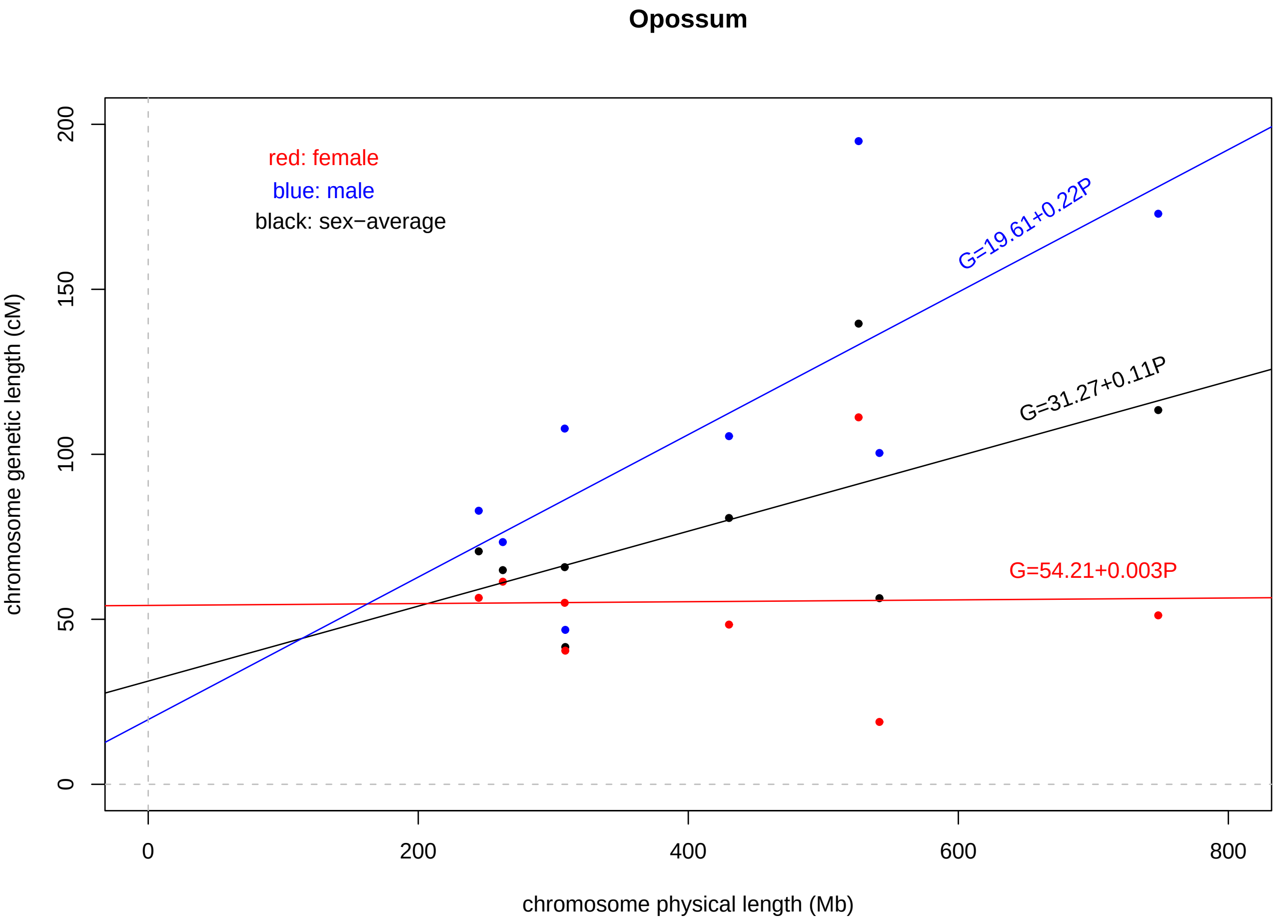

Figure S5: Two-parameter regression of Opossum *Monodelphis domestica* chromosome genetic length (cM) over physical length (Mb). The regression coefficients for female map (red) are: $G_{female} = 54.206 + 0.003P$, for male map (blue): $G_{male} = 19.610 + 0.216P$, and for sex-averaged map (black): $G_{sex-ave} = 31.275 + 0.114P$.

