

# TreefinderとPhylogearsを用いたParallel Likelihood Ratchetによる最尤系統推定

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率
  - 10回のコイントスで表1回裏9回になった

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率
  - 10回のコイントスで表1回裏9回になった
    - 「コイントスでは表裏の出る比率は1:9」というモデルの尤度は $(1/10) \times (9/10)^9 = 0.03874205$

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率
  - 10回のコイントスで表1回裏9回になった
    - 「コイントスでは表裏の出る比率は1:9」というモデルの尤度は $(1/10) \times (9/10)^9 = 0.03874205$
    - 「コイントスでは表裏は同じ比率で出る」というモデルの尤度は $(1/2)^{10} = 0.0009765625$

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率
  - 10回のコイントスで表1回裏9回になった
    - 「コイントスでは表裏の出る比率は1:9」というモデルの尤度は $(1/10) \times (9/10)^9 = 0.03874205$
    - 「コイントスでは表裏は同じ比率で出る」というモデルの尤度は $(1/2)^{10} = 0.0009765625$
- モデルのデータへの当てはまりの良さを表す

# 尤度ってナニ?

- あるモデルが正しい状況で、手元のデータが実現する確率
  - 10回のコイントスで表1回裏9回になった
    - 「コイントスでは表裏の出る比率は1:9」というモデルの尤度は $(1/10) \times (9/10)^9 = 0.03874205$
    - 「コイントスでは表裏は同じ比率で出る」というモデルの尤度は $(1/2)^{10} = 0.0009765625$
- モデルのデータへの当てはまりの良さを表す
- コンピュータに扱いやすい対数尤度を用いる

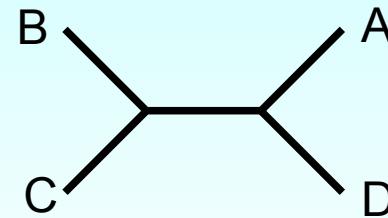
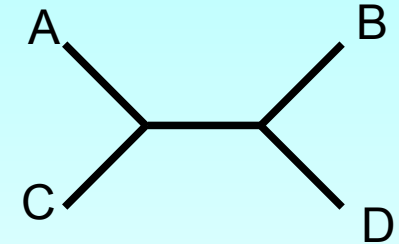
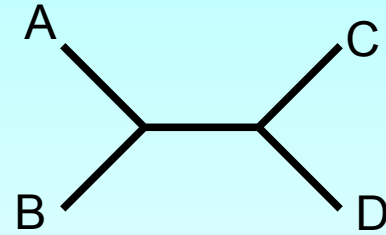
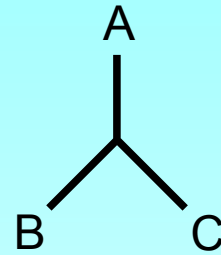
# 最尤系統推定におけるパラメータ

- 分子進化モデル
- 系統モデル
  - 枝長 (と枝長比)



# 最尤系統推定におけるパラメータ

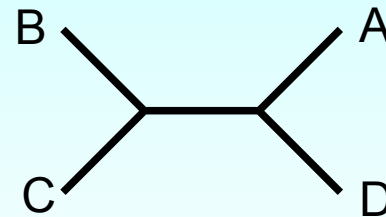
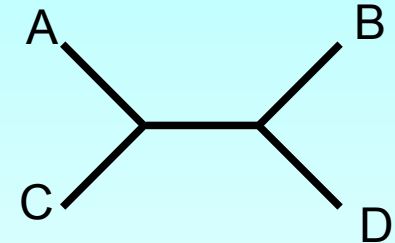
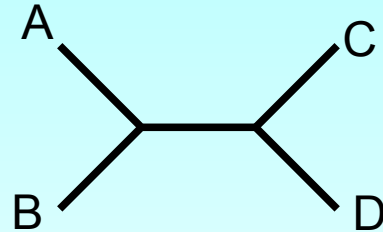
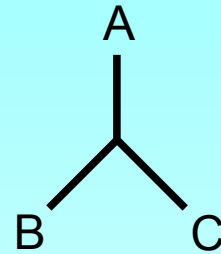
- 分子進化モデル
- 系統モデル
  - 枝長 (と枝長比)
  - 樹形
    - 30TU : 1通り
    - 40TU : 3通り
    - 50TU : 3×5通り
    - 60TU : 3×5×7通り
    - 1000TU :  $1700459 \times 10^{176}$



# 最尤系統推定におけるパラメータ

- 分子進化モデル
- 系統モデル
  - 枝長 (と枝長比)
  - 樹形

- 30TU : 1通り
- 40TU : 3通り
- 50TU : 3×5通り
- 60TU : 3×5×7通り
- 1000TU :  $1700459 \times 10^{176}$



尤度が最大になるこれらを決定する = 最尤系統推定

# 最尤系統推定における初期値

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 最尤系統推定における初期値

- 尤度最大化を始めるには良い初期値が必要

# 最尤系統推定における初期値

- 尤度最大化を始めるには良い初期値が必要
- 分子進化モデルの初期値は初期樹形が決まれば最節約的に決定可能

# 最尤系統推定における初期値

- 尤度最大化を始めるには良い初期値が必要
- 分子進化モデルの初期値は初期樹形が決まれば最節約的に決定可能
- 初期樹形をどのように与えるか?

# 最尤系統推定における初期値

- 尤度最大化を始めるには良い初期値が必要
- 分子進化モデルの初期値は初期樹形が決まれば最節約的に決定可能
- 初期樹形をどのように与えるか?
  - 近隣結合法
  - 段階的配列付加法

# 最尤系統推定における局所探索

樹形空間

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する


田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用



# 最尤系統推定における局所探索

樹形空間



初期樹形

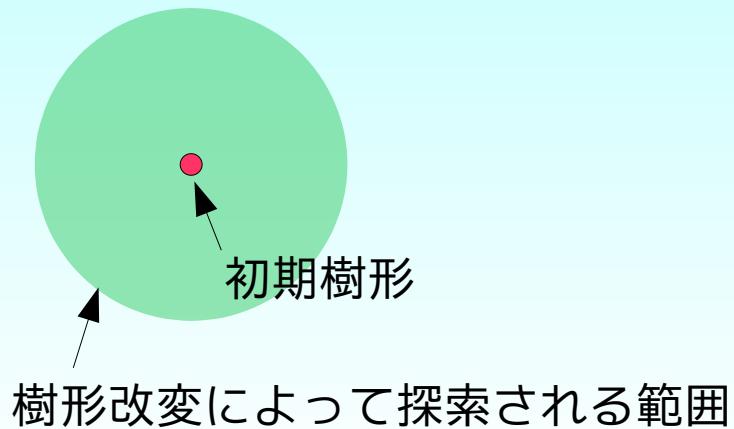
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 最尤系統推定における局所探索

## 樹形空間



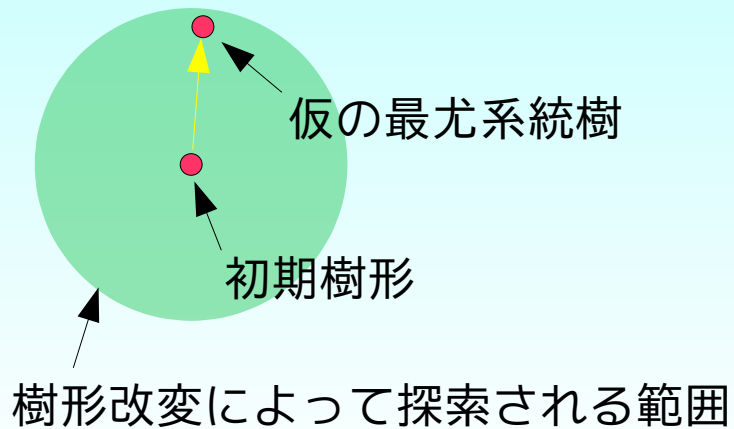
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

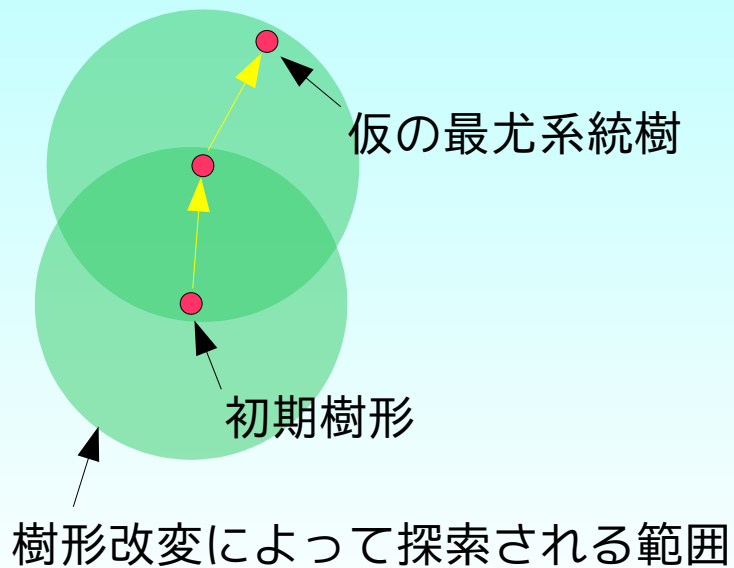
# 最尤系統推定における局所探索

## 樹形空間



# 最尤系統推定における局所探索

## 樹形空間



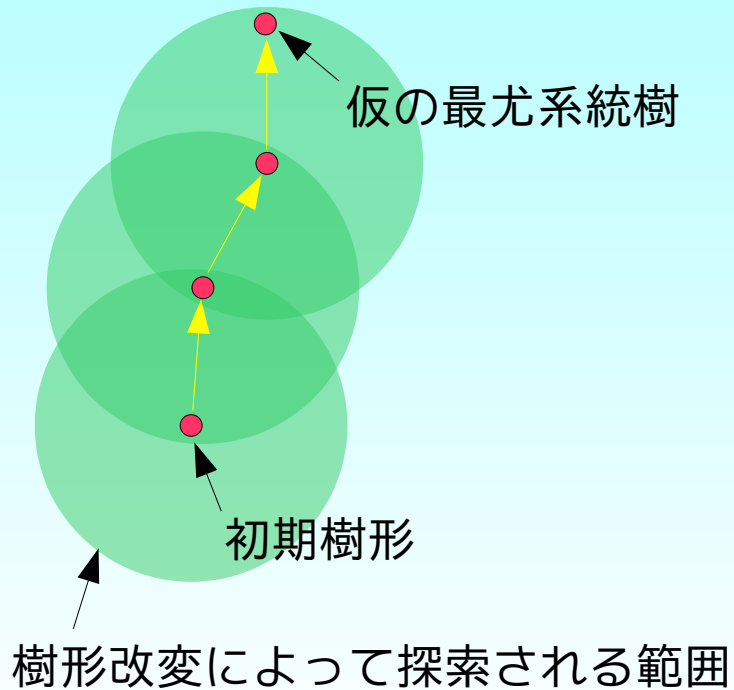
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 最尤系統推定における局所探索

