

遺伝マーカーを用いた樹木集団内遺伝構造の 空間解析手法

高橋 誠¹ ・ 島谷 健一郎²

(受付 2002 年 7 月 30 日 ; 改訂 2002 年 10 月 29 日)

要 旨

分子遺伝学の急速な発展に伴い、樹木集団内における個体ベースでの遺伝変異の空間構造研究が世界各地で活発に行われている。遺伝変異データは、対立遺伝子レベル、遺伝子型レベルおよび multilocus レベルという 3 層の階層構造から成っている。さらに、空間的遺伝構造の解析では、ここに各個体の位置情報が付与される。このような高度な情報量を有するデータを解析するための統計的手法は多岐に渡り、それぞれに固有の特性を有している。本稿ではこれまでの空間的遺伝構造の解析に用いられてきた統計量を 3 層のカテゴリーに区分して、それらの特性や問題点、並びに生物学的背景や統計量間の関係について論じる。

キーワード：生態学、遺伝マーカー、樹木集団、集団遺伝学、空間統計、対立遺伝子、遺伝子型。

1. はじめに

遺伝マーカーを用いた樹木集団に関する研究は 1970 年代から盛んに行われて来ているが、当初は互いに数 km ~ 数 100 km 離れた複数の天然林集団から調査対象樹種の個体をランダムに抽出し、その遺伝変異を地理的なスケールで解析するという地理的変異に関する研究が多かった。ところが 1980 年代後半から、天然林内に設けた調査区内の全個体の位置を測定し、それら全ての遺伝変異を調べる(数 m ~ 数 10 m という、より細かいスケールでの)研究が行われ始めた。その後の分子遺伝学や DNA 分析技術の急速な発展により、遺伝変異のデータをより迅速に、より廉価に、より単純な作業で大量に集められるようになり、1990 年代には世界各地でこのような個体ベースの研究が精力的に行われるに至った。それにより、これまでの研究手法では知り得なかった森林や樹木集団についての事実が見出されたり、推定することが不可能(困難)とされていた事象について、その推定値(あるいは実測値)が得られる可能性が開けてきた。さらには、天然林資源管理のための新たな知見を得るべく、天然林研究は新局面を迎えつつ今世紀に入った。

めざましい発展を続ける分子遺伝学とは裏腹に、労力と経費をかけて得られた全個体についての位置情報付き遺伝変異データを解析するための統計的解析手法の方は、Sokal and Oden (1978a, 1978b)による空間統計の導入以来、さほどの進展を見せていない。それらの手法を用

¹(独) 林木育種センター 育種部育種課環境育種研究室：〒319-1301 茨城県多賀郡十王町伊師 3809-1;
makotot@affrc.go.jp

² 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7; shimatan@ism.ac.jp

いた多くの研究により新しい知見が得られている一方で、それらの問題点も指摘されている (Epperson (1995c), Loiselle et al. (1995), Shimatani and Takahashi (2003), Smouse and Peakall (1999)). 位置情報付き遺伝変異データが内包する情報量の膨大さからすると、現在の解析手法はそのごく一部を活用しているに過ぎず、貴重な情報を含んでいる(と期待される)データを精査していない感がある。そこで本稿では、今まで提唱された空間的遺伝変異データに対する解析手法の総括を試み、今世紀の研究の出発点を固めてみたい。

まず、現在空間的遺伝構造の研究に用いられている遺伝マーカーとは何かについて簡単に説明し、それをを用いた研究の意義と取り扱われるデータの構造について述べる。その後、これまで用いられてきた統計量を順次紹介し、それらの違いや特徴、生物学的背景や問題点について論じる。

1.1 遺伝マーカー

ゲノムプロジェクトに代表されるように、今日、遺伝子は世界で最も盛んに研究されている科学領域の一つである。生物の遺伝情報は DNA という長い鎖状の核酸と呼ばれる高分子に塩基配列という形で保持されており、DNA の塩基配列のどの部位がどのような機能を司っているかを解明することで、医療や薬学、あるいは農林業などの分野で応用されることが期待されている。その一方で、個々の遺伝子の機能を解明するのではなく、様々な分析手法によって検出される DNA 断片の塩基配列の差異を、個体間の遺伝的な違いを表す手がかりや遺伝的な意味での単なる記号として用いることがある。そのような目的で遺伝変異を調べる時、その個々の変異は遺伝マーカー (genetic marker; あるいは分子マーカー (molecular marker)) と呼ばれる。

本稿で扱う集団内遺伝構造の解析に用いられる遺伝マーカーは、通常、「中立マーカー (neutral marker)」と考えられている。ここで言う「中立」とは、自然淘汰に対して中立で個体の適応度の増減に影響を与えないということである。遺伝マーカーのこの性質は、例えば種や集団の遺伝的多様性を評価するのに適している。即ち、集団内の大多数の個体が少数の母樹に由来する兄弟であったり、血縁関係がある個体間での近親交配が続いた結果である場合、その集団においては個体間での遺伝的な近縁度が高まり、遺伝変異幅 (遺伝的多様性) が減少する。そのため、乾燥や病虫害、あるいは雪害などの特定の環境負荷に対して集団が脆弱になったり、近交弱勢により集団を構成する各個体の適応度が減少するといったことが起き、集団が消失する可能性が高まると考えられる。このため、集団遺伝学 (population genetics) や保全遺伝学 (conservation genetics) では、集団内の遺伝的多様性を推定することが重要な課題となっているが、中立的な遺伝マーカーを調べてそれらがどの程度の多様性を有しているかにより、集団内での近交の程度や各集団のもつ遺伝的多様性の比較や集団間での遺伝的分化の程度などを評価することができる (c.f. 根井 (1990))。この時、逆に適応的な遺伝子ではそういった推定はできない。なぜなら、適応的な遺伝子では、元来の遺伝変異の一部分が自然淘汰により失われている可能性が大きいからである。

1.2 植物の特性と空間的遺伝構造

1.2.1 固着性と更新動態

樹木など植物の空間的遺伝構造を特徴づける動物にはない性質として、植物の固着性を挙げることができる。すなわち、植物はある場所で発芽すると、通常その個体が枯死するまでその場所から移動することはない。これを集団内における静的な状態とすれば、集団内での動的な機会は、天然更新による世代交代 (更新動態とも呼ばれる) であろう。天然更新は、開花、花粉流動 (pollen flow) と交配 (mating)、結実、種子散布 (seed dispersal)、発芽といった複数のプロセスから成るが、この過程で (遺伝子を保有する植物自身は動かないが) 遺伝子は、花粉流動と

種子散布の2因子により世代から世代で移動する。この事を遺伝子流動 (gene flow) という。更新動態は、個体の空間的な再配置のプロセスであると同時に、遺伝変異の空間的再配置のプロセスでもある。

1.2.2 花粉流動、種子散布と空間的遺伝構造

種子散布の範囲は、しばしば一定範囲内に限定される。例えば、ブナ (*Fagus* 属) やナラ (*Quercus* 属) などの場合、種子は枝から重力で落下し、その一部はリスやネズミ、時にはカケスなどの動物や鳥によって運搬されるが、それでも種子の大半は林冠の縁からおよそ 20 m 以内に散布される (前田 (1988), 柳谷 他 (1969))。また、花粉流動は、雄蕊で生産された花粉が風あるいは虫や鳥の体に付着して送粉されることで、その範囲は一般に種子の場合よりも広範囲であるが、それでも一定の範囲内であると考えられている (数 10 ~ 数 100 m; Dow and Ashley (1996), Troggio et al. (1996), Isagi et al. (2000))。このように花粉流動や種子散布が空間的に限定されるため、ある世代での空間的遺伝構造は、その前の世代の遺伝構造に依存的であると考えられる。このため、花粉流動や種子散布の程度に応じて、世代を重ねるごとに徐々に空間的遺伝構造が形成されて行くと考えられる。Wright (1943) はこれを「距離による隔離 (isolation by distance)」と呼んだ。この際、花粉の飛散距離や種子の散布距離といったパラメーターが異なると、集団内に形成される遺伝構造は異なるだろうし、この他にも、交配様式やクローン繁殖の有無、近交弱勢 (inbreeding depression), 自然淘汰 (selection), 交配個体間の近縁度が高い (集団の有効なサイズが小さい) 事に起因する遺伝的浮動 (random genetic drift) の影響も遺伝構造に関与していると思われる。これらをモデル化した研究は、当初は集団を単位とし地理的スケールにおける isolation by distance を考察するものが大半であったが (例えば Sokal and Wartenberg (1983), Sokal et al. (1989), Slatkin and Arter (1991), Sokal and Jacquez (1991), Sokal and Oden (1991)), 1990 年代から、個体ベースによるより小さい空間スケールでの研究も主としてシミュレーションによって行われるようになり (Epperson (1990, 1995a, 1995b, 1995c), Berg and Hamrick (1995), Doligez et al. (1998), Hardy and Vekemans (1999)), 現実の天然林集団についての研究と並行して活発に進められている。

2. 空間的遺伝構造で取り扱うデータ

2.1 調査方法 — 野外作業と遺伝子分析実験

天然林は様々な樹種で構成されるが、生態学では複数種から成る個体の集合を群集 (community), その構成種中の任意の1種を固定し、その種の個体だけの集合を個体群 (population) と言うが、この population を集団遺伝学では「集団」と訳している。そして、空間的遺伝構造の解析では、ある1樹種について調査地を設けて分析・解析するのが一般的である。

集団内の個体密度によっても異なるが、林冠木が対象の場合、調査地は 100 m × 100 m (面積 1 ha) 程度以上とすることが多い。熱帯林のように樹種別の個体密度が低い場合には数 ~ 数 10 ha の調査地が必要な場合もある。実際、Loiselle et al. (1995) がコスタリカ北東部の *Psychotria officinalis* の集団を解析した場合には 2.25 ha の調査地を設定している。また、Konuma et al. (2000) に至っては、マレー半島の *Neobalanocarpus heimii* の集団を解析した際、実に 42 ha の調査地を用いている。

