研究速報

マウスゲノムとの比較遺伝子地図を用いた末マッピング 遺伝子の染色体の予測

和田 康彦^{*1}・岸野 洋久^{*2}・柏木 宣久^{*3}

^{*1} 農林水産省畜産試験場育種部家畜ゲノム研究チーム

^{*2} 東京大学大学院農学生命科学研究科生物測定学研究室

^{*3} 統計数理研究所調査実験解析研究系

Using Linkage Map of Mouse for Prediction of Gene Location in Animals

Yasuhiko Wada^{*1}, Hirohisa Kishino^{*2}, and Nobuhisa Kashiwagi^{*3}

^{*1} Department of Animal Breeding and Genetics, National Institute of Animal Industry, Ikenodai,
Kukisaki-machi, Inashiki-gun, Ibaraki 305–0901, Japan

^{*2} Laboratory of Biometrics, graduate school of Agricultural and Life Sciences,
University of Tokyo, 1–1–1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

^{*3} The Institute of Statistical Mathematics, 4–6–7, Minamiazabu,
Minato-ku Tokyo 106–8569, Japan

The high-density linkage map of mouse has been constructed. This map contains quantitative trait loci (QTL) and major genes which are not mapped on other animals. In this study, a Bayesian model was developed to predict the chromosome number of unmapped genes in animals using comparative mapping data in contrast with mouse linkage map. The comparative mapping data of mouse-human, mouse-rat and mouse-cattle is collected from the Animal Genome Database. In order to examine the accuracy of our method, we randomly divided the data into two subsets. The ratios of chromosomal translocation of mouse-human, mouse-rat and mouse-cattle were estimated by one of the subsets, and the chromosome numbers of mapped genes of human, rat and cattle were predicted for another subset.

Key words: Bayesian model; Comparative mapping; Genome evolution; Linkage map.

1. はじめに

分子生物学の急速な進歩によってヒトやマウスにおいては膨大な数の DNA マーカーや遺伝子が遺伝子地図上にマッピングされるようになり、それらのデータはデータベース化されてインターネット上で公開されている (Fasman *et al.* (1997), Blake *et al.* (1999)). 家畜、特にブタにおいて

は経済的に価値の高い形質に関与する QTL (Quantitative Trait Locus; 量的形質遺伝子座) を探索するために, DNA マーカーの開発とマーカー連鎖地図の作製, 遺伝子の染色体地図上へのマッピングが行われている (Andersson *et al.* (1994), Rohrer *et al.* (1996), Mikawa *et al.* (1999), Wada *et al.* (1999)).

複数の生物種の遺伝子地図を比較することによって, ゲノムの進化を研究しようという試みも行われている. そのなかで DeBry and Seldin (1996) は相同遺伝子をもとにヒトとマウスの遺伝子地図を比較して, 2 つの動物種の間で染色体上に同じ並び順で遺伝子がマッピングされているシンテニー領域が多数存在していることを報告した. また, 植物においてもイネ科植物 (Devos and Gale (1997)) やアブラナ科植物 (Lagercrantz (1998)) においても多くのシンテニー領域が保存されていることが明らかとなっている. これらの結果は, 高等生物においては染色体の再構成が起きる頻度が小さく, シンテニー領域が多く保存されていることを示唆している.

ところで, 家畜における遺伝子地図の作製はヒトやマウスと比較して遅れている. もし, 多くのシンテニー領域が家畜と他の哺乳動物との間で保存されているのであれば, ヒトやマウスなどの遺伝子地図作製が進んでいる動物種と家畜との間の比較遺伝子地図を作成し, 家畜においてまだマッピングされていない遺伝子の家畜ゲノム上でのおよその位置をある程度の確率で予測することが可能となる. さらに, その結果と QTL 解析の結果を比較することによって, QTL の原因遺伝子の解明にも道を開くことが期待される.

