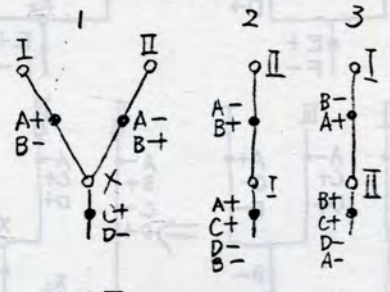


— Trees and Information —

Cladogramの機能は形質の出現に関する情報の要約であるから、いかにコンパクトに)要約しているかが複数のcladogramから選択する上での判断基準であった。この基準が Parsimony Criterion であることは明白である。それでは、Cladogramに unique component を付加した Trees の場合、

Trees

	A	B	C	D	情報はどのよう
Species I	+	-	+	-	に表現されるか? 左表の形
Species II	-	+	+	-	質が与えられた時、3つの 2-taxon trees は右の

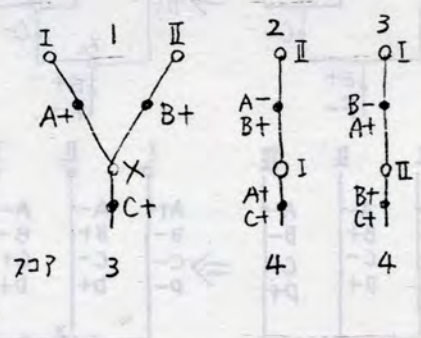


質が与えられた時、3つの 2-taxon trees は右のよう

の Trees から positive 及び negative occurrences を数えるに左表となり、3つとも同スコアになる。つまり		A	B	C	D	A	B	C	D	スコア
	1	1	1	1	0	1	1	0	1	6
	2	1	1	1	0	1	1	0	1	6
	3	1	1	1	0	1	1	0	1	6

これだけでは比較できないことになる。

そこで、positive occurrences (+) に重点を置いて上の形質分布を表現する。つまり、 unnecessary negative occurrences (-) を省いて、できるだけ + のみで表現すると、右図のようになり、



1の Tree が最も compact に情報を要約していることになる。

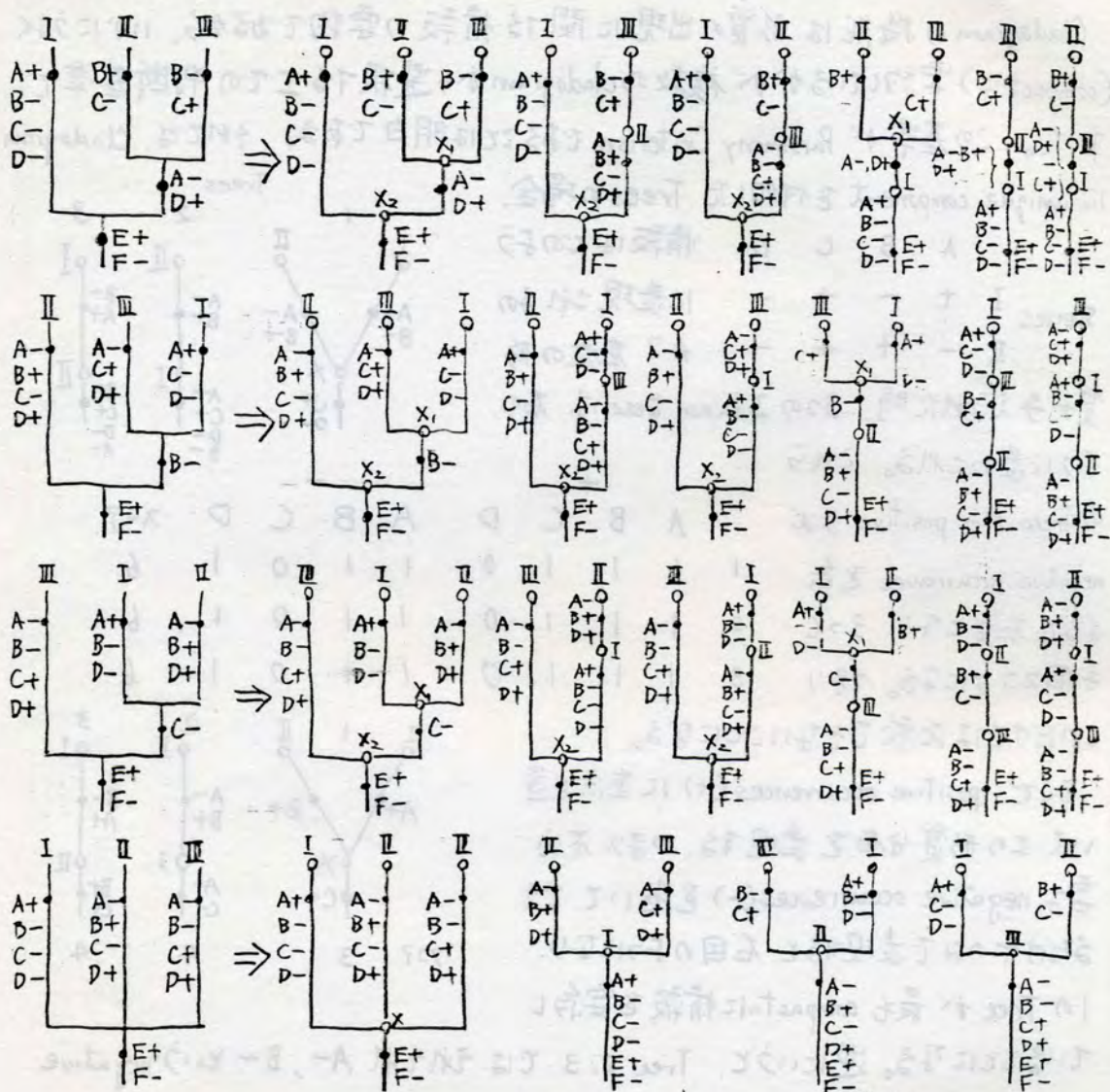
逆にいうと、Tree 2, 3 ではそれぞれ A-, B- という negative occurrence が不可欠なのである。

上からわかることは、Trees の比較は positive / negative occurrences のスコアでは判断できない。何故ならすべての Trees は同スコアを持つからである。だから、Trees の比較は、必要不可欠の negative score がいくつあるかを数え、これが最も小さい Tree を選ぶという parsimony criterion をもとに、行なわなければならない。

この議論は 3-taxon problem を考えることに(F), (F) 明確に理解できる。3-taxon cladogram は 4 つあり、3つの dichotomy のそれぞれに 2 つ、6 つの Trees が存在する。