調査地設定後、その中に生育している全ての個体をナンバーテープなどにより識別するが、林冠木に限る場合、通常胸高直径 (DBH; 地上高 1.2 m の樹幹の直径) で、 $DBH \geq 5$ cm や $DBH \geq 10$ cm の全個体を調査するといった方法が用いられる。研究者によっては、樹高によって個体サイズを規定し、樹高 3 m 以上の個体を調査するといった例もある。なお、研究目的によっては、樹高が 30 cm 以下の小さい個体のみや当年生実生 (その年に発芽した個体) あるいは灌木や下層植

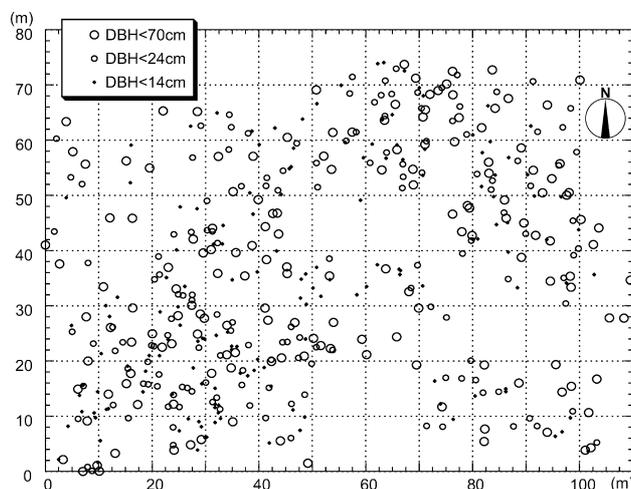


図1. 秋田県田沢湖町ブナ天然林内のブナの個体配置図(Takahashi et al. (2000)より改変).

生を調査対象とすることもある。

調査地の設定にあたって、まず各個体の位置を測量し、それらの立木位置図を作成する(図1)。全調査個体から葉や冬芽などの分析用試料を採取し、それらを実験室に持ち帰り、いくつかの遺伝マーカーについて分析し、全調査個体の遺伝子型データを得る。結果として、得られるデータは例えば表1のようにまとめられる。ここでは図1の集団に対して、*Mdh-3*, *Dia-1*, *Dia-2*, *Got*, *Amy-3*, *Aap-1* および *Pgi-1* と名付けられている7つのアイソザイム遺伝子座(これらは酵素に関わる遺伝子座である)の遺伝子型を分析している。よく知られているように、DNAは4種の塩基(アデニン(A), シトシン(C), グアニン(G), チミン(T))の配列によって規定されている。3つの塩基が1組でコドンを構成し、各コドンにより特定のアミノ酸が指定され、遺伝子はコドンの配列として捉えることもできる。突然変異により塩基配列が変更されると、アミノ酸配列やそれらで構成されるタンパク質電荷や分子量に変化が生じる。このため、アイソザイム分析では、試料をゲル中で電気泳動させ、塩基配列の差異に起因するアミノ酸配列の違いを、酵素を染色した際のバンドの移動度の違いとして検出する。

本稿では、野外での調査法や分析手法の詳細は省略する。上記のアイソザイム分析については津村(2001)を、他の遺伝マーカー、例えばマイクロサテライトDNA分析の仕組みや実験方法については、井鷲(2001)を参照されたい。

2.2 遺伝変異データとその意味

DNA鎖のある部位が特定の機能を司る遺伝子の生成に携わっている場合、あるいはDNA鎖のある断片が遺伝マーカーとして利用できる場合、その部位を遺伝子座(locus)という。ひとつの遺伝子座に塩基配列が異なる遺伝子が複数存在する時、それら是对立遺伝子(allele)と呼ばれる。植物などの真核生物は通常2倍体(diploid)である。すなわち、各遺伝子座は母性由来と父性由来の各1個の対立遺伝子から成り、1対の対立遺伝子により遺伝子型(genotype)が構成される。例えば、花弁の色を支配する遺伝子座において、対立遺伝子*a*は赤い花弁、*b*は白い花弁を形成するとした場合、この遺伝子座の遺伝子型が同一対立遺伝子から成るホモ接合体(homozygote) a/a の場合、実際の花の色(表現型 phenotype)と呼ばれる)は赤となり、 b/b の場

表 1. 空間的遺伝構造の解析に用いられるデータ例. 図 1 の集団に対する位置情報付き遺伝子型データの一部である. 個体の相対的な位置関係を表す座標はコンパス測量などによって求める. ここでは, x - y - z 座標を算出した例を示したが, 地形が均一な場合には x - y 座標で表せば十分であろう. 個体サイズを表す尺度として胸高直径(地上高 1.2 m の樹幹直径)を示している. 遺伝子型情報として, ここではアイソザイム遺伝子型を分析している.

No.	個体名	個体の座標			胸高直径 (cm)	遺伝子型データ						
		x	y	z		<i>Mdh-3</i>	<i>Dia-1</i>	<i>Dia-2</i>	<i>Got</i>	<i>Amy-3</i>	<i>Aap-1</i>	<i>Pgi-1</i>
1	黄B361	53.90	77.47	5.11	96	b/b	a/b	b/b	c/c	e/e	b/c	c/c
2	黄B362	48.96	108.94	8.52	38	b/c	a/b	b/b	c/c	b/e	b/b	c/c
3	黄B363	51.96	110.58	8.62	64	b/b	b/b	b/b	c/c	e/e	b/b	c/c
4	黄B364	54.64	115.66	9.17	87	b/c	b/b	b/b	c/c	e/e	b/b	c/c
5	黄B365	66.18	108.78	8.77	66	b/c	a/b	b/b	c/c	b/e	b/c	c/d
6	黄B366	50.19	125.96	10.08	62	b/b	b/b	b/b	c/c	e/e	a/b	c/c
7	黄B367	47.16	131.18	10.52	21	b/c	b/b	b/b	c/c	b/e	b/b	c/c
8	黄B368	44.83	126.39	9.98	42.6	b/b	b/b	b/b	c/c	e/e	b/c	c/c
9	黄B369	43.27	127.44	10.09	46	b/b	b/b	b/b	c/c	e/e	b/c	c/c
10	黄B370	42.40	132.83	10.77	58.6	b/b	a/a	b/b	c/c	b/e	b/b	c/c
11	黄B371	44.30	147.24	12.20	56.6	b/b	b/b	b/b	c/c	b/e	b/b	c/c
12	黄B372	49.65	145.29	11.96	56.2	b/b	b/b	b/b	c/c	b/b	b/b	c/c
13	黄B373	30.42	125.86	9.86	34.4	b/b	b/b	b/b	c/c	b/e	b/b	c/d
14	黄B374	29.85	126.58	9.93	36.9	b/b	b/b	b/b	c/c	b/e	b/b	c/c
15	黄B375	52.59	150.40	12.29	66	b/b	b/b	b/b	c/c	b/e	b/b	c/c
16	黄B376	55.73	159.89	13.32	56.2	b/b	b/b	b/b	c/e	b/b	b/b	c/c
17	黄B378	63.33	170.05	13.96	42.6	b/b	b/b	b/b	c/c	b/e	b/c	c/c
18	黄B379	57.66	176.80	14.68	54.6	b/c	b/b	d/d	c/c	b/e	b/b	c/c
19	黄B380	53.29	172.97	14.45	49	b/c	a/b	b/b	c/e	b/b	b/b	c/c
20	黄B381	47.02	177.96	15.06	38.4	b/c	b/b	b/b	c/c	b/b	b/b	c/c
21	黄B382	43.79	178.89	15.06	52	b/b	b/b	b/b	c/e	e/e	b/b	c/c
22	黄B383	47.28	159.75	13.01	46.6	b/b	b/b	b/b	c/c	e/e	b/b	c/c
23	黄B384	47.63	161.25	13.08	29	b/b	b/b	b/b	c/c	e/e	b/b	c/c
24	黄B385	25.54	165.13	14.20	62	b/b	a/b	b/c	c/c	e/e	b/b	c/c
25	黄B386	29.12	172.73	14.68	52.6	b/b	b/b	b/b	c/c	b/e	b/b	c/c
26	黄B387	22.59	166.52	14.46	26	b/c	b/b	b/b	c/c	e/g	b/b	c/c
27	黄B388	19.04	163.49	14.33	46.4	b/b	b/b	b/b	c/c	b/b	b/b	c/c
28	黄B391	21.10	171.85	14.96	33.2	b/b	b/b	b/b	c/c	b/d	b/c	c/c
29	黄B392	18.54	172.90	15.38	46.9	b/b	a/b	b/b	c/c	e/e	b/b	c/c
30	黄B394	18.37	176.89	15.64	41.3	b/b	b/b	b/b	c/c	e/e	b/b	c/c
31	黄B395	17.15	181.26	16.12	32.3	b/b	b/b	b/b	c/c	b/b	b/b	c/c
32	黄B396	16.13	181.87	16.21	41.2	b/c	b/b	b/b	c/c	b/b	b/b	c/c
33	黄B398	10.05	172.95	16.12	36.5	b/b	b/b	b/b	c/c	e/e	b/b	c/c
34	黄B399	11.84	167.28	15.47	50.8	b/b	b/b	b/b	c/c	b/e	b/b	c/c
35	黄B400	10.18	163.07	15.28	32.4	b/c	b/b	b/b	c/c	e/e	b/b	c/c
36	黄B402	5.38	167.92	16.24	51.8	b/c	b/b	b/b	c/c	b/e	b/b	c/c
37	黄B403	6.27	157.89	15.17	44.6	b/b	b/b	b/b	c/d	e/g	b/b	c/c
38	黄B404	10.63	158.40	14.80	26.4	b/b	b/b	b/b	c/c	b/b	b/b	c/c
39	黄B405	17.00	156.71	14.03	26	b/b	b/b	b/b	c/c	b/f	a/b	c/c
40	黄B408	12.08	150.49	13.77	51.6	b/c	b/b	b/b	c/c	b/e	b/b	c/c
41	黄B409	5.71	151.65	14.50	39	b/b	a/b	b/b	c/c	b/b	b/b	c/c
42	黄B410	7.36	142.46	13.42	48.8	b/b	b/b	b/b	c/c	b/e	b/b	c/c
43	黄B411	12.40	143.20	13.13	45.2	b/b	b/b	b/d	c/c	e/e	b/b	c/c
44	黄B412	15.61	144.56	13.13	23.4	b/b	b/b	b/b	c/c	b/e	b/b	c/c
45	黄B413	17.55	148.57	13.34	26.6	b/c	b/b	b/b	c/c	b/b	b/b	c/c
46	黄B414	25.42	147.04	12.61	36.2	b/b	b/b	b/b	c/c	b/b	b/c	c/c
47	黄B415	24.81	144.28	12.41	61	b/b	b/b	b/b	c/c	e/e	b/b	c/d
48	黄B416	15.59	134.64	12.07	28.4	b/b	a/b	b/c	c/e	b/e	b/c	c/c
49	黄B417	17.73	133.41	11.89	32	b/b	b/b	b/b	c/c	e/g	b/c	c/c
50	黄B418	13.41	133.98	12.15	35	b/b	b/b	b/b	c/c	e/e	b/c	c/c

