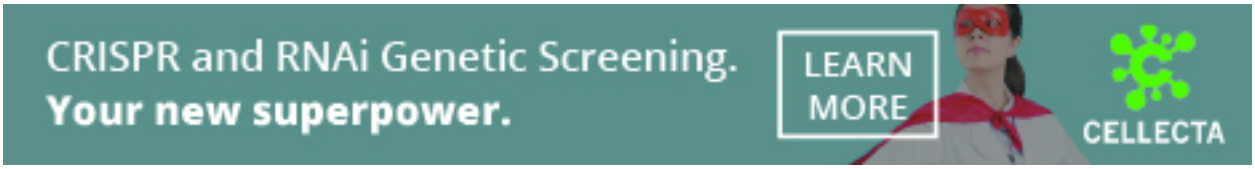
Karyotype distributions in a stochastic model of reciprocal translocation.

D Sankoff and V Ferretti

Genome Res. 1996 6: 1-9
Access the most recent version at doi:10.1101/gr.6.1.1



RESEARCH

Karyotype Distributions in a Stochastic Model of Reciprocal Translocation

David Sankoff[1] and Vincent Ferretti

Centre de Recherches Mathématiques, Université de Montréal, Québec H3C 3J7, Canada

A random process of reciprocal translocation for a fixed number k of chromosomes (or arms) will have an equilibrium distribution of chromosome lengths. In this paper we calculate this distribution, by analytical means for $k = 2$ and partially for $k = 3$, and simulate the means of the marginal distributions for higher k. We compare this with a random (i.e., ahistorical) distribution of genomic DNA among k chromosomes and to a selection of karyotypes of real organisms. The results motivate a revised model where translocations giving rise to undersize chromosomes are disadvantaged.

The number, size, and centromeric position of its chromosomes are the most evident properties of the karyotype of a species. Because overall genomic DNA content is rather variable and does not have systematic phylogenetic pertinence, the distribution of chromosome, or chromosome arm, length (measured cytogenetically, genetically, or as DNA content), normalized by total length, is a meaningful characteristic of a given organism for comparative purposes. Over the course of evolution, the gross characteristics of a karyotype are altered by processes such as genome fusion, chromosome fusion and fission, reciprocal translocation, paracentric inversions, duplication, deletion, and insertion of genomic material. It is a tenet of mammalian genomics that the distribution of conserved chromosomal segments evident in the comparison of two relatively divergent species can be accounted for by repeated reciprocal translocations, each involving two breakpoints occurring more or less at random along the arms of two chromosomes (Nadeau and Taylor 1984), though of course noncoding regions and heterochromatin, centromeric, and telomeric regions have all been cited as particularly susceptible to the breaking process.

From an evolutionary point of view, a reciprocal translocation occurs when arms of two chromosomes break simultaneously and are each rejoined to the "wrong" chromosome (for detailed descriptions, see Schulz-Schaeffer 1980; Swanson et al. 1981). A random process of reciprocal translocation for a fixed number k of chromosomes (or arms) will have an equilibrium distribution of chromosome lengths. In this paper we calculate this distribution, by analytical means for $k = 2$ and partially for $k = 3$, and simulate the density for higher k. We compare this with a random (i.e., ahistorical) distribution of genomic DNA among k chromosomes and with a selection of karyotypes of real organisms. The results motivate a revised model where translocations giving rise to undersize chromosomes are disadvantaged.

[1]**Corresponding author.**
E-MAIL sankoff@ere.umontreal.ca; FAX (514) 343-2254.

Random Reciprocal Translocations

We define a stochastic model for $k \geq 2$ chromosomes without taking into account the fact that the chromosomal segments exchanged by translocations do not contain centromeres. This same model can be used, and is perhaps more properly used, when k represents the number of arms. Let $l_1, \ldots, l_k$ be the lengths of the k chromosomes of a karyotype at time t, where $l_1 \geq \ldots \geq l_k$ and where $\Sigma_i l_i = 1$. Choose two different chromosomes, for example, the ith and the jth, according to some probability distribution $P(i,j)$, which is either uniform ($=1/k$) or depends on the lengths l_i. Pick a breakpoint at random on each of the two chromosomes, breaking them into segments of length Ul_i, $(1 - U)l_i$, Vl_j, $(1 - V)l_j$, respectively. Then we reform a karyotype at time $t + 1$ containing chromosomes of length $l_1, \ldots, Ul_i + Vl_j, \ldots, (1 - U)l_i + (1 - V)l_j, \ldots, l_k$, which then must be reindexed so that the lengths of the chromosomes are in a monotone nonincreasing order.

This process is repeated indefinitely. As the

number of iterations approaches infinity, the probability that the length of the ith longest chromosome is in a certain interval will converge. Let $q(l_1, \ldots, l_k)$ be the joint equilibrium probability density of the lengths of the longest, second longest, . . . , shortest chromosome, respectively. The following sections are devoted to the calculation of this density.