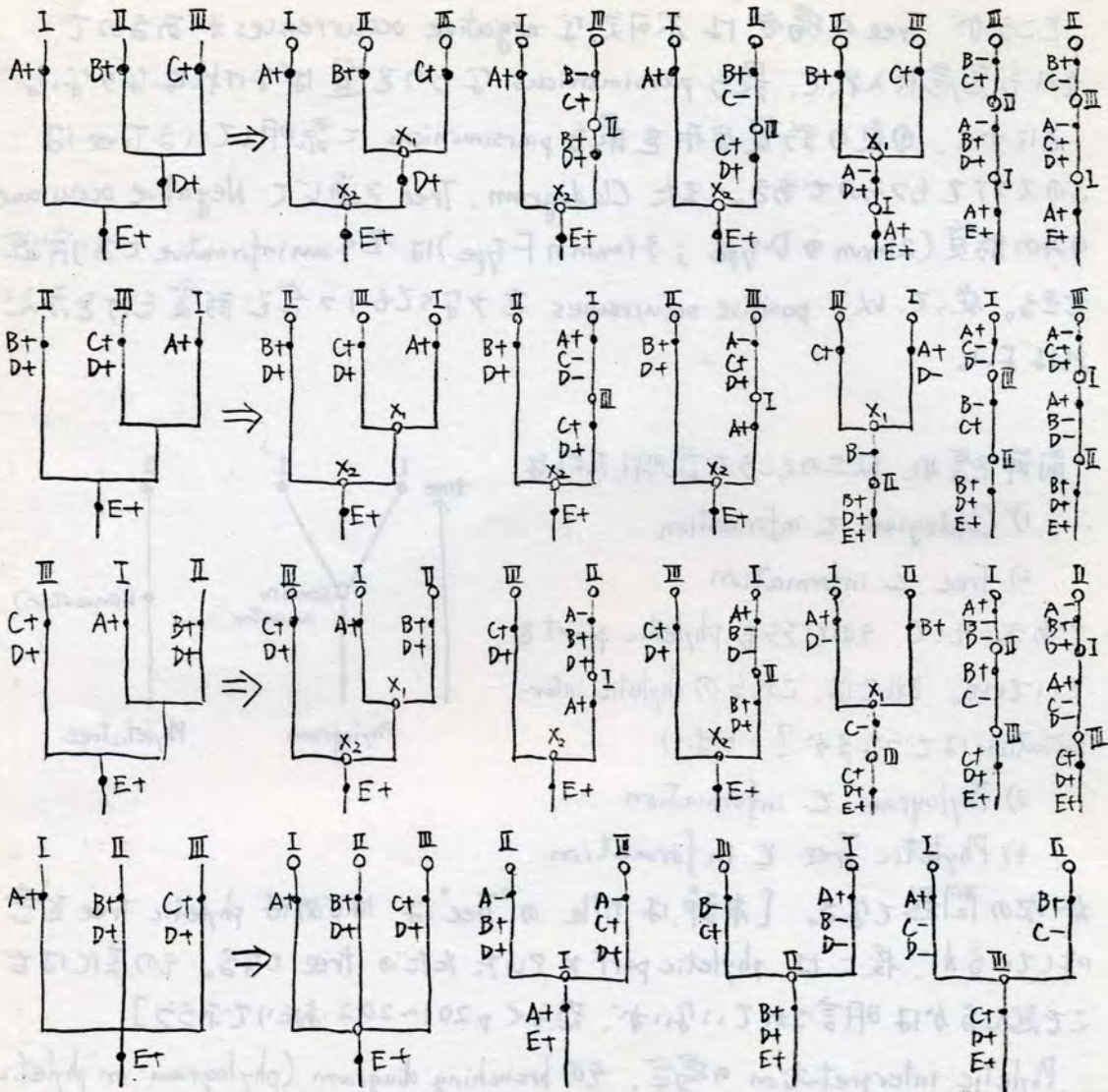
	A	B	C	D	E	F
I	+	-	-	-	+	-
II	-	+	+	+	+	-
III	-	-	+	+	+	-

そして1つの trichotomy に対しては 4つの Trees がある。そこで positive/negative occurrences を記入すると次のようになる。



Cladogram	Trees					
	1	2	3	4	5	6
12	12	12	12	10	11	11
13	13	13	13	10	11	12
13	13	13	13	10	11	12
14	14	12	11	11	X	X
	↓ negative scores partly eliminated					
5	5	6	6	6	7	7
6	6	8	7	7	8	9
6	6	7	8	7	8	9
6	6	8	8	8	X	X

以上の結果を表にすると左のようになる。それからわかるように、2-taxon とは異なり、3-taxon の場合は、必ずしも cladogram と tree のスコアが一致しない。前節で見た通り、cladogram の場合 negative score は不必要であるが、tree の場合は必要なものと不必要なものがある。そこで 3-taxon と同様、不可欠な negative score のみを考慮したのが左下表であり、次の図から算出される。



前頁の positive/negative occurrences の記入法はいくつかあり、それによって total score にいくらか変化があるが、この上図では total score に変化を起す可記入法の差はなく、それ故前頁最下の表の score は一意に定まる。またこの表と図とは p 212, 213 にあるものと同じである。 1982. 10. 21.

Cladogram の選択を、positive/negative occurrences を用いて、行なったのは、その記入法を統一していたからである。その場合でも、前表で示したように、決定的要因は positive occurrences (特に D-type character) であった。つまり、negative occurrences は Cladogram の選択には全く不必要であり、すべて削除できる。上図の左端の cladogram の表示はまさにこれであり、positive occurrences のみが記入されている。そして positive occurrences の数が最小である(この場合一番上の cladogram)ものが形勢分布を most parsimonious に説明しているとして選ばれているのである。

ところが Tree の場合は不可欠な negative occurrences があるので、それも考慮に入れて、最も parsimonious なものを選ぶなければならない。  
 とにかく、①頁の形質分布を最も parsimonious に説明している Tree は 5 のステップをもつものである。また Cladogram, Tree を通じて Negative occurrence のみの形質 (2-taxon の D-type ; 3-taxon の F-type) は全く uninformative であり削減できる。従って、以下 positive occurrences を少なくとも 1 つ含む形質だけを考えればよい。

前節を含め、以上のことで説明したのは

1) Cladogram と information

2) Tree と information

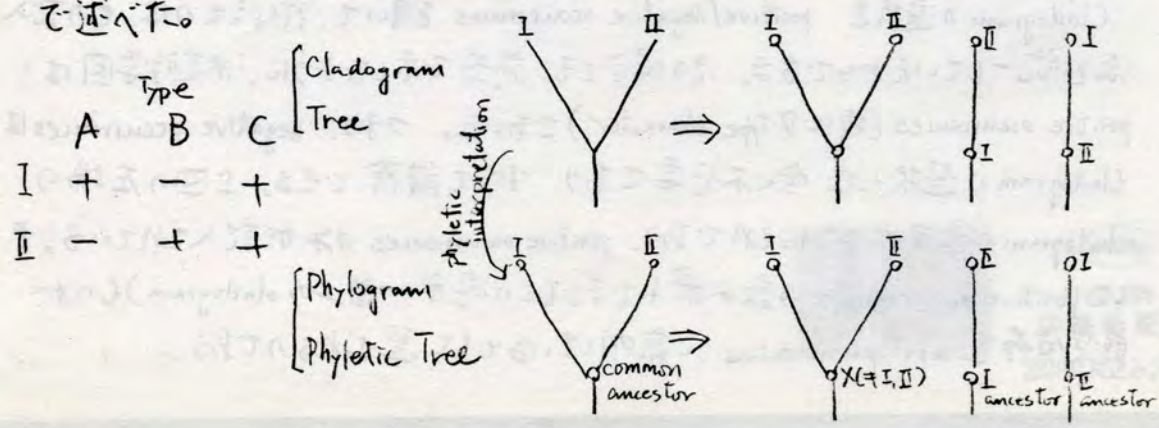
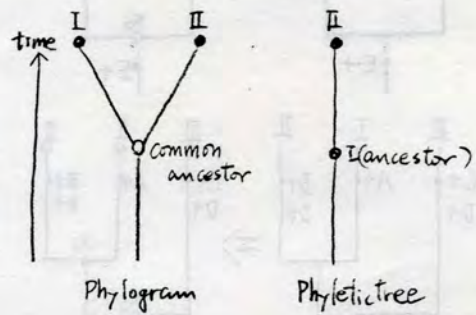
である。そして、そのどちらも phyletic part を欠いている。それでは、これらの phyletic interpretation はどうなるか? (つまり)