合には白となる．異なる対立遺伝子から成るヘテロ接合体 (heterozygote) a/b の場合には，いくつかの可能性が考えられる．対立遺伝子 a の効果だけが表れて赤い花をつける場合には， a は b に対して優性であり (逆に白い花の場合には， b が a に対して優性である)， b は a に対して劣性であると言う．ヘテロ接合体の表現型がそれぞれの表現型の中間型になる場合 (この例の場合にはピンク色など) もあり，この場合，対立遺伝子 a と b は互いに共優性であるという．

アイソザイムやマイクロサテライト DNA を用いて分析を行う場合，データの質を高めるために，通常，複数の遺伝子座について遺伝子型データを取る． L 個の遺伝子座について分析し，遺伝子座 l ($l = 1, 2, \dots, L$) で K_l 種類の異なる対立遺伝子 (K_l は座 l によって異なる)

$$\{a_{l,1} \ a_{l,2} \ \dots \ a_{l,K_l}\} = \{a_{l,k}\} \quad (k = 1, 2, \dots, K_l)$$

が検出されたとする．各個体の遺伝子型は一組の対立遺伝子から構成されるので，個体 i の遺伝子型 G_i は，

$$(2.1) \quad G_i = \prod_{l=1}^L A_{l,i,1} A_{l,i,2}$$

で表され， $A_{l,i,t}$ ($t = 1$ or 2) には $a_{l,1}$ から a_{l,K_l} のいずれか 1 種類の対立遺伝子が占める．それで

$$(2.2) \quad G_i = \prod_{l=1}^L A_{l,i,1} A_{l,i,2} = \prod_{l=1}^L a_{l,k(l,i,1)} a_{l,k(l,i,2)} \quad (1 \leq k(l,i,1) \leq k(l,i,2) \leq K_l)$$

と記述することも可能である．

以上のように，遺伝変異データは本質的に質的であって，量的ではない．また，式 (2.1) からわかるように，個体ベースの遺伝変異データは (1) 対立遺伝子 ($A_{l,i,t}$) (2) 単一遺伝子座での遺伝子型 ($A_{l,i,1} \ A_{l,i,2}$) (3) multilocus genotype (複数遺伝子座の遺伝子型を 1 セットのデータとして扱う； $\prod(A_{l,i,1}, A_{l,i,2})$) という 3 層の階層構造を有している．その結果， K_l 種類の階層 1 から始めて階層 2 ではその任意の組み合わせとして $K_l(K_l + 1)/2$ 種類，階層 3 では $\prod_{l=1}^L K_l(K_l + 1)/2$ 種類にも膨れ上がる．後節でも述べるが，質的データとして細かく分類すると multilocus 階層ではあまりに細かすぎ (個体の大半がどこか 1 ヶ所の違いを有し合っている場合もある)，それで階層 (1) (2) でカテゴリカルに荒く分類したり量的データに変換して集約を図ったりするのだが，その過程で何らかの情報の欠落が必然的に生じる．

植物は通常，花粉と種子によって移動し，次世代の新しい個体を形成する．この時，花粉により飛散する際の各花粉の保有している遺伝子のセット P_i は，花粉が半数体 (haploid) であるため，

$$(2.3) \quad P_i = \prod_{l=1}^L A_{l,i} \quad (A_{l,i} \in \{a_{l,1} \ a_{l,2} \ \dots \ a_{l,K_l}\})$$

となる．一方，種子は，花粉 (父) 親と母樹に由来する配偶体が受精して形成された二倍体であるため，種子により散布される際の遺伝子のセット S_i は式 (2.1) の G_i と同一のデータ構造となり

$$(2.4) \quad S_i = \prod_{l=1}^L (A_{l,i,1}, A_{l,i,2})$$

となる．個体 i の遺伝子座 l において，母親由来の対立遺伝子を $A_{l,i}^S$ ，父 (花粉) 親由来の対立

遺伝子を $A_{l,i}^P$ とした場合、式(2.1)及び(2.3)(2.4)は、

$$(2.1') \quad G_i = \prod_{l=1}^L (A_{l,i}^P, A_{l,i}^S)$$

$$(2.3') \quad P_i = \prod_{l=1}^L A_{l,i}^P$$

$$(2.4') \quad S_i = \prod_{l=1}^L (A_{l,i}^P, A_{l,i}^S)$$

と表現することもできる。花粉による飛散距離を d_P 、種子による散布距離を d_S とした場合(これらはある分布を有する確率変数と考えられる)、世代 t から世代 $t+1$ へ更新した際の移動距離 $D_{t,t+1}$ (これも確率変数であろう)は、母親由来(中立な)対立遺伝子の場合には、

$$(2.5) \quad D_{t,t+1}(A_i^S) = d_S$$

であるが、花粉親由来の対立遺伝子の場合には、

$$(2.6) \quad D_{t,t+1}(A_i^P) = d_S + d_P$$

となる(最終的な移動距離は、母親を経由させずに、花粉親との直接の直線距離とみなすべき)。このように、どちらの親に由来する対立遺伝子かにより、1世代間での対立遺伝子の移動距離は異なる。空間的な遺伝構造は d_P や d_S の値と密接な関係にあるため、集団内の遺伝変異の空間配置を考える上で、式(2.5)(2.6)の持つ意味は大きい。なお、多くの天然林解析の場合、個々の対立遺伝子がどちらの親から受け継いだ対立遺伝子かを明らかにすることは困難である。

以上、述べてきたように、個体を構成する遺伝子の単位は遺伝子型 (G_i) であるが、繁殖の際の移動の単位は半数体であり、花粉で移動する際の半数体 (P_i) の段階と種子で移動する2倍体 (S_i) の段階の、異なる2段階から成る。集団内における遺伝変異の空間構造を解析する際、遺伝子流動の際の移動の単位は対立遺伝子なので、対立遺伝子レベルでの解析を行うことは重要である。一方、ある時点における集団内での各個体の遺伝情報は複数遺伝子座の遺伝子型で与えられる。このため、その集団内での遺伝変異の空間構造を把握するためには遺伝子型レベルでの解析も欠かせない。次節以下に記すように、こういった異なる階層を有した遺伝変異データの構造が、解析並びに結果の解釈を著しく複雑なものにしている。

3. 遺伝的多様性を表す指標

ここでは、集団遺伝学研究で広く用いられている集団内の遺伝的多様性を表す指標について列記する。これらの数値は、集団全体としての遺伝的特性を記述するために用いられ、集団内での個体の位置情報には関与しない。しかし、空間的統計解析において、集団全体の指標は空間的遺伝構造の有無やその統計検定を行う際にしばしば帰無仮説として重要な役割を果たすので、今後の便宜のためにここでまとめておくことにする。

3.1 多型的な遺伝子座の割合

ある遺伝子座において単一の対立遺伝子しか検出されない場合、「変異が認められない」と表現される。変異が認められる場合でも、変異が一定水準以下の場合、その遺伝子座の変異は単型的(monomorphic)であるとされる。その逆は多型的(polymorphic)であると言う。単型と多

型の閾値には 95%水準や 99%水準が用いられることが多い。前者の場合、最も頻度の高い対立遺伝子の頻度が 0.95 を越すと、もはやその座は多型的であるとはみなされず、単型的として処理されることになる。

このように定義された多型的な遺伝子座の、調べた遺伝子座の中での割合は、遺伝的多様度を直接示してくれる。しかし、各遺伝子座の遺伝変異が多型的か否かのバイナリーデータに集約されてしまうため、特に分析座数が多くない場合には、以下に紹介する指標で見ると遺伝的多様性に差のある集団に対しても同程度の評価を与えてしまいかねず、遺伝的多様性指標としての精度は高くない。

3.2 遺伝子座あたりの平均対立遺伝子数 (N_a)

L 個の遺伝子座について分析を行い、 l 番目の遺伝子座において検出された対立遺伝子数が K_l である時、 N_a は、

$$(3.1) \quad N_a = \sum_{l=1}^L K_l / L$$

で表される。 N_a の値が高い程、集団内により多くの対立遺伝子が保有されており、遺伝的多様性が高いことを示す。但し、 N_a は分析個体数に依存的である。つまり、たくさんの個体を調べれば頻度が低い対立遺伝子 (rare allele) が検出される確率が上昇するため、同程度の遺伝的多様性を保有した集団であっても、分析個体数が多いほど N_a の値は高くなる傾向にある。

3.3 遺伝子座あたりの有効な対立遺伝子数 (N_e)

遺伝子座 l において、対立遺伝子 $a_{l,k}$ の頻度を $p_{l,k}$ とする時、 l 遺伝子座あたりの有効な対立遺伝子数 (N_e) は、

$$(3.2) \quad N_e = \sum_{l=1}^L ne_l / L$$

ただし、

$$(3.3) \quad ne_l = 1 / \sum_{k=1}^{K_l} (p_{l,k})^2$$

によって定義される (Kimura and Crow (1964))。遺伝子座 l における ne_l は、全ての対立遺伝子の頻度が $1/K_l$ の時に観察された対立遺伝子数に等しくなり ($ne_l = K_l$)、それ以外では、 $ne_l < K_l$ である。