The Two-chromosome Case

To simplify the notation, let $x = l_1$ and $1 - x = l_2$ be the lengths of the two initial chromosomes, and let U and V be two independent random numbers between 0 and 1. Then the two new chromosomes have lengths $A = Ux + V(1 - x)$ and $1 - A = (1–U),\ x + (1 - V)(1 - x)$, respectively. Let $Y = \text{Max}[A,1 - A]$ be the length of the longer of the two, and let $F_x(y) = \text{Prob}[Y \leq y|x]$.

Consider the two-dimensional square $[0, 1] \times (0, 1]$ that is the domain of (U,V). When $A \geq 1 - A$, then $Y \leq y$ if U is between the lines $Ux + V(1 - x) = ½$ and $Ux + V(1 - x) = y$, as indicated in Figure 1. This has area

$$\frac{2y-1}{2x} \text{ if } y \leq x \text{ or } \frac{1}{2} - \frac{(y-1)^2}{2x(1-x)}$$

if $x \leq y \leq 1$. When $A \leq 1 - A$, by symmetry an equal area is contributed to the probability that $Y \leq y$. Then

$$F_x(y) = \frac{2y-1}{x}, \text{ if } \frac{1}{2} \leq y \leq x$$

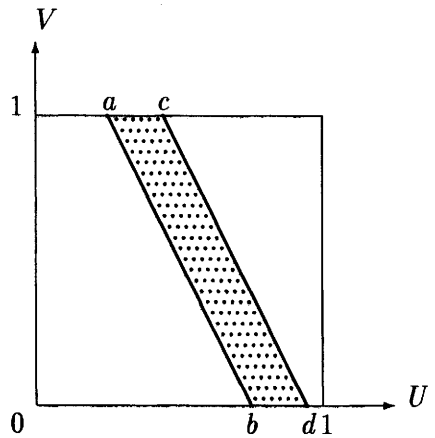


Figure 1 Areas corresponding to length distribution delimited by the line $Ux + V(1 - x) = ½$ joining points a and b and the line $Ux + V(1 - x) = y$ joining points c and d.

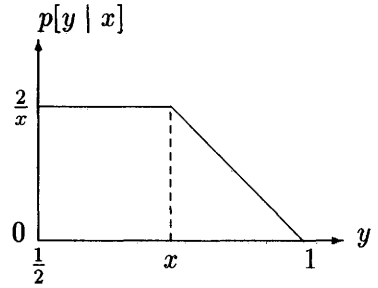


Figure 2 Probability density for length of longer chromosome.

$$= 1 - \frac{(y-1)^2}{x(1-x)}, \text{ if } x \leq y \leq 1.$$

The density of this probability is

$$p[y|x] = \frac{2}{x}, \text{ if } \frac{1}{2} \leq y \leq x$$
$$= \frac{2(1-y)}{x(1-x)}, \text{ if } x \leq y \leq 1.$$

as depicted in Figure 2.

Now that we know the density $p(y|x)$ for each x, we can look for the equilibrium density $q(y)$; in our original notation $q^{(1)} = 1 - q^{(2)}$. The equilibrium $q(y)$ must satisfy

$$q(y) = \int_{½}^{1} q(x)p(y|x)dx$$
$$= 2(1-y)\int_{½}^{y} \frac{q(x)}{x(1-x)}dx + 2\int_{y}^{1} \frac{q(x)}{x}dx.$$

Differentiating twice, we obtain the differential equation

$$y(1-y)q''(y) + 2q(y) = 0,$$

whose solution is

$$q(y) = 12y(1-y)$$

on the interval [½, 1]. The mean of the density q is 11⅛.

How do these results compare with other random processes for dividing the interval [0,1] into two segments? The simplest such process would cut the interval at a point randomly chosen in the interval and then take the largest piece as l_1 and the other as l_2. In this case the mean of the equilibrium density would be ¾, which is larger than 11⅛.

Is there biological evidence that might decide between the translocation model and the random lengths model? Unfortunately, there are not many species with only two chromosomes. One well-known example is the grass *Haplopappus gra-*