3) Phylogram と information

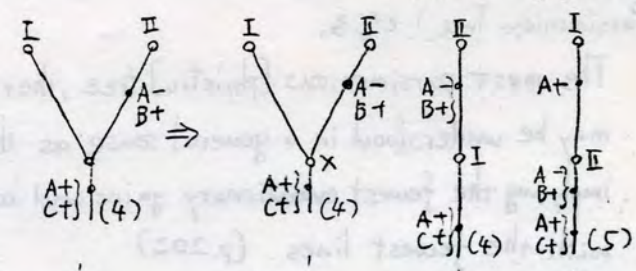
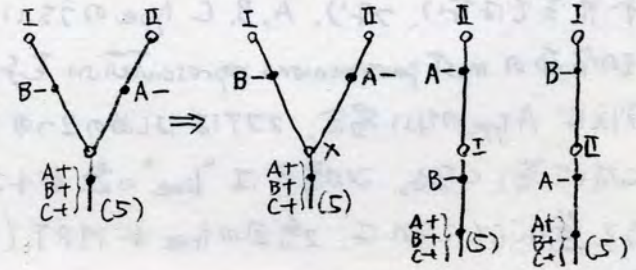
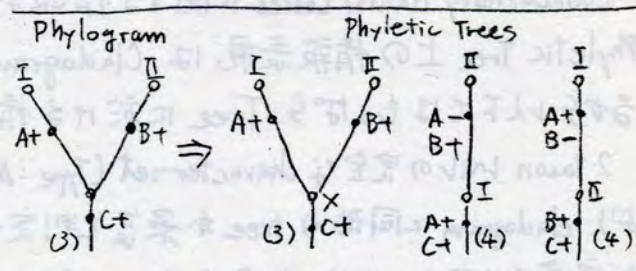
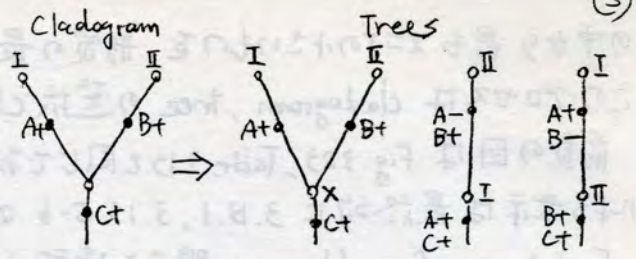
4) Phyletic Tree と information

が次の問題となる。[本節は title の "Tree" は はじめは phyletic tree を意味しているか、後には phyletic part を欠いたただの tree となる。その変化はどこで起こるかは明言されていないが、恐らく p.201~202 あたりであろう]

Phyletic interpretation の場合、その branching diagram (phylogram or phyletic tree) の "line" の意味が変わってくる。つまり、cladogram や tree のように、形質の存在を示す「場所」に加えて、その方向にそう「時間」成分が生じてくる。従って、この evolutionary connotation を反映して、information の表示は無数にある。2-taxon の場合を下記に示す。D-type を省略する理由は上で述べた。



つまり、Cladogram や Tree の場合、"line" は形質の positive/negative occurrences を表すだけである。ところが Phylogram や Phyletic Tree の "line" は、time component が加わったので、evolutionary gains/losses を示して示るのである (p201) ところが、形質分布の表からはこれらに関する情報は得られないので、余分な仮説を最小にする parsimony criterion を適用するしかない。



それぞれの Phylogram, Phyletic Tree のスコアを数えると括弧で示したようになる。これからわかることは

1) Phylogram (1の系列)

Most parsimonious phylogram はスコアの最初の phylogram

2) Phyletic Tree (2, 3, 4の系列)

Most parsimonious phyletic tree はそれぞれの系列の最初の phyletic tree

従って、それぞれの系列の最上段のものが形質分布の most parsimonious representation であり、これらは、cladogram, tree のそれと一致している。つまり、time component を考慮した時、無数の representation が可能であるが、parsimony を考えれば、最終的に cladogram や tree による表示と一致する (最上段) の phylogram, phyletic tree に帰着するのである。よって、phylogram, phyletic tree に於ける information の表現は、parsimony criterion ははたらく限り、cladogram, tree のそれと一致する。そして、これらの phylogram, phyletic tree

(endless)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

(4)

(4)

(4)

(5)

の中から最もスコアの小さいものを、形質の最節約表現として選ぶわけである。  
このプロセスは cladogram, tree の選択と同一である。

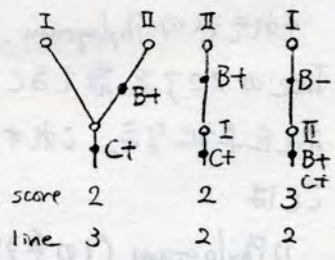
前頁の図は Fig. 3.23, Table 3.13 と同じであり、Fig. 3.23 1-8 の phyletic tree による表示は最終的に 3.13.1, 3.13.5-6 の3つからの選択に戻る。

Evolutionary Gains/Losses に関する情報が得られない限り、Phylogram, Phyletic Tree 上の情報表現は Cladogram, Tree のそれと一致する。であるから以下ではもっぱら Tree に於ける情報表現に注目する。

2-taxon 1x1 の完全な character-set (Type: A, B, C, [D]) に対しては、上の解析により、cladogram と同形の tree が最良と判定された。それでは、character set が完全ではない、つまり、A, B, C type のうちいくつかは欠落している集合に対してその分布の most parsimonious representation を与える tree はどうなるのか? (p.202ff.)

例えば A-type のない場合、スコアははじめの2つの tree に対して等しくなる。この場合は "line" の数が小さい方を選ぶことにすれば、2番目の tree が MPT (Most Parsimonious Tree) となる。

	A	B	C
I	0	-	+
II	0	+	+



The most parsimonious [phyletic] tree, therefore, may be understood in a general sense as that tree implying the fewest evolutionary gains and losses with the fewest lines. (p.202)

一般に A, B, C type が存在するか欠落するかによって、 $2^3=8$ 通りの場合があり、1つの complete set を除く 7つは incomplete sets となる。(Table 3-15, Fig. 3.25)

	A	B	C	MPT		A	B	C	MPT
1	I +	-	+		5	I +	-	0	
	II -	+	+			II -	+	0	
2	I 0	-	+		6	I 0	-	0	
	II 0	+	+			II 0	+	0	
3	I +	0	+		7	I +	0	0	
	II -	0	+			II -	0	0	
4	I 0	0	+		8	I 0	0	0	
	II 0	0	+			II 0	0	0	

1: complete set      2~7: incomplete sets

しかし十分な調査が行われれば、全ての character-type が見つかる時期  
される。これ故、character set 1 に対応する tree が「一般解」と言えるだろう。  
ただし、

That tree [3.25]1 seems to be the general solution to the problem of  
the MPT for any two taxa again does not prove that tree [3.25]1 is true,  
and the other trees are wrong, as historical statements of ancestry and  
descent. (p.205)

これ故 (p.205)

Historical truth must be judged by some criterion other than the MPT.  
なぜなら、

A MPT seems to offer no basis for inference beyond that of the  
corresponding cladogram for the reasons mentioned above. (p.205~206)  
以上をまとめれば、

There seems no reason to equate the notions of MPT and historical truth.  
Indeed, it may be best to divorce these two concepts and to inquire  
whether historical truth may be estimated by means other than the search  
for the MPT. (p.206)

つまり、Cladogramと同形の tree が一般的な MPT であるが、それが歴史的  
的眞実という保証にはない。だからこの二つは分けるのがよい、ということである。  
それではどういふ場合に一般解としての MPT が歴史的に見て眞実では  
ないのか？ それには次の二つの場合がある。(p.206ff.) ただし character set  
5-8 は考慮しない。

1) Character Type A 科には B の欠除が sampling error ではなく眞実であるとき。  
上述したように MPT 1 は偽であり 2 科には 3 が歴史的に眞である。

