

— Cladistic Analysis and Synthesis —

1984. 5. 22.

分類学の領域では、ある生物群について観察された形質の分布情報がデータとして与えられた時、それを最も単純に説明する species-cladogram を選択した ('cladistic component analysis')。一方、生物地理学の領域では、一連の地域において発見された固有種の分布情報を最も単純に説明する area-cladogram を選択した ('biogeographic component analysis')。つまり、与えられた分布情報を最も単純に説明する cladogram を選ぶことが、分岐解析 (cladistic analysis) の目的である。その結果得られたものを基本 cladogram (fundamental cladogram) と呼ぶ。

note 5 で述べたように、cladogram の持つ情報は component information (C), term information (T) の2つに分解できる。

k個の要素をもつある component について、

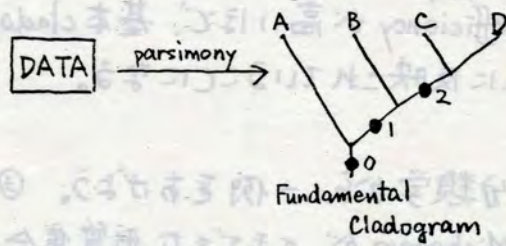
$$C = 1, T = k - 1$$

と定義される。C+T は total information (U) と呼ばれる。cladogram

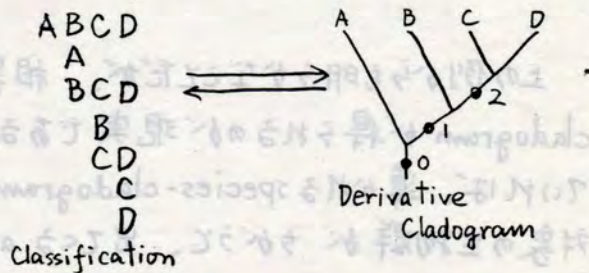
全体の情報量は、含まれている components の情報量の総和である。

分類学者は、その基本 cladogram をもとにして、階層分類を作るわけだが、その分類は cladogram の形で表現できる。分類システムから導かれた cladogram を、派生 cladogram (derivative cladogram) という。派生 cladogram についてもやはり、component, term, total information が定義できる。

cladist であるならば、基本 cladogram をそのまま分類システムにしてしまうわけだから、基本、派生の両 cladograms は完全に一致する。けれども、全ての分類学者が cladists とはいえない。数量分類学者もいれば、進化分類学者もいる。



Component	0	1	2	total
Information				
· Component	1	1	1	3
· Term	3	2	1	6
· Total	4	3	2	9



学派が違えば、基本cladogramの作り方(つまりclustering method)も異なり、基本cladogramをどの程度分類システムに反映させるかも異なってくる。ここでは第二の点に注目しよう。基本cladogramと派生cladogram(分類システム)との差の尺度として'efficiency'の概念を導入する。

基本cladogramの情報: C_f, T_f, U_f

派生cladogramの情報: C_d, T_d, U_d

< 'Efficiency' E >

component: $E_c = \frac{C_d}{C_f}$; term: $E_T = \frac{T_d}{T_f}$; total: $E_U = \frac{U_d}{U_f}$

効率 efficiencyが高いほど、基本cladogramの情報がより忠実に分類システムに反映されていることになる。

分類学から、一例をあげよう。③ Fig. 4.29 は、ハケの研究者である、C. D. Michener が、さまざまな形質集合に対して、cladistics あるいは phenetics を用いた結果得られた基本cladogram である。また④ Fig. 4.30 は、同じ Michener がそれらに基づいて与えた分類システムから導かれた派生cladogram である。これらの cladograms が含む情報、派生cladogramの効率等は④ Tab. 4.1 にまとめられている。効率の低さからわかるように、Michener は、自分が得た基本cladogramsとはかなり異なる分類システムを主張しているわけである。(特に、Fig. 4.30. 1 に含まれる components 19, 20 は、基本cladogramのどれにも含まれていない。Michenerはどこからこの結果を得たのか?)

上の例からも明らかだが、相異なるデータに対しては、異なる基本cladogram が得られるのが現実である。分類学では、扱う形質がちがうとすれば、導かれる species-cladogram もちがってくる。生物地理学では、対象の生物群がちがうと、出てくる area-cladogram もちがう。そのとき、異なる基本cladograms がもつ共通の components をとり出して、それから成るcladogram を作る事ができよう。このように、分岐解析の結果得られたいくつかの基本cladograms を、ある一つの一般cladogram (general cladogram) としてまとめることが、分岐統合 (cladistic synthesis) の目的といえる。

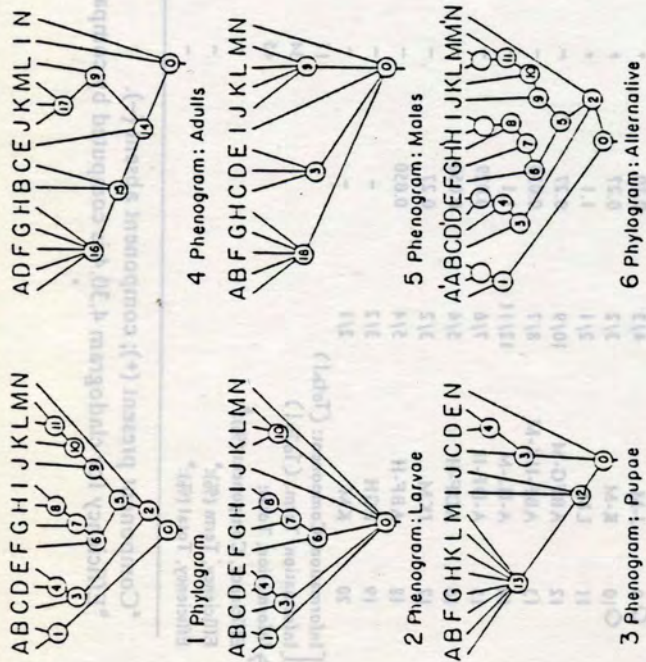


Figure 4.29. Cladograms based on information from Michener's (1977, figures 1, 7-10) study of allodapine bees (cladograms 1-5) and an alternative phylogram (cladogram 6).

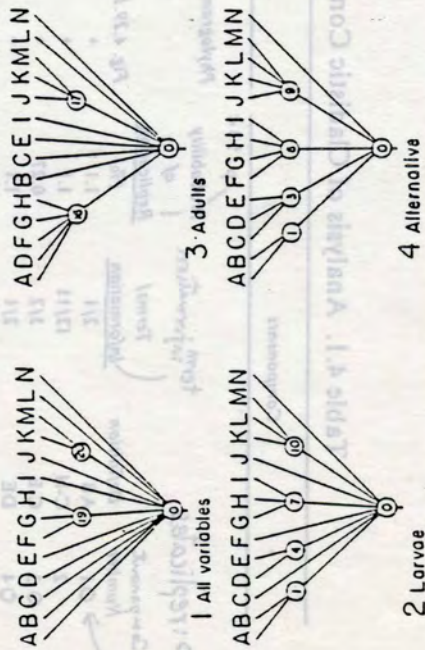


Figure 4.30. Cladograms based on classifications from Michener's (1977:51-52) study of allodapine bees (cladograms 1-3) and an alternative classification (cladogram 4).