## 樹形空間



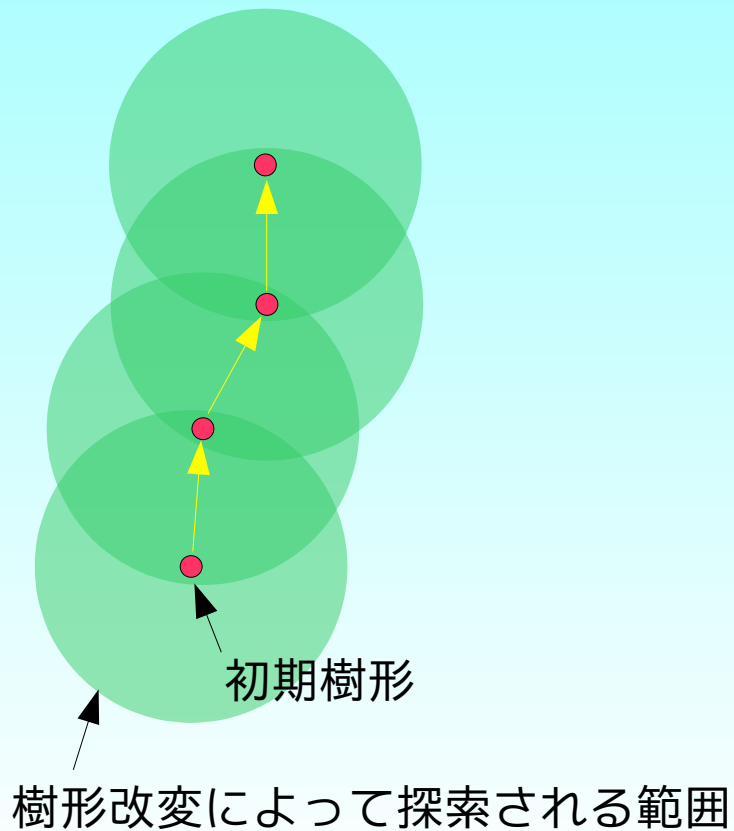
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 最尤系統推定における局所探索

## 樹形空間



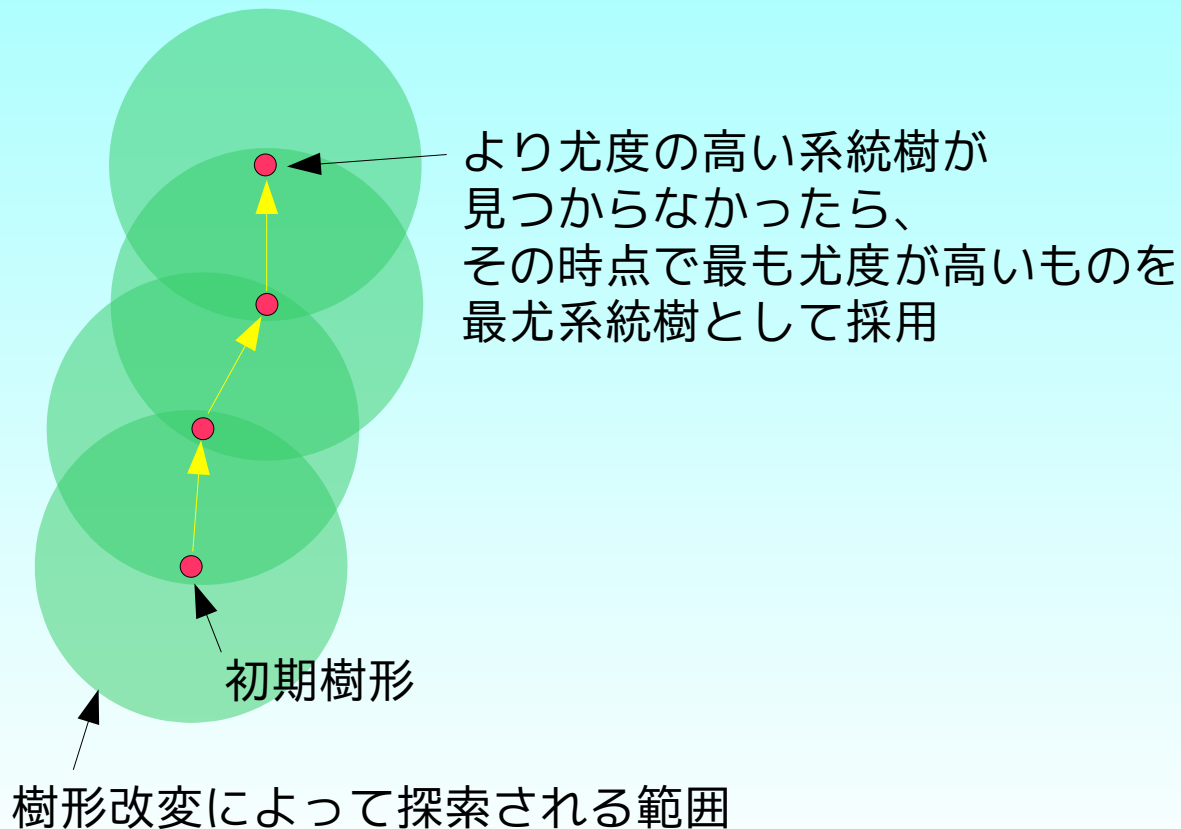
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