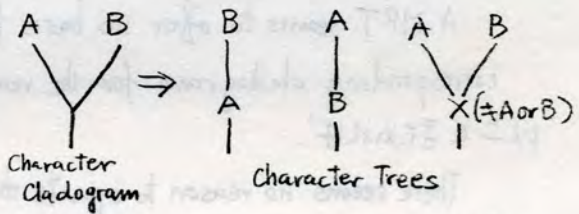
2) Character-Type A 科には B の全てが変換されている (Character Transformation) とき  
これはつまり A- が実 transformed A ("A(t)") であることがわかった場合であり、  
A → A(t) が成立し、I, II 兩種に分布するからやはり A-type ではなく C-type  
になる。(Table 3.16.1) このとき MPT 1 は MPT 2 になる。なぜなら、A(t)  
は II に存在するから A(t) = B+ と見なせば、これは B-type になる。更に A → A(t)  
は明白であるから、A, A(t) は C-type の形質となる従って、A-type はなくなり、  
B-type と C-type だけがあるから、Character set は 1 → 2 となり、これでも

MPTも 1→2 となるのである。Bが変換される場合も同じ (Table 3.16.2)  
 A, B が両方とも変換される。IIにある A(t) は B-type, Iにある B(t) は A-type  
 であり、さらに A, A(t) 及び B, B(t) は C-type だから Character-set は 1 のまま不  
 変。それ故 MPT I は不変である。 (Table 3.16.2A) 1982.10.22

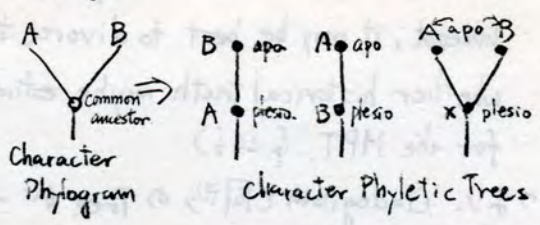
形質変換の問題は A-type と B-type の間の相同問題に帰着できる。  
 (p.207 中) つまり、 $A \rightarrow B, B \rightarrow A, X \xleftrightarrow{A} B$  のこれか起こる場合である。(Table  
 3.16.1-3) しか、この形質の派生関係 (Plesiomorphy / Apomorphy Relationship) は  
 実は形質変換を Phyletic に解釈 (た場合であり)、そう解釈しないならば  
 単に「種 I, II には状態 A, B (派生関係は問わない) をもつ C-type の形質が  
 存在する」と言っているだけである。形質 A, B を taxa と見ると、形質に関

H DAY!

する cladogram, tree, phylogram,  
 phyletic tree が書ける。これこれ  
 の解釈は次の通り)



1) Character Cladogram  
 状態 A, B をもつ C-type の形質が  
 存在する。



2) Character Phylogram  
 状態 A, B をもつ C-type の形質が  
 存在し、A, B は 0 の共有祖先

形質に由来する (Character Cladogram の Phyletic Interpretation)

3) Character Trees  
 状態 A, B をもつ C-type の形質が存在し、Unique Component A, B, X が  
 指定されている。

4) Character Phyletic Trees  
 状態 A, B をもつ C-type の形質が存在し、直接の祖先 (Plesiomorphy) A, B  
 X に由来している。(Character Tree の Phyletic Interpretation)

Character Cladogram が Character Trees の集合であり、Character Phylogram  
 が Character Phyletic Trees の集合であるのは、2-taxon problem の場合と同じ  
 である。この場合の Phyletic Interpretation は Character Transformation に基  
 づく解釈である。

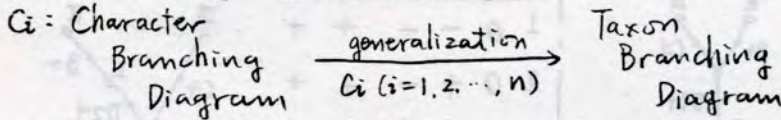
この所を説明したのが次の文である。



A homology statement merely invokes the notion of a character in common or, in this example, a character of type C. A statement of evolutionary homology adds the possibility, but not the necessity, that a character previously thought to be of type A, or of type B, is in reality, a character of type C. In an evolutionary sense, therefore, a homology statement may be understood generally to imply the possibility of a particular character transformation: that, of two characters, one is primitive relative to the other, which is advanced. (p. 209)

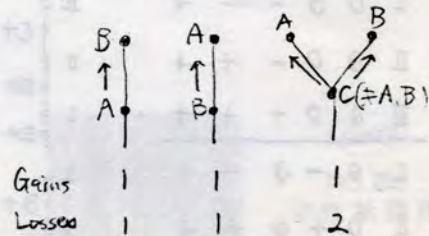
Fig. 3.26.1-3 は "Character Tree" とされているが、実は character phyletic tree であり、Fig. 3.26.4 は 1, 2 を組み合わせたものとして、実は ~~より~~ 一般的に character phylogram と考えられる。

では、一般的に見て、Taxon Branching Diagram (Cladogram, Tree, Phylogram, Phyletic Tree) と Character Branching Diagram (Character Cladogram, etc.) との関係はどうなっているのか、一見して明らかであるが Character Branching Diagram は単一の C-type の形質 ( $C_i$ ) について、その状態 ( $A_i, B_i$ ) が存在しているかとか  $A_i \rightarrow B_i$  とか論じているのに対し、Taxon Branching Diagram では複数の形質 ( $C = \{C_1, C_2, \dots, C_n\}$ ) に関する問題が扱われている。従って

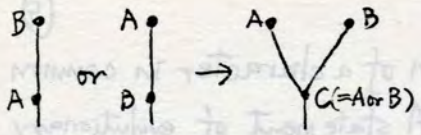


という図式が成り立つ (p. 210) 1982.10.23.

更にまた、unique component (或は plesiomorphy の指定) をしていないだけ、Character Cladogram (或は Character Phylogram) の方が一般的である。いかえれば全ての可能性 (Character [Phyletic] Trees に包含される) を含んでいる。従って、Character Cladogram (Character Phylogram) に対応する [Taxon] Cladogram (Phylogram) が一般解といえる。それでは Tree (Phyletic Tree) の場合はどうなるか? Character [Phyletic] Trees の比較は、Taxon [Phyletic] Trees と同じく、Evolutionary Gains/Losses のスコアに基づく。この場合 "Loss" は "Transformation" のことである。そのスコアからわかるように最初の 2 つが同スコアで Parsimonious Tree といえる。そこでこの 2 つの Tree をあわせること次の Fig. になる。



(Table. 3 17)



右の character tree は character phylogram の unique component を  $X=A \text{ or } B$  と定められたものである。

そして、この character tree が most parsimonious であり、これに対応する下の Tree が 2-taxon の場合の MPT である。

つまり形質変換が起こることが起こるまいが、右の形の Tree (cladogram と同形) が一般解である。これが歴史的真実である保証はない。



3-taxon の場合 ① 頁の形質分布に対しては、これに対応する cladogram と同形の Tree が MPT であった。2-taxon の場合、形質の組合せは  $2^2=8$  通りあったが 3-taxon の場合はどうなるか? Character-Type は A, B, C, D, E (F は省略。の理由は 2-taxon と同じ) のちで、D には D1, D2, D3 の 3 通りがある。そして D-type があれば Primary Tree でありなければ Tertiary Tree になる。

