note 5　p.238 ff.

—— Information and Components ——

今までは、2-taxon 或は 3-taxon について考えてきたが、4以上(一般にn) taxonについてはどうなるか? 例えば、4-taxonの場合、完全なcharacter-setに於ては右表のcharacter-typesが存在する。そして3-taxonに於けるD-type(D1, D2, D3)が果した役割(最適cladogramの決定)は、4-taxonの場合 E-type(E1〜E6)及び F-type(F1〜F4)によって行なわれる。

	A	B	C	D	E1	E2	E3	E4	E5	E6	F1	F2	F3	F4	G
I	+	−	−	−	+	+	+	−	−	−	+	+	+	−	+
II	−	+	−	−	+	−	−	+	+	−	+	+	−	+	+
III	−	−	+	−	−	+	−	+	−	+	+	−	+	+	+
IV	−	−	−	+	−	−	+	−	+	+	−	+	+	+	+

(p.239 Table 3.26)

そして3-taxonの場合cladogramの選択はD1, D2, D3の相対的な数により簡単にできるが、4-taxonの場合はそれほど単純ではない。

要するに、3-taxonでは極めて単純であったものが4(以上) taxonに於てはおそろしく難しくなるのである。けれども非常に嬉しいことに

Any complex problem, involving four or more taxa, *may* be reduced to a series of 3-taxon problems. (p.238)

だから、3-taxonを出発点として高次の問題は逐次的に解けるのであり、本節ではその過程を取扱うのである。　1982. 10. 28.

今、4-taxonの形質分布が下のようであったとする。表の数値はそれぞれのTypeの出現回数である。

(Table 3.28参照)

この表をもとに直接解析を行なおうとすると、上で述べたようにかなり困難が生じてくる。

	A	B	C	D	E1	E2	E3	E4	E5	E6	F1	F2	F3	F4	G
I	a	−	−	−	e_1	e_2	e_3	−	−	−	f_1	f_2	f_3	−	g
II	−	b	−	−	e_1	−	−	e_4	e_5	−	f_1	f_2	−	f_4	g
III	−	−	c	−	−	e_2	−	e_4	−	e_6	f_1	−	f_3	f_4	g
IV	−	−	−	d	−	−	e_3	−	e_5	e_6	−	f_2	f_3	f_4	g

そこでこの4-taxon problemを3-taxon problemに「還元」する必要がある。4-taxonから3-taxonを選ぶ組合せの数は $_4C_3 = 4$ であるから、上表から一行ずつ削除することにより、以下の表が得られる。

1) Ⅰ, Ⅱ, Ⅲ :

cf. Table 3.21

	A	B	C	D1	D2	D3	E
Ⅰ	$a+e_3$	—	—	—	e_1+f_2	e_2+f_3	f_1+g
Ⅱ	—	$b+e_5$	—	e_4+f_4	e_1+f_2	—	f_1+g
Ⅲ	—	—	$c+e_6$	e_4+f_4	—	e_2+f_3	f_1+g

2) Ⅰ, Ⅱ, Ⅳ :

cf. Table 3.30

	A	B	C	D1	D2	D3	E
Ⅰ	$a+e_2$	—	—	—	e_1+f_1	e_3+f_3	f_2+g
Ⅱ	—	$b+e_4$	—	e_5+f_4	e_1+f_1	—	f_2+g
Ⅳ	—	—	$d+e_6$	e_5+f_4	—	e_3+f_3	f_2+g

3) Ⅰ, Ⅲ, Ⅳ :

cf. Table 3.31

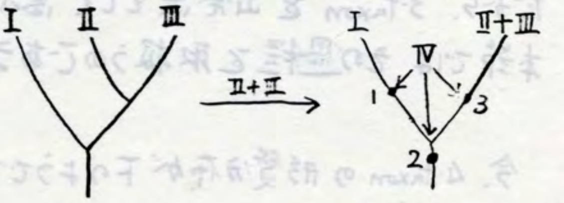
	A	B	C	D1	D2	D3	E
Ⅰ	$a+e_1$	—	—	—	e_2+f_1	e_3+f_2	f_3+g
Ⅲ	—	$c+e_4$	—	e_6+f_4	e_2+f_1	—	f_3+g
Ⅳ	—	—	$d+e_5$	e_6+f_4	—	e_3+f_2	f_3+g

4) Ⅱ, Ⅲ, Ⅳ :

cf. Table 3.32

	A	B	C	D1	D2	D3	E
Ⅱ	$b+e_1$	—	—	—	e_4+f_1	e_5+f_2	f_4+g
Ⅲ	—	$c+e_2$	—	e_6+f_3	e_4+f_1	—	f_4+g
Ⅳ	—	—	$d+e_3$	e_6+f_3	—	e_5+f_2	f_4+g

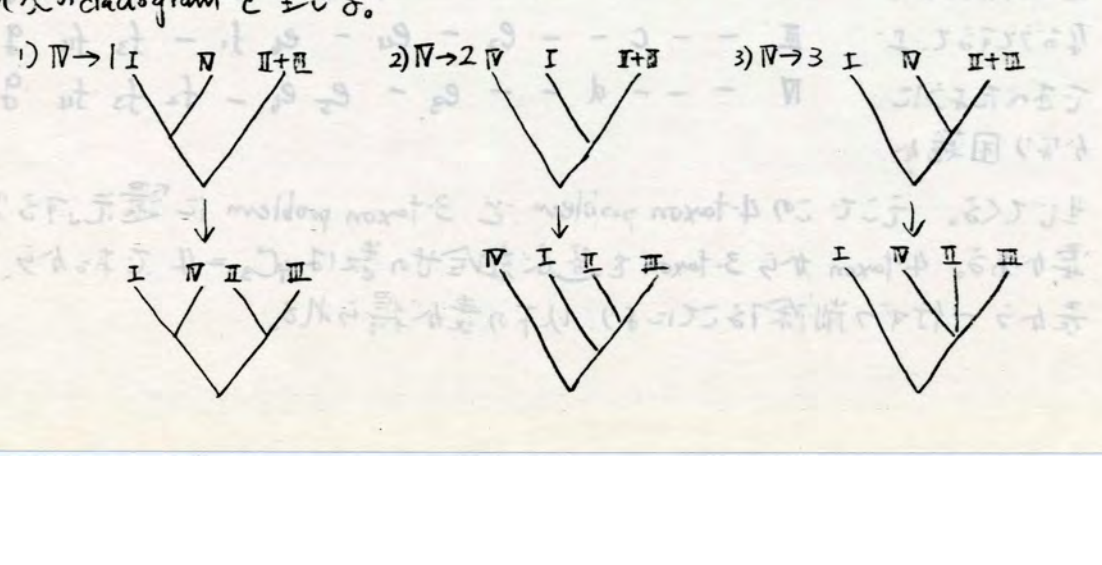
いま解析の出発点として、Ⅰ, Ⅱ, Ⅲ の 3-taxon に於て D1 > D2, D3 であったとする。このとき、左の 3-taxon cladogram が書けるが、その図から Ⅱ と Ⅲ は 1つの group を作っていると見なせる。そこで Ⅱ+Ⅲ という group を作ると結局 Ⅰ と Ⅱ+Ⅲ という 2-taxon cladogram ができる。



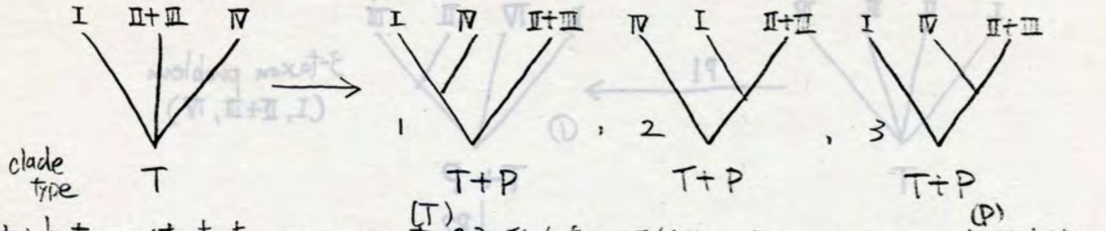
だから、4-taxon problem はまずはじめに

「3-taxon problem (Ⅰ, Ⅱ+Ⅲ, Ⅳ) を解くこと」 —— ①

を行なわねばならない。Ⅳ が入りこめる位置は 1, 2, 3 の 3ヶ所だけであり、それぞれ次の cladogram を生じる。



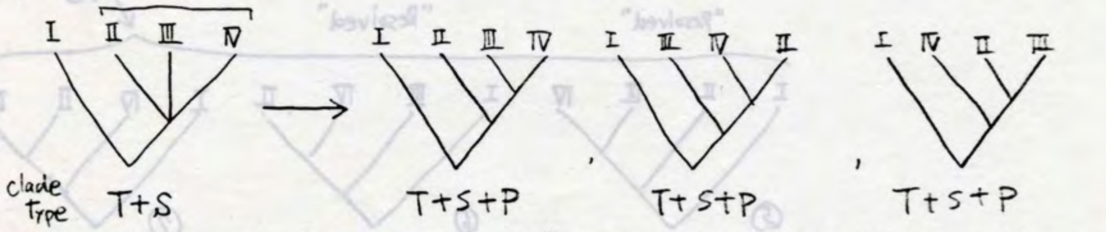
①の問題は trichotomy を dichotomy に移す問題に他ならない。



trichotomy は tertiary component (T) のみをもち、それに primary component (P) を付加すると dichotomy になるから、trichotomy は dichotomy の集合である。そして dichotomy のうち 1, 2 の cladogram は完全に解かれているが、3 は II, III, IV に関する trichotomy をもつ。だから 3 の場合に解くべき問題は

「3-taxon problem (II, III, IV) を解くこと」 —— ②

である。



問題②を解くには 2頁の表を見れば(表4)わかるが、①を解くにはもとの表の II, III のスコアを足して右表を導いてみなければならない。(Table 3.29)

	A	B	C	D1	D2	D3	E
I	a	—	—	—	$e_1+e_2+f_1$	e_3	f_2+f_3+g
II+III	—	$b+c+e_4$	—	$e_5+e_6+f_4$	$e_1+e_2+f_1$	—	f_2+f_3+g
IV	—	—	d	$e_5+e_6+f_4$	—	e_3	f_2+f_3+g

以上をまとめると次のようになる。4-taxon problem に対して解くべき 3-taxon problem を

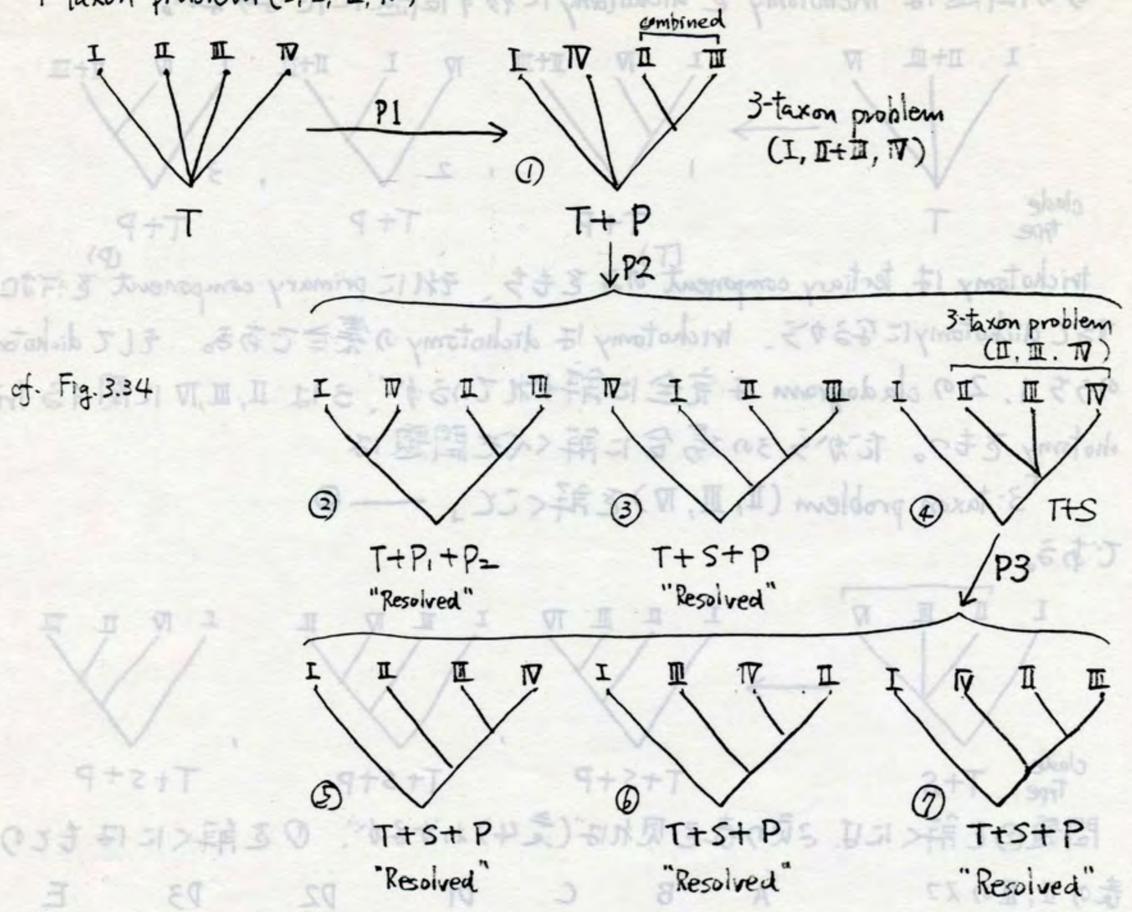
P1 : 3-taxon problem (I, II, III) を解く (出発点)

