

— Biogeographic Component Analysis —

1984. 2. 4.

1) Biogeography, like evolutionary theory, desperately needs to have the study of pattern rigidly divorced from a priori notions of process, and to study processes underlying pattern by using pattern to test specific theories of process.

(N. Eldredge, p. 36 in 'Vicariance Biogeography: A Critique' [1981])

2) A cladistic approach to biogeography presupposes that there are areas of endemism that are interrelated among themselves. Area-relationship is to be understood at a general level, in the sense that area-relationship includes, and can be broken down into, species-relationships among the members of particular groups of organisms that naturally occur in the areas. (Nelson and Platnick [1981], p. 468)

3) If there is geological agreement among cladograms for different groups of organisms, then there is a reason to infer a common causal factor (historical explanation). (Ibid., p. 221)

4) The significance of biogeography has been merely that biogeography has raised the possibility of agreement between cladograms as considered in the geographic dimension — a possibility that has been little studied, but one worthy of further investigation. (Ibid., p. 223)

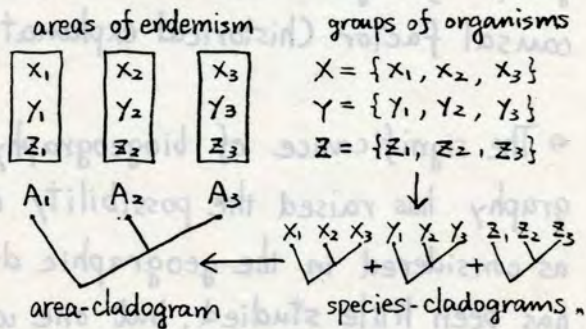
< Biogeographic Pattern >

生物のさまざまな形質の分布情報が与えられた時、その情報を最も単純に (parsimoniously) 説明する cladogram を選択する。cladistic component analysis とは、この最適な cladogram (以下では、'species-cladogram' あるいは 'component-analysis' と呼ぶ) を求めることにより、その生物群中に存在する分岐関係 — species-relationship

— を確立することである。その結果として構築される分岐構造が、
‘systematic pattern’ である。

生物地理学 (biogeography) とは、生物の地理的分布を研究する分野である。ミクロのレベルでの生物分布 (たとえば、ある地域の中での空間的・時間的分布) は、その地域の物理的環境 あるいは ‘habitat’, ‘niche’ の問題とかがかかってくるから、これはまさに生態学の領域 — ecological biogeography — に属している。それでは、マクロのレベルでの生物分布 (たとえば、アジアやアフリカの生物相) については、どのような問題が提起され得るだろうか? その一つは地域間の生物相の類似度、相互関係である。従来から用いられている「生物区」(たとえば「旧北区」「新北区」「エチオピア区」など) の概念は、生物相の類似度に基づいている。つまり、マクロの生物分布では、地域間の「類縁関係」を決めるという問題が生じ、それを探るためには生物相を担う地域そのものの「歴史」を調べなければならない。これは historical biogeography の領域に入る。それでは地域の「類縁関係」とは何なのか? また、systematics と biogeography とは、(従来はほとんど何の関連性も与えられていなかったのだが) いったいどんな関係にあるのか?

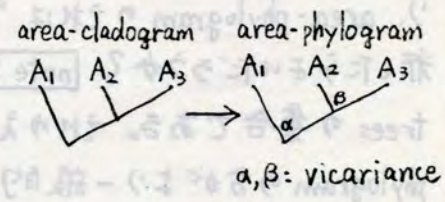
いま、3つの地域 A_1, A_2, A_3 に分布する3つの生物群 X, Y, Z があつたとし、各群に含まれる x_i, y_i, z_i は地域 A_i に「固有」(endemic) であるとする。 X, Y, Z の各群に cladistic component analysis を行なつた



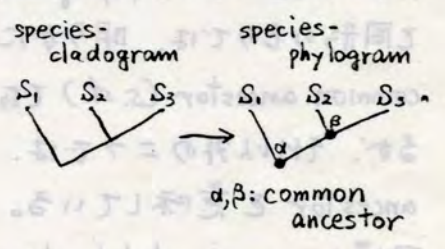
ところ、 $X_1(x_2x_3), Y_1(y_2y_3), Z_1(z_2z_3)$ という分岐構造を持つ species-cladograms が得られたと仮定する。この3つの species-cladograms は構造が一致しているので、一つにまとめて書くことができる。つまり個々の生物群ごとに species-cladogram を書くかわりに、それらが分布している「地域」(area) そのものの cladogram (‘area-cladogram’) を書くことができる。つまり、地域の類縁関係とは、そこに分布する固有種の類縁関係である、と考えられる。この area-cladogram は、そこに分布する別の生物群の species-cla

dogramをつくらって一致を確かめることにより、いくらでもテストできる。最節約原理 (parsimony) の立場からいえば、[例]の area-cladogram は与えられた3つの生物群の地理的分布情報を最も単純に説明している。こうして、生物地理学に対して cladistics が提起する問題は「地域間の類縁 (=「分岐」) 関係をいかにして探るか?」ということになる。biogeographic component analysis とは、固有種群の分布情報を最も単純に説明する最適 area-cladogram を求めることにより、地域間に存在する分岐関係 — area-relationship — を確立することである。その結果として構築される分岐構造が、「biogeographic pattern」である (冒頭の引用(2) 参照)。