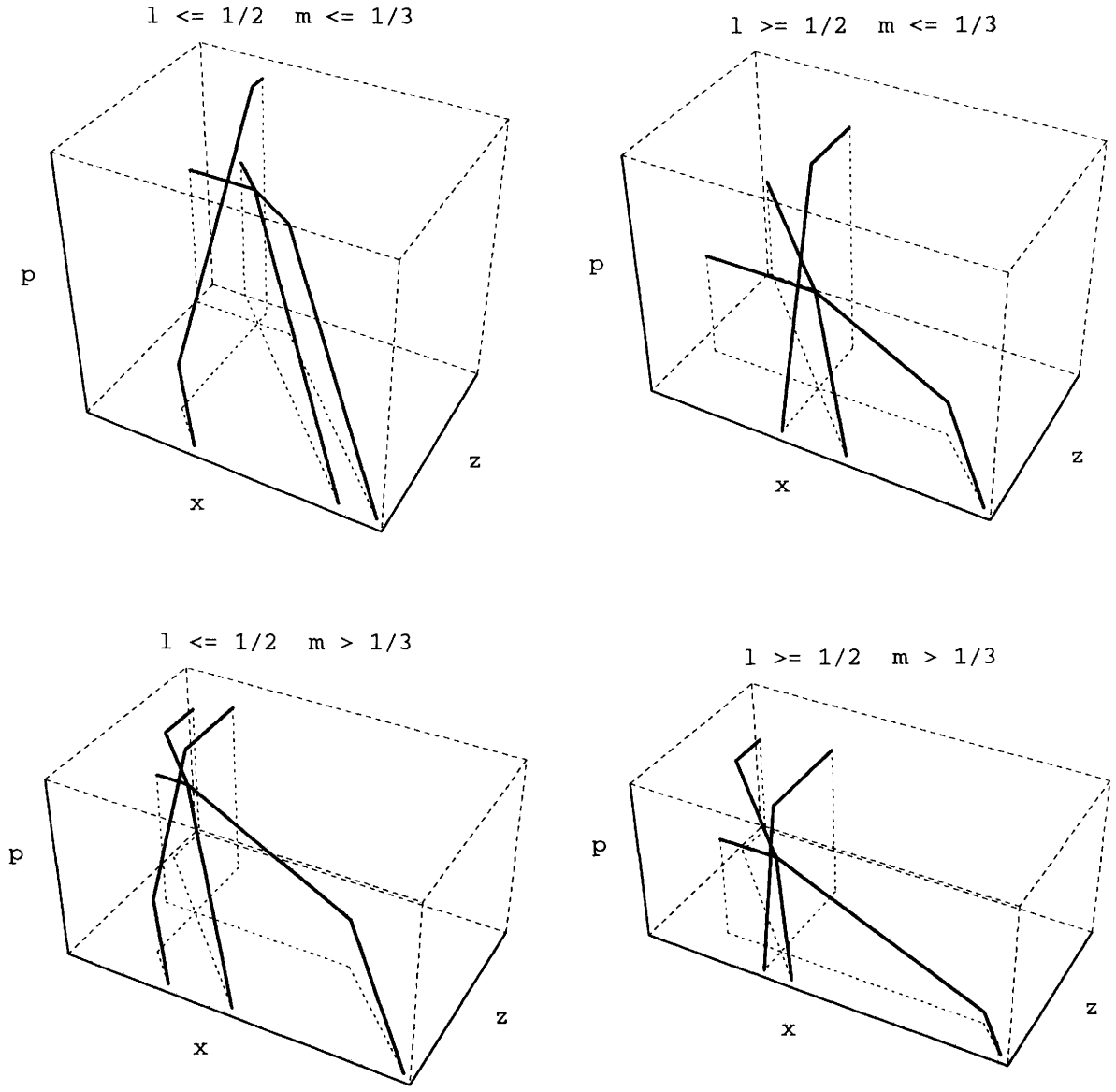


Figure 3 Joint probability densities for longest and shortest chromosomes.

cilis (Jackson 1957), where the sizes of the larger and smaller chromosomes are in the ratio of 5:3 (or 62.5:37.5). Thus, the translocation model (69:31) fits better than the random lengths model (75:25), though we cannot place too much weight on this single case.

Three Chromosomes

Because each translocation involves just two chromosomes, the analysis for three or more chromosomes reduces in some aspects to the case $k = 2$. Complications arise, however, because the two new chromosomes resulting from a translocation involving the ith and the jth largest chromosome may change the rank of the lengths of several or all of the chromosomes unaffected by the translocation itself.

To model the translocation process, we need to specify how pairs of chromosomes are chosen for each event. The most natural postulate is that the probability $P(i,j)$ of choosing the ith and the jth largest chromosome is proportional to their lengths:

$$P(i,j) = l_i \frac{l_j}{1-l_i} + l_j \frac{l_i}{1-l_j}$$
$$= l_i l_j \left(\frac{1}{1-l_j} + \frac{1}{1-l_i} \right),$$

Table 1. Simulated Mean Chromosome Lengths l_i for Karyotypes of Varying Numbers of Chromosomes k, Based on the Proportional Model (M_1), the Uniform Model (M_2), and Random Fragmentation (M_3)

$k = 2$			
l_i	M_1	M_2	M_3
l_1	0.313	0.312	0.250
l_2	0.687	0.688	0.750

$k = 3$			
l_i	M_1	M_2	M_3
l_1	0.122	0.160	0.111
l_2	0.298	0.304	0.277
l_3	0.580	0.536	0.611

$k = 4$			
l_i	M_1	M_2	M_3
l_1	0.067	0.101	0.062
l_2	0.153	0.180	0.146
l_3	0.279	0.275	0.271
l_4	0.501	0.443	0.520

$k = 5$			
l_i	M_1	M_2	M_3
l_1	0.040	0.070	0.040
l_2	0.092	0.121	0.090
l_3	0.161	0.177	0.156
l_4	0.260	0.250	0.257
l_5	0.446	0.381	0.457