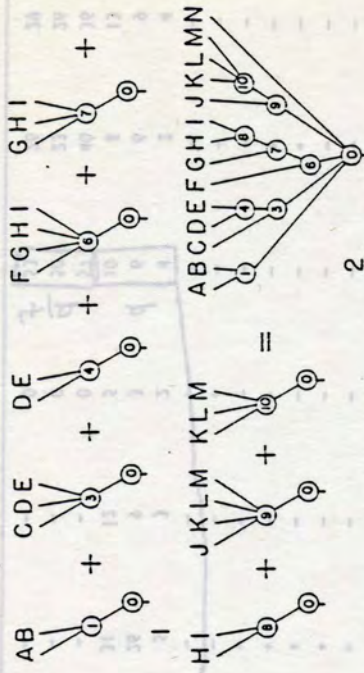


Figure 4.31. Replicated components (1) and their resulting cladogram (2).

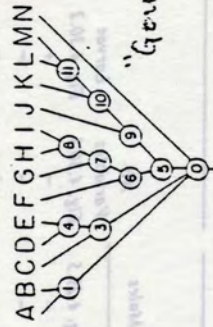


Figure 4.32. The general cladogram derived from the fundamental cladograms of figure 4.29.

Table 4.1. Analysis of Cladistic Components of Cladograms, and of Their Information and Efficiency^a

Component	Fundamental Cladograms					Derivative Cladograms				
	Phylogram	Larvae	Pupae	Adults	Males	All Variables	Larvae	Adults	Alternative	
	Fig. 4.29.1	Fig. 4.29.2	Fig. 4.29.3	Fig. 4.29.4	Fig. 4.29.5	Fig. 4.30.1	Fig. 4.30.2	Fig. 4.30.3	Fig. 4.30.4	
Component Number	1	2	3	4	5	6	7	8	9	
Definition	AB	C-N	C-E	DE	F-M	F-I	G	H	I	
Term/Information	2/1	12/11	3/2	2/1	8/7	4/3	3/2	2/1	1/1	
Probability of Replication (%)	1.1	1.1	0.27	1.1	0.033	0.10	0.27	1.1	0.10	
Efficiency, Component (%) ^b	11	34	45	11	11	11	11	11	11	
Efficiency, Term (%) ^b	11	34	45	11	11	11	11	11	11	
Efficiency, Total (%) ^c	11	34	45	11	11	11	11	11	11	

^aComponent present (+); component absent (-).

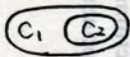
^bEfficiency for cladogram 4.30.4 is computed by comparison with cladogram 4.29.1.

複数の cladograms の一致性は、ふつう taxonomic congruence または taxonomic consensus と呼ばれる。それに基づいてある一般 cladogram を導くには、まず components 間の相互関係を明らかにしなければならない。(component = cluster ということも再度確認しておく)

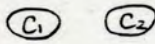
< component relations >

・合併可能：一つの cladogram に同時に含まれる
(combinability)

・包含 (inclusion)



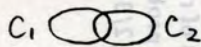
・排除 (exclusion)



・同一 (replication)



・合併不能：一つの cladogram に同時に含まれない。
(noncombinability)



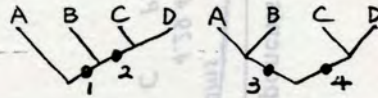
例えば、右の2つの cladograms について、



1, 2 } 包含
1, 4 }



2, 3 } 排除
3, 4 }



2, 4: 同一,

1, 3: 合併不能

replication の関係は特に重要である。なぜなら、偶然性のみによって、それが生じる確率はさわめて小さいからである。n個の要素の中からk個をとって作る component の場合の数は nC_k だから、ある component が、偶然に再び生じる確率は $1/nC_k$ であり、非常に小さい値となる (④ Tab. 4.1 の 'Probability of Replication' の項)。

上述の Michener の基本 cladograms について、どの component が同一であるかをまとめたのが ⑥ Tab. 4.2 である。確率のくわしい計算方法は「原典版」(~) で論じているのでここでは省略する。その表の最終結果である 'P cumulative' の数値からわかるように、基本 cladograms に生じた同一の components が偶然に生じたのだと結論することは、もはやできない。

1984. 5. 22.

この同一関係にある components ('replicates') に着目すると、③ Fig. 4. 29 に生じた 18 components は次の3カテゴリーに分けられる。

I Replicates (1, 3, 4, 6, 7, 8, 9, 10)

Table 4.2. Probabilities (P) of Replication of Components (C) in Fundamental Cladograms*

Number of Terms (i)	Number of Possible Components (μCi)	Fundamental Cladograms					P for components of same number of terms, %			P cumulative %
		4.29.1	4.29.2	4.29.3	4.29.4	4.29.5	C	P: %	C	
2	91	1 4 8 11	4.40 4.55 4.44	4 4 4	4.40					3.90 × 10 ⁻⁴
3	364	3 7 10	3.75 4.08 3.85	3 7 10	1.29	3	1.10	3	8.37 × 10 ⁻⁷	3.27 × 10 ⁻¹²
4	1001	6 9	16.7	6		6	0.59	9	5.73 × 10 ⁻⁴	1.87 × 10 ⁻¹⁷
5	2002									10 ⁻¹⁷
7	3432									10 ⁻¹⁷
8	3003	5								10 ⁻¹⁷
10	1001									10 ⁻¹⁷
12	91	2								10 ⁻¹⁷

*For a component that replicates a previously resolved component, the probability is that of replication due to chance alone, under the assumption that the components are resolved one at a time in the order listed for each cladogram. The probability fractions for replicates are: cladogram 4.29.2 (4/91, 3/66, 2/45, 3/80, 2/49, 1/26, 2/12); 4.29.3 (4/91, 3/232); 4.29.4 (2/341); 4.29.5 (4/364, 2/341). A component that is resolved for the first time has 0 probability of replicating a previously resolved component.