species-cladograms が地理的に一致 ('geological agreement' 引用(3)) するのはどういうことだろうか? species-cladogram それ自身は、parsimony のもとで形質分布を最も単純に説明しただけのものである。また area-cladogram それ自身は、parsimony のもとで地理的分布を最も単純に説明しただけのものである。つまり、'systematics' は species-cladogram を、'biogeography' は area-cladogram を (引用(4)) をつくるだけであって、それ以上のことは何もしていない。言い換えれば、両方とも 'pattern' を確立しようとしているのであって、その 'pattern' を生みだした 'process' については何も言っていない (引用(1))。



ここで、cladogram を進化的に解釈してみよう。つまり cladogram に時間軸を付加した 'phylogram' を考える。species-phylogram の場合、与えられた形質分布を生んだ 'common factor' としては、以前述べたように、'common ancestor' を想定すればよい。つまり common ancestor がいたから、複数の形質間に「一致」が見られるのだと考えることができる。それでは area-phylogram の方は、与えられた地理的分布を (つまり、複数の species-cladograms 間の地理的「一致」を) 生んだ 'common factor' は何を意味しているのだろうか?



この 'common factor' は、何らかの方法で、地域間を「分断」したのである。
つまり、前頁の図でいえば、この「分断現象」 ('vicariance event') により、

$$\begin{aligned} \text{vicariance } \alpha &: A_1 + A_2 + A_3 \longrightarrow A_1 | A_2 + A_3 \\ \beta &: A_2 + A_3 \longrightarrow A_2 | A_3 \end{aligned}$$

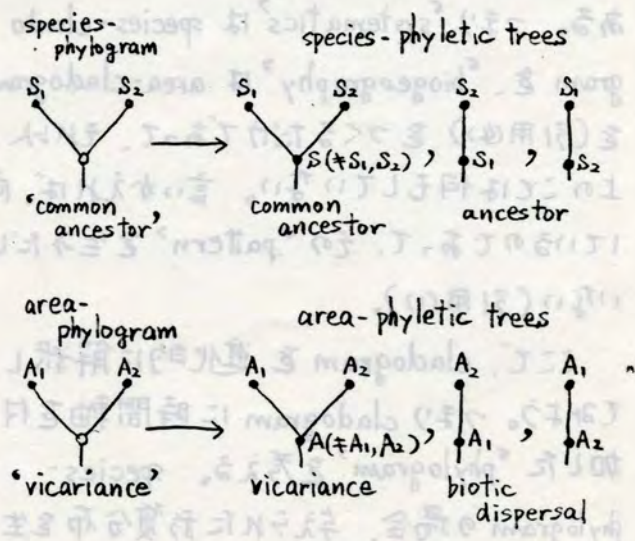
というように、地域 A_1, A_2, A_3 の生物相が分断された、と考えられる。

このように、進化的に解釈するならば、systematic 及び biogeographic 'pattern' を生んだ 'process' (= 'common factor') を考えることができ、もししたらその仮説をテストできるかもしれない。しかし、systematics 及び biogeography そのものは、そういう 'process' とは何の関係もない。systematics は複数の形質間の「一致」を示すだけであり、biogeography は複数の生物群の分布間の「一致」を示すだけである。

Appendix 1: 'common factor' の解釈について

上述のように species-phylogram の 'common factor' は 'common ancestor' であり、area-phylogram のそれは 'vicariance' であるが、この二つをどのように解釈したらよいだろうか？ note 2 で明らかにした通り、phylogram は phyletic trees

trees の集合である。それゆえ、phylogram の方がより一般的である。phyletic tree について考えてみよう。species-phyletic tree の場合、phylogram と同形のものであれば、明らかに common ancestor (s. st.) であるが、それ以外の二つでは、ancestor を意味している。



同様に、area-phyletic tree の場合、phylogram と同形のものであれば vicariance (s. st.) であるが、それ以外の二つでは、biotic dispersal を意味している。したがって、'common factor' の正しい解釈は次のとおりである。

- 'common ancestor' = common ancestor or ancestor
- 'vicariance' = vicariance or biotic dispersal

< Biogeographic Component Analysis > 1984. 2. 6.

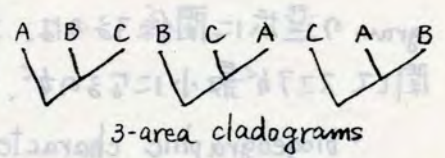
cladistic component analysis (以下'CCA'と略記) で見たように、2-taxon の場合 cladogram は一つしかないから選択の余地がない。同様に、biogeographic component analysis (以下'BCA'と略記) でも、2-area の場合は area-cladogram は一つしかないからやはり選択の余地がない:

As a result, the most basic possible unit of analysis in biogeography, as in phylogenetics, must be a three-taxon statement.