$k = 10$			
l_i	M_1	M_2	M_3
l_1	0.010	0.023	0.010
l_2	0.021	0.038	0.021
l_3	0.034	0.051	0.033
l_4	0.048	0.065	0.048
l_5	0.065	0.079	0.064
l_6	0.085	0.095	0.084
l_7	0.110	0.113	0.109
l_8	0.143	0.136	0.143
l_9	0.193	0.169	0.193
l_{10}	0.290	0.231	0.293

$k = 20$			
l_i	M_1	M_2	M_3
l_1	0.002	0.008	0.002
l_2	0.005	0.012	0.005
l_3	0.008	0.016	0.008
l_4	0.011	0.019	0.011
l_5	0.013	0.023	0.013
l_6	0.017	0.026	0.017
l_7	0.021	0.030	0.021
l_8	0.025	0.033	0.025
l_9	0.029	0.037	0.029
l_{10}	0.033	0.040	0.033
l_{11}	0.038	0.044	0.038
l_{12}	0.044	0.049	0.044
l_{13}	0.050	0.053	0.050
l_{14}	0.057	0.059	0.057
l_{15}	0.066	0.065	0.066
l_{16}	0.076	0.071	0.076
l_{17}	0.088	0.080	0.088
l_{18}	0.105	0.090	0.105
l_{19}	0.130	0.106	0.130
l_{20}	0.180	0.136	0.180

where l_i and l_j are the lengths of the two chromosomes. In Simulations (below) we also discuss the model where this probability is $1/\binom{k}{2}$, independent of the lengths of the chromosomes.

In the case $k = 3$, given initial chromosome lengths $l \geq m \geq n$, the joint probability distribution of the length X of the longest and Z of the shortest of the three new chromosomes after a single translocation event[1] is

$$F_{l,n}(x,z) = \sum_{1 \leq i < j \leq 3} P(i,j) F_{l,n}^{(i,j)}(x,z),$$

where $F_{l,n}^{(i,j)}(x,z)$ is the distribution of these lengths given that ith and the jth largest chromosomes are involved in the translocation.

The quantity $F_{l,n}^{(i,j)}(x,z)$, is calculated in much the same way as $F_x(y)$ in The Two-chromosome Case (above), except that keeping track of the ranks of the lengths is more complicated. Consider for example the case $(i,j) = (2,3)$, where the second and third largest chromosomes, of length m and n, respectively, are involved in the translocation. Then $X = \text{Max}[Um + Vn, (1-U)m + (1-V)n, l]$, $Z = \text{Min}[Um + Vn, (1-U)m + (1-V)n, l]$, and two subcases are to be considered:

(1) $l \geq \frac{1}{2}$. Here, $X \equiv l$, so

$$F_{l,n}^{(2,3)}(x,z) = 0, \text{ for } x < l,$$

and

$$\begin{aligned} F_{l,n}^{(2,3)}(x,z) &= \text{Prob}[Z \leq z],\ x \geq l, \\ &= z^2/mn,\ 0 \leq z \leq n \\ &= \frac{2z - n}{m},\ n \leq z \leq \frac{m+n}{2} \\ &= 1,\ \frac{m+n}{2} \leq z \leq \tfrac{1}{3}, \end{aligned}$$

as can be calculated in much the same way as in The Two-chromosome Case.

(2) $l < \frac{1}{2}$. Here $l \leq X \leq m + n$, so

$$F_{l,n}^{(2,3)}(x,z) = 0, \text{ for } x < l$$

and

$$F_{l,n}^{(2,3)}(x,z) = P[Z \leq z], \text{ for } x > m + n,$$

where $P[Z < z]$ is given in case 1 above. For $l \leq x \leq m + n$, $F_{l,n}^{2,3}(x,z)$ corresponds to the area of the set of points $(U,V) \in [0,1] \times [0,1]$ for which $X \leq x$ and $Z \leq z$.

[1]Given that the lengths of the chromosomes sum to 1, the length Y of the second largest new chromosome is determined by X and Z.

