Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1421–1422, February 1994
Medical Sciences

Disomic locus content mapping

(radiation hybrid/genetic map)

NEWTON E. MORTON

Cancer Research Campaign, Genetic Epidemiology Research Group, Department of Child Health, University of Southampton, Level G, Princess Anne Hospital, Southampton SO9 4HA, United Kingdom

Contributed by Newton E. Morton, October 25, 1993

ABSTRACT **A linear genetic map may be constructed from segregation of markers in DNA fragments broken by radiation or shearing and/or incorporated into a vector: this is locus content mapping. Theory for monosomic locus content mapping is extended to disomy, making the phase of heterozygous markers informative for linkage and each clone potentially informative for every chromosome. Because there is no interference among breaks and phase may be reliably inferred for all heterozygous loci, multiple pairwise mapping is especially appropriate.**

In a Poisson process the probability of an unbroken segment is exponentially related to the mean number of breaks w. By analogy with meiotic recombination we want to express this probability in terms of a parameter θ that takes the value 0.5 when $w = \infty$, corresponding to random segregation. Random breakage implies additivity of w, which may be scaled to physical units such as kilobases and constitutes a genetic map. If breakage is nonrandom, as with restriction fragments, the method gives order, but estimates of location are unreliable.

Genetic maps and sequences are characterized by resolution, connectivity, and reliability. Linkage maps have the highest connectivity but the lowest resolution, and the converse is true for sequences; physical maps are intermediate in both respects. The Human Genome Initiative is currently preoccupied with connectivity of physical maps. *In situ* hybridization gives order but only a crude estimate of location (1). For sufficiently large samples radiation hybrid mapping gives high connectivity and reliable estimates of location on the assumptions of random breakage and monotonic recovery. (Radiation hybrid mapping panels are much less informative because of clonal instability, multiple breaks, and the low resolution of mapping panels.) Current radiation hybrid maps have been based on rodent cells monosomic for a single human chromosome (2). An older method, introduced before DNA markers, used human cells disomic for all autosomes (3); this method has the advantages that each derived clone is potentially informative for every chromosome and the phase of heterozygous loci is informative about location. We develop here the requisite theory, which is also applicable to other types of locus content mapping (4).

The Monosomic Model

Let the probability of no break between loci i and j be $e^{-w_{ij}} = 1 - 2\theta_{ij}$, where w_{ij} is the mean number of breaks and θ_{ij} is the frequency of recombination (5, 6). In a given region there is a singular point (which may be a centromere, telomere, or locus) that is retained with highest frequency because of positive selection or a locally low density of expressed human loci. This point is usually the centromere, and we shall refer to it as such. Let L be the conditional probability that a locus separated from a proximal locus by a break be retained. Then the probability that both loci are retained is

$$\begin{aligned} P_{ij} &= P_i[e^{-w_{ij}} + (1 - e^{-w_{ij}})L] \\ &= P_i(1 - 2\theta_{ij} + 2\theta_{ij}L), \end{aligned} \quad [1]$$

where i is proximal to j and P_i is the marginal probability that i is retained. The probability that only i is retained is

$$\begin{aligned} P_{i\cdot} &= P_i(1 - e^{-w_{ij}})(1 - L) \\ &= P_i(2\theta_{ij})(1 - L). \end{aligned} \quad [2]$$

The probability that only j is retained is

$$\begin{aligned} P_{\cdot j} &= (1 - P_i)(1 - e^{-w_{ij}})L \\ &= (1 - P_i)2\theta_{ij}L. \end{aligned} \quad [3]$$

The probability that neither is retained is

$$\begin{aligned} P_{\cdot\cdot} &= (1 - P_i)[e^{-w_{ij}} + (1 - e^{-w_{ij}})(1 - L)] \\ &= (1 - P_i)(1 - 2\theta_{ij}L). \end{aligned} \quad [4]$$