II) replicates と合併可能な components (2, 5, 11, 12)

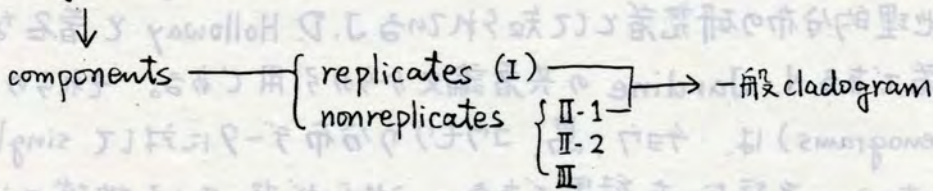
II-1) 互いに合併可能なもの (5, 11)

II-2) 互いに合併不可能なもの (2, 12)

III) replicates と合併不可能な components (13, 14, 15, 16, 17, 18)

上の3カテゴリの中から replicates のみを取りだして cladogram を作り、③の Fig. 4.31 のようになる。上述したように replicates が偶然に生じる確率は小さいわけだから、Fig. 4.31 は一般 cladogram の骨格をあらわしている。次いで、この replicates と合併可能であり、同時に互いに合併可能な components (5, 11) も Fig. 4.31 に記入できる。その結果得られるのが ③ Fig. 4.32 であり、それが「一般 cladogram」と呼ばれるものである。基本 cladograms から一般 cladogram に至る過程を図示すれば、

基本 cladograms : 1, 2, ..., n



いったん、一般 cladogram ができてしまえば、今度は逆にそれを基準として基本 (or 派生) cladograms を比較することが出来る。(④ Tab. 4.2 では基本 cladogram を基準としていたことに注意) 一般 cladogram に含まれる components (I+II-1) を 'true' と呼ぶ、II を 'ambiguous', III を 'false' (一般 cladogram から見た時の命名) ということにする。このとき、基本 (or 派生) cladograms の情報も、それによって True (T), Ambiguous (A), False (F) の3カテゴリに分かれる。

C_T, C_A, C_F : component information

T_T, T_A, T_F : term information

U_T, U_A, U_F : total information

ある cladogram が含む情報は、一般 cladogram を基準とすると、例えば次式で計算される: 'True' - 'False' + $\frac{1}{2}$ 'Ambiguous'。この情報を一般情報 (general information) と呼ぶ。とすると、

general component information : $C^* = C_T - C_F + \frac{1}{2} C_A$

general term information : $T^* = T_T - T_F + \frac{1}{2} T_A$

general total information : $U^* = U_T - U_F + \frac{1}{2} U_A$

また、一般 cladogram の情報が、基本 (or 派生) cladograms にどの程度反映されているかを測る尺度として、「一般効率」(general efficiency) が定義できる。

C, T, U : 一般 cladogram の情報

general component efficiency : $E_C^* = C^*/C$

general term efficiency : $E_T^* = T^*/T$

general total efficiency : $E_U^* = U^*/U$

Michener の cladograms (③ Fig. 4.29~30) に対する、③ Fig. 4.32 の一般 cladogram を基準とする解析の結果は ④ Tab. 4.3 にまとめられている。

次に、生物地理学から一ツ例をあげよう。⑩ Fig. 8.21, 23, 24 は、4ヨウの地理的分布の研究者として知られている J. D. Holloway と著名な数量分類学者である N. Jardine の共著論文からの引用である。これらの cladograms (phenograms) は、4ヨウ、鳥、コウモリの分布データに対して single-linkage clustering を行なった結果である。これらが扱っている地域のうち、3 cladograms に共通するものを Regions 1~9、コウモリの cladogram (⑩ Fig. 8.24) において widespread となっているものを Regions 10~18 として以下のように解析した。

1. Regions 1~9

それらに対する area-cladograms (基本 cladograms) は、⑪ Fig. 8.25 ~27 に示されている。それらに含まれている components 及び replication の割合を調べると、⑫ Tab. 8.2 のようになる。

カテゴリ-I : 2, 3, 4, 5, 6, 7

カテゴリ-II-1 : なし

一般 area-cladogram は ⑪ Fig. 8.28 である。

2. Regions 10~18

area-cladograms は、⑭ Fig. 8.29~31 である。コウモリに関する cladogram (Fig. 8.31) は widespread taxa を多く含んでいるが、「仮定 1」(正確には 仮定 2 — cf. note 1b) のもとで解析した。components と replication の解析結果は、⑮ Tab. 8.3 にまとめられている。

Table 4.3. Analysis of General Information and General Efficiency of Cladograms

Cladograms	Component Information				Term Information				Total C + T True	Efficiency (%)				
	Total Units		Units That Are:		Total Units		Units That Are:			Component	Term	Total		
	True	False	True	Ambiguous	True	False	Ambiguous							
Fundamental Cladogram	4.29.1	11	10	0	1	10.5	34	23	0	11	39	105	124	118
	4.29.2	7	7	0	0	7	12	12	0	0	19	70	52	58
	4.29.3	4	2	1	1	1.5	19	3	7	9	2	15	2	6
	4.29.4	5	1	4	0	-3	26	3	23	0	-23	-30	-87	-70
	4.29.5	3	2	1	0	1	9	5	4	0	2	10	4	6
Derivative Cladogram	4.30.1	2	0	2	0	-2	3	0	3	0	-5	-20	-13	-15
	4.30.2	4	4	0	0	4	6	6	0	0	10	40	26	30
	4.30.3	2	0	2	0	-2	6	0	6	0	-8	-20	-26	-24
	4.30.4	4	4	0	0	4	9	9	0	0	13	40	39	39
General	10	10	0	0	10	23	23	0	0	23	33	100	100	100

Cladogram

B. Horsfly

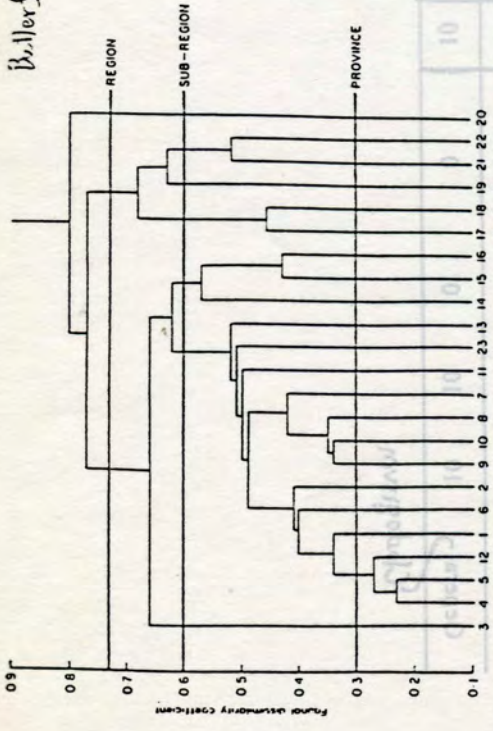


Figure 8.21. Cladogram for 22 areas (with Moluccas united), derived by single-link cluster analysis from coefficients of faunal similarity, based on distributions of 870 species of butterflies (Rhopalocera). After Holloway and Jardine (1968), figure 1, p. 161.

Bird

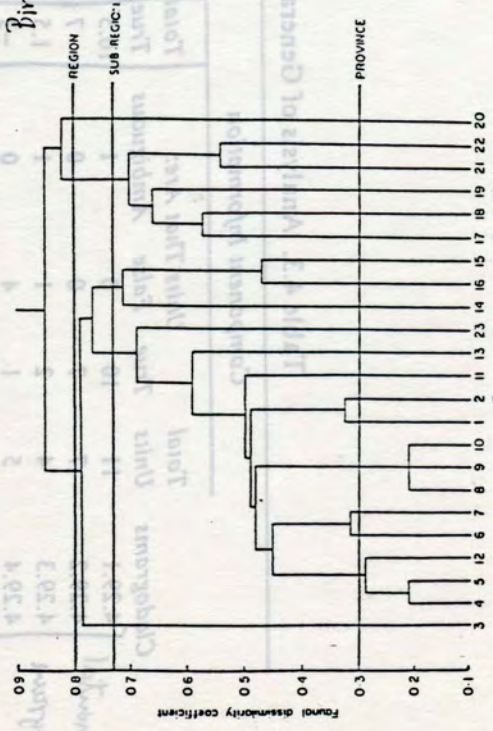


Figure 8.23. Cladogram for 22 areas (with Moluccas united) derived by single-link cluster analysis from coefficients of faunal similarity, based on distributions of 876 species of birds. After Holloway and Jardine (1968), figure 2, p. 162.