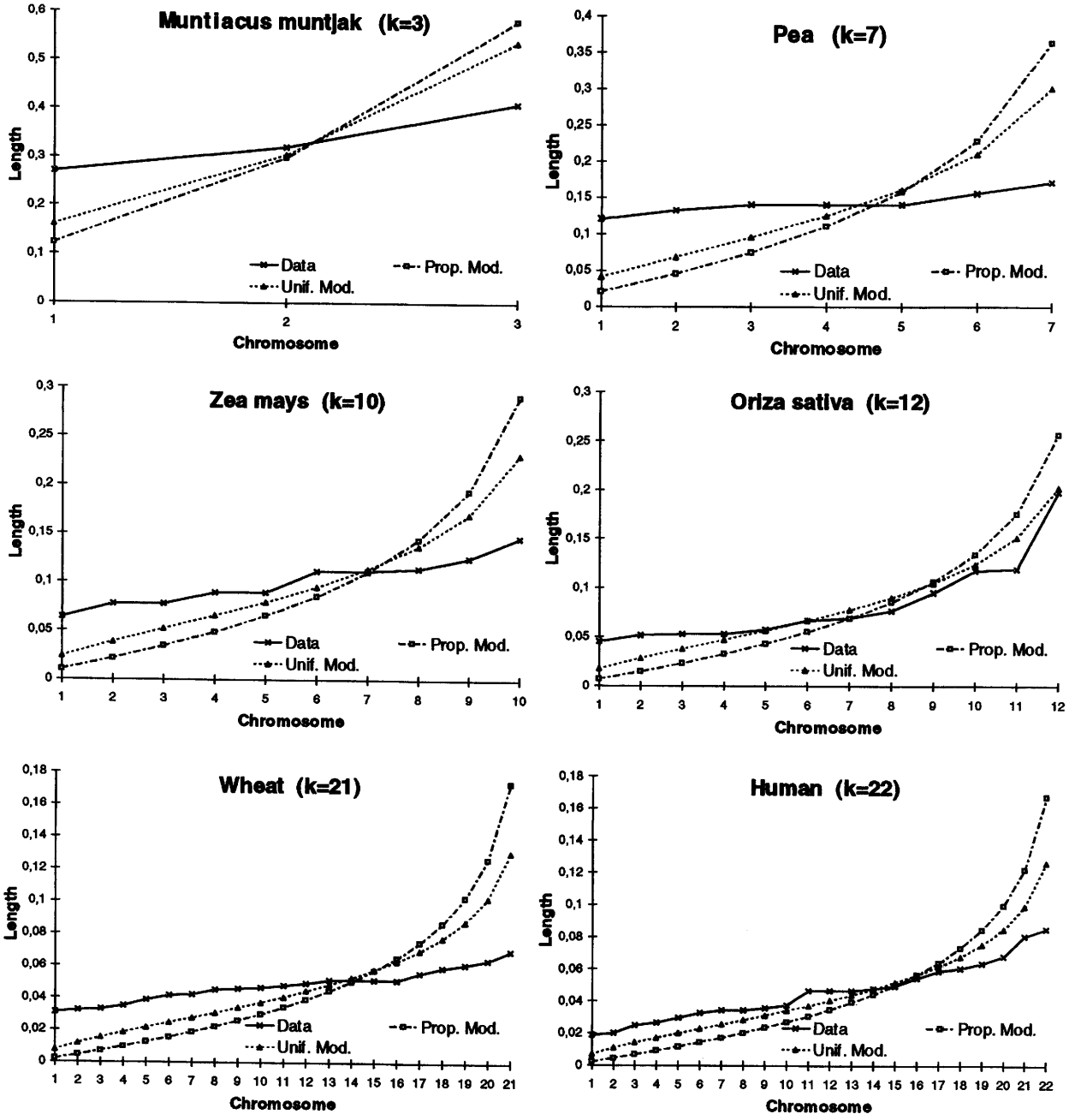


Figure 4 Comparison of simulated mean chromosome lengths, based on the proportional and uniform models, with karyotypes from six species. The corresponding *NSS* values for the proportional and uniform model are, respectively, *Muntiacus muntjak*, 0.052, 0.028; pea, 0.061, 0.031; *Zea mays*, 0.033, 0.014; *Oriza sativa*, 0.011, 0.004; wheat, 0.021, 0.009; human, 0.010, 0.003.

$$F_{l,n}^{(2,3)}(x,z) = 0,\ 0 \leqslant z \leqslant m+n-x$$

$$= \frac{z^2 - x^2 + 2x(m+n) - (m+n)^2}{mn},\quad m+n-x \leqslant z \leqslant n$$

$$= \frac{2nz - x^2 + 2x(m+n) - n^2 - (m+n)^2}{mn},\quad n \leqslant z \leqslant \frac{m+n}{2}$$

$$= \frac{2x(m+n) - x^2 + mn - (m+n)^2}{mn},\quad \frac{m+n}{2} \leqslant z \leqslant \text{⅓}$$