To eliminate the P_i as nuisance parameters, let the probability of no break between i and the centromere be $e^{-w_{0i}}$ and P_0 be the probability that the centromere is retained. Then

$$\begin{aligned} P_i &= P_0 e^{-w_{0i}} + (1 - e^{-w_{0i}})L \\ &= 1 - 2\theta_{0i} + 2\theta_{0i}L, \end{aligned} \quad [5]$$

where $1 - 2\theta_{0i} = e^{-w_{0i}+h}$ and $h = \ln[(P_0 - L)/(1 - L)]$ for $P_0 > L$, or zero otherwise.

Of several methods for monosomic radiation hybrid mapping (7), all but one assume that $P_i = P_j = P_0 = L$ with a resultant loss of efficiency that increases with the gradient of P_i but depends on the method used. For multiple pairwise mapping of chromosome 21 it has been shown that neglect of the monotonic decline of P_i from P_0 to L inflates χ^2 from 990 to 1508 and, therefore, has a relative efficiency of 66% (6). There is no reason to prefer the simpler but falsified assumption. "Pushmi-pullyu" hybrids selected for presence of one flanker and absence of the other are the special case $P_0 = 1$, $L = 0$, which gives efficient mapping of the interval. Other types of locus content mapping are presumably consistent with $P_i = L$ (4).

Disomic Probabilities

For each pair of loci there are four possibilities to consider: double heterozygotes, double homozygotes, proximal heterozygote, and distal heterozygote.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Case 1. Double Heterozygotes. Denote the two alleles at locus *i* by A, a and the two alleles at locus *j* by B, b. Without loss of generality we assume that the phase is AB/ab and further that allelic recovery is random, so that $P(\mathrm{AB}) = P(\mathrm{ab})$, etc. There are 10 probabilities to consider, each the product of two monosomic probabilities:

$$p_1 = P(\mathrm{AB/ab}) = (P_{ij})^2$$

$$p_2 = P(\mathrm{AB/a\cdot}) = P(\mathrm{ab/A\cdot}) = P_{ij}P_{i\cdot}$$

$$p_3 = P(\mathrm{AB/\cdot b}) = P(\mathrm{ab/\cdot B}) = P_{ij}P_{j}$$

$$p_4 = P(\mathrm{AB/\cdot\cdot}) = P(\mathrm{ab/\cdot\cdot}) = P_{ij}P_{\cdot\cdot}$$

$$p_5 = P(\mathrm{A\cdot/a\cdot}) = P_{i\cdot}^2$$

$$p_6 = P(\mathrm{A\cdot/\cdot b}) = P(\mathrm{a\cdot/\cdot B}) = P_{i\cdot}P_{j}$$

$$p_7 = P(\mathrm{A\cdot/\cdot\cdot}) = P(\mathrm{a\cdot/\cdot\cdot}) = P_{i\cdot}P_{\cdot\cdot}$$

$$p_8 = P(\mathrm{\cdot B/\cdot b}) = P_{\cdot j}^2$$

$$p_9 = P(\mathrm{\cdot B/\cdot\cdot}) = P(\mathrm{\cdot b/\cdot\cdot}) = P_{\cdot j}P_{\cdot\cdot}$$

$$p_{10} = P(\mathrm{\cdot\cdot/\cdot\cdot}) = P_{\cdot\cdot}^2.$$

Case 2. Double Homozygotes. This case represents both genuine homozygotes and failure to score fragment length (e.g., on an agarose gel). Because A = a and B = b, there are only four probabilities that are different from monosomy, although the phenotypes are the same:

$$P(\mathrm{AB}) = p_1 + 2(p_2 + p_3 + p_4 + p_6)$$

$$P(\mathrm{A\cdot}) = p_5 + 2p_7$$

$$P(\mathrm{\cdot B}) = p_8 + 2p_9$$

$$P(\mathrm{\cdot\cdot}) = p_{10}.$$

Case 3. Proximal Heterozygote. Because B = b, there are eight probabilities neglecting dosage:

$$P(\mathrm{AaB}) = p_1 + 2p_2$$

$$P(\mathrm{AB}) = p_3 + p_4 + p_6$$

$$P(\mathrm{Aa\cdot}) = p_5$$

$$P(\mathrm{aB}) = p_3 + p_4 + p_6$$

$$P(\mathrm{A\cdot}) = p_7$$

$$P(\mathrm{a\cdot}) = p_7$$

$$P(\mathrm{\cdot B}) = p_8 + 2p_9$$

$$P(\mathrm{\cdot\cdot}) = p_{10}.$$

Case 4. Distal Heterozygote. Because A = a, this is similar to the preceding case:

$$P(\mathrm{ABb}) = p_1 + 2p_3$$

$$P(\mathrm{AB}) = p_2 + p_4 + p_6$$

$$P(\mathrm{A\cdot}) = p_5 + 2p_7$$

$$P(\mathrm{Ab}) = p_2 + p_4 + p_6$$

$$P(\mathrm{\cdot Bb}) = p_8$$

$$P(\mathrm{\cdot B}) = p_9$$

$$P(\mathrm{\cdot b}) = p_9$$

$$P(\mathrm{\cdot\cdot}) = p_{10}.$$

Discussion

To convert these probabilities into an algorithm for disomic mapping it is necessary to have a convention for allelic representation. Because it is convenient in monosomic mapping to indicate for each clone presence of a locus by 1, absence by 0, and no observation by 8, it is only necessary to modify this convention by designating presence of only the first allele by 1, presence of only the second allele by 2, and presence of both alleles by 3. For double heterozygotes, if the number of clones classified as AB or ab exceeds the number classified as Ab or aB the phase AB/ab is supported, and this evidence is strong when A and B are close. For a specified order the alleles at the proximal locus may be denoted arbitrarily by A = 1, a = 2, while alleles at the distal locus are denoted by B = 1, b = 2 if AB are in coupling and by b = 1, B = 2, otherwise. The calculations lend themselves to multiple pairwise mapping by logarithms of odds (lods) (8), which is especially appropriate in this case because there is no interference among breaks (6) and phase may be reliably inferred for all heterozygous loci.

Locus content mapping is only one source of evidence, which may be weakened by sampling errors, nonuniform breakage, and chimerism. Full use of the evidence requires integration into a composite location that summarizes all the data, genetic and physical (9). When this information is delegated to a subjective consensus at a brief workshop, without reference to a location data base or use of integration algorithms, the resultant map is poorly documented, sparse, needlessly imprecise, and of little value for construction of high-resolution maps, positional cloning, or sequencing. Each advance in gene mapping makes integration more imperative for the many thousands of loci assigned by diverse methods to chromosomes of humans and other well-studied organisms.

1. Morton, N. E. (1991) *Ann. Hum. Genet.* **55,** 235–241.
2. Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990) *Science* **250,** 245–250.
3. Goss, S. J. & Harris, H. (1975) *Nature (London)* **255,** 1445–1458.
4. Dear, P. H. & Cook, P. R. (1993) *Nucleic Acids Res.* **21,** 13–20.
5. Haldane, J. B. S. (1919) *J. Genet.* **8,** 299–309.
6. Lawrence, S., Morton, N. E. & Cox, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7477–7480.
7. MacCluer, J. W., Chakravarti, A., Cox, D. R., Bishop, D. T., Bale, S. J. & Skolnick, M. H. (1992) *Cytogenet. Cell Genet.* **59,** 71–111.
8. Morton, N. E. & Andrews, V. (1989) *Ann. Hum. Genet.* **53,** 263–269.
9. Morton, N. E., Collins, A., Lawrence, S. & Shields, D. C. (1992) *Ann. Hum. Genet.* **56,** 223–232.