Bat

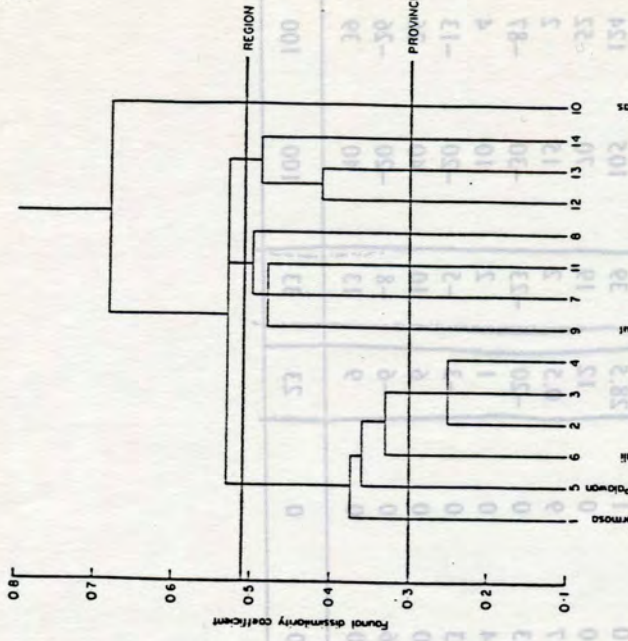


Figure 8.24. Cladogram for 14 areas, derived by single-link cluster analysis from coefficients of faunal similarity, based on distributions of 124 species of bats. After Holloway and Jardine (1968), figure 3, p. 163.

- Regions 1 ~ 9 (o) : 1. Australia ; 2. Moluccas ; 3. Solomons ;
 4. New Guinea ; 5. Celebes ; 6. Philippines ; 7. Indo-China ;
 8. Sumatra ; 9. Malaya
 Regions 10 ~ 18 (x) : 10. Formosa ; 11. Palawan ; 12. Timor ;
 13. Borneo ; 14. Java ; 15. Bali ; 16. Lombok ;
 17. Flores ; 18. Sumbawa

Butterfly

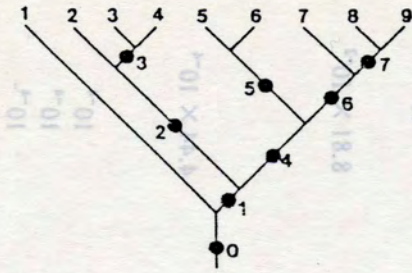


Figure 8.25. Simplified area-cladogram based on butterflies (cf. figure 8.21).

Bird

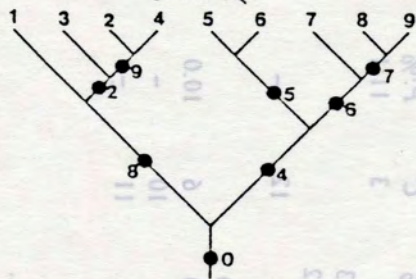


Figure 8.26. Simplified area-cladogram based on birds (cf. figure 8.23).

Bat

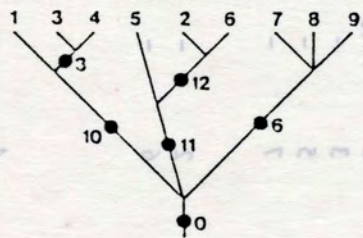
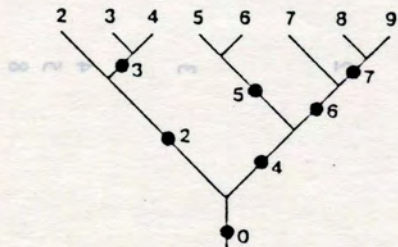


Figure 8.27. Simplified area-cladogram based on bats (cf. figure 8.24).

Regions 1-9



General Area-Cladogram (Butterfly + Bird + Bat.)

Figure 8.28. General area-cladogram (cf. figures 8.25-8.27).

Regions 1-9

Table 8.2. Probabilities (P) of Replication of Components (C) in Area-Cladograms

Number of Terms	Number of Possible Components	Cladograms			P for Components of Same Number of Terms, %			P. Cumulative %
		C	P:%	8.27	C	P:%	8.81 × 10 ⁻²	
2	36	3	8.33	3	11.1	8.81 × 10 ⁻²	8.81 × 10 ⁻²	
3	84	5 7	9.52	12	-	8.81 × 10 ⁻²	8.81 × 10 ⁻²	
4	126	2 6	20.0	6	10.0	0.50	4.44 × 10 ⁻⁴	
5	126	8	25.0	10	-	-	10 ⁻⁴	
8	9	4 1	100.	11	-	-	10 ⁻⁴ 10 ⁻⁴ 10 ⁻⁴	

NOTE. For a component that replicates a previously resolved component, the probability is that of replication due to chance alone, under the assumption that the components are resolved one at a time in the order listed for each cladogram. The probability fractions for replicates are: cladogram 8.26 (3/36, 2/21, 2/10, 1/4, 1/1), cladogram 8.27 (4/36, 2/20). (cf. Table 4.2)

(4)

カテゴリ-I : 13, 14, 15, 18

カテゴリ-II-1 : 19

一般 area-cladogram は (14) Fig. 8.31 となる。

∴ Regions 1 ~ 18

area-cladograms は (16) Fig. 8.33 ~ 35 である。含まれている components および replication は (17) Tab. 8.4 に示されている。

カテゴリ-I : 2, 3, 5, 7, 18, 25, 26, 27, 28, 29, 32,

カテゴリ-II-1: 19, 33

一般 area-cladogram は (16) Fig. 8.36 となる。

上の二つの例では、clustering method そのものは必ずしも cladistics ではなかった。この clustering 法が最良であるかは、それだけで一つの大問題である。しかし、この note での主眼はむしろ、複数の相異なる cladograms があるとき、それらからどのようにして consensus (つまり一般 cladogram) を得るのか、ということに置かれている。複数の情報から、ある一つの結論を導くための方法は、clustering 法の議論とはまたちがった意味で、重要な問題なのである。

一つの大きな疑問が残る。上述の方法で得られた分岐構造 (ある一般 cladogram) は、自然界に存在するある pattern を現わしているのだろうか？ それとも、方法論自身がでっちあげた単なる虚構にすぎないのだろうか？ さらにまた、その分岐構造は、何らかの方法で「仮説検定」され得るものなのか？ こういった諸問題は、分類学や生物地理学の根幹にかかわるものだけに、深く考えてみる必要がある。

1984. 5. 23.