(Platnick and Nelson [1978], Syst. Zool. 27(1), p. 9)

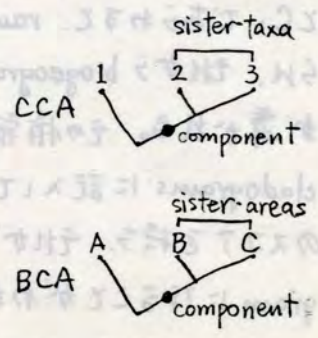
(1) 3-area case

3 areas A, B, C に対して可能な dichotomous area-cladograms は右の3つである。それぞれの area-cladogram からは、各地域の固有種に関する species-cladogram が演繹される。ここでは、最節約原理に基づいて、ある与えられた地理的分布情報を上の3つの area-cladograms のどれが最も単純に表現しているかを考えてみよう(もちろん、ここで trichotomy を含めて考えてもかまわない)。



まず問題になるのは、与えられた species-cladogram に含まれる地理的分布情報をいかにして area-cladogram に表現するかということ、及び、その表現の「単純性」をいかにして判定するか、である。

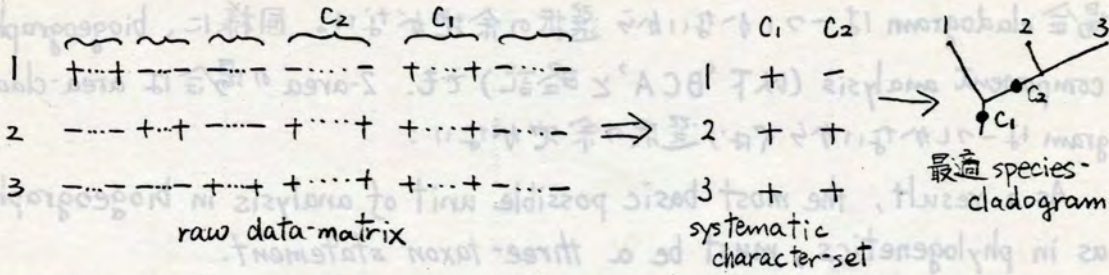
「分布情報」そのものの意味を明確にしておこう。CCA に於ては、形質の「種類」(形態学的・生化学的・遺伝的 etc) に全く関係なく、その生物群間での「分布」パターンのみが問題だった。それと同様に、BCA に於ては、生物群の「種類」(動物・植物) に全く関係なく、その地域間での「分布」パターンのみが問題になる。また「情報」の単位である 'component' が持つ意味は、species-cladogram では general synapomorphy に基づく「姉妹群」('sister taxa') の存在であったが、area-cladogram では、それに対応する species-cladogram に含まれる 'sister taxa' の分布に基づく「姉妹地域」('sister areas') の存在を意味する。



ここで、CCA 及び BCA に於て説明されるべ

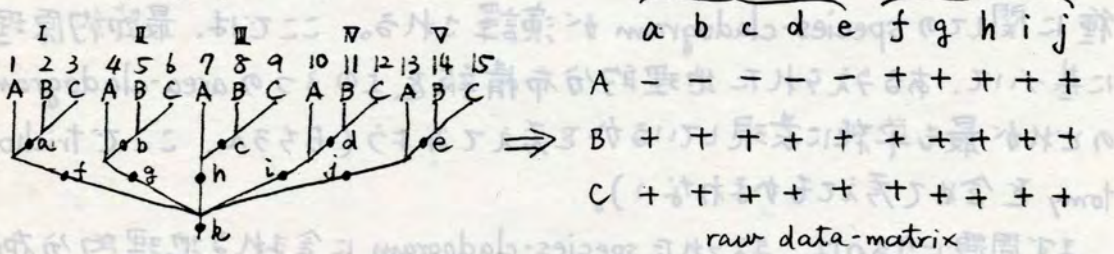
②
 ① 「形質集合」 (character-set) について具体的に考えてみよう。

° systematic character-set

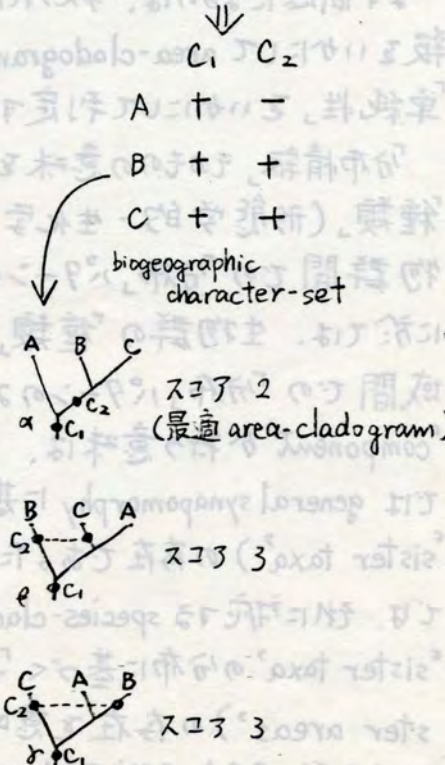


同一の分布パターンを示す多くの形質があるが、それらを「形質タイプ」でまとめると systematic character-set になる。しかし note 3 で見たとおり、このうち cladogram の選択に関係するのは、ごく少なく、上の場合 C₂ だけである。そして C₂ に関してスコアが最小になるのが、最適 species-cladogram である。1984. 2. 6.

