分子系統樹を用いた比較法と祖先形質復元：膨大な生物多様性情報を活用するために

奥山雄大

国立科学博物館　植物研究部

系統樹は何を語るか？

- 種間比較のツール：多様な生物をつなぐ『架け橋』
- 進化史上のイベントが、いつ、どこで、どのように起こったかが推定できる。
- 多様な種の形質を統計学的に扱うことで、形質進化の背景にある要因や一般則を理解できる。

Tree of Life Web Project
http://tolweb.org/tree/

本日のお話

- Mesquiteを使ってみよう
- Mesquiteによる祖先形質復元法
- （補足）BayesTraitsによる祖先形質復元法
- Mesquiteによる系統的独立比較法
- （補足）Rによる系統的独立比較法

Mesquiteをインストールしよう

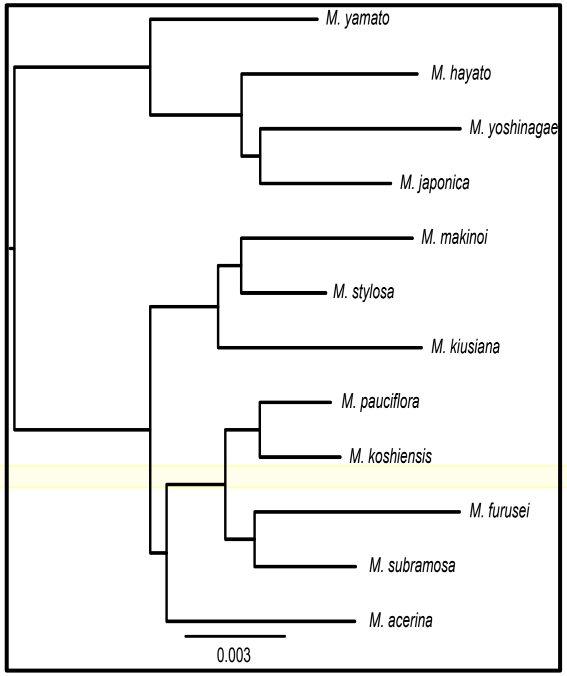
- http://mesquiteproject.org/mesquite/mesquite.html (“mesquite”でgoogle検索)
- 色んなことが手軽に出来る反面、一部の解析では専門のソフトウェアに劣る場合も

Mesquiteを走らせるために最低限必要なデータ

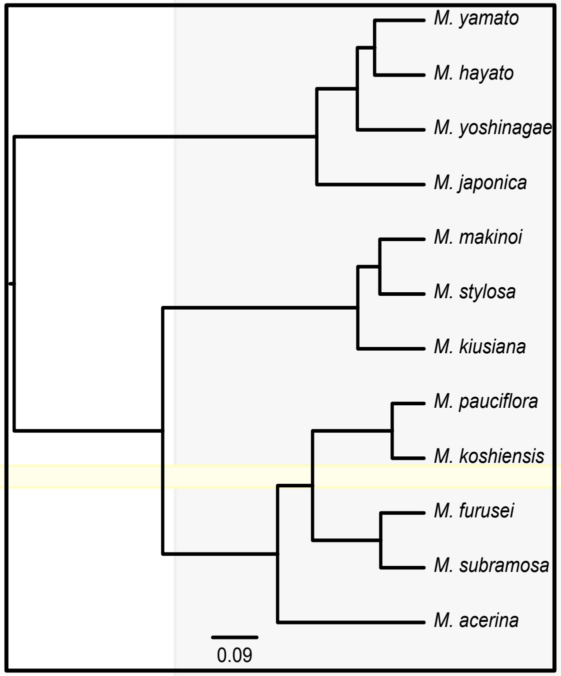
- 分子系統樹のツリーファイル（#Nexus形式の有根系統樹。タクソンブロック入り）
- これをMesquiteで開けばOK!

```
#NEXUS
Begin taxa;
    Dimensions ntax=12;
    Taxlabels
        M._YAMATO_A2
        M._japonica_T
        M._HAYATO_I
        M._yoshinagae_KW2
        M._stylosa_var.makinoi_Od1
        M._stylosa_var.stylosa_KZ1
        M._kiusiana_I
        M._pauciflora_A
        M._koshiensis_M
        M._furusei_var.furusei_MS
        M._acerina_AS
        M._furusei_var.subramosa_AS1
        ;
End;

BEGIN TREES;
    TRANSLATE
        1 M._YAMATO_A2,
        2 M._japonica_T,
        3 M._HAYATO_I,
        4 M._yoshinagae_KW2,
        5 M._stylosa_var.makinoi_Od1,
        6 M._stylosa_var.stylosa_KZ1,
        7 M._kiusiana_I,
        8 M._pauciflora_A,
        9 M._koshiensis_M,
        10 M._furusei_var.furusei_MS,
        11 M._acerina_AS,
        12 M._furusei_var.subramosa_AS1;
    TREE PAUP_1 = ((((1:0.1006189955,3:0.1006189955):0.03511267323,4:0.1357316687):0.08238870409,2:0.2181203728):0.6152104563,(((5:0.
END;
```



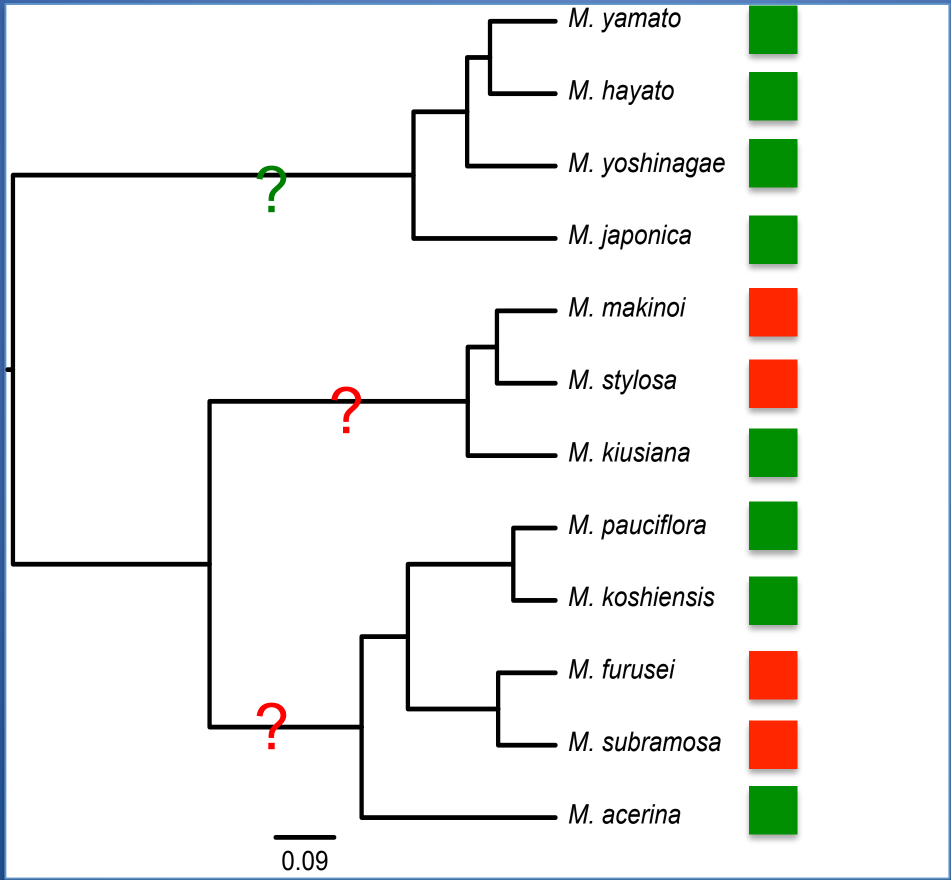
phylogram



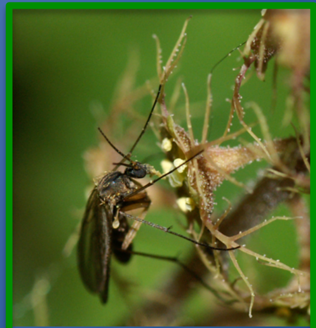
chronogram

祖先形質の復元（推定）

- ある形質がどの枝で、何度進化したかを推定する



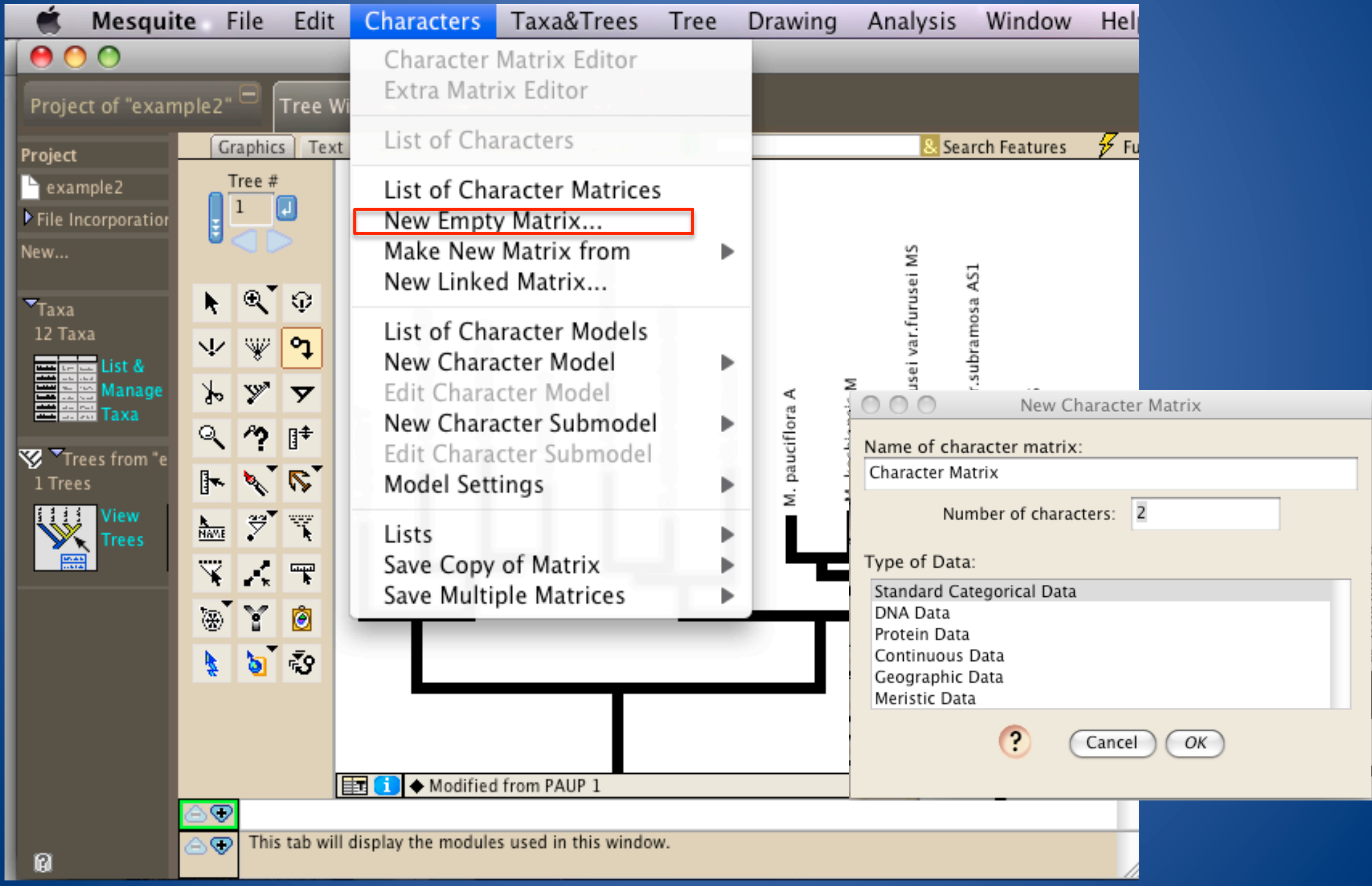

釣鐘型の花


皿型の花

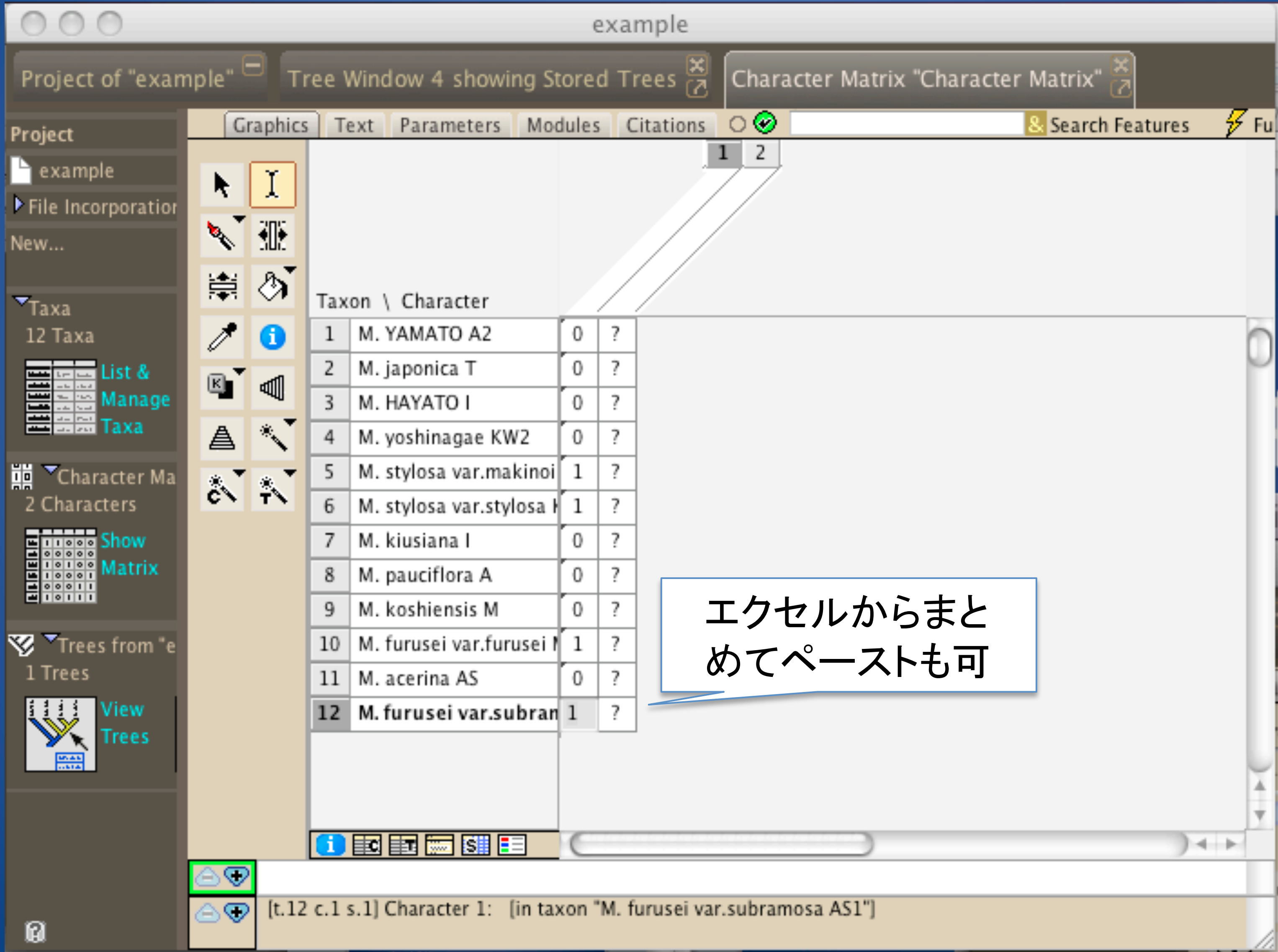
祖先形質の復元（推定）

- ある形質がどの枝で、何度進化したかを推定する
- 最節約法による方法→形質の変化が最少となるように祖先形質を推定
- 最尤法による方法→系統樹全体に適用される形質進化のモデル（連続時間マルコフモデル[Pagel, 1994]）のパラメータを、最もデータ（分かっている種の形質状態＋系統樹）に当てはまりがよいように最適化して祖先形質を推定