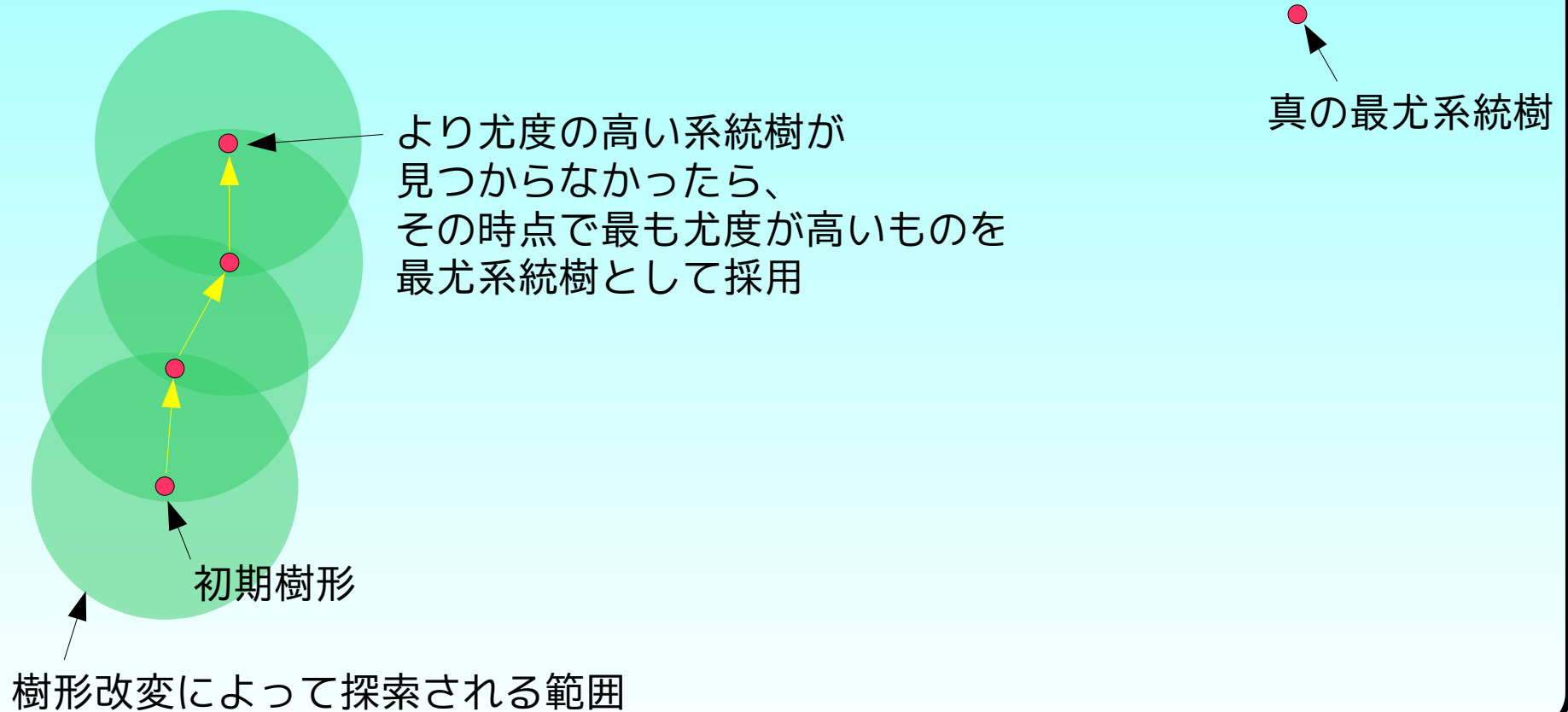
# 最尤系統推定における局所探索

## 樹形空間



# 最尤系統推定における局所探索

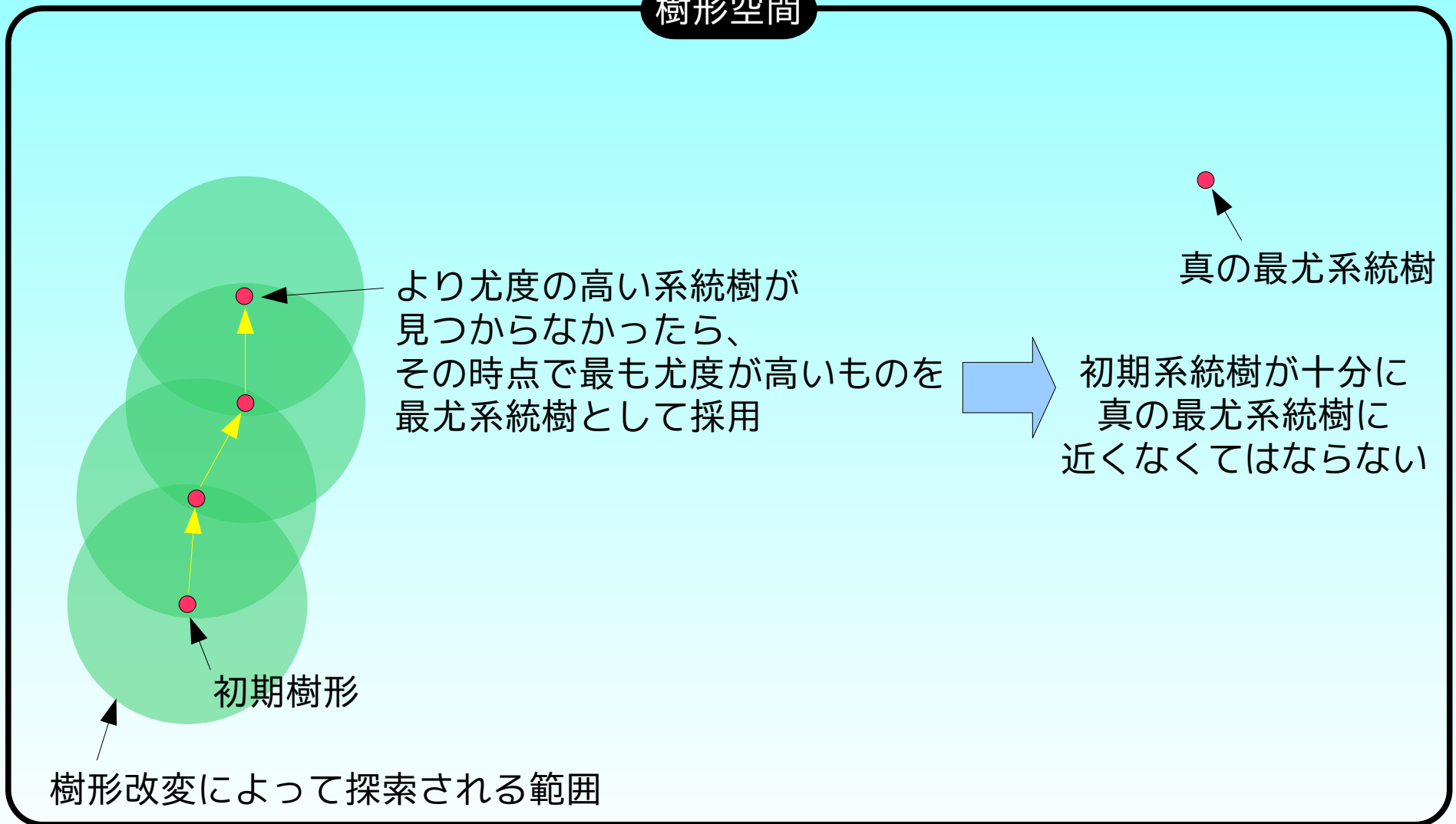
## 樹形空間





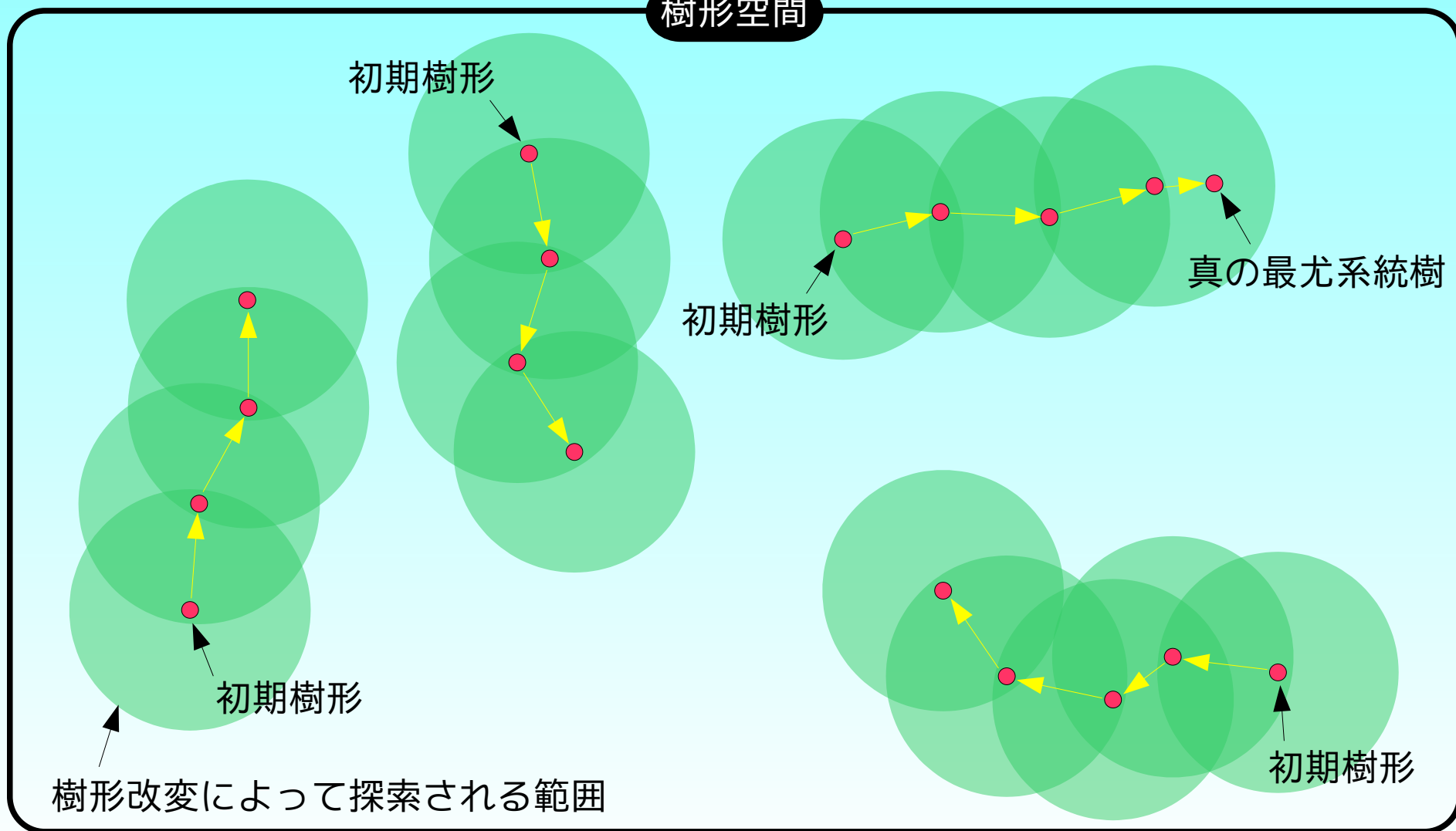
# 最尤系統推定における局所探索

## 樹形空間



# ショットガンの局的探索

樹形空間



進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

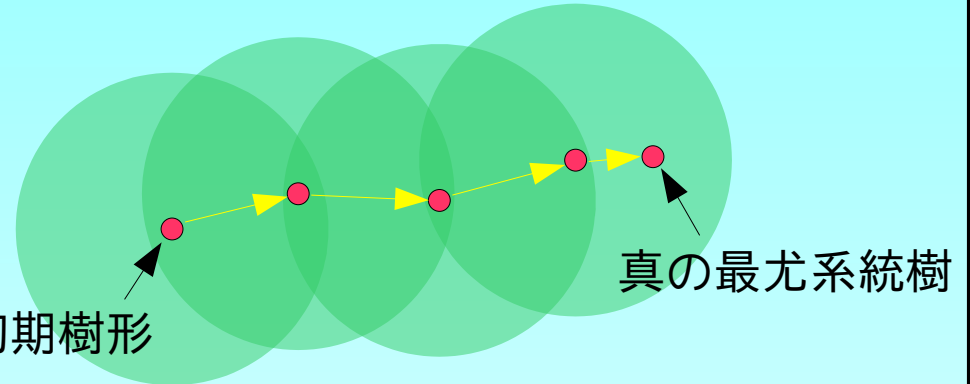
田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# ショットガンの局的探索