P2 : 3-taxon problem (I, II+III, IV) を解く (上の①)

P3 : 3-taxon problem (II, III, IV) を解く (上の②)

とおく。P1 の解は後の解析を進めるための出発点であるから、それは不可欠である。この「出発点」はこれ以外の (II, III, IV) (I, III, IV) (I, II, IV) のどれであってもよい。次の P2 の解は IV の「位置」を決める上で不可欠である。けれどもそこで決まった IV の位置により P3 が必要になるかどうかが変わってくる。以上をまとめたのが次の流れ図である。

4-taxon problem (I, II, III, IV)



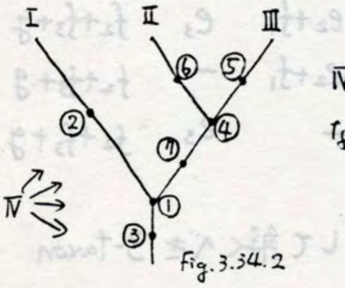
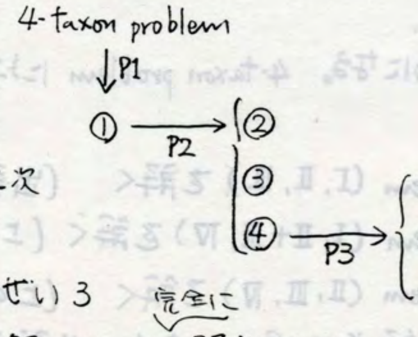


Fig.3.34.2

上の process は 結局 3-taxon cladogram のどの位置に IV が入るかということである。それを図示すれば左のようになる。番号はそれぞれ上の cladogram と対応している。



この process からわかることは次の2つである。

1) 4-taxon problem はせいぜい 3つの 3-taxon problem を解くことにより(完全に)解決される。つまり

cladogram ②,③ → P1, P2 を解く。

cladogram ⑤,⑥,⑦ → P1, P2, P3 を解く。(p.244)

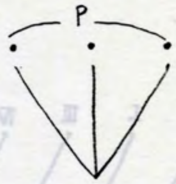
2) nonprimary "cladogram" ①,④ は primary cladograms の集合である。

これは p.177ff で既に説明されている。今の場合

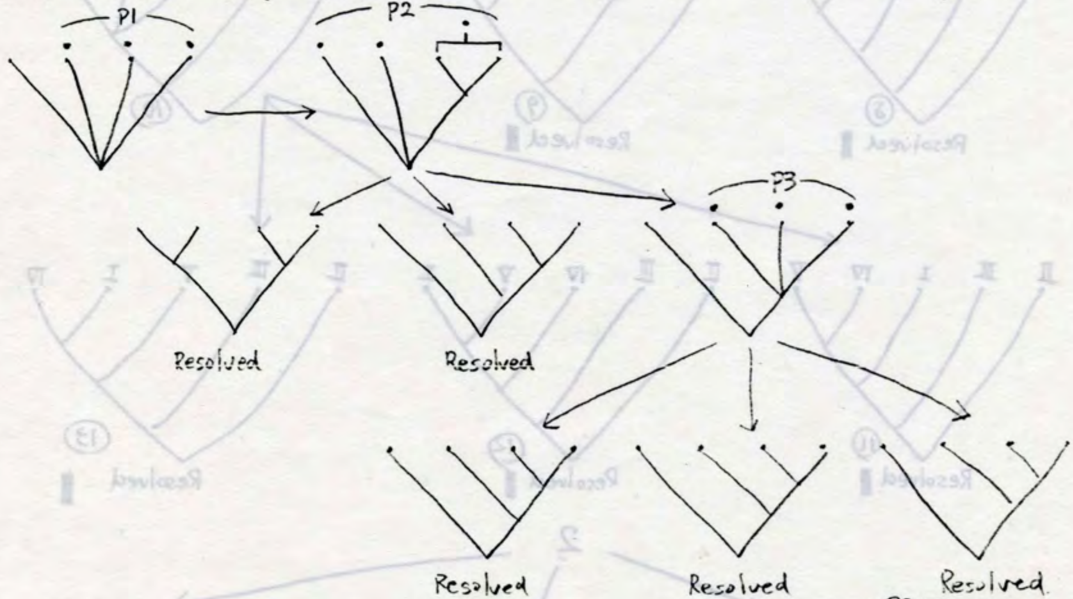
④ = {⑤, ⑥, ⑦}

① = {②, ③, ④} = {②, ③, ⑤, ⑥, ⑦}

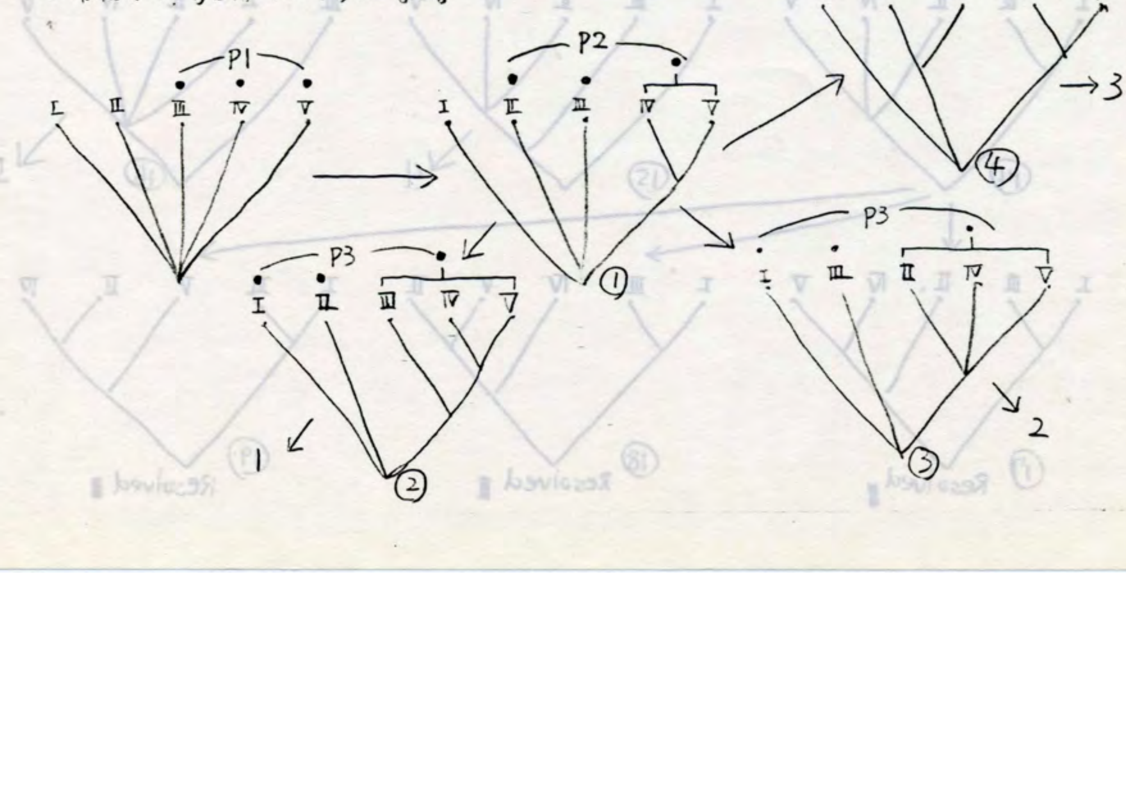
であり、確かに ①, ④は primary cladogramの集合である。これを information content の点から見ると (cf. p.182~183)、nonprimary cladogram は それ自身の actual information content (①: T+P ; ④: T+S) に加えて、それが規定する集合のmember がもち得る potential information content (set①: S, P ; set④: P) を含むことになる。(p.243~244)　1982.10.29