Mesquiteに形質の打ち込み



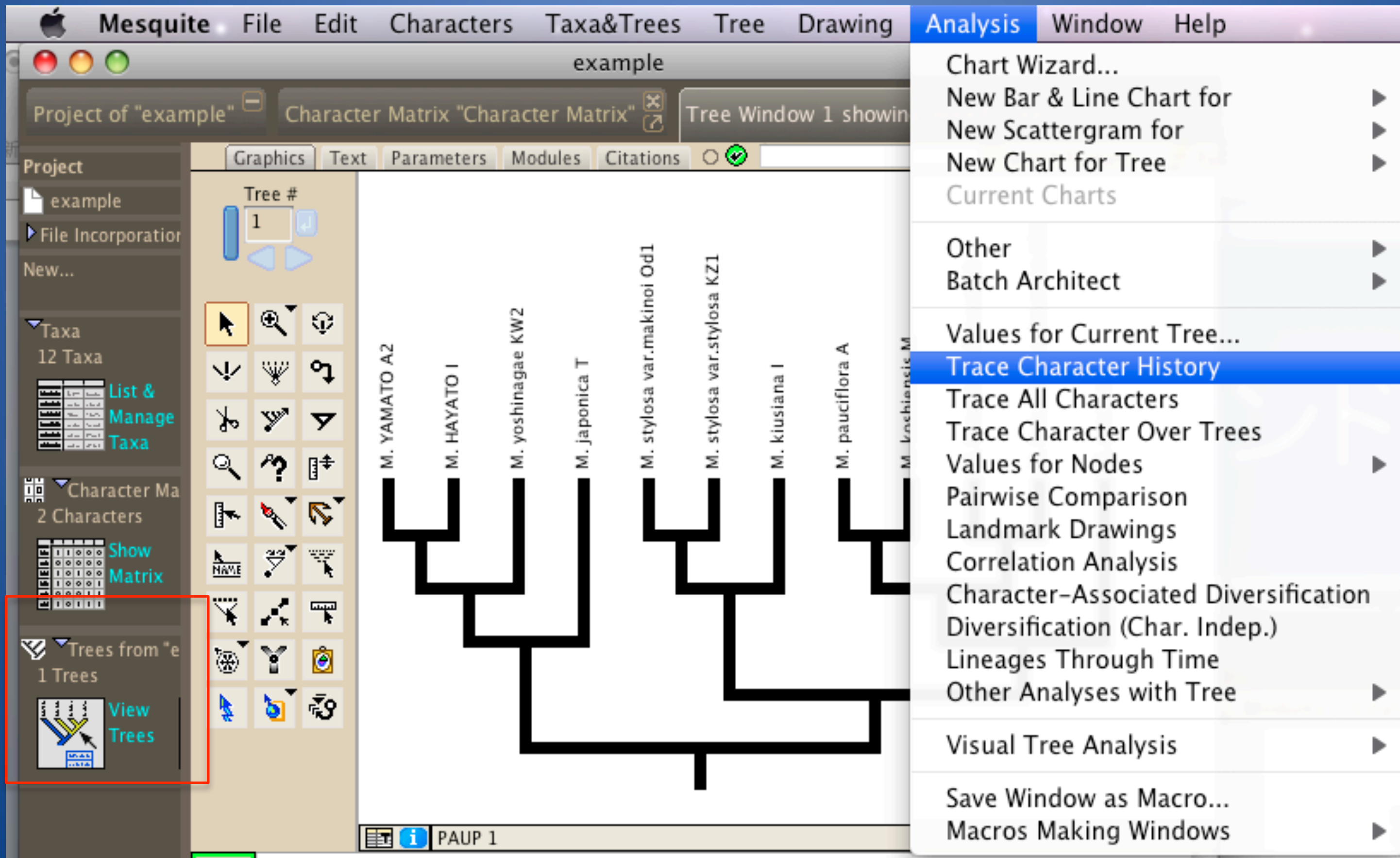
Mesquiteに形質の打ち込み



エクセルからまと
めてペーストも可

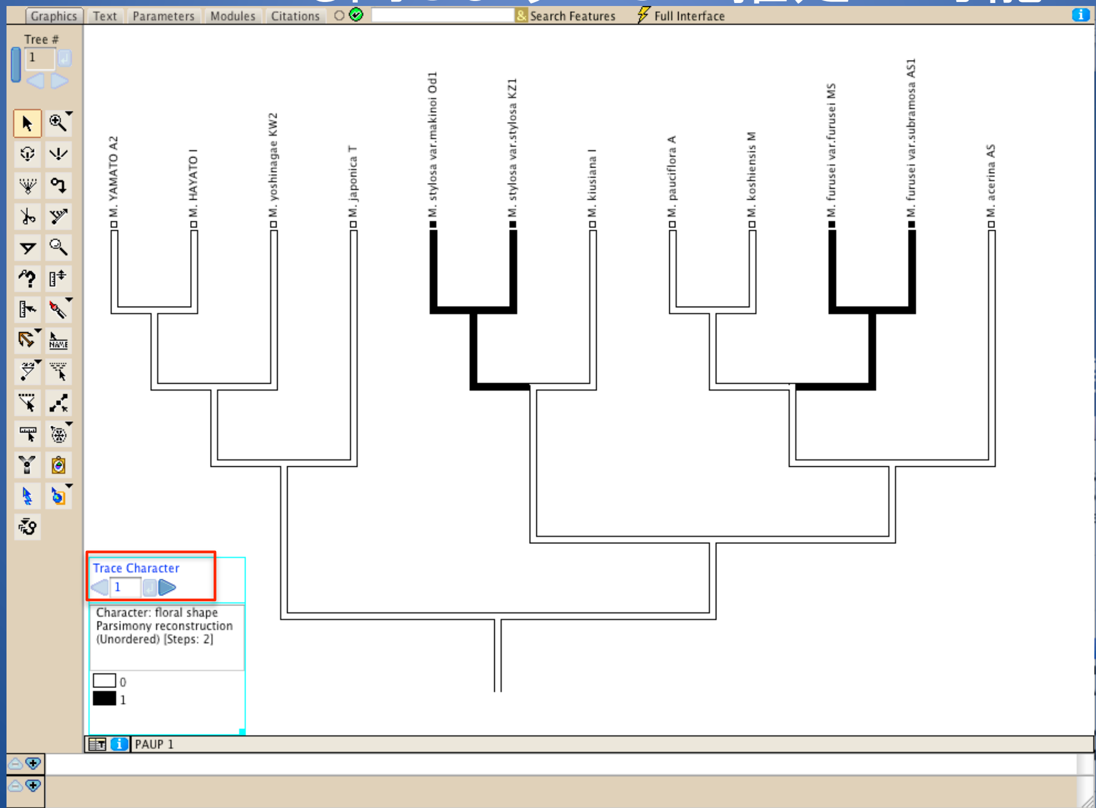
祖先形質の推定

- Tree Windowを開きTrace Character HistoryからStored Characterを選択
- Parsimony/Likelihoodによる祖先形質復元を選べる



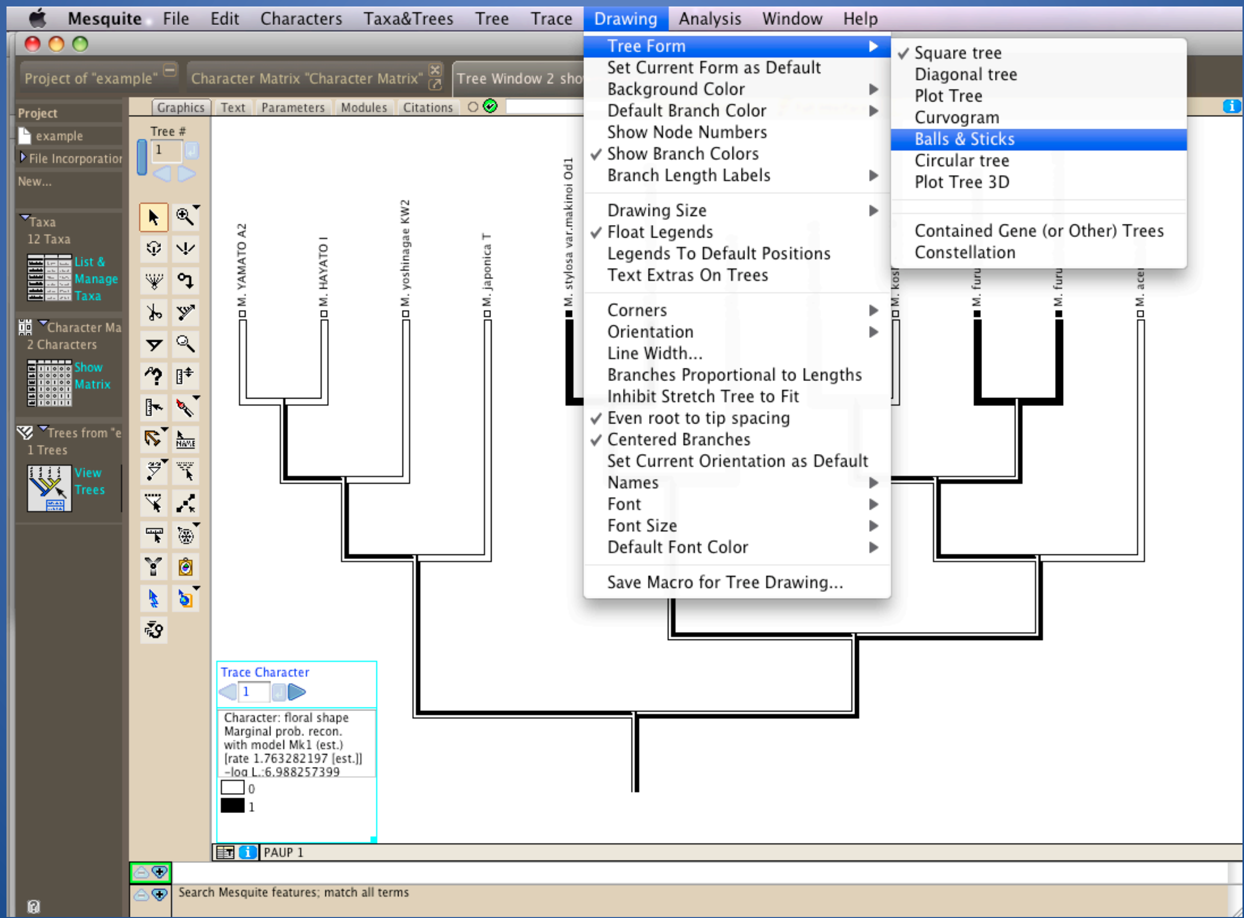
祖先形質の復元（最節約法）

- 直感的に分かりやすい結果が得られる
- 形質の変化が多い場合は誤りやすく、また推定不可能な枝が増える
- Continuous dataでも同じようにして推定が可能



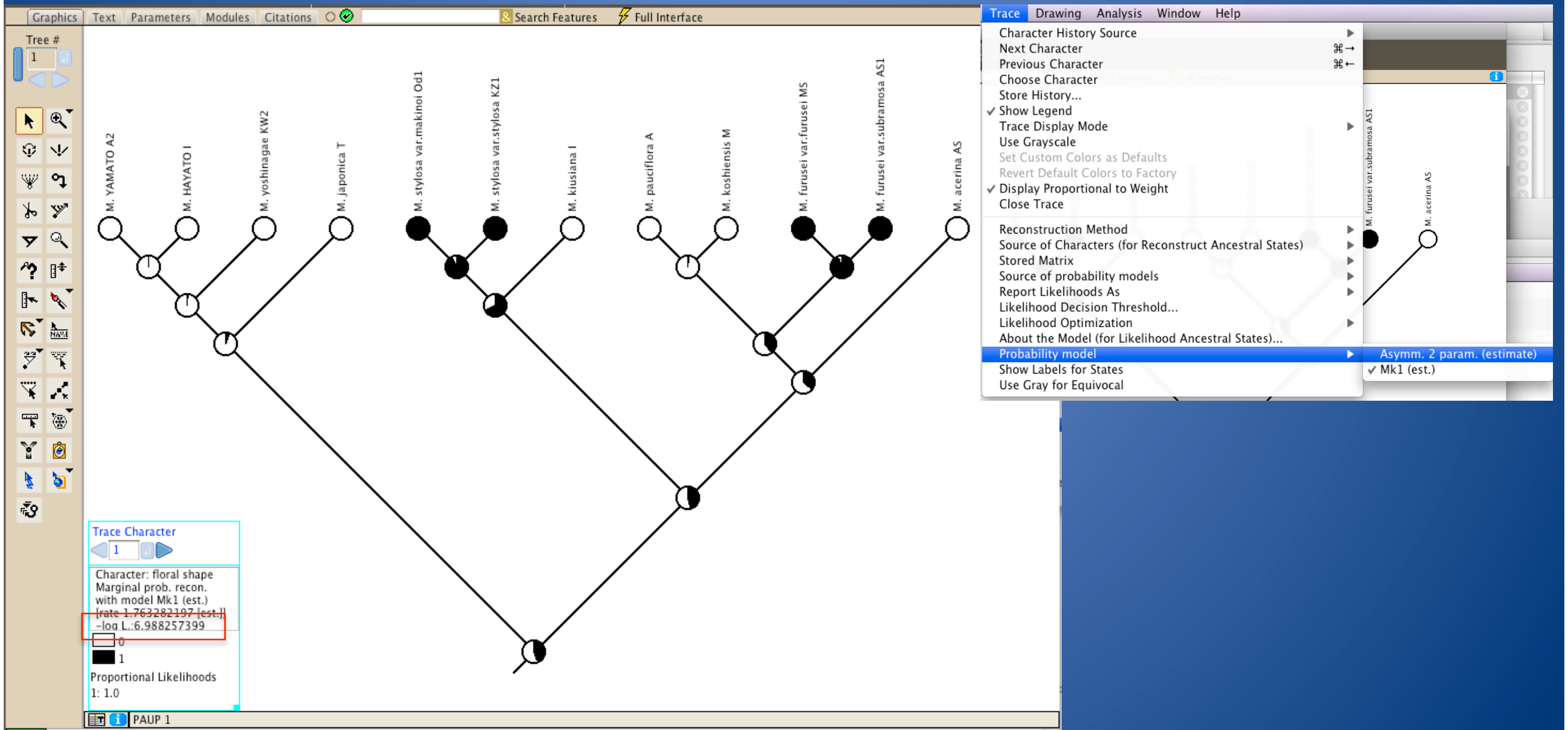
祖先形質の復元（最尤法）

- Current Probability Modelを選択
- Square Treeでは結果が見にくいので、Tree FormのBalls&Sticksを選ぶ