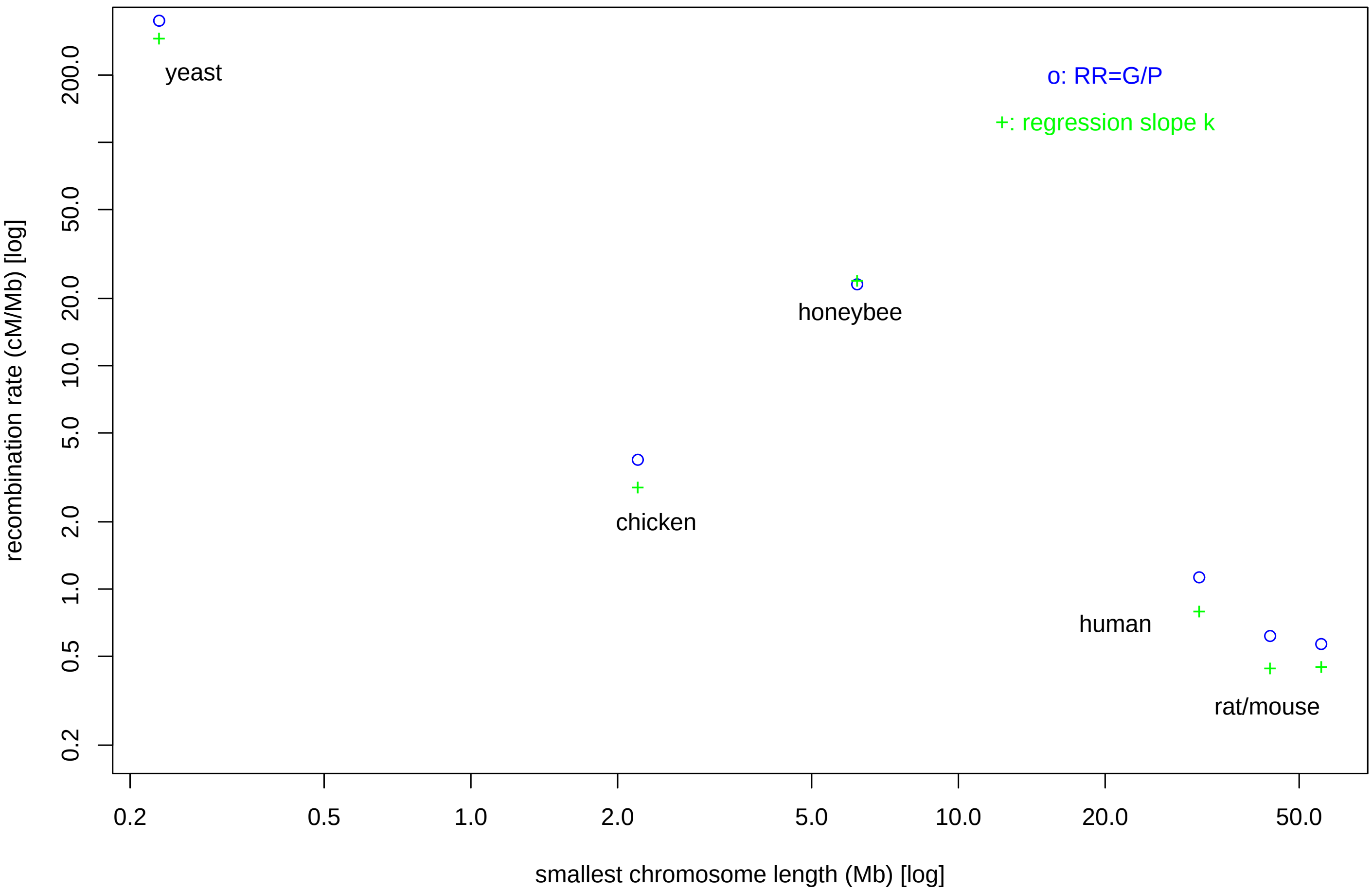

Figure S6: Recombination rates of six genomes as a function of the smallest chromosome size. Genome-wide recombination rate is measured both by the genome-wide genetic-to-physical length ratio ($RR = G/P$) (circles) and by the regression slope k (pluses). The plot is in log-log scale.