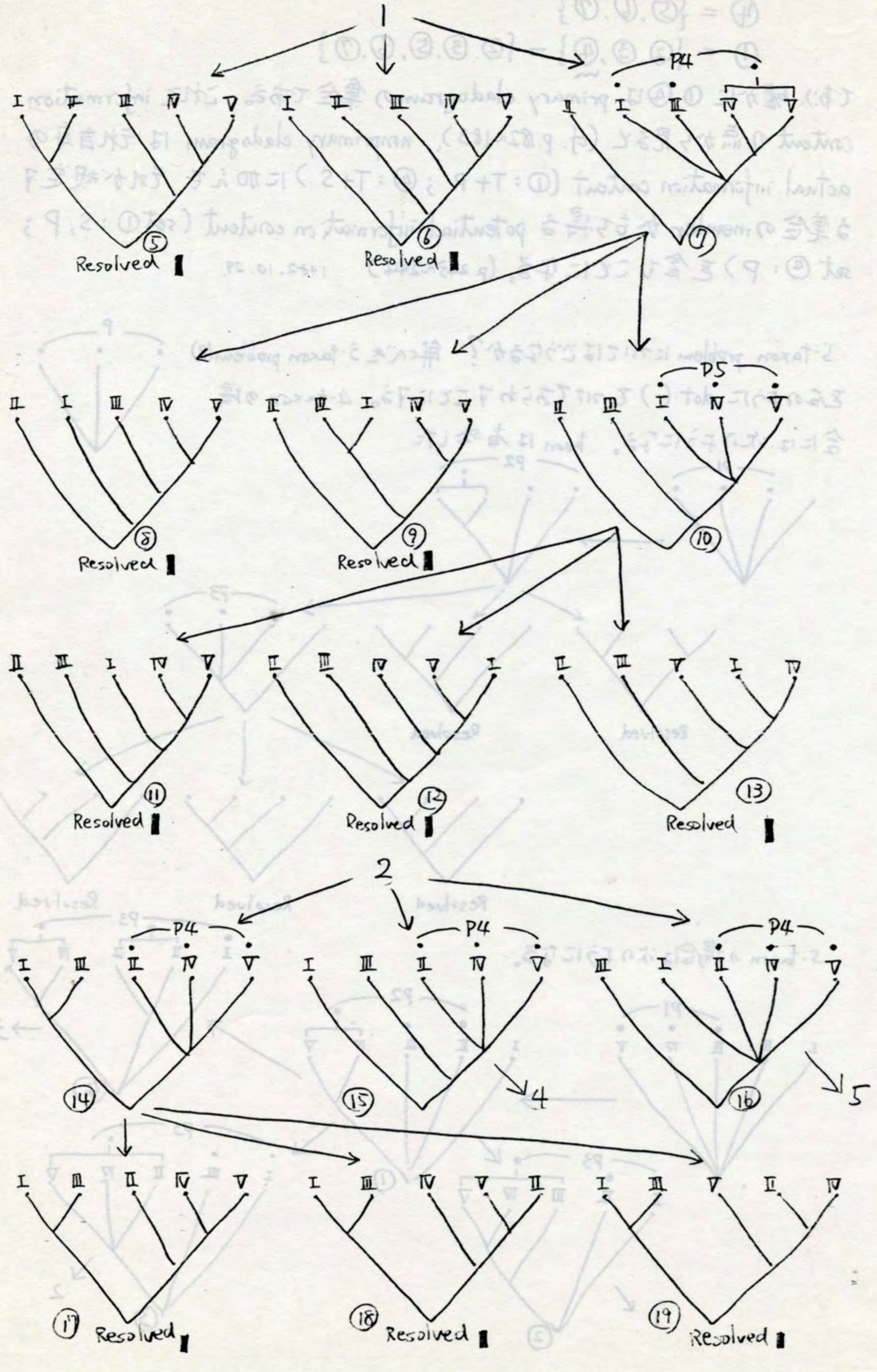


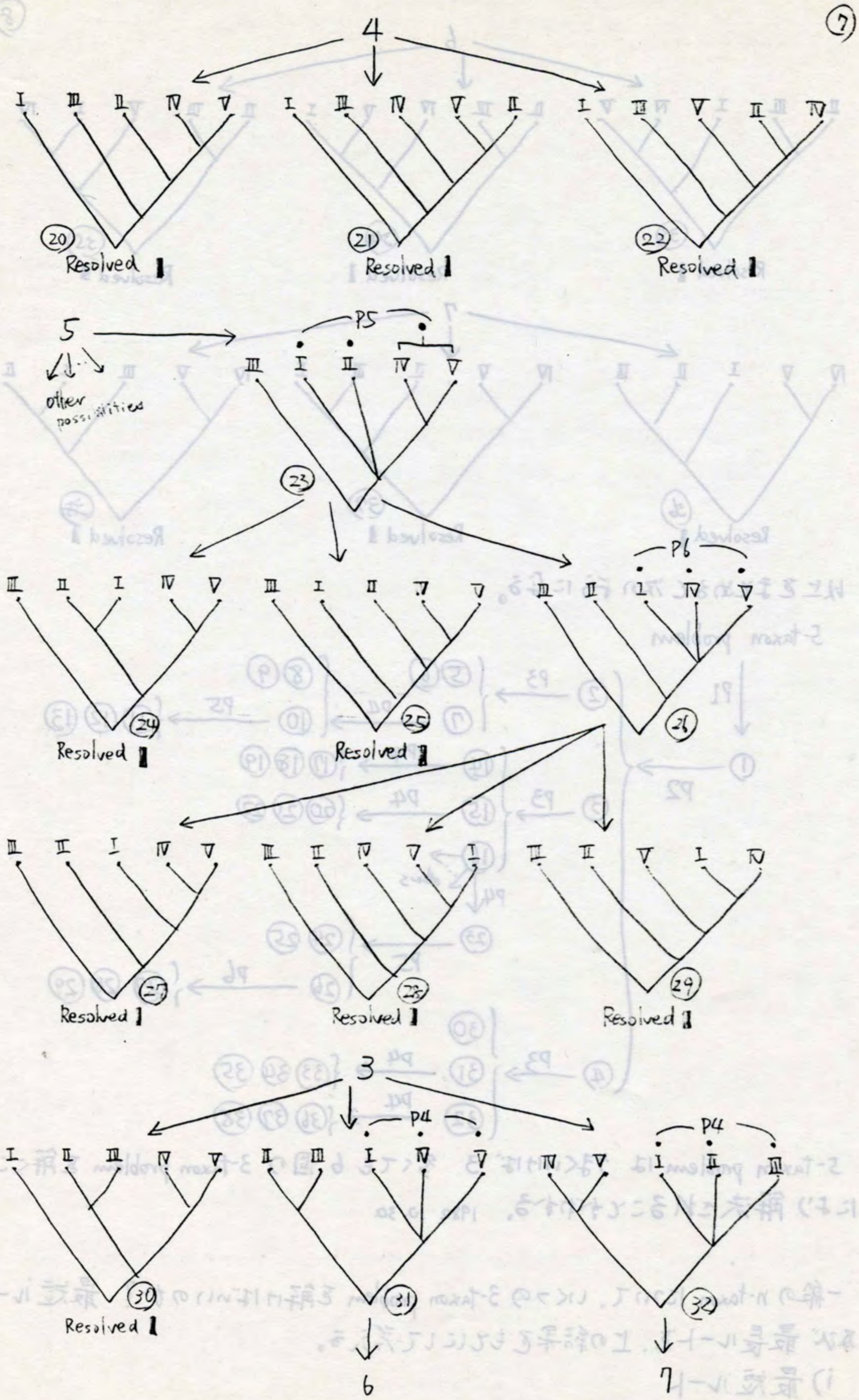
5-taxon problem についてはどうなるか? 解くべき 3-taxon problem(P) を左のように dot(・)をつけてあらわすことにする。4-taxon の場合には 次のようになる。taxon は省略した

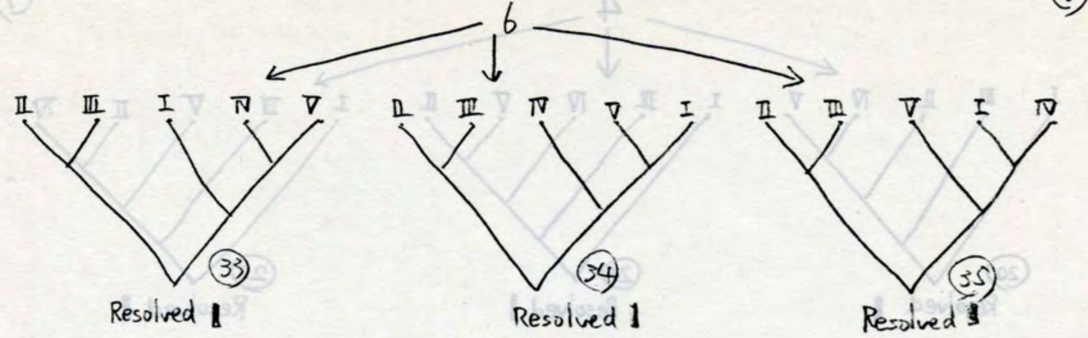


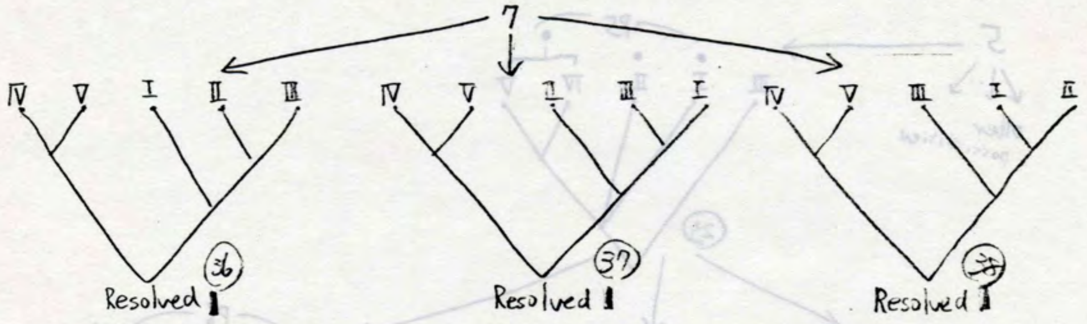
5-taxon の場合は次のようになる。




1
P4
I II III IV V
I II III IV V
II I III IV V
5 Resolved
6 Resolved
7
II I III IV V
II IV I III V
II III I IV V
P5
8 Resolved
9 Resolved
10
II III I IV V
II III IV V I
II III V I IV
11 Resolved
12 Resolved
13 Resolved
2
P4
I III II IV V
I III II IV V
III I II IV V
14
15
16
4
5
I III II IV V
I III IV V II
I III V II IV
17 Resolved
18 Resolved
19 Resolved

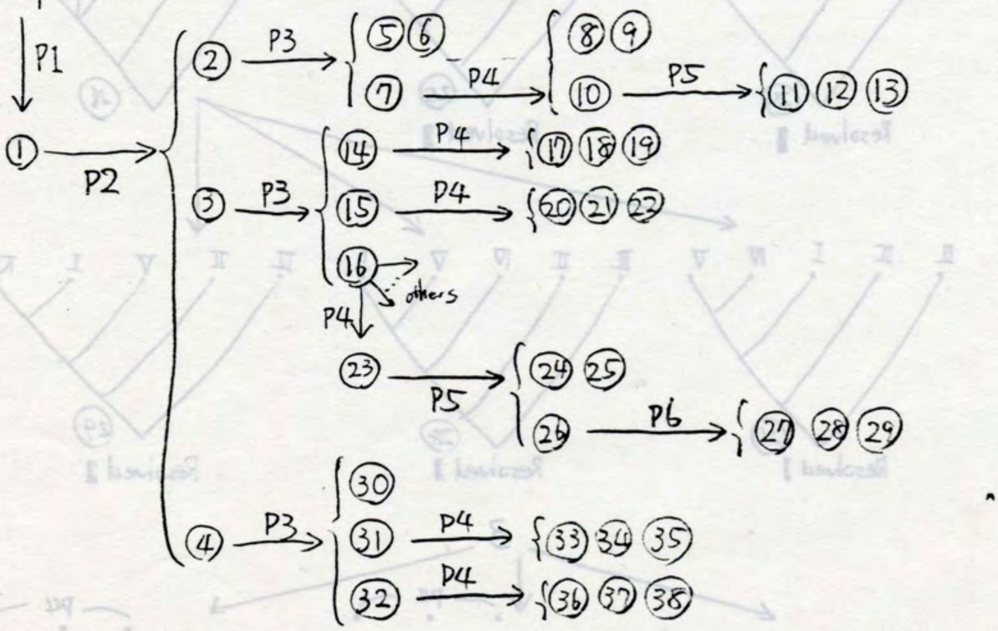

4
20
Resolved
21
Resolved
22
Resolved
5
other possibilities
P5
23
P6
24
Resolved
25
Resolved
26
27
Resolved
28
Resolved
29
Resolved
3
P4
P4
30
Resolved
31
32
6
7





以上をまとめると次のようになる。

5-taxon problem

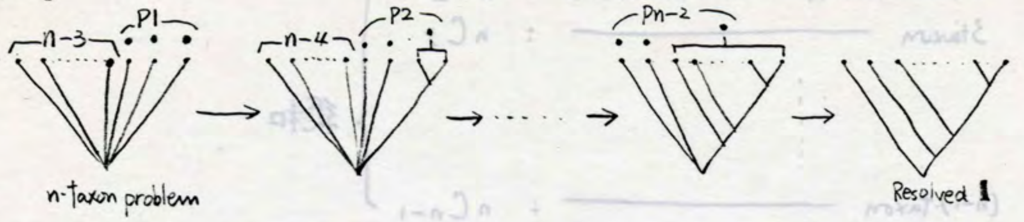


5-taxon problem は うまくいけば 3 多くても 6 個の 3-taxon problem を解くことにより解決されることがわかる。 1982. 10. 30.