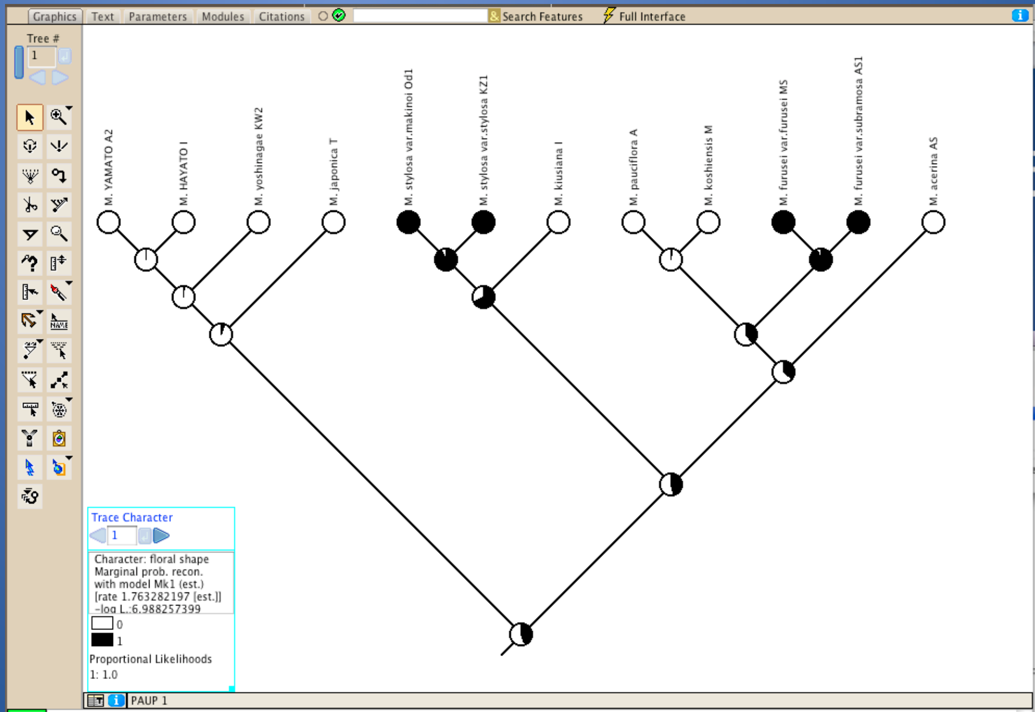
祖先形質の復元（最尤法）

- 形質進化のモデルは0→1と1→0の起きる確率が同じとするモデルMk1と、異なるとするモデルAsymm. 2 param.を選べる（尤度比検定などで選択）



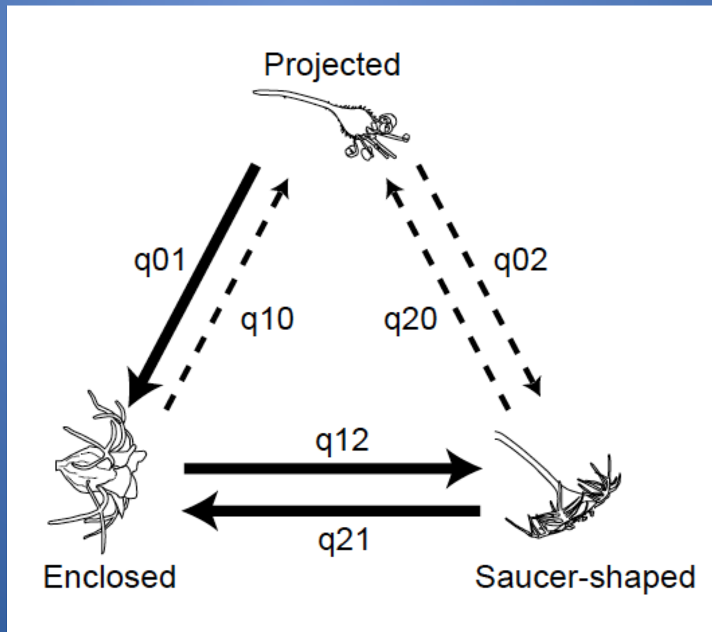
祖先形質の復元（最尤法）

- 曖昧な結果が出やすい
- (Mesquiteでは)形質が3種類以上になった時にもパラメータは2までしか増やせない
- (Mesquiteでは)祖先形質の推定の確からしさ（パイグラフ）が過大評価される“global estimator”と呼ばれる推定法しかできない



BayesTraits

- 形質が3種類以上になった時にも全てのパラメーターを動かせる
- 大量のツリーを並行して解析できる
- より厳密な手法(“local estimator”)が使える



BayesTraitsの入手

- http://www.evolution.rdg.ac.uk/BayesTraits.html

Reading Evolutionary Biology Group - BayesTraits

Home

People

Research

Selected Publications

Software

Data Sets

Description

BayesTraits is a computer package for performing analyses of trait evolution among groups of species for which a phylogeny or sample of phylogenies is available. This new package incoporates our earlier and separate programes Multistate, Discrete and Continuous. BayesTraits can be applied to the analysis of traits that adopt a finite number of discrete states, or to the analysis of continuously varying traits. Hypotheses can be tested about models of evolution, about ancestral states and about correlations among pairs of traits.

Manual (PDF)

BayesTraits - Windows
BayesTraits - Linux
BayesTraits - OS X PPC
BayesTraits - OS X Intel

Continuous Manual (PDF)

The bms_runner script uses BayesTraits to predict functional links among genes. bms_runner helps configure BayesTraits on the basis of known, user-supplied data, and launches BayesTraits repeatedly to seek correlated gain and loss within pairs of genes.

bms_runner - Linux

The Bayesian-Multistate Perl scripts (Buschbom & Barker 2006, Syst Biol 55:471-484) are no longer distributed here. For information on those scripts, please contact Daniel Barker by email: **db60@st-andrews.ac.uk**

Back to Software

BayesTraitsの実行

- UNIXプログラム
- ツリーファイルと形質を記入したテキストファイルの2つを必要とする
- 改行コード等に注意

```
M._YAMATO_A2                      0
M._japonica_T                     0
M._HAYATO_I                       0
M._yoshinagae_KW2                 0
M._stylosa_var.makinoi_Od1        1
M._stylosa_var.stylosa_KZ1        1
M._kiusiana_I                     0
M._pauciflora_A                   0
M._koshiensis_M                   0
M._furusei_var.furusei_MS         1
M._acerina_AS                     0
M._furusei_var.subramosa_AS1      1
```

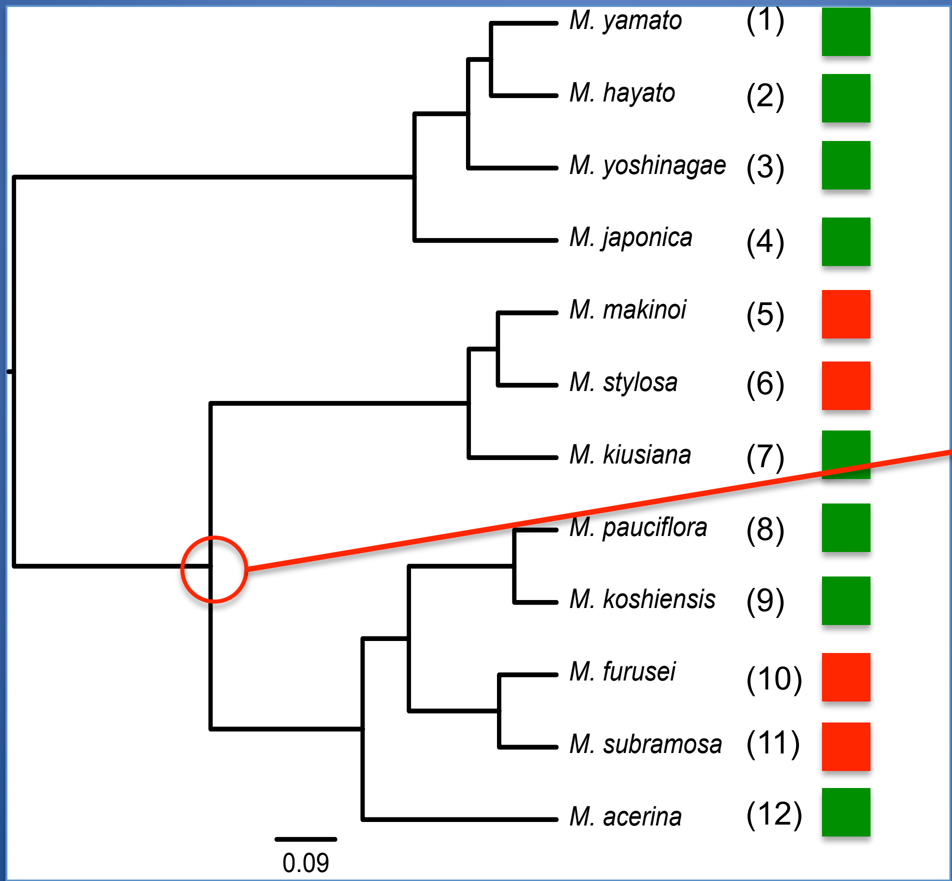
BayesTraitsの実行

ターミナル — BayesTraits — 84×36

```
Rand Seed       1280669581
Err: Does have a #NEXUS tag in.
tc64098:BayesTraits 2 yudai$ ./BayesTraits example2.tre example.txt
Rand Seed       1280669632
Please Select the model of evolution to use.
1)      MultiState.
2)      Discrete: Independent
3)      Discrete: Depend
4)      Continuous: Random Walk (Model A)
5)      Continuous: Directional (Model B)
1
Please Select the analsis method to use.
1)      Maximum Likelihood.
2)      MCMC
1
Options:
Model:                          Multistates
Tree File Name:                 example2.tre
Data File Name:                 example.txt
Log File Name:                  example.txt.log.txt
Summary:                        False
Analsis Type:                   Maximum Likelihood
ML attempt per tree:            10
No of Rates:                    2
Base frequency (PI's)           None
Character Symbols               0,1
Using a covarion model:         False
Restrictions:
    q01                         None
    q10                         None
Tree Information
     Trees:                     1
     Taxa:                      12
     Sites:                     1
     States:                    2
>
```