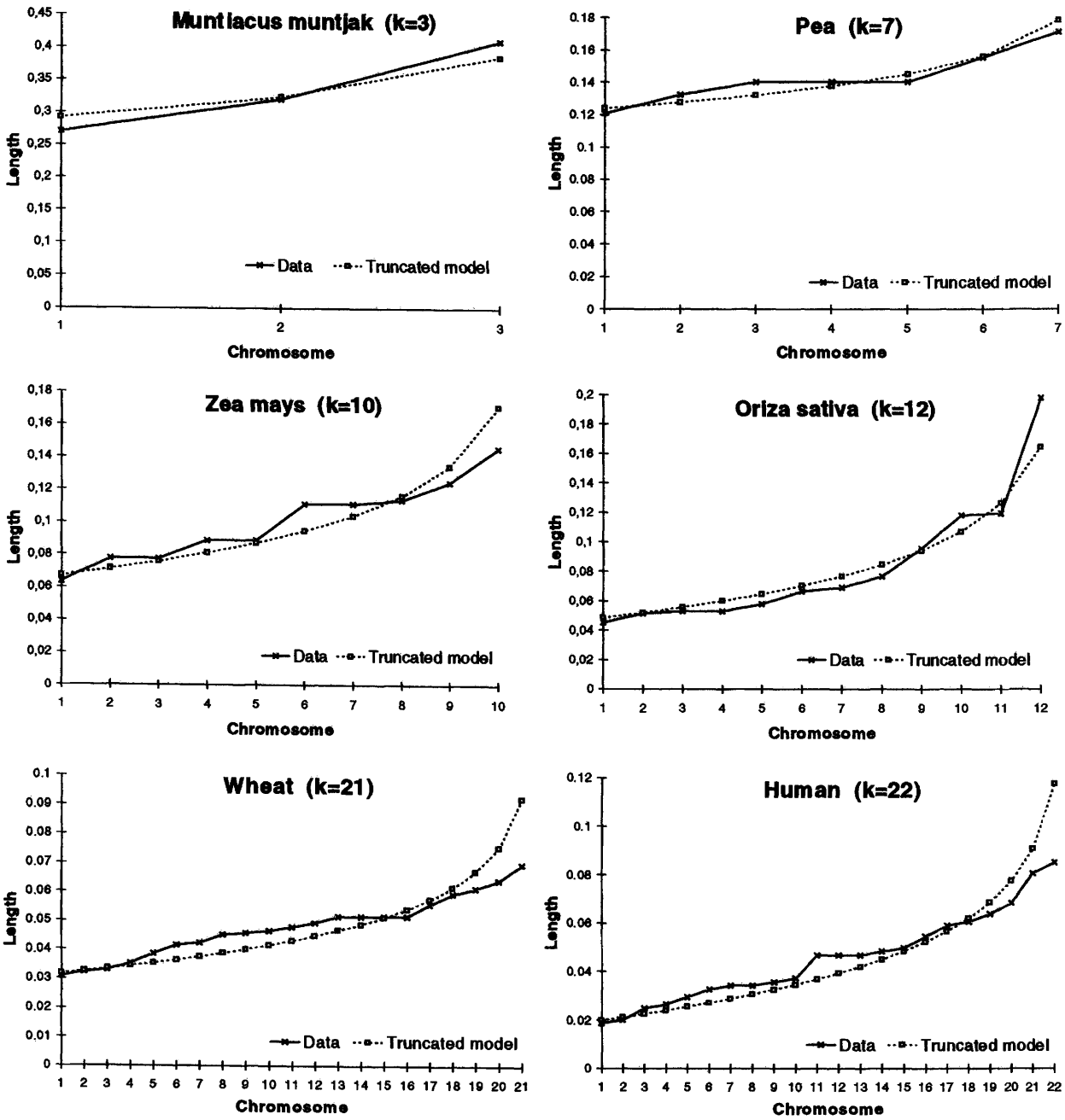


Figure 5 Comparison of simulated mean chromosome lengths, based on the truncated proportional model, with karyotypes from six species. The corresponding *NSS* values for this model are *Muntiacus muntjak*, 0.001; pea, 0.002; *Zea mays*, 0.001; *Oriza sativa*, 0.001; wheat, 0.0007; human, 0.001.

The density $p^{(2,3)}(x,z|l,n)$ of this probability distribution vanishes everywhere in the domain $[⅓, 1] \times [0, ⅓]$ except on the lines $x = m + n - z$ and $x = l$. More precisely, in case 1

$$p^{(2,3)}(x,z|l,n) = 0, \text{ if } x \neq l$$

and

$$p^{(2,3}(l,z|m,n) = \frac{2z}{mn}, \text{ if } 0 \leqslant z \leqslant n$$
$$= \frac{2}{m}, \text{ if } n \leqslant z \leqslant \frac{m+n}{2}$$
$$= 0, \text{ if } \frac{m+n}{2} \leqslant z \leqslant ⅓ .$$

In case 2

$$p^{(2,3)}(x,z|l,n) = 0, \text{ if } x \neq l \text{ or } x \neq m+n-z,$$

$$p^{(2,3)}(m+n-z,z|l,n) = \frac{2z}{mn},\ 0 \leq z \leq m+n-l,$$

$$p^{(2,3)}(l,z|m,n) = \frac{2z}{mn},\ m+n-l \leq z \leq n$$

$$= \frac{2}{m},\ n \leq z \leq \frac{m+n}{2}$$

$$= 0,\ \frac{m+n}{2} \leq z \leq \tfrac{1}{3}.$$

Similar analyses yield $p^{(1,2)}$ and $p^{(3,1)}$. Each of Figure 3a-d depicts the three conditional densities for one of the four regions created by the two boundaries $l = \frac{1}{2}$, $n = \frac{1}{3}$. Weighting these three densities by $P(i,j)$ and summing them yields $p(x,z|l,n)$. Because the three conditional densities are concentrated on one-dimensional subspaces of the (x,z) space, which are disjointed except for one point at which all three intersect, $p(x,z|l,n)$ has essentially the composite form of $p^{(1,2)}$,$p^{(2,3)}$, and $p^{(3,1)}$.