一般の n-taxon について、いくつの 3-taxon problem を解けばいいのか？ 最短ルート及び 最長ルートを、上の結果をもとにして考える。

i) 最短ルート

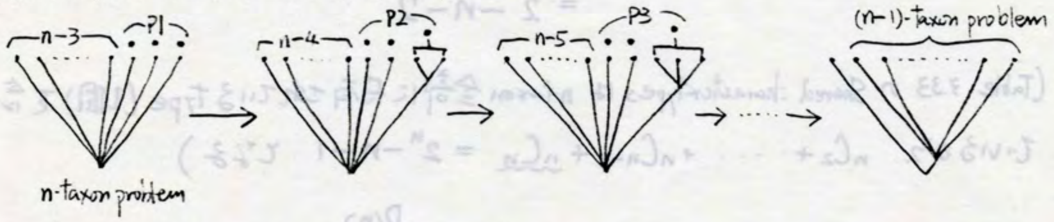
5-taxon の場合 ①→②→⑤or⑥ のルートがこれにあたり 3つの problem を解けばよい。



n-taxon の場合、上に示したように (n−2)個の 3-taxon problem を解けばよい。

ii) 最長ルート

5-taxon の場合 ①→③→⑯→㉓→㉑→㉗or㉘or㉙ のルートであって、6つの problem を解かなければならない。



n-taxon の場合 (n−2)個の 3-taxon problem を解いてはじめて右端の (n−1) taxon problem に導かれる。いま n-taxon problem を完全に解くのに必要な 3-taxon problem の最小値・最大値をそれぞれ $N_{min}(n)$, $N_{max}(n)$ とすると、

$$N_{min}(n) = n-2 \qquad n = 3, 4, \cdots$$

$$N_{max}(n) = n-2 + N_{max}(n-1)$$

後の漸化式は次のように解ける

$$\begin{aligned}
N_{max}(n) - \cancel{N_{max}(n-1)} &= n-2 \\
\cancel{N_{max}(n-1)} - \cancel{N_{max}(n-2)} &= n-3 \\
&\vdots \\
\cancel{N_{max}(4)} - \cancel{N_{max}(3)} &= 2 \\
+)\ \cancel{N_{max}(3)} - N_{max}(2) &= 1 \\
\hline
N_{max}(n) - \underbrace{N_{max}(2)}_{=0} &= \sum_{k=1}^{n-2} k = \frac{1}{2}(n-1)(n-2)
\end{aligned}$$

$$\therefore N_{max}(n) = \frac{1}{2}(n-1)(n-2)$$

3-taxon の場合 cladogram の選択に影響をもつのは D-type であり、その数は ${}_3C_2 = 3$ (Δ D1, D2, D3) であった。同じく 4-taxon ならば E-type (${}_4C_2 = 6$) と F-type (

$_4C_3=4$) の合計 10 であった。n-taxon ならば次のようになる。

2-taxon に共有されている character : $_nC_2$
3taxon ——————— : $_nC_3$
⋮
(n-1)-taxon ——————— : $_nC_{n-1}$

総和

二項展開式で $x=1$ とおくと

$$(1+1)^n = {}_nC_0 + {}_nC_1 + \underbrace{{}_nC_2 + \cdots\cdots + {}_nC_{n-1}} + {}_nC_n$$

$$\therefore {}_nC_2 + {}_nC_3 + \cdots + {}_nC_{n-1} = 2^n - {}_nC_0 - {}_nC_1 - {}_nC_n$$
$$= 2^n - n - 2$$

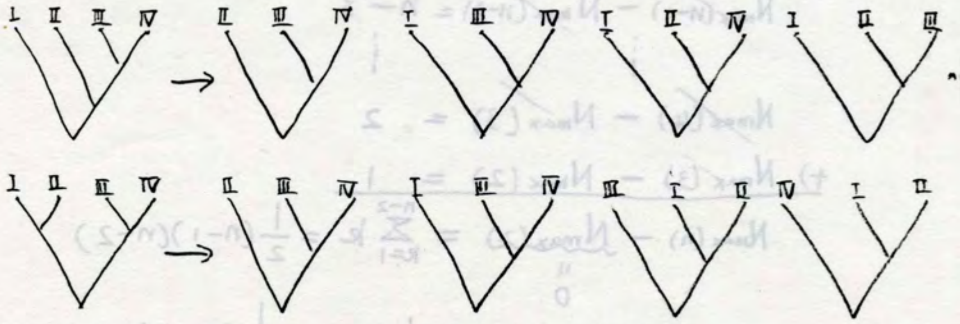
(Table 3.33 の Shared character types は n-taxon 全部に共有されている type (1個) を含んでいるから $_nC_2 + \cdots + {}_nC_{n-1} + \underline{{}_nC_n} = 2^n - n - 1$ となる)

n-taxon に対応する dichotomous cladogram の数($D(n)$)は Felsenstein [1978] によって次式で与えられている。

$$D(n) = (2n-3)\,D(n-1) \qquad \therefore D(n) = \frac{(2n-3)!}{2^{n-2}(n-2)!}$$

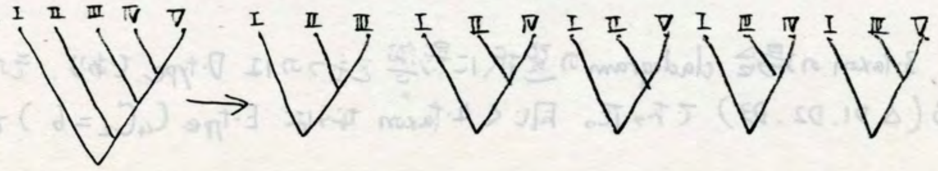
これでは n-taxon cladogram にはいくつの 3-taxon cladogram が含まれて(imply)いるか? 4-taxon の場合、$_4C_3 = 4$ 通の 3-taxon cladogram がある。

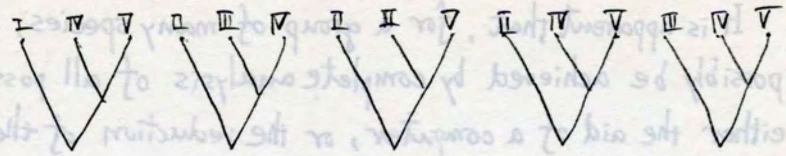
clade type



5-taxon の場合 $_5C_3 = 10$ 通の 3-taxon cladogram がある。

clade type





であるから何故 Table 3.33 の"Implied 3-taxon Cladograms"の項目に最大値最小値が存在し、その数値が表のようになるのかは現時点では不明である。12 ただし、そのような値の説明式は次のように書ける。

$$M_{max}(n) = {}_nC_3 - 1$$
$$= \frac{1}{6}n(n-1)(n-2) - 1$$
$$M_{min}(n) = {}_nC_3 - (n-2)$$
$$= \frac{1}{6}n(n-1)(n-2) - (n-2)$$
$$= \frac{1}{6}(n-2)[n^2 - n - 6]$$
$$= \frac{1}{6}(n+2)(n-2)(n-3)$$

以上から、Table 3.33 の各項目の一般式は次のようになる。

n-taxon problem に於ける

1) 3-taxon problem の数 $N_{max}(n)$, $N_{min}(n)$：

$$N_{max}(n) = \frac{1}{2}(n-1)(n-2), \quad N_{min}(n) = n-2$$

2) 暗示されている 3-taxon cladogram の数 $M_{max}(n)$, $M_{min}(n)$：

$$M_{max}(n) = \frac{1}{6}n(n-1)(n-2) - 1, \quad M_{min}(n) = \frac{1}{6}(n+2)(n-2)(n-3)$$

3) 共有されている character type の数 $C(n)$：

$$C(n) = 2^n - n - 1$$

4) 可能な dichotomous cladogram の数 $D(n)$：

$$D(n) = \frac{(2n-3)!}{2^{n-2}(n-2)!}$$

以上からわかることは、

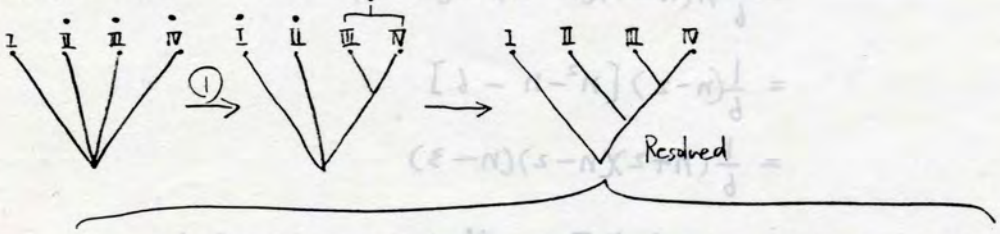
It is apparent that, for a group of many species, no cladogram could possibly be achieved by complete analysis of all possibilities without either the aid of a computer, or the reduction of the many species to a series of 3-taxon problems. (p.244~p.246). 1982.10.31.

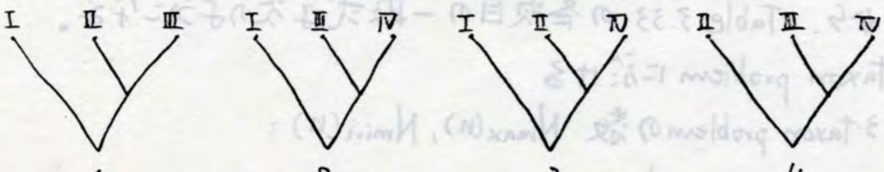
"Implied 3-Taxon Cladogram"について：

定義："Implied 3-Taxon Cladograms"

あるn-taxon problemがm個の3-taxon problemsの系列によって完全なn-taxon dichotomous cladogramになったとする。このn-taxon cladogramから演繹される$_nC_3$個の3-taxon cladogramsのうち、そのn-taxon cladogramの導出に用いられたl個の3-taxon cladogramsを除いた、$(_nC_3 - l)$個の3-taxon cladogramsを、このn-taxon cladogramの"Implied 3-Taxon Cladograms"という。

4-taxonの場合

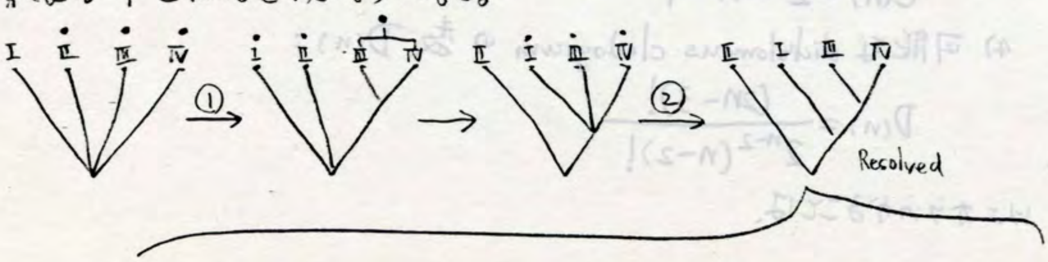


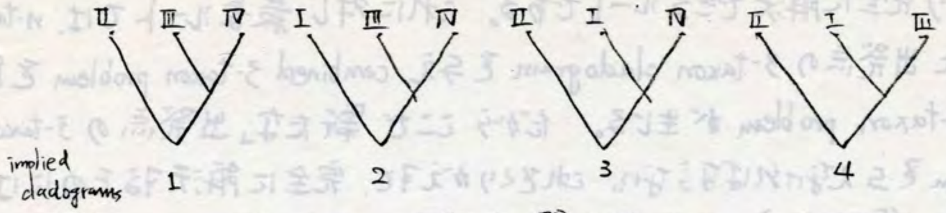


Implied Cladograms 1 2 3 4

この4-taxon dichotomous cladogramから演繹される$_4C_3=4$個の3-taxon cladograms (1—4)のうち4は除外できる。なぜなら、cladogram 4はもとの4-taxon cladogramの導出(problem ①)に関わっているからである。だから$l=1$であり、implied cladogramの数は$4-1=3$となる。このprocessは最短ルートをたどっている。

最長ルートをたどると次のようになる。



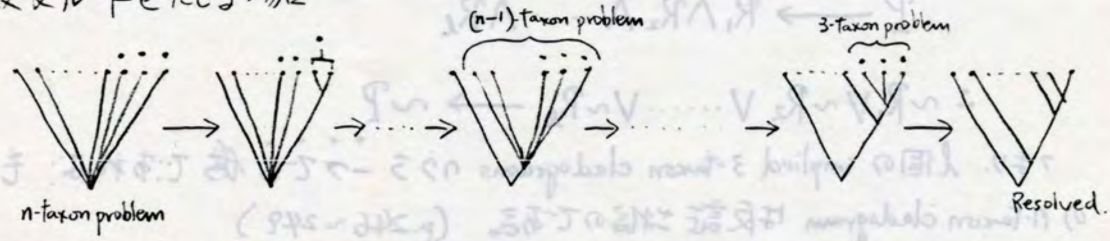


この場合、4-taxon cladogram の導出に関わる cladogram 1 (problem①) 及び cladogram 2 (problem②) は除外できる。従って implied cladogram の数は 4−2=2 である。

一般に n-taxon problem がその最短ルート ($N_{min}(n)$ の 3-taxon problems) により解決されたならば、出発点の 3-taxon cladogram を1つ与えればそれ以後は combined 3-taxon problem を解くことにより、完全に解決される。従って結果の n-taxon cladogram の $_nC_3$ 個の 3-taxon cladogram から出発点の cladogram を1つ除いたものが implied 3-taxon cladogram であり、このとき最大の数 $M_{max}(n)$ をとる。

$$M_{max}(n) = {}_nC_3 - 1 = \frac{1}{6}n(n-1)(n-2) - 1$$

最長ルートをたどる場合



つまり、出発点の 3-taxon cladogram を1つ与えて後の変形を与えると (n−1)-taxon problem が生じる。これを含めて (n−3) 個の 3-taxon cladogram を与えると、polychotomy が1つずつ減り、3-taxon problem が1つのこるから、更に1つ（総計 n−2）の 3-taxon cladogram を与えることにより完全に解決される。従って解決された n-taxon cladogram の $_nC_3$ 個の 3-taxon cladograms から、その導出に関わる (s-taxon problem (s=n, n−1, …, 3) の「出発点」である) (n−2) 個の cladogram を除外したものが implied 3-taxon cladogram であり、このとき最小の数 $M_{min}(n)$ をとる。

$$M_{min}(n) = {}_nC_3 - (n-2) = \frac{1}{6}(n+2)(n-2)(n-3)$$

まとめれば次のようになる。n-taxon problem (n-chotomy) が与えられたとき、その解決の「出発点」として最低1つの 3-taxon cladogram を与えなければならない。最短ルートとはこれさえ与えれば、その後は combined 3-taxon problem を解