Local Estimator

- 祖先形質を知りたい節(ノード)について、取りうるそれぞれの形質に固定した場合の全体の尤度を求める

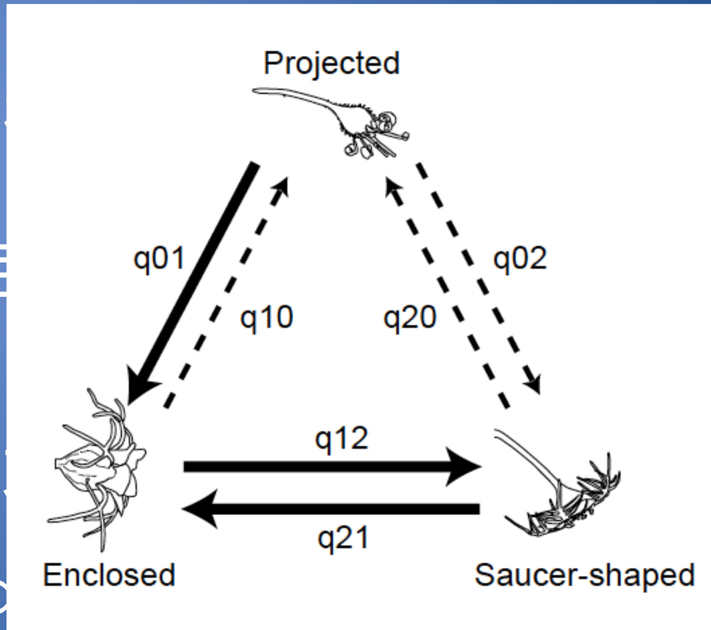


ここが■の場合
logL=-6.290513

ここが■の場合
logL=-6.550491

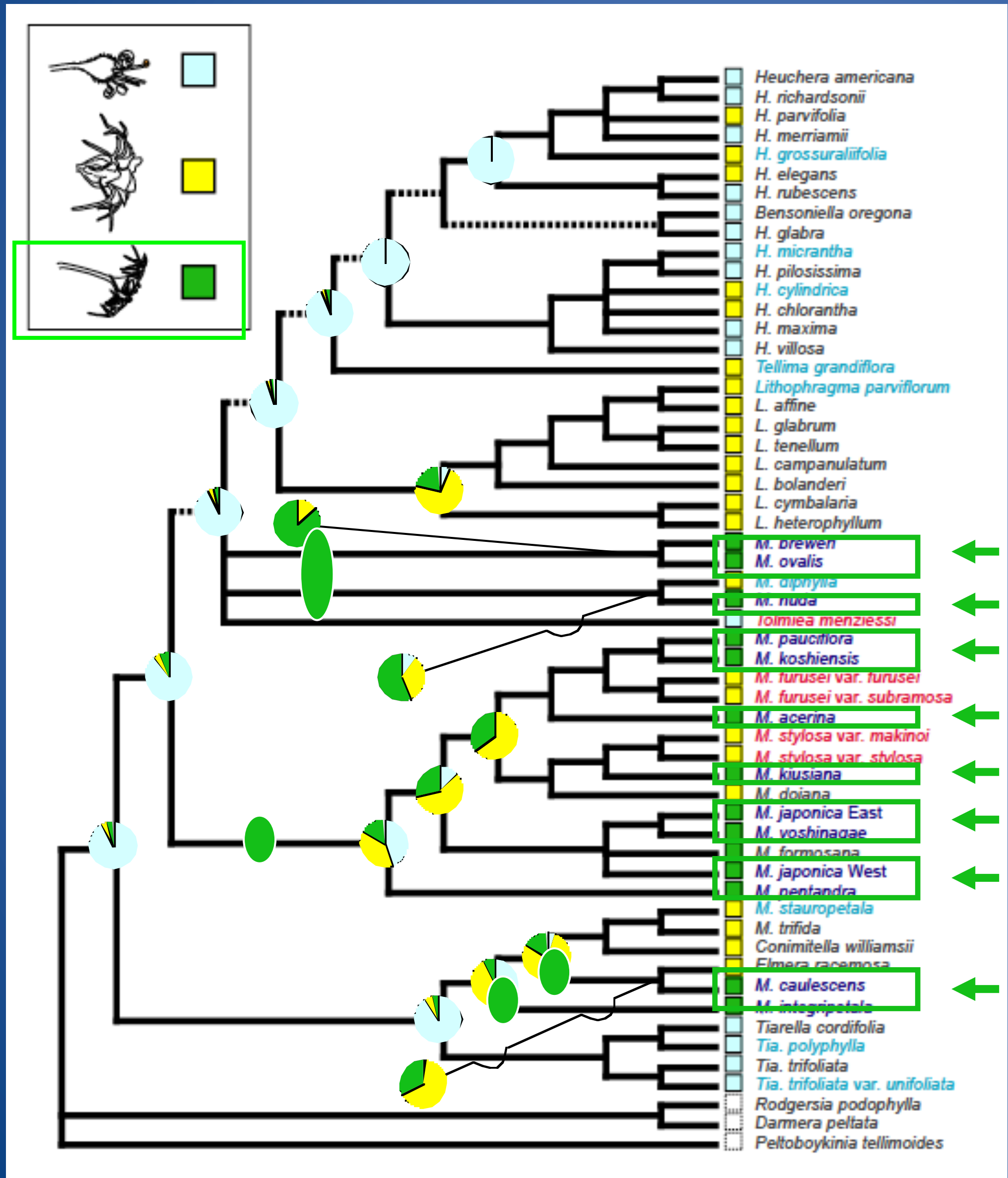
BayesTraitsの実行

- 初期状態ではパラメータは全て自由に動くので、必要に応じてrestrictコマンドを用いてパラメータ数を減らす（例：restrict q01 q10）
- local estimatorによる取りうる 比
較にはfossilコマンドを用いる 最
近共通祖先の形質を0に固定 10
12）
- 全て設定したらrunコマンドで
- 同じnodeでもう一度fossil Nod
行し、前の結果と尤度を比較する
- 尤度の比をパイグラフなどで表現する



BayesTraitsによる祖先形質推定の応用例

皿型花

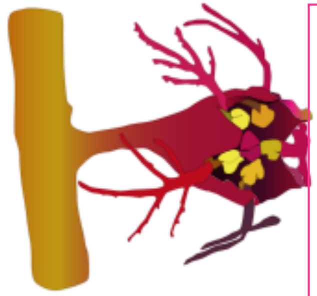


キノコバエ媒

- チャルメルソウの仲間において、皿型花とキノコバエ媒の間に進化上の対応があることが明らかになった

チャルメルソウ類の種分化は
どのように起こったか？

チャルメルソウ類の送粉者となる
キノコバエ類には2タイプある



筒状ー釣鐘状の花

- ミカワチャルメルソウ
- チャルメルソウ
- タキミチャルメルソウ
- シコクチャルメルソウ
- *Tolmiea menziesii* (US)

皿状ー杯状の花

- モミジチャルメルソウ
- オオチャルメルソウ
- ツクシチャルメルソウ
- コシノチャルメルソウ
- ハヤトチャルメルソウ（新種）
- *Mitella pentandra* (US)
- *Mitella breweri* (US)
- *Mitella caulescens* (US)
- コチャルメルソウ
- ヤマトチャルメルソウ（新種）
- マルバチャルメルソウ
- エゾノチャルメルソウ

シギキノコバエ

口吻が伸長していない
キノコバエ

チャルメルソウ節の種と送粉者との関係の厳密性

チャルメルソウの種	シギキノコバエ	その他のキノコバエ	その他（双翅目）
チャルメルソウ	102	0	0
ヤマトチャルメルソウ	0	69	1

三重県赤目四十八滝の2種混生個体群で、
2人の観察者による計6時間の連続観察により得た訪花回数

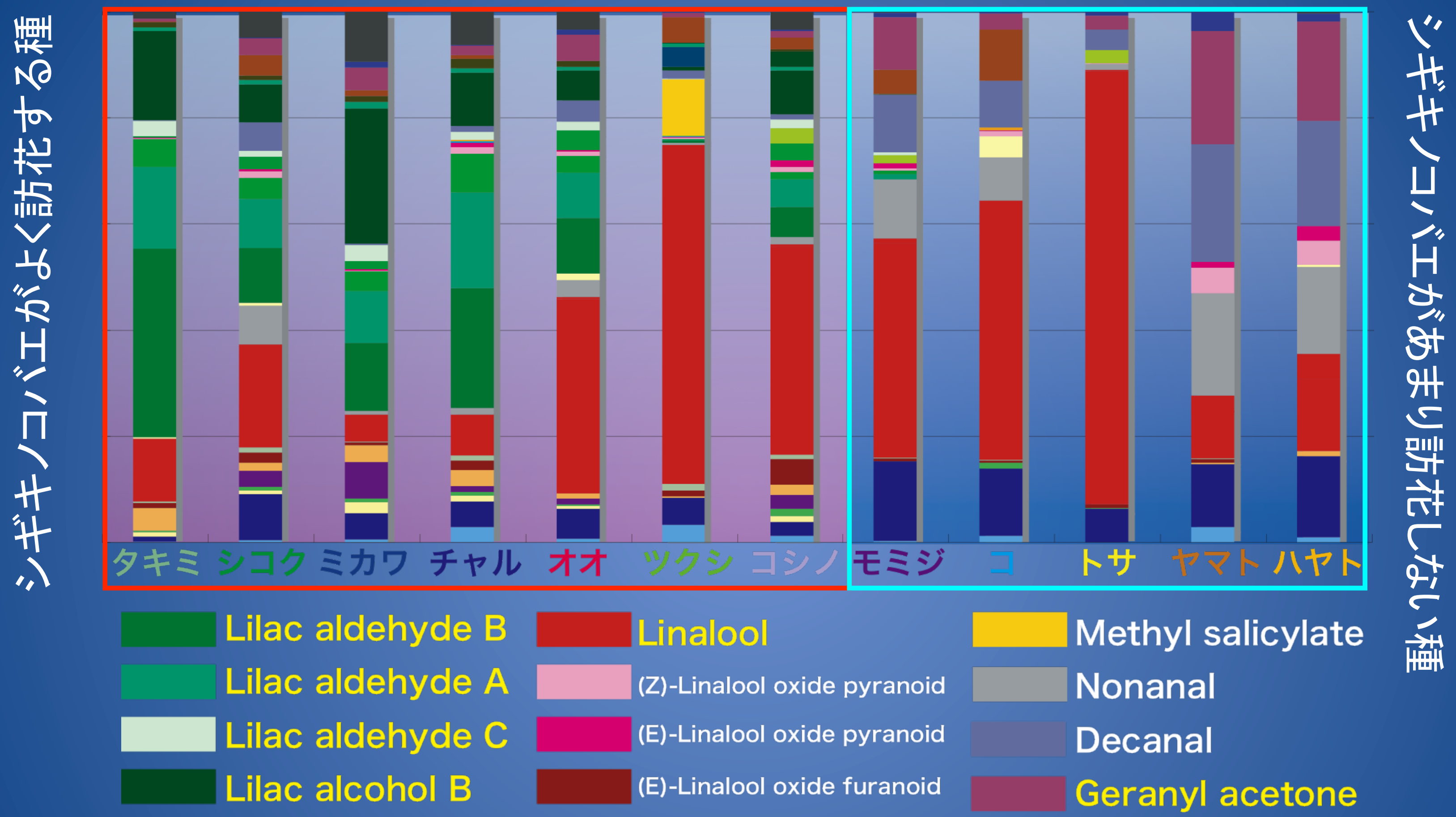
チャルメルソウの種	シギキノコバエ	その他のキノコバエ
シコクチャルメルソウ	39	0
トサチャルメルソウ	0	6

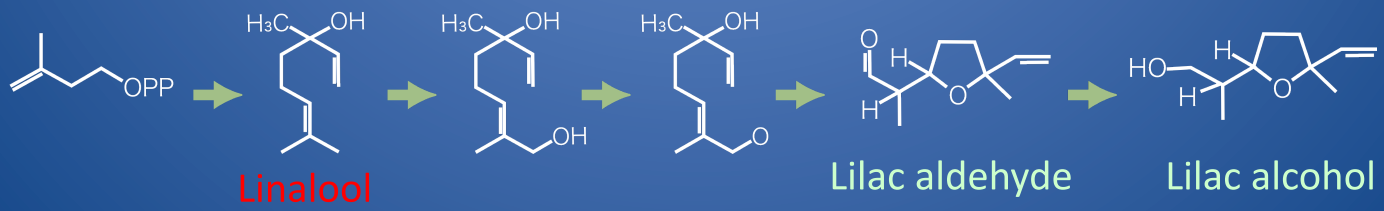
徳島県穴吹町の2種混生個体群で、
2人の観察者による計2時間の連続観察により得た訪花回数