そこで, 著者らは家畜の遺伝子地図情報をデータベース化した家畜ゲノムデータベースを開発, 公開するとともに, 家畜との比較遺伝子地図を簡単に利用するためのシステムを開発してきた (Wada and Yasue (1996), Wada *et al.* (1997), 和田・安江 (1997), <http://ws4.niai.affrc.go.jp/>). これらのシステムは一見して理解できるようにおもにコンピュータビジュアライゼーションの手法を利用しているので, コンピュータや統計手法に不慣れな分子生物学者にも利用しやすいという利点がある. しかし, 遺伝子のマッピングには多くの労力と予算が必要になるために, 研究計画を詳細につめていく段階ではその計画にどの程度の可能性やリスクがあるのかを知ることが重要になる. その場合には, ねらいとする遺伝子が家畜のどの染色体にどの程度の確率で存在するのかを直感ではなく, 数値として予測できるとリスク管理をすることができて, 研究管理上でのメリットが多い.

Sankoff *et al.* (1992) は複数の種における染色体領域の逆位や転座, 挿入や欠失の情報から染色体領域の再構成距離を求める方法を示した. 彼らはその方法を用いて 16 種のミトコンドリア領域における塩基配列のデータから系統樹を作成している. その後, 逆転, 挿入, 欠失などの染色体におこった操作を推定して, その情報から種間の分子系統樹を再現する試みが行われている (Sankoff and Blanchette (1998)). しかしながら, これらの手法は種が増加するにつれて, また, ゲノムサイズが大きくなるにつれて, 計算の困難さが飛躍的に増大するという欠点がある. さらに, これらの手法はゲノムの進化を再構成するという観点からはきわめて興味深いものであるが, ねらいとする遺伝子がマウス以外の当該動物種においてどの染色体のどの位置にどの程度の確率で存在するのかを統計学的に予測することはできない.

そこで, まず第 1 段階として比較遺伝子地図のデータから, ねらいとする遺伝子が他の動物種においてどの染色体にどの程度の確率で存在するのかを予測する方法について検討した. データとしてはマウスの MGD (Mouse Genome Database) の連鎖地図とヒト, ラットおよびウシの染色

体物理地図などのデータを利用して、マウスのゲノムと他の動物種のゲノムとの間での染色体転座確率をデータから推定し、マウスの MGD 連鎖地図上にはマッピングされているが、他の動物種ではマッピングされていない遺伝子座について、それが存在するであろう染色体を予測するベイズモデルを開発した。そして、実際のデータをランダムに 2 分割して、片方のデータを用いて染色体転座確率を推定し、もう片方のデータを用いてマウス以外の動物種でもマッピングされているすべての遺伝子について各染色体について予測確率を算出し、存在しているであろう染色体を予測し、その有用性を検証した。

2. ベイズモデルを用いた予測法

簡単のために常染色体上の遺伝子のみに限定して考える。マウス MGD 連鎖地図の同一染色体上で距離が d モルガンだけ離れた 2 つの部位が染色体転座で切断される確率がその間の組み換え率と比例するというモデルを考えてみた。組み換え率としては Kosambi (1944) のもの

$$\left(\theta(d) = \frac{1}{2} \frac{\exp(4d) - 1}{\exp(4d) + 1} \right) \text{を用い, 染色体転座は}$$

$$\eta(d) = \alpha \theta(d) \quad (1)$$

の確率で起こるとする。

ある動物種でまだマッピングされていない遺伝子の存在する染色体を予測することを考える。この動物種の半数染色体の数を M とすると、マウスゲノムで距離が d モルガンだけ離れた 2 つの遺伝子が当該動物種でも同一の染色体に存在する確率は、各染色体に対する事前確率は等確率とすると、

$$p(d) = (1 - \eta(d)) + \frac{\eta(d)}{M} \quad (2)$$

となる。第 2 項は、染色体転座の結果として同一の染色体上に移ったことを意味する。染色体転座を起こし、別の特定の染色体の上に存在する確率は

$$q(d) = \frac{\eta(d)}{M} \quad (3)$$

となる。これを推移確率のように、表現することにする。すなわち、マウスゲノム上のある遺伝子 A の当該動物種における相同遺伝子 A' が第 i 染色体に存在しているという条件の下で、A と MGD 連鎖地図上で d モルガンだけ離れた遺伝子 B の相同遺伝子 B' が第 j 染色体に存在している条件付確率を

$$p_{ij}(d) = \begin{cases} p(d) & (i=j \text{ のとき}) \\ q(d) & (i \neq j \text{ のとき}) \end{cases} \quad (4)$$

とする。

図 1 に示すように、マウスのある染色体上の目的の遺伝子 C から最近接の両側の 2 つの遺伝

図 1. Illustration of the procedure of prediction.