° biogeographic character-set



いま、I~Vの5つの生物群の構成種が地域A, B, Cに固有であって、その species-cladogram が上のものであったとする。これには a~k という components が含まれているが informative components といえるのは k を除く a~j である。a~e と f~j はそれぞれ分布パターンが同じであり、そのタイプを C₂ と C₁ であらわすと、raw data-matrix が得られ、これから biogeographic character-set が導かれる。その情報を可能な3つの area-cladograms に記入してみると、α が最小のスコアを持ち、それが最適 area-cladogram になることがわかる。



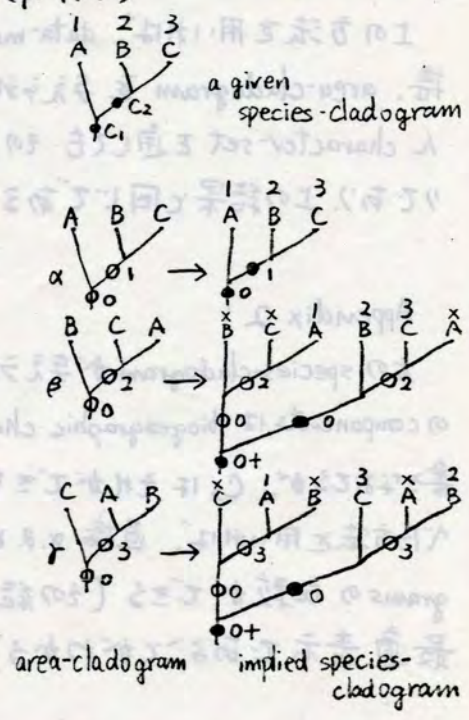
しかし、ここで一つの問題が生じる。

CCAならば、raw data-matrixを出発点として解析を進めることができるが、BCAの場合、それに先立って species-cladogram を data-matrix に変換しておかねばならない。上に与えた species-cladogram はかなり単純な形だから、すぐに data-matrix に変換できたのである。それがもう少し複雑になると、その変換は容易ではない。どうすればいいだろうか？

The problem may be analyzed under the assumption that all errors are due to missing species. (p. 417)

右図の species-cladogram が与えられているとしよう。そこには 2つの components C_1 と C_2 が含まれている。area-cladogram α は 2つの components 0 と 1 を含んでいるが、 $0 = C_1$, $1 = C_2$ と考えることにし、 α と同形の species-cladogram 上に与えられた分布情報を表現することができる。 β はやはり 2つの components 0 と 2 を含んでいるが、これを用いて data を表すには、欠損種 (missing species) を仮定する、より大きな species-cladogram をつくらなければならない。そして図に示したとおり、 $0+ = C_1$, $0 = C_2$ と考え、欠損種の位置 (x) をうまく決めると与えられた分布情報を表現できる。 γ についても、 β と同様に、欠損種を含む species-cladogram で、 $0+ = C_1$, $0 = C_2$ と考えることにし、与えられた分布情報を表現できる。

欠損種 (及び欠損 components) を仮定すれば、どんな area-cladogram でも与えられた分布情報を表現できる。これは、収斂 (convergence) と仮定すれば、どんな species-cladogram でも与えられた形質情報を表現できるのと同じである。右表からわかるように、欠損種及び欠損 components が少ない area-cladogram (この場合 α) ほど



	area-cladograms		
missing species	α	β	γ
A	0	1	1
B	0	1	1
C	0	1	1
missing components			
0+	-	0	0
0	0	1	1
1	0	-	-
2	-	2	-
3	-	-	2
total	0	6	6

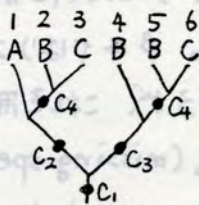
より単純な表現である。なぜなら欠損種 (欠損 components) という余分な仮定 (ad hoc statements) が少ないのでテストしやすいからである。このように、ad hoc な欠損種の仮定を最少にとどめて、最も単純な area-cladogram を選ぶという方法は、最節約原理の適用にはかならない。これは、ちょうど ad hoc な収斂の仮定を最少にとどめて、最も単純な species-cladogram を選ぶという CCA における最節約原理の適用と対応している。

上の方法を用いれば、data-matrix 及び character-set を通してではなく、直接、area-cladogram を与えられた species-cladogram と比較できる (もちろん character-set を通してもその比較はできるわけで、その結果は ④ の右下の通りであり、上の結果と同じである)。

1984. 2. 6

Appendix 2 :

左の species-cladogram が与えられた時、 C_1, C_2, C_4 の components は biogeographic character-set の形式に書きなおせるが、 C_3 はそれができない。しかし、上で述べた方法を用いれば、直接 α, β, γ の 3 area-cladograms の比較ができる (その結果、 α -pattern が最適表示であることがわかる)。

