

ブロック分割を用いた多遺伝子座のハプロタイプ推定に関する一提案
 A Proposal of Haplotype Estimation for Many SNPs Inputs Using Block Division

下里二郎 甲藤二郎
 Jiro Shimosato Jiro Katto

早稲田大学 大学院 理工学研究科
 Graduate School of Science and Engineering, Waseda University

1. はじめに

約30億個のヒトゲノム塩基配列の解読が急ピッチで進む中、塩基配列に疾患などの様々な形質を対応付ける形質マッピングとよばれる作業に関心が集まっている[1]

形質マッピングでは、相同染色体のそれぞれに遺伝子がどのように並んでいるかを示すハプロタイプを推定する必要がある。本稿では100遺伝子座に対応できると考えられる新たなハプロタイプ推定アルゴリズムの提案を行う。

2. 遺伝子型とハプロタイプ推定

相同染色体の同じ遺伝子座にある2つの遺伝子の組み合わせである遺伝子型データは、ハプロタイプに比べ容易に得られるため、遺伝子型データからハプロタイプを推定する手法が近年研究されている。これには個人遺伝子型データを用いてEM手法に基づく最尤推定手法[2][3]や、筆者らが提案したマルコフモデルの最適経路による手法[4]などが有効だが、前者は20遺伝子座を超える多遺伝子座に適用した場合、候補となるハプロタイプ数が膨大になり、許容できる時間内にハプロタイプを推定することは困難であり、また後者は多遺伝子座のハプロタイプ推定ではその推定結果に誤差が生じてしまう。そこで、遺伝子型データをハプロタイプブロックと呼ばれるブロックにわけ、そのブロック内でハプロタイプを[4]の手法で推定する手法を提案する。

3. ハプロタイプブロック分割

長い塩基配列では、アレルは染色体上で連鎖と言う遺伝子座同士の結びつきの強いいくつかの塊に分かれる。この塊をハプロタイプブロック[5]という。任意の2遺伝子座間の連鎖の尺度となる値として次式で導かれる D' 、 ρ^2 がある。

$$D = p_{ac} - p_a p_c$$

$$\rho^2 = D^2 / (p_a p_b p_c p_d)$$

$$D' = D / D_{max} \quad (D > 0 \text{ の場合})$$

$$D' = D / D_{min} \quad (D < 0 \text{ の場合})$$

$$\text{ただし } \begin{cases} D_{max} = \max(p_a p_d, p_b p_c) \\ D_{min} = \min(-p_a p_c, p_b p_d) \end{cases}$$

ここで、1遺伝子座アレルを a,b、2遺伝子座アレルを c,d とし、それぞれのアレル頻度を p_a, p_b, p_c, p_d 、2遺伝子座におけるそれぞれのハプロタイプ頻度を $p_{ac}, p_{ad}, p_{bc}, p_{bd}$ とした。

4. 提案手法の概要

本手法は大きく分けて2つの処理からなる。第1の処理は同一集団の遺伝子型データから全ての2遺伝子座間の D' 、あるいは ρ^2 を求め、それをもとに遺伝子座をハプロタイプブロックに分割するというものである。第2の処理は、分割された全てのハプロタイプブロック内において各遺伝子座をノードと見立てた図1のようなマルコフモデルを自動構築し、[4]の手法を用いて個人が最も取り得るハプロタイプを推定し、それらを全て繋ぐプロセスである。

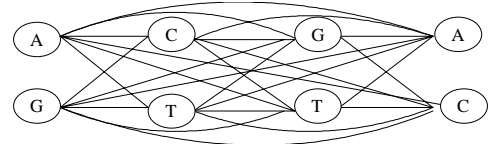


図1 遺伝子座のマルコフモデル

5. 実験結果

実験には[6]の103遺伝子座のデータを用いた。

(1)ハプロタイプブロック分割

全ての2遺伝子座間の D' 、あるいは ρ^2 をもとめ、連鎖の強いと思われる0.7以上の値の部分を図2に示した。対角線上にいくつかのブロックが確認できる。

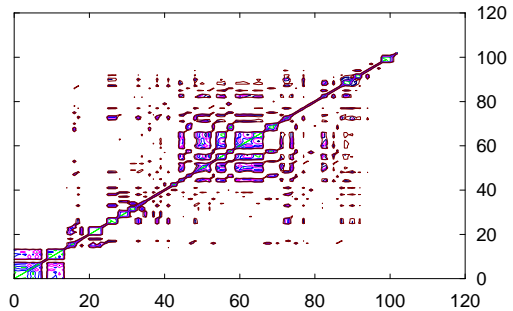


図2 2遺伝子座間の D'

(2)ハプロタイプ推定

分割されたブロック内において[4]の手法で推定された各ハプロタイプの頻度を表1に示す。また同じデータを用いて実験的に得られたハプロタイプ頻度の結果[5]と比較したところほぼ等しい結果が得られた。

表1 ハプロタイプブロック内のハプロタイプ頻度

block	0	1	2	3	4	5	6	7
locus ID	0-7	8-12	13-23	24-43	44-75	76-81	82-94	95-102
Haplotype frequency [%]	76 18	79 19	36 25	38 12	40 12	29 14	44 9	32 17
			28	17	29	17	35	50
				25	10	36	4	

6. おわりに

本稿ではハプロタイプブロックによる分割とマルコフモデルにおける辺重み最大化により多遺伝子座の遺伝子型データに対応可能なハプロタイプ推定手法の提案を行った。

参考文献

[1]鎌谷直之(編): "ポストゲノム時代の遺伝統計学"羊土社,2001
 [2]Excoffier,L.& Slatkin,M.:Mol.Biol.,12: 921-927,1995
 [3]Fallin,D.& Schork,N.: Am.J.Hum.Genet.,67 :947-959,2000
 [4]下里,甲藤.: FIT,1:29-30,2002
 [5]Mark,J. et al. : Nat.Genet.,29:229-232,2001
 [6] "IBD5 data" <http://www-genome.wi.mit.edu/humgen/IBD5/>