N_e は N_a と同様に l 遺伝子座あたりの遺伝子数を表しているが、この際、個々の対立遺伝子頻度が考慮されているため、 N_e は N_a と異なり分析個体数の影響を受けにくい。

3.4 平均ヘテロ接合度 (H)

遺伝子座 l における遺伝子型が対立遺伝子 $a_{l,k}$ のホモ接合体である個体の頻度を G_k とする時、遺伝子座 l でのヘテロ接合度 (heterozygosity) の観察値 (ho_l) と期待値 (he_l) は、

$$(3.4) \quad ho_l = 1 - \sum_{k=1}^{K_l} G_k$$

$$(3.5) \quad he_l = 1 - \sum_{k=1}^{K_l} (p_{l,k})^2$$

と定義される．数学的には，ヘテロ接合度の期待値 (he_l) は 2 つの対立遺伝子を復元抽出した際に異なる対立遺伝子が得られる確率 (ヘテロ接合体となる確率) を意味し，値が大きいほど遺伝的多様性が高いと判断される．通常，分析した全遺伝子座のヘテロ接合度から算出する平均ヘテロ接合度の観察値 ($H_o = \sum_{l=1}^L ho_l/L$) と期待値 ($He = \sum_{l=1}^L he_l/L$) により遺伝的多様性は評価される．なお， $ne_l = 1/(1 - he_l)$ である．

3.5 NAC (Number of Alleles in Common)

遺伝子座 l において個体 i と j が共通して持っている対立遺伝子数を $nac_l(i, j)$ とする．例えば，表 1 で個体 1 と 2 の遺伝子座 *Mdh-3* における遺伝子型は，それぞれ b/b と b/c なので， $nac_{Mdh-3}(1, 2) = 1$ であり，*Pgi-1* については共に c/c なので $nac_{Pgi-1}(1, 2) = 2$ である．同様に，個体 20 と 21 の *Amy-3* の遺伝子型は b/b と e/e なので， $nac_{Amy-3}(20, 21) = 0$ である．遺伝子座 l の NAC_l と L 個の遺伝子座の平均 NAC はそれぞれ

$$(3.6) \quad NAC_l = \sum_{i < j} nac_l(i, j) / (n(n-1)/2)$$

$$(3.7) \quad NAC = \sum_{l=1}^L NAC_l / L$$

で定義される (Surles et al. (1990))． NAC はヘテロ接合度と逆で，値が大きいほど集団内の個体間の遺伝的類似度が高いことを意味し，遺伝的多様性は小さい事を示す．

3.6 遺伝的多様性を表す指標の不偏推定値

上に列挙した指標や統計量は，集団からのランダムサンプリングにより推定する場合がほとんどであるが，その統計量のサンプリングに関する取り扱いを決して単純なものではない．例えば個体数 n のサンプルからヘテロ接合度 (he_l) を推定する時，その不偏推定値 (\hat{he}_l) は，

$$(3.8) \quad \hat{he}_l = 2n \left(1 - \sum_{k=1}^{K_l} \hat{p}_{l,k}^2 \right) / (2n - 1)$$

で与えられる ($\hat{p}_{l,k}$ は対立遺伝子 $a_{l,k}$ の頻度の推定値) (根井 (1990))．さらに個体のランダムサンプリングに加えて，遺伝子座の方も，染色体内の全ての遺伝子座ではなくいくつかの遺伝マーカーを抽出しており，2重のサンプリング構造という複雑な側面を有している (根井 (1990))．ヘテロ接合度や Ne の統計学的取り扱いについては Kimura and Crow (1964) や根井 (1990) で扱われているが， NAC の不偏推定値については，これまでのところ検討はなされていないようである．

4. 空間的遺伝構造を表す統計量

Sokal and Oden (1978a, 1978b) は，集団間の地理的変異を量的に記述するため，各集団を (不規則な) 格子点とみなし，それらの対立遺伝子頻度に対して格子 (lattice) 上の空間自己相関 (spatial autocorrelation) を適用した．集団内の個体レベルの空間遺伝統計は，それを個体レベルに援用する方向で行われてきている．

ここでは最初に，以下で紹介する空間統計に共通して用いられる距離階級について述べ，その後，2.2 節でも述べたように，遺伝変異データには 3 層の階層構造があるので，対立遺伝子レベル，遺伝子型レベル，multilocus レベルに区分して，統計量を列記して行く．これらの区分は単なる数学的取り扱いだけでなく生物学的背景の違いもあり，解析結果の位置付けや解釈をする上で注意を要する．

4.1 距離階級

ある集団内に n 個体が分布している時、2 個体の組み合わせ(ジョイン; join)数は全部で $n(n-1)/2$ 通りある。これらのジョインを一定の基準に基づいて H 個の距離階級(distance class)に区分する。2 個体間のユークリッド距離に基づいて、距離階級が設られる場合が圧倒的に多いが、個体間距離によらない距離階級を使った論文やその是非を論じている文献もある(Cliff and Ord(1981), Sokal and Oden(1978a), Smouse and Peakall(1999))。

個体 i と j の座標をそれぞれ X_i, X_j とし(ここで $X_i = (x_i, y_i)$ あるいは $X_i = (x_i, y_i, z_i)$ である), 個体 i, j 間の距離を $d_{i,j} = \|X_i - X_j\|$ とする。例えば、0-5 m, 5-10 m, 10-15 m, ... 45-50 m の 10 個の距離階級を設ける場合、 $h = 1, 2, \dots, 10$ に対し各階級の上限を $D(h) = 5h$ (但し $D(0) = 0$) として、階級 h における個体 i, j のジョインに対して $w_{i,j}^h$ を

$$(4.1) \quad w_{i,j}^h = \begin{cases} 1 & \text{if } D(h-1) \leq d_{i,j} < D(h) \\ 0 & \text{otherwise} \end{cases}$$

とおく。

4.2 対立遺伝子レベルの統計量

4.2.1 Moran's I statistics の遺伝子頻度への適用

遺伝子座 l の任意のひとつの対立遺伝子 $a_{l,k}$ に着目する。個体 i に対し $z_{k,i}$ を(正確には遺伝子座番号 l にも依るが、以後、特に混乱の恐れのないときは l を省略し、同様に $a_k, A_{i,1}, A_{i,2}$ などと略記する),

$$(4.2) \quad z_{k,i} = \begin{cases} 1 & \text{if } A_{i,1} = A_{i,2} = a_k \\ 0.5 & \text{if } A_{i,1} = a_k \text{ or } A_{i,2} = a_k \\ 0 & \text{if } A_{i,1} \neq a_k \text{ and } A_{i,2} \neq a_k \end{cases}$$

と定める。距離階級 h ごとに、 I_k^h を

$$(4.3) \quad I_k^h = \frac{\sum_{i,j} w_{i,j}^h (z_{k,i} - p_k)(z_{k,j} - p_k)}{\sum_{i,j} w_{i,j}^h \cdot \sum_{i=1}^n (z_{k,i} - p_k)^2 / n}$$

と定義する(Sokal and Oden(1978a))。これは格子上の空間自己相関(Moran's I statistics)を、各個体が保有している対立遺伝子 a_k のコピー数の $1/2$ (1 本の木という集団における対立遺伝子 a の頻度)として適用したものである。 I_k^h は $-1 \sim +1$ の範囲で、期待値は $-1/(n-1)$ になる(Cliff and Ord(1981))。なお、設定された調査区内の全個体に関する統計と考えると、分母の分散項 $\sum_{i=1}^n (z_{k,i} - p_k)^2 / n$ は個体数 n で割り、 $n-1$ を用いる不偏推定値は用いない。

元々 Sokal and Oden(1978a)は、複数個の集団がある時に集団 i の対立遺伝子 a_k の頻度を $z_{k,i}$ に代入して Moran's I statistics を用いることを提唱したが、天然林集団内の個体を対象とした空間的遺伝構造の解析では、各個体を(最小ながら)集団とみなして、その対立遺伝子 $a_{l,k}$ の頻度を用いているわけである。集団の対立遺伝子頻度であれば $z_{k,i}$ は $0 \sim 1$ の連続的変量となるが、1 個体という集団では、 $z_{k,i}$ は $1, 0.5, 0$ という離散的な値しか取らない。

階級 h に属するジョイン(2 個体の組み合わせ)中に、両者とも a_k のホモ接合体($z_{k,i} = 1$)である、あるいはどちらも a_k を保有していない($z_{k,i} = 0$) 組み合わせが多い場合、 I_k^h はプラスに傾く。逆にジョインの片側が a_k ホモで、もう一方は a_k を保有していないジョインが多いと、 I_k^h の値はマイナスに傾く。なお、 I_k^h におけるヘテロ接合体の寄与の仕方は興味深い。個体 i

が a_k のヘテロの場合 $z_{k,i} = 0.5$ であるが、 $(z_{k,i} - p_k)$ は対立遺伝子の頻度 p_k に依存するため、 $p_k < 0.5$ の時は a_k ヘテロは a_k ホモと正の相関を作り出すが、 $p_k > 0.5$ では a_k ヘテロと a_k ホモは「似ていない」と判断され負の相関を示し、逆に a_k を保有していない個体と正の相関を生み出す。

距離階級を x 軸に、 I_k^h を y 軸にした折れ線グラフはコアレログラム (correlogram) と呼ばれるが、これを作成することで、距離階級の増大と共に I_k^h がどのように変化するか、言い換えれば個体間距離に応じた (個体内) 遺伝子頻度の相関の強さの変化を視覚的に把握することができる。種子散布距離の小さい樹種では、近接個体は互いに同一母樹から生まれた兄弟である確率が高く、それらの母樹が、対立遺伝子 a_k を保有していればその娘たちは a_k を共有する可能性が高いため、 I_k^h は有意な正の相関を示し、コアレログラムは単調に減少して (または正から負に転落した後に微増して) 0 に落ち着く事が期待される。実際、天然林データでは、この傾向を示す例がいろいろ報告されている (Berg and Hamrick (1995), Loiselle et al. (1995), Streiff et al. (1998), Takahashi et al. (2000))。図 2 に、表 1、図 1 の集団に対する I_k^h のコアレログラムの例を示す。