Regions 10-18



General Area-Cladogram (Buckley + Bird + Bor)

Figure 8.35 General area-cladogram (cf. figures 8.29-8.31)

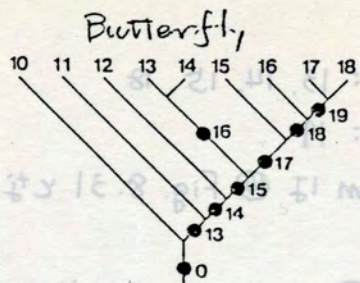


Figure 8.29. Simplified area-cladogram based on butterflies (cf. figure 8.21).



Figure 8.30. Simplified area-cladogram based on birds (cf. figure 8.23).

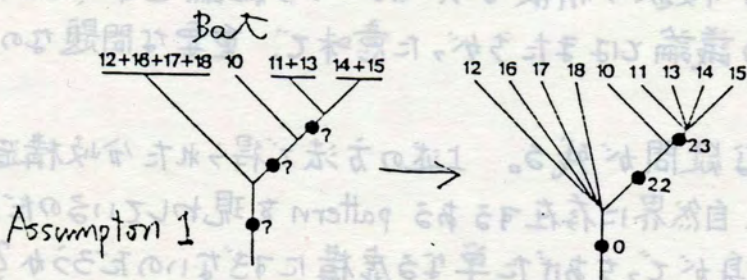
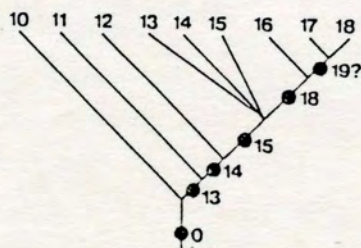


Figure 8.31. Simplified area-cladogram based on bats (cf. figure 8.24).

General Area-Cladogram (Butterfly + Bird + Bat)



Regions 10-18

Figure 8.32. General area-cladogram (cf. figures 8.29-8.31).

Regions 10-18

Table 8.3. Probabilities (P) of Replication of Components (C) in Area-Cladograms

Number of Terms	Number of Possible Components	Cladograms		P for Components of Same Number of Terms, %		P Cumulative %
		C	P: %	C	P: %	
2	36	16	-	8.31	-	2.86
3	84	19	-	23	2.86	2.86
4	126	18	-	22	2.86	2.86
5	126	17	-	22	2.86	2.86
6	84	15	-	22	2.86	2.86
7	36	14	-	22	2.86	2.86
8	9	13	-	22	2.86	2.86

NOTE. For a component that replicates a previously resolved component, the probability is that of replication due to chance alone, under the assumption that the components are resolved one at a time in the order listed for each cladogram. The probability fractions for replicates are: cladogram 8.30 (1/35, 1/3, 1/3, 1/2). (cf. Table 4.2)

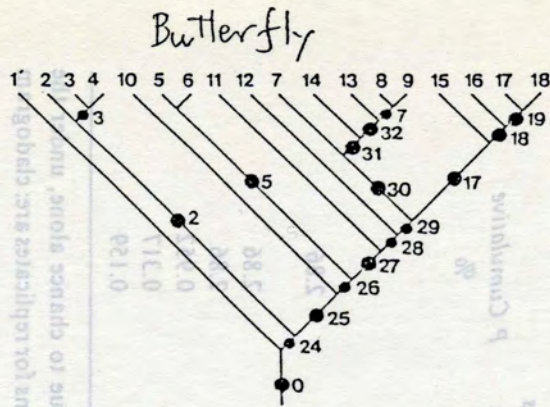


Figure 8.33. Area-cladogram based on butterflies (cf. figure 8.21).

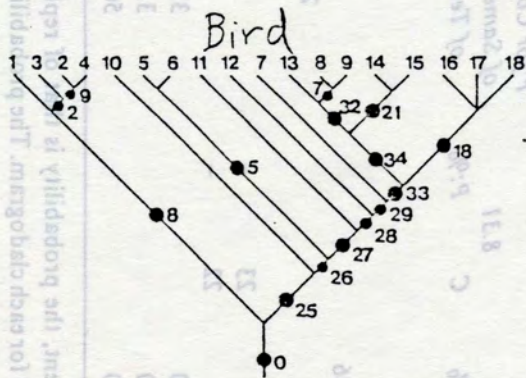


Figure 8.34. Area-cladogram based on birds (cf. figure 8.23).

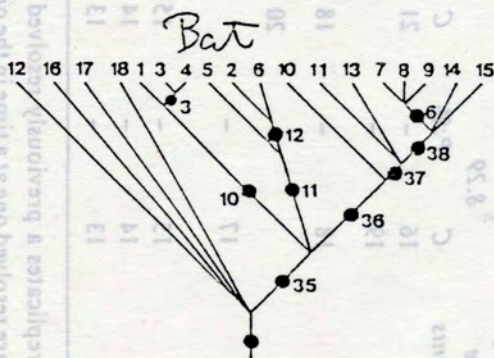
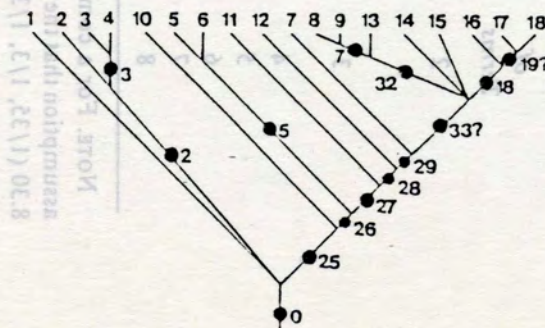


Figure 8.35. Area-cladogram based on bats (cf. figure 8.24).

Regions 1-18



General Area-Cladogram (Butterfly + Bird + Bat)

Figure 8.36. General area-cladogram (cf. figures 8.33-8.35).

Regions 1-18

Table 8.4. Probabilities (P) of Replication of Components (C) in Area-Cladograms

Number of Terms	Number of Possible Components	8.33		8.34		8.35		P for Components of Same Number of Terms, %	P. Cumulative %
		C	P. %	C	P. %	C	P. %		
2	153	3	-	5	2.61	3	3.92		
		5	-	7	2.50			2.56×10^{-1}	
		7	-	9	-	12	-		
		19	-	21	-			2.56×10^{-1}	
3	816	2	-	2	1.88	6	-		
		18	-	18	2.38	10	-	2.23×10^{-1}	
		32	-	32	5.00	11	-		
4	3,060	17	-	8	-				10^{-7}
		31	-						10^{-7}
5	8,568	30	-	34	-	38	-		10^{-7}
7	31,824					37	-		10^{-7}
8	43,758			33	-	36	-		10^{-7}
9	48,620	29	-	29	25.0			1.43×10^{-8}	
10	43,758	28	-	28	33.3			4.77×10^{-9}	
11	31,824	27	-	27	50.0			2.38×10^{-9}	
13	8,568	26	-	26	100.			10^{-9}	
14	3,060	25	-	25	100.			10^{-9}	
17	18	24	-					10^{-9}	