子 A および B までの距離をそれぞれ d_1, d_2 モルガンとし, それら 2 つの遺伝子の相同遺伝子 A', B' が当該動物種ではそれぞれ第 j_1, j_2 染色体に存在しているとする, 目的の遺伝子 C の相同遺伝子 C' が当該動物種の第 i 染色体に存在している予測確率は

$$P(i|d_1, d_2, j_1, j_2) = \frac{p_{j_1, i}(d_1) p_{i, j_2}(d_2)}{\sum_{k=1}^M p_{j_1, k}(d_1) p_{k, j_2}(d_2)} = \frac{p_{j_1, i}(d_1) p_{i, j_2}(d_2)}{p_{j_1, j_2}(d_1 + d_2)} \quad (5)$$

となる.

この手法を用いて, 目的の遺伝子が当該動物種の第 i 染色体に存在する時の予測確率を計算するには, まず α を各動物種ごとに推定する必要がある. この α を推定するには, マウスの MGD 連鎖地図上にマッピングされている遺伝子で, 他の種においても存在する染色体が特定されている遺伝子をゲノムデータベースから抜き出す. そして, 相手の動物種で同じ染色体に 2 つの遺伝子がともに存在しているかどうかをチェックする. すべての遺伝子の組について 1 cM 間隔で当該動物種で同一染色体上に存在しているかどうかをカウントして, その割合を計算する. その値を染色体転座確率 $\eta(d)$ として α を推定することができる.

α を推定できれば, MGD 連鎖地図上で目的の遺伝子の両側で当該動物種でもマッピングされている遺伝子を探し, その 2 つの遺伝子と目的の遺伝子との MGD 連鎖地図上での距離と当該動物種での染色体番号がわかれば, (5)式から各染色体の予測確率を算出できる. この予測確率が一番高く予測された染色体に目的の遺伝子が存在すると推定できる.

3. パラメータ α の推定

上記の手法がどの程度の予測精度を持っているかを調べるために, 各遺伝子と目的の遺伝子の MGD 連鎖地図での位置を中心とした一定の範囲内の遺伝子との組を 2 つに分割し, 一方の

表 1. Estimate and standard error of α .

Animal	N. of genes	Estimate of α	Standard error
Human	22,212	0.9815	0.03919
Rat	310	0.8285	0.2299
Cattle	185	1.3140	0.1499

表 2. Results of prediction of chromosome number.

Animal		Probability of posterior distribution											Total
		0.0-	0.1-	0.2-	0.3-	0.4-	0.5-	0.6-	0.7-	0.8-	0.9-	1.0	
human	N. of correct	0	0	0	1	11	10	17	17	12	15	599	682
	N. of error	87	1	0	2	7	8	8	11	9	2	48	183
	Ratio of correct prediction	0.0	0.0	0.0	0.33	0.61	0.56	0.68	0.61	0.57	0.88	0.93	0.79
rat	N. of correct	0	0	0	0	0	2	0	0	3	15	45	65
	N. of error	6	1	0	0	1	5	3	1	0	3	1	21
	Ratio of correct prediction	0.0	0.0	0.0	0.0	0.0	0.29	0.0	0.0	1.0	0.83	0.98	0.76
cattle	N. of correct	0	0	0	1	2	3	2	2	3	7	27	47
	N. of error	0	1	0	0	4	5	7	2	4	1	3	27
	Ratio of correct prediction	0.0	0.0	0.0	1.0	0.33	0.38	0.22	0.50	0.43	0.88	0.90	0.64