↓

はっきりと誘引する送粉者を決める因子が存在する

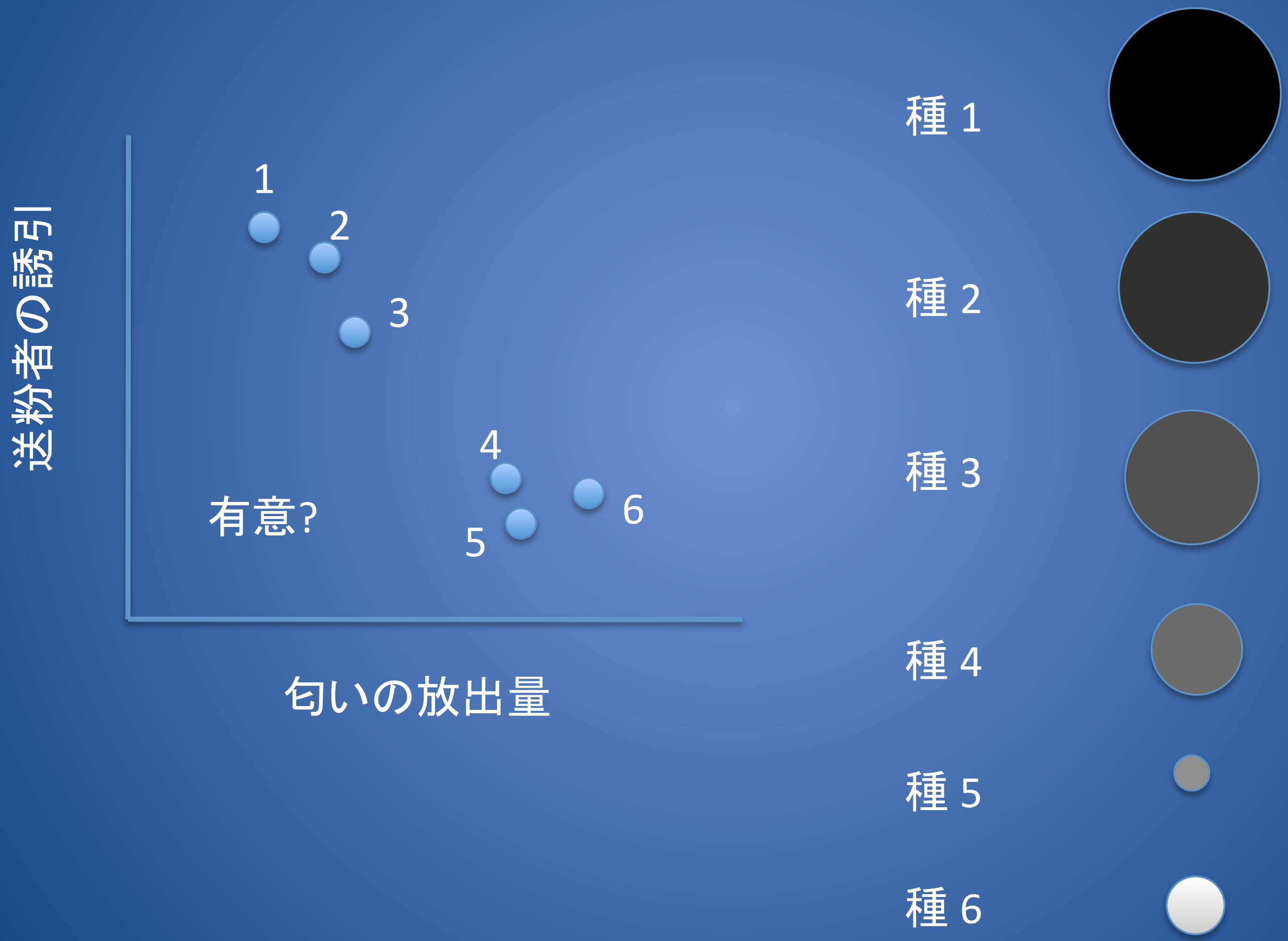
チャルメルソウ全種の花香成分組成





Okamoto et al.，未発表データ

種間比較：種間で形質間の相関を調べる際に生じる問題

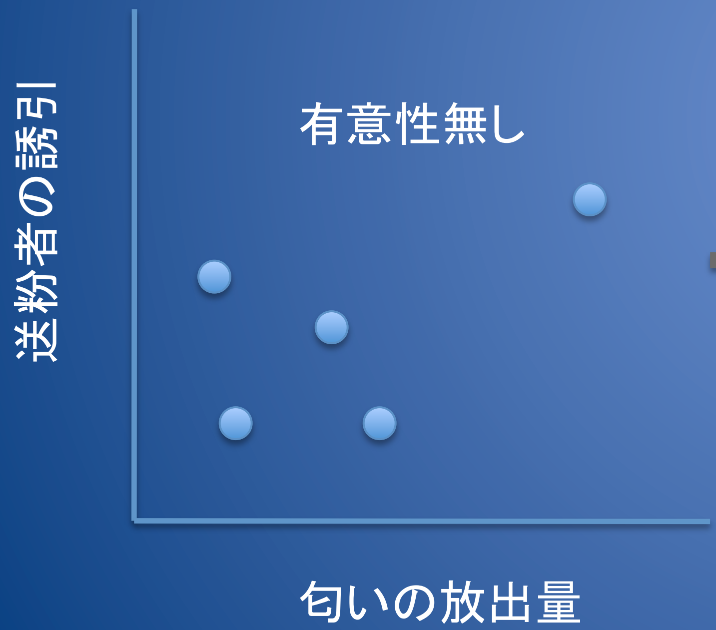


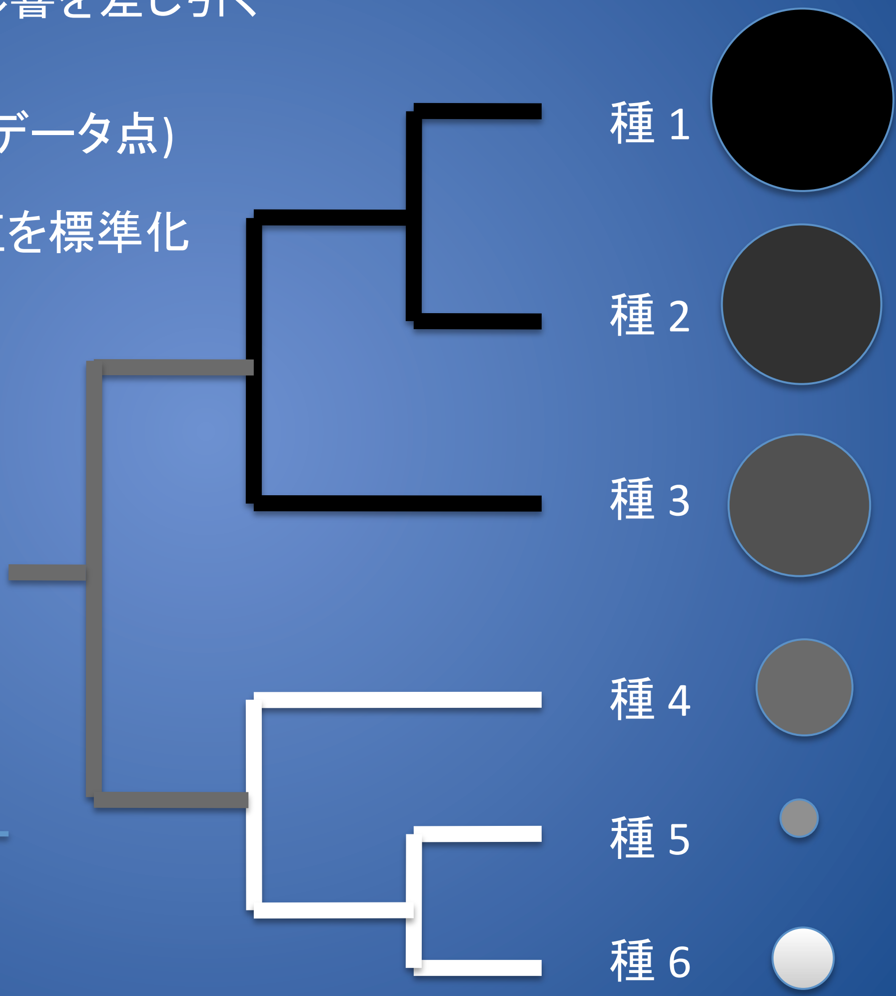
系統学的独立比較: PIC

データ点の値から系統関係の影響を差し引く

データ点の数は1つ減少する
（n種を比較した場合はn-1個のデータ点)