Setting

$$p(x,z|l,n) = \sum_{1 \leq i < j \leq 3} P(i,j)p^{(i,j)}(x,z|l,n),$$

the equilibrium density q should satisfy the integral equation

$$q(x,z) = \int_{\frac{1}{3}}^{\frac{1}{2}} \int_{1-2l}^{1-\frac{1}{2}} p(x,z|l,n)q(l,n)dn\,dl + \int_{\frac{1}{2}}^{1} \int_{0}^{1-\frac{1}{2}} p(x,z|l,n)q(l,n)dn\,dl.$$

The solution to this equation requires investigating separately the dozens of regions within which each of the $p^{(i,j)}$ does not change form, and it is not known whether there is a simple expression for the solution analogous to the case $k = 2$.

Simulations

The difficulties already encountered for $k = 3$ oblige us to undertake computer simulations to estimate the expected length of the longest, second longest, . . . , kth longest chromosome, for $k \geq 3$. If $q(l_1, \ldots, l_k)$ is the equilibrium joint density function on the domain $l_1 \geq \ldots \geq l_k$, our task was to estimate $E_q(l_i)$, for $i = 1, \ldots, k$. Our approach was simply to carry out the experiment described in the Random Reciprocal Translocations (above) for 100,000 steps and to average the lengths of $l_1, \ldots, l_k$ over all the steps.

The experiments were carried out with two choices of weight function $P(i,j)$. First, we postulated that $P(i,j)$ is proportional to the lengths l_i and l_j:

$$P(i,j) = l_i \frac{l_j}{1-l_i} + l_j \frac{l_i}{1-l_j} = l_i l_j \left(\frac{1}{1-l_j} + \frac{1}{1-l_i} \right).$$

A second set of runs assumed this probability to be $1/\binom{k}{2}$, independent of the lengths of the chromosomes, and we will call this the uniform model.

In addition, the results of the translocation experiments were compared with the coutcome of simply fragmenting the unit interval into k segments, using $k-1$ random breakpoints selected according to the uniform distribution.

Table 1 shows that aside from small values of k the proportional translocation model is very close to the random fragmentation model. We also see in Table 1 that the length-independent translocation model results in a more uniform distribution of expected lengths, whereas the proportional model predicts a wider range of lengths.

Comparisons with Some Known Karyotypes and a Truncated Model

In The Two-chromosome Case (above), we showed how the proportional translocation model fits the *H. gracilis* data better than the random lengths model. Similarly, we compared karyotypes (chosen for illustrative purposes from among those depicted in King 1975; Lima-de-Faria 1980; Swanson et al. 1981) from species with a range of values of k (Fig. 4) with the simulations in Simulations (above). As measured by a normalized sum of squares

$$NSS = \frac{1}{k} \sum_{i=1}^{k} \frac{(l_i - L_i)^2}{L_i},$$

where L measures the empirical lengths, the uniform model fits somewhat more closely than either the proportional model or the random fragmentation model. It can be seen, however, that the predictions of all translocation models are systematically biased toward too large a range of chromosome lengths and that this bias is more important than the differences between the models.

Physical chemical considerations of rates of chromosome transport during mitosis and meiosis suggest that genomes combining very large and very small chromosomes might be at

a disadvantage. From the point of view of modeling, this could be handled by prohibiting any translocation resulting in a chromosome of length below a certain threshhold. This "truncation" approach is also justified at the cytogenetic level where a viable and functional chromosome must minimally contain a centromere and two telomeres (and at least one gene whose function is not duplicated elsewhere in the genome). This imposes a lower bound on the size of a chromosome, on a purely structural basis. Finally, from the genetic viewpoint, there is reason to believe that for meiosis to be completed successfully, each chromosome must be of length sufficient for at least one crossover to be expected among the four aligned strands before they segregate into two pairs.

We redid the simulations of the proportional model corresponding to each empirical data set, fixing a threshhold equal to the smallest observed chromosome size. As seen in Figure 5, this results in a great improvement in the fit of the models, greater than might have been expected simply by virtue of adding an additional parameter to the model.

It can be seen that except for the very largest chromosomes in most of the species, the fit is much improved. Given the rather preliminary nature of this exercise, including the choice of karyotypes based only on their fortuitous availability to the authors, no attempt was made to optimize the truncation threshold. We did, however, compare a model with truncation of awkwardly large chromosomes instead of excessively reduced ones. Though the fit with the real data was of course better for the longest chromosomes, it was much worse than the lower bound truncation when it came to the smallest chromosomes, and the overall fit tended to be worse, as measured by the same normalized sum of squares used in Figure 4. Similarly, a comparison with a truncated uniform model was no improvement over the results in Figure 5.