NOTE. For a component that replicates a previously resolved component, the probability is that of replication due to chance alone, under the assumption that the components are resolved one at a time in the order listed for each cladogram. The probability fractions for replicates are: cladogram 8.34 (4/153, 3/120, 3/160, 2/84, 1/20, 1/4, 1/3, 1/2, 1/1, 1/1, 1/1), cladogram 8.35 (6/153). (cf. Table 4.2)

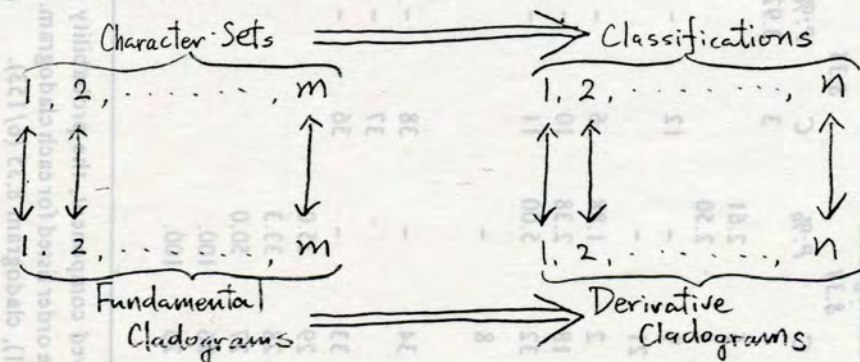
— Classification and General Cladograms —

ある一組の character-set に対して grouping (phyletic, phenetic, gradistic) を行おうと一つの branching diagram が生じる。これを "Fundamental Cladogram" とする。一方分類体系 (Linnaean Hierarchy) のその集合関係 (cladistic aspect) もまた一つの branching diagram を派生する。これを "Derivative Cladogram" とする。(p.307)

Fundamental Cladograms: which summarize data about the interrelationships of organisms

Derivative Cladograms (Classifications): which specify the cladistic aspect of classifications (p.307)

この2つの間の関係は下のようである。



Systematic study of the interrelationships of the species of a given group generally results first in one [or more] fundamental cladogram, which is used as a basis for constructing one [or more] derivative cladogram. (p.307)

Cladogram の情報内容については、

$$\text{Total Information (U)} = \text{Component Information (C)} + \text{Term Information (T)}$$

が成り立つことは既に見た。この式は上述の2種の cladogram のそれぞれについて成り立つから、

$$\text{Fundamental Cladogram} : U_F = C_F + T_F$$

$$\text{Derivative Cladogram} : U_D = C_D + T_D$$

となる。ここで Classification (Derivative Cladogram) がどの程度もとの Data (Fundamental Cladogram) の情報を反映しているかを示す指標として "Efficiency"

を考へる。上の記号を用ひて

Component Efficiency: $E_C = \frac{C_D}{C_F}$

Term Efficiency : $E_T = \frac{T_D}{T_F}$

Total Efficiency : $E_U = \frac{U_D}{U_F}$

x100 の%表示 (cf. Table 41)

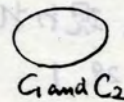
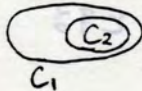
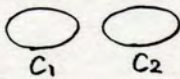
(p.310)

Fundamental Cladogram が複数個ある場合、それらを含む Components の関係は次の通りである。(p.312) Set-Relationship

A. Combinability

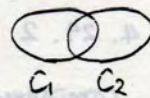
ある一つの cladogram に同時に存在し得る。

- 1) Exclusion
- 2) Inclusion
- 3) Replication



B. Noncombinability

ある一つの cladogram に同時に存在し得ない。



もしも自然が chaos のみをもつならば、相異なる fundamental cladograms は全て正当化される。けれども order の存在を前提とするならば、それらは全てある単一の natural system を記述していると考えられる。(p.310-311) ここで、複数の fundamental cladograms を何らかの方法でまとめて、一つの cladogram が出来るならば、それは natural system の Descriptor と考えられる。この cladogram を "General Cladogram" とする。(p.311)

上の component relationship のうち特に "Replication" に注目したい。

Replicated components are significant because the probability of replication for any given component, due to chance alone, is small. (p.312)

例えば、n-taxon problem に生じた m-taxon component が他の character-set に関して replicate される確率を考へる。m-taxon component の場合の数は nC_m だから、ある component が replicate される確率は $1/nC_m$ である。それ故

$$\frac{1}{nC_m} = \frac{1}{nC_1} + \frac{1}{nC_2} + \dots + \frac{1}{nC_m}$$

The cladogram specified by the replicates may be considered a first step toward a general cladogram. (p.314)

つまり、異なる character-set に対してくり返し生ずる component (replicate) をもとにして general cladogram を作るわけである。例として Fig. 4.29 を挙げる。(p.306)

Fig. 4.29 の 1~5 の 5 つの fundamental cladogram は 18 個の informative components を含む。この components がいつに replicate されているかを示したのが Table 4.2 (p.313) である。cladogram が複数の component を含む時、それらが replicate される確率を求めるには、いくつかの方法があるが、(1) がある。

the assumption that the components are resolved one at a time in the order listed for each cladogram. (Table 4.2. caption)

として、Probability of Replication (P) を計算する。まず、component が「初めて」現われた時は $P=0$ とする

1) 4.29.1

それぞれの component に対して $P=0$ が成り立つ。

2) 4.29.2

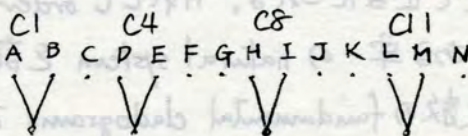
1-11

component が生ずる順序は 2-taxon, 3-taxon, ... の順で、~~taxon の場合は Table 4.2 の順である~~。

o 2-taxon component: C1, C4, C8, C11

の 4 つがある。

4.29.1 には



14 taxa から 2 taxa をとる場合の数は

${}_{14}C_2 = 91$

4.29.1

であり、そのうち C1, C4, C8, C11 は 4.29.1

で生じている、ここで Replication の確率 P は

[1] C1, C4, C8, C11 が 4.29.2 ~ 4.29.5 で示された順序で起る確率

[2] C1, C4, C8, C11 が 4.29.2 ~ 4.29.5 で示された数だけ起る確率

の 2通りに解釈できる。[1], [2] に対応する確率をそれぞれ P_1, P_2 とする。

[1] に従うならば、4.29.2 では C1, C4, C8 がこの順で発生するから、それが replicate される確率は

$$C1: \frac{1}{{}_{14}C_2}, C4: \frac{1}{{}_{14-2}C_2}, C8: \frac{1}{{}_{14-2-2}C_2}$$

$\therefore P_1 = \frac{1}{91} \cdot \frac{1}{66} \cdot \frac{1}{45}$ C1が生ずると、C4はC1を除く12 taxaから計算される
 C8はC1, C4を除く10 taxaから計算される