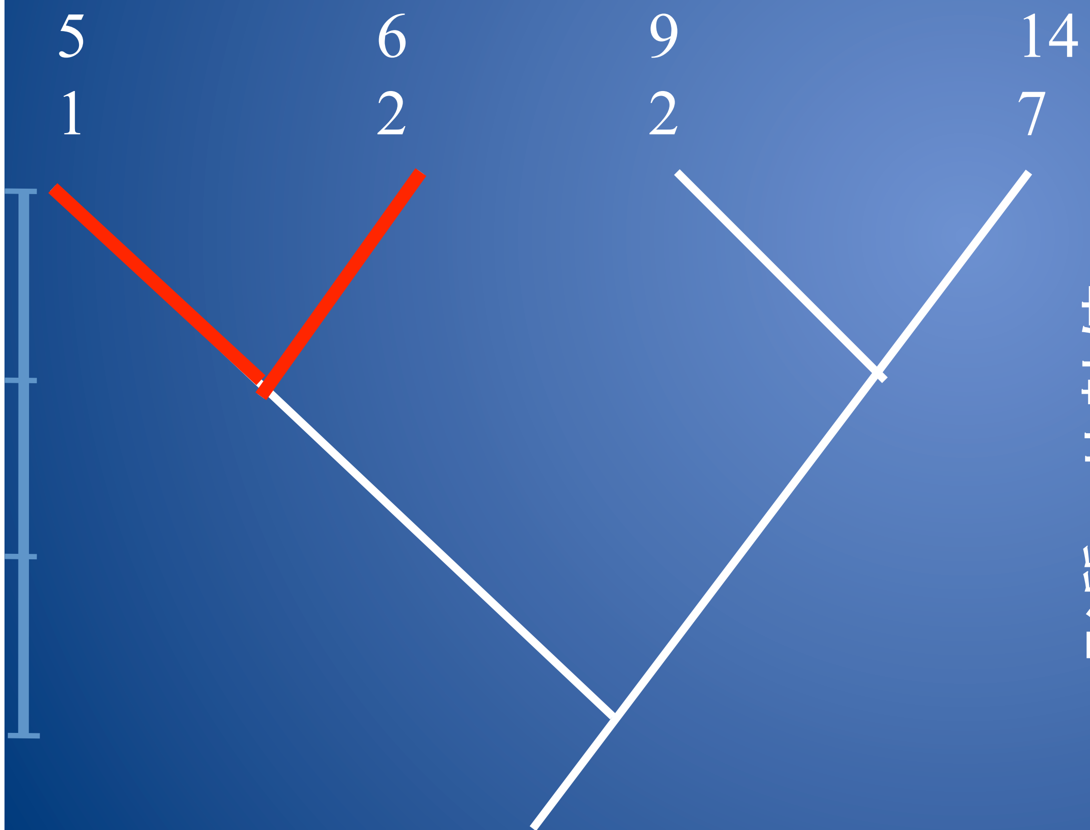
進化のチャンス＝樹長で比較値を標準化



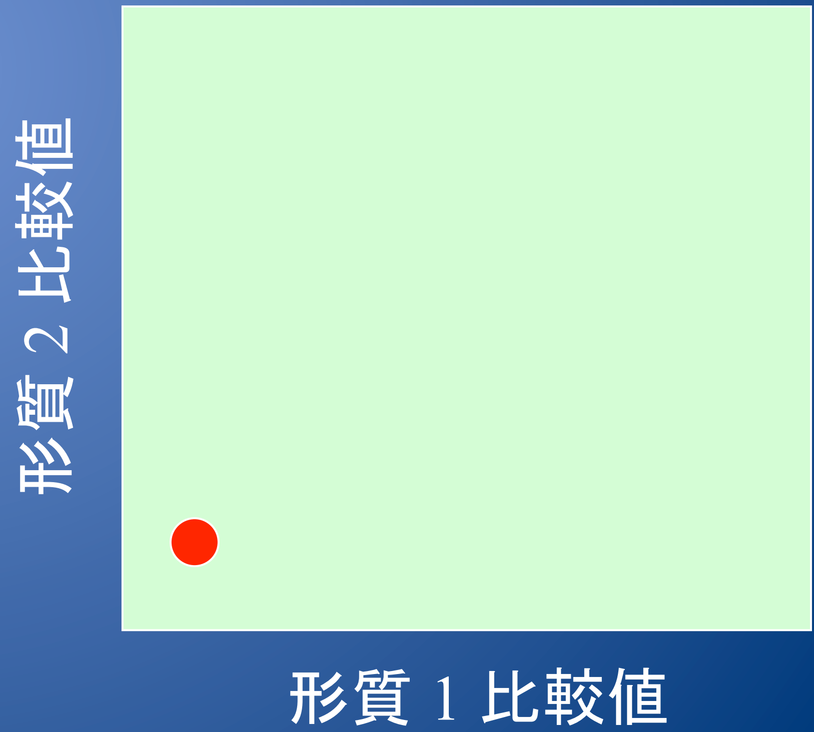


系統的独立比較の実際

形質 1: (6-5=1)
形質 2: (2-1=1)
比較値: (1,1)/√2

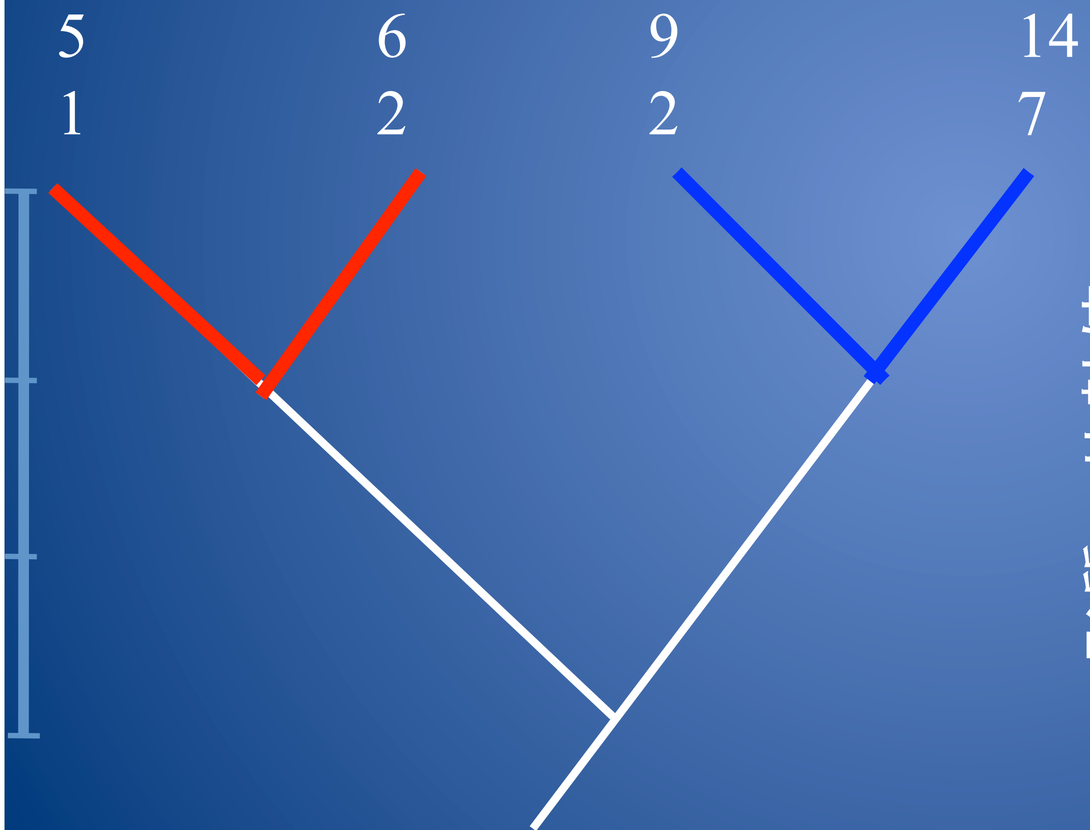


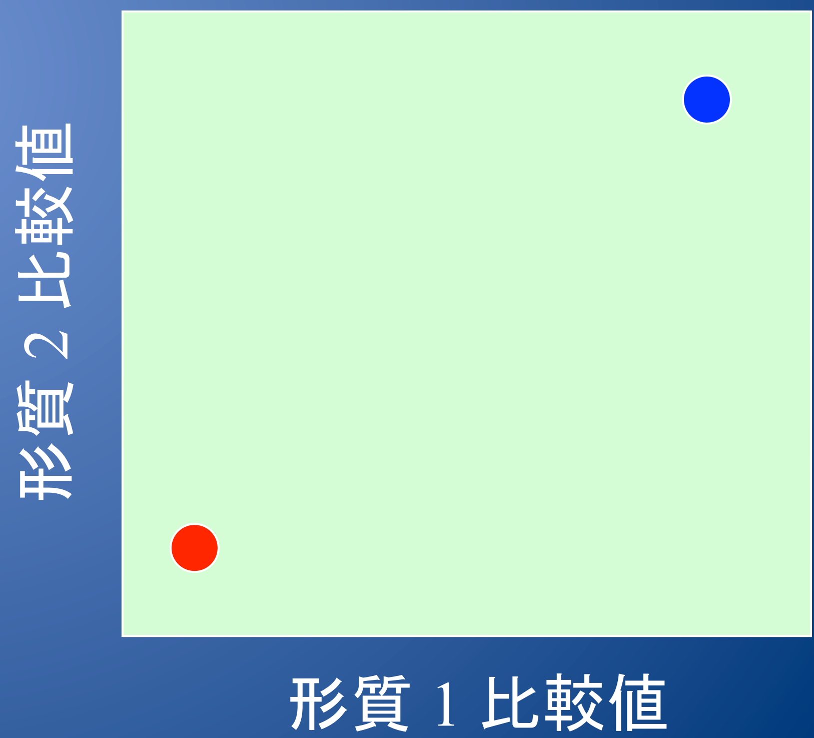
Felsenstein 1985



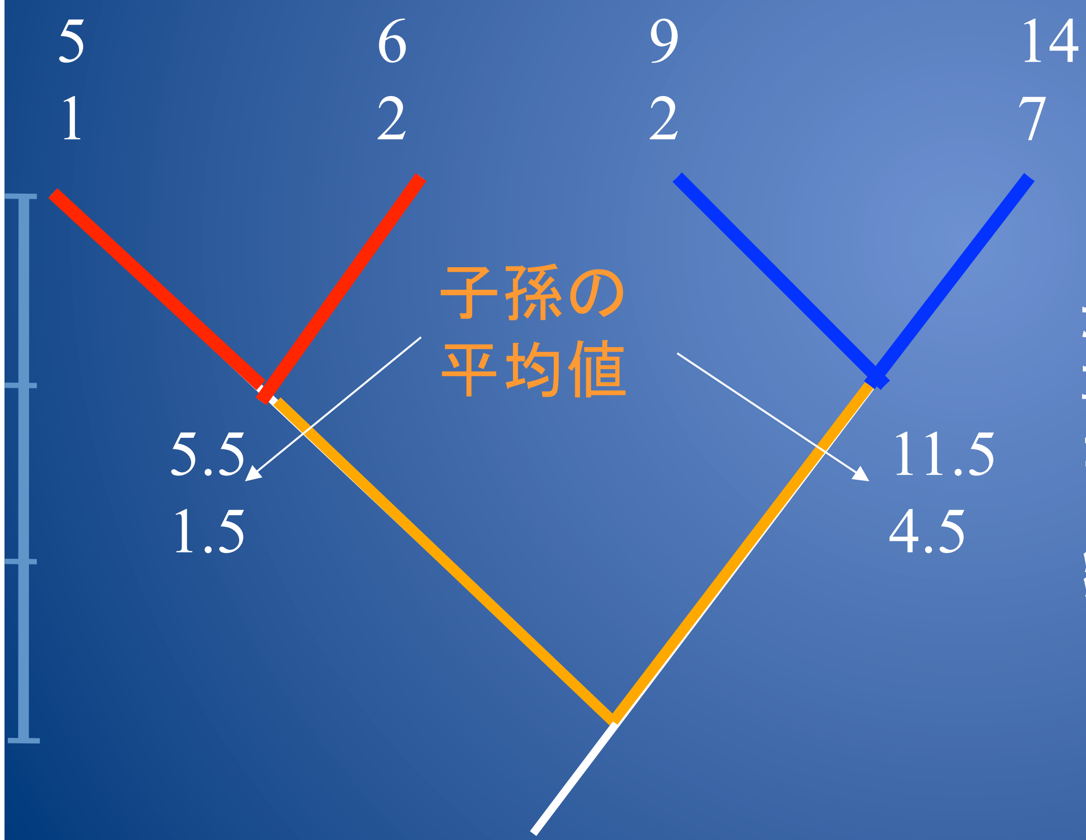
系統的独立比較の実際

形質 1: (14-9=5)
形質 2: (7-2=5)
比較値: (5,5)/√2

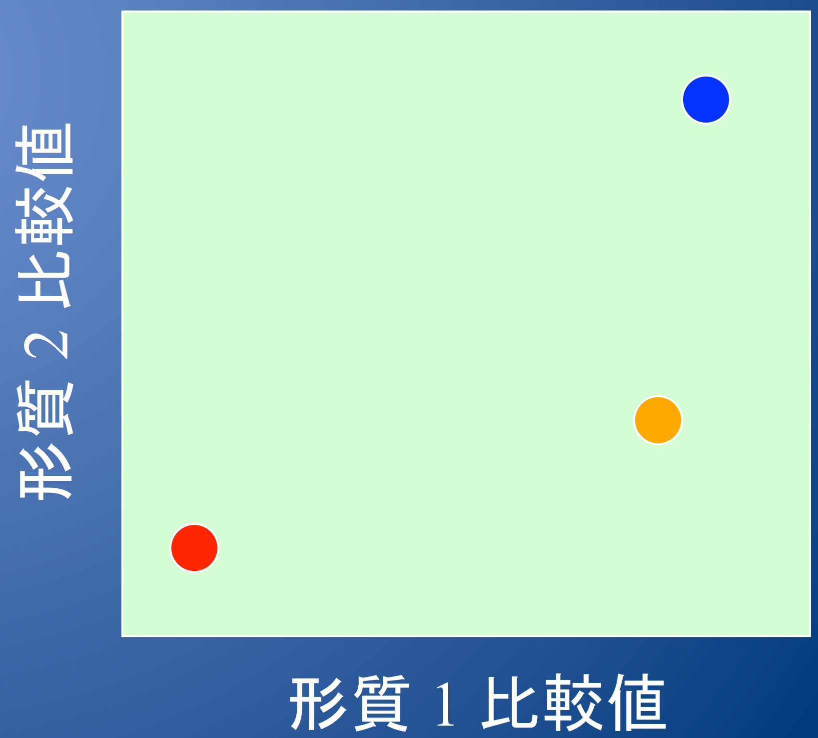




系統的独立比較の実際



形質 1: 11.5-5.5=6
形質 2: 4.5-1.5=3
比較値: (6,3)/√4

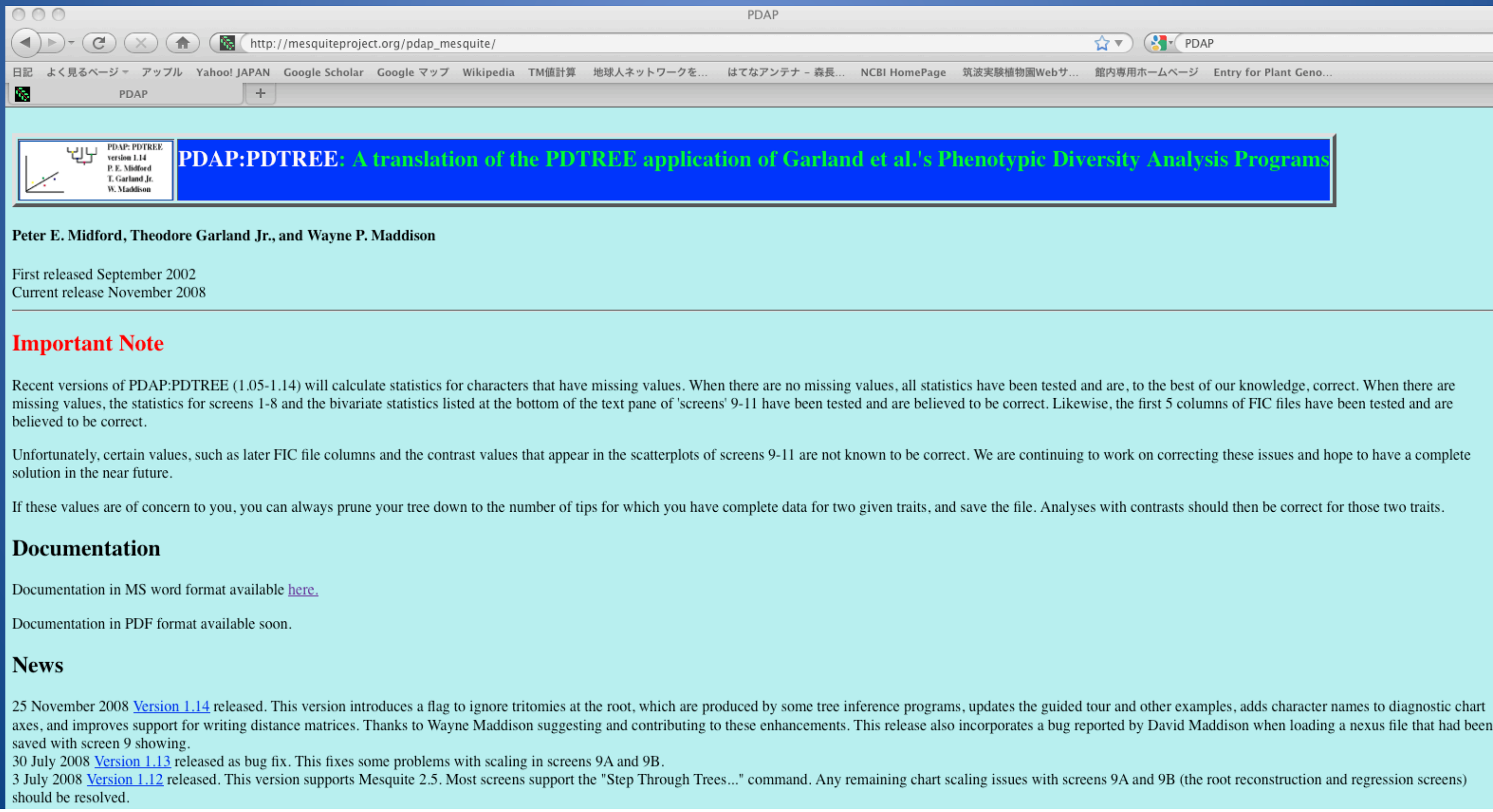


系統的独立比較の実際

- 比較の方向は任意
- 比較値が一意に決まらないので、片方の値を正にする(positivized)
- 符号検定による形質間の相関の検定も行える（xを正に固定した時にyが正になるものと負になるものの数で検定）
- 回帰をする場合は原点から(切片=0）

Mesquiteで実行するには

- http://mesquiteproject.org/pdap_mesquite/から PDAP:PDTREEをダウンロード、解凍
- Put in Mesquite Directoryフォルダの中の、PDAPフォルダをMesquiteのmesquiteフォルダに入れて、Mesquiteを起動



Peter E. Midford, Theodore Garland Jr., and Wayne P. Maddison

First released September 2002
Current release November 2008

Important Note

Recent versions of PDAP:PDTREE (1.05-1.14) will calculate statistics for characters that have missing values. When there are no missing values, all statistics have been tested and are, to the best of our knowledge, correct. When there are missing values, the statistics for screens 1-8 and the bivariate statistics listed at the bottom of the text pane of 'screens' 9-11 have been tested and are believed to be correct. Likewise, the first 5 columns of FIC files have been tested and are believed to be correct.

Unfortunately, certain values, such as later FIC file columns and the contrast values that appear in the scatterplots of screens 9-11 are not known to be correct. We are continuing to work on correcting these issues and hope to have a complete solution in the near future.

If these values are of concern to you, you can always prune your tree down to the number of tips for which you have complete data for two given traits, and save the file. Analyses with contrasts should then be correct for those two traits.

Documentation

Documentation in MS word format available here.

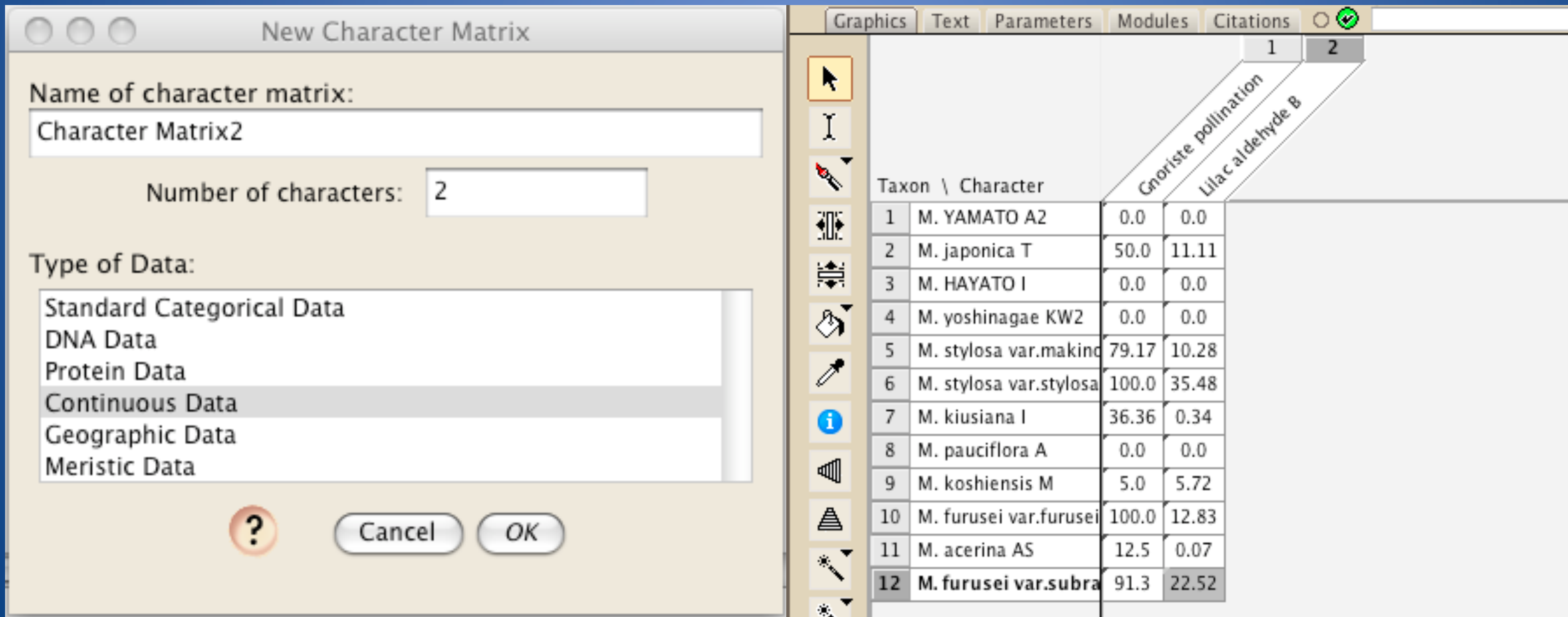
Documentation in PDF format available soon.

News

25 November 2008 Version 1.14 released. This version introduces a flag to ignore tritomies at the root, which are produced by some tree inference programs, updates the guided tour and other examples, adds character names to diagnostic chart axes, and improves support for writing distance matrices. Thanks to Wayne Maddison suggesting and contributing to these enhancements. This release also incorporates a bug reported by David Maddison when loading a nexus file that had been saved with screen 9 showing.
30 July 2008 Version 1.13 released as bug fix. This fixes some problems with scaling in screens 9A and 9B.
3 July 2008 Version 1.12 released. This version supports Mesquite 2.5. Most screens support the "Step Through Trees..." command. Any remaining chart scaling issues with screens 9A and 9B (the root reconstruction and regression screens) should be resolved.