I_k^h が正 (または負) の相関を示した場合、それが本当に何らかの生物学的要因による強い相関なのか、偶然でも起こりうる程度の弱いものか、有意性を検定する必要がある。それは、個体位置を固定した上で遺伝子データのみをランダムに取りかえる randomization test を例えば 999 回行い、実際のデータから求めた I_k^h が 999 + 1 個の I_k^h の中で上位または下位 5% 以内に入るか否かにより 95% 水準で検定する。なお、より簡略に、 I_k^h を正規近似する方法もある (Sokal and Oden (1978a), Cliff and Ord (1981))。

ところで、 I_k^h は H 個の距離階級で L 個の遺伝子座での K_l 個の対立遺伝子のすべてについて計算できるので、理論的には合計 $H \sum_{l=1}^L K_l$ 個の I_k^h が算出される。しかし、それらが独立に意味を有しているわけではない。まず、 $K_l = 2$ の遺伝子座については $z_{1,i} = 1 - z_{2,i}$ かつ $p_1 = 1 - p_2$ より $(z_{1,i} - p_1) = -(z_{2,i} - p_2)$ となり、対立遺伝子 1 を用いても 2 を用いても、両者の I_k^h は全く同一となる。 $K_l > 3$ の遺伝子座においても、例えばアイソザイム遺伝子座の場

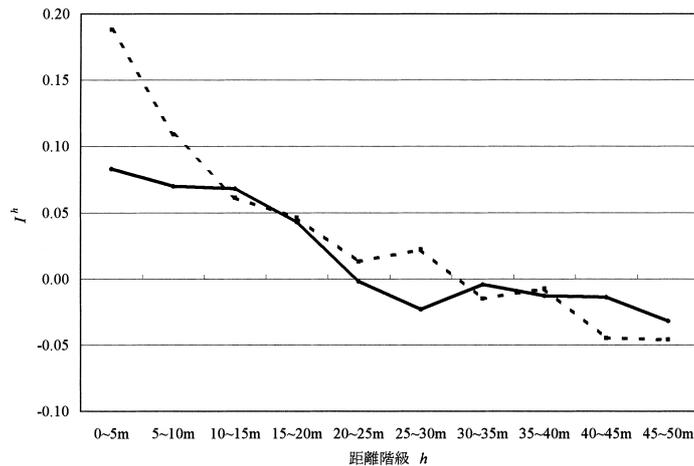


図 2. 表 1・図 1 の集団に対するコアレログラムの例。距離階級は 5 m ごと。実線は $Mdh-3$ 遺伝子座における対立遺伝子 b を用いた I_k^h 、点線は $Pgi-3$ 遺伝子座の対立遺伝子 c を用いた場合の I_k^h を示す。最初の方の距離階級で高い正の値を示している (Takahashi et al. (2000) より改変)。

合, 最も頻度の高い対立遺伝子の頻度が 0.9 くらいであることが多い. それで, 例えば対立遺伝子頻度が $p_1 = 0.9, p_2 = 0.08, p_3 = 0.02$ の場合, 対立遺伝子 2 を用いた時に $z_{2,i} = 0$ となる個体には, a_1 ホモ ($z_{1,i} = 1$) に加えて a_1/a_3 のヘテロ接合体 ($z_{1,i} = 0.5$) や a_3/a_3 のホモ接合体 ($z_{1,i} = 0$) も含まれ, 全ての i に対して $z_{1,i} = 1 - z_{2,i}$ が成り立つわけではないが, $p_3 = 0.02$ を鑑みれば大半の個体について $z_{1,i} = 1 - z_{2,i}$ が成立し, a_1 と a_2 は酷似したコアレログラムを生成する. このように (同一遺伝子座内の) 対立遺伝子間で $I_1^h, I_2^h, \dots, I_{K_l}^h$ は, 独立でないどころかほとんど同値な統計量である場合すらある. 本来, 対立遺伝子は染色体内からランダムサンプリングされていると考えたいわけだが, 同一遺伝子座内の異なる対立遺伝子はこの仮定を全く満たしていない. さらに, 各距離階級間も, 同一個体を異なる階級にまたがる join の形で繰り返し使用するため, やはり独立ではない. このような 2 重の意味で独立でない統計量を空間的遺伝変異解析では扱わなければならない.

4.2.2 Coancestry

4.2.1 節と同じ記号の元で,

$$(4.4) \quad \rho_k^h = \frac{\sum_{i,j} w_{i,j}^h (z_{k,i} - p_k)(z_{k,j} - p_k)}{\sum_{i,j} w_{i,j}^h \cdot p_k(1 - p_k)} + \frac{2}{\left(8 \sum_{i,j} w_{i,j}^h + 1\right)^{0.5} - 1}$$

を距離階級 h における対立遺伝子 a_k についての coancestry という (Loiselle et al. (1995)). 第 2 項は, 有限個の格子点を扱う事に伴う補正項と思ってよからう. この ρ_k^h は元来, 集団における対立遺伝子頻度をもとに集団内での identity-by-descent (任意の 2 個体のそれぞれから任意に選んだ遺伝子が昔は同一祖先内の同一遺伝子だった確率. 例えば全兄弟なら $1/4 = 0.25$) を推定する統計量として用いられていた (Cockerham (1969)). それを Loiselle et al. (1995) は, 個体ベースの空間構造解析に用いた.

ρ_k^h が大きい正の値である時, そのような 2 個体は共通祖先を有し血縁関係にある事を示す. 従って, 花粉流動と種子散布によって血縁関係が集団内でどのように広がっているかを見るには最適な統計量に見える. それで Loiselle et al. (1995) では, I_k^h と ρ_k^h は統計量としては似通ったものではあるが, 集団遺伝学的な根拠がある分 ρ_k^h の方が優れていると判断して ρ_k^h を用いている. ところで, Hardy-Weinberg 平衡状態下にある集団のように, 頻度 p_k の対立遺伝子 a_k のホモ接合体の割合が p_k^2 , ヘテロが $2 p_k(1 - p_k)$, a_k を含まない個体の頻度が $(1 - p_k)^2$ なら, I_k^h の分母の分散項 (式 4.3) は,

$$(4.5) \quad \sum_{i=1}^n (z_{k,i} - p_k)^2 / n = p_k^2 \cdot (1 - p_k)^2 + 2 p_k(1 - p_k)(0.5 - p_k)^2 + (1 - p_k)^2 \cdot (-p_k)^2 \\ = p_k(1 - p_k)/2$$

となり, $\rho_k^h = I_k^h/2$ という単純な関係が導かれる. Hardy-Weinberg 平衡でない集団でも (集団に依存する定数倍だけの関係式で ρ_k^h と I_k^h はむすばれる (Hardy and Vekemans (1999)). 従って, ひとつの集団に対してでは ρ_k^h と I_k^h はほとんど空間統計として同値である. しかし, 異なる集団を比較する場合は, 何らかの差が生じる可能性を否定できない.

統計的な有意性の検討は, I_k^h と同様な randomization test で行う.

4.3 遺伝子型レベルの統計量

空間的遺伝構造は花粉流動や種子散布といった遺伝子流動と密接な関係がある. 2.2 節にも記したが, そのような遺伝子流動の単位は, 対立遺伝子であって, その組み合わせの遺伝子型

ではない。しかし、実際の分析で得られるのは各個体の遺伝子型データである(表 1)。遺伝子型レベルの解析では、質的データである遺伝子型に基づいて各個体が分類される。

遺伝変異データを遺伝子型レベルで行う統計量に Standard Normal Deviates (SND) がある。ある一つの遺伝子座についてみた場合、そこには同一遺伝子型間のジョイン (like join, aa と aa , ab と ab , など) と異なる遺伝子型間のジョイン (unlike join, aa と ab , aa と bc , など) がある。個体数 n の集団内の遺伝子型 G_A の個体数を n_A とする時、 G_A の like join の数の期待値は、

$$(4.6) \quad \mu_{AA}^h = \sum_{j>i} w_{i,j}^h \cdot \frac{n_A(n_A - 1)}{n(n - 1)}$$

である。距離階級 h における G_A の like join の観察数が J_{AA}^h の時、距離階級 h の G_A の like join の SND は、

$$(4.7) \quad \text{SND}_{AA}^h = \frac{J_{AA}^h - \mu_{AA}^h}{\sqrt{\text{Var}}}$$

によって求められる。ここで Var は (有限格子上で) like join の数の分散だが、その算出式は長くなるのでここでは省略する (Sokal and Oden (1978a) を参照されたい)。

遺伝子型 G_A と G_B の個体数がそれぞれ n_A , n_B の時、 G_A と G_B の unlike join の数の期待値は、

$$(4.8) \quad \mu_{AB}^h = \sum_{j>i} w_{i,j}^h \cdot \frac{2n_A n_B}{n(n - 1)}$$

となり、距離階級 h における G_A と G_B の unlike join の観察数が J_{AB}^h の時、距離階級 h の G_A と G_B の unlike join の SND は、like join の時と同様な式 (4.7) によって求められる。SND はジョイン数をもとに算出されるため、join-count statistics と呼ばれる。遺伝子型分布がランダムな時、SND の期待値はゼロとなる。期待値よりも多く観察されるときにはプラスで、その遺伝子型の個体数がその距離階級で多く見られることを示唆する。逆に少ないときにはマイナスに傾き、解釈も逆になる。それらが統計的に有意か否かは ρ_k^h や I_k^h と同様に randomization test で行えるが (4.7) は正規近似できるため、簡便に行う事も可能である (Sokal and Oden (1978a), Cliff and Ord (1981))。