くことにより完全に解決できるルートである。これに対し最長ルートでは、n-taxon problem に出発点の 3-taxon cladogram を与え combined 3-taxon problem を解くと (n−1)-taxon problem が生じる。だからここで「新たな」出発点の 3-taxon cladogram を与えなければならない。これをくりかえすて 完全に解決するためには (n−2)個の(「出発点」)3-taxon cladogram が必要なのである。

以上により ⑫ の疑問点は解決した。n-taxon dichotomous cladogram の導出に関わる 3-taxon cladogram の数を ℓ とすると 上述したように $1 \leq \ell \leq n-2$ だから、

$$\text{n-taxon problem} \xrightarrow{\ell \text{ 3-taxon cladograms}} \text{n-taxon cladogram} \xrightarrow{\text{deduction}} ({}_nC_3 - \ell) \text{ implied 3-taxon cladograms}$$

後半部分を命題として書くと

P : n-taxon cladogram は真である。

R_i: i-th implied 3-taxon cladogram は真である。$(i = 1, 2, \cdots, \ell)$

$$P \longrightarrow R_1 \wedge R_2 \wedge \cdots \wedge R_\ell$$

$$\therefore \sim R_1 \vee \sim R_2 \vee \cdots \vee \sim R_\ell \longrightarrow \sim P$$

つまり、ℓ 個の implied 3-taxon cladograms のうち一つでも偽であれば、もとの n-taxon cladogram は反証されるのである。(p.246〜249)

The minimum and maximum numbers of 3-taxon problems [$N_{min}(n)$ and $N_{max}(n)$] characterize what may be termed minimum and maximum "modes" of resolving dichotomous cladograms. (p.250)

ここでいう minimum mode や maximum mode は それぞれ 上で使ってきた最短ルート及び最長ルートに対応する。[なお Fig. 3.37 以下では 3-taxon problem の「結果」を示すための "•" が使われたが、上の記法では "•" は単に 3-taxon problem の「存在」を示すだけであった。統一をはかるため 以下では 本の記法に従う。]

6-taxon の場合 $N_{min}(n) = 6-2 = 4$, $N_{max}(n) = \frac{1}{2}(6-1)(6-2) = 10$ だから、minimum mode 及び maximum mode は それぞれ 4 及び 10 の 3-taxon problems を解く必要があり、Fig. 3.37 − 3.40 がそれを示している。それからわかることは、

The minimum mode is a unique and stepwise resolution, for it consists of particular 3-taxon problems solved in a particular order. (p.250)

The maximum mode is not a unique and stepwise resolution, as is the minimum mode, for the steps toward resolution need not follow in the same sequence. (p.251)

ある n-taxon problem が与えられた時、それを minimum mode で解決することが望ましいが、それを見分ける方法はあるのか？ 残念なことに、

Resolution of a cladogram in the minimum mode, if such could be done without advance knowledge of the final resolution, would require intuition equivalent in its effect of such knowledge. (p.252)

Minimum-mode resolution is sufficiently complex so that it probably can never be consistently achieved in practice. (p.253)

つまり、minimum-mode resolution は「直感」以外の方法では不可能ということである。けれども

Minimum-mode resolution seems always specifiable retrospectively. (p.253)

つまり、greater-than-minimum mode resolution によりある n-taxon cladogram に到達したならば、逆にその cladogram を与える minimum-mode resolution を指定できるのである。Cladogram はその components により指定されるのであるが、その component が解となるような 3-taxon problem を与えることにより、minimum-mode resolution に到達する。(Table 3-40) であるから、

The minimum mode, then, is a suite of 3-taxon problems that, once solved, result in the informative components of the cladogram that, with respect to a certain sample of information, is the true and final resolution. Thus a cladogram is definable in two different, but related, senses:

(1) as a suite of components;

(2) as a suite of 3-taxon problems for which the solutions are the suite of components. (p.254-255)

上のことは、完全に解決された dichotomous cladogram についていえることである。Minimum-mode resolution の場合、3-taxon problem と component とは一対一対応をする。そして未解決の cladogram では component の数はより少ないから、ある n-taxon dichotomous cladogram の components の最大数 $C'_{max}(n)$ は

$$C'_{max}(n) = N_{min}(n) + 1 = n - 1$$

component(1,2…n) (対応する problem がない)

Nondichotomous (未解決) cladogram の場合はどうなるか? (p.255ff.)
p.255の説明にあるように、nondichotomy の場合、ある component を解として与える 3-taxon cladogram が必ずしも存在しない。だから、(solution)

Thus it would seem that definition of cladograms is not generally possible in terms of 3-taxon problems, but is generally possible in terms of the solutions to such problems. If the solutions and components are compared, it is easy to see that the components are merely abbreviated forms of the solutions, with the addition of one component including all taxa. (p. 255)

つまり、dichotomous 及び nondichotomous cladogram を通して考えると、3-taxon problem の組として cladogram を表現するのは無理である。けれども component によればその表現は可能である。そして component とは solution の省略表現形に他ならない。(Table 3.40)

"Information of Cladogram"

n-taxon cladogram (dichotomous or not) の components のうち全ての taxa 1, 2, … n を含む component は、何ら informative ではないので これを除外すると "Informative Components" の数 $C(n)$ が得られる。このとき

$C(n)$: Component Information Table. 3.41

と定義すると、$C(n) \geq 0$ であり且つ

$$C_{max}(n) = C'_{max}(n) - 1 = N_{min}(n) = n - 2$$

つまり、Component Information は minimum-mode resolution と関係する。

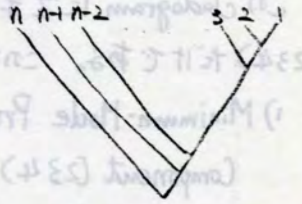
しかし Component Information $C(n)$ だけでは Cladogram の全情報 (Total

Information）を表現していない。何故ならC(n)はclade typeに関する情報のみを表わしているから。だからTaxaに関する情報が必要である。この情報を"Term Information" T(n)とする。その定義はどうするか？

もとにもどると、

Whereas the components relate to the number of minimum-mode problems [$N_{min}(n)$], the terms of a component relate to the number of maximum-mode problems [$N_{max}(n)$]. (p. 256) 1982.11.1.

n-taxon problemが完全に解決され且つそれがmaximum mode resolutionであったとすると、左の形のdichotomous cladogramができる。このときのcomponentsは



component 0 : $n, n-1, \cdots, 2, 1$

component 1 : $n-1, \cdots, 2, 1$

$\vdots$

component n-2 : $2, 1$

informative

このうちComponent 0はuninformativeであり除外される。とするとcomponentsの総数（Δ component information）C(n)は

$$C(n) = n-2 = N_{min}(n)$$

であり確かにcomponent informationはminimum-mode problemsの数$N_{min}(n)$と「関係」する。いま、Total Informationを

$$\text{Total Information} = \sum_i \text{No. of Terms of the } i\text{-th Component}$$

と定義すると、上のcladogramでは

$$\text{No. of Terms of the } i\text{-th Component} = (n-i)$$

$$\therefore \text{Total Information} = \sum_{i=1}^{n-2}(n-i) = (n-1)+(n-2)+\cdots+3+2 = \frac{1}{2}(n+1)(n-2)$$

そして更に

$$\text{Term Information} = \text{Total Information} - \text{Component Information}$$

と定義すると

$$\text{Term Information} = \frac{1}{2}(n+1)(n-2) - (n-2) = \frac{1}{2}(n-1)(n-2) = N_{max}(n)$$

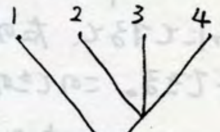
となり、確かにTerm Information maximum-mode problemsの数$N_{max}(n)$と「関係」する。

$$\text{Term Information} = \sum_i (\text{No. of Terms of Component } i - 1)$$

従って Component i の Term Information は Taxa (Terms) の数から1を引いたものである。

Nondichotomous cladogram も含め、より一般的に考える。まずはじめに、上の引用文の "minimum-mode problems" や "maximum-mode problems" とは一体何を意味するのか? 一例として右の cladogram を考える。(Fig. 3.42.1)

この cladogram に含まれる informative component は (234) だけである。このとき、

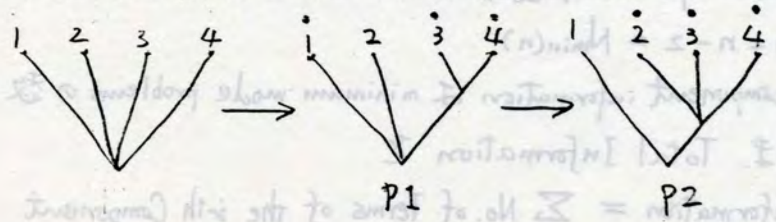


1) Minimum-Mode Problem (最短ルート)

Component (234) を与える 3-Taxon Problem は unspecifiable (Table 3.40) であるが、その解は 1(234) として存在する。とにかくこの解を与える Problem は1つだけである。これは Component (234) の存在のもとで retrospective にのみ考えられる。

2) Maximum-Mode Problem (最長ルート)

Component (234) の存在がわからないとき、上の cladogram に到達する為に必要な 3-Taxon Problems は、(前述したとおり)



となり 2つの 3-Taxon Problems を解かねばならない。

つまり、上の与えられた cladogram に至るまでには

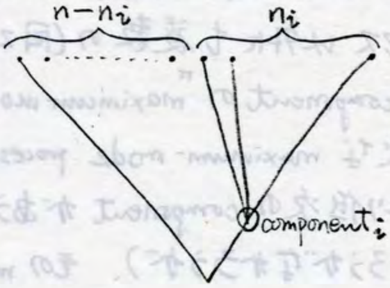
1) Minimum-Mode ⟶ 1つの 3-taxon problem (unspecified) を解く
2) Maximum-Mode ⟶ 2つの 3-taxon problems を解く

必要がある。そして、1で与えられた 3-taxon problem(s) を "Minimum-Mode Problem(s)"、2で与えられたものを "Maximum-Mode Problem(s)" と定義するのである。

上の例からもわかるが、特定の component (上例では (234)) に関し、minimum-mode problem の数は component の数と等しく1であり、maximum-mode problems の数は その component が含む Taxa (Terms) の数から1を引いたもの(上

例では 3−1=2）に等しい。言い換えれば、特定の component に関し component information は 1であり、minimum-mode problem の数に等しい。他方 term information（taxa − 1）は、maximum-mode problem の数に等しい。これは上の引用文の意味に他ならない。

一般的に言えば次のようになる。n-taxon cladogram (dichotomous or not) があるとき、その i 番目 ($i=1, 2, \ldots, n-2$) の Component が n_i 個の Terms (Taxa) を持つとする。