## 樹形空間

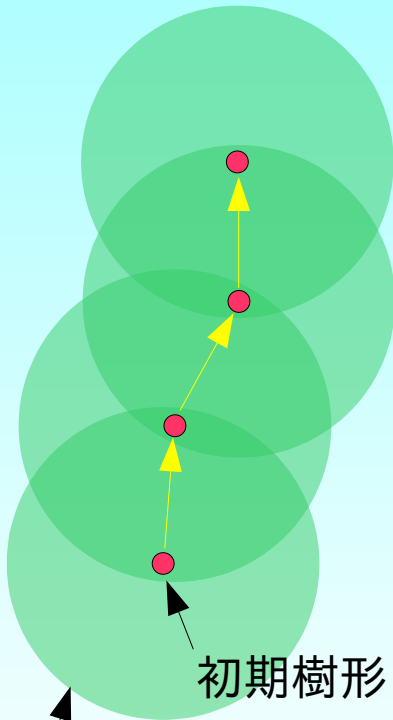
初期樹形



初期樹形

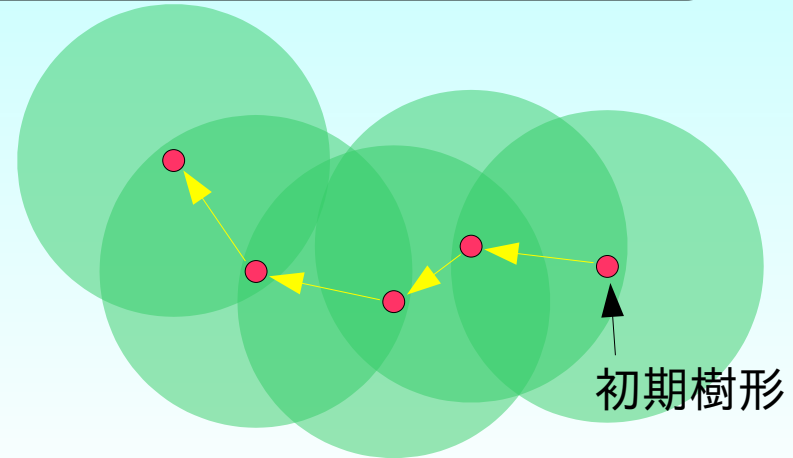
真の最尤系統樹

何らかの方法でばらつかせた初期樹形  
からの局的探索を繰り返す



初期樹形

樹形改変によって探索される範囲



初期樹形

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用



# Likelihood Ratchet

OTU1	TGTTT	...	TTTTTC
OTU2	AGTAC	...	TTTTTC
OTU3	AGTAT	...	TTGTC
⋮	⋮		⋮
⋮	⋮		⋮
⋮	⋮		⋮
OTUN	AGTAT	...	ATTTC

# Likelihood Ratchet

OTU1	TGTTT	...	TTTTTC
OTU2	AGTAC	...	TTTTTC
OTU3	AGTAT	...	TTGTC
⋮	⋮		⋮
OTUN	AGTAT	...	ATTTC
weight	11111	...	11111

無作為に重複を許してサイト数の20~25%加重

# Likelihood Ratchet

OTU1	TGTTT	...	TTTTTC
OTU2	AGTAC	...	TTTTTC
OTU3	AGTAT	...	TTGTC
⋮	⋮		⋮
OTUN	AGTAT	...	ATTTC
weight	12113	...	12112

無作為に重複を許してサイト数の20~25%加重

# Likelihood Ratchet

1. 無作為に重複を許してサイト数の20~25%を加重



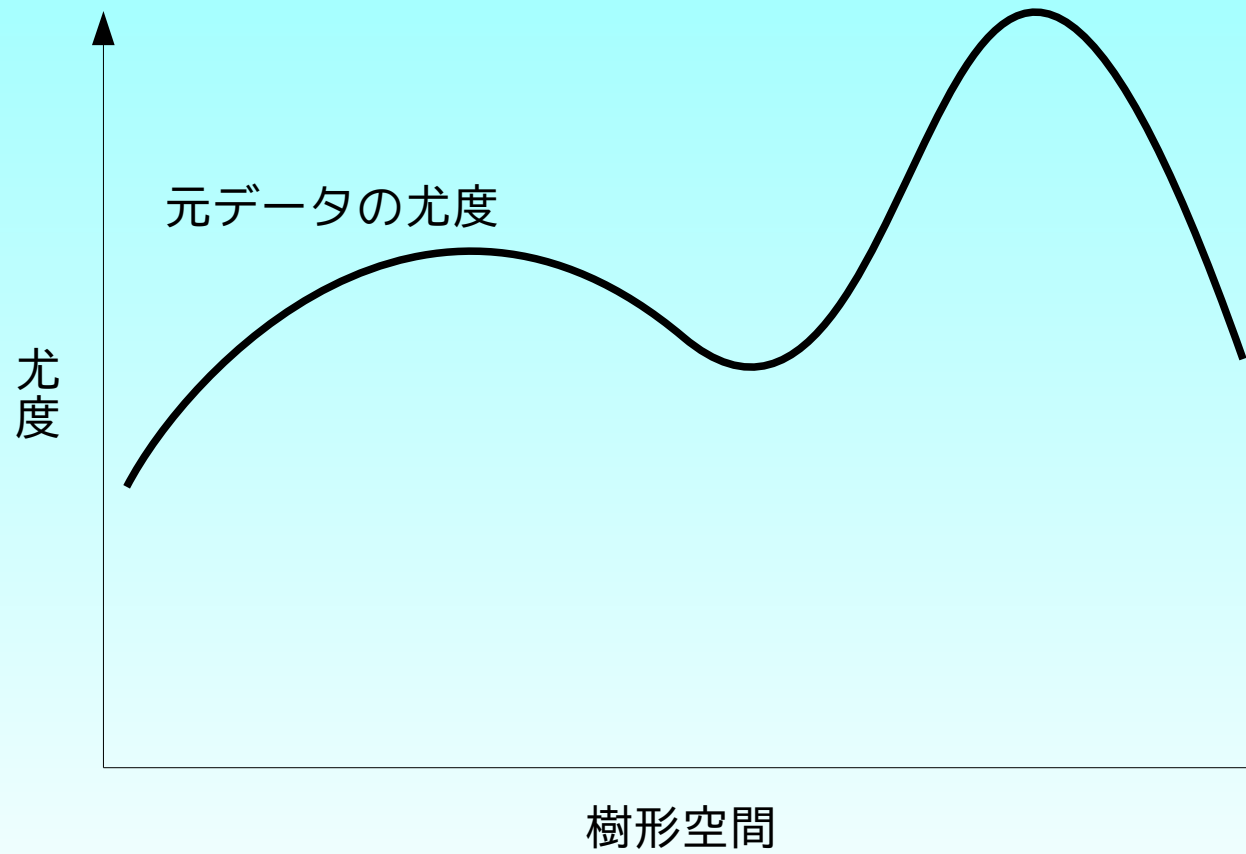
# Likelihood Ratchet

1. 無作為に重複を許してサイト数の20~25%を加重
2. 加重データで系統推定or初期樹形生成

# Likelihood Ratchet

1. 無作為に重複を許してサイト数の20~25%を加重
2. 加重データで系統推定or初期樹形生成
3. その結果を初期樹形として元データで系統推定

# Likelihood Ratchet

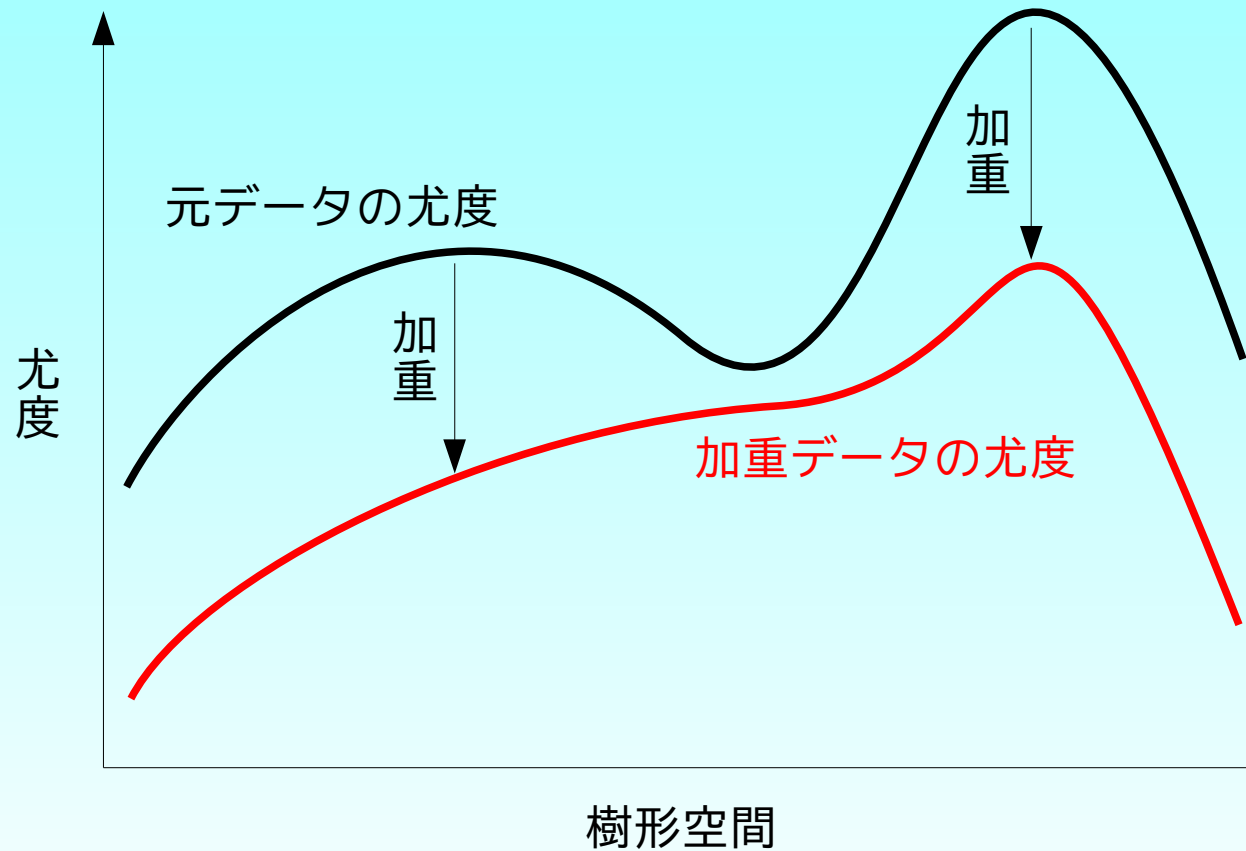


進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

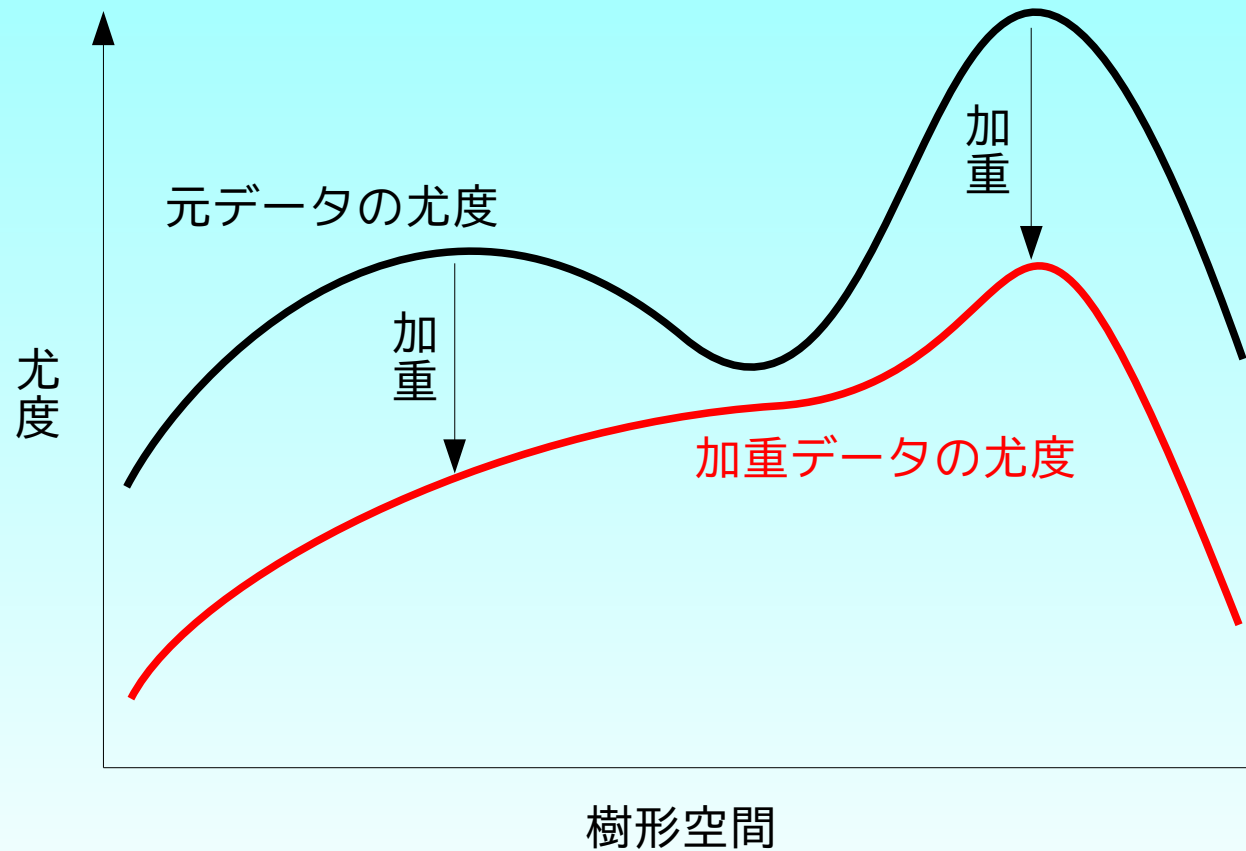
2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# Likelihood Ratchet