ホモ接合体 $a_k a_k$ の like join における有意に正な SND はその遺伝子型の集中分布を表し、また対立遺伝子 a_k の集中も意味すると言えよう。ところが、 $a_k a_{k'}$ ヘテロの like join や $a_k a_k$ ホモと $a_k a_{k'}$ ヘテロの unlike join では、 I_k^h や ρ_k^h と同じように遺伝子頻度 p_k に応じた解釈が必要と思われる。また、 $a_k a_{k'}$ ヘテロと $a_k a_{k''}$ ヘテロの unlike join や a_k ホモと $a_{k'}$ ホモの unlike join など多種多様な組み合わせがあり、詳細に検討すると様々な面白い知見が見出されるかもしれない。実際、Epperson (1995a, b, c) などの simulation study では、種子及び花粉が近隣個体間でしか交換されない (isolation by distance) という仮定の下では、世代数を経るとともに、ホモ接合体のパッチ (genetic patch) が形成されて行き、さらにヘテロもある種のパターン形成に寄与する。その結果に基づき、遺伝子型レベルでのきめ細かな解析が天然林内に形成される空間的遺伝構造では重要であると主張している。しかし、そのような解析は、対立遺伝子が二つしかない遺伝子座という仮想集団だからこそ可能であったわけで、前述のように対立遺伝子 K_l 個の遺伝子座において検出される遺伝子型の最大数は $K_l(K_l + 1)/2$ 種類あり、それが L 個の遺伝子座と H 個の距離階級で算出できるのだから、理論的には $H \sum_{l=1}^L (K_l(K_l + 1)/2) \cdot (K_l(K_l + 1)/2 + 1)/2$ 個という膨大な数の SND が算出可能となる。この膨大な数の SND 一つ一つについて詳細な解析を行い適切な解釈を与える事は至難の業であるし、紙面の都合上、公表も不可能である。

そのため, like join と unlike join という区分ごとに, 各距離階級で有意であった SND の割合で論じられることが多い. また, 遺伝子型頻度が 0.05 以下の遺伝子型や $\min(\mu^h) < 1$ のジョインは, 解析から除外されることが多い (Epperson and Alvarez-Buylla (1997), Takahashi et al. (2000)).

本来, SND では遺伝子型データをそのまま計算に用いるので, ρ_k^h や I_k^h の対立遺伝子レベルの時のように, 対象とする対立遺伝子以外をひとまとめにするといったような遺伝変異データの集約が行われていない. そのため, SND の解析解像度は, より高いといえる. しかし, 得られた解析結果を有意だった割合として要約してしまうのでは, SND の持つ特性がその段階で損なわれているように思えてならない. また, 座内の遺伝子型間 (即ち SND 間) には何らかの相関が伴うと考えられる. 例えば両親の遺伝子型が $a_k a_k$ と $a_k a_{k'}$ なら $a_k a_k$ ホモと $a_k a_{k'}$ ヘテロの子供が誕生するから, a_k ホモと a_k ヘテロの個体間には交配や親集団の遺伝構造を背景に持つ相関があると思われる. この点については, simulation による人工的集団で若干の検討がなされている (Epperson (1995a, b, c)) 程度である.

4.4 Multilocus レベルの統計量

multilocus レベルの解析では, 各個体の複数遺伝子座の遺伝子型をひとまとまりのデータとして扱う. ρ_k^h や I_k^h にしる SND にしる, 対立遺伝子や遺伝子型のすべてについて計算するため膨大な数の互いに相関を持つであろう統計量が算出される. そして, 有意なものの割合を求めたり (相関を伴うにもかかわらず) 単純平均をとったりするが, こういった集約になんらかの統計的根拠があるわけではない.

multilocus レベルの解析は, 遺伝データを同時分布として扱い, 統計量は各距離階級 h につき最初から 1 つに集約される.

4.4.1 Number of Alleles in Common (NAC)

3.5 節で取り上げた NAC に距離階級概念をあてはめ, 空間統計としても活用できる (Berg and Hamrick (1995)). 空間統計量としての NAC は,

$$(4.9) \quad NAC_l^h = \frac{\sum_{i < j} w_{i,j}^h n_{ac_l}(i, j)}{\sum_{i < j} w_{i,j}^h}$$

$$(4.10) \quad NAC^h = \sum_{l=1}^L NAC_l^h / L$$

と定義される. NAC_l^h と NAC^h の期待値は (multilocus) 遺伝子型空間分布がランダムなら, 集団全体の平均である式 (3.6) と (3.7) に一致する. 遺伝変異が空間的にランダムの際の期待値が SND ではゼロ, 4.2.1 節の I_k^h では, $-1/(n-1)$ であるが, NAC の場合にはその集団特有の値となる. それで複数集団で空間構造を比較したい場合には, NAC^h/NAC の 1 からのズレ具合 (または $NAC^h - NAC$ の 0 からのズレ具合) を見る.

4.4.2 Smouse & Peakall の r^h

Smouse and Peakall (1999) は以下のような統計量 r^h を提唱した. まず, 個体 i の遺伝子座 l における遺伝子型を, 集合 $\{(y_1, y_2, \dots, y_{K_l}) \mid y_k \in \{0, 1, 2\}, \sum_{k=1}^{K_l} y_k = 2\}$ 上に,

$$(4.11) \quad \mathbf{y}_{l,i} = (y_{l,i,1}, y_{l,i,2}, \dots, y_{l,i,K_l})$$

と表す事にする. 例えば座 l での遺伝子型が $a_k a_{k'}$ ヘテロなら, $(0, \dots, \frac{1}{k}, \dots, \frac{1}{k'}, \dots, 0)$, $a_k a_k$ ホ

もなら $(0, \dots, \frac{2}{k}, \dots, 0)$ となる．次に (遺伝子座 l での) 個体 i と j の遺伝子型の差異をまず

$$(4.12) \quad g_{l,i,j} = \frac{1}{2K_l} \sum_{k=1}^{K_l} (y_{l,i,k} - y_{l,j,k})^2$$

と定めるが，低頻度の対立遺伝子の保有数の違いは高頻度のものより 2 個体間の遺伝的差異をより顕著に示すと考え，頻度の逆数でウェイトをかけて

$$(4.12') \quad \tilde{g}_{l,i,j} = \frac{1}{2K_l} \sum_{k=1}^{K_l} \frac{(y_{l,i,k} - y_{l,j,k})^2}{p_{l,k}}$$

としてもよい．4.4.1 NAC ではこのような遺伝子型間の差異 $nac_l(i, j)$ を直接用いて空間統計にしているが (4.9 式), Smouse and Peakall (1999) では，この差異に関する重心からの離れ具合の積 (共分散) $c_{l,i,j}$ に直して用いている．即ち， $\{y_{l,i}\}$ の重心は

$$(4.13) \quad \bar{y}_l = (y_{l,1}, y_{l,2}, \dots, y_{l,K_l}) = 1/n \cdot \sum_{i=1}^n (y_{l,i,1}, y_{l,i,2}, \dots, y_{l,i,K_l})$$

で与えられる (これは遺伝子頻度を使って $\bar{y}_l = 2(p_{l,1}, p_{l,2}, \dots, p_{l,K_l})$ と表す事もできる) が (遺伝子座 l での) 個体 i と j の遺伝型差異の共分散

$$(4.14) \quad c_{l,i,j} = \frac{1}{2K_l} \sum_{k=1}^{K_l} (y_{l,i,k} - \bar{y}_{l,k})(y_{l,j,k} - \bar{y}_{l,k})$$

$$\left(\text{または } \tilde{c}_{l,i,j} = \frac{1}{2K_l} \sum_{k=1}^{K_l} \frac{(y_{l,i,k} - \bar{y}_{l,k})(y_{l,j,k} - \bar{y}_{l,k})}{p_{l,k}} \right)$$

の，すべての遺伝子座にわたる平均 $C_{i,j} = \sum_{l=1}^L c_{l,i,j}$ (または $\tilde{C}_{i,j} = \sum_{l=1}^L \tilde{c}_{l,i,j}$) で，個体 i と j の遺伝型差異を数量化する．さらに，この遺伝的差異を NAC のように単純に join の数で割って

$$(4.15) \quad C^h = \frac{\sum_{i < j} w_{i,j}^h C_{i,j}}{\sum_{i < j} w_{i,j}^h}$$

の形の平均を使うのではなく，また I_k^h のように集団全体の分散で正規化するのではなく，距離階級 h におけるジョインの中で個体 i が寄与した回数を数え，その回数による重み付けを行った分散を用いる．すなわち，

$$(4.16) \quad X_i^h = \sum_{\substack{j=1 \\ j \neq i}}^n w_{i,j}^h, \quad c_{l,i} = \sum_{k=1}^{K_l} \frac{(y_{l,i,k} - \bar{y}_{l,k})^2}{2K_l}, \quad C_i = \sum_{l=1}^L c_{l,i} \quad \text{と} \text{お} \text{い} \text{て},$$

$$r^h = \frac{\sum_{i < j} w_{i,j}^h C_{i,j}}{\sum_{i < j} X_i^h C_i} \quad \left(\text{または対立遺伝子頻度 } p_{l,k} \text{ で割る } \tilde{C} \text{ を用いる形} \right)$$

と定義するのである．本稿ではこれを Smouse & Peakall の r^h と呼ぶことにする．

4.4.3 Rogers の遺伝距離に基づく統計量

Furnier et al. (1987) は，各座での Rogers の遺伝距離 (元々集団間の地理的遺伝的差異を数量化するため各集団の遺伝子頻度を使って定義された) を (4.2 節と同じように) 1 個体という集団に適用

し、遺伝子座間で平均することにより multilocus レベルでの解析とした。個体 i と j の Rogers の遺伝距離 $R_{i,j}$ は (4.11) 式の遺伝子型表記 $\{(y_1, y_2, \dots, y_{K_l}) \mid y_k \in \{0, 1, 2\}, \sum_{k=1}^{K_l} y_k = 2\}$ を 2 で割って $\{(y_1, y_2, \dots, y_{K_l}) \mid y_k \in \{0, 0.5, 1\}, \sum_{k=1}^{K_l} y_k = 1\}$ に直し、ヘテロが $(0, \dots, 0.5, \dots, 0.5, \dots, 0)$ 、ホモが $(0, \dots, 1, \dots, 0)$ と表されるようにして(結果として(4.2)の形になる)、2 遺伝子型間のユークリッド距離の $1/\sqrt{2}$ で定義される。

$$(4.17) \quad R_{i,j} = \frac{1}{L} \sum_{l=1}^L \sqrt{\frac{1}{2} \cdot \sum_{k=1}^{K_l} (y_{l,i,k} - y_{l,j,k})^2}$$