α) A B C D1 E

I	+	-	-	-	+
II	-	+	-	+	+
III	-	-	+	+	+

---

I	0	-	-	-	+
II	0	+	-	+	+
III	0	-	+	+	+

---

I	+	0	-	-	+
II	-	0	-	+	+
III	-	0	+	+	+

---

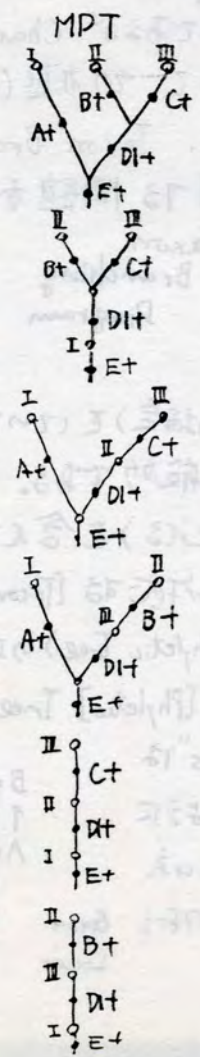
I	+	-	0	-	+
II	-	+	0	+	+
III	-	-	0	+	+

---

I	0	0	-	-	+
II	0	0	-	+	+
III	0	0	+	+	+

---

I	0	-	0	-	+
II	0	+	0	+	+
III	0	-	0	+	+



β) A B C D2 E

I	+	-	-	+	+
II	-	+	-	+	+
III	-	-	+	-	+

---

I	0	-	-	+	+
II	0	+	-	+	+
III	0	-	+	-	+

---

I	+	0	-	+	+
II	-	0	-	+	+
III	-	0	+	-	+

---

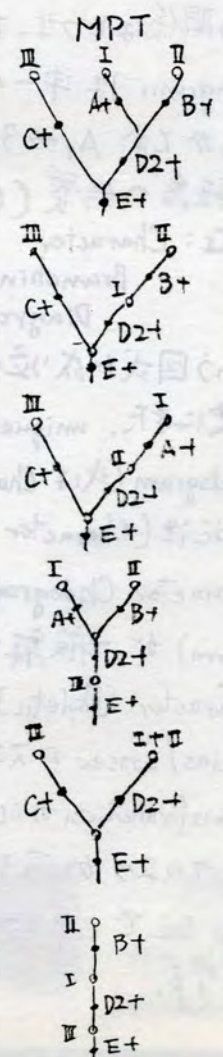
I	+	-	0	+	+
II	-	+	0	+	+
III	-	-	0	-	+

---

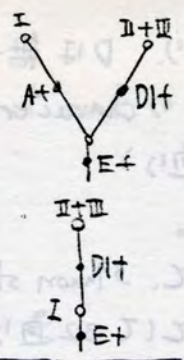
I	0	0	-	+	+
II	0	0	-	+	+
III	0	0	+	-	+

---

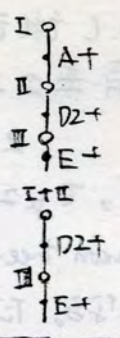
I	0	-	0	+	+
II	0	+	0	+	+
III	0	-	0	-	+



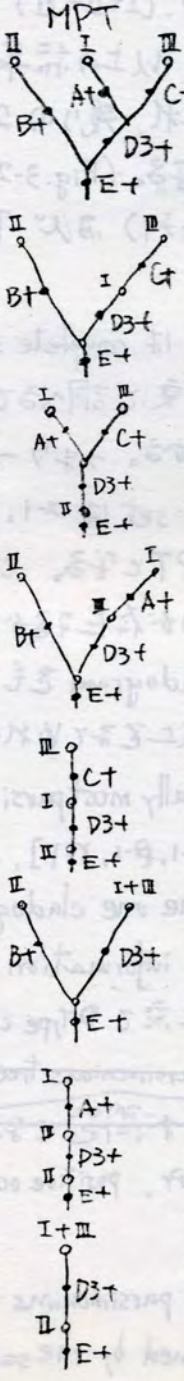
A	B	C	D1	E	
I	+	0	0	-	+
II	-	0	0	+	+
III	-	0	0	+	+
I	0	0	0	-	+
II	0	0	0	+	+
III	0	0	0	+	+



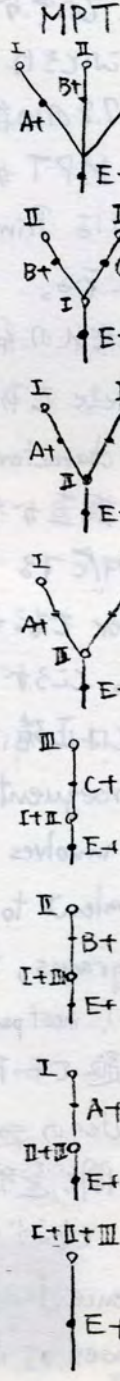
A	B	C	D2	E	
I	+	0	0	+	+
II	-	0	0	+	+
III	-	0	0	-	+
I	0	0	0	+	+
II	0	0	0	+	+
III	0	0	0	-	+



A	B	C	D3	E	
I	+	-	-	+	+
II	-	+	-	-	+
III	-	-	+	+	+
I	0	-	-	+	+
II	0	+	-	-	+
III	0	-	+	+	+
I	+	0	-	+	+
II	-	0	-	-	+
III	-	0	+	+	+
I	+	-	0	+	+
II	-	+	0	-	+
III	-	-	0	+	+
I	0	0	-	+	+
II	0	0	-	-	+
III	0	0	+	+	+
I	0	-	0	+	+
II	0	+	0	-	+
III	0	-	0	+	+
I	+	0	0	+	+
II	-	0	0	-	+
III	-	0	0	+	+
I	0	0	0	+	+
II	0	0	0	-	+
III	0	0	0	+	+



A	B	C	D	E	
I	+	-	-	0	+
II	-	+	-	0	+
III	-	-	+	0	+
I	0	-	-	0	+
II	0	+	-	0	+
III	0	-	+	0	+
I	+	0	-	0	+
II	-	0	-	0	+
III	-	0	+	0	+
I	+	-	0	0	+
II	-	+	0	0	+
III	-	-	0	0	+
I	0	0	-	0	+
II	0	0	-	0	+
III	0	0	+	0	+
I	0	-	0	0	+
II	0	+	0	0	+
III	0	-	0	0	+
I	+	0	0	0	+
II	-	0	0	0	+
III	-	0	0	0	+
I	0	0	0	0	+
II	0	0	0	0	+
III	0	0	0	0	+



A, B, C それぞれについて「有」「無」の2通り。 Dは無・D1・D2・D3の4通り  
Eは有・無の2通りであるから、3-taxon の character-set の場合の数は、  
 $2 \times 2 \times 2 \times 4 \times 2 = 64$  (通り)

したがって p. 211 の「54」は誤りである。

2-taxon tree でも見たように、Eがないと、3-taxon statement としての重畳性はなくなる。だから、上では E+を前提として 32通りの場合を回示した。更に、その中から (I+II), (I+III), (II+III), (I+II+III) を含むものを除く。なぜなら、これは 2-taxon statement だから。以上の結果  $\alpha-7, 8, \beta-5, 8, \gamma-6, 8$  5-5, 6, 7, 8 の10組の character-set は削除され、残りの 22個の set のそれぞれについて MPT が1つずつ対応することになる。(Fig. 3-27, 28; Table 3-18) このうちわけは Primary Tree が18 ( $\alpha, \beta, \gamma$  系列) 及び Tertiary Tree が4 ( $\delta$  系列) である。 1982. 10. 24.