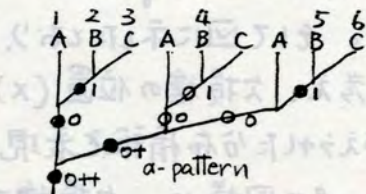


以上の議論をもとに、CCA と BCA の分析手順を次のようにまとめることができる。

CCA:
possible species-cladograms
{1, 2, ..., n}

表示
↑
datamatrix
(character-set)

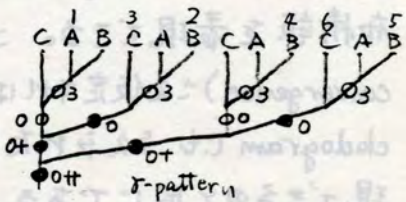
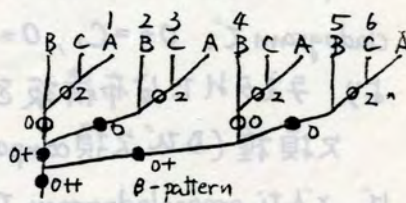
optimal species-cladogram
{i}



BCA:
possible area-cladograms
{1, 2, ..., n}

↓
implied species-cladograms
{1, 2, ..., n}

↑
optimal implied species-cladogram
{i}

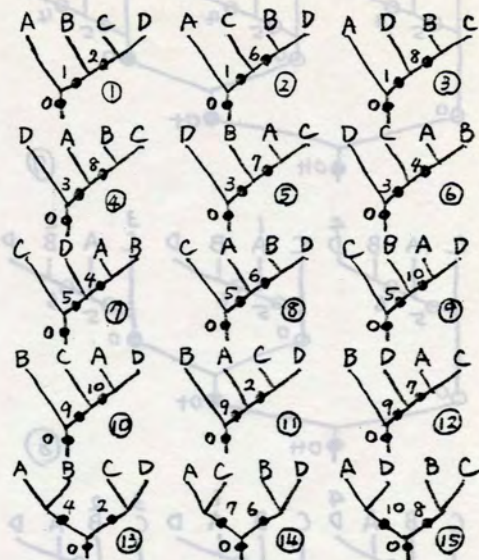


1984. 2. 8.

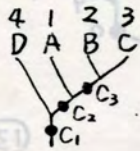
(2) 4-area case

4 areas A, B, C, D に対して可能な dichotomous area-cladograms は右の 15 個である。いま、この 4 地域の固有種に関する species-cladogram が与えられているとする。目的は、3-area の場合と同じく、その species-cladogram に含まれている地理的分布情報を、最も単純に表示する area-cladogram を右の 15 個から選択することである。

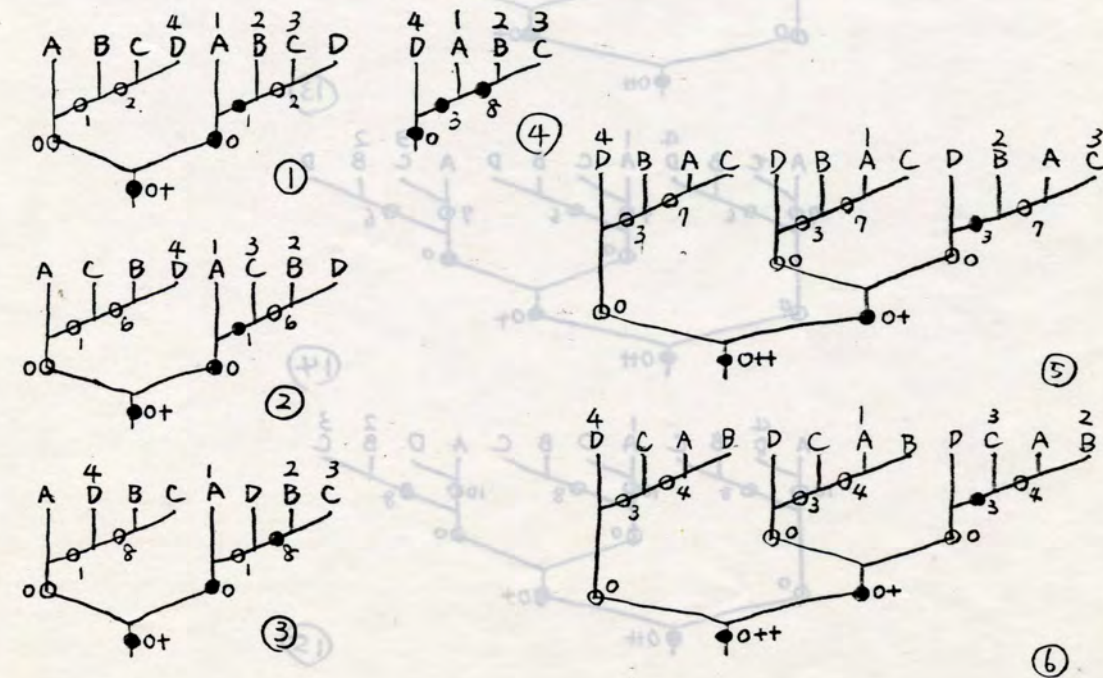
1984. 2. 12.