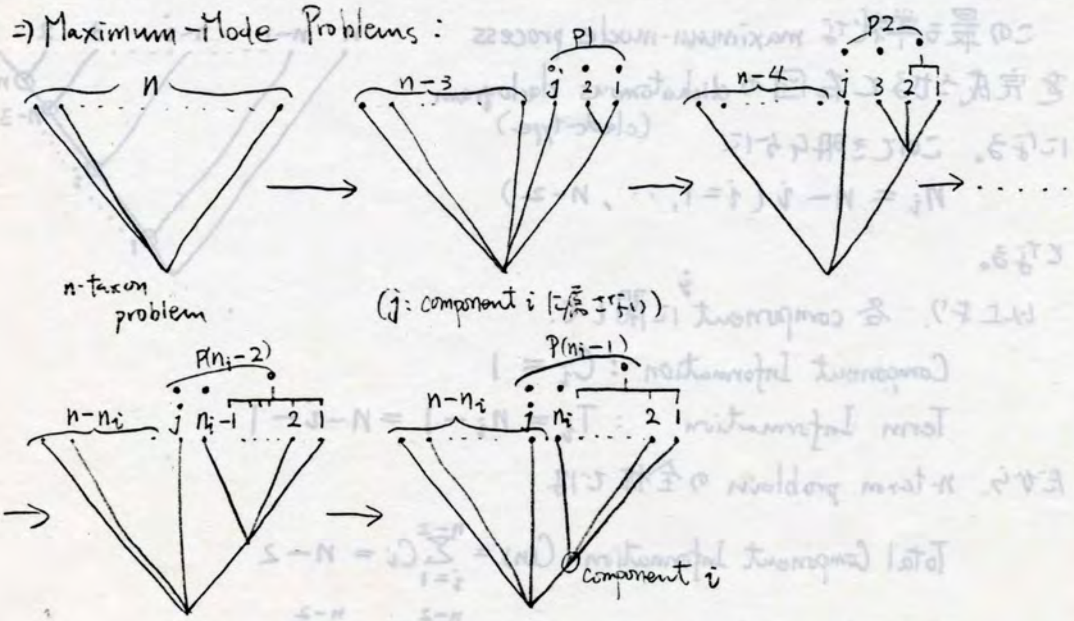


この Component i に関して

1) Minimum-Mode Problem：

Component i を解として与える 3-taxon problem を指定すればよいのだから、その数は 1 である。これは component の数 (1) と一致

2) Maximum-Mode Problems：

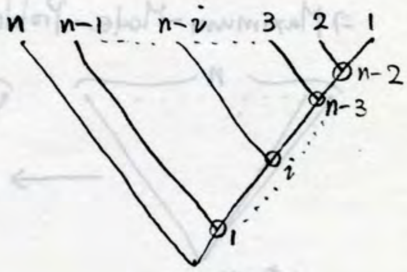


上からわかるように、maximum-mode で component i に到達するには (n_i-1) 個の 3-taxon problems が必要である。これは component i の terms 数 (n_i) から 1 を引いた数に等しい。ここでは component i に属する低次の components は存在しない（△ component i に関し n_i-taxon problem が生じる）場合を考えた。低次の components が存在するならば、component i に到達するまでのステップ数（Problem の数）は n_i-1 よりも大きくなる。けれどもここで注意すべきことは "maximum-mode" という場合、n-taxon problem を出発点と

して(if any)完全な dichotomous cladogram を到達点とする一連のプロセスの中で最もステップ数の多いものを指している。そして、ある特定の component の maximum-mode problems というのは、n-taxon problem から出発して(上述したように)、n_i-taxon problem に到る「最大」ステップ数を指している。maximum-mode process は(minimum-mode とは異なり)unique ではないから、今述べた最も単純なプロセス以外にも複数の(同ステップ数を持つ)ものが存在する。しかし上である component の"maximum-mode problems"というときは、もっぱらこの最も単純な maximum-mode process を指しているのである。だから、component i により低次の component があろうがなかろうが(△ component i が n_i-chotomy であろうがなかろうが)、その maximum-mode problems の数は $n_i - 1$ であると考えられる。つまり、低次の component に関係するステップは考えなくてよいのである。

HDAY!

この最も単純な maximum-mode process を完成させると右図の dichotomous cladogram (clade-type) になる。このとき明らかに



$$n_i = n - i \quad (i = 1, \cdots, n-2)$$

となる。

以上より、各 component i に関して、

Component Information : $C_i = 1$

Term Information : $T_i = n_i - 1 = n - i - 1$

だから、n-taxon problem の全体では

Total Component Information : $C(n) = \sum_{i=1}^{n-2} C_i = n - 2$

Total Term Information : $T(n) = \sum_{i=1}^{n-2} T_i = \sum_{i=1}^{n-2} i - (n-2)$

$$= \frac{1}{2}(n-1)(n-2)$$

となり、確かに次式が成り立つ。

$$C(n) = N_{min}(n), \quad T(n) = N_{max}(n)$$

ただし この等式は上の clade-type の時にのみ成立する。それ以外の clade-type の場合は一般に

$$C(n) < N_{min}(n), \quad T(n) < N_{max}(n)$$

となる。つまり、$N_{min}(n)$, $N_{max}(n)$ は Component Information 及び

Term Informationの上限を与えている。

もし、problemが完全に解決されていなくてnondichotomyがある場合、そのnondichotomous cladogramのComponent Information及びTerm Information $C'(n)$, $T'(n)$は次のようになる。

$$C'(n) < C(n) \leq N_{min}(n)$$

$$T'(n) < T(n) \leq N_{max}(n)$$

$$\therefore C'(n) + T'(n) < C(n) + T(n)$$

注：$T(n) \geq n-1$だから、明らかに
$T(n) > N_{min}(n)$

つまり、Total Information (C+T) に関し、完全に解決されたcladogramの方が未解決(一部解決)のものよりも大きい。

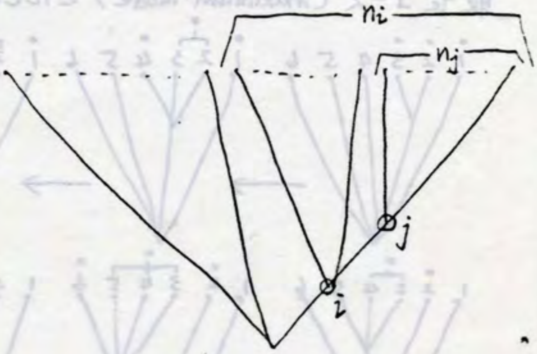
ある特定のn-taxon cladogram (dichotomous or not) に対して、

$C(n)$：そのcladogramに到達する最小ステップ数 (minimum-mode)
$T(n)$：―――――――――最大ステップ数 (maximum-mode)

であるから、$C(n)$と$T(n)$の平均は、何ステップでそのcladogramで到達するかの期待値をあらわす。この数を"Average Information"という。

$$\text{Average Information} = \frac{C(n) + T(n)}{2}$$

Component iとjがそれぞれn_i, n_jのtaxaを含むとする。component jはcomponent iに含まれているから、iに到る最長プロセスは$T_i = n_i - 1$で、更にjに到るには、それに加えて$T_j = n_j - 1$のプロセスを経る必要がある。だから、component iとjに到る最長プロセスは

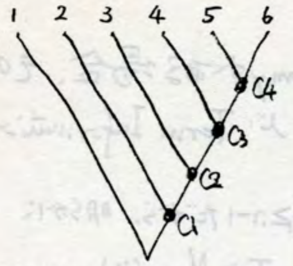


$$T_i + T_j = n_i + n_j - 2$$

である。つまりComponent間にはnested relationshipがあるから、$i<j$ならばComponent iに到るプロセスと独立にcomponent jに到るプロセスを考えられる。だから、相加性が生じるのである。

低次から高次のcomponentへプロセスを進めると最短コースになる。逆に高次から低次に進むと最長になる。

$n=6$の場合は次のようになる。(Fig. 3.37, 3.39)



Component 1 : 2−6

2 : 3−6

3 : 4−6

4 : 5−6

最短コース(minimum-mode)をたどると (Fig.3.37)

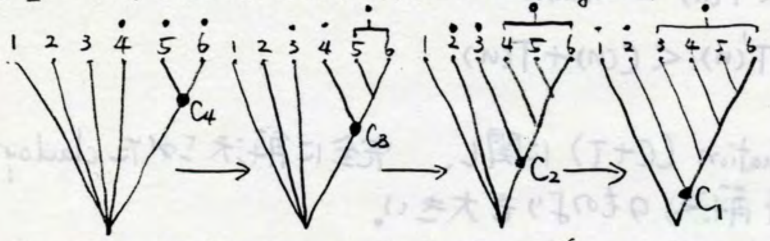
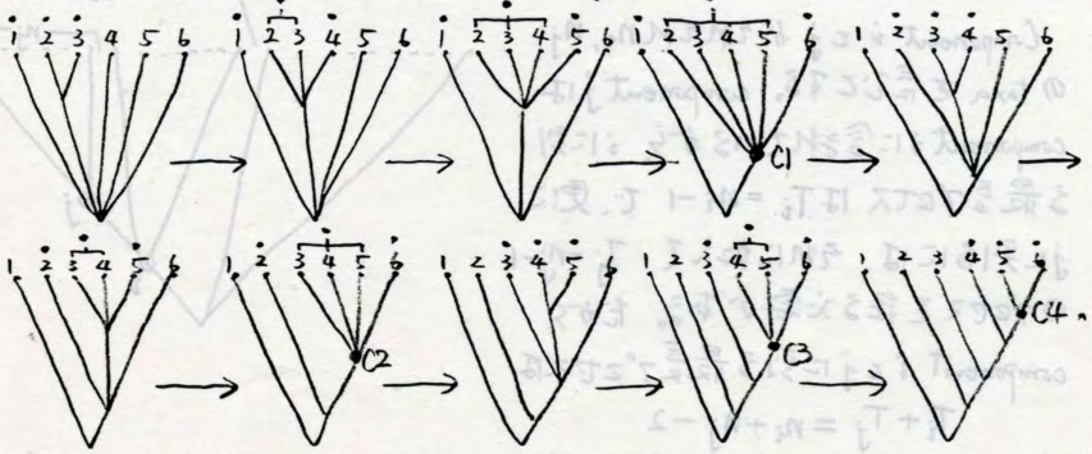


Table 3.40

これらより、それぞれのComponentを作る為の3-taxon problemを与えられる。

component	Solution	Problem	Ci minimum-mode problem
C4	4(56)	4.5,6	1
C3	3,(456)	3,4(5,6)	1
C2	2(3456)	2,3,(4,5,6)	1
C1	1(23456)	1,2,(3,4,5,6)	1

最長コース(maximum-mode)をたどると (Fig.3.39)

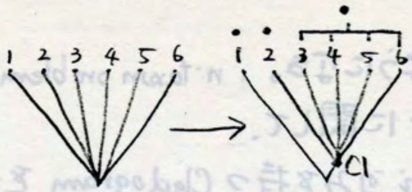


これらをもとにすると、それぞれのComponentを与えるmaximum-mode problemsの数 T_iがわかる

component C1 : $T_1 = 4$
C2 $T_2 = 3$
C3 : $T_3 = 2$
C4 : $T_4 = 1$

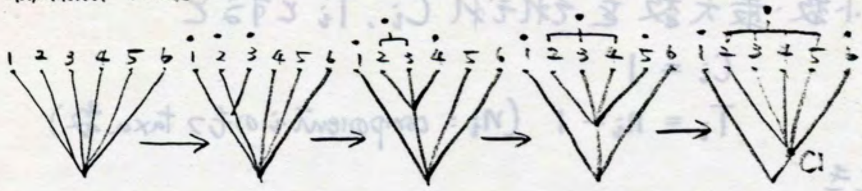
これらの値は上のprocessから数えたものであるがこれらはすべて独立に数えた時の値と一致する。

C1 Minimum-mode



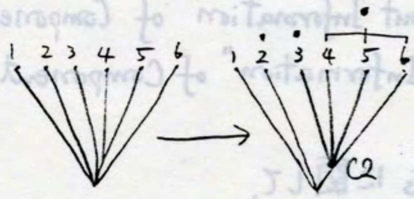
$C_1 = 1$

Maximum-mode



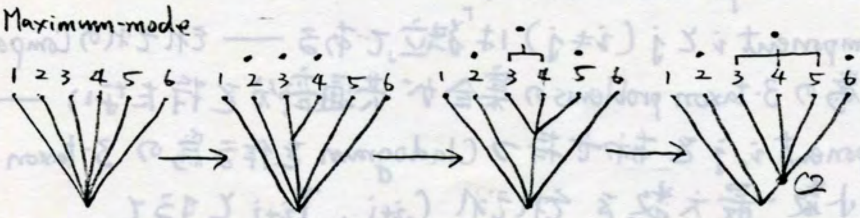
$T_1 = 4$

C2 Minimum-mode



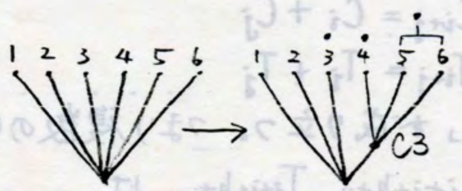
$C_2 = 1$

Maximum-mode



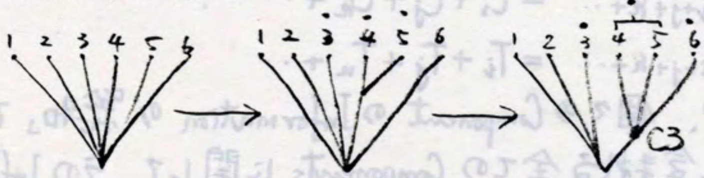
$T_2 = 3$

C3 Minimum-mode



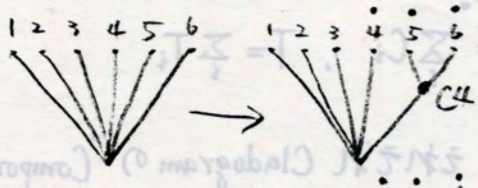
$C_3 = 1$

Maximum-mode



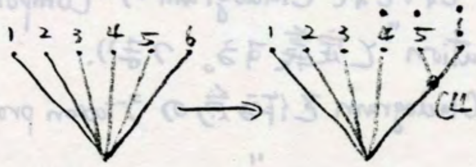
$T_3 = 2$

C4 Minimum-mode



$C_4 = 1$

Maximum-mode



$T_4 = 1$

確かに相加性に存在する。

以上をまとめると次のようになる。 n-taxon problem について

1) 個々の Component に関して、

ある Component i のみを持つ Cladogram を作る為の 3-taxon problems の最小数・最大数 をそれぞれ C_i, T_i とすると

$$C_i = 1$$
$$T_i = n_i - 1 \quad (n_i : \text{component } i \text{ のもつ taxa 数})$$