それぞれの系列について  $\alpha-1, \beta-1, \gamma-1$  は complete set であり、それ以外は incomplete である。この complete set について更に調べると、 $\alpha-1, \beta-1, \gamma-1$  の差は D-type character の差に還元できることがわかる。つまり一般的にいえば (△ 充分に形質の調査が行われたならば)、character set は  $\alpha-1, \beta-1, \gamma-1$  のどれかになり、それに対応する Tree が一般解としての MPT となる。この決定要因は D-type character であるから、どの D-type (D1, D2, D3) が存在するかに依り、一意的に MPT が定まる。ところがこの D-type character は Cladogram をも定めるから、MPT と Cladogram とは正確に一致することがわかる。以上をまとめれば

Consequently the problem of the generally most parsimonious tree for three taxa involves only three primary trees [ $\alpha-1, \beta-1, \gamma-1$ ], and in that sense is equivalent to the problem of choosing the one cladogram, of three possible cladograms, that best summarizes the information. (p. 211-213)

つまり、most parsimonious cladogram の選取に於て D-type character の positive occurrence が問題であったが、これと同様に ~~most parsimonious tree を選取するときには D-type character の positive/negative occurrences が問題である。特に  $\alpha-1, \beta-1, \gamma-1$  の MPT 間の選取ならば、cladogram と同様に、positive occurrences のみを取り入れれば~~ であるから。

Hence cladograms and generally most parsimonious trees turn out, at least in cases of three taxa, to be determined by the same factors [D-type character]

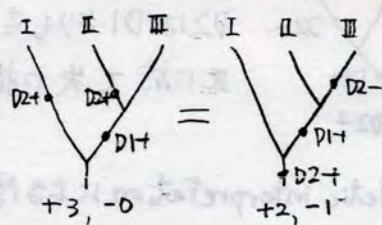
Cladograms and generally most parsimonious trees, therefore, seem merely to imply alternative strategies for arriving at the same result — namely a generalization about the distribution of character-types among species in the real world. (p. 213-215)

既に何度も言っていることだが、D-type character により決定される <sup>G</sup>MPT は必ずしも歴史的眞実ではない。それでは GMPT と歴史的説明との関係はどうなっているのか？ それは、GMPT は歴史的説明の範囲に制限を与えるということである (p. 215) つまり、ある D-type (D1, D2, D3) により 1 つの GMPT ( $\alpha-1, \beta-1, \gamma-1$ ) が決まれば、そのもとで可能な Trees はその D をもつ系列 ( $\alpha, \beta, \gamma$ ) に限定されるということである。例えば D1 type が存在して  $\alpha-1$  が GMPT とは選ばれたならば、歴史的説明の可能な数は 22 (全体) から 6 ( $\alpha, \gamma, \delta$  を除く  $\alpha$  系列の数) に減少するのである。

次に D-type character が conflict している場合を考える。(p. 215 ~ 216)

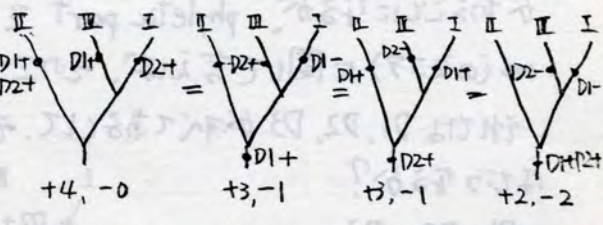
	A	B	C	D1	D2	E
I	+	-	-	-	+	+
II	-	+	-	+	+	+
III	-	-	+	+	-	+

Tree 1  
S=33

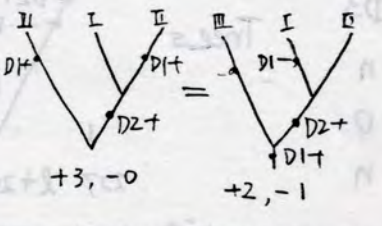


上の形質分布に於いては、Tree 1, 3 が同 S=33 である。(この Trees は 3 taxon における MPT と考えられる) したがって Tree 4 が選択されるわけはない。(何れから Tree 2 は (III) を含意するから誤りでない)。同様に D1, D2, D3 が全て起っている場合は Tree 1, 2, 3 が同 S=33 からやはり Tree 4

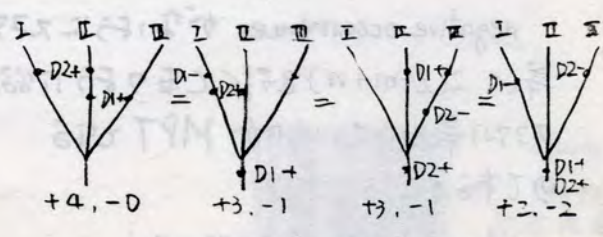
Tree 2  
S=34



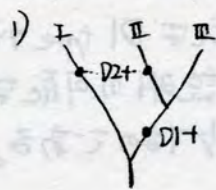
Tree 3  
S=33



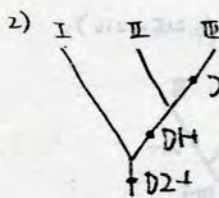
Tree 4  
S=34



を違はなければならぬ。これは cladogram の場合と同じである。(Table 3-20)  
 [ここで注意的なのは、上でも図示したように D-type の occurrences の記入法にはいくつかのやり方がお (Tree 1, 3 : 2通り ; Tree 2, 4 : 3通り)。けれどもこの方法を違えても全体のスコアには全く影響しない。つまり、"line" が形質の存在・欠除を示すだけ (cladogram, treeの場合) ならば、すべての方法は等価である。ところが phyletic interpretation によって "line" が時間軸を表わすようになる。このやり方を違ふかによって、解釈がすべて異なる。つまりその形質の出現時期が変わってくるのである。例えば tree 1 に対応する phyletic interpretation は次の2つである。



1) D1 は II, III の共通祖先に於て出現した。  
 D2 は I, II に於て、独立に生じた。



2) D1 は II, III の共通祖先に於て出現した。  
 D2 は D1 よりも古く、I, II, III の共通祖先に於て出現し III に於て失われた。

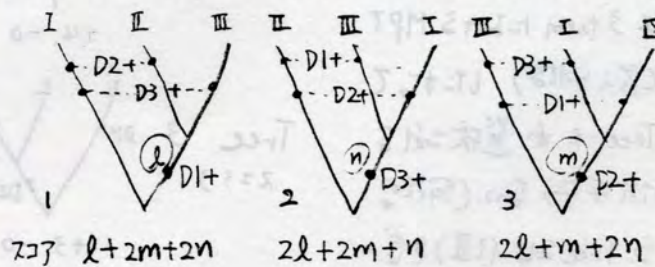
異なる

つまり、phyletic interpretation による限り、上の2つの evolutionary gains/losses が異なることになるが、phyletic part を除けば、即ち positive/negative occurrences (のスコア) に関して言えば、その2つは全く等価である。]

それでは、D1, D2, D3 がすべてありとして、その数の比率が  $l:m:n$  である場合はどうなるか?

	D1	D2	D3
I	0	m	n
II	l	m	0
III	l	0	n

Trees



スコア  $l+2m+2n$        $2l+2m+n$        $2l+m+2n$

negative occurrences が無い方にスコアをとると上のようになる。このスコアから等しく  $2(l+m+n)$  を引くと右の方になる。

スコアの最も小さいものが MPPT となる

のであるから、

$l > m, n \Rightarrow -l < -m, -n$

$\therefore$  Tree 1 を選ぶ

	Tree 1	Tree 2	Tree 3
スコア	$l+2m+2n$	$2l+2m+n$	$2l+m+2n$
	$-2(l+m+n)$	$\downarrow$	$\downarrow$
スコア	$-l$	$-n$	$-m$

2)  $n > l, m \Rightarrow -n < -l, -m \therefore$  Tree 2 を選ぶ

3)  $m > l, n \Rightarrow -m < -l, -n \therefore$  Tree 3 を選ぶ

つまり、数が最も多い D-type character が決定要因となり、これに対応する tree が GMPT である。(Table 3.22) complete or incomplete  
[GMPT と MPT の使い分けについて: ある特定の character set をもってよく説明するのが MPT であり、その関係は 1対1 である。(Table 3.19)  
これに対し、ある character set が data とて与えられた時、それをもっともよく説明 complete

明らかなら GMPT である] Table 3.22.

hybridization が許されるなら reticulate tree が生じ、スコアの点だけからいえば、こちらの方がより parsimonious である。しかし、それは情報を要約しているのではなく、只反復しているだけだから、GMPT の候補にはならない (p.218, Fig. 3.29.)