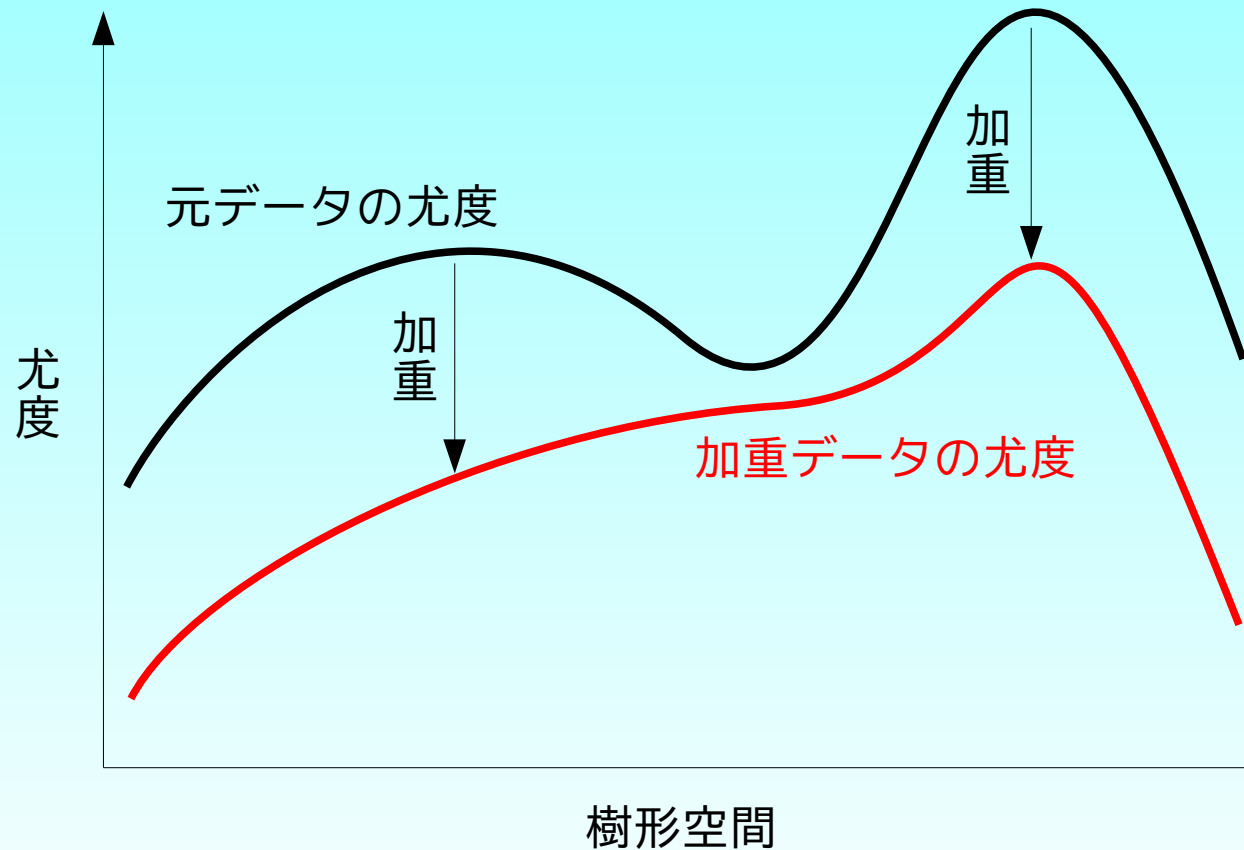
- 無作為加重により尤度のLandscapeが変化する

# Likelihood Ratchet



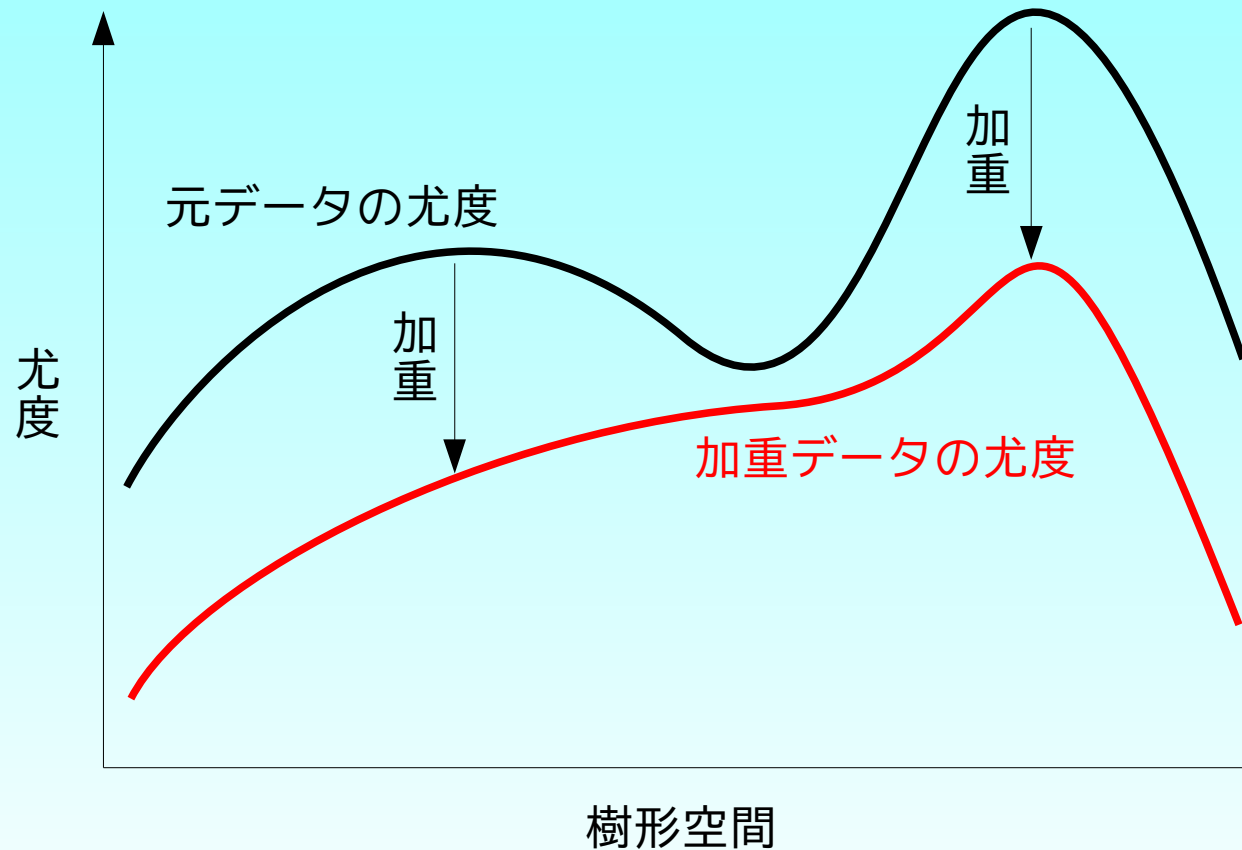
- 無作為加重により尤度のLandscapeが変化する
- Landscapeの変化によって、尤度の谷が無くなったり越えやすくなることもある

# Likelihood Ratchet



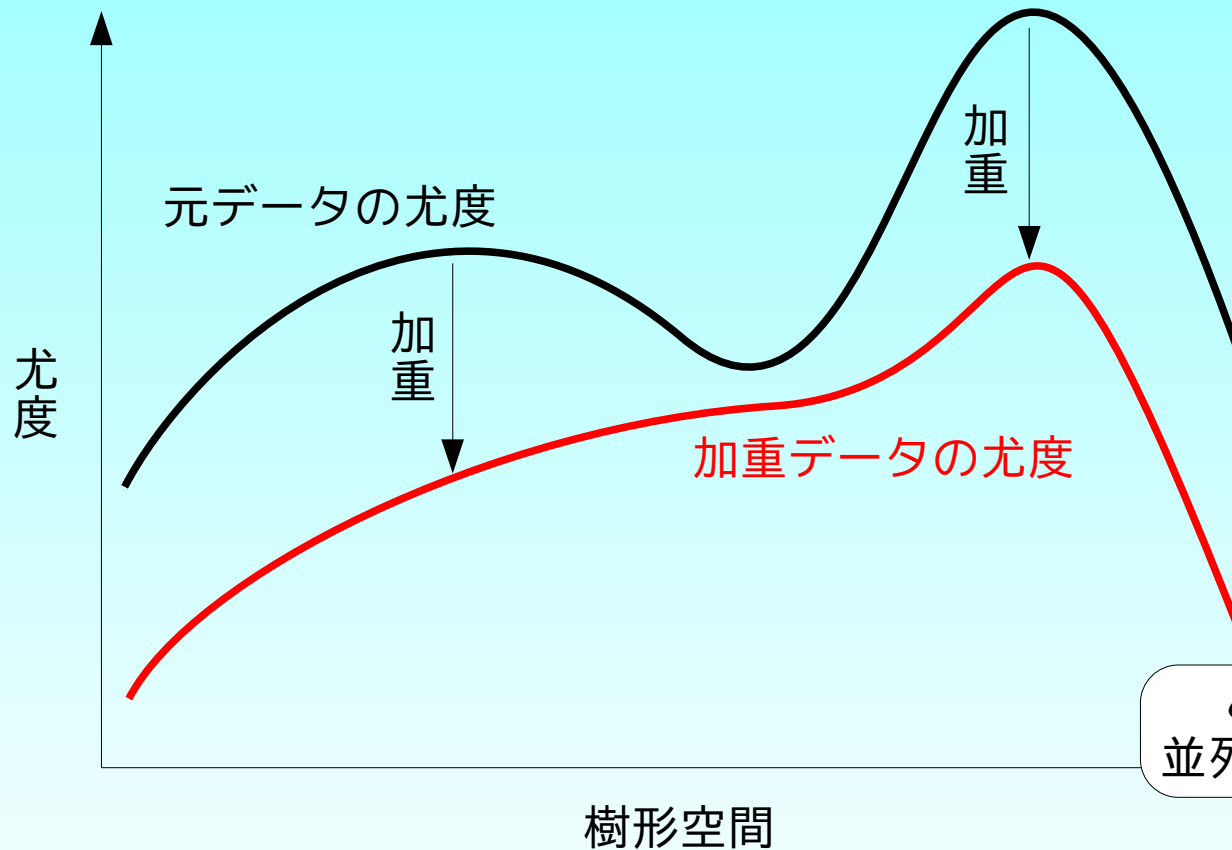
- 無作為加重により尤度の Landscape が変化する
- Landscape の変化によって、尤度の谷が無くなったり越えやすくなることもある
- 谷を越えた(かもしれない)結果を初期樹形として与えて元データの系統推定を行うことを繰り返す

# Likelihood Ratchet



- 無作為加重により尤度のLandscapeが変化する
- Landscapeの変化によって、尤度の谷が無くなったり越えやすくなることもある
- 谷を越えた(かもしれない)結果を初期樹形として与えて元データの系統推定を行うことを繰り返す
- これにより、より多くの尤度の山を把握する