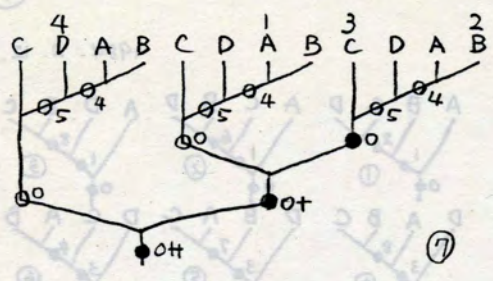
4-area cladograms



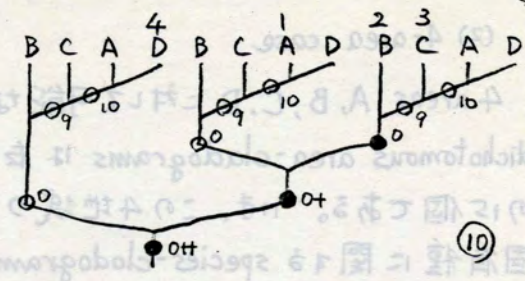
いま A, B, C, D に固有の 4 種 1, 2, 3, 4 に対して、左図に示した species-cladogram が得られたとする。それは、各々の area-cladogram から導かれた implied species-cladogram を用いて、次のように表示される。



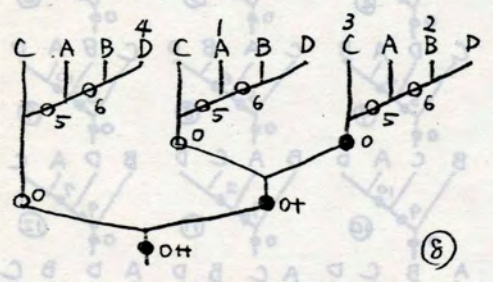
10



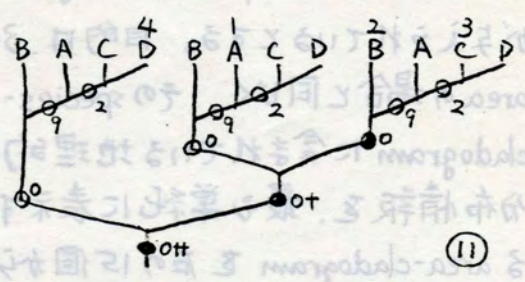
7



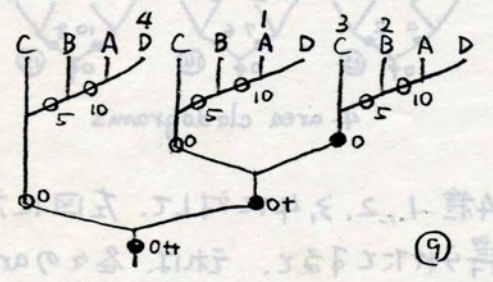
10



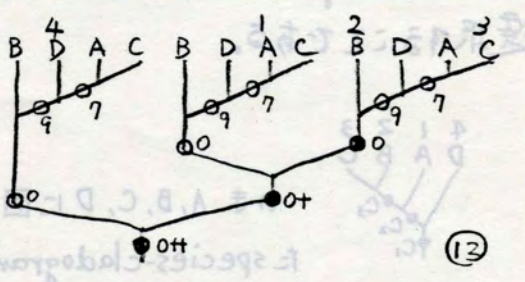
8



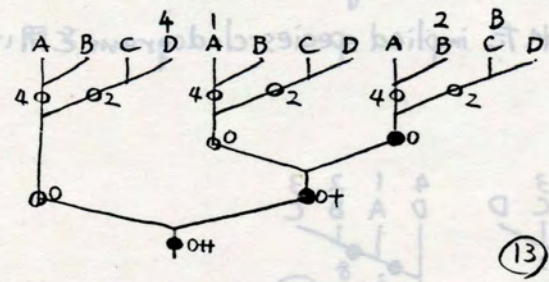
11



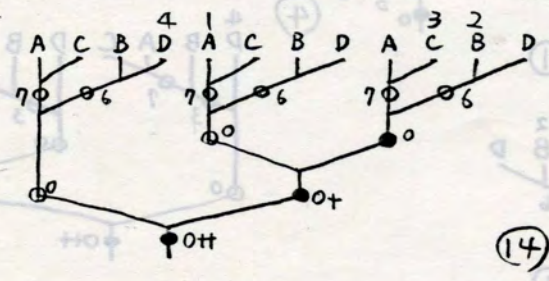
9



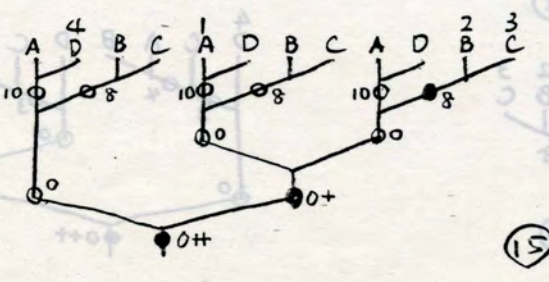
12



13



14



15

以上の結果を表にすると次のようになる。

area-cladogram	'complexity' of implied species-cladogram	missing species	missing components
1	2	4	4
2	2	4	4
3	2	4	4
4 (Optimal)	1	0	0
5	3	8	8
6	3	8	8
7	3	8	8
8	3	8	8
9	3	8	8
10	3	8	8
11	3	8	8
12	3	8	8
13	3	8	8
14	3	8	8
15	3	8	8