ルートをとる点で Smouse & Peakall の r^h と異なり、距離としてはより自然な感じを与えるが、そのため座間での和は一度座内で集約された数値の和になり、Smouse & Peakall の r^h のように全体が各成分の総和という関係はない。ちなみに NAC は、遺伝子型を集合 $\{(y_1, y_2, \dots, y_{K_l}) \mid y_k \in \{0, 1, 2\}, \sum_{k=1}^{K_l} y_k = 2\}$ 上で表した時のマンハッタン距離を用いた場合に相当する。

4.4.4 $I_{l,k}^h$ の multilocus 平均

multilocus に渡って多数、算出された $I_{l,k}^h$ 値は、様々に平均されている。Mahy et al. (1999) などは 0.04 以上の頻度を有する)すべての対立遺伝子についての $I_{l,k}^h$ 値の単純平均のコアレログラムを表示している。Takahashi et al. (2000) では、各座から最も頻度の高いものだけを選び、bootstrap も活用して平均を議論している。また、Streiff et al. (1998) では、分子と分母を別々に multilocus で積算した

$$I^h = \frac{\sum_{l=1}^L \sum_{k=1}^{K_l} \sum_{i,j} w_{i,j}^h (z_{l,k,i} - p_{l,k})(z_{l,k,j} - p_{l,k})}{\sum_{i,j} w_{i,j}^h \cdot \sum_{l=1}^L \sum_{k=1}^{K_l} \sum_{i=1}^n (z_{l,k,i} - p_{l,k})^2 / n}$$

を用いている。

対立遺伝子レベルや遺伝子型レベルにそれぞれ生物学的な背景が伴っていたように、multilocus レベルにも独自の背景がある。まず、3 節のヘテロ接合度や NAC のように遺伝子座間の平均を取ると言う事は暗に座間が独立である事(メンデルの独立の法則)を仮定する。その場合、座 1 での $a_{1,k}a_{1,k'}$ 型の頻度を $G_{1,kk'}$ 、座 2 での $a_{2,m}a_{2,m'}$ 型の頻度を $G_{2,mm'}$ とすると、multilocus での遺伝子型 $a_{1,k}a_{1,k'} \times a_{2,m}a_{2,m'}$ 型の頻度は $G_{1,kk'}G_{2,mm'}$ となる。しかし実際にはこれを満たさない場合もあり、古くから研究されている例が、その 2 座が同一染色体上の近傍に位置しているために交配の際に連鎖が生じ、特定の multilocus 型が独立な場合より多くなる事例である。これを連鎖不平衡(linkage disequilibrium)という。ところがもっと単純に、限られた個体数の繁殖個体(集団の有効なサイズが小さい)集団から次代集団が形成される場合でも連鎖不平衡が起こり得る。例えば母樹として繁殖に貢献できた個体の大半が multilocus で、 $a_{1,1}a_{1,1}/a_{2,2}a_{2,2'}$ 型と $a_{1,2}a_{1,2}/a_{2,1}a_{2,1'}$ 型だったとすると(父方として貢献した個体には他の遺伝子型も存在するとする)、次世代では第 1 遺伝子座に a_1 を持っている個体は第 2 遺伝子座に $a_{2,2}$ を持つ確率が高くなる。また、第 1 遺伝子座に $a_{1,2}$ を持つなら第 2 座に $a_{2,1}$ を持つ個体が多いだろう。それまでその樹種が天然分布していなかった地域に少数ながら種子が拡散して、発芽・成長・繁殖した場合、これを親集団とする次の世代は親集団の有効なサイズが著しく小さいため、上に記したような、任意の 2 個体が兄弟(sibling)などの血縁関係のある個体である確率(類似した multilocus haplotype を共有している確率)が高い集団を形成する。これを創始者効果(founder effect)と言うが、このような場合、上に記したような状況が起こりやすい。

このように、遺伝子流動における移動の単位は1個の対立遺伝子でもなければ遺伝子型でもなく、multilocus レベルでの haplotype(式(2.3))である。そのため、multilocus レベルで解析することにより、更新の際の遺伝子流動について間接的ながら知ることができるはずで、また座ごとに独立して計算するヘテロ接合度等では検出できない遺伝的多様性の消失を見出せるかもしれない(Takahashi et al.(2000))。multilocus レベルでの解析は、単に統計量間の従属性や同時分布としての取り扱いだけでなく、生物学的にも他の2つのレベルとは異なった情報を提供してくれるのである。

4.5 空間的遺伝構造の検定

ここまで、繰り返して「個体位置を固定した上で遺伝子だけをランダムに換える randomization test により遺伝子分布がランダムか否か検定する」と記述したが、これにも対立仮説の設定に応じた異なる randomization が考えられる。一般には各個体の遺伝子型を遺伝子座ごとにランダムに換えるようだが、対立遺伝子レベルで換える事も考えられるし、各個体の multilocus 遺伝子型を保ちながら取り換える事も考えられるし、これらの併用も可能である(Berg and Hamrick (1995))。すべて、その研究が遺伝子の何を調べるためか、その目的に応じて対立仮説が違ってくるのは当然であるが、例えば座間の相関(連鎖不平衡)はその有無が最初からわかっている仮定できるものではなく、データから判断されるべきものであろう。結局、異なるレベルでの randomization を併用して順に帰無仮説を棄却して行くしかないのだろうか。Berg and Hamrick (1995)以来、大切な問題のはずなのにあまり議論されていないようである(なお個体の位置まで動かして考える研究の類は Shimatani(2002), Shimatani and Takahashi(2003)で始まったばかりである)。

5. 考察

ここまでの論述で明らかなように、表1のような遺伝変異データを手にした時、解析者には対立遺伝子ごとに解析するか、各遺伝子座で遺伝子型ごとに取り扱うか、multilocus レベルで集約した1つの統計量を用いて解析するかという多岐にわたる選択肢がある。対立遺伝子レベルや遺伝子型レベルで多数の統計量を算出し、そこから得られた統計的に有意な値について、逐一意味付けを行う事は労力を要する。その点、multilocus レベルの解析では、一つの統計量によって遺伝構造を定量化できるので、すっきりと結論を導けるように見える。というのは、対立遺伝子レベルや遺伝子型レベルで多数の統計量を算出した場合、算出された(特に同一座内の対立遺伝子の)統計量間には相関があるので、遺伝構造を過剰あるいは過小評価してしまう可能性がある。multilocus 統計量では複数の確率変数の同時分布とみなしているため、一見この問題を回避しているように見える。特に、Smouse & Peakall の r^h は、単一对立遺伝子レベル $((y_{l,i,k} - \bar{y}_{l,k})(y_{l,j,k} - \bar{y}_{l,k}))$ 、単一遺伝子座レベル $C_{l,i,j} = \sum_{k=1}^{K_l} (y_{l,i,k} - \bar{y}_{l,k})(y_{l,j,k} - \bar{y}_{l,k})/2K_l$ または $\tilde{C}_{l,i,j}$ 、multilocus レベル $C_i = \sum_{l=1}^L C_{l,i}$ 並びに $r^h = \sum_{i<j} w_{i,j}^h C_{i,j} / \sum_{i<j} X_i^h C_i$ の3つのレベルで算出することが可能で、しかも、単一对立遺伝子レベルの解析結果を合成することにより単一遺伝子座レベルの統計量となり、単一遺伝子座レベルの値を合成することにより multilocus レベルの値となる。すなわち、これらの異なるレベルで遺伝変異データを数値化する際に同一のルールに従っているため、例えばどのレベルのどの成分が最終結果 r^h にどのように寄与しているかを考察できる。一方、同じ3階層の統計量の結果ではあっても、 I_k^h と SND と NAC の3者では、遺伝変異データを数値化する段階で既に異なっており、3者を有機的に包括して考察を進めることは難しく、単に多数の統計量を羅列しただけに終わりがかねない。このように、異なるレベルでの解析を、統一的に一つの統計量で取り扱える点で、Smouse & Peakall の r^h を用いるメリットは大きい(なお、このような解析は、NAC でも可能である)。

しかし実際のところ, Smouse & Peakall の r^h にせよ NAC にせよ, 要するに対立遺伝子ごとのその保有数 (またはその半分) の差を加算しているに過ぎず, 相関の問題を解決したことにはなっていない点は認識しておくべきであろう.

本稿で紹介した空間的遺伝構造の解析に用いられる統計量には, 共通した構造が見られる. 1) 距離階級を用いる (なお個体分布とそれらの遺伝子型分布を同時に扱う方法では, 一定幅で区切られた距離階級ごとの統計量でなく, 任意の距離に対して定義される関数を用いる (Shimatani (2002), Shimatani and Takahashi (2003))). 2) 分子は, 各距離階級でジョインを構成する 2 個体の遺伝的類似度を表す数値やそれらの共分散である. 3) 分母は, その距離階級に属する join の数で割って 1 join あたりの平均値に直すか, 統計量によっては用いているシステムでの遺伝変異の分散等を組み入れて正規化する.

例えば, Smouse and Peakall (1999) における対立遺伝子レベルの r^h (4.4.2 節) と I_k^h (4.2.1 節) は非常に似ており, どちらの分子も対立遺伝子における個体 i と個体 j の重心 (すなわち対立遺伝子頻度) からの偏差の積となっている. 両者の異なる点は, 後者の分母が全サンプルについての分散であるのに対して, r^h の分母は各距離階級における各個体の出現度数によって重み付けした分散になっている点 (と, r^h ではすべての対立遺伝子について加算していく点) である. 4.2.2 節でも記したように coancestry ρ_k^h も基本的に同じような形で, 単に正規化する分母が違うだけである. 一方 NAC では, 個体間の遺伝的類似度は, 個体 i と個体 j が共有している対立遺伝子の数になっており, 共分散の形ではない. 直接の遺伝的類似度を用いるか共分散の形に直して用いるか. これは Smouse and Peakall (1999) が明示した問題であるが, これからも議論が続くものと思われる.