# Likelihood Ratchet

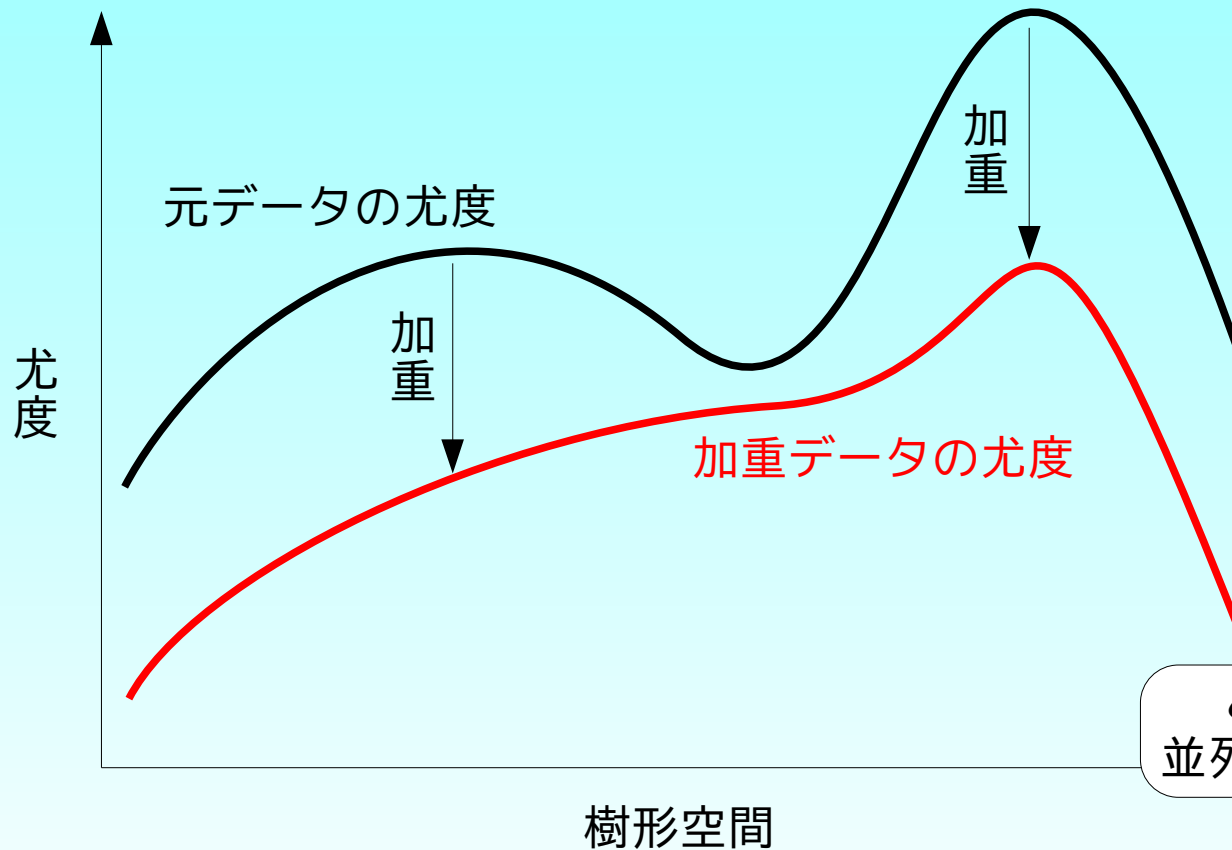


- 無作為加重により尤度のLandscapeが変化する
- Landscapeの変化によって、尤度の谷が無くなったり越えやすくなることもある
- 谷を越えた(かもしれない)結果を初期樹形として与えて元データの系統推定を行うことを繰り返す
- これにより、より多くの尤度の山を把握する

これまでのPAUP\*4+PAUPRatは  
並列処理や比例・分離モデルに非対応



# Likelihood Ratchet



- 無作為加重により尤度のLandscapeが変化する
- Landscapeの変化によって、尤度の谷が無くなったり越えやすくなることもある
- 谷を越えた(かもしれない)結果を初期樹形として与えて元データの系統推定を行うことを繰り返す
- これにより、より多くの尤度の山を把握する

これまでのPAUP\*4+PAUPRatは  
並列処理や比例・分離モデルに非対応

Treefinder+Phylogears

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# Phylogearsの概要

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# Phylogearsの概要

- Likelihood Ratchetの並列処理
  - 無作為加重データでの初期樹形生成を並列処理
  - 多数の初期樹形からの最尤系統樹探索を並列処理

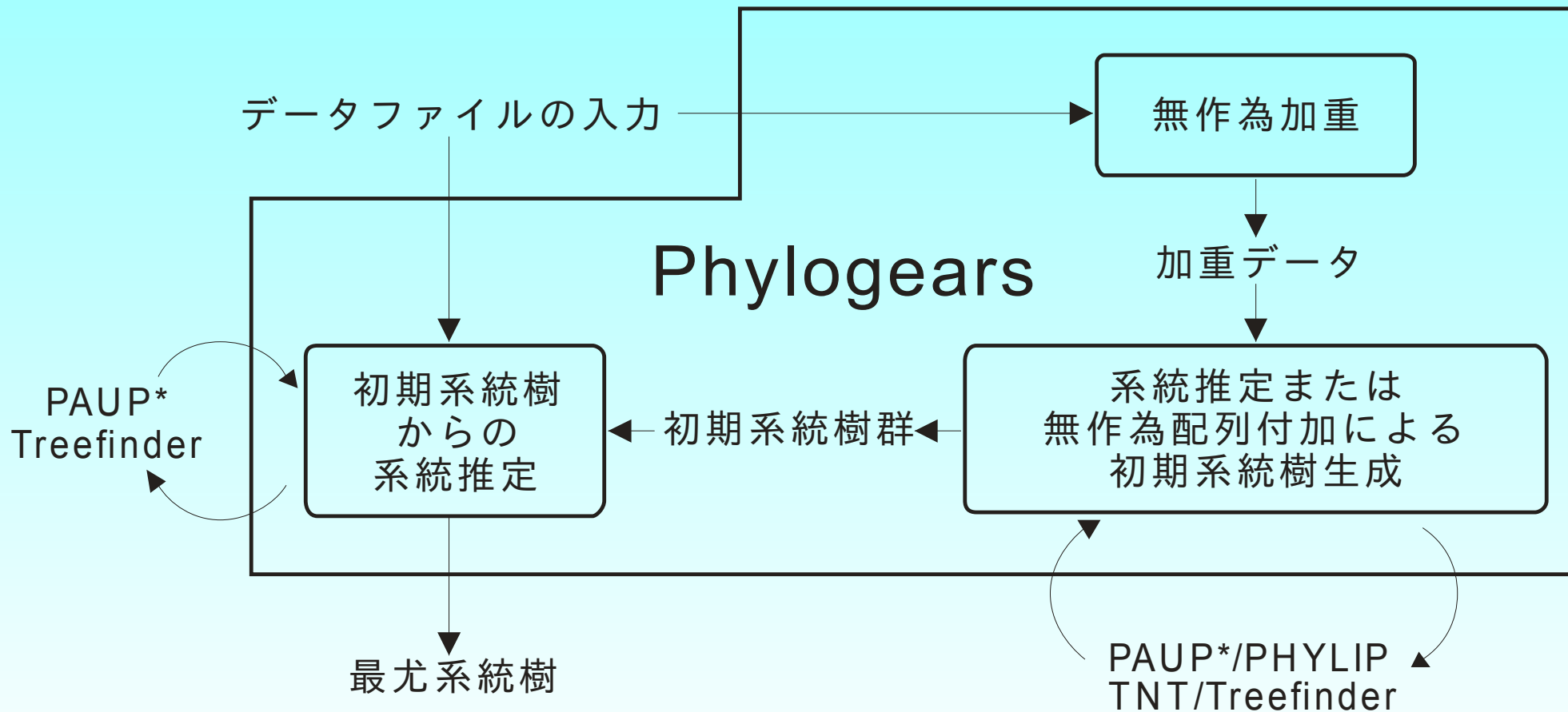
# Phylogearsの概要

- Likelihood Ratchetの並列処理
  - 無作為加重データでの初期樹形生成を並列処理
  - 多数の初期樹形からの最尤系統樹探索を並列処理
- ブートストラップ解析を並列処理
  - 多数のリサンプリングデータにおける最尤系統樹探索を並列処理

# Phylogearsの概要

- Likelihood Ratchetの並列処理
  - 無作為加重データでの初期樹形生成を並列処理
  - 多数の初期樹形からの最尤系統樹探索を並列処理
- ブートストラップ解析を並列処理
  - 多数のリサンプリングデータにおける最尤系統樹探索を並列処理
- その他色々

# Phylogearsが中でやっていること



進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 一括処理ファイルのコピー

ファイル(E) 編集(E) 表示(V) ツール(I) ヘルプ(H)

整理 表示 書き込む

フォルダ

- phylogears-1.0.2009.03.07\_for\_Windows
  - bin
  - share
  - phylogears
    - PAUP
    - PHYLIP
    - TE
      - AIC**
        - codonproportional
        - codonseparate
        - codonshared
        - proportional
        - proportional\_codonproportional**
        - proportional\_codonshared
        - separate
        - separate\_codonproportional
        - separate\_codonseparate
        - separate\_codonshared
        - single
        - AICc1

名前	更新日時	種類	サイズ	属性
whole_AIC_proportional_co...		Windows パッチ ファイル	whole_AIC_proportional_co...	TL ファイル
whole_AIC_proportional_co...		TL ファイル	whole_AIC_proportional_co...	NEX ファイル
whole_AIC_proportional_co...		TL ファイル	whole_AIC_proportional_co...	Windows パッチ ファイル
whole_AIC_proportional_co...		Windows パッチ ファイル	whole_AIC_proportional_co...	TL ファイル
whole_AIC_proportional_co...		TL ファイル	whole_AIC_proportional_co...	