D-type の

上で述べたように、複数の evolutionary gains/losses は同スコアを与え、しかも D-type についての losses を 0 にする解が存在する。このとき、

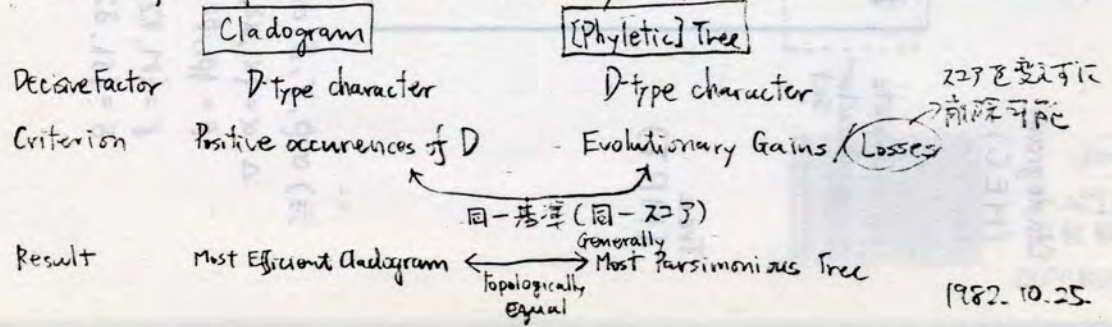
For three taxa, counting positive occurrences [of D-type] and counting evolutionary gains and losses [of D-type] give the same result, for no pattern of loss can give a total more parsimonious than a pattern of multiple gains. (p.216-217)

つまり、cladogram と [phyletic] tree とは全く同一基準で選択される。このとき、

The cladogram that is the most efficient summary merely represents the greatest number of characters of type D by the single line available for such representation. (p.218)

この基準はその子 tree にも応用できる。一言でいえば、

... the generally most parsimonious tree duplicates the cladogram previously accepted as the best summary. (p.216)

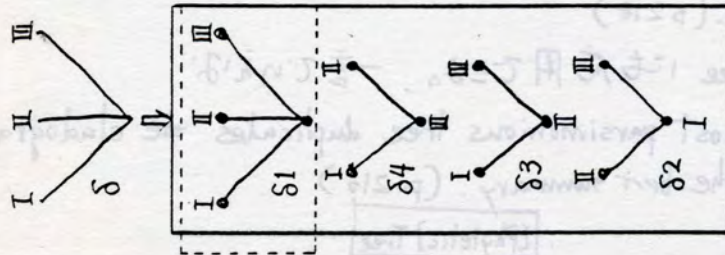
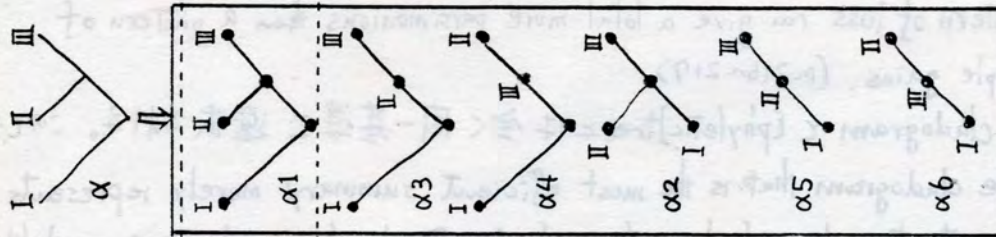
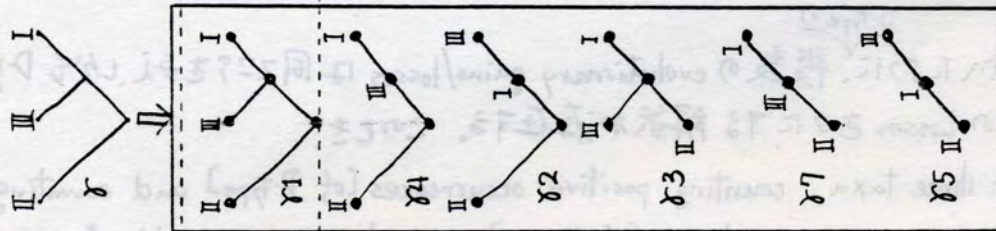
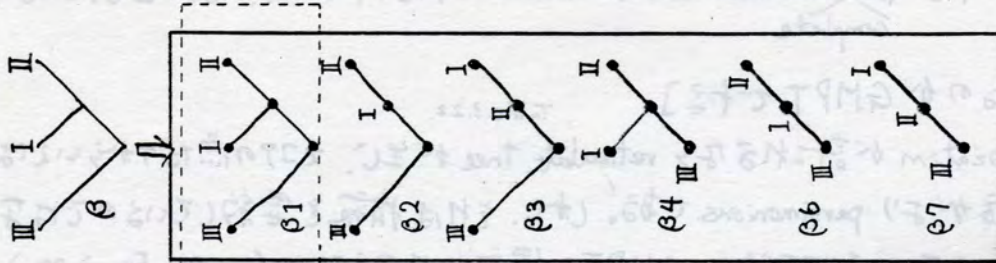


1982. 10. 25.

< Cladistic Analysis と Tree Analysis >

3-taxon problem に関して. cladistic analysis と tree analysis の比較を可る。  
 ⑩~⑫で見た通り. trivial なものを除く. 22個の character-sets (⑩~⑪)の表  
 に従い  $\alpha_1, \beta_1$  etc. と表記する)のそれぞれに対し. ある MEC と MPT が一つ  
 ずつ対応する. それを以下に示る.

MEC: Most Efficient Cladogram  
 MPT: Most Parsimonious Tree



Cladogram  
(MEC)

Complete  
Character-  
Set

Tree  
(MPT)

注)  $\alpha, \beta, \gamma, \delta$ : character-sets の集合

$\Delta \alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6\}$

$\beta = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6\}$

$\gamma = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6\}$

$\delta = \{\delta_1, \delta_2, \delta_3, \delta_4\}$



cladistic analysis と tree analysis について説明する。ある character-set (CS<sub>ij</sub>) が data として与えられた時、その情報を最も効率よく要約する cladogram (C<sub>i</sub>) — ΔMEC — が存在する ("cladistic analysis")。また、その data に対する最適な tree (T<sub>ij</sub>) — ΔMPT — も存在する ("tree analysis")。そして、ここで T<sub>ij</sub> ∈ C<sub>i</sub> であることが証明できる。何故なら、CS<sub>ij</sub> が含む D-type character (つまり、α, β, γ 系列ならば、それぞれ D1, D2, D3 type, δ 系列ならば、無し) によって選ばれた C<sub>i</sub> に属する全ての T<sub>i</sub> は、その D-type character のスコアが等しく且つ他系列に属するものよりも低いからである。だから、D-type 以外の形質のスコアを調べれば、T<sub>ij</sub> ∈ C<sub>i</sub> が全形質に関し最小のスコアを持っているはずである。それ故 T<sub>ij</sub> ∈ C<sub>i</sub> 即ち、MPT ∈ MEC となる。[なお、上で定義した character-sets の集合 α, β, γ, δ は次のように再定義できる。