[2]に従うならば、

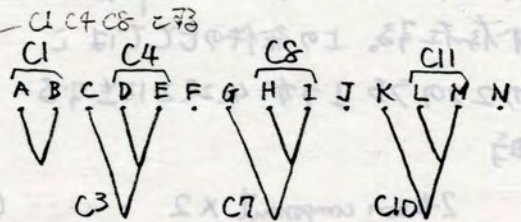
C1, C4, C8, C11 から1つが生じる確率 $\dots \dots \dots \frac{4}{14} C_2$
 上の残りから1つが生じる確率 $\dots \dots \dots \frac{3}{14-2} C_2$
 上の残りから1つが生じる確率 $\dots \dots \dots \frac{2}{14-2-2} C_2$

$\therefore P_2 = \frac{4}{91} \cdot \frac{3}{66} \cdot \frac{2}{45}$ $\therefore P_1 < P_2$

つまり [1] の解釈は 4.29.2 で C1, C4, C8 がこの順で生ずる確率を求めているのに対し、[2] は 4.29.2 で C1, C4, C8, C11 からどれか3個が生ずる確率を求めている。だから、 P_1 は順序対 (C1, C4, C8) の生ずる確率であるが P_2 はそれを含む被数個 ($4 \times 3 \times 2 = 24$ 個) の順序対のどれかが生ずる確率である。だから $P_1 < P_2$ となる。本書では [2] の解釈が採用されている。その理由は、ここで問題となっているのは反復された component の「数」であって「種類」ではないからである。これについては以下でも述べる。

とにかく以下では P は P_2 を意味するものとする。

• 3-taxon component: 4.29.1 では C3, C7, C10 の3つがある。ここで既に C1, C4, C8, C11 のうち3つが存在しているからその 2-taxa を 1 taxon と見なす。3-taxon component が1つ生じる時、



4.29.1

2-taxon component の数 $3 \times 8 = 24$
 component 以外の taxon 数

1 taxon $\times 3 \dots \dots \dots 8C_3 = 56$
 Total $24 + 56 = 80$

だから

C3, C7, C10 から1つが生じる確率 $\dots \dots \dots \frac{3}{80} \text{ --- ①}$

次に2番目の 3-taxon component が生じる時、①で生じた1つの 3-taxon component は除外されるから

2-taxon component + 1 taxon

$$2 \times 7 = 14$$

残された 1-taxon の数

1-taxon x 3

$${}^7C_3 = 35$$

Total $14 + 35 = 49$

上の残りから 1つが生じる確率 $\dots 2/49$ ②

最後に、3番目の 3-taxon component が生じる時

2-taxon component + 1 taxon

$$1 \times 6 = 6$$

1-taxon x 3

$${}^6C_3 = 20$$

Total $6 + 20 = 26$

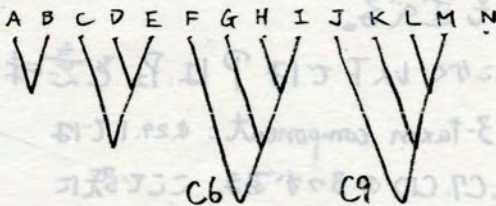
上の残りから 1つが生じる確率 $\dots 1/26$ ③

①②③より

$$P = \frac{3}{80} \cdot \frac{2}{49} \cdot \frac{1}{26}$$

・ 4 taxon component : 4.29.1 に

は C6, C9 の 2つの 4 taxon component が存在する。上の条件のもとではこの2つのうち 1つが 4.29.2 に生じる時



2-taxon component x 2

0 3 taxon component の数

3-taxon component + 1 taxon

$$3 \times 3 = 9$$

残された 1 taxon の数

2-taxon component + 1 taxon x 2

$$1 \times 3 = 3$$

Total $9 + 3 = 12$

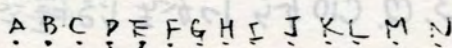
C6, C9 のうち 1つが生じる確率 $\dots 2/12$

以上で 4.29.2 は終り。

③ 4.29.3 2-taxon component C4 (1個だけ)

4.29.1 における C1, C4, C8, C11 のうち 1つが replicate される確率 $\dots 4/9$

3 taxon component は C3 のみで



ある。C4 の制限条件のもとでは

3-taxon component の生じる場合は 4.29.3 に

4.29.3

2-taxon component + 1-taxon ... (1) x (2) = 12

1-taxon x 3 ... 12 C3 = 220

Total 12 + 220 = 232

C3, C7, C10のうち1つが生じる確率 ... 3/232

その他 4.29.3は 8-taxon component (C13), 10-taxon component (C12) を含むが、これらは初出だから P=0 である。

4) 4.29.4 : 2-taxon component はなく 3-taxon component は C17 が1個だけあるが初出だから P=0。4-taxon component は C9 が replicate される。制限条件は C17 だから、4.29.4 に 4-taxon component が生じる場合は、

3-taxon component + 1-taxon ... (1) x (11) = 11

1-taxon x 4 ... 11 C4 = 330

Total 11 + 330 = 341

C6, C9のうち1つが生じる確率 ... 2/341

5-taxon, 7-taxon, 12-taxon component (C16, C15, C14) は全て初出。

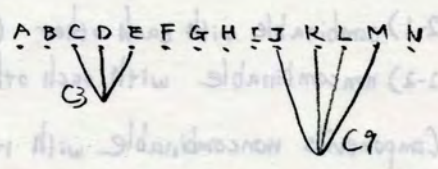
5) 4.29.5 : 3-taxon component C3 が存在する。制限条件はないから、場合の数は 14 C3 = 364。4.29.1で C3, C7, C10, さらに 4.29.4で C17 が生じたから、

C3, C7, C10, C17のうち1つが生じる確率 ... 4/364

4-taxon component C9 が存在する。

場合の数は

3-taxon component + 1-taxon
1 x 11 = 11



1-taxon x 4 ... 11 C4 = 330

Total 11 + 330 = 341

C6, C9のうち1つが生じる確率 ... 2/341

5-taxon component C18 は初出。

以上をもち Table 4.2の各項目の説明は終わった。上のこゝは component の replication probability を直接計算することにより簡単になる。その結果を 2-taxon component ごとに集計したのが "P for components of same number"

of terms" の項^目であり、これを component に関して累積したのが "P cumulative" の項目である。これからわかるように偶然性のみにより、4.29.1~5 が生ずる確率は実に 10^{-19} ($10^{-17}\%$) の order なのである。
 2) として正確に 4.29.1~5 が生ずる確率は 10^{-19} であり、更に小さいのである ($\therefore P_1 < P_2$)