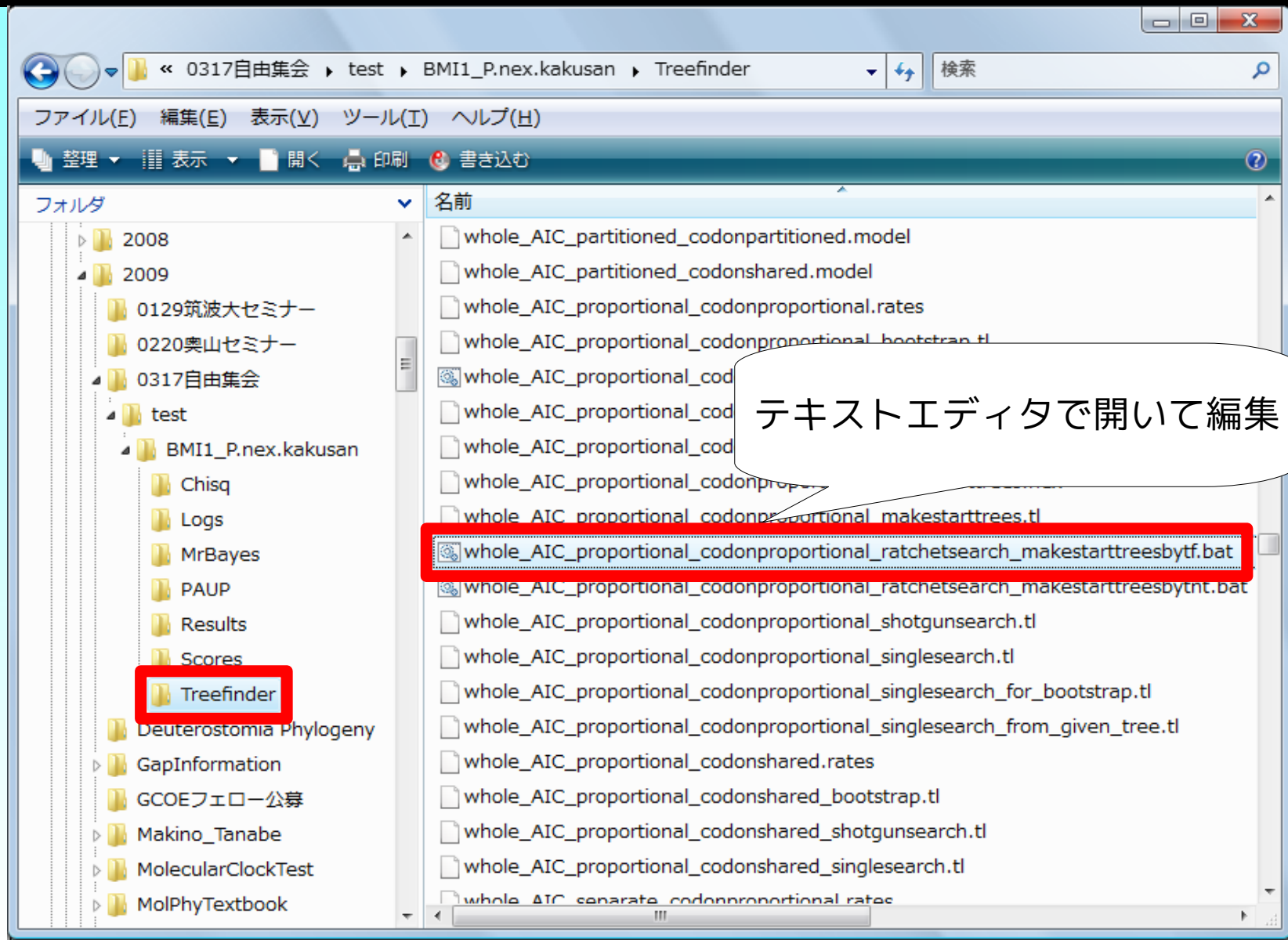
→ Kakusan3の出力したTreefinderフォルダへコピー

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 一括処理ファイルの編集



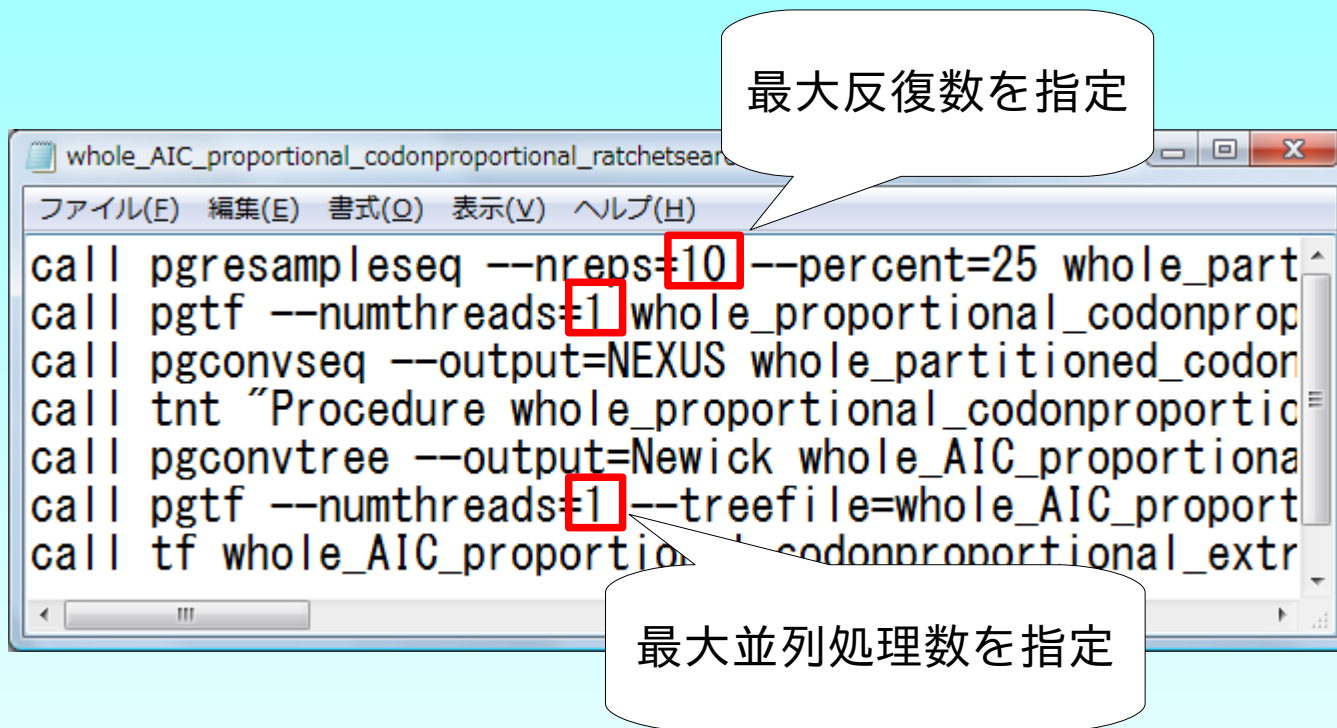
進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

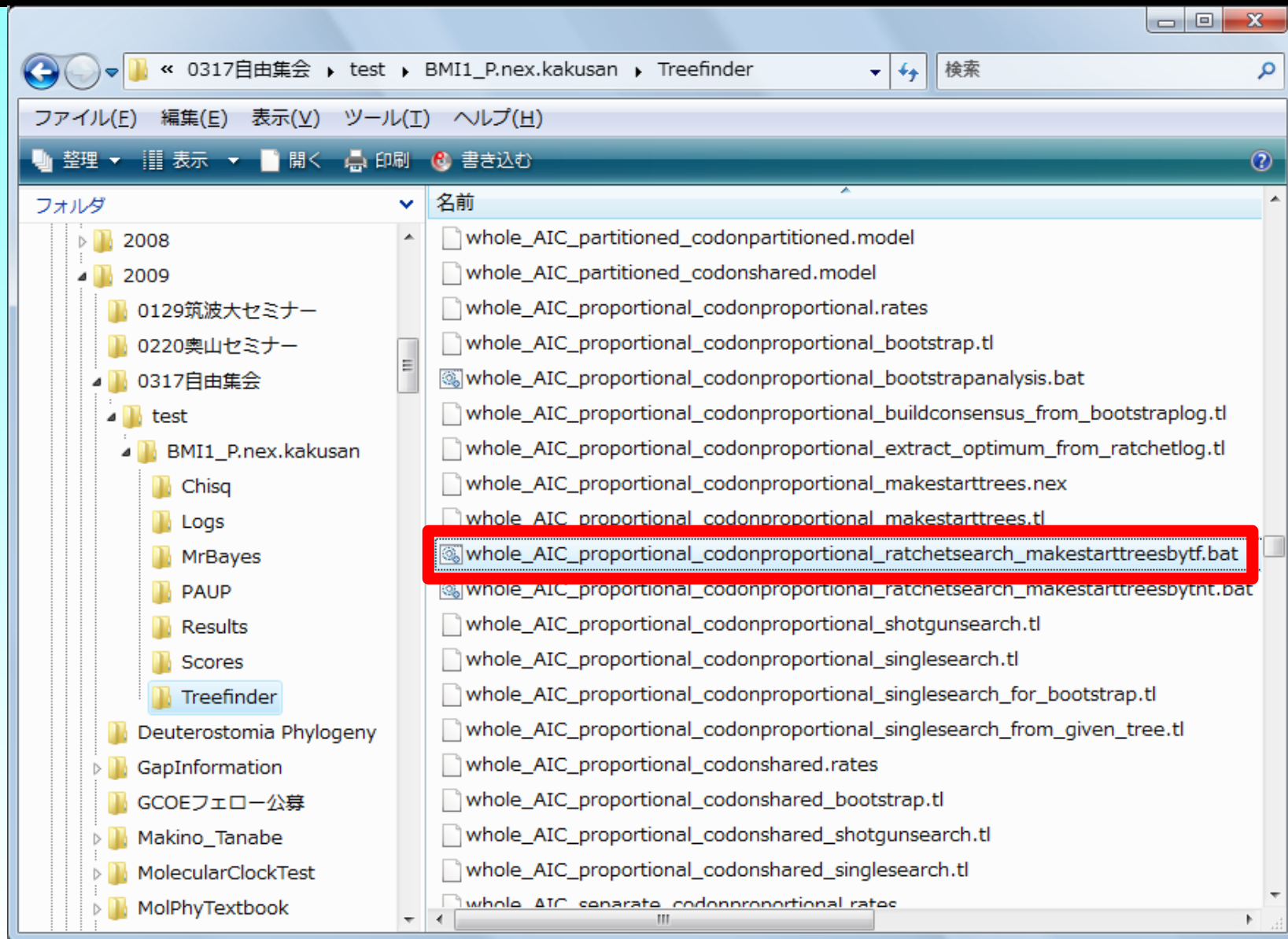
2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用



# 一括処理ファイルの編集



# 一括処理ファイルの実行



進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 実行画面

```
C:\Windows\system32\cmd.exe

pgtf 1.0.2009.03.07
=====

Official web site of this script is
http://www.fifthdimension.jp/products/phylogears/ .
To know script details, see above URL.

Copyright (C) 2008 Akifumi S. Tanabe

This program is free software; you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation; either version 2 of the License.

This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License along
with this program; if not, write to the Free Software Foundation, Inc.,
51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA.

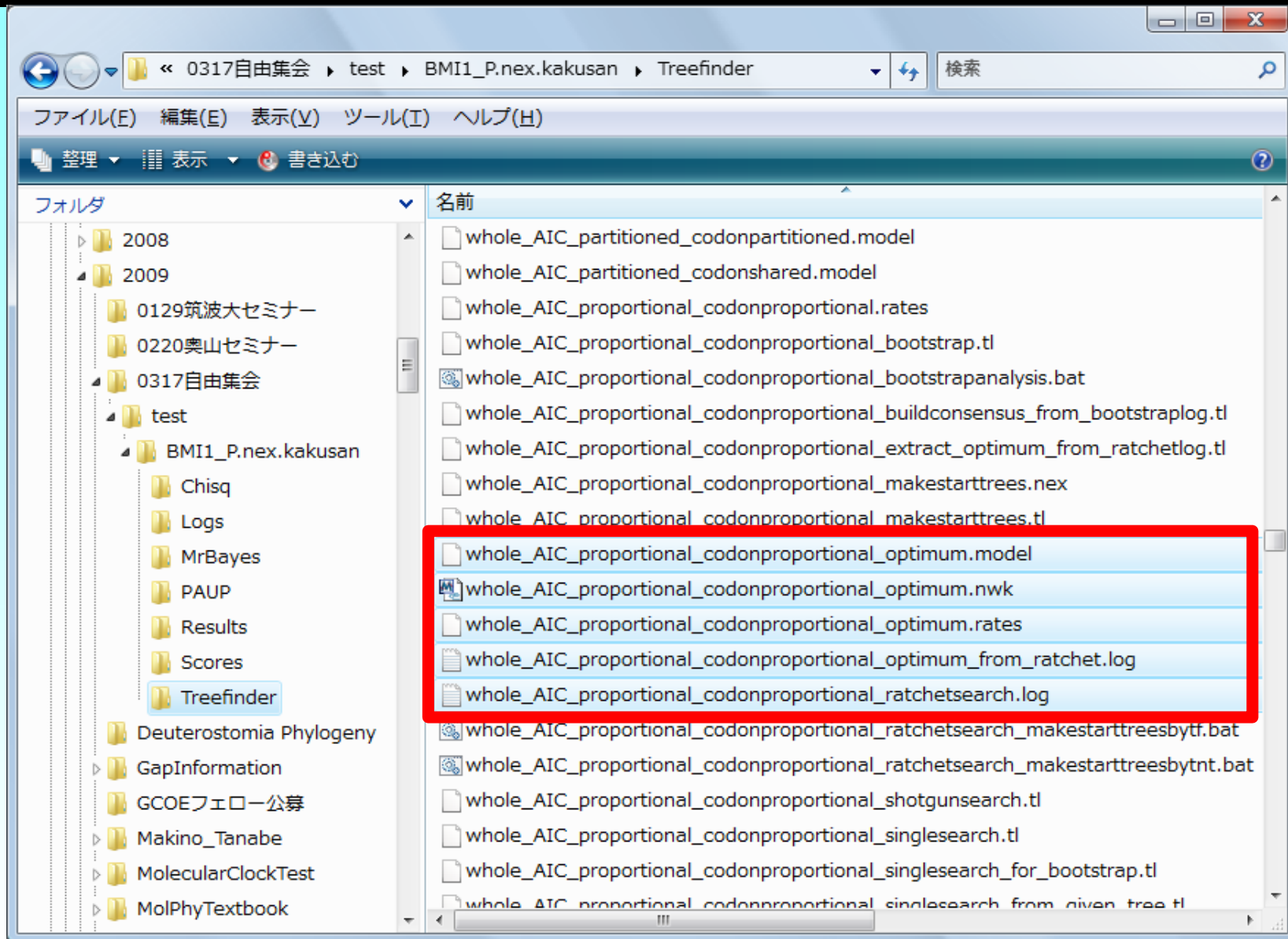
processing replicate 1...
```