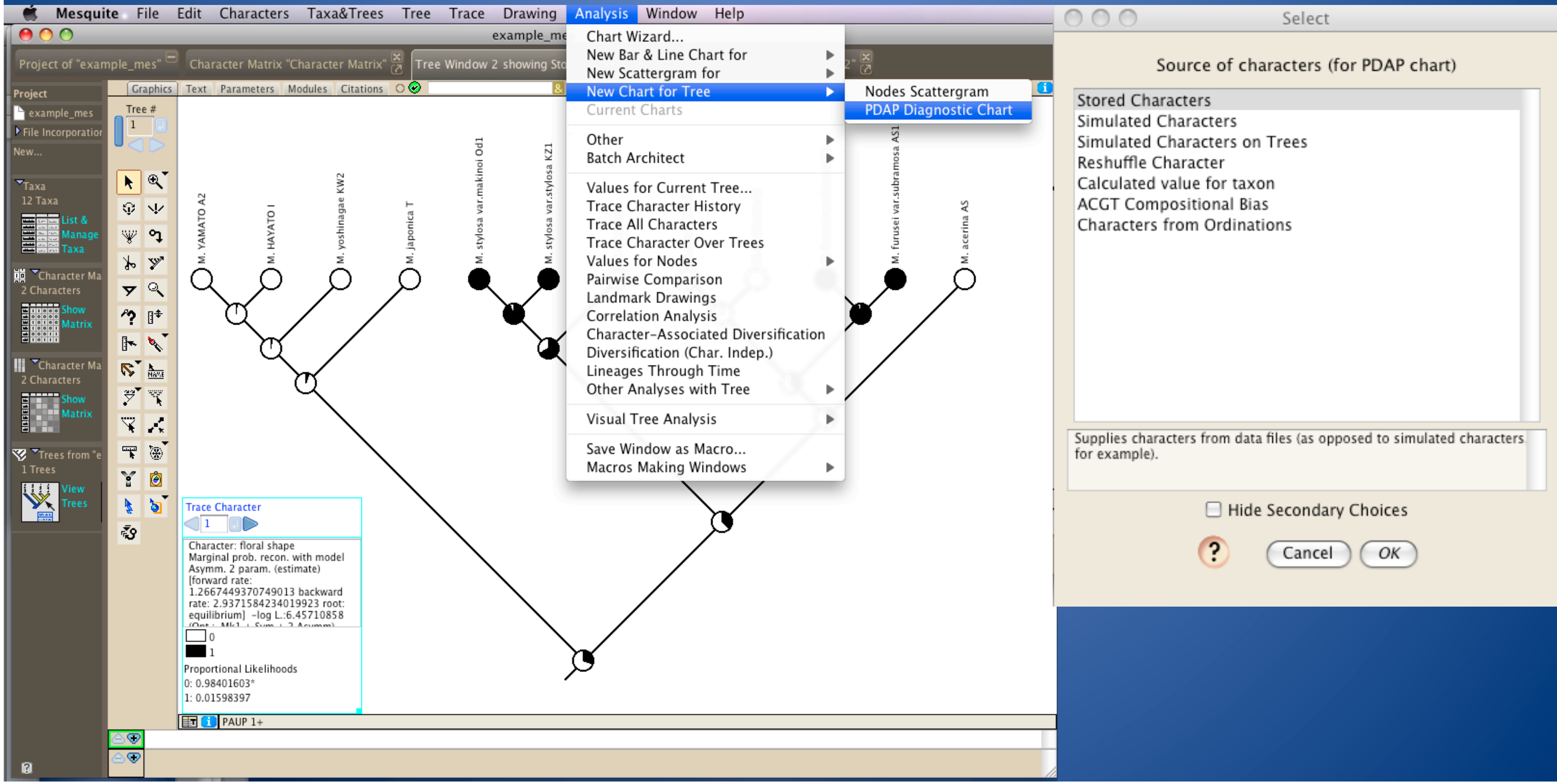
Mesquiteに形質の打ち込み

- 上のバーからCharacters>New Empty Matrixを選択
- Continuous Dataを選択



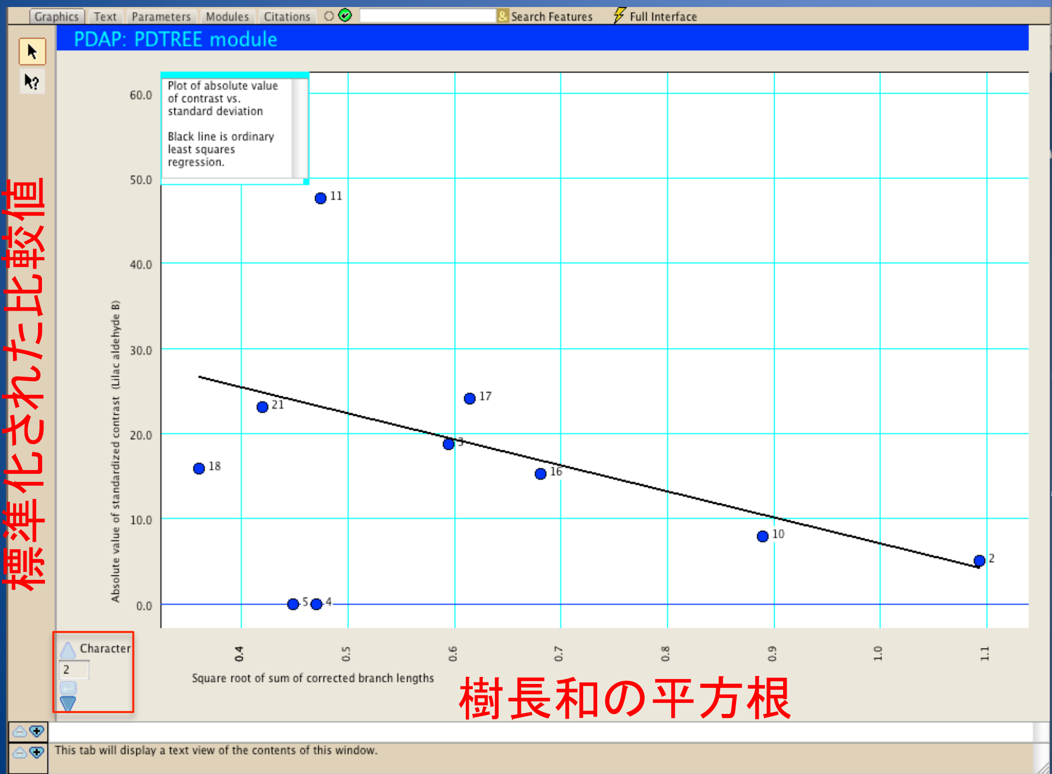
系統的独立比較の実行

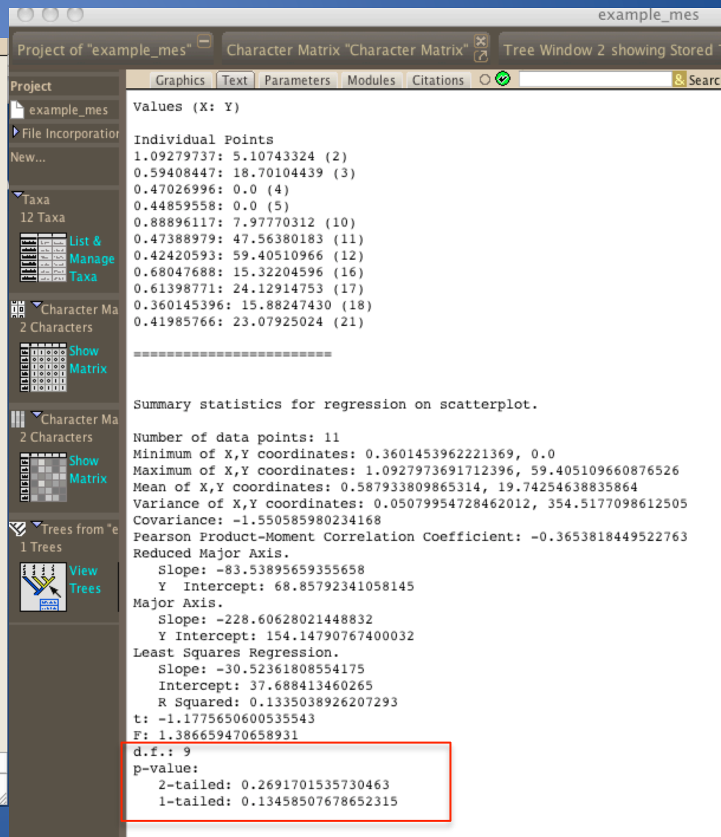
- Tree windowからNew Chart for Tree>PDAP Diagonistic Chartを選択



系統的独立比較の仮定と適合するかを確認
(PDAP diagonistic chart)

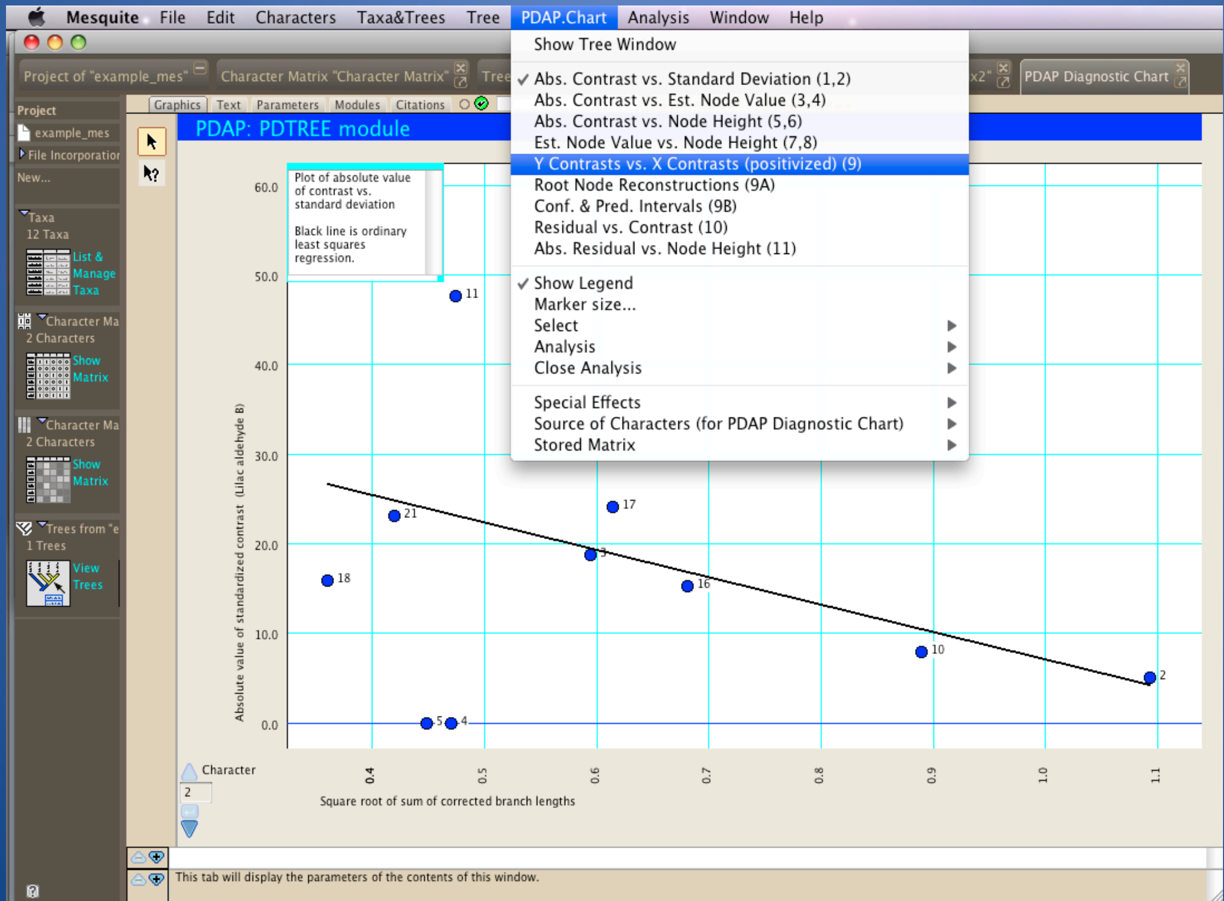
- 標準化した比較値と樹長が相関しない
- 比較値（Y軸の値）は正規分布している
- 明らかな外れ値が存在しない