Type
cladogram

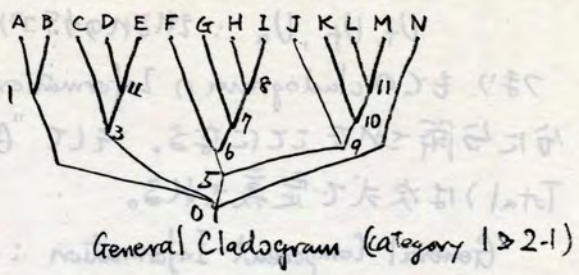
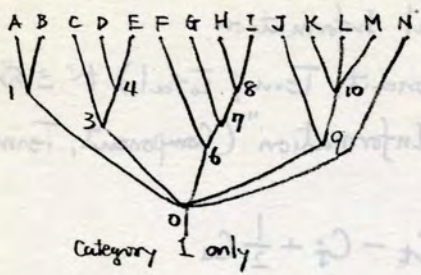
- (1), (2) の解釈をもう一度まとめる。
- [1]: 4.29.1~5 が、偶然性のみにより、正確に生ずる確率つまり、4.29.2 で 2-taxon component C1, C4, C8 がこの順で生じること
 - [2]: 4.29.1~5 型の反復 component の数の分布 pattern をもつ cladogram が、偶然性のみにより、生ずる確率つまり、ex 4.29.2 で 2-taxon component が (C1, C4, C8, C11 のうち) 3つ replicate されること

この結果 Components 1, 4, 6, 7, 8, 9, 10 が replicate されており、それが偶然性のみにより反復される確率は非常に小さいことがわかった。そこでこれらの Replicates をもってさらに考察を進める (p.314 下) Table 4.1 参照。Component は次のカテゴリーに分けられる。

- 1) Replicates (1, 3, 4, 6, 7, 8, 9, 10)
- 2) Components combinable with replicates (2, 5, 11, 12)
 - 2-1) combinable with each other (5, 11)
 - 2-2) noncombinable with each other (2, 12)
- 3) Components noncombinable with replicates (13, 14, 15, 16, 17, 18)

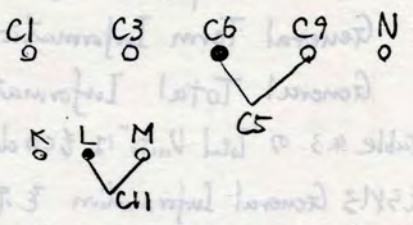
先に述べたように、Replicates をもって general cladogram をつくる出発点とするならば、まずはじめに replicates と相入れない (noncombinable) components 即ち カテゴリー-3 は除かねばならない。次に replicates と combinable であるも互いに combinable でなければ一つの cladogram 内で共存できないから、カテゴリー-2-2 もまた除かねばならない。

このようにして Replicates (カテゴリー-1) をもってした cladogram が Fig. 4.31.2 であり、それに カテゴリー-2-1 を付加したものが Fig. 4.32 でこれが "General Cladogram" とされる。



Category 2-1 に含まれる C5, C11 を考える

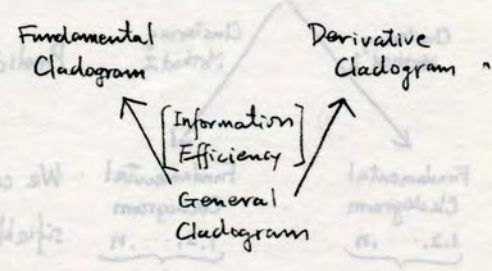
C5: $C5 = C6 + C9$ であるが、これは C6 が C1, C3, C9, N とともに作り得る $4C_1 + 4C_2 + 4C_3 + 4C_4 = 15$ 個の components の 1 つである。一方 C11



は、C10 の K, L, M が作り得る $2C_1 + 2C_2 = 3$ 個の components の 1 つである。従って、Replicates と共存し得る Category 2-1 の components の可能な組合せは $15 \times 3 = 45$ 通りであり、C5, C11 はその 1 つであるから、確率は $1/45 \approx 0.022$

- Category 1 } True Components (General Cladogram)
- Category 2-1 }
- Category 2-2 } Ambiguous Components
- Category 3 } False Components (p. 315) 1982.11.12

Table 4-1 では Cladogram (Fundamental, Derivative) の Information (Component, Term, Total) は それぞれの cladogram について計算され、Efficiency (Component, Term, Total) も Derivative Cladogram に対応する Fundamental Cladogram をもとにして計算されていた。けれども上で述べた General Cladogram を基準とするならば、Fundamental 及び Derivative



Cladogram の Information と Efficiency は 同一の基準で比較できる (Table 4-3) である。General Cladogram を基準とする Component の 3 カテゴリー (True, False, Ambiguous) を それぞれの cladogram に適用する。このとき

- ① C_t, C_f, C_a : それぞれのカテゴリーの Component Information
- T_t, T_f, T_a : の Term Information

U_t, U_f, U_a : これこれの「成分」の Total Information
 7手). もとの cladogram の Information (Component, Term, Total) が 3 成分
 に分解されたことになる。それ "General Information" (Component, Term
 Total) は次式で定義される。

General Component Information : $C_g \equiv C_t - C_f + \frac{1}{2} C_a$

General Term Information : $T_g \equiv T_t - T_f + \frac{1}{2} T_a$

General Total Information : $U_g \equiv U_t + U_f + \frac{1}{2} U_a$ 322 "Total C+T"

[Table 4.3 の "Total Units" は もとの cladogram の Information を示し。"Total True" は 式で示
 される General Information を示す。]

General Cladogram の Information をそれぞれ C_g, T_g, U_g とする。
 "General Efficiency" は次式で定義される。

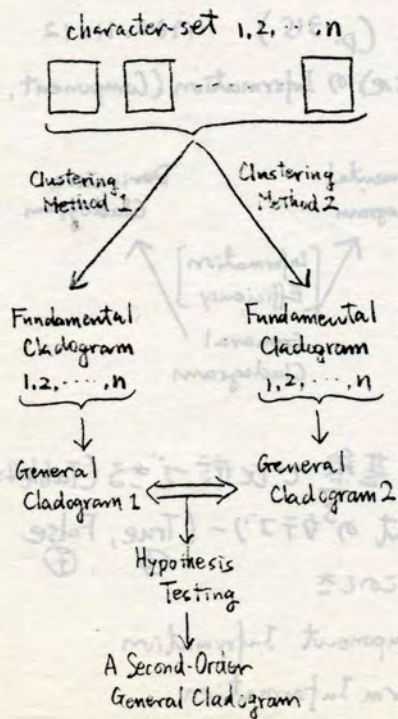
General Component Efficiency : $E_c \equiv C_g / C_t$

General Term Efficiency : $E_T \equiv T_g / T_t$

General Total Efficiency : $E_U \equiv U_g / U_t$

General Information, General Efficiency の "General" とは "General
 Cladogram を基準にした" という意味である。

General cladogram specifies a nonrandom pattern of interrelationships. (p.322)
 1982. 11. 13.



Hypothesis: General cladograms estimate the same
 dadistic parameter.

Prediction: Two or more general cladograms [1, 2, etc] are
 not significantly different.

We conclude that it (a general cladogram) is not [fal-
 sifiable] in a general sense, but that it is in a particular
 sense. (p.323.)

The theory of general cladograms includes phenetics,
 phyletics, gradistics, and any other theory that might
 be proposed. (p.323)

1982. 11. 14.