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 完了後に出力されるファイル群

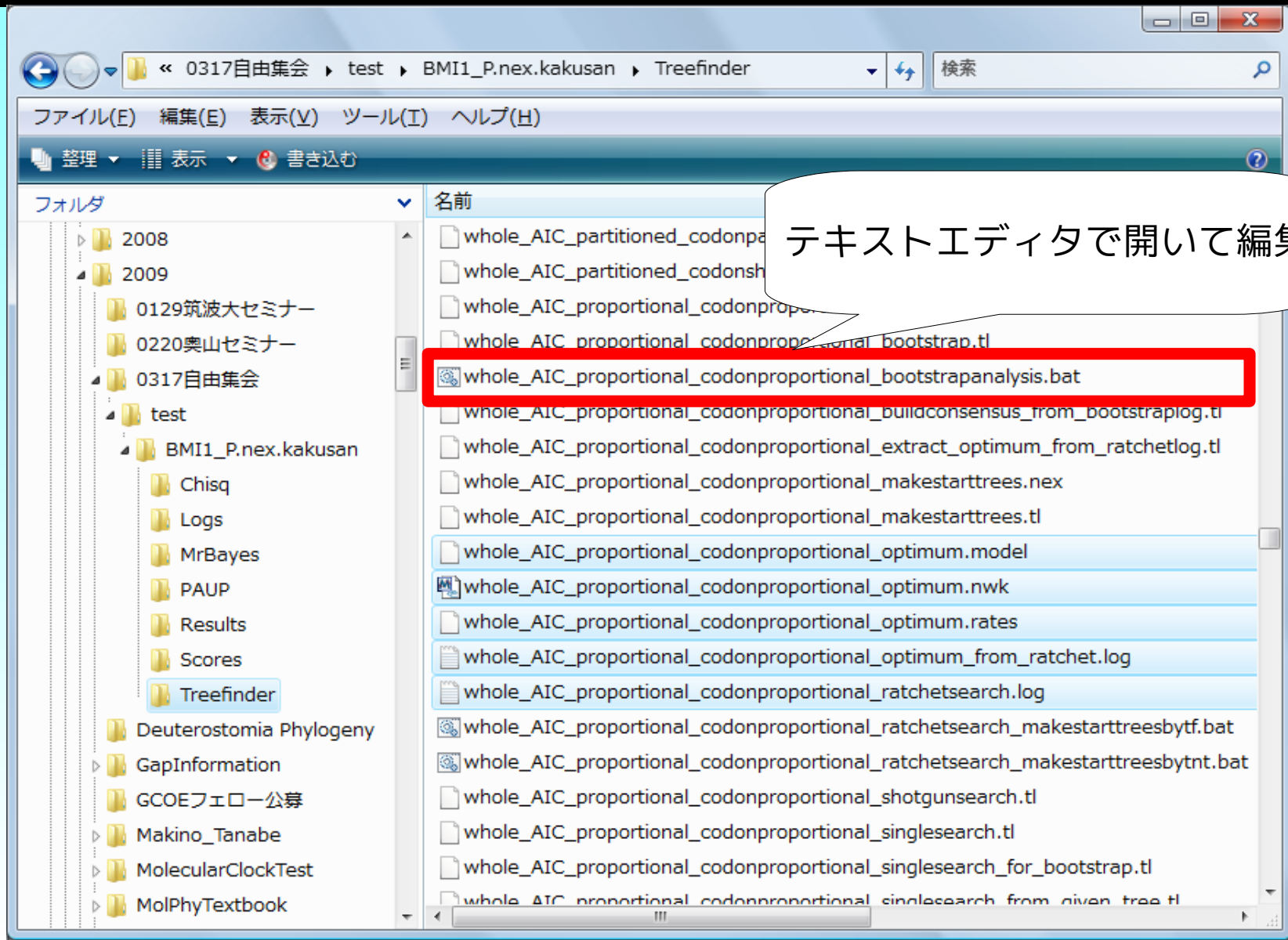


進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 一括処理ファイルの編集 ブートストラップ解析



進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# 一括処理ファイルの編集 ブートストラップ解析

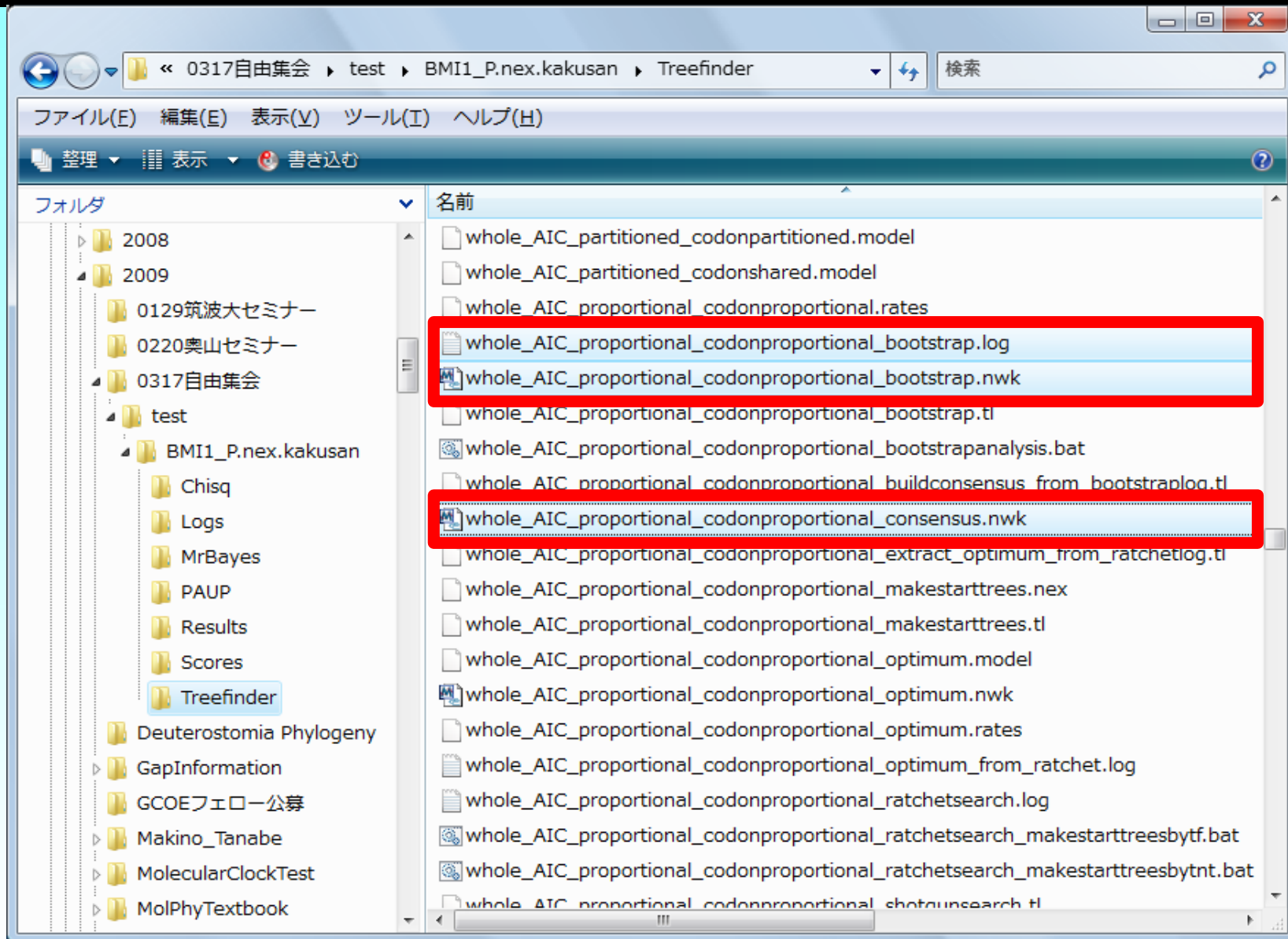
The screenshot shows a text editor window with the following content:

```
whole_AIC_proportional_codonproportional_bootstrapan
ファイル(E) 編集(E) 書式(O) 表示(V) ヘルプ(H)
call pgresampleseq --nreps=10 whole_partitioned_codon
call pgtf --numthreads=1 whole_proportional_codonprop
call tf whole_AIC_proportional_codonproportional_buil
```

Callout 1: 最大反復数を指定 (Maximum number of repetitions specified) - points to the value 10 in the first command.

Callout 2: 最大並列処理数を指定 (Maximum number of parallel processes specified) - points to the value 1 in the second command.

# 完了後に出力されるファイル群 ブートストラップ解析



進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用

# まとめ

進化モデル選択とLikelihood Ratchet、系統樹から進化速度の変化を検出する

田辺 晶史 (東北大・院・生命科学)

2009年生態学会自由集会：進化生態学のための分子系統樹の推定と応用



# まとめ

- OTUの多いデータでは樹形空間は広大になってしまう

# まとめ

- OTUの多いデータでは樹形空間は広大になってしまう
- そんなデータでも効率的に樹形空間内を探索する方法が Likelihood Ratchet

# まとめ

- OTUの多いデータでは樹形空間は広大になってしまう
- そんなデータでも効率的に樹形空間内を探索する方法が Likelihood Ratchet
- Likelihood Ratchetは、尤度のLandscapeを人為的に改変することで尤度の谷を越えやすくする

# まとめ

- OTUの多いデータでは樹形空間は広大になってしまう
- そんなデータでも効率的に樹形空間内を探索する方法がLikelihood Ratchet
- Likelihood Ratchetは、尤度のLandscapeを人為的に改変することで尤度の谷を越えやすくする
- TreefinderとPhylogearsを組み合わせることで、この樹形探索を並列処理することが可能

# まとめ

- OTUの多いデータでは樹形空間は広大になってしまう
- そんなデータでも効率的に樹形空間内を探索する方法がLikelihood Ratchet
- Likelihood Ratchetは、尤度のLandscapeを人為的に改変することで尤度の谷を越えやすくする
- TreefinderとPhylogearsを組み合わせることで、この樹形探索を並列処理することが可能
- ブートストラップ解析も同様に並列処理が可能