この表から、area-cladogram 4 が最適表現であることがわかる（また、その area-cladogram は与えられた species-cladogram と「同形」である）。上表において 'complexity' とは、その implied species-cladogram に含まれている 'area-cladogram' の「個数」である。data として与えられた species-cladogram の 'species' および 'components' の「個数」をそれぞれ x , y (~~complete dichotomy の場合は、 $y = x - 1$; non dichotomy があれば $y < x - 1$~~) とし、implied species-cladogram の 'complexity' を M とすると、

$$[\text{no. of missing species}] = NM - x$$

$$[\text{no. of missing components}] = (N-1)m + x - y$$

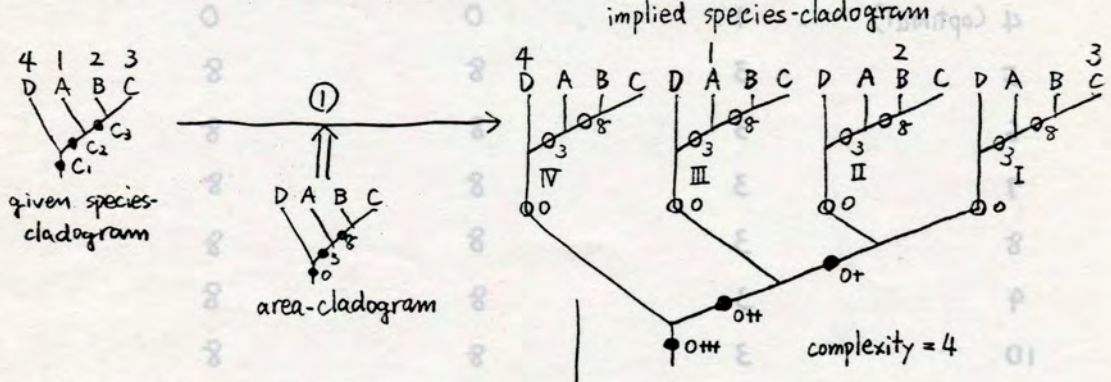
(式の導出は Appendix 3 参照)

となり、上の第 2 及び 第 3 列の数値は 第 1 列の 'complexity' に完全に

依存する。そこで最適 area-cladogram の選択の際は、implied species-cladogram の 'complexity' のみを基準とし、それが最小のものを選ぶ
はよい。 1984. 2. 12.

'implied species-cladogram' はどのようにして求められるのか？

data として与えられるあらゆる 717 の species-cladogram を、area-cladogram から派生する implied species-cladogram を用いて表わすため、次の操作をする。

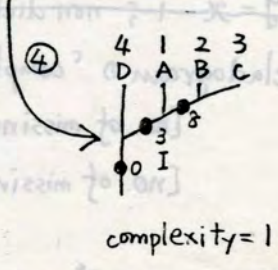
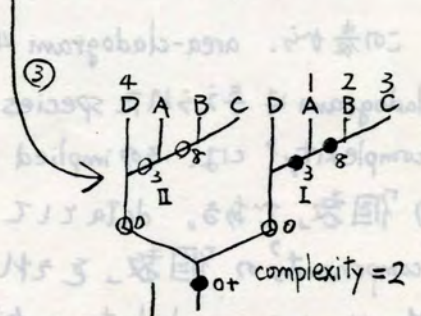
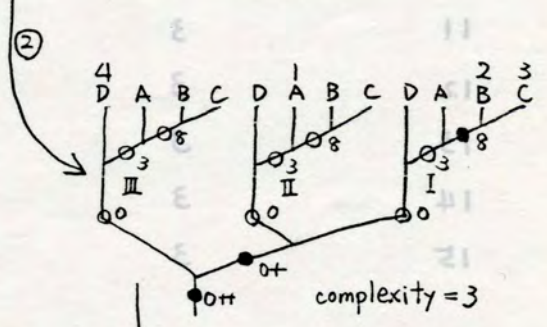


1) まず、与えられた species-cladogram の各 species の部分を ある area-cladogram で置換する。その結果 complexity = 4 の implied species-cladogram ができる。そこには、4つの area cladograms (I~IV) が含まれており、各々は 1つの species を担っている。

2) species の移項 : 2_B (II → I) ; 1_A (III → II) ; 4_D (IV → III)。この結果、IV は不用となり削除される。complexity = 3 となる。

3) species の移項 : 1_A (II → I) ; 4_D (III → II)。この結果 III は削除される。complexity = 2 となる。