$$\alpha = \{ \text{D1-type character を含んでいる CS} \}$$

$$\beta = \{ \text{D2-type character を含んでいる CS} \}$$

$$\gamma = \{ \text{D3-type character を含んでいる CS} \}$$

$$\delta = \{ \text{D-type character を含まない CS} \}$$

次に、character-sets の集合 CS<sub>i</sub> を、次式で定義する。

$$CS_i = \{ CS_{ij} \mid j=1, 2, \dots, n_i \}$$

そして、

$$j=1 \Rightarrow CS_{ij} \text{ は complete CS}$$

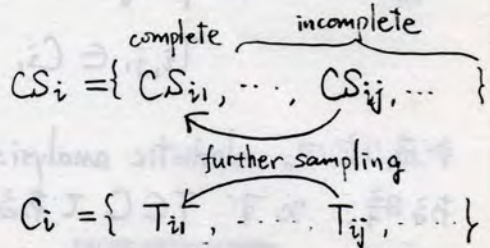
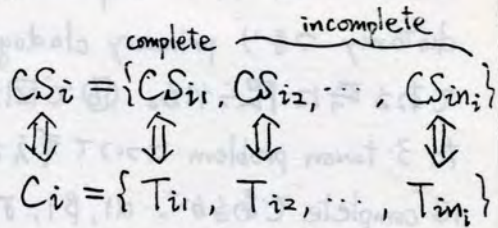
$$j \neq 1 \Rightarrow CS_{ij} \text{ は incomplete CS}$$

とする。右図からわかるように、 $\forall i$

, CS<sub>i</sub> に対し、ある MEC, C<sub>i</sub> が存在し、 $\forall j$ , CS<sub>ij</sub> に対し、ある MPT, T<sub>ij</sub> が存在する。 $\forall i$  に対し、complete CS, CS<sub>i1</sub> に対応する T<sub>i1</sub> と CS<sub>i</sub> に対応する C<sub>i</sub> とが位相的に一致することは既に見た。

形質の sampling をより広範に行なうならば、incomplete CS に於て欠けていたタイプの形質が発見され、complete CS に移行する可能性がある。それによつて T<sub>ij</sub> は C<sub>i</sub> と位相同形の T<sub>i1</sub> に変化する。

けれども、逆の場合、つまり、complete CS が further sampling により incomplete CS に変わることはあり得ない。complete CS に対応する T<sub>i1</sub>



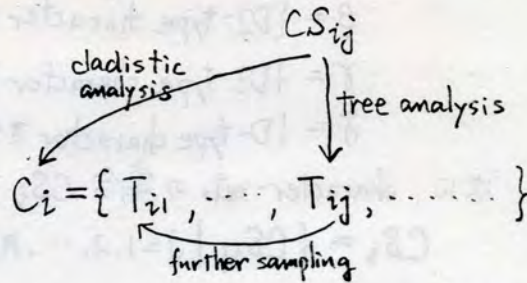
が GMPT (generally most parsimonious tree) と呼ばれ「一般解」とみなされるのは、その  $T_{i1}$  が、その時点で incomplete である  $T_{ij}$  ( $j \neq 1$ ) が将来の sampling により移行し得る「到達点」を示しているからである。

けれども、重要なことは、たとえ sampling の拡大により CS がより完全になつたとしても、対応する  $T_{ij}$  は全て同じ  $C_i$  に属するということである。つまり、sampling により  $j_1 \rightarrow j_2$  になつても、 $T_{ij_1} \in C_i$  ならば必ず  $T_{ij_2} \in C_i$  なのである ( $\rightarrow \textcircled{*}$ )。そして、Linnaeus 式の階層分類と 1対1の対応をもつのは cladogram であつて tree ではないのであるから、cladistic analysis は tree analysis よりも、より普遍性が高いのである。階層分類を前提とする限り、そして tree analysis の結果 (MPT) が常に MEC と位相同形の MPT を到達点とする限り、tree analysis は cladistic analysis を越えられない。右図参照

$\textcircled{*}$  sampling :  $j_1 \rightarrow j_2$  のとき、

$$T_{ij_1} \in C_i \Rightarrow T_{ij_2} \in C_i$$

いえるのは、 $C_i$  が complete dichotomy つまり primary cladogram である時に限られる。⑭で図示した



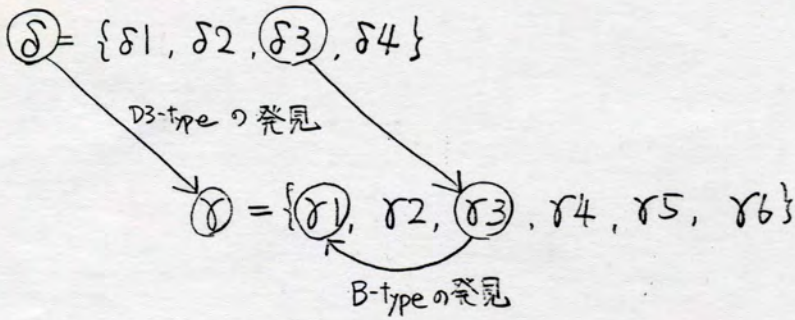
た 3-taxon problem について考えると、character-set  $\delta 1$  は  $\delta 2, \delta 3, \delta 4$  より complete であるが、 $\alpha 1, \beta 1, \gamma 1$  は complete ではない。何故なら、 $\delta 1$  は D-type character を持っていないからである。だから、D-type が発見されれば  $\delta$  系列の tree は その他の系列の tree に変わることになる。けれどもこの時は、cladogram そのものも変わり、その結果、変化した tree を含むことになる。つまり一般的に、sampling :  $i_1 \rightarrow i_2, j_1 \rightarrow j_2$  のとき、

$$T_{i_1 j_1} \in C_{i_1} \Rightarrow T_{i_2 j_2} \in C_{i_2}$$

が成り立つ。cladistic analysis の結果  $C$  と tree analysis の結果  $T$  があつた時、必ず  $T \in C$  であるという事実は何の変わりもない。

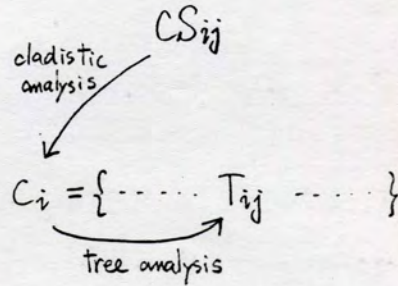
3-taxon の場合 tree  $\alpha 1, \beta 1, \gamma 1$  はそれ以上変化が起きない「到達点」であるが、 $\delta 1$  は更に  $\alpha 1, \beta 1, \gamma 1$  への移行が生じ得る。tertiary cladogram  $\delta$  は note 2 で見たとおり、「不完全な cladogram であり、 $\alpha, \beta, \gamma$  へ

移行し得る。だから、 $\delta$ 系列はその他の $\gamma$ 系列と区別して考えなければならぬ。



$\delta \ni \delta_3 \xrightarrow{\text{D3-type}} \gamma \ni \gamma_3 \xrightarrow{\text{B-type}} \gamma \ni \gamma_1$

character-set がどんなに変化しようと、常に  $MPT \in MEC$  が成立する。とすると、解析の順序としては、まずはじめに、 $CS_{ij}$  に対し cladistic analysis を行なって、 $C_i$  を決定し、(分類としては、ここまでで充分である) 次いで、必要ならば、他の情報(あるタイプの形質は存在しない)という証拠)をもとに、tree analysis を行なって、特定の  $T_{ij} (\in C_i)$  を決定するのが、自然である。



cladistic analysis があらかじめなされてこそ tree analysis は可能になる。この順序を逆転するのは不合理である。

1983. 10. 16.