このとき

C_i : "Component Information" of Component i

T_i : "Term Information" of Component i

と定義する。

2) 複数の Components に関して、

Component i と j $(i \neq j)$ は「独立」である —— それぞれの Component を作る為の 3-taxon problems の集合が共通部分を持たない —— から、Component i, j をあわせ持つ Cladogram を作る為の 3-taxon problems の最小数・最大数を それぞれ C_{i+j}, T_{i+j} とすると

$$C_{i+j} = C_i + C_j$$
$$T_{i+j} = T_i + T_j$$

という「相加性」が成り立つ。つまり複数の (2 or more) Components の Information $C_{i+j+k+\cdots}$, $T_{i+j+k+\cdots}$ は

$$C_{i+j+k+\cdots} = C_i + C_j + C_k + \cdots$$
$$T_{i+j+k+\cdots} = T_i + T_j + T_k + \cdots$$

で示される通り、個々の Component の Information の「総和」である。

Cladogram に含まれる全ての Components に関して、その Information の総和をとる。

$$C = \sum_i C_i \quad , \quad T = \sum_i T_i$$

この C, T をそれぞれ Cladogram の "Component Information" 及び "Term Information" と定義する。つまり、

C : その Cladogram を作る為の 3-taxon problems の最小数

T : 〃 の最大数

そして更に

$$C \le N_{min}(n) = n-2$$

$$T \le N_{max}(n) = \frac{1}{2}(n-1)(n-2)$$

が成り立つ。つまり、CとTには上限がある。そして、n-taxon problemが完全に解決された — dichotomous cladogramが得られた — 時にのみ上限値 $N_{min}(n)$, $N_{max}(n)$が得られる。そして完全に解決されたcladogramは最も多くのInformationを持っているのだから㊟、もしnon-dichotomyがあれば、それだけInformationは少なくなることになる。このInformationの尺度がまさにC, Tである。この2つの尺度を加えて、そのcladogram (dichotomous or not)の"Total Information"と定義すれば

$$\text{Total Information}: C+T \le N_{min}(n)+N_{max}(n) = \frac{1}{2}(n+1)(n-2)$$

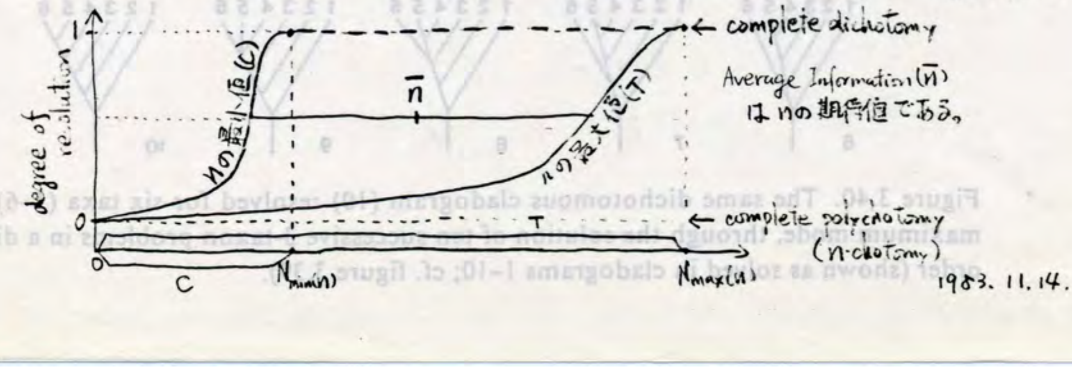
が成り立つ。cladogramが解決されていればいるほど(nondichotomyが少ないほど)、このTotal Informationは上限値に近づくのである。[もっとも dichotomous cladogramの"clade-type"によっては、上限値 $N_{min}(n)$, $N_{max}(n)$が得られないことがある。例えば Fig 3.41. 4 or 8 (cf. Table 3.43)。けれども、それぞれのclade-typeに対応する上限値を設定しておけば上の議論はそのまま成立する。]

あるcladogramを作る為の3-taxon problemsの最小数がC、最大数がTであるから、その平均をとれば一体いくつの3-taxon problemsを解けばいいのかという期待値が得られる。これを"Average Information"という。

$$\text{Average Information}: \frac{C+T}{2}$$

㊟ Cladogramに於ける「情報量」とは解かれた3-taxon problemsの「数」に他ならない。そしてComponent Information (C) 及び Term Information (T)はそれぞれこの「数」の最小値と最大値を与えているのである。

1982. 11. 2.



Average Information ($\bar{n}$) は nの期待値である。

1983. 11. 14.

Minimum Mode (4 steps)

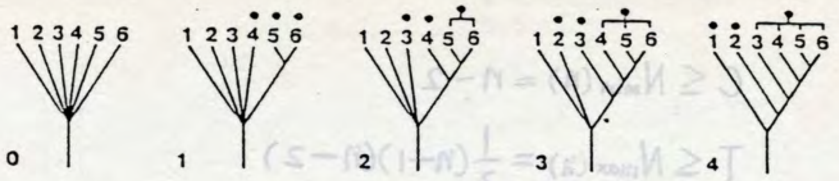


Figure 3.37. A dichotomous cladogram (4) resolved for six taxa (1–6) in the minimum mode, through the solution of four successive 3-taxon problems (shown as solved in cladograms 1–4).

Greater-than-Minimum Mode (5 steps)

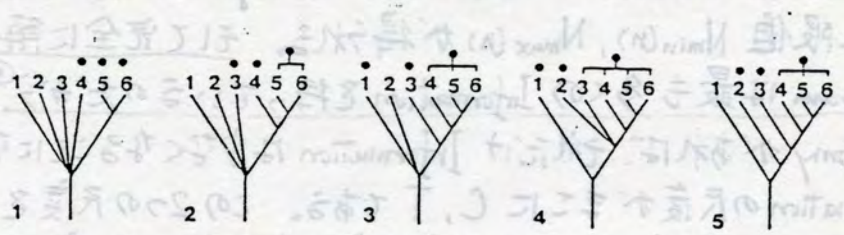


Figure 3.38. The same dichotomous cladogram (5) resolved for six taxa (1–6) in a greater-than-minimum mode, through the solution of five successive 3-taxon problems (shown as solved in cladograms 1–5; cf. figure 3.37).

Maximum Mode (10 steps)

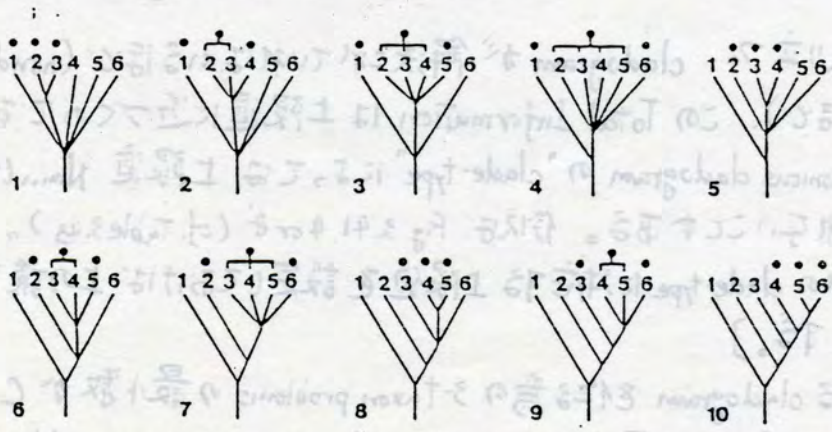
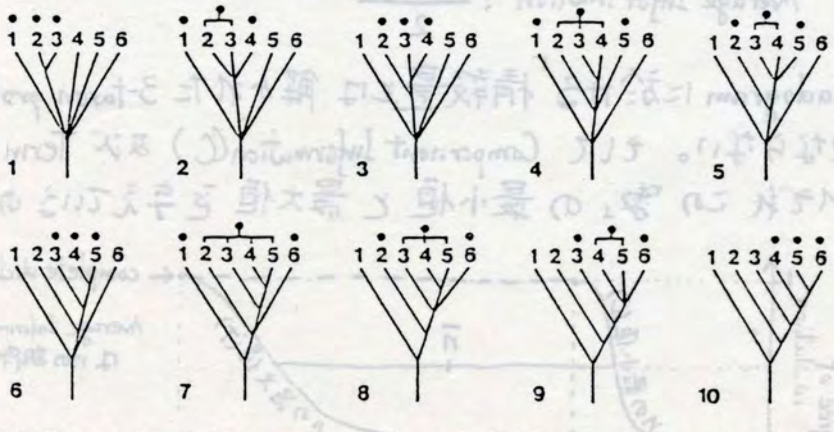


Figure 3.39. The same dichotomous cladogram (10) resolved for six taxa (1–6) in the maximum mode, through the solution of ten successive 3-taxon problems (shown as solved in cladograms 1–10; cf. figures 3.37–3.38).



Maximum Mode (10 steps)

Figure 3.40. The same dichotomous cladogram (10) resolved for six taxa (1–6) in the maximum mode, through the solution of ten successive 3-taxon problems in a different order (shown as solved in cladograms 1–10; cf. figure 3.39).

2 minimum mode

Fig. 37 ~ 3-40 と比 clade-type が多い

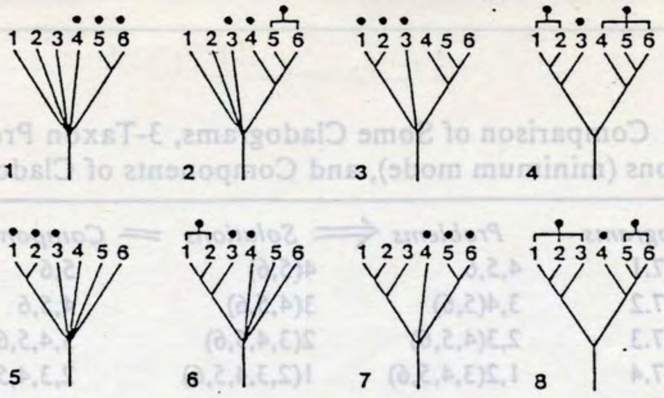


Figure 3.41. Cladograms (4, 8) resolved for six taxa (1–6) in the minimum mode, through the solution of suites of four successive 3-taxon problems (shown as solved in cladograms 1–4 and 5–8).

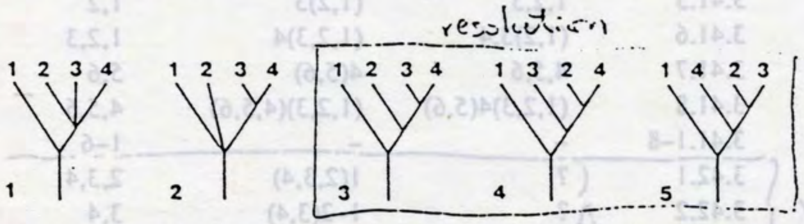


Figure 3.42. Cladograms (1–2) that represent solutions to unspecifiable 3-taxon problems in the minimum mode, with (1) or without (2) reference to final resolutions (3–5).