Discussion

Recently, there has been much work on genomic distances (Sankoff et al. 1992; Sankoff 1992, 1993a,b) inferred through the number of inversions (Kececioglu and Sankoff 1994, 1995; Hannenhalli 1995; Hannenhalli and Pevzner 1995), transpositions (Bafna and Pervzner 1995), and/or translocations (Hannenhalli and Pevzner 1995; Kececioglu and Ravi 1995) necessary to transform one observed genome into another. Little work has been done, however, on quantifying the incidence and chromosomal scope of these processes, especially on a comparative basis. For example, the algorithmic inference literature implicitly assumes that all rearrangement events of a given type are equally likely, independent of how large a segment they affect. Further modeling should compare the results of this type of assumption, versus other empirically-motivated weighting schemes, so that inference problems can be formulated and solved in a biologically more meaningful way. Thus, our demonstration of the plausibility of the truncation model should have consequences for the problems studied in Hannenhalli and Pevzner (1995); Kececioglu and Ravi (1995).

It must be acknowledged that no truncation model can be universally satisfactory, for a number of reasons. First, some genomes, for example, in *Aves*, contain large numbers of very small "dot" chromosomes, so that no threshold mechanism seems operative, at least in these cases. Second, and more importantly, translocations resulting in very small chromosomes, especially with any remaining genes duplicated elsewhere, seem just as likely to appear as chromosome fusions, reducing k, and it seems essential to incorporate this possibility into the model.

We have mentioned the necessity of eventually applying our models to chromosome arms, rather than entire chromosomes. This task will be complicated by the process of centromere movement in the course of evolution, often in a systematic way across all chromosomes, as in the mouse genome.

Another direction for research involves the incorporation of heterogeneity of breaking susceptibility of chromosomes along their lengths from the telomeric to centromeric zones and from heterochromatic to euchromatic regions.

ACKNOWLEDGMENTS

We thank Gopalakrishnan Sundaram for his help in setting up the simulation experiments. Thanks are also due to Erica Jen for encouragement and suggestions for the mathematical analysis, to William F. Grant for pointers on the cytogenetics literature and for the references to *H. gracilis* and *M. muntjak*, and to David Baillie, Bronya Keats, and Joseph H. Nadeau for discussions of the truncation model. Research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and

the Canadian Genome Analysis and Technology Program. D.D. is a Fellow of the Canadian Institute for Advanced Research.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bafna, V. and P.A. Pevzner. 1995. Sorting by transpositions. *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 614–623.

Hannenhalli, S. 1995. Polynomial algorithm for computing translocation distance between genomes. Proceedings of the 6th Symposium on Combinatorial Pattern Matching, *Springer-Verlag Lecture Notes Comput. Sci.:* 162–176.

Hannenhalli, S. and P.A. Pevzner. 1995. Transforming cabbage into turnip. (polynomial algorithm for sorting signed permutations by reversals). In *Proceedings of the 27th Annual ACM-SIAM Symposium on the Theory of Computing,* pp. 178–189. ACM, New York, NY.

Jackson, R.C. 1957. New low chromosome number for plants. *Science* **126:** 1115–1116.

Kececioglu, J. and R. Ravi. 1995. Of mice and men. Evolutionary distances between genomes under translocation. *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 604–613.

Kececioglu, J. and D. Sankoff. 1994. Efficient bounds for oriented chromosome inversion distance. *Proceedings of the Fifth Symposium on Combinatorial Pattern Matching,* (Springer Verlag Lecture Notes in Computer Science) **807:** 307–325.

———. 1995. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. *Algorithmica* **13:** 180–210.

King, R.C. 1975. *Handbook of genetics*. Plenum Press, New York, NY.

Lima-de-Faria, A. 1980. How to produce a human with 3 chromosomes and 1000 primary genes. *Hereditas* **93:** 47–73.

Nadeau, J.H. and B.A. Taylor. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Nat. Acad. Sci.* **81:** 814.

Sankoff, D. 1992. Edit distance for genome comparison based on non-local operations. *Proceedings of the Third Symposium on Combinatorial Pattern Matching,* (Springer Verlag Lecture Notes in Computer Science) **644:** 121–135.

———. 1993a. Analytical approaches to genomic evolution. *Biochimie* **75:** 409–413.

———. 1993b. Models and analyses of genomic evolution. In *Second International Conference on Bioinformatics, Supercomputing and Complex Genome Analysis*.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B.F. Lang, and R. Cedergren. 1992. Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. *Proc. Nat. Acad. Sci.* **89:** 6575–6579.

Schulz-Schaeffer, J. 1980. *Cytogenetics*. Springer-Verlag, New York, NY.

Swanson, C.P., T. Merz, and W.J. Young. 1981. *Cytogenetics*, 2nd ed. Prentice Hall, Englewood Cliffs, NJ.

Received May 11, 1995; accepted in revised form December 14, 1995.