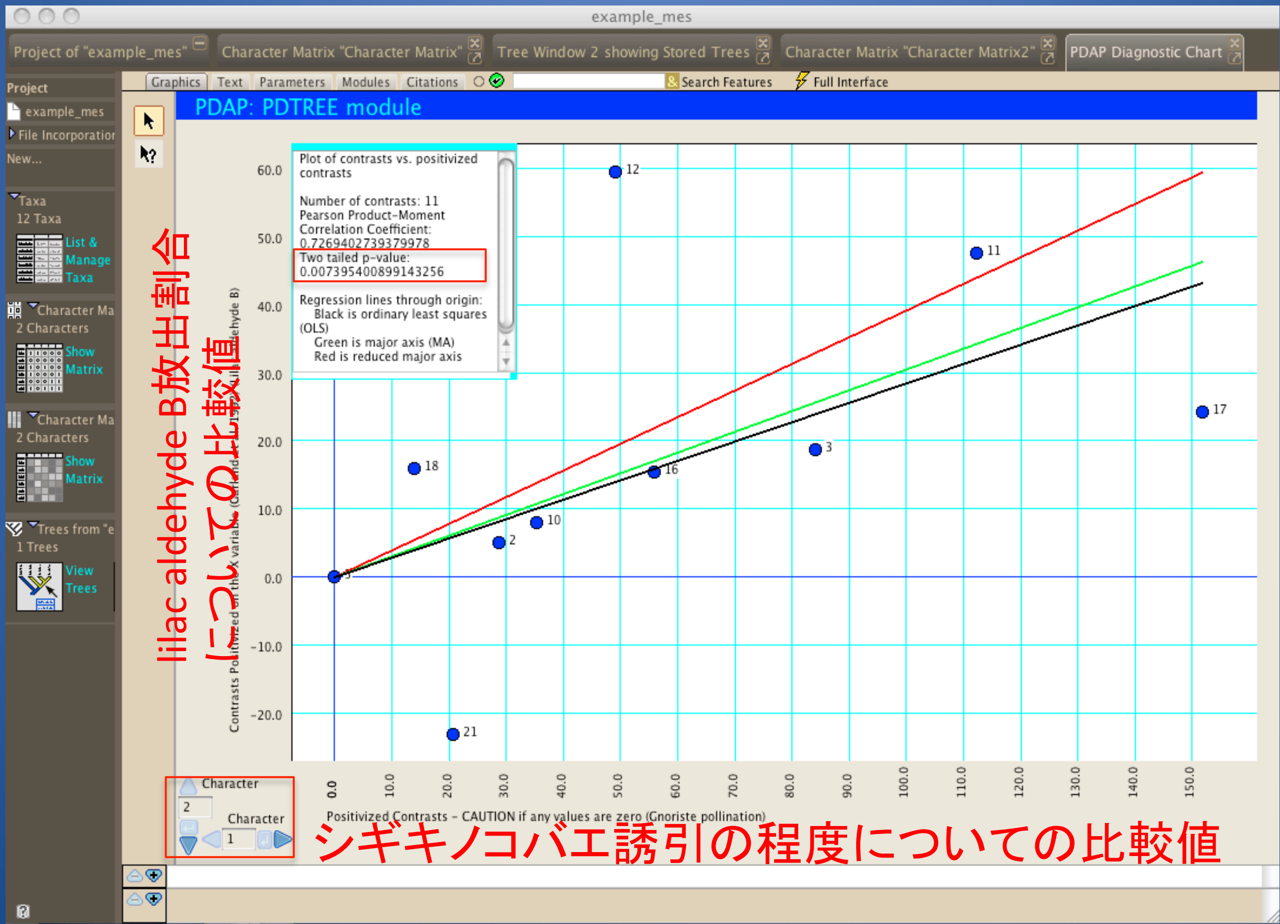
系統的独立比較の実行

- 上のバーからY Contrast vs. X Contrast (positivized)を選択



系統的独立比較の実行

- 回帰係数についての検定結果を見て、形質間の相関の有無を見る

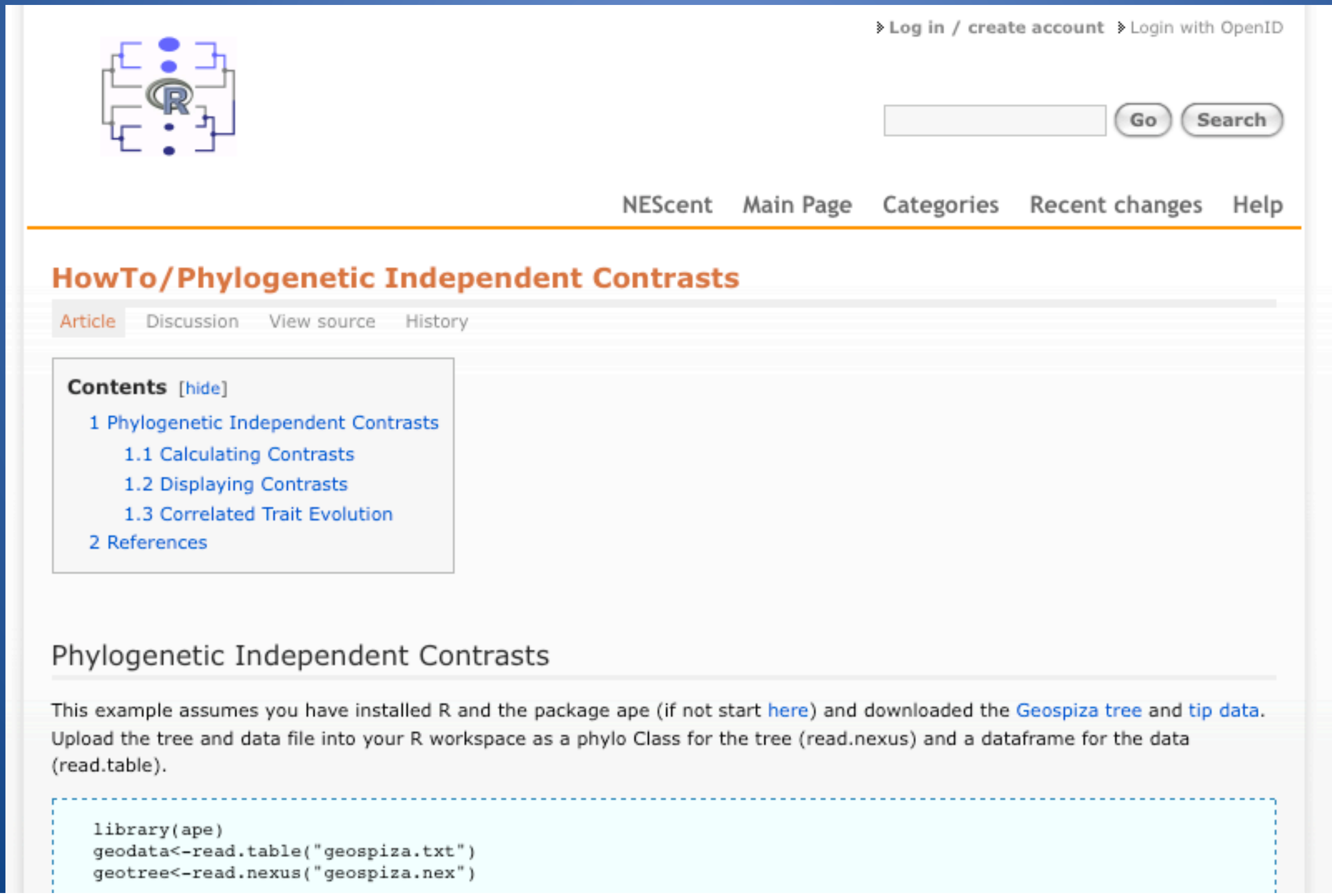


PDAP/PDTREEによる系統的独立比較

- 通常のデータ量では解析内容に特に問題ない
- 大量の形質データを扱うには不向き
- Rのパッケージ“ape”の利用
- apeはそのままではpositivized処理に非対応なので、必ずしもMesquiteの結果と一致しない

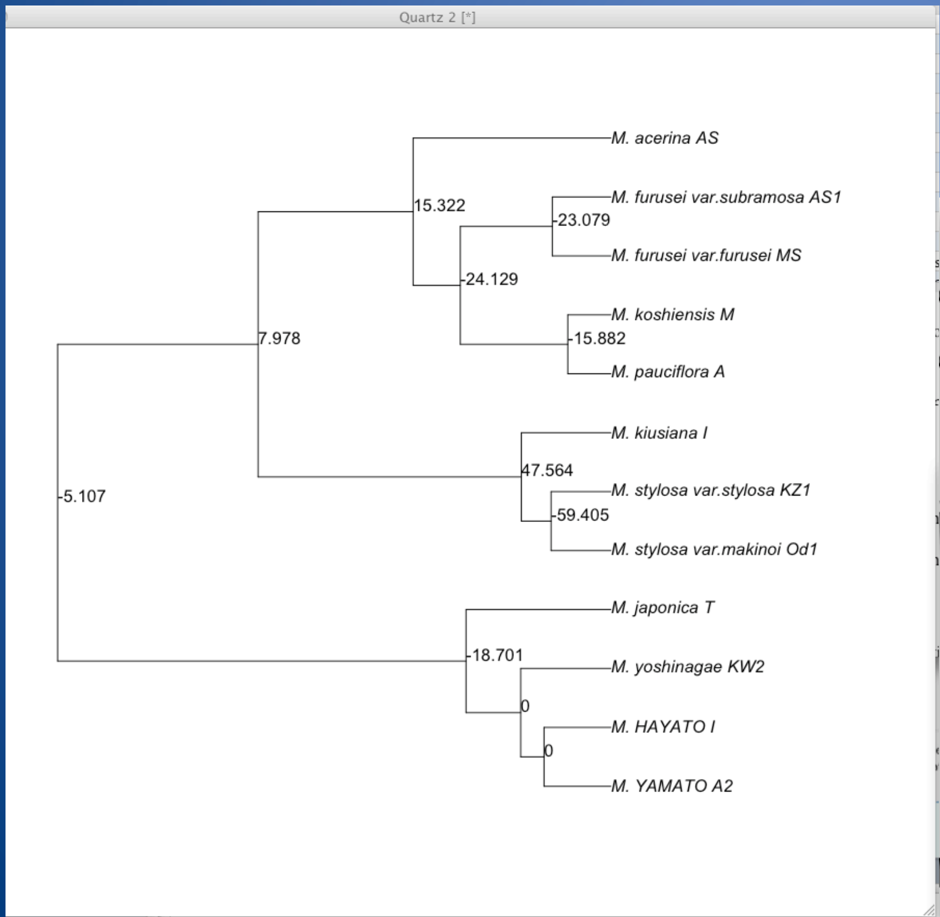
apeによる系統的独立比較

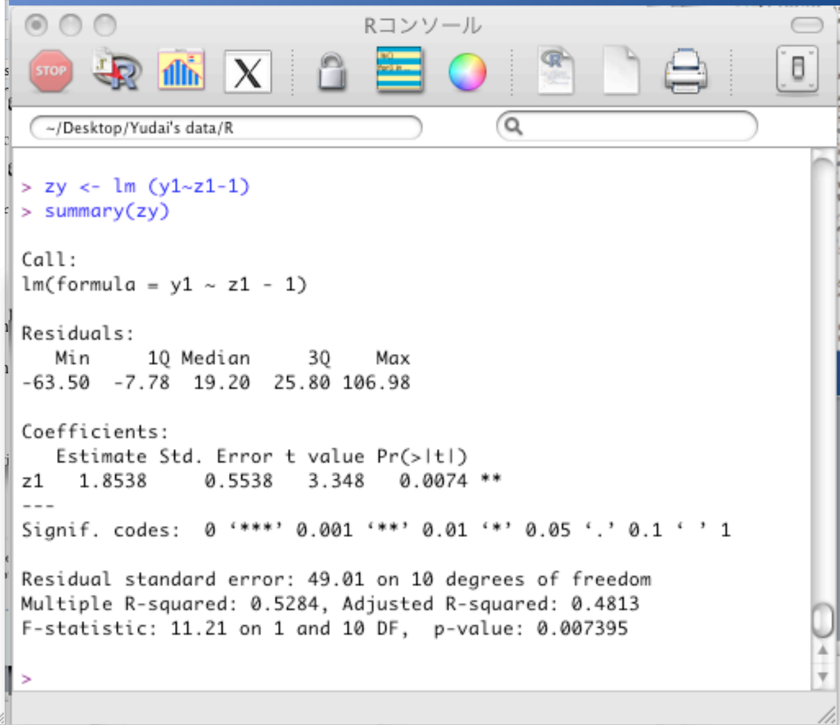
- Rをインストールした後、パッケージgee、apeの順でインストール
- http://www.r-phylo.org/wiki/HowTo/Phylogenetic _Independent_ Contrastsに詳しく解説あり



apeによる系統的独立比較

- 結果の出力が自在にできる
- 自動化による大量データの処理に向いている





大量データを用いた応用例：
遺伝子発現と表現型の間での連関解析

遺伝子のリスト(5000以上)

種のリスト

Tag sequence	*M. s. makinoi*	*M. s. makinoi*	*M. stylosa stylosa*	*M. f. subramosa*	*M. f. subramosa*	*M. furusei furusei*	*M. japonica*	*M. kiushiana*	*M. acerina*	*M. acerina*	*M. pauciflora*	*M. yoshinagae*	*M. y. hayato*	*M. j. yamato*
CATGTAATAAAATTACCTACCCATCT	1141	4875	1	0	3	10	4	1329	999	1131	14	1303	4	1
CATGACAACTCGCGGTGGAAGACCCA	3	0	2	313	251	55	1467	0	690	1135	904	233	2440	353
CATGTATCTGCTTCAGAGAAATACAG	472	703	772	262	508	448	60	473	224	269	1000	322	4	110
CATGGAACTGTTAAATAAATGCTTTC	233	493	283	718	485	227	4	608	0	0	833	1	4	0
CATGTAGGGGTAACTCCCCCATTTTC	95	135	4	72	246	356	143	7	114	18	44	723	508	1302
CATGACAACTCGTGGTGGAAGACCCA	860	358	931	2	0	0	4	1329	1	3	2	0	0	1
CATGTCAAGTCCTGTTTCTTGGGTTG	359	527	213	83	80	41	507	80	48	48	60	341	254	201
CATGGGCATATGCCACCAACTTTGTC	604	171	251	159	113	102	92	2	227	166	102	34	53	177
CATGTCCTCCGAGGTATTTTCCCGTG	410	306	483	152	0	174	4	7	2	9	257	1	44	19
CATGTGTGCGAGTCAACGGGCTAGTA	121	65	118	72	97	112	214	96	146	74	247	71	372	4
CATGACAACGTGGGTGCTGAAGTTGC	157	81	166	103	139	41	0	227	163	95	102	37	140	352
CATGGACCCGGTGAAATTGAATTCTC	229	254	118	378	115	51	22	43	44	33	114	102	263	28
CATGTATCTTTTTTTTTATCAGCAAA	18	520	38	32	0	4	94	46	63	215	29	47	679	5

.
.
.

×

物質のリスト(32種類)

種のリスト

	M. s. makinoi	*M. s. makinoi*	*M. stylosa stylosa*	*M. f. subramosa*	*M. f. subramosa*	*M. furusei furusei*	*M. japonica*	*M. kiushiana*	*M. acerina*	*M. acerina*	*M. pauciflora*	*M. yoshinagae*	*M. y. hayato*	*M. j. yamato*
	N=7	*N*=7	*N*=9	*N*=20	*N*=20	*N*=3	*N*=18	*N*=8	*N*=9	*N*=9	*N*=13	*N*=8	*N*=9	*N*=12
α-Pinene[b]	0.40	0.40	0.16	2.92	2.92	0.61	0.75	3.24	0.22	0.22	1.26	0.00	0.90	3.07
6-methyl-5-hepten-2-one[b]	8.63	8.63	0.93	4.78	4.78	4.95	6.33	5.16	15.05	15.05	12.66	6.38	15.30	12.08
Phellandrene	0.77	0.77	0.82	1.07	1.07	1.98	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(*E*)-β-ocimene	1.60	1.60	4.34	3.13	3.13	3.22	0.89	0.28	0.00	0.00	0.09	0.00	1.05	0.33
(*E*)-Linalool oxide (furanoid)[b]	1.86	1.86	0.95	1.77	1.77	0.56	0.00	1.02	0.61	0.61	0.39	0.72	0.00	0.78
(*Z*)-Linalool oxide (furanoid)[b]	1.04	1.04	0.28	0.92	0.92	0.16	0.00	1.22	0.14	0.14	0.07	0.00	0.00	0.12
Linalool[b]	19.45	19.45	11.83	7.71	7.71	5.05	37.06	63.95	40.44	40.44	48.86	81.73	13.48	8.48
Hotrienol	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.00	0.80	0.80	0.00	0.19	4.83	0.19
Lilac aldehyde B	10.28	10.28	35.48	22.52	22.52	12.83	11.11	0.34	0.07	0.07	0.00	0.00	0.00	0.00
Lilac aldehyde A	9.24	9.24	15.42	18.08	18.08	9.75	7.89	0.20	0.94	0.94	0.00	0.00	0.00	0.00

.
.
.

まとめ

- 正確な（分子）系統樹は、現生する生物の多様な形質を進化の軸でつなぎ、解析するために必須なツールである
- Mesquiteは祖先形質復元や系統的独立比較を手軽に行える便利なソフトウェアである
- 最尤法を用いた厳密な祖先形質復元にはPagelのBayesTraitsを用いるべきである
- 系統的独立比較において大量の形質データを処理する際には、Rのパッケージapeが利用できる
- 系統樹を考慮した比較解析は、近年発展が目覚ましい「オミクス」的研究にも応用できる