Table 3.33. Number of Taxa in Relation to Number of Possible 3-Taxon Problems (for completely dichotomous resolution), Number of Implied 3-Taxon Cladograms, Number of Character-Types Shared by Two or More Taxa, and Number of Possible Dichotomous Cladograms

	3-Taxon Problems		*Implied 3-Taxon Cladograms* → p. 247ff			
Taxa	*Minimum*	*Maximum*	*Maximum*	*Minimum*	*Shared Character-Types*	*Dichotomous Cladograms*
2	0	0	0	0	1	1
3	1	1	0	0	4	3
4	2	3	3	2	11	15
5	3	6	9	7	26	105
6	4	10	19	16	57	945
7	5	15	34	30	120	10,395
8	6	21	55	50	247	135,135
9	7	28	83	77	502	2,027,025
10	8	36	119	112	1013	34,459,425

Table 3.40. Comparison of Some Cladograms, 3-Taxon Problems and Solutions (minimum mode), and Components of Cladograms

Cladograms	*Problems* $\Leftarrow$	*Solutions* $=$	*Components*
3.37.1	4,5,6	4(5,6)	5,6
3.37.2	3,4(5,6)	3(4,5,6)	4,5,6
3.37.3	2,3(4,5,6)	2(3,4,5,6)	3,4,5,6
3.37.4	1,2(3,4,5,6)	1(2,3,4,5,6)	2,3,4,5,6
3.37.1–4	–	–	1–6
3.41.1	4,5,6	4(5,6)	5,6
3.41.2	3,4(5,6)	3(4,5,6)	4,5,6
3.41.3	1,2,3	(1,2)3	1,2
3.41.4	(1,2)3(4,5,6)	(1,2,3)(4,5,6)	1,2,3
3.41.5	1,2,3	(1,2)3	1,2
3.41.6	(1,2)3,4	(1,2,3)4	1,2,3
3.41.7	4,5,6	4(5,6)	5,6
3.41.8	(1,2,3)4(5,6)	(1,2,3)(4,5,6)	4,5,6
3.41.1–8	–	–	1–6
nondichotomy 3.42.1	?	1(2,3,4)	2,3,4
3.42.2	?	1–2(3,4)	3,4
3.42.1–2	–	–	1–4

unspecifiable

Table 3.41. Total Components, Informative Components, and Component Information of Some Cladograms (cf. figure 3.41)

Cladograms	*Total Components* $\longrightarrow$	*Informative Components*	*Component Information*
3.41.1	2: $\boxed{1–6}$; 5,6	1:5,6	1
3.41.2	3: $\boxed{1–6}$; 4–6; 5,6	2: 4–6; 5,6	2
3.41.3	4: $\boxed{1–6}$; 1,2; 4–6; 5,6	3: 1,2; 4–6; 5,6	3
3.41.4	5: $\boxed{1–6}$; 1–3; 1,2; 4–6; 5,6	4: 1–3; 1,2; 4–6; 5,6	4

omitted

Table 3.42. Total Components, Informative Components, Total Terms, and Term Information of Some Cladograms (cf. figure 3.42)

Cladograms	*Total Components*	*Informative Components*	*Total Terms*	*Term Information*
3.42.1	2: $\boxed{1–4}$; 2–4	1:2–4	3:2,3,4	2
3.42.2	2: $\boxed{1–4}$; 3,4	1:3,4	2:3,4	1

omitted

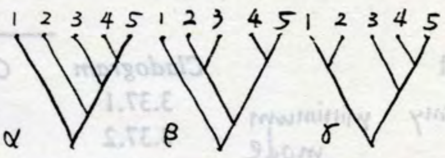
Table 3.43. Component, Term, Total, and Average Information of Some Cladograms (cf. figures 3.37–3.42)

	Cladogram	Information			
		Component	Term	Total	Average
6-taxon o: dichotomy minimum mode	3.37.1	1	1	2	1.0
	3.37.2	2	3	5	2.5
	3.37.3	3	6	9	4.5
	0 3.37.4	4	10	14	7.0
Intermediate	3.38.1	1	1	2	1.0
	3.38.2	2	3	5	2.5
	3.38.3	3	6	9	4.5
	3.38.4	3	7	10	5.0
	0 3.38.5	4	10	14	7.0
maximum-mode(1)	3.39.1	1	1	2	1.0
	3.39.2	1	2	3	1.5
	3.39.3	1	3	4	2.0
	3.39.4	1	4	5	2.5
	3.39.5	2	5	7	3.5
	3.39.6	2	6	8	4.0
	3.39.7	2	7	9	4.5
	3.39.8	3	8	11	5.5
	3.39.9	3	9	12	6.0
	0 3.39.10	4	10	14	7.0
maximum mode(2)	3.40.1	1	1	2	1.0
	3.40.2	1	2	3	1.5
	3.40.3	2	3	5	2.5
	3.40.4	2	4	6	3.0
	3.40.5	2	5	7	3.5
	3.40.6	3	6	9	4.5
	3.40.7	3	7	10	5.0
	3.40.8	3	8	11	5.5
	3.40.9	3	9	12	6.0
	0 3.40.10	4	10	14	7.0
minimum mode(1)	3.41.1	1	1	2	1.0
	3.41.2	2	3	5	2.5
	3.41.3	3	4	7	3.5
	o 3.41.4	4	6	10	5.0
(2)	3.41.5	1	1	2	1.0
	3.41.6	2	3	5	2.5
	3.41.7	3	4	7	3.5
	o 3.41.8	4	6	10	5.0
4-taxon nondichotomy	3.42.1	1	2	3	1.5
	3.42.2	1	1	2	1.0
dichotomy	3.42.3	2	3	5	2.5
	3.42.4	2	3	5	2.5
	3.42.5	2	3	5	2.5

$\boxed{\text{note 5}}$ (continued from ㉕)

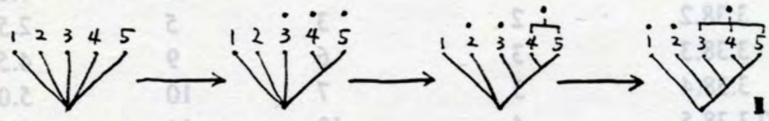
cladogram の "clade-type" が異なると、それに対応する "minimum-mode" 及び "maximum-mode" process はどのように変わってくるのか？ $n=5$ について調べる。

$n=5$ のとき、clade-type は右の α, β, γ である。
complete dichotomous clade-types
それぞれに対する minimum- and maximum mode は次の通り。



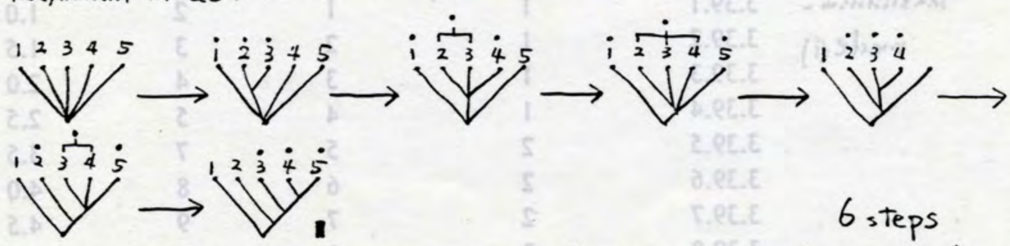
1) clade-type α

<minimum-mode>



3 steps

<maximum-mode>



6 steps

minimum- 及び maximum-mode problems の数をそれぞれ $N^{\alpha}_{min}(5)$, $N^{\alpha}_{max}(5)$ とすると、上より、

$$N^{\alpha}_{min}(5)=3, \quad N^{\alpha}_{max}(5)=6 \quad \text{(cf. Table 3.33)}$$

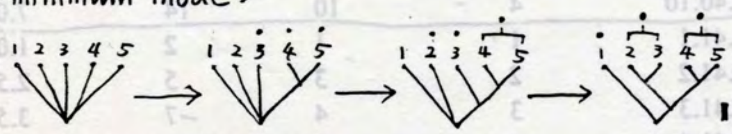
component 及び term information をそれぞれ $C^{\alpha}(5)$, $T^{\alpha}(5)$ とすると、

$$C^{\alpha}(5)=3, \quad T^{\alpha}(5)=6\ (=3+2+1)$$

$$\therefore\ C^{\alpha}(5)=N^{\alpha}_{min}(5), \quad T^{\alpha}(5)=N^{\alpha}_{max}(5)$$

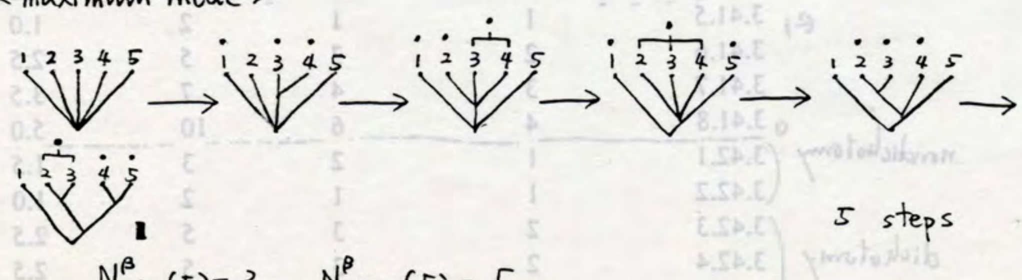
2) clade-type β

<minimum-mode>



3 steps

<maximum-mode>



5 steps

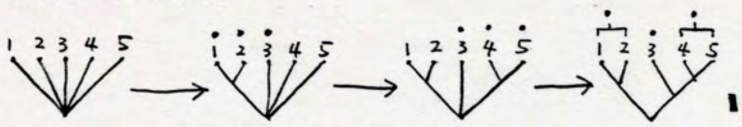
$$N^{\beta}_{min}(5)=3, \quad N^{\beta}_{max}(5)=5$$

$C^{\beta}(5) = 3, \quad T^{\beta}(5) = 5 \ (= 3+1+1)$

$\therefore C^{\beta}(5) = N^{\beta}_{min}(5), \quad T^{\beta}(5) = N^{\beta}_{max}(5)$

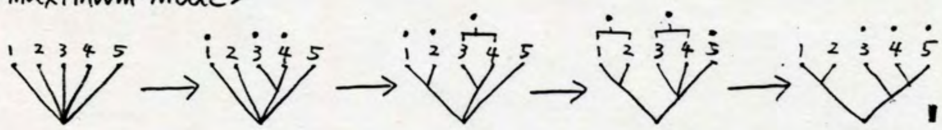
3) clade-type γ

<minimum-mode>



3 steps

<maximum-mode>



4 steps

$N^{\gamma}_{min}(5) = 3, \quad N^{\gamma}_{max}(5) = 4$

$C^{\gamma}(5) = 3, \quad T^{\gamma}(5) = 4.$

$\therefore C^{\gamma}(5) = N^{\gamma}_{min}(5), \quad T^{\gamma}(5) = N^{\gamma}_{max}(5)$

以上からわかることは、

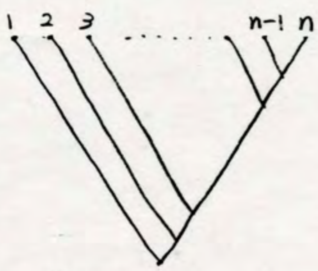
1. 各clade-typeの中では $C = N_{min}$, $T = N_{max}$ が成り立つ。つまり、component informationはminimum-modeに、そしてterm-informationはmaximum-modeに等しい。

2. complete dichotomous clade-typesでは、type間で $C (= N_{min})$ は等しいが、$T (= N_{max})$ は異なる。上の場合、

$$C^{\alpha}(5) = C^{\beta}(5) = C^{\gamma}(5), \quad T^{\alpha}(5) > T^{\beta}(5) > T^{\gamma}(5)$$

3. non-dichotomous clade-typesは、complete dichotomous clade-typesにくらべて、CとTがともにより小さい (Table 3-43)。だから、CとTはcladogramの「複雑さ」(complexity)を表わす尺度と考えられる。

4. 一般のnに対して、CとTが最大になる（それ故最も「複雑」な）clade-typeは右のものである。(n=5では上のtype α がこれにあたる) Table 3-33にはこの最も複雑なclade-typeに対応する数値が書かれており、それらはclade-typesを全て調べた場合の上限値となっている。



$C = n-2$

$T = \frac{1}{2}(n-1)(n-2)$

1983. 11. 16.