データで α を推定し、もう一方のデータと推定された α を用いて予測された染色体が正しい染色体であったかどうかを検討した。

動物種としては、ヒト、ラット、ウシの3種を用いた。マッピングデータは家畜ゲノムデータベース (Wada and Yasue (1996)) から抜きだし、マウスおよびヒト、ラット、ウシにおいてそれぞれ常染色体上にマッピングされている遺伝子のみを抜きだした。

このデータをランダムに2等分し、片方のデータを用いて(1)式より α を推定した。推定にはマルカート法を用い、計算には農林水産研究計算センターのアプリケーションサーバー上にあるSASを使用した。

3種の動物における α の推定値と標準誤差と推定に用いた遺伝子の数を表1に示した。 α の推定値はラット、ヒト、ウシの順に大きくなった。これはラットが他の2種よりもマウスと遺伝的に近いことと一致している。しかし、ラットとウシにおいてはデータが少ないためヒトよ

りも標準誤差が大きくなっていた。

4. 染色体の予測

α の推定時に用いなかった方の比較遺伝子地図データと前項で推定した α の値から(5)式を用いて各遺伝子について染色体ごとの予測確率を算出した。予測確率最大の染色体と実際にマッピングされている染色体が同じ時を正解, 異なる時を不正解として, 予測確率を 0.1 間隔ごとに増やしていき, 正解数と不正解数をカウントした。これらをヒト, ラットおよびウシについて計算した結果を表 2 に示した。

トータルでの正解率はヒトで 0.88, ラットで 0.76, ウシで 0.64 であった。ヒトでは利用可能な正解率であったが, ラット, ウシの順に正解率が低くなった。これはマッピングされている遺伝子の数が少ないためであろうと考えられた。ただ, ラットでは予測確率 0.8 以上で正解率 0.83 以上となっており, ウシでは予測確率 0.9 以上で正解率 0.88 以上となっていた。このことから, マッピングの進んでいない動物種でも予測確率の高いものであれば十分に使用できるものと考えられる。

5. 討論

本研究は, 家畜においてマッピングされていない遺伝子について, 染色体でも予測できないか, という観点で始められたものである。家畜においては染色体上にマッピングされている遺伝子はまだ少なく, 必ずしも現時点では実用的な手法とは言えない面もある。しかし, ブタ (Yerle *et al.* (1998)) やウシ (Yang *et al.* (1998)) においては RH (radiation hybrid) マップが開発されており, 数年後には RH マップ上にマッピングされている遺伝子が飛躍的に増加すると考えられている。その時には, 本研究で述べたような手法も一定の役割を果たすことができるであろう。

マウスにおいては種々の系統が作出され, いろいろな形質が連鎖地図上に遺伝子座としてマッピングされている。これらの多くについては原因遺伝子が未知であるが, 本手法を用いてそれらの遺伝子座のヒトでの染色体を推定することによって, ヒトにおける分子生物学上の膨大な知識とマウスにおける連鎖地図作成の努力とを連係することができるのではないかと考えられる。

本手法の応用分野への利用を考えるとときには, 染色体を推定するだけではなくて, 染色体上のより狭い範囲を推定できる方が価値が高い。その場合には, シンテニーの詳細情報を利用して, 比較ゲノムデータから相同な領域を量的に予測することになる。今回の手法はマッピングデータのみを利用したものであった。しかし, マッピングされている遺伝子のアミノ酸配列や塩基配列は入手可能であり, それらのデータを含めることによって, もっと精度の高い予測が可能になるかもしれない。

分子生物学の進展によって, すべてのゲノム領域でランダムに乗換えが起こっているわけではなく, ホットスポットと呼ばれる乗換えが起こりやすい領域や, 反対に乗換えがほとんど生じない領域があることがわかってきた (Lindahl (1991))。ゲノム進化においてもすべてのゲノム領域において等確率で染色体転座するとは限らないことが考えられ, 今後, この点についても検討する必要がある。

謝辞

本研究は統計数理研究所共同研究 (10-共研 A-53) にもとづくものである。本研究の遂行にあたって格段のご配慮をいただいた農林水産省畜産試験場育種部の安江 博上席研究官に謝意を表する。