遺伝構造を解析するための空間統計は, ただ単に数値を算出しランダム性を検定するために用いるのではなく, 集団遺伝学的な, あるいは生態学的なプロセスを明らかにするために計算するのであって, 統計量の数値そのものに何らかの生物学的意味が伴うならさらに好ましいであろう. 例えば, 4.2.2 Coancestry では, 任意の 2 個体から任意に選んだ 2 遺伝子が (何代か前の) 祖先では同一個体内の同一遺伝子であった確率 (identity by descent) を推定しようとしている. 従って, その値から, どのくらい接近しているとそれらが兄弟である確率が高くなるか, 直接に判断できる. 推定方法や信頼性には問題があるようだが, 統計量が古くから集団遺伝学で使われてきている集団内の遺伝的多様性指標や集団間の遺伝的距離と密接な関係を持っているのなら, その数値を単にランダム性の検定のみを用いるのはもったいなく, 統計量の優劣は単に統計的検定能力の正確さで評価してよいものではない.

集団内の遺伝構造を解析する際に取り扱う遺伝変異データは, 2.2 節でも述べたように複雑な階層構造を持ったデータセットである. このようなデータセットの性質をたったひとつの数値で余すことなく表現する事など不可能であろう. 実際の天然林を解析した多くの研究では, ランダムとは考えられない遺伝構造, 明白な遺伝子の集中分布が見られる集団が多く報告されている. そのような場合, どの対立遺伝子が, どの遺伝子型が, どの multilocus 遺伝子型がどのように全体の空間構造に寄与しているか? そしてその背後にはどのような生態学的なプロセスがあるのか? また, それぞれの統計量の特徴 (良し悪し) は, その数式や少数のデータに対する適用だけでは不明な部分が多く, 多様なデータに利用されることで次第に明らかにされて行くものである. 結局のところ, 現段階では, 最善と思われる単一の統計量だけを用いて一気に結論を導くより, 計算機の速度向上もあるので, 様々な統計量を試み, どのような違いが生じるか. 生じたのなら何が原因か. ひとつひとつ慎重に精査して行く事が面白いように思えるし, 時間はかかっても新しい発見への道も開きやすいのではないだろうか.

6. おわりに

今回は空間的遺伝構造を中心にまとめたが、それは森林集団内での更新動態に対する空間的アプローチの一つの側面でしかない。この他にも、父性解析(種子, 実生, 稚幼樹の親木を同定する)や個体群生態学的に主体を置くなど、遺伝マーカーを利用する研究は多様化している。空間的遺伝構造は、その中で比較的古くから始まり、今なおその中心に位置する問題と言えよう。これまでも、空間的遺伝構造の解析手法については様々な研究論文の中で論じられているが、それらをデータの階層構造の観点からまとめ直した事は一定の成果をもたらすと期待したい。

今回、森林樹木集団内の空間的遺伝構造の解析に並々ならぬ興味を寄せる、異なる研究バックグラウンドを持つ二人により、その解析に用いられる統計量について考察が試みられた。本総合報告をきっかけに、天然林における空間データを伴った遺伝変異データの解析に関心を寄せる研究者が、フィールド生態学、分子遺伝学、統計学の垣根を越えて一人でも多く現れれば、著者らとしては望外の喜びである。

引用文献

- Berg, E. E. and Hamrick, J. L. (1995). Fine-scale genetic structure of a turkey oak forest, *Evolution*, **49**, 110–120.
- Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: Models and applications*, Pion, London.
- Cockerman, C. C. (1969). Variance of gene frequencies, *Evolution*, **23**, 72–84.
- Doligez, A., Baril, C. and Joly, H. I. (1998). Fine-scale spatial genetic structure with nonuniform distribution of individuals, *Genetics*, **148**, 905–919.
- Dow, B. D. and Ashley, M. V. (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*, *Molecular Ecology*, **5**, 615–627.
- Epperson, B. K. (1990). Spatial autocorrelation of genotypes under directional selection, *Genetics*, **124**, 757–771.
- Epperson, B. K. (1995a). Spatial structure of two-locus genotype under isolation by distance, *Genetics*, **140**, 365–375.
- Epperson, B. K. (1995b). Spatial distributions of genotypes under isolation by distance, *Genetics*, **140**, 1431–1440.
- Epperson, B. K. (1995c). Fine-scale spatial structure: Correlations for individual genotypes differ from those for local gene frequencies, *Evolution*, **49**, 1022–1026.
- Epperson, B. K. and Alvarez-Buylla, E. A. (1997). Limited seed dispersal and genetic structure in life stages of *Cecropia obtusifolia*, *Evolution*, **51**, 275–282.
- Furnier, G. R., Knowles, P., Clyde, M. A. and Dancik, B. P. (1987). Effects of avian seed dispersal on the genetic structure of whitebark pine populations, *Evolution*, **41**, 607–612.
- Hardy, O. J. and Vekemans, X. (1999). Isolation by distance in a continuous population: Reconciliation between spatial autocorrelation analysis and population genetics models, *Heredity*, **83**, 145–154.
- 井鷲裕司 (2001) マイクロサテライトマーカー分析法, 『森の分子生態学～遺伝子が語る森林のすがた～』(西脇亜也, 陶山佳久, 村上哲明 編) 275–304, 文一総合出版, 東京.
- Isagi, Y., Kanazashi, T., Suzuki, W., Tanaka, H. and Abe, T. (2000). Microsatellite analysis of the regeneration process of *Magnolia obovata*, *Heredity*, **84**, 143–151.
- Kimura, M. and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population, *Genetics*, **49**, 725–738.

- Konuma, A., Tsumura, Y., Lee, C.-T., Lee, S.-L. and Okuda, T. (2000). Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis, *Molecular Ecology*, **9**, 1843–1852.
- Loiselle, B. A., Sork, V. L., Nason, J. and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae), *American Journal of Botany*, **82**, 1420–1425.
- 前田禎三 (1988) ブナの更新特性と天然更新技術に関する研究, 宇都宮大学農学部学術報告特輯, **46**, 1–79.
- Mahy, G., Vekemans, X. and Jacquemart, A.-L. (1999). Patterns of allozymic variation within *Calluna vulgaris* populations at seed bank and adult stages, *Heredity*, **82**, 432–440.
- 根井正秋 (1990) 『分子進化遺伝学』, 培風館, 東京.
- Shimatani, K. (2002). Point processes for fine-scale spatial genetics and molecular ecology, *Biometrical J.*, **44**, 325–352.
- Shimatani, K. and Takahashi, M. (2003). On methods of spatial analysis for genotyped individuals, *Heredity* (in press).
- Slatkin, M. and Arter, H. E. (1991). Spatial autocorrelation methods in population genetics, *American Naturalist*, **138**, 499–517.
- Smouse, P. E. and Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure, *Heredity*, **82**, 561–573.
- Sokal, R. R. and Jacquez, G. M. (1991). Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis, *Evolution*, **45**, 152–168.
- Sokal, R. R. and Oden, D. L. (1978a). Spatial autocorrelation in biology 1. Methodology, *Biological Journal of the Linnean Society*, **10**, 199–228.
- Sokal, R. R. and Oden, D. L. (1978b). Spatial autocorrelation in biology 2. Some biological implications and four applications of evolutionary and ecological interest, *Biological Journal of the Linnean Society*, **10**, 229–249.
- Sokal, R. R. and Oden, D. L. (1991). Spatial autocorrelation analysis as an inferential tool in population genetics, *American Naturalist*, **138**, 518–521.
- Sokal, R. R. and Wartenberg, D. E. (1983). A test of spatial autocorrelation analysis using an isolation-by-distance model, *Genetics*, **105**, 219–237.
- Sokal, R. R., Jacquez, G. M. and Wooten, M. C. (1989). Spatial autocorrelation analysis of migration and selection, *Genetics*, **121**, 845–855.
- Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Glössl, J. and Kremer, A. (1998). Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Assessed with isozymes and microsatellites, *Molecular Ecology*, **7**, 317–328.
- Surles, S. E., Arnold, J., Schnabel, A., Hamrick, J. L. and Bongarten, B. C. (1990). Genetic relatedness in open-pollinated families of two leguminous tree species, *Robinia pseudoacacia* L. and *Gleditsia triacanthos* L., *Theoretical and Applied Genetics*, **80**, 49–56.
- Takahashi, M., Mukouda, M. and Koono, K. (2000). Differences in genetic structure between two Japanese beech (*Fagus crenata* Blume) stands, *Heredity*, **84**, 103–115.
- Troggio, M., DiMasso, E., Leonardi, S., Ceroni, M., Bucci, G., Piovani, P. and Menozzi, P. (1996). Inheritance of RAPD and I-SSR markers and population parameters estimation in European beech (*Fagus sylvatica* L.), *Forest Genetics*, **3**, 173–181.
- 津村義彦 (2001) アロザイム実験法, 『森の分子生態学～遺伝子が語る森林のすがた～』 (西脇亜也, 陶山佳久, 村上哲明 編) 183–219, 文一総合出版, 東京.
- Wright, S. (1943). Isolation by distance, *Genetics*, **28**, 114–138.

柳谷新一, 金豊太郎, 小西 明(1969). 低木類を刈払いしたブナ保残木作業における更新初期の成績, 林業試験場東北支場年報, 10, 124-135.

Spatial Statistics for Within-population Genetic Structure of Tree Species Using Genetic Markers

Makoto Takahashi

(Breeding Department, Forest Tree Breeding Center)

Kenichiro Shimatani

(The Institute of Statistical Mathematics)

Recent development of molecular genetic techniques have involved intensive studies of within-population spatial genetic structures for forest tree species. Genetic variation data have stratified structures of three levels; allele, genotype, and multilocus genotype. Studies of spatial genetic structures should simultaneously deal with the stratified information together with the locations of individuals. This article reviews spatial statistical methods for analyzing spatial genetic datasets and discusses their properties, problems, biological background, and the relationships among the statistics.