4) species の移項 : 4_D (II → I)。この結果 II は削除される。complexity = 1 となる。



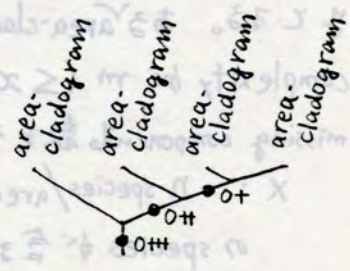
以上の操作からわかるように、すべての area-cladogram について、出発点は等しく complexity = 4 の implied species-cladogram である。重要なことは、種的位置を変更することにより、complexity をどこまで下げられるかである。

④頁の番号に従えば、次のようになる。

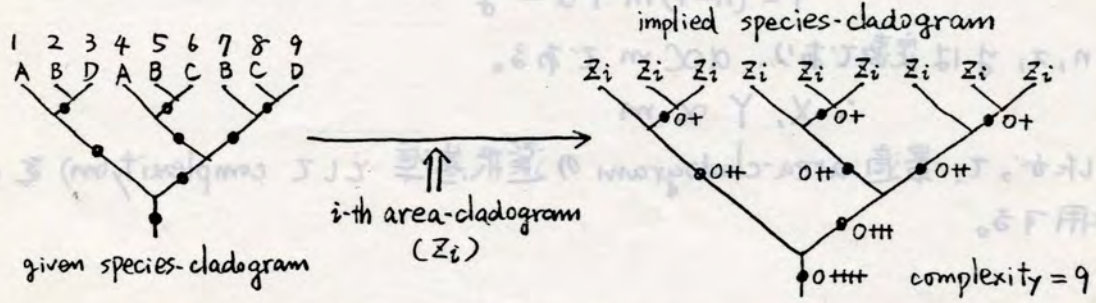
- area-cladogram ①~③ : complexity 4 → 2
- ④ : 4 → 1 (optimal)
- ⑤~⑮ : 4 → 3

つまり、各々の area-cladogram に対する complexity を最小にしておいて、その上で area-cladogram 間の比較をして、最小の complexity を持つ area-cladogram を選択することになる。上の場合、それは ④ である。

[complexity 4 の implied species-cladogram は、上のデータに対しては、右の形となる。そこに表現されている分岐構造は、任意の area-cladogram が与えられたデータを表示する上で必要かつ十分であることを注意しよう。つまり、右の分岐構造を含んでいなければデータを表現できないし、同時にそれを含んでいるより複雑な cladogram (complexity ≥ 5) は、不必要である。]



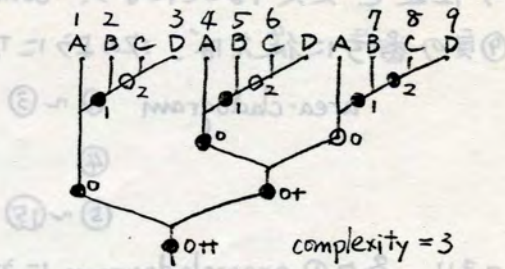
上例は、あまりに単純であって、最適 area-cladogram は見ただけでもわかってしまうから、次に少し複雑な例をあげる。



4 areas A~D に固有種 1~9 が分布していて、species-cladogram が図のようであったとする。出発点の implied species-cladogram は、i 番目の areadcladogram (Zi) を置換したもので、complexity = 9 となる。それぞれの Zi に対し、この complexity はどこまで減らせるだろうか？ その操作の手順をすべて書き下してもいいのだが、その気もないので、結果のみを書くと次のようになる。

area-cladogram	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉	Z ₁₀	Z ₁₁	Z ₁₂	Z ₁₃	Z ₁₄	Z ₁₅
minimal complexity	3	4	4	5	6	6	6	5	6	6	6	6	5	5	5

complexity が最小 (3) である。最適 area-cladogram Z₁ を用いて、与えられた分布情報を表現すると、右図のようになる。



Appendix 3: 'complexity'

与えられた species-cladogram の species 数, components 数を, それぞれ x , y とする。ある area-cladogram に導かれる implied species-cladogram の complexity が m ($\leq x$) であるとすると。このとき, missing species 数と missing components 数をそれぞれ X , Y とすると。

X : n species/area-cladogram だから, complexity = m のとき, nm 個の species が含まれる。与えられた x species あるわけだから、

$$X = nm - x$$

Y : $(n-1)$ components/area-cladogram だから, complexity = m ならば $(n-1)m$ 個の components がある。さらに, $0+$, $0+$, ... 等の components の個数を α ($\leq y$) とする。与えられた y components あるから、

$$Y = (n-1)m + \alpha - y$$

n, x, y は定数であり, $\alpha < m$ である。

$$\therefore X, Y \propto m$$

したがって、最適 area-cladogram の選択基準として complexity (m) を採用する。

以上論じてきたのは、きわめて単純な空間分布についてだけだった。1つあるいは複数の species がいくつかの areas に拡散している場合、つまり 'widespread species' を含むより複雑な場合を、note 15 で扱う。

1984. 2. 13.