参考文献

- Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. and Lundstrom, K. (1994). Genetic Mapping of Quantitative Trait Loci for Growth and Fatness in Pigs. *Science*, **263**, 1771-1774.
- Blake, J. A., Richardson, J. E., Davisson, M. T., Eppig, J. T. and The Mouse Genome Database Group (1999). The Mouse Genome Database (MGD): Genetic and genomic information about the laboratory mouse. *Nucleic Acids Res.*, **27**, 95-98.
- DeBry, R. W. and Seldin, M. F. (1996). Human/mouse homology relationships. *Genomics*, **33**, 337-351.
- Devos, K. M. and Gale, M. D. (1997). Comparative genetics in the grasses. *Plant Mol. Biol.*, **35**, 3-15.
- Fasman, K. H., Letovsky, S. I., Cottingham, R. W. and Kingsbury, D. T. (1997). The GDB(tm) Human Genome Database Anno 1997. *Nucleic Acids Res.*, **25**, 72-80.
- Hannenhalli, S., Chappey, C., Koonin, E. V. and Pevzner, P. A. (1995). Genome Sequence Comparison and Scenarios for Gene Rearrangements: A Test Case. *Genomics*, **30**, 299-311.
- Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Ann. Eugen.*, **12**, 172-175.
- Lagercrantz, Ulf. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**, 1217-1228.
- Lindahl, K. F. (1991). His and hers recombinational hotspots. *Trends Genet.*, **7**, 273-276.
- Mikawa, S., Akita, T., Hisamatsu, N., Inage, Y., Ito, Y., Kobayashi, E., Kusumoto, H., Matsumoto, T., Mikami, H., Minezawa, M., Miyake, M., Shimanuki, S., Sugiyama, C., Uchida, Y., Wada, Y., Yanai, S. and Yasue, H. (1999). A linkage map of 243 markers in an intercross of Gottingen miniature and Meishan pigs. *Animal Genetics*, **30**, 407-417.
- Rohrer, G. A., Alexander, L. J., Hu, Z., Smith, T. P. L., Keele, J. W. and Beattie, C. W. (1996). A comprehensive map of the porcine genome. *Genome Research*, **6**, 371-391.
- Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B.F. and Cedergren, R. (1992). Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA*, **89**, 6575-6579.
- Sankoff, D. and Blanchette, M. (1998). Multiple genome rearrangement and breakpoint phylogeny. *J. Comp. Biol.*, **5**, 555-570.
- Wada, Y. and Yasue, H. (1996). Development of an animal genome database and its search system. *Comp. Applic. Biosci.*, **12**, 231-235.
- Wada, Y., Inoue, K., Ohga, K. and Yasue, H.: Software tool for gene mapping: granch. (1997). *Comp. Applic. Biosci.*, **13**, 323-324.
- 和田康彦, 安江 博 (1997). Java Applet を用いた遺伝子地図情報に関するビジュアライゼーションの試み. コンピュータサイエンス, **4**, 111-115.
- Wada, Y., Akita, T., Furukawa, T., Hisamatsu, N., Inage, Y., Ishii, K., Ito, Y., Kobayashi, E., Kusumoto, H., Matsumoto, T., Mikawa, S., Minezawa, M., Miyake, M., Shimanuki, S., Sugiyama, T., Uchida, Y., Yanai, S. and Yasue, H. (2000). Quantitative trait loci (QTL) analysis in A Meishan X Gottingen cross population. *Submitted to Animal Genetics*.

- Yang, Y. P., Rexroad, C. E. 3rd., Schlapfer, J. and Womack, J. E. (1998). An integrated radiation hybrid map of bovine chromosome 19 and ordered comparative mapping with human chromosome 17. *Genomics*, **48**, 93–99.
- Yerle, M., Pinton, P., Robic, A., Alfonso, A., Palvadeau, Y., Delcros, C., Hawken, R. J., Alexander, L., Beattie, C., Schook, L., Milan, D. and Gellin, J. (1998). Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet. Cell Genet.*, **82**, 182–188.